

Weihua Han  
Zhiming M. Wang *Editors*

# Toward Quantum FinFET

# Lecture Notes in Nanoscale Science and Technology

## Volume 17

### *Series Editors:*

Zhiming M. Wang  
University of Electronic Science and Technology of China, Chengdu, China

Andreas Waag  
Institut für Halbleitertechnik, TU Braunschweig, Braunschweig, Germany

Gregory Salamo  
Department of Physics, University of Arkansas, Fayetteville, AR, USA

Naoki Kishimoto  
Quantum Beam Center, National Institute for Materials Science, Tsukuba,  
Ibaraki, Japan

Stefano Bellucci  
Laboratori Nazionali di Frascati, Istituto Nazionale di Fisica Nucleare,  
Frascati, Italy

Young June Park  
School of Electrical Engineering, Seoul National University, Shinlim Dong,  
Kwanak-Gu, Seoul, Korea

For further volumes:  
<http://www.springer.com/series/7544>



Weihua Han • Zhiming M. Wang  
Editors

# Toward Quantum FinFET



Springer

*Editors*

Weihua Han  
Engineering Research  
Center for Semiconductor  
Integrated Technology  
Chinese Academy of Sciences  
Beijing, China, People's Republic

Zhiming M. Wang  
Engineering Research  
Center for Semiconductor  
Integrated Technology  
Chinese Academy of Sciences  
Beijing, China, People's Republic

ISSN 2195-2159

ISBN 978-3-319-02020-4

DOI 10.1007/978-3-319-02021-1

Springer Cham Heidelberg New York Dordrecht London

ISSN 2195-2167 (electronic)

ISBN 978-3-319-02021-1 (eBook)

Library of Congress Control Number: 2013953547

© Springer International Publishing Switzerland 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Preface

The rapid development of micro–nano fabrication is driving the MOSFET downscaling trend that evolves from planar channel to nonplanar FinFET. However, silicon-based CMOS technology is expected to face fundamental limits in the near future, and therefore, new types of nanoscale devices for the quantum effect of the carrier transport are being investigated aggressively. The main objective of this book is to create a platform for knowledge sharing and dissemination of the latest advances in novel areas of the quantum FinFET and to provide a comprehensive introduction to the field and directions for further research.

“Toward Quantum FinFET” reviews a range of quantum phenomena in FinFET, including quantized conductance of 1D transport, single electron effect, tunneling transport, and so on. Therefore, the book is organized as follows. Chapter 1 provides the basic tools involved in the nonequilibrium Green’s function formalism (NEGF) simulation of the quantum ballistic transport in FinFETs. The NEGF provides a rigorous description of quantum transport in nanoscale devices. Chapters 2 and 3 analyze the electron mobility in the FinFET structure. Chapter 2 focuses on the fundamentals of this mobility in the framework of the SNM (simple, novel, malleable) model for quantum-confined nanowires. Chapter 3 systematically investigates the impact of the surface orientation, strain, fin doping, and gate stack on SOI double-gate FinFET mobility.

Chapters 4–8 cover various phenomena with an influence on the characteristics of FinFET. Chapter 4 focuses on full Two-Dimensional (2D) Quantum Mechanical (QM) analytical modeling in order to evaluate the 2D potential profile within the active area of FinFET structure. This approach is applied to a detailed study of the threshold voltage and its variation with the process parameters. Chapter 5 analyzes the impact of line-edge roughness (LER) and work-function variations (WFV) on FinFET electrical performance through extensive Monte Carlo (MC) ensemble simulations compared with simplified models for variability estimation. Chapter 6 simulates a 16 nm multi-fin FinFET device and its circuit characteristics by solving a set of 3D quantum-mechanically corrected transport equations coupled with circuit nodal equations self-consistently. Chapter 7 provides an introductory overview of process variability in modern fabrication and investigates in detail

how LER and RDF affect inversion-mode (IM) and junctionless (JL) FinFET performance variability for sub-32 nm technology nodes. Chapter 8 mainly discusses the impacts of nonnegligible quantum confinement in multi-gate devices on random telegraph noise (RTN) characteristics. Chapter 9 reports transport properties of poly-Si nanowire transistors which are fabricated by a simple and low-cost method.

Chapters 10 and 11 explore FinFET technology in future applications. Chapter 10 highlights the importance of the drain extended class of high voltage MOS devices for advanced implementations of System on Chip (SoC) applications using FinFET or other tri-gate technologies. Chapter 11 reports a compact model for FinFETs with double-gate configuration and a physics-based hot carrier effect model for prediction of FinFETs performance degradation due to the interface state.

Chapters 12–15 talk about single-electron transistors from silicon to graphene and their applications. Chapter 12 reports a successful implementation of a CMOS-compatible room-temperature single-electron transistor by ultrascaling a FinFET structure down to an ultimate limiting form, resulting in the reliable formation of a sub-5 nm silicon Coulomb island. Chapter 13 focuses on transport characteristics arising from single-electron tunneling via individual dopant atoms, the basic operation mode of single-dopant transistors. Chapter 14 talks about all kinds of graphene-based quantum dot devices, including single dot, single dot with integrated single electron transistor (SET) charge detector, double dot in series, and double dot in parallel, and investigates the properties of devices by doing the low temperature quantum transport measurement. Chapter 15 analyzes the terahertz response in Schottky warp-gate controlled single electron transistors.

The Editors are extremely grateful to all of the chapter authors for their great efforts and outstanding chapters. It has taken us nearly one year to complete the manuscript of the book, but we are confident that the result is a fundamental bridge between quantum FinFET and nanotechnology that can stimulate readers' interest in developing new types of nanoscale transistors for semiconductor technology.

Beijing, China, People's Republic  
Beijing, China, People's Republic

Weihua Han  
Zhiming M. Wang

# Contents

<b>1 Simulation of Quantum Ballistic Transport in FinFETs . . . . .</b>	1
Yasser M. Sabry, Mohammed M. El-Banna, Tarek M. Abdolkader, and Wael Fikry	
<b>2 Model for Quantum Confinement in Nanowires and the Application of This Model to the Study of Carrier Mobility in Nanowire FinFETs . . . . .</b>	25
Arif Khan, Saeed Ganji, and S. Noor Mohammad	
<b>3 Understanding the FinFET Mobility by Systematic Experiments . . . . .</b>	55
Kerem Akarvardar, Chadwin D. Young, Mehmet O. Baykan, and Christopher C. Hobbs	
<b>4 Quantum Mechanical Potential Modeling of FinFET . . . . .</b>	81
Balwinder Raj	
<b>5 Physical Insight and Correlation Analysis of Finshape Fluctuations and Work-Function Variability in FinFET Devices . . . . .</b>	99
Emanuele Baravelli	
<b>6 Characteristic and Fluctuation of Multi-fin FinFETs . . . . .</b>	125
Hui-Wen Cheng and Yiming Li	
<b>7 Variability in Nanoscale FinFET Technologies . . . . .</b>	159
Greg Leung and Chi On Chui	
<b>8 Random Telegraph Noise in Multi-gate FinFET/Nanowire Devices and the Impact of Quantum Confinement . . . . .</b>	205
Runsheng Wang, Changze Liu, and Ru Huang	

<b>9 Investigations on Transport Properties of Poly-silicon Nanowire Transistors Featuring Independent Double-Gated Configuration Under Cryogenic Ambient . . . . .</b>	227
Wei-Chen Chen and Horng-Chih Lin	
<b>10 Towards Drain Extended FinFETs for SoC Applications . . . . .</b>	247
Mayank Shrivastava, Harald Gossner, and V. Ramgopal Rao	
<b>11 Modeling FinFETs for CMOS Applications . . . . .</b>	263
Lining Zhang, Chenyue Ma, Xinnan Lin, Jin He, and Mansun Chan	
<b>12 Enhanced Quantum Effects in Room-Temperature Coulomb Blockade Devices Based on Ultrascaled finFET Structure . . . . .</b>	285
Jung B. Choi	
<b>13 Single-Electron Tunneling Transistors Utilizing Individual Dopant Potentials . . . . .</b>	305
Daniel Moraru and Michiharu Tabe	
<b>14 Single-Electron Transistor and Quantum Dots on Graphene . . . . .</b>	325
Lin-Jun Wang, Tao Tu, Li Wang, Cheng Zhou, and Guo-Ping Guo	
<b>15 Terahertz Response in Schottky Warp-Gate Controlled Single Electron Transistors . . . . .</b>	351
Weihua Han and Seiya Kasai	
<b>Index . . . . .</b>	361

# Contributors

**Tarek M. Abdolkader** Department of Electrical Engineering, Umm Al-Qura University, Makkah, Saudi Arabia

**Kerem Akarvardar** Technology Research Group, Globalfoundries Inc., Albany, NY, USA

**Emanuele Baravelli** University of Bologna, Viale Risorgimento Bologna, Italy

**Mehmet O. Baykan** Portland Technology Development, Intel Corporation, Santa Clara, CA, USA

**Mansun Chan** Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

**Wei-Chen Chen** Emerging Central Laboratory, Macronix International Co., Ltd, Hsinchu, Taiwan, ROC

**Hui-Wen Cheng** Department of Electrical and Computer Engineering, National Chiao Tung University, Hsinchu, Taiwan, ROC

**Jung B. Choi** Department of Physics & Research Institute for Nano Science & Technology, Chungbuk National University, Cheongju, Korea

**Chi On Chui** Department of Electrical Engineering, University of California, Los Angeles, Los Angeles, CA, USA

**Mohammed M. El-Banna** Department of Physics and Mathematical Engineering, Faculty of Engineering, Ain Shams University, Cairo, Egypt

**Wael Fikry** Department of Physics and Mathematical Engineering, Faculty of Engineering, Ain Shams University, Cairo, Egypt

**Saeed Ganji** Electrocom Corporation, Potomac, MD, USA

**Harald Gossner** Intel Mobile Communication, Neubiberg, Germany

**Guo-Ping Guo** Key Lab of Quantum Information, Chinese Academy of Science, University of Science and Technology, Hefei, Anhui, People's Republic of China

**Weihua Han** Engineering Research Center of Semiconductor Integrated Technology, Institute of Semiconductors, Chinese Academy of Sciences, Beijing, People's Republic of China

**Jin He** Peking University Shenzhen SOC Key Laboratory, PKU-HKUST Shenzhen Institute, Shenzhen, China

**Christopher C. Hobbs** Front End Processes, SEMATECH Inc., Albany, NY, USA

**Ru Huang** Institute of Microelectronics, Peking University, Beijing, China

**Seiya Kasai** RCIQE, Hokkaido University, Sapporo, Hokkaido, Japan

**Arif Khan** Department of Physics, Neotia Institute of Technology, Management, and Science, Amira, 24-Parganas (South), West Bengal, India

Electrocom Corporation, Potomac, MD, USA

**Greg Leung** Department of Electrical Engineering, University of California, Los Angeles, Los Angeles, CA, USA

**Yiming Li** Parallel and Scientific Computing Laboratory, Electrical and Computer Engineering, National Chiao Tung University, Hsinchu, Taiwan, ROC

**Horng-Chih Lin** Department of Electronics Engineering and Institute of Electronics, National Chiao Tung University, Hsinchu, Taiwan, ROC

National Nano Device Laboratories, Hsinchu, Taiwan, ROC

**Xinnan Lin** School of Electronic and Computer Engineering, Peking University, Shenzhen, China

**Changze Liu** Institute of Microelectronics, Peking University, Beijing, China

**Chenyue Ma** Graduate School of Advanced Sciences of Matter, Hiroshima University, Hiroshima, Japan

**S. Noor Mohammad** Sciencotech, Washington, DC, USA

**Daniel Moraru** Research Institute of Electronics, Shizuoka University, Naka-ku, Hamamatsu, Shizuoka-ken, Japan

**Balwinder Raj** Department of Electronics and Communication Engineering, National Institute of Technology Jalandhar, Jalandhar, Punjab, India

**V. Ramgopal Rao** Electrical Engineering Department, IIT Bombay, Mumbai, Maharashtra, India

**Yasser M. Sabry** Department of Electronics and Communications Engineering, Faculty of Engineering, Ain Shams University, Cairo, Egypt

**Mayank Shrivastava** Department of Electronic Systems Engineering, Indian Institute of Science, Bangalore, Karnataka, India

**Michiharu Tabe** Research Institute of Electronics, Shizuoka University, Naka-ku, Hamamatsu, Shizuoka-ken, Japan

**Tao Tu** Key Lab of Quantum Information, Chinese Academy of Science, University of Science and Technology, Hefei, Anhui, People's Republic of China

**Runsheng Wang** Institute of Microelectronics, Peking University, Beijing, China

**Lin-Jun Wang** Hefei National Laboratory for Physical Sciences at the Micro-scale, University of Science and Technology, Hefei, Anhui, People's Republic of China

**Li Wang** Key Lab of Quantum Information, Chinese Academy of Science, University of Science and Technology, Hefei, Anhui, People's Republic of China

**Hao Wang** Engineering Research Center of Semiconductor Integrated Technology, Institute of Semiconductors, Chinese Academy of Sciences, Beijing, People's Republic of China

**Chadwin D. Young** Materials Science and Engineering Department, UT Dallas, Richardson, TX, USA

**Lining Zhang** Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

**Cheng Zhou** Key Lab of Quantum Information, Chinese Academy of Science, University of Science and Technology, Hefei, Anhui, People's Republic of China

# Chapter 1

## Simulation of Quantum Ballistic Transport in FinFETs

**Yasser M. Sabry, Mohammed M. El-Banna, Tarek M. Abdolkader,  
and Wael Fikry**

**Abstract** Quantum effects play a vital role in determining the transistor characteristics of FinFET devices. Quantum confinement, coherent ballistic transport, and quantum mechanical tunneling are a few examples. The nonequilibrium Green's function formalism (NEGF) provides a rigorous description of quantum transport in nanoscale devices. Depending on the chosen space for representation of the wave function, real-space and mode-space representations are widely used. In this chapter, the basic tools involved in the NEGF simulation of the quantum ballistic transport in FinFETs are provided. The different techniques applied in either the real- or the mode-space representations are discussed. In this chapter, a comparison of the NEGF methods in the real-space representation considers the recursive Green's function method, the Gauss elimination method, and the contact block reduction method and then highlights the computational efficiency of these methods. A comparison between the fully coupled, the partially coupled and the uncoupled methods in the mode-space representation is also given considering their accuracy and computational efficiency.

---

Y.M. Sabry (✉)

Department of Electronics and Communication Engineering, Faculty of Engineering,  
Ain Shams University, Cairo, Egypt  
e-mail: [ysabry@ieee.org](mailto:ysabry@ieee.org)

M.M. El-Banna • W. Fikry

Department of Physics and Mathematical Engineering, Faculty of Engineering  
Ain Shams University, Cairo, Egypt  
e-mail: [mm.elbanna@eng.asu.edu.eg](mailto:mm.elbanna@eng.asu.edu.eg); [wael\\_fikry@ieee.org](mailto:wael_fikry@ieee.org)

T.M. Abdolkader

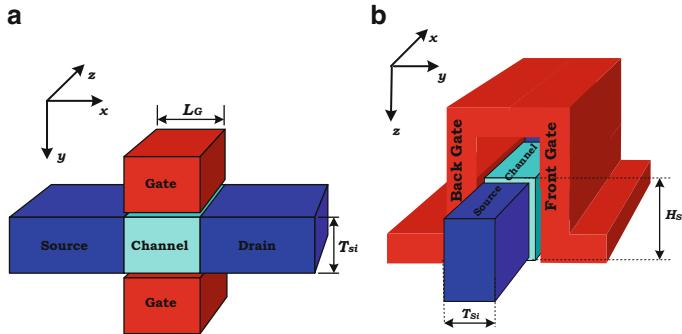
Department of Electrical Engineering, Umm Al-Qura University, Makkah, Saudi Arabia  
e-mail: [tarik\\_mak@hotmail.com](mailto:tarik_mak@hotmail.com)

## 1.1 Introduction

Scaling the CMOS technology down to and less than the 22-nm gate length regimes unveiled the limitations of the planar bulk MOSFETs and made it increasingly sound. Short-channel effects (SCEs), variability in transistor performance, poor electrostatic control of the gate on the channel, and high leakage currents are a few examples [1]. Consequently, device and process engineers started to search for possible successors to the planar bulk transistor. Early in the 1990s a UC Berkeley team led by Chenming Hu proposed the double-gate (DG) MOSFET structure shown in Fig. 1.1a where the SCEs are significantly suppressed by keeping the gate capacitance in closer proximity to the channel capacitance [2]. To satisfy a better performance, the structure shown is rotated 90° and the two sides of the gate are contacted forming a “fin” as shown in Fig. 1.1b. This results in what is called a FinFET structure. The FinFET transistors are promising candidates to pursue the scaling taking advantage of their superior features. They have thin channel region eliminating subsurface leakage paths. Their DG structure sandwiching the channel strengthens the gate control on the channel suppressing the SCEs, sets the threshold voltage of the transistor independent of the channel doping [3], and consequently reduces the variability due to random dopant fluctuation effect [1]. Indeed, the electrostatic integrity of the FinFET is higher owing to the tighter control of the channel potential by multiple gates wrapped around channel body [4]. From the switching application perspective, the FinFETs can be 37 % faster while consuming smaller amount of dynamic power and cutting down the static leakage as much as 90 % [2]. The FinFETs made their way to the industry in 2011 when Intel started its 22-nm FinFET technology [5]. On the other hand, it should be noted that FinFETs have high parasitic resistances and capacitances that may lead to a degradation of their analog performance [6].

Scaling the transistor dimensions to a few tens of nanometers necessitates accurate device simulation with predictive capability to guide future designs [7, 8]. Significant Quantum Mechanical Effects (QMEs) arise [9], which need a quantum-mechanical-based device simulation tools. Even if the classical simulation tools with sophisticated quantum corrections are available, their predictability is limited by a continuous need for calibration [7]. Numerical device simulation is performed mainly by the self-consistent solution of the electrostatic Poisson’s equation and the transport model equations. The nonequilibrium Green function formalism (NEGF) provides a distinct and an accurate method for the solution of the quantum transport problem in nanoscale devices [10]. The quantum mechanical phenomena are incorporated in an intrinsic way. The solution is, however, done with a huge computational burden unless some applicable approximations are assumed.

In this chapter, we will explain the different techniques used in the NEGF to simulate quantum FinFETs. In Sect. 1.2, a brief overview of the quantum-mechanical effects in FinFETs is given. In Sects. 1.3 and 1.4, the self-consistent solution method and the elements of the NEGF are introduced and applied on the FinFETs in the real-space domain. A comparison between the different techniques used in the real space



**Fig. 1.1** (a) The planar DG-MOSFET as originally proposed by Chenming Hu. (b) The FinFET structure

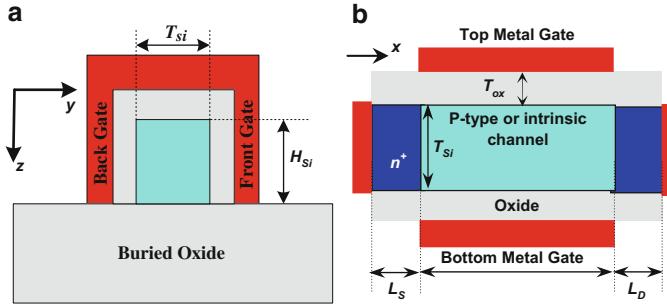
is carried out in Sect. 1.5. In Sect. 1.6, the mode-space representation is introduced and the methods based on full coupling, partial coupling, or uncoupling of the modes are explained and compared. Finally, Sect. 1.7 provides a conclusion for the chapter.

## 1.2 Quantum Effects in FinFETs

In FinFET devices, the transport of the charge carriers occurs in a silicon film of a thickness in the order of a few nanometers. This nanometric film is surrounded by an isolation layer of oxide, i.e., a high potential barrier. This results in significant QMEs which cannot be ignored in the device modeling and simulation. In fact, these QMEs can be grouped into three categories: (1) Quantum confinement across the transistor channel, (2) quantum-mechanical tunneling across the gate oxide, and (3) quantum transport of the charge carriers along the channel. These effects are briefly reviewed hereinafter.

### 1.2.1 Quantum Confinement

FinFETs with fin thickness below 20 nm exhibit QM confinement of the charge carriers across the channel [11]. Figure 1.2a shows a vertical cross section, and Fig. 1.2b shows a horizontal cross section of a FinFET device. The confinement of carriers originates from both structural [12] and electrical confinement [13]. The confinement leads to a dramatic change in the device behavior. The energy levels available for motion in the y- and z-direction are quantized with a continuum for motion in the x-direction. This quantization is accompanied with the presence of the



**Fig. 1.2** FinFET structure: (a) vertical cross section and (b) horizontal cross section

modes (or subbands), each mode with a certain band energy distribution along the channel and a certain charge carriers probability distribution across the channel. These modes are the heart of the mode-space approach that will be described in Sect. 1.4. The separation between these modes in energy as well as the energy of the lowest order mode becomes larger for smaller values of  $T_{Si}$  and  $H_{Si}$  and for larger gate bias voltage. Therefore, the threshold voltage can be controlled using  $T_{Si}$  and, thus, a lightly doped or intrinsic channel can be used. In this chapter, we will assume the channel is intrinsic while the source and the drain are heavily doped such that their intrinsic series resistance can be ignored. In many practical cases and in order to have a sufficient current flowing along the channel,  $H_{Si}$  is made much larger than  $T_{Si}$  such that the confinement effect can be considered in one direction only.

### 1.2.2 Quantum-Mechanical Tunneling

Two QM tunneling effects are likely to occur in FinFETs, tunneling through the gate oxide and source-to-drain tunneling along the transistor channel. Gate oxide tunneling is unavoidable for oxide thickness smaller than 7 nm [14]. For the FinFETs, the oxide thickness is typically less than 2 nm, and the direct tunneling mechanism dominates the other mechanisms [15]. For this reason, high-k dielectric material is used instead of silicon oxide such that their physical thickness can be larger while their effective thickness, corresponding to the oxide capacitance, is the same [15]. The gate leakage current can be accounted for within the NEGF quantum transport simulation in the real-space representation. The source-to-drain tunneling current becomes important for channel lengths below 5 nm [16]. Indeed, this mechanism is automatically accounted for within the NEGF calculation of the source to drain current without paying any extra attention.

### 1.2.3 Ballistic Transport and Quantum Interference

The phase breaking length is defined as the distance over which the electron wave's phase is destroyed by some process. It was estimated to be in the range of 50 nm for silicon [13]. The phase breaking process arises from the interaction of the electron with phonons (lattice vibrations), photons (electromagnetic vibrations), or other electrons. For macroscopic devices, phase randomizing scattering dominates, and electrons wave's phase is randomized by collisions and lose their phase during the transport process. Therefore, quantum interference effects can be neglected in macroscopic devices, and a semiclassical approach based on Boltzmann transport equation can be used to describe the transport [17]. In the nanometric devices the dimensions of which are much larger than the atoms but smaller than scattering events, the electrons may transport ballistically from one side of the device to the other side without scattering. In this case, the phase of the electrons wave nature plays an important role in the transport process because electrons can interfere constructively or destructively. It was shown that carrier density oscillations can be found near the channel barrier edges especially at low temperature [14]. At room temperature or higher, the interference effect is washed out by the statistics. Ballistic transport and carrier interference are intrinsically accounted for in the NEGF transport simulation.

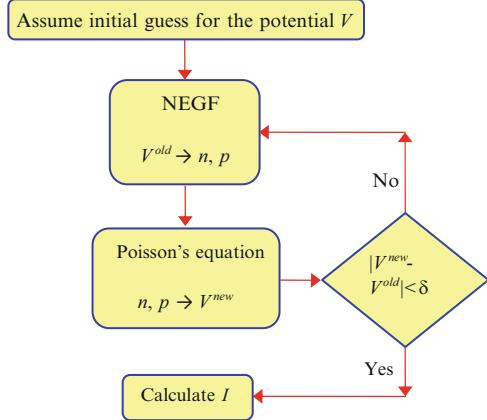
## 1.3 Self-Consistent Field Method

Device simulation process is based on solving two main problems, the electromagnetic problem and the carrier transport problem. At steady state, electromagnetic becomes electrostatic and treated by solving Poisson's equation. Poisson's equation describes the dependence of the electrostatic potential in the device on the distribution of fixed charges (e.g., dopant ions) and free carriers (electrons and holes). For inhomogeneous material device, like a FinFET, Poisson's equation is written as:

$$\nabla \cdot \nabla(\epsilon V) = -q(p - n + N_D^+ - N_A^-), \quad (1.1)$$

where  $V$  is the electrostatic potential,  $q$  is the electronic charge,  $\epsilon$  is the permittivity of the medium,  $p$  and  $n$  are the hole and electron concentrations, respectively, and  $N_D^+$  and  $N_A^-$  are the ionized donor and acceptor doping concentrations, respectively. The electrostatic and carrier transport problems are coupled together; i.e., to determine the electrostatic potential by Poisson's equation, we need the carrier distribution and at the same time, the carrier distribution is obtained from transport equations, which depend on the electrostatic potential. Therefore, the solution of the NEGF and Poisson's equation is carried out by the self-consistent field method [18]. It is an iterative method that starts by assuming an initial guess for the potential distribution in the device as shown in Fig. 1.3. According to this potential, the NEGF

**Fig. 1.3** Flow chart illustrates the self-consistent field method used in NEGF device simulation



is used to calculate the electron and hole concentrations in the device. With the known electron and hole concentrations, Poisson's equation can be solved yielding a new potential distribution. The new potential is compared to the old potential, and the solution cycle is repeated until a self-consistent solution for the potential is obtained. The self-consistency criterion is that the difference in potential between two successive iterations drops below a certain tolerance.

## 1.4 The NEGF in Real-Space Representation

The rigorous description for the NEGF can be found in the literature where it is described using an advanced language in the quantum mechanics, namely the second quantization language [19]. Fortunately, a simpler description can be also found in the literature [20–22]. Here, we are following the latter description and implementing it on the FinFET directly. Consider the FinFET structure shown in Fig. 1.2. The following assumptions are made:

1. The channel length is short enough such that the device is operating in the ballistic limit [21]
2. The top-gate oxide thickness is assumed to be much thicker than the side gate oxide such that channels are formed under the side-gate oxides only. Consequently, the simulation domain is assumed to be two-dimensional (2D)
3. Huge metal contacts; i.e., reservoirs, where thermal equilibrium is maintained and the Fermi level in these regions is determined by the applied voltage [21]
4. n-Channel transistor where holes contribution, to both the transport and the electrostatic problems, can be neglected
5. No electron penetration in the insulator region
6. A single band effective mass Hamiltonian [22, 23] is used to model the electron transport, and the concept of the “envelope” wave function is applied

In the NEGF treatment, a suitable space is chosen where the operators and the physical quantities are represented. The envelope wave function  $\psi(x,y,z)$  is expanded in terms of the orthonormal basis  $\psi(x,y)\exp(jk_z)/\sqrt{H_{\text{Si}}}$  where the quantum number  $k_z$  corresponds to the transverse eigenenergy  $E_{kz} = k_z^2/2m_z^*$ , and  $m_z^*$  is the electron effective mass in the  $z$ -direction. The eigenstates in the  $z$ -direction are plane waves because of the assumption of invariant potential in that direction. The 2D wave function  $\psi(x,y)$  is obtained from the solution of the 2D Schrödinger equation:

$$\left[ -\frac{\hbar^2}{2} \left( \frac{1}{m_x^*} \frac{\partial^2}{\partial x^2} + \frac{1}{m_y^*} \frac{\partial^2}{\partial y^2} \right) + E_c(x,y) \right] \psi(x,y) = E_l \psi(x,y), \quad (1.2)$$

where  $m_x^*$  and  $m_y^*$  are the electron effective mass in  $x$ - and  $y$ -directions, respectively,  $E_c$  is the conduction band edge, and  $E_l$  is the longitudinal energy due to motion in  $x$ - and  $y$ -directions. Upon finite difference discretization of (1.2) using  $N_x$  and  $N_y$  mesh points in the  $x$ - and  $y$ -directions, respectively, a set of linear equations is obtained and can be cast in the matrix form:

$$[\mathbf{H}_l]\{\psi\} + [\mathbf{E}_c]\{\psi\} = [E_l \mathbf{I}]\{\psi\}, \quad (1.3)$$

where

$$\begin{aligned} \mathbf{H}_l &= \begin{bmatrix} \alpha & \beta & \mathbf{0} & \cdots & \mathbf{0} \\ \beta & \alpha & \beta & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \ddots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \beta & \alpha \end{bmatrix}_{N_x N_y \times N_x N_y} & \alpha &= \begin{bmatrix} 2t_x + 2t_y & -t_y & 0 & \cdots & 0 \\ -t_y & 2t_x + 2t_y & -t_y & \cdots & 0 \\ 0 & \cdots & \ddots & \cdots & \vdots \\ 0 & 0 & \cdots & -t_y & 2t_x + 2t_y \end{bmatrix}_{N_y \times N_y} \\ \beta &= \begin{bmatrix} -t_x & 0 & \cdots & 0 \\ 0 & -t_x & \vdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & \cdots & -t_x \end{bmatrix}_{N_y \times N_y} & E_c &= \begin{bmatrix} Ec_1 & 0 & \cdots & \cdots & 0 \\ 0 & Ec_2 & 0 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & Ec_{N_x N_y} \end{bmatrix}_{N_x N_y \times N_x N_y} \\ \psi &= \begin{Bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{N_x N_y} \end{Bmatrix}_{N_x N_y \times 1} & t_x &= \frac{\hbar^2}{2m_x^*(\Delta x)^2} & t_y &= \frac{\hbar^2}{2m_y^*(\Delta y)^2}, \end{aligned}$$

where  $\Delta x$  and  $\Delta y$  are the finite-difference grid spacing in the  $x$ - and  $y$ -directions, respectively. The solution of (1.3) gives the closed boundary solution of the FinFET device without interaction with the outside world. The real problem has an open boundary and a current flowing into and drawn from the device. The NEGF accounts for that open boundary using the self-energy matrix  $\Sigma$ . The self-energy has two main differences that distinguish it from normal potential energy encountered in quantum mechanics [20]. The first difference is that  $\Sigma$  is energy dependent. Because of that, it is more convenient to think of the energy as an independent

variable and we want to know the response of the device to the incident electrons with different energies. This is always the case for an open system rather than for the isolated one which has certain eigenenergies. The second difference between  $\Sigma$  and normal potential is that  $\Sigma$  is not Hermitian. This makes the effective Hamiltonian [ $H - \Sigma$ ] has complex eigenvalues and the imaginary part of these eigenvalues broadens the density of states and gives the eigenstates a finite lifetime. In case of non-ballistic transport where a scattering events occur within the device, a scattering self-energy matrices have to be incorporated into the simulation [20].

The retarded Green's function of the active device represents the response of the device at one point due to a unit impulse excitation at another point. The device is susceptible to an excitation due to its open boundaries. The retarded Green's function is given by:

$$\mathbf{G} = [E_l \mathbf{I} - \mathbf{H} - \boldsymbol{\Sigma}_S - \boldsymbol{\Sigma}_D]^{-1}, \quad (1.4)$$

where  $\mathbf{H} = \mathbf{H}_l + \mathbf{E}_c$ ;  $\boldsymbol{\Sigma}_S$  and  $\boldsymbol{\Sigma}_D$  are the self-energy matrices accounting for the open boundary at the source and drain contacts, respectively, and given by [25]:

$$\boldsymbol{\Sigma}_S = \begin{bmatrix} \beta g_S \beta & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \vdots \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} \end{bmatrix}_{N_x N_y \times N_x N_y} \quad \boldsymbol{\Sigma}_D = \begin{bmatrix} \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \vdots & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \beta g_D \beta \end{bmatrix}_{N_x N_y \times N_x N_y}, \quad (1.5)$$

where  $\mathbf{g}_S$  and  $\mathbf{g}_D$  are the surface Green's functions of the source and drain contacts, respectively [26]. In (1.4), no electron penetration in the insulator region was assumed such that the boundary is assumed close at the gates. More details about the inclusion of the gate self energy can be found in [27].

In the NEGF, the broadening function  $\Gamma$  determines the electron exchange rates between the contact and the device. It is responsible for the broadening of the energy levels and the states finite lifetime. If the self-energy matrices are purely real, then there is no exchange rate between the device and the contacts, and the effective Hamiltonian is Hermitian with real eigenvalues. In this case, the real self-energy just shifts the eigenvalues of the device as if it was a normal potential. The broadening functions,  $\Gamma_S$  and  $\Gamma_D$ , are calculated by:

$$\Gamma_S = i[\boldsymbol{\Sigma}_S - \boldsymbol{\Sigma}_S^+] \quad \Gamma_D = i[\boldsymbol{\Sigma}_D - \boldsymbol{\Sigma}_D^+]. \quad (1.6)$$

The retarded Green's function  $\mathbf{G}$  is obtained using (1.4) and the spectral functions, which can be considered the NEGF version of the density of states, are filled by the source/drain contacts and can be afterwards obtained as:

$$\mathbf{A}_S = \mathbf{G} \Gamma_S \mathbf{G}^+ \quad \mathbf{A}_D = \mathbf{G} \Gamma_D \mathbf{G}^+. \quad (1.7)$$

Then, the energy resolved electron density is given by the diagonal of the correlation function where the latter is calculated by:

$$\mathbf{G}^n(E_l) = \mathbf{A}_S(E)F(E_l, E_{f_S}) + \mathbf{A}_D(E)F(E_l, E_{f_D}), \quad (1.8)$$

where  $E_{f_S}$  and  $E_{f_D}$  are the Fermi levels of the source and drain contact, respectively, and the function  $F$  is given by:

$$F(E, E_f) = \sqrt{\frac{2m^*z k_B T}{\pi \hbar^2}} \mathfrak{I}_{-1/2}\left(\frac{E_f - E}{k_B T}\right), \quad (1.9)$$

where  $\mathfrak{I}_{-1/2}$  is the Fermi–Dirac integral of order  $-1/2$ . The longitudinal-energy-resolved electron density at a grid point  $i$  is obtained by:

$$n(i; E_l) = \frac{\mathbf{G}^n(i, i; E_l)}{2\pi\Delta x\Delta y}. \quad (1.10)$$

The longitudinal energy resolved electron density  $n(i; E_l)$  is, further, summed over the silicon six conduction band valleys [28] and, finally, the total electron density  $n(i)$  is obtained by integration over the longitudinal energy.

The transmission coefficient from the source contact to the drain contact is defined in terms of the Green’s function and the broadening function as:

$$T_{SD} = \text{Trace}[\mathbf{G}_S \mathbf{G} \mathbf{G}_D \mathbf{G}^+]. \quad (1.11)$$

The longitudinal-energy-resolved terminal current in the ballistic limit is, afterwards, obtained as:

$$I(E_l) = \frac{q}{2\pi\hbar} T_{SD} [F(E_l, E_{f_S}) - F(E_l, E_{f_D})]. \quad (1.12)$$

The terminal current is, further, summed over the six conduction band valleys and, finally, integrated over the longitudinal energy.

## 1.5 Computationally Efficient Methods in the Real Space

The retarded Green’s function is a central quantity in the NEGF. As seen from (1.4), it is calculated by means of matrix inversion for the effective Hamiltonian matrix. The effective Hamiltonian matrix size is the same as the number of points in the grid which is  $N_{\text{grid}} = N_x N_y$  for 2D device. Numerical matrix inversion consumes a large number of operations in the order of  $N_{\text{grid}}^3$ . Moreover, the Green’s function should be calculated for each energy point considered in the simulation. This makes total number of operations scales as  $N_{\text{op}} = N_E N_{\text{grid}}^3$ , where  $N_E$  is the number of

energy points. Therefore, efforts have been exerted to develop computationally efficient methods to reduce the computational burden. In this section, three methods are considered: the Recursive Green's Function (RGF) algorithm [21, 29, 30], the Contact Block Reduction (CBR) method [23, 31], and Gauss Elimination (GE) method [24].

### 1.5.1 The Recursive Green's Function Algorithm

The RGF algorithm builds up the Green's function recursively without full inversion of the Hamiltonian matrix [21]. It can be used only if the effective Hamiltonian matrix [ $EI - H - \Sigma$ ] is block tri-diagonal. This means it allows only nearest neighbor layers coupling in the RS. Unfortunately, a lead couples all the layers connected to it [31] and, therefore, the RGF works when the device has no more than two contacts and, therefore, the gate leakage current cannot be accounted for. The RGF detailed algorithm steps can be found in [21]. The operation count of this algorithm scales approximately as  $N_y^3 N_x$ . The dependence on  $N_y^3$  arises because matrices of the sub-Hamiltonian of the device vertical layers should be inverted, and the dependence on  $N_x$  corresponds to one such inversion for each of the layers. These operations are carried out for each energy step and, therefore, the total number of operations is estimated as  $N_{\text{op}} = N_E N_y^3 N_x$ .

### 1.5.2 The Contact Block Reduction Method

There are three key points in the CBR method that makes it computationally efficient relative to the traditional NEGF [31] (1) Dyson's equation is used together with a clever splitting of the simulation domain that makes the elements of the Green's function to be calculated with inversion of a relatively small matrix, (2) the isolated device eigenstates are used as a basis for the transport problem, and the number of eigenstates needed to maintain acceptable accuracy is greatly reduced by applying von Neumann boundary condition for the isolated device Hamiltonian, and (3) the use of the leads propagating modes as a basis instead of the real-space basis in single band case where only propagating modes contribute to the current.

In the CBR method, the Greens' function of the isolated device  $\mathbf{G}^0$  is given by its spectral representation [31]:

$$\mathbf{G}^0(i, j, E) = \sum_{\alpha=1}^{N_{\text{eigen}}} \frac{\psi_{\alpha}(i)\psi_{\alpha}^*(j)}{E - \varepsilon_{\alpha} + i\eta}, \quad \eta \rightarrow 0, \quad (1.13)$$

where  $\eta i \psi_{\alpha}$  are the eigenfunctions of the isolated device,  $\varepsilon_{\alpha}$  are the corresponding eigenenergies,  $i$  and  $j$  are the grid point's indices in real space. The simulation

domain is spitted into two sub-domains, the interior part of the device and the boundary region that connects the interior parts to the contacts. The spectral function filled by the source/drain contacts,  $A_{S,D}$  is given by [31]:

$$A_{S,D}(i,j;E) = \sum_{\alpha,\beta} \psi_\alpha(i)\psi_\beta^*(j) \frac{\text{Trace}\left(\psi_\beta\psi_\alpha^+ \mathbf{B}_C^{-1} \Gamma_C^{S,D} (\mathbf{B}_C^{-1})^+\right)}{(E - \epsilon_\alpha + i\eta)(E - \epsilon_\beta + i\eta)}, \quad \eta \rightarrow 0, \quad (1.14)$$

where  $\mathbf{B}_C = I - \Sigma_C \mathbf{G}_C^0$  and the subscript C means calculated for the boundary region. The aforementioned splitting of the simulation domain and the application Von Neumann boundary condition requires the modification of the Hamiltonian matrix given in (1.3) to:

$$\mathbf{H}_l = \begin{bmatrix} \alpha + \beta & \mathbf{0} & \beta & \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \alpha + \beta & \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \cdots & \beta \\ \beta & \mathbf{0} & \alpha & \beta & \mathbf{0} & \cdots & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \beta & \alpha & \beta & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \vdots & \mathbf{0} & \mathbf{0} & \ddots & \ddots & \ddots & \ddots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \cdots & \cdots & \cdots & \ddots & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & \cdots & \cdots & \cdots & \cdots & \mathbf{0} & \beta & \alpha & \beta \\ \mathbf{0} & \beta & \mathbf{0} & \cdots & \cdots & \mathbf{0} & \beta & \alpha & \beta \end{bmatrix}. \quad (1.15)$$

And the corresponding self-energy matrix is given by:

$$\sum = \left[ \begin{bmatrix} \Sigma_S & \mathbf{0} \\ \mathbf{0} & \Sigma_D \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{N_C \times N_C} \quad \mathbf{0} \right]_{N_{\text{grid}} \times N_{\text{grid}}}, \quad (1.16)$$

where  $\Sigma_S = \beta g_S \beta - \beta$  and  $\Sigma_D = \beta g_D \beta - \beta$ .

The double summation on the eigenstates in (1.14) is composed of two terms, the first one is energy independent but position dependent and the second one is the opposite. Therefore, the number of operations can be estimated as  $N_{\text{op}} = N_{\text{eigen}}^2 N_E + N_{\text{eigen}}^2 N_{\text{grid}}$  [31] where  $N_{\text{eigen}}$  is the number of eigenstates to be used. The Numerical calculation effort of the transmission function and the spectral function can be further reduced in the single-band case by transforming the Green's function and the self-energy into a basis of the leads mode space [31]. The idea behind choosing these modes as a basis is that since the potential is constant inside a given lead, then its modes are truly uncoupled which results in diagonal self-energy matrices. Besides diagonal self-energy matrices, only few modes contribute to the transmission at a given energy [31]. Therefore, the number of operations can be estimated as  $N_{\text{op}} = N_E N_{\text{eigen}} N_{\text{modes}} N_{\text{grid}}$  [23] in which higher orders of  $N_{\text{grid}}$  or  $N_{\text{eigen}}$  are absent.

### 1.5.3 The Gauss Elimination Method

The idea in this method is based on the sparse nature of the broadening function which can be seen from (1.5) and (1.6). Consequently, the entire Green's function isn't needed to calculate the spectral functions in (1.6) and (1.7). Instead, the spectral functions are calculated using the following equations [24]:

$$\mathbf{A}_S = \mathbf{G}_S [\beta(g_S - g_S^+) \beta] \mathbf{G}_S^+ \quad \mathbf{A}_D = \mathbf{G}_D [\beta(g_D - g_D^+) \beta] \mathbf{G}_D^+, \quad (1.17)$$

where  $\mathbf{G}_S$  and  $\mathbf{G}_D$  are submatrices of the retarded Green's function and can be obtained using Gauss elimination method by the following equations:

$$\mathbf{G}_S = [E_l \mathbf{I} - \mathbf{H} - \boldsymbol{\Sigma}_S - \boldsymbol{\Sigma}_D] \setminus \mathbf{I}_S \quad \mathbf{G}_D = [E_l \mathbf{I} - \mathbf{H} - \boldsymbol{\Sigma}_S - \boldsymbol{\Sigma}_D] \setminus \mathbf{I}_D, \quad (1.18)$$

where  $A \setminus B$  denotes division of the  $B$  by  $A$  using Gauss elimination method;  $\mathbf{I}_S$  and  $\mathbf{I}_D$  are given by:

$$\mathbf{I}_S = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ 0 & 0 & \ddots & 0 \\ \vdots & \ddots & \ddots & 1 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix}_{N_x N_y \times N_y} \quad \mathbf{I}_D = \begin{bmatrix} 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \vdots & \ddots & 1 & \vdots \\ 0 & \cdots & \cdots & 1 \end{bmatrix}_{N_x N_y \times N_y}. \quad (1.19)$$

In fact, using this technique, we calculate only  $N_x N_y \times 2N_y$  elements of the Green's function instead of calculating  $N_x N_y \times N_x N_y$  elements.

### 1.5.4 Computational Efficiency Comparison

The three methods discussed previously were implemented and integrated into the FETMOSS simulator [18]. Their computational efficiency is calculated and compared relative to the traditional NEGF. For this purpose, the four methods (the traditional NEGF, the RGF algorithm, the CBR method, and the GE method) were used to simulate the device with the parameters given in Table 1.1. The drain voltage was kept constant at 0.7 V, and the gate voltage was swept from 0.0 to 0.7 V with a step of 0.1 V. Thus, we have eight bias points. For the first bias point; i.e.,  $V_{GS} = 0.0$  V, the initial guess was taken to be the zero potential at various grid points in the device. The initial guess for any other bias point was taken from the solution of the preceding bias point, for example, initial guess for  $V_{GS} = 0.1$  V was taken from the solution of  $V_{GS} = 0.0$  V. It is important to mention that, the traditional NEGF, the RGF algorithm, and the GE method are all giving exactly the same current for the same applied voltage. This is because neither the RGF

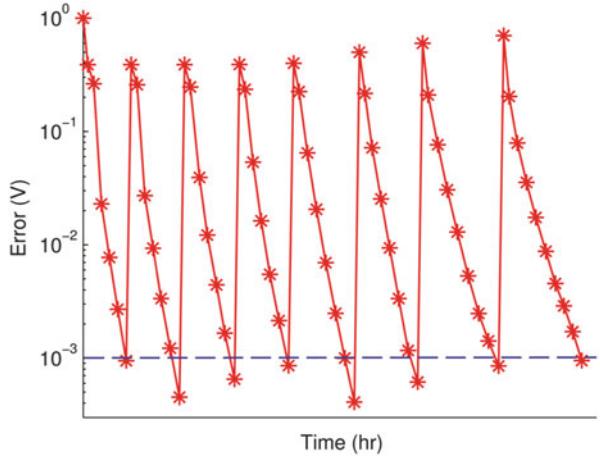
**Table 1.1** The simulated devices dimensions, doping concentration, material parameters, simulator options, the finite difference grid spacing, and the supply voltage

Category	Parameter	Value
Dimensions	Channel length $L$	5 nm
	Source and drain length $L_S, L_D$	5 nm
	Oxide thickness $T_{ox}$	1 nm
	Silicon (body) thickness $T_{Si}$	2 nm
Doping	Channel doping	$10^{10} \text{ cm}^{-3}$
	Source and drain doping	$2 \times 10^{20} \text{ cm}^{-3}$
	Junction doping profile	Step
Material	Silicon relative permittivity $\epsilon_{Si}$	$11.7 \epsilon_0$
	Oxide relative permittivity $\epsilon_{ox}$	$3.9 \epsilon_0$
	Top and bottom gate work function $\phi_m$	4.5 eV
	Longitudinal electron effective mass $m_l^*$	$0.91 m_o$
	Transverse electron effective mass $m_t^*$	$0.19 m_o$
	Self-consistence tolerance $\delta$	$10^{-3}$ V
	Poisson's tolerance	$10^{-6}$ V
Grid	Vertical node spacing	0.1 nm
	Horizontal node spacing	0.2 nm
Supply voltage	$V_{DD}$	0.7 V

algorithm nor the GE methods trades off the accuracy with the simulation speed while the CBR method does. A key parameter in the CBR method is the number of eigenstates  $N_{\text{eigen}}$  used in the simulation. The lesser the eigenstates, the faster the simulation and the lesser accurate are the results. It has been demonstrated that the needed percentage of eigenstates for a given acceptable accuracy (less than 5 % in the terminal current) is bias dependent and can vary from 6 % in the on-state to 40 % in the off-state [13]. For this reason  $N_{\text{eigen}}$  was decreased gradually from 40 % at  $V_{GS} = 0.0$  V to 6 % at  $V_{GS} = 0.7$  V.

Figure 1.4 depicts the self-consistent error versus simulation time for the traditional NEGF, the RGF algorithm, and the GE method while the CBR method has a slightly different error [32]. A solution is found when the error drops below 1 mV. Once this criterion is met, the terminal current is calculated and a new bias point is initiated. This causes the error to jump to a larger value, and the error starts decreasing again with the iterations until the solution of the new bias point is found. The cycle was repeated until the eight bias points were completed. The simulation time differs considerably from one method to another. The traditional NEGF with full matrix inversion has the greatest simulation time while the CBR method has the smallest one. A summary of the total simulation time ( $t_{\text{total}}$ ), average simulation time per bias point ( $t_{\text{bias}}$ ), and the average simulation time per iteration ( $t_{\text{iteration}}$ ) is presented in Table 1.2. These simulations were carried out on a home PC: Intel® Pentium 4 CPU 2.4 GHz, 768 MB RAM. Using the RGF algorithm or the GE method introduces about 1 order of magnitude reduction in simulation time below that traditional NEGF. The CBR method yields the smallest simulation time with about 2 orders of magnitude reduction. By such a great reduction in the simulation

**Fig. 1.4** The self-consistent error versus time using the traditional NEGF, the RGF algorithm, and the GE method



**Table 1.2** The simulation time comparison between the traditional NEGF, the GE method, the RGF method, and the CBR method

Method	$t_{\text{total}}$ (h)	$t_{\text{bias}}$ (h)	$t_{\text{iteration}}$ (h)
Traditional NEGF	75.7218	9.4652	1.2022
The GE method	11.6525	1.4566	0.1850
The RGF algorithm	5.0872	0.6359	0.0807
The CBR method	0.8661	0.1108	0.0135

These simulations were carried out on a home PC: Intel® Pentium 4 CPU 2.4 GHz, 768 MB RAM

time, the CBR method makes it practical to simulate and design FINFETs using the RS simulations. The main disadvantage of the CBR method is the necessity to dynamically determine the number of eigenstates to achieve the desired accuracy [33]. This is not necessary in either the RGF algorithm or the GE method.

## 1.6 The NEGF in Mode-Space Representation

Although the real-space representation is considered the most accurate approach within the NEGF, the simulation time is a main obstacle for its usage in device simulation due to the huge size of the Hamiltonian matrix of the device region for 2 or 3D calculations. Alternatively, the Mode-Space (MS) approach has been proposed in which the wave functions are expanded in terms of the device eigenfunctions and the transport calculation can be simplified to a 1D problem along the transport direction [25, 34, 35]. The Uncoupled-Mode Space (UMS) is the fastest approach ignoring the coupling between all the device's modes and treating the transport problem of each mode separately [36, 37]. The Schrödinger equation is calculated based on analytical approximation [38], numerically for one slice perpendicular to the transport direction [39], or for many slices [37]. Unfortunately,

the approach is limited to ultrathin devices in which  $T_{\text{SI}}$  is less than 5 nm [35, 36]. More accurately, the Coupled-Mode Space (CMS) fully accounts for the coupling terms between the modes resulting in more accurate simulation where the CMS computational burden is in between the real-space and uncoupled MS [39, 40]. Recently, a novel approach was proposed for solving the NEGF transport problem by incorporating partial coupling between the modes. In this way, the proposed approach combines the advantage of the CMS in accuracy and uncoupled MS in the same efficiency [41–42]. The proposed as well as the CMS approach were implemented into the FETMOSS simulator [18]. In this section, we first review the CMS approach theory in brief; then the recently proposed Partial-Coupled Mode Space (PCMS) is introduced and benchmarked against the CMS.

### 1.6.1 Coupled Mode-Space Approach

In the CMS representation, the wave function  $\psi(x,y)$  is expanded in the form:

$$\psi(x,y) = \sum_n \phi_n(x) \chi_n(x,y), \quad (1.20)$$

where  $\phi_n(x)$  are the expansion coefficients and  $\chi_n(x,y)$  are the modes associated with confinement in the  $y$ -direction. The modes are obtained by solving a 1D wave equation in the  $y$ -direction within each vertical slice of the device along the  $x$ -direction:

$$-\frac{\hbar^2}{2m_y^*} \frac{\partial^2 \chi_n(x,y)}{\partial y^2} + E_c(x,y) \chi_n(x,y) = E_n(x) \chi_n(x,y), \quad (1.21)$$

where  $E_n(x)$  represents the bottom of the mode  $n$ . Substituting (1.20) into (1.21), multiplying by  $\chi_m^*(x,y)$ , and integrating over  $y$  to get [40]:

$$\begin{aligned} & -\frac{\hbar^2}{2m_x^*} \frac{\partial^2 \phi^{(m)}(x)}{\partial x^2} + [E^{(m)}(x) - E_1] \phi^{(m)}(x) \\ &= \sum_{n=1}^{N_m} [a_{mn}(x) + b_{mn}(x)] \phi^{(n)}(x), \end{aligned} \quad (1.22)$$

where  $a_{mn}(x) = -\frac{\hbar^2}{2m_x^*} \left[ \int dy \chi^{*(m)}(x,y) \frac{\partial^2}{\partial x^2} \chi^{(n)}(x,y) \right]$   
and  $b_{mn}(x) = -\frac{\hbar^2}{2m_x^*} \left[ 2 \int dy \chi^{*(m)}(x,y) \frac{\partial^2}{\partial x^2} \chi^{(n)}(x,y) \right]$ .

$N_m$  is maximum number of subbands (modes). For each  $m$ th mode, the right-hand side of (1.22) involves a summation over all other modes including the  $m$ th mode itself. This summation gives rise to coupling between the modes. Only a few of the lowest modes are occupied and need to be included in the simulation due to quantum confinement. Equation (1.22) implies a set of  $N_m$  equations that is in the matrix form:

$$[\mathbf{H}_{\text{CMS}}] \left\{ \phi^{(m)} \right\} = [E_l \mathbf{I}] \left\{ \phi^{(m)} \right\},$$

where

$$[\mathbf{H}_{\text{CMS}}] = \begin{bmatrix} h_{11} & h_{12} & h_{13} & \cdots & h_{1N_m} \\ h_{21} & h_{22} & h_{23} & \cdots & h_{2N_m} \\ h_{31} & h_{32} & h_{33} & \cdots & h_{3N_m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{N_m 1} & h_{N_m 2} & h_{N_m 3} & \cdots & h_{N_m N_m} \end{bmatrix}_{N_m \times N_m}, \quad \left\{ \phi^{(m)} \right\} = \begin{Bmatrix} \phi^1(x) \\ \phi^2(x) \\ \phi^3(x) \\ \vdots \\ \phi^{N_m}(x) \end{Bmatrix}_{N_m \times 1} \quad (1.23)$$

and

$$h_{mn}(x) = \left[ -\frac{\hbar^2}{2m_x^*} \frac{\partial^2}{\partial x^2} + E_m(x) \right] \delta_{m,n} - a_{mn}(x) - b_{mn}(x) \frac{\partial}{\partial x}, \quad (1.24)$$

where  $h_{mn}$  matrix dimension is  $(N_x \times N_x)$ ,  $N_x$  is the number of grid points in  $x$ -direction, and  $\delta_{m,n}$  is the Kronecker delta function. Indeed, the CMS formalism is equivalent in accuracy to the real space, if all the modes are included in the summation, i.e.,  $N_m = N_y$ . However, by using a smaller number of modes, the Hamiltonian matrix size can be reduced from  $(N_x N_y \times N_x N_y)$  to  $(N_x N_m \times N_x N_m)$  and a considerable gain in the simulation time is achieved. Finally, the NEGF framework calculates the electron density and current after the device Hamiltonian is obtained [14, 28, 39, 43]. The CMS fully account for the different terms of (1.22) while the UMS completely ignores the right-hand side of this equation and set to zero. In the next section, we show that only partial coupling between the modes is needed, for accurate simulation, depending on whether the mode shape (or index) is odd or even.

### 1.6.2 Partial-Coupled Mode-Space Approach

The PCMS approach combines the advantage of the CMS and the UMS based on two criteria [44]. The first one is the choice of the suitable number of modes contributing to the charge density and terminal current calculations. The second one is the elimination of some coupling terms between modes that have no contribution in calculating the charge and the current.

**Table 1.3** Device parameters list used in simulation

Device	1	2	3	4	5
Year of production	2013	2015	2017	2019	2021
Channel length $L$ (nm)	13	10	8	6	5
Si thickness $T_{\text{Si}}$ (nm)	7.5	6	4.5	3.8	3.2
Oxide thickness $T_{\text{ox}}$ (nm)	0.6	0.6	0.55	0.5	0.5
Supply voltage $V_{\text{DD}}$ (V)	0.9	0.8	0.7	0.7	0.65

The coupling effect of the  $m$ th mode on the  $n$ th mode is defined by the coupling term that was defined in (1.22):

$$C_{mn}(E_1, x) = [a_{mn}(x) + b_{mn}(x)]\phi^{(n)}(x, E_1). \quad (1.25)$$

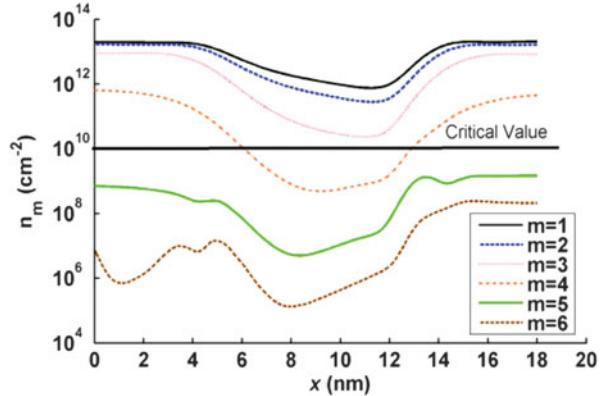
It depends on both the longitudinal energy and the  $x$ -position along the transistor channel. The coupling terms are used as indicator of whether the coupling between the modes is important or can be neglected. The simulator FETMOSS was modified to calculate these terms for assessing their importance [18, 32, 37].

For this study, we select five devices to cover the ITRS targets from the year 2013 up to the year 2021 [45]. Table 1.3 lists the parameters of these devices. All devices were subjected to PCMS and CMS simulation. The transistor channel is intrinsic and the gate metal work function is 4.65 eV. For source and drain regions, doping was selected to be  $2 \times 10^{20} \text{ cm}^{-3}$  while length was 5 nm.

As previously mentioned, the quantum confinement creates discrete steps between the subband energy of one mode and the higher order one. For high energy levels, Fermi–Dirac statistics tells that the mode is almost not occupied by electrons and, therefore, has negligible effect on the simulation results. Thus, a criterion for determining the suitable number of modes is needed. After several iterations, we found that taking modes with charge density  $\geq 10^{-3}$  of the lowest order mode's charge density is sufficient for charge and current calculations. An example is shown in Fig. 1.5 where device 3 is simulated and the 2D charge density for different modes is depicted. The needed number of modes in this case is 4. By using the same criteria, the numbers of modes are 7, 6, 4, and 3 for devices 1, 2, 4, and 5, respectively. The use of larger number of modes will increase the simulation time without adding value to the accuracy.

To determine the eliminated coupling terms according to the second criteria, we will consider the maximum absolute values of the coupling terms  $|C_{mn}|_{\max}$ . The value of these quantities was studied at different bias conditions. The on-state ( $V_{\text{GS}} = V_{\text{DS}} = V_{\text{DD}}$ ) of the device is presented here because the same observations were found for different conditions. The coupling terms  $|C_{mn}|_{\max}$  are calculated and presented in Table 1.4 for device 1 in the on-state. The common observation among all modes is that there is (are) always one or two dominant term(s) where the other terms are order of magnitude smaller and can be neglected. For example,  $C_{13}|_{\max}$  is the dominate term for the first order mode,  $C_{24}|_{\max}$  for the second order mode,  $C_{13}|_{\max}$  and  $C_{35}|_{\max}$  for the third order mode, and so on. In fact the same observation was found for all other devices under study.

**Fig. 1.5** 2D electron density of different modes along the channel in  $x$ -direction for device 3 in on-state



**Table 1.4**  $|C_{mn}|_{\max}$  for device 1 in on-state with dominant terms highlighted

$n \backslash m$	1	2	3	4	5	6	7
1	0.01	0.07	0.31	0.02	0.05	0.01	0.01
2	0.07	0.01	0.02	0.36	0.01	0.01	0.00
3	0.33	0.02	0.02	0.03	0.33	0.01	0.06
4	0.02	0.35	0.03	0.02	0.01	0.26	0.00
5	0.05	0.01	0.34	0.01	0.02	0.01	0.23
6	0.01	0.07	0.01	0.30	0.01	0.02	0.01
7	0.02	0.00	0.07	0.00	0.26	0.01	0.01

The existence of a few dominant terms among the coupling terms in symmetric FinFETs structure indicates that there is no need for considering full coupling among the modes. One can deduce that the problem can be decoupled into two smaller problems, where modes number 2, 4, and 6 have to be solved together due to their significant coupling. At the same time, modes number 1, 3, 5, and 7 follow the same case. Therefore, we have one problem in three unknowns and another one in four unknowns. Generally, the problem with  $N_m$  unknowns can be divided into two smaller problems: one for the odd modes and the other one for the even modes. The size of these problems is  $N_m/2$  and  $N_m/2$ , if  $N_m$  is even or  $(N_m - 1)/2$  and  $(N_m + 1)/2$ , if  $N_m$  is odd. Since the solution of a linear system of  $N$  unknowns involves  $N^3$  operations [46], the proposed PCMS relative operations count with respect to the CMS is:

$$N_{\text{rel}} = 2(N_m/2)^3/N_m^3 = 25\%, \quad \text{if } N_m \text{ is even}$$

$$N_{\text{rel}} \left\{ [(N_m - 1)/2]^3 + [(N_m + 1)/2]^3 \right\} / N_m^3, \quad \text{if } N_m \text{ is odd.} \quad (1.26)$$

In the language of the NEGF, the PCMS involves the division of the Hamiltonian matrix in (1.23) into two separate Hamiltonians  $H_{\text{odd}}$  and  $H_{\text{even}}$  given by:

$$H_{\text{odd}} = \begin{bmatrix} h_{11} & h_{13} & h_{15} & \cdots & h_{1i} \\ h_{31} & h_{33} & h_{35} & \cdots & h_{3i} \\ h_{51} & h_{53} & h_{55} & \cdots & h_{5i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{i1} & \cdots & \cdots & \cdots & h_{ii} \end{bmatrix}_{p \times p}, \quad H_{\text{even}} = \begin{bmatrix} h_{22} & h_{24} & h_{26} & \cdots & h_{2j} \\ h_{42} & h_{44} & h_{46} & \cdots & h_{4j} \\ h_{62} & h_{64} & h_{66} & \cdots & h_{6j} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{j2} & \cdots & \cdots & \cdots & h_{jj} \end{bmatrix}_{q \times q}. \quad (1.27)$$

where  $i = N_m - 1$ ,  $j = N_m$ ,  $p = N_m/2$  and  $q = N_m/2$ , if  $N_m$  is an even number, while  $i = N_m$ ,  $p = (N_m + 1)/2$ ,  $j = (N_m - 1)$ , and  $q = (N_m - 1)/2$ , if  $N_m$  is an odd number. The terms connecting even and odd modes are all neglected. The electron density and terminal current can be then calculated normally within the NEGF [43], one time for the odd modes and a second time for the even modes, and their contributions are added.

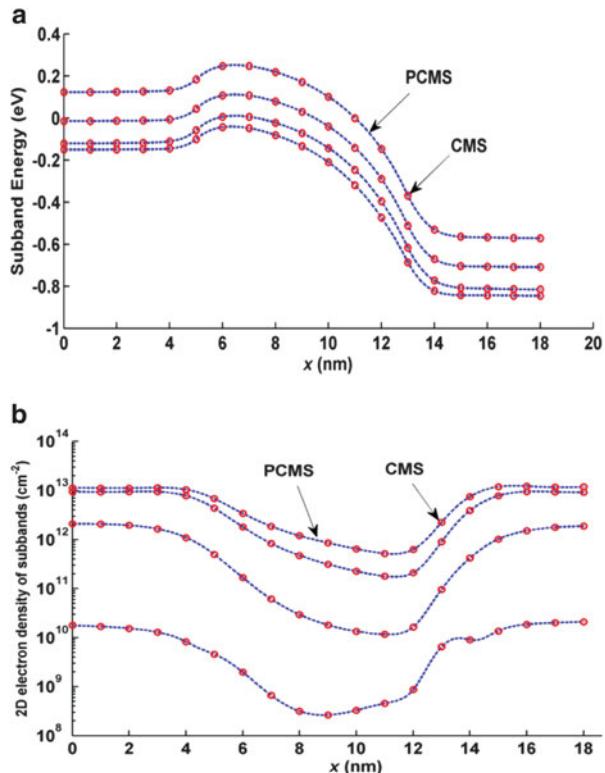
### 1.6.3 Validation of the PCMS Approach

A broad study was carried out on the devices given in Table 1.3 with the purpose of validating the PCMS with respect to the CMS and to evaluate the practical reduction in simulation time with (1.26). The simulation was carried out on a home PC with 3 GHz, Core 2 Quad 64 bit AMD processor, and 8 GB RAM memory. The solution of Schrödinger equation in the transverse direction ( $y$ -direction) results in subband energies which values vary in the channel direction. Figure 1.6a depicts well matching of the profiles of subband edges along the channel ( $x$ -direction) for device 3 in on-state for the four lowest subbands in the primed valley using PCMS and CMS approaches. The four subbands are chosen according to the first criteria mentioned before.

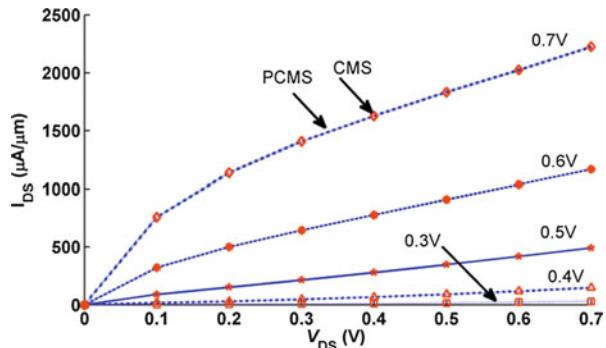
The occupation of the different subbands with carriers is calculated from the 2D subband carrier density  $N_{2D}$  ( $\text{cm}^{-2}$ ). Figure 1.6b illustrates the 2D electron density of the lowest subbands along the channel for device 3 in on-state. In comparing PCMS with respect to CMS approaches, the maximum percentage error equal 0.2 % with 72.6 % reduction in simulation time according to second criteria.

For example, Fig. 1.7 illustrates a sample  $I_{\text{DS}} - V_{\text{DS}}$  characteristics at different  $V_{\text{GS}}$  for device 3. The drain voltage was swept from 0 to 0.7 V with steps of 0.1 V while the gate voltage was swept from 0.3 to 0.7 V in steps of 0.1 V. According to the first criteria, the curves predict matching results between the PCMS and the CMS approaches.

**Fig. 1.6** (a) Subband energy profile and (b) 2D electron density along the channel for device 3 in on-state for the four lowest subbands in the primed valley using PCMS and CMS approaches



**Fig. 1.7**  $I_{DS}$ - $V_{DS}$  family of curves for device 3 at different  $V_{GS}$  using PCMS and CMS approaches



The percentage difference in terminal current and electron charge density (3D distribution) between PCMS and CMS were calculated and tabulated in Table 1.5 for the various simulated devices. They are defined as:

$$\Delta I = \left| \frac{I_{\text{CMS}} - I_{\text{PCMS}}}{I_{\text{CMS}}} \right| \times 100, \quad \Delta n = \max \left| \frac{n_{\text{CMS}} - n_{\text{PCMS}}}{n_{\text{CMS}}} \right| \times 100. \quad (1.28)$$

**Table 1.5** Accuracy and computations reduction of the PCMS relative to CMS

Device	1	2	3	4	5
$N_m$	7	6	4	4	3
$\Delta n$ off-state	$7 \times 10^{-1}$	$5 \times 10^{-1}$	$1 \times 10^{-1}$	$2 \times 10^{-3}$	$3 \times 10^{-3}$
$\Delta n$ on-state	$1 \times 10^{-1}$	$3 \times 10^{-2}$	$2 \times 10^{-2}$	$2 \times 10^{-3}$	$2 \times 10^{-3}$
$\Delta I$ off-state	$6 \times 10^{-3}$	$9 \times 10^{-4}$	$3 \times 10^{-5}$	$10 \times 10^{-4}$	$3 \times 10^{-4}$
$\Delta I$ on-state	$1 \times 10^{-2}$	$5 \times 10^{-3}$	$3 \times 10^{-5}$	$2 \times 10^{-4}$	$3 \times 10^{-5}$
$N_{\text{rel}}$	27.2	25	25	25	33.3
$t_{\text{rel}}$ off-state	26.8	28.6	27.6	26.5	34.9
$t_{\text{rel}}$ on-state	26.9	26.1	27.4	26.4	35.1

The simulation was repeated many times for different bias conditions to check the relative accuracy of PCMS. Excellent agreement was found between the PCMS and the CMS in all bias conditions and for all the simulated devices. Among these bias conditions, the cases of off- and on-states are given in Table 1.5. We define the relative time consumed by the PCMS simulation and the CMS simulation by:

$$t_{\text{rel}} = \frac{t_{\text{PCMS}}}{t_{\text{CMS}}} \times 100. \quad (1.29)$$

The simulation time in both cases was recorded, and the relative was calculated and listed in Table 1.5. The predicted value from (1.26) was also calculated and tabulated. The computational burden reduction ranges from 75 to 65 %. The predicted and recorded values from the simulation agree well. The small difference between them can be due to other operations that don't scale well with  $N^3$ .

## 1.7 Conclusion

Simulation of quantum ballistic transport in FinFET devices using the nonequilibrium Green's function formalism (NEGF) was discussed in details. The NEGF provides a rigorous method for simulation of the various phenomena occurring in FinFET devices. Thus, the NEGF enables design optimization of the FinFETs and accurate prediction of their characteristics with further futuristic scaling. A comparison was carried out between the different simulation methods in both the real-space representation of the wave function and the mode-space representation. In the real space, the contact block reduction (CBR) method was shown to be the most computationally efficient approach enabling about 2 orders of magnitude reduction in the simulation time with respect to the traditional NEGF. Therefore, one expects that the CBR method will make it practical to simulate and design true 2D and 3D devices on a home PC. The mode-space representation offers a greater computational efficiency but on the expenses of its inability to account for all type of scattering within the device or for the gate leakage current. A recent approach was presented that allows for partial coupling between the eigenfunctions and

dividing the large problem into two smaller ones. Using this method, a reduction of 70 % in the simulation time was achieved compared to the full coupling method, without loss of accuracy.

## References

1. Ho, B., et al.: Design optimization of multigate bulk MOSFETs. *IEEE Trans. Electron Dev.* **60**, 28–33 (2013)
2. Kawa, J., Biddle, A.: FinFET: the promises and the challenges. *Synopsys Insight Newslett.* (2012). <http://www.synopsys.com/Company/Publications/SynopsysInsight/Pages/Art2-finfet-challenges-ip-IssQ3-12.aspx>
3. Abdolkader, T.M., Fikry, W.: Semi-empirical quantum correction model for electron concentration in symmetric double gate MOSFETs. In: Proceedings of the 2004 international conference on electrical, electronic and computer engineering, ICEEC'04, pp. 549–552 (2004)
4. Bhoj, A.N., et al.: 3-D-TCAD-based parasitic capacitance extraction for emerging multigate devices and circuits. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. PP, pp. 1–1 (2013)
5. Intel® 22nm Technology. <http://www.intel.com/content/www/us/en/silicon-innovations/intel-22nm-technology.html>
6. Tinoco, J.C., et al.: Impact of extrinsic capacitances on FinFET RF performance. *IEEE Trans. Microw. Theor. Tech.* **61**, 833–840 (2013)
7. Martinez, A., et al.: The non-equilibrium green function approach as a TCAD tool for future CMOS technology. In: International conference on simulation of semiconductor processes and devices (SISPAD), pp. 95–98 (2011)
8. Sabry, Y.M., et al.: Quantum transport based simulation and design optimization of a 10 nm FinFET. Dts: 2009. In: Proceedings of the 4th IEEE international conference on design & technology of integrated systems in nanoscale era (DTIS), pp. 125–129 (2009)
9. Abdolkader, T.M.: Modeling and 2D simulation of double-gate SOI devices. Ph.D. Dissertation, Engineering Physics and Mathematics Department, Ain Shams University, Cairo (2005)
10. El-Banna, M.M., et al.: Simulation of quantum transport in double gate MOSFETs: coupled-mode space versus real space. In: 2012 International conference on engineering and technology (ICET), pp. 1–5 (2012)
11. Venugopalan, S., et al.: Compact models for real device effects in FinFETs. In: The international conference on simulation of semiconductor processes and devices (SISPAD), Denver, CO, USA (2012)
12. Mingchun, T., et al.: Quantum compact model for ultra-narrow body FinFET. In: 10th International conference on ultimate integration of silicon. ULIS 2009, pp. 293–296 (2009)
13. Ferry, D.K., Goodnick, S.M.: Transport in nanostructures. Cambridge University Press, Cambridge (1997)
14. Ren, Z.: Nanoscale MOSFETs: physics, simulation, and design. Ph.D. Thesis, Purdue University, West Lafayette (2001)
15. El Hakim, M.A., et al.: Gate leakage in low standby power 16 nm gate length double-gate MOSFETs. In: 26th National radio science conference (NRSC 2009), 17–19 Mar 2009, Cairo, Egypt, pp. 1–9 (2009)
16. Sverdlov, V.A., et al.: Nanoscale silicon MOSFETs: a theoretical study. *IEEE Trans. Electron Dev.* **50**, 1926–1933 (2003)
17. Lundstrom, M.: Fundamentals of carrier transport. Cambridge University Press, Cambridge (2009)
18. Abdolkader, T.M., et al.: FETMOSS: a software tool for 2D simulation of double-gate MOSFET. *Int. J. Numer. Model. Electron. Netw. Dev. Fields* **19**, 301–314 (2006)

19. Tröster, P., et al.: Transport calculations based on density functional theory, Friedel's sum rule, and the Kondo effect. *Phys. Rev. B* **85**, 115409 (2012)
20. Datta, S.: Electronic transport in mesoscopic systems. Cambridge University Press, Cambridge (1997)
21. Anantram, M.P., et al.: Modeling of nanoscale devices. *Proc. IEEE* **96**, 1511–1550 (2008)
22. Datta, S.: Quantum phenomena. Addison-Wesley, Reading, MA (1989)
23. Khan, H., et al.: Self-consistent treatment of quantum transport in 10 nm FinFET using contact block reduction (CBR) method. *J. Comput. Electron.* **6**, 77–80 (2007)
24. Sabry, Y.M., et al.: A computationally efficient method for quantum transport simulation of double-gate MOSFETs. In: 2009 National radio science conference, NRSC 2009, 17–19 Mar 2009, New Cairo, Egypt (2009)
25. Venugopal, R., et al.: Simulating quantum transport in nanoscale transistors: real versus mode-space approaches. *J. Appl. Phys.* **92**, 3730–3739 (2002)
26. Damle, P.S.: Nanoscale device modeling: from MOSFETs to molecules. Citeseer (2003)
27. Elmaghraby, Y.I.: Simulation of leakage gate current in multi-gate nano-scale transistors. M.Sc. Dissertation, Electronics and Communications Department, Ain Shams University, Cairo (2011)
28. Pierret, R.F.: Advanced semiconductor fundamentals. Prentice Hall, Englewood Cliffs, NJ (2003)
29. Svizhenko, A., et al.: Two-dimensional quantum mechanical modeling of nanotransistors. *J. Appl. Phys.* **91**, 2343–2354 (2002)
30. Klimeck, G., et al.: Development of a nanoelectronic 3-D (NEMO 3-D) simulator for multi-million atom simulations and its application to alloyed quantum dots. *Comput. Model. Eng. Sci.* **3**, 601 (2002)
31. Mamaluy, D., et al.: Contact block reduction method for ballistic transport and carrier densities of open nanostructures. *Phys. Rev. B* **71**, 245321 (2005)
32. Sabry, Y.M., et al.: Simulation of quantum transport in double-gate MOSFETs using the non-equilibrium Green's function formalism in real-space: a comparison of four methods. *Int. J. Numer. Model. Electron. Netw. Dev. Fields* **24**(4), 322–334 (2011)
33. Sabry, Y.M., et al.: Inspection of the contact block reduction method for quantum transport simulation of FinFETs. In: 26th National radio science conference (NRSC 2009), 17–19 Mar 2009
34. Luisier, M., et al.: Quantum transport in two-and three-dimensional nanoscale transistors: coupled mode effects in the nonequilibrium Green's function formalism. *J. Appl. Phys.* **100**, 043713 (2006)
35. Kurniawan, O., et al.: Ballistic calculation of nonequilibrium Green's function in nanoscale devices using finite element method. *J. Phys. D Appl. Phys.* **42**, 105109 (2009)
36. Ren, Z., et al.: NanoMOS 2.5: a two-dimensional simulator for quantum transport in double-gate MOSFETs. *IEEE Trans. Electron Dev.* **50**, 1914–1925 (2003)
37. Sabry, Y.M., et al.: Uncoupled mode-space simulation validity for double gate MOSFETs. In: 2007 International conference on microelectronics, pp. 158–161 (2007)
38. Kurniawan, O., et al.: Simplified model for ballistic current–voltage characteristic in cylindrical nanowires. *Microelectron. J.* **41**, 155–161 (2010)
39. Polizzi, E., et al.: A three-dimensional quantum simulation of silicon nanowire transistors with the effective mass approximation. *J. Appl. Phys.* **96**, 2192–2203 (2004)
40. Fiori, G., et al.: Coupled mode space approach for the simulation of realistic carbon nanotube field-effect transistors. *IEEE Trans. Nanotechnol.* **6**, 475–480 (2007)
41. El-Banna, M.M., et al.: Partial-coupled mode space for quantum transport simulation in nanoscale double-gate MOSFETs. In: 2010 International conference on microelectronics (ICM), pp. 303–306 (2010)
42. Sabry, Y.: Simulation of quantum transport in nanoscale devices. M.Sc. Dissertation, Electronics and Communication Engineering Department, Ain Shams University, Cairo (2009)
43. Datta, S.: Nanoscale device modeling: the Green's function method. *Superlattice Microst.* **28**, 253–278 (2000)

44. El-Banna, M.M., et al.: “Partial-coupled mode-space” a new approach for efficient simulation of ballistic quantum transport in multi-gate devices. *J. Am. Sci.* **9**, 329–338 (2013)
45. International Technology Roadmap for Semiconductor (ITRS) [online]. Available <http://www.itrs.net>
46. Carter, T.A., et al.: Linear algebra, an introduction to linear algebra for pre-Calculus students. Available <http://ceee.rice.edu/publications.html>, May 1995

# **Chapter 2**

## **Model for Quantum Confinement in Nanowires and the Application of This Model to the Study of Carrier Mobility in Nanowire FinFETs**

**Arif Khan, Saeed Ganji, and S. Noor Mohammad**

**Abstract** FinFETs represent exciting new technology. Nanowire FinFETs are more promising than the bulk-silicon FinFETs. As they have the gate capacitance in closer proximity to the whole of the channel, they control the short-channel effects very well and also suppress the leakage current. Key to the superior performance of these devices is high carrier mobility. The focus of this chapter is the fundamental of this mobility in the framework of the SNM (simple, novel, malleable) model for quantum-confined nanowires. Extensive investigation has been carried out to address the role of quantum confinement and dielectric confinement on mobility enhancement in nanowire FinFETs. Impacts of ionized impurity scattering, acoustic phonon scattering, and dislocation scattering on the carrier mobility have been examined. Calculated results have been compared with available experiments. These results have also been used to suggest possible modifications in the design of nanowire FinFETs.

---

A. Khan

Department of Physics, Neotia Institute of Technology, Management, and Science,  
Diamond Harbour Road, Amira 743 368, 24-Paraganas (South), West Bengal, India

Electrocom Corporation, P. O. Box 60317, MD 20859-0317, USA  
e-mail: [karif123@yahoo.com](mailto:karif123@yahoo.com)

S. Ganji

Electrocom Corporation, P. O. Box 60317, MD 20859-0317, USA  
e-mail: [sganji@electrocomcorp.com](mailto:sganji@electrocomcorp.com)

S.N. Mohammad (✉)

Sciencotech, 780 Girard Street Northwest, Washington, DC 20001, USA  
e-mail: [snmohammad2002@yahoo.com](mailto:snmohammad2002@yahoo.com)

## 2.1 Introduction

FinFET is a typical nonplanar transistor built first at the University of California, Berkeley, on semiconductor-on-insulator. It is a generic fin-based transistor [1] regardless of the number of gates. The conducting channel of it is wrapped around thin silicon “fin,” which constitutes the main structure of the device. The effective length  $L_{\text{chnl}}$  of this channel is about 25 nm or less; it is determined by the dimension of the fin. Among various FinFETs, nanowire FinFETs are very promising. They have capability to suppress short channel effects, to have reduced drain-induced barrier lowering, and to provide excellent scalability. Although nanowires are one-dimensional in character, nanowire FinFETs can have three-dimensional architecture. They can be configured by depositing nanowire on an insulating substrate surface and then creating source and drain contacts at the ends of the nanowire. The Si-FinFETs are quite unique; they have high heat dissipation at the Si substrate, provide no floating body effect, and exhibit low defect density. They possess key advantages of the silicon-on-insulator architecture. They have generally top gate, sidewall gates, and special gate extensions under the silicon substrate. They are field-effect transistors with the gate covering almost the entire body. They are fabricated by employing conventional CMOS-like processes.

Nanowires are attractive building blocks for ultra large-scale hierarchical assembly of integrated (ULSI) nanoelectronic devices [2, 3]. Billions of these nanowire devices can be packed together in an area of a few square centimeters to manufacture ULSI circuits. The thinner the nanowires, the denser is the integration and the higher is the speed of the circuits. Depending on the nanowire surface, the carrier scattering during carrier transport through these nanowires can be suppressed. The size, composition, morphology, and electronic properties of the nanowires are therefore superior enough to ensure the best ULSI operations. Electrical transport would obviously be the center of the ULSI operations. A number of investigations [4–12] carried out so far demonstrate conflicting electrical transport through thin nanowires. The correlation between electrical transport and nanowire structural characteristics has not been established well. There is a lack of consistency among various data. This can however be alleviated with fundamental understanding of the structure–property relationship of nanowires.

The first-principle simulations of carrier momentum relaxation rates for various carrier scatterings [4–12] are the most widely employed routes for the study of carrier mobility and related properties. Despite some variations, they make use of self-consistent Poisson–Schrödinger–Monte Carlo solver to account for carrier scatterings. A recent model for quantum confinement in nanowires [13] is distinctly different from them. This model is new and quite relevant to experiments. It is simple and malleable (extendable). We call it simple, novel, malleable (SNM) model. The Focus of this model has primarily been to establish a conceptual framework between nanowire surface amorphicity and quantum confinement. The SNM model is quite straightforward and suitable even for nonexperts in the field. Our focus in this chapter is therefore the SNM model [13]. The objective of this investigation is to understand scattering by charged impurities, dislocations, and lattice vibrations, and to determine

resistivity and carrier mobility in nanowires in the framework of the SNM model. To quantify this objective, the dependence of carrier mobility on temperature  $T$ , doping density  $N_D$ , and the nanowire diameter  $d_{\text{NW}}$  is examined. For these, surfaces and interfaces, which are key nanowire elements, are carefully studied. It is noteworthy that only a relatively small percentage of atoms are at or near a surface or interface of a bulk material. In contrast, many atoms, even half or more of them, are near the surface (interface) of a nanowire. Because of this, parameters such as energy levels, electronic structure, and reactivity of a nanowire surface are all quite different from those in the interior of a bulk material. Indeed the material properties of a nanowire are quite different from those of a bulk material. For the present study, it is assumed that the nanowires are cylindrical and have radius  $r_D = d_{\text{NW}}/2$ .

## 2.2 Surface Energy

Nanowires are grown on nanoparticles [14, 15]. Surface energy of these nanoparticles is considered to be important for nanowire properties such as carrier transport in nanowires. These nanoparticles could be metals (Fe, Ni, Co, Al, Au, etc.), semiconductors (Si, Ge, SiGe, etc.), oxides ( $\text{Al}_2\text{O}_3$ ,  $\text{SiO}_2$ ,  $\text{CaO}$ , etc.), ceramics, or polymers. Based on available experiments [16–18], the melting temperature  $T_M$  of a nanoparticle is larger for larger diameter. The surface energy of this nanoparticle is dependent on also the nanoparticle diameter. It is substantiated by available experiments. For example, Lamber et al. [19] showed that the surface energy ( $6.0 \pm 0.9 \text{ J/m}^2$ ) of Pd seed in a polymer matrix is three times larger than the surface energy ( $1.808 \text{ J/m}^2$ ) of the Pd bulk. Goldstein et al. [20] found that the surface energy ( $1.74\text{--}2.50 \text{ J/m}^2$ ) of CdS seed is two to three times larger than the surface energy ( $0.75 \text{ J/m}^2$ ) of the bulk CdS. The surface energy of PbS seed is  $2.45 \text{ J/m}^2$ , but the surface energy of PbS bulk [21] only  $0.0383 \text{ J/m}^2$ . The surface energy of free Ag seed [22] is  $7.2 \text{ J/m}^2$ , but the surface energy of the Ag bulk is only  $1.065\text{--}1.54 \text{ J/m}^2$ . All these demonstrate that the surface energy increases with decreasing nanoparticle diameter. For the present study, the surface energy of a nanoparticle is modeled by making use of coordination number  $Z_b$  of the nanoparticle's bulk atoms, coordination number  $Z_s$  of the nanoparticle's surface atoms, and the diameter  $d_{\text{atom}}$  of the nanoparticle atoms to arrive at the formula [23]

$$\gamma_{\text{nano}} = \gamma_{\text{atom}}(1 + \alpha_{\text{nano}}T) \left[ 1 + \frac{\lambda_{\text{nanob}}d_{\text{atom}}}{D_{\text{nano}}} \right], \quad (2.1)$$

where [24, 25]

$$\gamma_{\text{atom}} = \left[ \frac{2 - (Z_s/Z_b) - (Z_s/Z_b)^{1/2}}{2N_T S_{\text{area}}} \right] E_{\text{coh}}. \quad (2.2)$$

**Table 2.1** Various parameters used for FECAs for the present calculations

Nanoparticle	$E_{\text{coh}}$ (kJ/mol)	$a_L$ (Å)	$d_{\text{atom}}$ (Å)	$\gamma_{\text{nano}}$ (J/m <sup>2</sup> )	$T_{\text{melt}}$ (°C)	$\Omega_{\text{nano}}$ (cm <sup>3</sup> /mol)
Ni	428	3.58	2.48	2.38	1,453	6.59
Fe	413	2.87	2.52	2.42	1,535	7.09
Co	424	2.53	2.50	2.80	1,495	6.67
Re	775	2.76	2.74	2.00	3,180	8.81
Ag	284	4.18	2.88	1.25	962	10.335
Mn	282	3.53	2.70	1.54	1,245	7.35
Pt	564	4.02	2.78	2.49	1,772	9.10
Pd	376	3.85	2.74	2.00	1,552	8.78
Au	368	4.20	2.92	1.51	1,064	13.60
Al	327	4.05	2.86	1.14	660	9.50
Ca	178	5.62	3.94	0.50	839	29.9
Si	446	7.71	2.64	1.14	1,410	12.058
Ge	372	8.10	2.74	0.88	937	13.62

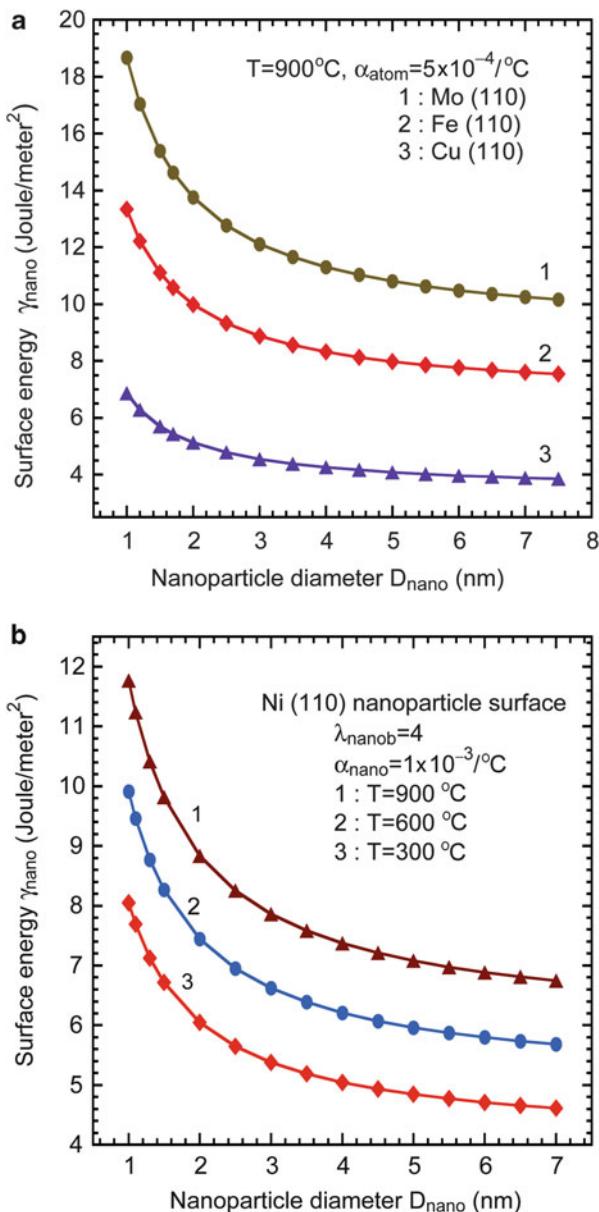
Bulk surface energy is  $\gamma_{\text{nano}}$ , bulk cohesive energy is  $E_{\text{coh}}$ , lattice constant is  $a_L$  (which, is the lattice constant  $a$  of the conventional lattice constants,  $a$ ,  $b$ , and  $c$ ), atom diameter is  $d_{\text{atom}}$ , nanoparticle diameter is  $D_{\text{nano}}$ , melting temperature is  $T_{\text{melt}}$ , and atomic volume is  $\Omega_{\text{nano}}$

$E_{\text{coh}}$  is the cohesive energy of an atom in the bulk,  $\lambda_{\text{nano}}$  is a suitable parameter,  $\alpha_{\text{nano}}$  is the coefficient of thermal stress of the seed, and  $N_T$  is the Avogadro number. Also,  $Z_s = 9$ ,  $Z_b = 12$ , and  $S_{\text{area}} = \sqrt{3}a_L^2/4$  for the (111) surface;  $Z_s = 6$ ,  $Z_b = 12$ , and  $S_{\text{area}} = \sqrt{2}a_L^2/2$  for the (110) surface;  $Z_s = 8$ ,  $Z_b = 12$ , and  $S_{\text{area}} = a_L^2/2$  for the (100) surface; and  $Z_s = 9$ ,  $Z_b = 12$ , and  $S_{\text{area}} = \sqrt{3}a_L^2/2$  for the (0001) surface, where  $a_L$  is the lattice constant of the atoms of the nanoparticle. We calculated the surface energy of nanoparticles by making use of (2.1) and (2.2) and the parameters listed in Table 2.1. Variation of surface energy with the nanoparticle diameter  $D_{\text{nano}}$  for Mo (110), Fe (110), and Cu (110) nanoparticles are shown in Fig. 2.1a. Variations of surface energy with the nanoparticle diameter  $D_{\text{nano}}$  for Ni (110) nanoparticle at 300, 600, and 900 °C, respectively, are shown in Fig. 2.1b. Both Figs. 2.1a, b indicate that the smaller the nanoparticle, the larger is the surface energy, and that the surface energy increases with increasing temperature. They are all consistent with the available experiments [19–22].

## 2.3 Thermodynamic Imbalance

Nanoparticles experience thermodynamic imbalance  $\Xi_{\text{surf}}$  during nanowire growth at the growth temperature  $T$ . Owing to very small dimension, nanoparticles have large surface-to-volume ratio. Also, they have bulk atoms surrounded all around by neighboring atoms. Obviously, the surface atoms have fewer neighboring atoms, but the bulk atoms have larger number of neighboring atoms. All the atoms of a

**Fig. 2.1** Dependence of surface energy  $\gamma_{\text{nano}}$  on the diameter  $D_{\text{nano}}$  of (a) Mo (110), Fe (110), and Cu (110) nanoparticles at 300 °C, 600 °C and 900 °C and (b) Ni (110) nanoparticle at 900 °C



nano particle have cohesive energy, which is the energy shared by a bond of an atom with a bond of the neighboring atom. Surface atoms, with fewer neighboring atoms and fewer bonds, have lower cohesive energy. So the cohesive energy of a surface atom is lower even than the surface energy of an atom in the core of this nanoparticle. The cohesive energy of an atom is proportional as well to the thermal energy required

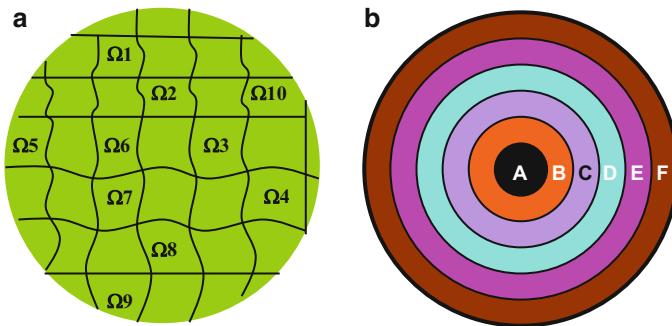
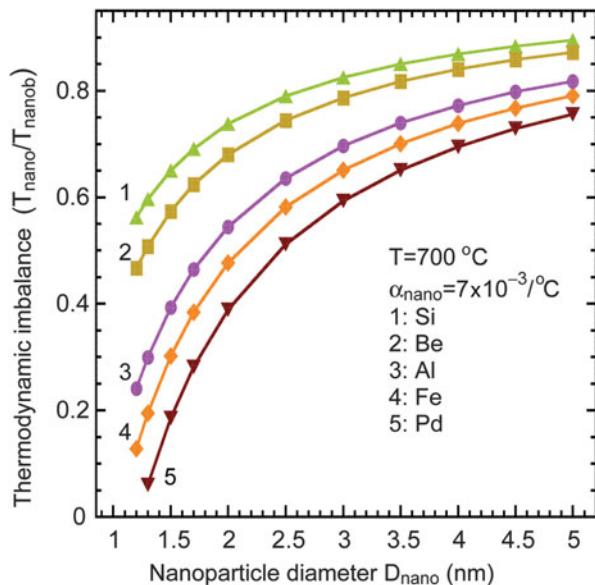
to free this atom from the nanoparticle. This means the surface atoms with lower cohesive energy have higher tendency to dissociate from the nanoparticle. Following Lindemann's criterion [26], cohesive energy of a nanoparticle is proportional also to the melting temperature of this nanoparticle. Therefore, a nanoparticle with lower cohesive energy has lower melting point at the peripheral surface than in the core of it. This nanoparticle may, in general, be assumed to be spherical and to have a diameter  $D_{\text{nano}}$ . The nanoparticle atoms may also be spherical. Let the number of these atoms be  $n_0$ , the coefficient of average surface energy of the nanoparticle, as given by (2.1),  $\gamma_{\text{nano}}$ , the melting point of the nanoparticle bulk  $T_{\text{nanob}}$ , and the melting point of the nanoparticle surface  $T_{\text{nano}}$ . Then the ratio  $T_{\text{nano}}/T_{\text{nanob}}$  may be given by [27]

$$\frac{T_{\text{nano}}}{T_{\text{nanob}}} = 1 - \frac{\pi \gamma_{\text{nano}} d_{\text{atom}}^3 (1 + \alpha_{\text{nano}} T)}{E_{\text{coh}} D_{\text{nano}}}. \quad (2.3)$$

We used (2.3) to study the variations of the thermodynamic imbalance  $T_{\text{nano}}/T_{\text{nanob}}$  with the nanoparticle diameter  $D_{\text{nano}}$  for Si, Be, Al, Fe, and Pd nanoparticles. The results are presented in Fig. 2.2. Various parameters used for the calculations for  $T_{\text{nano}}/T_{\text{nanob}}$  are listed in Table 2.1. Figure 2.2 indicates that the thermodynamic imbalance of a nanoparticle varies inversely with the nanoparticle diameter  $D_{\text{nano}}$ . Although not shown in this figure, the thermodynamic imbalance varies however directly with the temperature  $T$  and the surface energy  $\gamma_{\text{nano}}$ . And these are in line with the observations by Coombes [28], who found  $T_{\text{nano}}/T_{\text{nanob}}$  to increase with increasing diameter  $D_{\text{nano}}$ . Coombes found, for example, for Pb nanoparticle, that  $T_{\text{nano}}/T_{\text{nanob}} \approx 0.65$  for  $D_{\text{nano}} = 10$  nm, but  $T_{\text{nano}}/T_{\text{nanob}} \approx 1.0$  for  $D_{\text{nano}} = 80$  nm. Similarly, Nanda [29] found for Bi nanoparticle that  $T_{\text{nano}}/T_{\text{nanob}} \approx 0.7$  for  $D_{\text{nano}} = 6$  nm, but  $T_{\text{nano}}/T_{\text{nanob}} \approx 0.93$  for  $D_{\text{nano}} = 20$  nm.

If the thermodynamic imbalance leads to lattice disturbance and lattice disorder in the entire seed, the nanoparticle may be fractured into grains, which are shown schematically as  $\Omega 1, \Omega 2, \Omega 3, \Omega 4, \Omega 5, \Omega 6, \Omega 7, \Omega 8$ , etc. in Fig. 2.3a. Alternatively, the nanoparticle, particularly of circular cross-section, may have concentric regions A, B, C, D, E, and F (see Fig. 2.3b), and each of these regions experiences some distinct thermodynamic imbalance. No doubt, the thermodynamic imbalance of the F region at the peripheral surface is the highest, and the thermodynamic imbalance of the A region at the core is the lowest. The E, D, C, and B regions have decreasing thermodynamic imbalance. The A, B, C, D, E, and F regions may be distributed uniformly throughout the nanoparticle or just in locations close to its peripheral surface. They may also be distributed nonuniformly. Interestingly, these are all supported by available experiments. We cite one example. While carrying out laser ablation of a compressed mixture of carbon, BN, SiO, and Li<sub>3</sub>N, Zhang et al. [30] observed the formation of a number of concentric shells. They also observed that each of these shells has distinct material composition and plausibly thermodynamic imbalance and electrical potential.

**Fig. 2.2** Variation of thermodynamic imbalance  $T_{\text{nano}}/T_{\text{nanob}}$  with the nanoparticle diameter  $D_{\text{nano}}$  for Si, Be, Al, Fe, and Pd nanoparticles at 700 °C



**Fig. 2.3** Schematic diagram of nanoparticles mediating nanowire growth. Nanoparticle (a) fractured under the influence of thermodynamic imbalance and (b) exhibiting concentric regions distributed uniformly throughout the seed structure. The thermodynamic imbalance and surface energy are different in regions A, B, C, D, E, and F. Each of the regions A, B, C, D, E, F feels the same temperature as differently

## 2.4 Nanowire Surface Disorder

We assumed that the melting temperature, the diffusion activation energy, the Debye temperature, and the vacancy formation energy of a nanoparticle are  $T_{\text{melt}}$ ,  $E_{\text{act}}$ ,  $\Theta_{\text{nano}}$ , and  $E_{\text{vac}}$ , respectively. Similarly, the melting temperature, the diffusion activation energy, the Debye temperature, and the vacancy formation energy of the corresponding bulk (core) of the nanoparticle are  $T_{\text{meltb}}$ ,  $E_{\text{actb}}$ ,  $\Theta_{\text{nanob}}$ , and  $E_{\text{vacb}}$ , respectively. The Lindemann's criteria [26] is then given by

$$\frac{T_{\text{nano}}}{T_{\text{nanob}}} = \frac{T_{\text{melt}}}{T_{\text{meltb}}} = \frac{E_{\text{act}}}{E_{\text{actb}}} = \frac{\Theta_{\text{nano}}}{\Theta_{\text{nanob}}} = \frac{E_{\text{vac}}}{E_{\text{vacb}}}. \quad (2.4)$$

Due to thermodynamic imbalance, the degree of fluctuation  $\Xi_{\text{surf}}$  at the nanoparticle surface is related to the degree of fluctuation  $\Xi_{\text{core}}$  at the nanoparticle core.  $\Xi_{\text{core}}/\Xi_{\text{surf}}$  is given by [23]

$$\frac{T_{\text{nano}}}{T_{\text{nanob}}} = \frac{\Xi_{\text{core}}}{\Xi_{\text{surf}}}. \quad (2.5)$$

For the sake of convenience, a parameter  $\mathfrak{R}$  may be defined as [23]

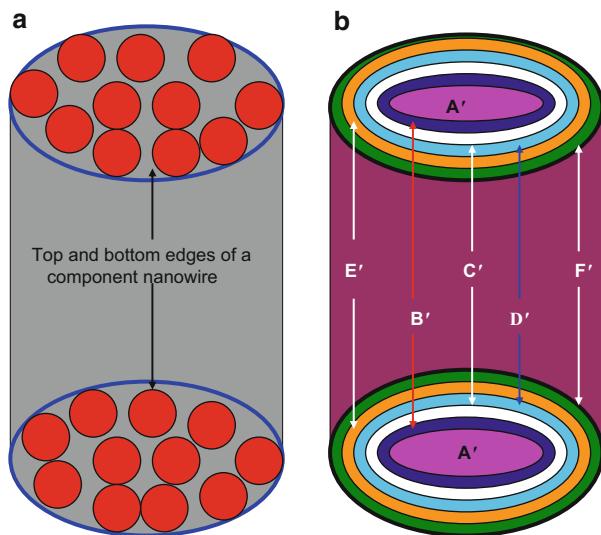
$$\mathfrak{R} = \pi \gamma_{\text{nano}} d_{\text{atom}}^3 \left( \frac{1 + \alpha_{\text{nano}} T}{E_{\text{cohb}} D_{\text{nano}}} \right). \quad (2.6)$$

This allows us to critically examine (2.5) and (2.6) in the light of the finding that the surface energy  $\gamma_{\text{nano}}$  at the peripheral surface of a nanoparticle is larger than that at the bulk of this nanoparticle. What we find from this is that even though the temperature in the entire nanoparticle is identical, the surface periphery of this nanoparticle feels this temperature to be higher than the temperature of its central core. Equation (2.5) indicates that the surface periphery of the nanoparticle has higher degree of fluctuation than the core of this nanoparticle. Due to this fluctuation, the lattice at the peripheral surface of the nanoparticle may be looser than that of its core. As  $E_{\text{vac}} < E_{\text{vacb}}$  [see (2.4)], the peripheral surface of the nanoparticle may have lattice vacancies. The peripheral surface may therefore be disturbed, disordered, amorphous, semi-amorphous, or amorphous-like, which is induced onto the nanowire thus produced on it. By disturbed, disordered, amorphous X region of a nanowire we mean that a region, which appears even to be single-crystalline, has interatomic interactions weaker than those in the central bulk (core). If the amorphicity of the X region ( $X = A, B, C, D, E$ , and  $F$ ) be denoted in general by  $\alpha_{\text{amor}}(X)$ , then in Fig. 2.2b,  $\alpha_{\text{amor}}(F) > \alpha_{\text{amor}}(E) > \alpha_{\text{amor}}(D) > \alpha_{\text{amor}}(C) > \alpha_{\text{amor}}(B) < \alpha_{\text{amor}}(A)$ .

Equation (2.5) indicates that the melting point of the peripheral surface of a nanoparticle is much lower than the melting point of the central core of it. It suggests that, while the peripheral surface of the nanoparticle is semimolten, the central core of it is solid. The bulk diffusion of the nanowire species (e.g., Si atoms for Si nanowire growth, Ga and N atoms for GaN nanowire growth, and Zn and O atoms for ZnO nanowire growth) into the nanoparticle depends on the nanoparticle's structural condition. This diffusion is higher in the molten (semimolten) peripheral surface than in the solid central core of a nanoparticle.

Recall that nanowires are created on nanoparticle when it is converted into droplet. Depending on growth condition, this droplet may though have many different characteristics [31]. If the nanoparticle employed for nanowire growth is fractured and has grains as shown in Fig. 2.3a, thin component nanowires are grown from each of the grains  $\Omega_1, \Omega_2, \Omega_3, \Omega_4$ , etc. of the nanoparticle. The nanowire produced by the entire nanoparticle thus consists of a number of thin component

**Fig. 2.4** Schematic diagrams of nanowires produced on disturbed (disordered) nanoparticles; nanowires (**a**) comprising thin, interacting component nanowires and (**b**) having concentric shells of varied amorphosities. These concentric shells are denoted by A', B', C', D', E', and F', respectively



nanowires, which may be aligned in the same direction, as shown in Fig. 2.4a. However, if the nanoparticle seed has concentric A, B, C, D, E, and F regions of surface energies  $E_{AS}$ ,  $E_{BS}$ ,  $E_{CS}$ ,  $E_{DS}$ ,  $E_{ES}$ , and  $E_{FS}$ , respectively ( $E_{AS} < E_{BS} < E_{CS} < E_{ES} < E_{FS}$ ), as shown in Fig. 2.4b, the characteristics of the A, B, C, D, E, and F regions are induced into the nanowire grown from this nanoparticle. The nanowires produced by the nanoparticle have different lattice structure and interatomic bonding in different concentric regions A', B', C', D', E', and F', as shown in Fig. 2.4b. Each of them has stress and amorphicity (semi-amorphicity, amorphous-like feature), different from those of others.

Wang et al. [32] carried out a comparative study of the CVD growth of carbon nanotubes employing iron/tantalum (Fe/Ta) and iron/silicon dioxide (Fe/SiO<sub>2</sub>) catalysts. Among these catalysts, Ta has a high surface energy [e.g.,  $\sim 5.01 \text{ J/m}^2$  in the (111) surface], but SiO<sub>2</sub> has a low surface energy (e.g.,  $\sim 0.32 \text{ J/m}^2$  at 1,000 °C). Yang et al. [32] found that under identical growth conditions, densities, size distributions, and morphologies of the Fe nanoparticles on Ta were distinctly different from those of the Fe nanoparticles on SiO<sub>2</sub>. Obviously Ta and SiO<sub>2</sub> had different surface energies. Fe nanoparticles on Ta support had higher surface energy induced by the Ta support. But Fe nanoparticles on SiO<sub>2</sub> support had lower surface energy induced by the SiO<sub>2</sub> support. As a result, the carbon nanotube growth catalyzed by Fe/SiO<sub>2</sub> had a low growth rate of less than 100 nm/min. In contrast, the carbon nanotube growth catalyzed by Fe/Ta had a high growth rate exceeding 1 μm/min. This experiment justifies the basic tenet of the SNM model that a nanowire may have different characteristics in different regions if it is produced on a nanoparticle having different characteristics in different regions.

Obviously, the thermodynamic states and the degree of amorphicity of the nanowire would depend on experimental conditions. They would, in a FinFET environment, depend also on the highly tensile stress and strain. While the thermodynamic states may be high in some nanowires, they may be low in some other nanowires. They may even be absent in some other nanowires. Ho et al. [33] grew nanowires made of thin component nanowires. On the other hand, Cui et al. [34], Dan et al. [35], and Bogart et al. [36] produced nanowires exhibiting crystalline core, but amorphous shell. The amorphous shell had plausibly component shells of different amorphicities.

Available results from photoemission and X-ray absorption studies [37] shed light on the long-range order of Si. These studies show that this long-range order undergoes gradual degradation in going from bulk Si to Si nanowires to porous silicon. Microscopic ab initio calculations by Bruno et al. [38] on the diameter-dependent electronic and excitonic band gaps of Si nanowires ( $d_{\text{NW}} < 2 \text{ nm}$ ) indicate that the dimension-dependent Si nanowire excitonic band gap is in good agreement with the dimension-dependent excitonic band gap of porous silicon derived from photoluminescence experiments [39, 40]. The calculations by Bruno et al. [38] made use of density functional theory (DFT) within the local density approximation. These calculations suggest that thin nanowires do exhibit disturbed, disordered, amorphous, semi-amorphous, or amorphous-like porous structure. They are more amorphous if they are thinner.

The energy band gap of a semiconductor material is dependent on its amorphicity. In general, the higher the amorphicity, the larger is the energy band gap of this material. Street [40] found that single-crystal Si has room-temperature energy band gap of 1.12 eV; but highly amorphous Si (a-Si) has room-temperature energy band gap of 1.6 eV. Experiments by Ma et al. [41], and calculations by Nolan et al. [42] and Scheel et al. [43] indicate that the energy band gap of Si nanowires is higher for thinner nanowire. These investigations (e.g., experiments and calculations) imply that thinner nanowires are more amorphous or amorphous-like than thicker nanowires.

## 2.5 Quantum Confinement

Note that the electron motion in thin nanowires is restricted in directions normal to the nanowire axis. Also, the quantum confinement quantizes the energies associated with the in-plane motion of carriers. The lowest level of these energies may, roughly be given by [44]

$$\Delta E \approx \frac{\pi^2 \hbar^2}{m_e^* d_{\text{NW}}^2}, \quad (2.7)$$

where  $m_e^*$  is the in-plane effective mass of electrons. The motion of carriers is not restricted along the nanowire axis, and as a result, the electrons have a dispersion relation [44] for small  $k_l$

$$E_{nm}(k_l) = \varepsilon_{nm} + \frac{\hbar^2 k_l^2}{2m_l^*}, \quad (2.8)$$

where  $\varepsilon_{nm}$  is the quantized energy level (at  $k_l = 0$ ) determined by two quantum numbers ( $n, m$ ),  $k_l$  is the wave vector of the electron wave functions, and  $m_l^*$  is the effective mass for electrons in the  $l$ th subband for electrons flowing along the nanowire axis. Again, the quantized subband energy  $\varepsilon_{nm}$  and the effective mass  $m_l^*$  are along the nanowire axis; they are the two most important parameters governing almost every electronic property of nanowires. The effective-mass density of states  $D_{\text{elec}}$  for electrons, as a function of the energy  $E$ , may be [45, 46]

$$D_{\text{elec}}(E) = \frac{\sqrt{2m_e^*}}{\pi\hbar} \left[ \frac{1}{E - E_{nm}} \right]^{1/2}. \quad (2.9)$$

There can be an analogous formula for the density of states for holes.

The structural disorder [37] of a nanowire influences the luminescence efficiency of this nanowire. This luminescence efficiency is higher in amorphous silicon nanowire than in crystalline silicon nanowire. The luminescence depends also on the structural dimension. If this dimension is very small, there is enhancement in emission intensity. And this enhancement may be explained by quantum confinement [47–50], which is quite significant for dimension smaller than the thermal de Broglie wavelength of carriers. If  $k_B$  is the Boltzmann constant and  $T$  is the temperature, the de Broglie wavelength is  $\lambda_{\text{DB}} = \hbar/\sqrt{2m_e^*k_B T}$ ;  $\lambda_{\text{DB}}$  is approximately 150 Å at room temperature. We emphasize again that, in quantum-confined nanowires, a reduction in the dimension accompanies restructuring of the density of states. Due to this dimension reduction, there occurs an upward shift of the conduction band edge  $E_C$  and a downward shift of the valence band edge  $E_V$ . The energy band gap  $E_G(d_{\text{NW}}) = E_C - E_V$  of amorphous Si (e.g., a-Si) nanowire is, as a result, be given by [13]

$$E_G(d_w) = E_{\text{GB}} + \frac{\pi^2 \hbar^2}{2d_{\text{NW}}^b} \left( \frac{1}{m_e^*} + \frac{1}{m_h^*} \right), \quad (2.10)$$

where  $m_e^*$  and  $m_h^*$  are the in-plane electron and hole effective masses,  $b$  is a parameter, and  $E_{\text{GB}}$  is the energy band gap of the corresponding bulk. First principle calculations are generally based on  $b = 2$ . To be specific, Fonoferov and Balandin [8] considered bulk-like phonons and observed that the decrease in room-temperature electron mobility  $\mu_{\text{FE}}$  with decreasing nanowire diameter follows the proportionality:

$$\mu_{\text{FE}} \propto d_{\text{NW}}^2.$$

Scheel et al. [43], however, found that  $b$ , for Si nanowires, may be 1.4 for the [001] orientation, 1.8 for the [110] orientation, and 1.3 for the [112̄] orientation, depending on effective mass of the orientation.

An important attribute of the SNM model [13] is that quantum confinement influences the energy levels differently in different concentric regions of a nanowire. It is true particularly in regions at and near the peripheral surface. An increase in quantum confinement of the concentric regions at and near this peripheral surface arises from an increase in surface energies and thermodynamic imbalances of the concentric regions. In principle, quantum confinement influences every electronic state within a nanowire. However, due to differences in surface energies and thermodynamic imbalances of the different concentric regions of the nanowire, quantum confinement differently influences the electronic state of different concentric regions within the nanowire. Also, carriers are confined differently in different concentric regions of a nanowire that have different energy levels within the nanowire. Recently Yi et al. [51] studied the impact of quantum confinement on channel conductance and transconductance in ultrathin Si nanowire transistors. They observed that one-dimensional (1D) subband structures are quantized and have one-dimensional density of states. They also observed regions of different channel conductance within the nanowire. Niquet et al. [52] found quasiparticle subband structure as key to the understanding of charge transport in semiconductor nanowires. They noted that current–voltage or conductance characteristics are directly related to the subband energies. These may all be consistent with the basic concept of concentric regions in thin nanowires.

Dunstain and Boulitrop [53] suggested that the enhanced luminescence process in amorphous materials is the result of recombination of hole–electron pairs in the deepest band-tail states, and that it is distributed randomly in space. It is exponential in energy within a critical volume. If  $E_L$  is the luminescence energy,  $E_U$  is the Urbach energy, and  $a_U$ ,  $b_U$  are constants ( $a_U = 50.56 \pm 0.06$  eV and  $b_U = 3.11 \pm 0.40$ ), which depend on the critical radius, then following Searle and Jackson [54]

$$E_G - E_L = a_U + b_U E_U. \quad (2.11)$$

This equation, together with (2.10), yields

$$E_L(d_w) = E_{G0} + \frac{\pi^2 \hbar^2}{2d_W^b} \left( \frac{1}{m_e^*} + \frac{1}{m_h^*} \right) - a_U - b_U E_U(d_{NW}). \quad (2.12)$$

Equation (2.12) indicates that the luminescence varies inversely with the nanowire dimension, but directly with the quantum confinement. Interestingly, if the increase in  $E_U$  with decrease in nanowire diameter  $d_{NW}$  is linear due to enhanced structural disorder, then choosing  $m_e^* = 0.2$  and  $m_h^* = 0.1$ , the luminescence energy predicted by (2.12) is in line with the experimental results given by Giorgis et al. [55]. This is a good demonstration of a relationship between amorphicity and quantum confinement, which is a key to the SNM model [13].

## 2.6 Energy Band Gap as Function of Nanowire Diameter

Amorphicity of the peripheral surface regions, together with the size and surface effects dictate physical properties of nanowires. Also note that the broken bond at the surface and the surface-relaxation-induced strain at the surface contribute to the surface reconstruction [56], which increase the cohesive energy. There is an increase in the elastic energy of inner atoms [57] by the surface stress. Electrons in the conduction band and holes in the valence band are, as a result, spatially confined by the potential barrier of the surface. A decrease in the nanowire diameter leads to an increase in the surface-to-volume ratio  $\delta$  and the internal pressure  $P_{\text{in}}$ . The immediate result of this is the disturbance, weakening of interatomic interactions, and even generation of thin parallel component nanowires, each of them of distinct electronic properties, such as energy level, energy band, and density of states, and all of them within the parent nanowire. Each of these component nanowires is quantum confined. The energetic state of the nanowire atoms near the nanowire surface is also modified. And it gives rise to an increase in the transition energy from the valence band to the conduction band leading to an increase in the energy band gap. The mean cohesive energy  $E_{\text{coh}}$  and the bulk cohesive energy  $E_{\text{coh}}^b$  per atom of a nanomaterial, based on the liquid drop model [58, 59], are related by

$$E_{\text{coh}} = E_{\text{coh}}^b - \left( \frac{3v_0\gamma_{\text{surf}}}{r_{\text{nano}}} \right), \quad (2.13)$$

where  $\gamma_{\text{surf}}$  is the surface energy per unit area,  $r_{\text{nano}}$  is the radius, and  $v_0$  is the atomic volume of the nanomaterial. If this material is spherical,  $E_{\text{coh}}$  is in eV,  $r_{\text{nano}}$  in nm, and  $\gamma_{\text{surf}}$  is the surface energy in eV/nm<sup>2</sup>, then  $v_0 = 4\pi r_A^3/3$ , where  $r_A$  as the radius of each atom of the nanomaterial. Also, the total surface energy is

$$E_{\text{surf}} = 4\pi\gamma_{\text{surf}}r_A^2. \quad (2.14)$$

If the nanomaterial is nanowire,  $r_{\text{nano}} = r_D$ . Equations (2.13) and (2.14), together with the relation  $v_0 = 4\pi r_A^3/3$ , give [13]

$$\frac{E_{\text{coh}}}{E_{\text{coh}}^b} = 1 - \left( \frac{r_A E_{\text{surf}}}{r_{\text{nano}} E_{\text{coh}}^b} \right). \quad (2.15)$$

Equation (2.15) is similar to the equation given by Rose et al. [60, 61]. Based on the bond energy model [62], Rose et al. obtained the equation

$$\frac{E_{\text{coh}}}{E_{\text{coh}}^b} = 1 - 4.5n_A^{-1/3} = 1 - 4.5r_A/r_{\text{nano}}, \quad (2.16)$$

where  $n_A$  is the total number of atoms in the nanomaterial. Equation (2.16) is very similar to (2.17) given by Sun et al. [63], which is

**Table 2.2** List of various parameters defining the diameter dependence ( $b = 1$ ) of nanowire band gap

Nanowire	Energy band gap $E_{\text{GB}}$ (eV)	$E_{\text{coh}}$ (kcal/mol)	$\vartheta$ (eV nm)
GaAs	1.424	155	1.8
InP	1.344	159	1.3
InN	1.9	277	1.8
CdS	2.5	132	1.2
CdTe	1.61	96	1.7
ZnSe	2.822	125	1.5
CdSe	1.74	114	0.8
ZnS	3.68	147	1.2
ZnTe	2.394	106	1.4

$$\frac{E_{\text{coh}}}{E_{\text{coh}}} = 1 - \frac{6r_A}{r_{\text{nano}}}. \quad (2.17)$$

The activation energy  $E_{\text{act}}$  of (2.4) is related to the energy band gap of a nanomaterial, including nanowire, as [56]:

$$\frac{E_{\text{act}}}{E_{\text{actb}}} = 1 - \frac{\Delta E_G}{E_{\text{GB}}}, \quad (2.18)$$

where  $\Delta E_G$  is the change in energy band gap:  $\Delta E_G = E_G - E_{\text{GB}}$ . A comparison of various equations presented above yields

$$\frac{\Delta E_G}{E_{\text{GB}}} = \frac{r_A E_{\text{surf}}}{r_D E_{\text{coh}}}. \quad (2.19)$$

Equation (2.19) may further be simplified to

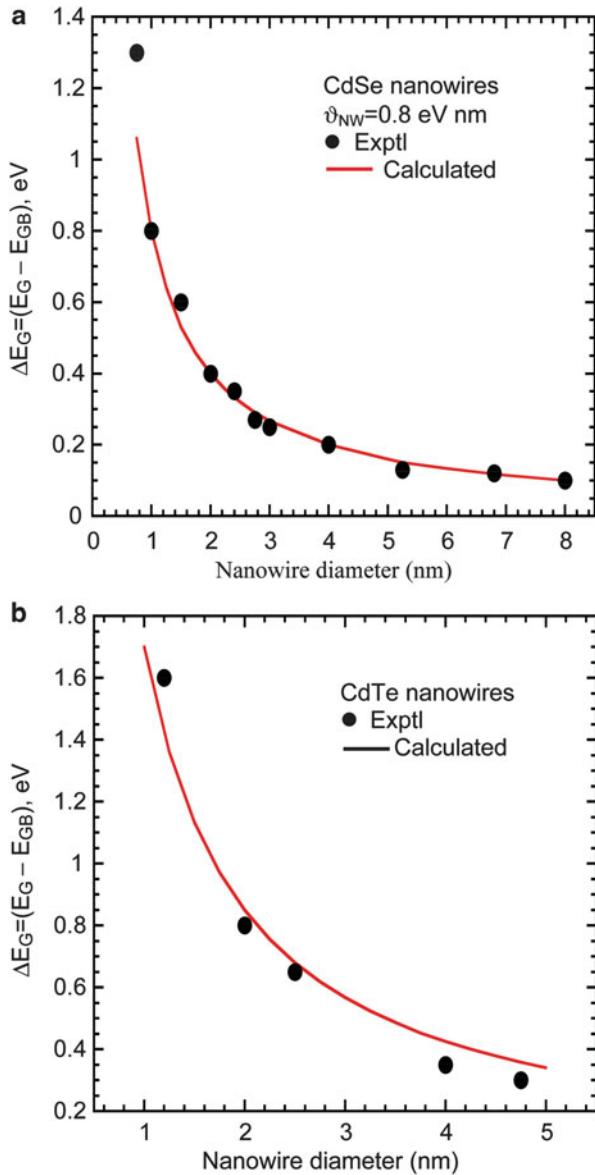
$$E_G = E_{\text{GB}} + \frac{\vartheta_{\text{NW}}}{d_{\text{NW}}^b}, \quad (2.20)$$

where  $b = 1$  and

$$\vartheta_{\text{NW}} = \frac{2E_{\text{GB}}E_{\text{surf}}r_A}{E_{\text{coh}}}. \quad (2.21)$$

Equation (2.20) is identical to the one derived for Si nanowires by Delerue et al. [64] ( $\vartheta_{\text{NW}} = 3.73$  eV nm and  $b = 1.39$ ) and Wu et al. [65]. The observations by Delerue et al. [64] and Wu et al. [65] thus validate (2.20). A striking feature of this equation is that the energy band gap of a nanowire is inversely proportional to the nanowire diameter  $d_{\text{NW}}$ . The parameter  $\vartheta_{\text{NW}}$  has its genesis in  $E_{\text{GB}}$ ,  $E_{\text{surf}}$ ,  $r_A$ , and  $E_{\text{coh}}$ . It may, for example, be obtained from a comparison of (2.20) with the available experimental  $E_G$  vs.  $d_{\text{NW}}$  plots [66, 67]. The  $\vartheta_{\text{NW}}$  values thus obtained with  $b = 1$  are listed in Table 2.2. Some other parameters [68] relevant to  $\vartheta_{\text{NW}}$  for nanowires are also

**Fig. 2.5** Comparison of the observed and the calculated variations of the energy band gap with nanowire diameter ( $b = 1$ ) for  
**(a)** CdSe nanowires and  
**(b)** CdTe nanowires



listed in this table. The experimental and theoretical plots for  $\Delta E_G = E_G - E_{GB}$  for CdSe and CdTe nanowires are compared in Figs. 2.5a, b. These figures show a very good agreement between the calculated and the experimental results. They demonstrate that (2.20) is indeed reasonably good. There is no fixed value for  $b$ . While Wu et al. [65] found it to be 1.1–1.6; Ma et al. [41] found it to be 2.1, and Scheel et al. [43]

found it to be between 1.3 and 1.8. Nevertheless,  $b = 1$  yields results in good agreement with the available experiments.

## 2.7 Formula for Amorphicity

As suggested in Section 2.6, surface amorphicity of a nanowire is a major cause of quantum confinement of this nanowire. An approximate formula for amorphicity is therefore needed to quantify the relationship between amorphicity and quantum confinement. There are many different experiments showing the variation of the amorphicity  $\alpha_{\text{amor}}$  with the energy band gap  $E_G$ . These variations are either linear or near-linear. One of them due to Rotaru et al. [69] for silicon is near-linear and is given by

$$\alpha_{\text{amor}} \approx \frac{(E_G - C_2)}{\sigma_2}, \quad (2.22)$$

where  $C_2$  and  $\sigma_2$  are suitable parameters:  $C_2 \approx 1$  eV and  $\sigma_2 = 0.4$  eV for Si. Equation (2.22) indicates that  $\alpha_{\text{amor}} = 0$  for  $E_G = 1.12$  eV, but  $\alpha_{\text{amor}} = 1$  (e.g., maximum) for  $E_G \approx 1.6$  eV. These are in line with experiment. The highest limit of amorphicity  $\alpha_{\text{amor}}$  for nanowires of smallest diameter ( $d_{\text{NW}} \leq 2$  nm) may be  $\sim 1.0$ . However, even very thick nanowires may have some residual amorphicity, suppose  $\alpha_{\text{amor}} \approx 0.1$ . Taking all these into consideration, we obtain

$$\alpha_{\text{amor}} \approx \sigma_{\text{NW}}^{-1} \left[ E_{\text{GB}} - C_{\text{NW}} + \frac{\vartheta_{\text{NW}}}{d_{\text{NW}}^b} \right], \quad (2.23)$$

where  $\sigma_{\text{NW}}$  and  $C_{\text{NW}}$  are suitable parameters:  $\sigma_{\text{NW}} = 1$  eV,  $b = 1$ .  $C_{\text{NW}} = 1.0$  eV for Si nanowires and 3.3 eV for GaN nanowires. This equation is approximate but reasonably good for  $d_{\text{NW}} \geq 2$  nm. A close look at (2.23) indicates that the amorphicity  $\alpha_{\text{amor}}$  of a nanowire increases with decreasing nanowire diameter, and this increase in the amorphicity  $\alpha_{\text{amor}}$  may be a reflection of the increase in quantum confinement. Indeed Chen et al. [70] noted that the dependence of  $\alpha_{\text{amor}}$  on  $d_{\text{NW}}^b$  [see (2.20)] is essentially identical to the dependence of quantum confinement on  $d_{\text{NW}}^{-b}$ . Chen et al. [70] observed that the quantum confinement in ZnO nanorods depends on  $d_{\text{NW}}^{-1.0}$ , which is identical to that in (2.20) and (2.23). The finding by Chen et al. [70] justifies the validity of (2.20) and (2.23).

The dielectric environment of a nanowire is also important for confinement, which is measured by its dielectric constant. The dielectric constant  $\epsilon_{\text{NW}}$  of a nanowire describes the ability of the nanowire to concentrate electric flux, and this electric flux changes with changes in dielectric environment. A dielectric coating of a thin nanowire blocks the field lines that tend to leak out of it into the surrounding. If  $n_{\text{rin}}$  is the refractive index,  $N_{\text{ion}}$  is the number of ions per unit volume,  $q$  is the electronic charge,  $Z_{\text{tec}}$  is the transverse effective charge,  $m_{\text{ion}}$  is the reduced ion mass, and  $\omega_{\text{TO}}$  is

the frequency of the transverse optical phonon, the dielectric constant  $\epsilon_{\text{NW}}$  of the coated nanowire may be given by [71, 72]

$$\epsilon_{\text{NW}} = n_{\text{rin}}^2 + \frac{N_{\text{ion}}q^2Z_{\text{tec}}^2}{m_{\text{ion}}\omega_{\text{TO}}^2}. \quad (2.24)$$

The dielectric coating tends also to suppress the vibration of the transverse optical phonon of the nanowire. As a result, the frequency  $\omega_{\text{TO}}$  is reduced and the dielectric constant  $\epsilon_{\text{NW}}$  is increased. This increase depends though on the nanowire material and the conditions for the growth of this material. It may therefore be different for different nanowire materials and have different functional forms. One plausible form among them may be [13]

$$\epsilon_{\text{NW}} \approx \epsilon_{\text{NW}0}(1 + \nu\alpha_{\text{amor}}), \quad (2.25)$$

where  $\epsilon_{\text{NW}0}$  is the dielectric constant of the nanowire in the absence of nanowire coating, and  $\nu$  is a parameter.

## 2.8 Models for Carrier Scattering

There may be five different scattering mechanisms for carriers in thin nanowires. These are the ionized impurity scattering, acoustic phonon scattering, dislocation scattering, deformation potential scattering, and surface roughness scattering. Among them, the deformation potential scattering is sensitive to the temperature range  $170 \text{ K} \leq T \leq 280 \text{ K}$ , and the surface roughness scattering is ineffective for very thin nanowires. It is similar in nature as the phonon scattering. The acoustic phonon scattering should also have similar characteristics as the electron–phonon scattering. Considering that the nanowire devices operate at  $T > 273 \text{ K}$ , one may consider primarily the ionized impurity scattering, acoustic phonon scattering, dislocation scattering, and deformation potential scattering for carrier transport in nanowire FinFETs. Carrier mobility due to these scatterings should though be modified by assuming that the amorphicity  $\alpha_{\text{amor}}$  is a rough measure of quantum confinement.

Making use of the model by Lee and Spector [73], the electron momentum relaxation rate for the ionized impurity scattering may be calculated by

$$\tau_{\text{imp}}^{-1}(k_l) = \frac{2\pi m_l^* N_{\text{ion}} r_{\text{D}}^2}{\hbar^3 k_l} \left( \frac{z_{\text{ion}} q^2}{2\pi \epsilon_{\text{NW}}} \right) [\ln(k_l r_{\text{D}})]^2, \quad (2.26)$$

where  $N_{\text{ion}}$  is the concentration of the impurities, and  $z_{\text{ion}}$  is the integral ionic charge,  $z_{\text{ion}} = 1$ . Salfi et al. [74] and Motayed et al. [75] observed that temperature-dependent carrier mobility by impurity scattering in nanowires varies

as  $T^{3/2}$ , which is the same as the temperature-dependent carrier mobility by impurity scattering in the bulk. Cimpoiasu et al. [76] also observed that the doping-dependent carrier mobility in nanowires may be described by a model for the bulk. These are interesting findings suggesting that carrier mobility due to ionized impurity scattering in nondegenerate amorphous-like semiconductor nanowires may also be given by [77]

$$\mu_{\text{ion}} = \frac{128\alpha_{\text{amor}}\sqrt{2}\epsilon_{\text{NW}}^2(k_B T)^{3/2}}{N_{\text{ion}}z_{\text{ion}}^2q^3m_n^{1/2}[\ln(1 + \lambda_{\text{ndg}}) - \lambda_{\text{ndg}}/(1 + \lambda_{\text{ndg}})]}. \quad (2.27)$$

Equation (2.27) is a classical model but takes the effect of quantum confinement into account. Also,

$$\lambda_{\text{ndg}} = \frac{24e_{\text{NW}}m_n(k_B T)^2}{\hbar^2 q^2 n_{\text{sm}}}, \quad (2.28)$$

where  $n_{\text{sm}}$  is the carrier concentration in semiconductor nanowires. If  $\lambda_{\text{deg}}$  is defined by [77]

$$\lambda_{\text{deg}} = \frac{4e_{\text{NW}}\hbar^2(3n_{\text{sm}}\pi^8)^{1/3}}{q^2 m_n}, \quad (2.29)$$

the carrier mobility due to ionized impurity scattering in degenerate semiconductor nanowires would be [77]

$$\mu_{\text{ion}} = \frac{24\alpha_{\text{amor}}\pi^3\epsilon_{\text{NW}}^2\hbar^3 n_{\text{sm}}}{N_{\text{ion}}z_{\text{ion}}^2q^3m_n^2[\ln(1 + \lambda_{\text{deg}}) - \lambda_{\text{deg}}/(1 + \lambda_{\text{deg}})]}. \quad (2.30)$$

Note that  $N_{\text{ion}} \approx N_D$  for a fully ionized doping level, but  $N_{\text{ion}} < N_D$  for partially ionized doping level. The intervalley acoustic phonon scattering rate is given by [78]

$$\tau_{ac}(k_l) = \left[ \frac{k_B T \wp_{ac}^2 \sqrt{m_l^*}}{\sqrt{2}\hbar^2 \rho_{\text{crys}} v_{\text{sound}}^2} \right] \frac{1 + 2\xi_{npb} E_{\text{final}}}{\sqrt{E_{\text{final}}(1 + \xi_{npb} E_{\text{final}})}} \mathfrak{I}_{nm} \Theta(E_{\text{final}}) \quad (2.31)$$

where  $\wp_{ac}$  is the acoustic deformation potential,  $\rho_{\text{crys}}$  is the crystal density,  $v_{\text{sound}}$  is the sound velocity,  $\xi_{npb}$  is the non-parabolity parameter,  $E_{\text{final}}$  is the final kinetic energy of the carriers, and  $\Theta(E_{\text{final}})$  is the Heaviside step function. Also,  $\mathfrak{I}_{nm}$  is the electron–phonon wave function overlap integral. The mobility due to acoustic phonon scattering may alternatively be given by [77]

$$\mu_{\text{aco}} = \frac{\sqrt{8\pi}\epsilon_{\text{NW}}q\hbar^4C_{\text{lec}}}{3\alpha_{\text{amor}}E_{\text{ds}}m_n^{5/2}(k_B T)^{3/2}}, \quad (2.32)$$

where  $C_{\text{lec}}$  is the average longitudinal elastic constant of the nanowire and  $E_{\text{ds}}$  is the displacement of the edge of the band per unit dilation of the lattice.

Nanowires may suffer from dislocation density. The electron mobility due to scattering by dislocations may be given by [79–81]

$$\mu_{\text{disl}} = \frac{30\sqrt{2\pi}\alpha_{\text{amor}}\epsilon_{\text{NW}}^2d_{\text{acep}}^2(k_B T)^{3/2}}{N_{\text{disl}}q^3f_{\text{rate}}^2\lambda_{\text{screen}}m_n^{1/2}}, \quad (2.33)$$

where  $n_{\text{sm}}$  is the net carrier concentration,  $N_{\text{disl}}$  is the dislocation density,  $d_{\text{acep}}$  is the distance between acceptor centers (dangling bonds),  $f_{\text{rate}}$  is the occupation rate of the acceptor centers, and  $\lambda_{\text{screen}}$  is the Debye screening length given by

$$\lambda_{\text{screen}} = \left( \frac{\epsilon_{\text{NW}}k_B T}{q^2 n_{\text{sm}}} \right)^{1/2}. \quad (2.34)$$

Generally  $n_{\text{sm}} = N_D$ . A close look of (2.33) indicates that the electron mobility due to scattering by dislocations increases with increase in net carrier concentration.

## 2.9 Calculated Carrier Mobility

The carrier mobility is often governed by more than one scattering mechanism. Following Matthiesen's rule, the possible carrier mobilities may therefore be calculated with

$$\mu_{\text{inac}}^{-1} = \mu_{\text{ion}}^{-1} + \mu_{\text{aco}}^{-1}, \quad (2.35a)$$

$$\mu_{\text{dsin}}^{-1} = \mu_{\text{disl}}^{-1} + \mu_{\text{ion}}^{-1}, \quad (2.35b)$$

$$\mu_{\text{dsac}}^{-1} = \mu_{\text{disl}}^{-1} + \mu_{\text{aco}}^{-1}, \quad (2.35c)$$

and

$$\mu_{\text{dsacin}}^{-1} = \mu_{\text{disl}}^{-1} + \mu_{\text{aco}}^{-1} + \mu_{\text{ion}}^{-1}. \quad (2.35d)$$

Various parameters used for the calculations of mobilities for silicon nanowires are listed in Table 2.3.

The influence of quantum confinement and dielectric confinement on the carrier mobilities  $\mu_{\text{ion}}$ ,  $\mu_{\text{aco}}$ , and  $\mu_{\text{disl}}$  in silicon nanowires is shown in Fig. 2.6a. One can see from this figure that  $\mu_{\text{ion}}$  and  $\mu_{\text{disl}}$  decrease, while  $\mu_{\text{aco}}$  increases with increasing

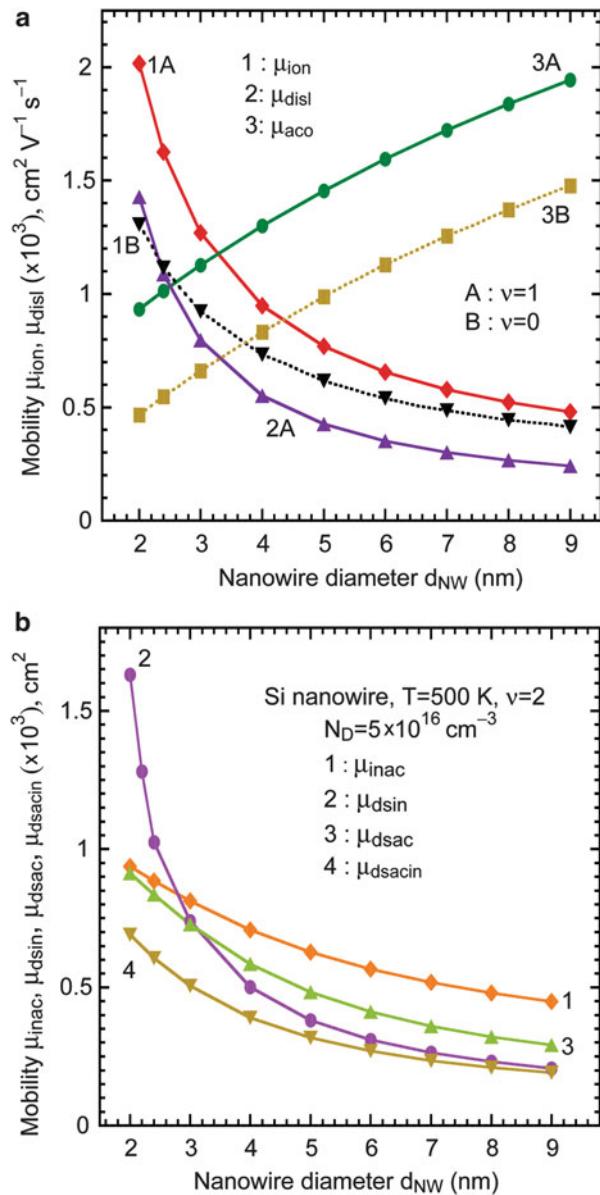
**Table 2.3** Various parameters used for the calculation of carrier mobilities in silicon nanowires

Parameter	Value
Effective electron mass $m_n$	0.98
Dielectric constant in absence of coating $\epsilon_{\text{NW}0}$	11.7
Energy band gap of bulk Si	1.12 eV
Amorphicity parameter $\sigma_{\text{NW}}$	1 eV
Amorphicity parameter $C_{\text{NW}}$	1.0 eV
Integral ionic charge $z_{\text{ion}}$	2.0
Occupation rate of acceptor centers $f_{\text{rate}}$	1.0
Distance between acceptor centers $d_{\text{acep}}$	1.0 nm
Nanowire dislocation density $N_{\text{disl}}$	$1.0 \times 10^{10} \text{ cm}^{-2}$
Nanowire doping density $N_D$	$5 \times 10^{16} \text{ cm}^{-3}$
Ionized impurity level $N_{\text{ion}}$	$5 \times 10^{16} \text{ cm}^{-3}$
Carrier concentration $n_{\text{sm}}$	$5 \times 10^{16} \text{ cm}^{-3}$
Average longitudinal elastic constant $C_{\text{lec}}$	$1.0 \times 10^{17} \text{ cm}^{-3}$
The parameter $\beta_{\text{ds}}$	200 K/T
Displacement of the edge of the band $E_{\text{ds}}$	$(3.0 \times 10^{-5} \beta_{\text{ds}}) \text{ eV}$
Band gap reduction parameter $\vartheta_{\text{NW}}$	1.76 eV nm
Avogadro number $N_T$	$6.02214179 \times 10^{23} \text{ mol}^{-1}$

nanowire diameter  $d_{\text{NW}}$ . The  $\mu_{\text{ion}}$  and  $\mu_{\text{disl}}$  have small values for  $d_{\text{NW}} > 5 \text{ nm}$  but large values for  $d_{\text{NW}} < 3 \text{ nm}$ . The  $\mu_{\text{aco}}$  has, on the other hand, smaller value for  $d_{\text{NW}} < 3 \text{ nm}$  and larger values for  $d_{\text{NW}} > 5 \text{ nm}$ . They are larger for nanowires with quantum confinement (see, Fig. 2.6a, curves 1A and 3A) and dielectric confinement but smaller for nanowires with only quantum confinement (see, Fig. 2.6a, curves 1B and 3B). These curves show that the dielectric confinement and quantum confinement lead together to elevate enhancement in carrier mobility than that just by the quantum confinement. Figure 2.6a depicts that the mobilities  $\mu_{\text{ion}}$  and  $\mu_{\text{disl}}$  increase indeed significantly for  $d_{\text{NW}} < 4 \text{ nm}$ , which is attributed to the Coulomb and the dislocation potentials tuned by the dielectric environment more in thinner nanowire than in thicker nanowire. These are in line with experiments by Cui et al. [82] and Koo et al. [83], who found carrier mobility higher in thin nanowires than in thin films. These are in line with also the calculations by Sakaki et al. [9] and measurements by Takagi et al. [84]. The very trend of the increase in  $\mu_{\text{aco}}$  with increasing nanowire diameter resembles the Monte Carlo calculations by Fonoberov and Balandin [8]. Electron–phonon scattering rate is higher in the presence of spatial confinement than that in the absence of spatial confinement. Thinner nanowire is affected more than thicker nanowire by lattice vibration. Also, the displacement  $E_{\text{ds}}$  is directly proportional to the lattice vibration but inversely proportional to the mobility  $\mu_{\text{aco}}$ . Carrier mobility in nanowire devices depends on electric field. An increase in mobility is caused by decrease in carrier concentration under the influence of this electric field.

Figure 2.6a demonstrates that the carrier mobility in a nanowire is determined by a competition of the ionized impurity scattering and dislocation scattering on one hand and the acoustic phonon scattering on the other. The carrier mobility increases

**Fig. 2.6** Diameter-dependent carrier mobilities. Comparison of (a)  $\mu_{\text{ion}}$ ,  $\mu_{\text{disl}}$ , and  $\mu_{\text{aco}}$  and (b)  $\mu_{\text{inac}}$ ,  $\mu_{\text{dsin}}$ ,  $\mu_{\text{dsac}}$ , and  $\mu_{\text{dsacin}}$  in the absence and the presence of dielectric confinement  $\epsilon_{\text{NW}} = \epsilon_{\text{NW}0}(1 + v\alpha_{\text{amor}})$ ,  $\epsilon_{\text{NW}0} = 11.7$ . Various parameters used for the calculations of  $\mu_{\text{ion}}$  and  $\mu_{\text{disl}}$  are  $z_{\text{ion}} = 2$ ,  $f_{\text{rate}} = 1$ ,  $d_{\text{acep}} = 1 \text{ nm}$ ,  $N_{\text{disl}} = 1.0 \times 10^{10} \text{ cm}^{-2}$ , and  $N_D = 5.0 \times 10^{16} \text{ cm}^{-3}$ . Various parameters used for the calculations of  $\mu_{\text{aco}}$  are  $N_D = 5.0 \times 10^{16} \text{ cm}^{-3}$ ,  $C_{\text{ale}} = 1.0 \times 10^{17} \text{ cm}^{-3}$ ,  $\beta_{\text{ds}} = 200 \text{ K}/T$ , and  $E_{\text{ds}} = (3.0 \times 10^{-5} \beta_{\text{ds}}) \text{ eV}$ , respectively



with increasing nanowire diameter if the acoustic phonon scattering dominates over the ionized impurity scattering and dislocation scattering. It decreases, in contrast, with increasing nanowire diameter. These are all in line with experiments by Motayed et al. [75] and Ford et al. [85]. Both of these experiments indicate that carrier mobility increases with increasing nanowire diameter. Experiment by

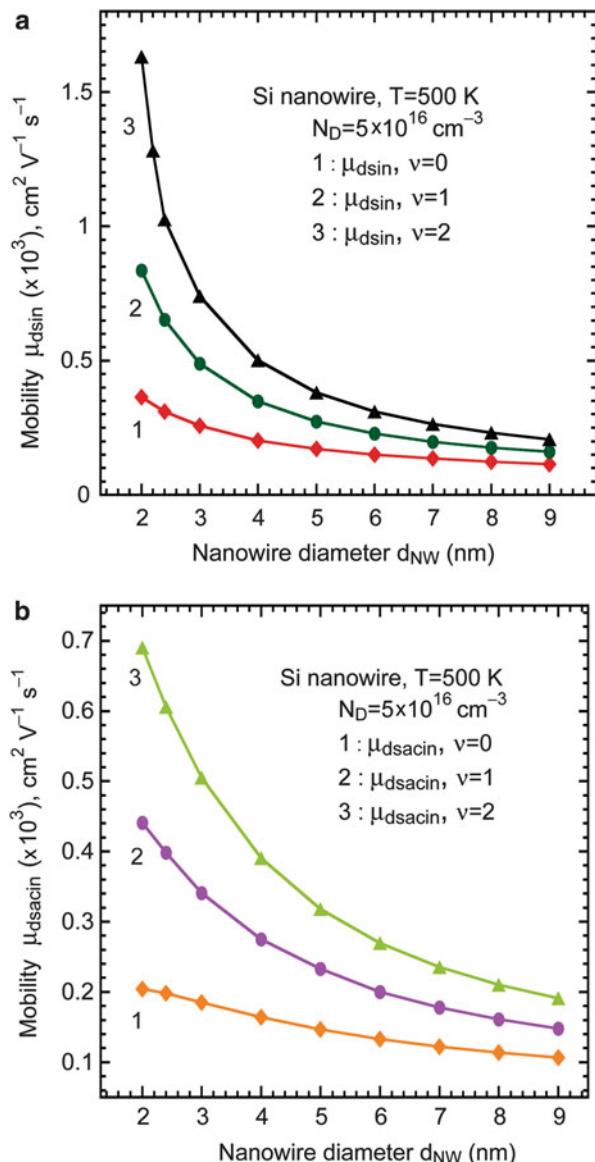
Trivedi et al. [86] depicts the same exact phenomena (e.g., increase in carrier mobility with decreasing nanowire diameter). To be more specific, Trivedi et al. showed that the mobility is  $1,235 \text{ cm}^2/\text{V s}$  for  $d_{\text{NW}} = 3.4 \text{ nm}$ ,  $737 \text{ cm}^2/\text{V s}$  for  $d_{\text{NW}} = 4.3 \text{ nm}$ , and  $300 \text{ cm}^2/\text{V s}$  for the bulk. Figure 2.6b compares calculated mobilities  $\mu_{\text{inac}}$ ,  $\mu_{\text{dsin}}$ ,  $\mu_{\text{dsac}}$ , and  $\mu_{\text{dsacin}}$  as function of the nanowire diameter  $d_{\text{NW}}$ . It indicates that  $\mu_{\text{aco}}$  can indeed be significant, and the mobilities  $\mu_{\text{inac}}$ ,  $\mu_{\text{dsac}}$ , and  $\mu_{\text{dsacin}}$  can be lower than the mobility  $\mu_{\text{dsin}}$  for nanowires of diameter  $d_{\text{NW}} < 4 \text{ nm}$ . The mobility  $\mu_{\text{dsacin}}$  due to the combined effect of ionized impurity scattering, dislocation scattering, and the acoustic phonon scattering is quite low for nanowires of diameter  $d_{\text{NW}} < 4 \text{ nm}$ . It is true even at temperature as high as  $T = 500 \text{ K}$ . These findings are interesting; they suggest that, any nanowire used for FinFETs must be grown in a way that its acoustic phonon scattering is low.

The impact of dielectric confinement on nanowire mobilities is shown in Fig. 2.7a, b. While Fig. 2.7a is for the mobility  $\mu_{\text{ion}}$ , Fig. 2.7b is for the mobility  $\mu_{\text{dsacin}}$ . Both of these figures show that carrier mobilities increase with increasing dielectric confinement [see (2.25)], and this increase is though higher for smaller  $d_{\text{NW}}$  due to quantum confinement.

The influence of doping on the mobilities  $\mu_{\text{ion}}$  and  $\mu_{\text{disl}}$  of nanowires of two different diameters ( $d_{\text{NW}} = 2 \text{ nm}$  and  $5 \text{ nm}$ ) is shown in Fig. 2.8a. One can see that, in both of these nanowires, the higher the doping density, the lower is the mobility  $\mu_{\text{ion}}$ . But in both of them, the higher the doping density, the higher is the mobility  $\mu_{\text{disl}}$ . This supports the measurements by Cimpoiasu et al. [76] which underscored the dominance of impurity scattering over phonon scattering. And this is important for nanowire devices, because these devices from lightly doped nanowires ( $N_D \leq 5 \times 10^{16} \text{ cm}^{-3}$ ) are preferred [87] for ultra large-scale integration. Interestingly, quantum confinement plays an important role in lowering the maximum achievable doping concentration in thin nanowires [88]. Since quantum confined nanowires may not be doped heavily, these nanowires can gratifyingly be free from scattering by dislocations. Figure 2.8b shows the dependence of the carrier mobilities  $\mu_{\text{inac}}$ ,  $\mu_{\text{dsin}}$ ,  $\mu_{\text{dsac}}$ , and  $\mu_{\text{dsacin}}$  on the nanowire doping. It indicates that, due to quantum-confinement-induced suppression of impurity scattering and phonon scattering, lowly doped nanowires ( $N_D < 5 \times 10^{16} \text{ cm}^{-3}$ ) are very suitable for very high carrier mobility  $\mu_{\text{inac}}$ . These lowly doped nanowires may however have low mobility  $\mu_{\text{dsac}}$  due to the combined influence of  $\mu_{\text{aco}}$  and  $\mu_{\text{disl}}$ . The mobility  $\mu_{\text{dsacin}}$  is lowly dependent on the doping concentration, and it is consistent with the observations by Cimpoiasu et al. [76] and Huang et al. [89], which show that mobility versus doping concentration plots are quite insensitive to the choice of doping level.

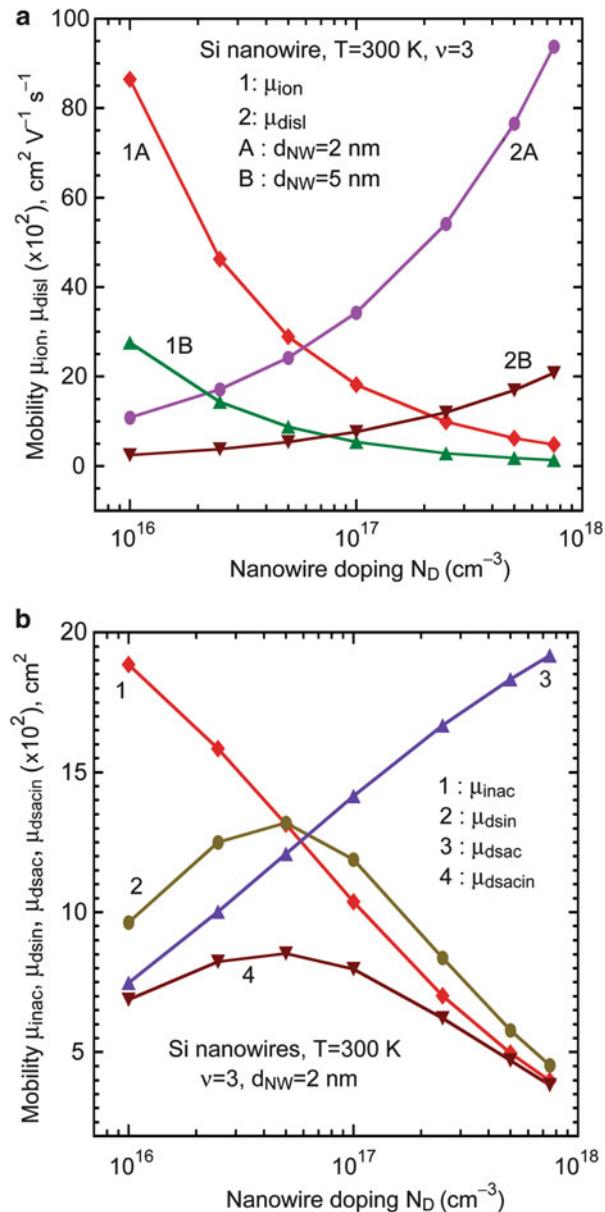
The influence of temperature on carrier mobilities is shown in Fig. 2.9a, b. Figure 2.9a shows that increase in temperature accompanies increase in the mobilities  $\mu_{\text{ion}}$  and  $\mu_{\text{disl}}$  but decrease in the mobility  $\mu_{\text{aco}}$ . Recall that the doping becomes increasingly nondegenerate at higher temperature. The inevitable result of this is the increase in mobility  $\mu_{\text{ion}}$  with increasing temperature. Remarkably, Ford et al. [85], Chang et al. [90], and Stern et al. [91] observed decrease in mobility with increasing temperature. The characteristics of the mobility versus temperature plot

**Fig. 2.7** Impact of dielectric confinement on the carrier mobility in Si nanowires. (a)  $\mu_{dsin}$  vs.  $D_{nano}$  plots and (b)  $\mu_{dsacm}$  vs.  $D_{nano}$  plots for  $v = 0, 1$ , and  $2$ , respectively [see (2.25)]



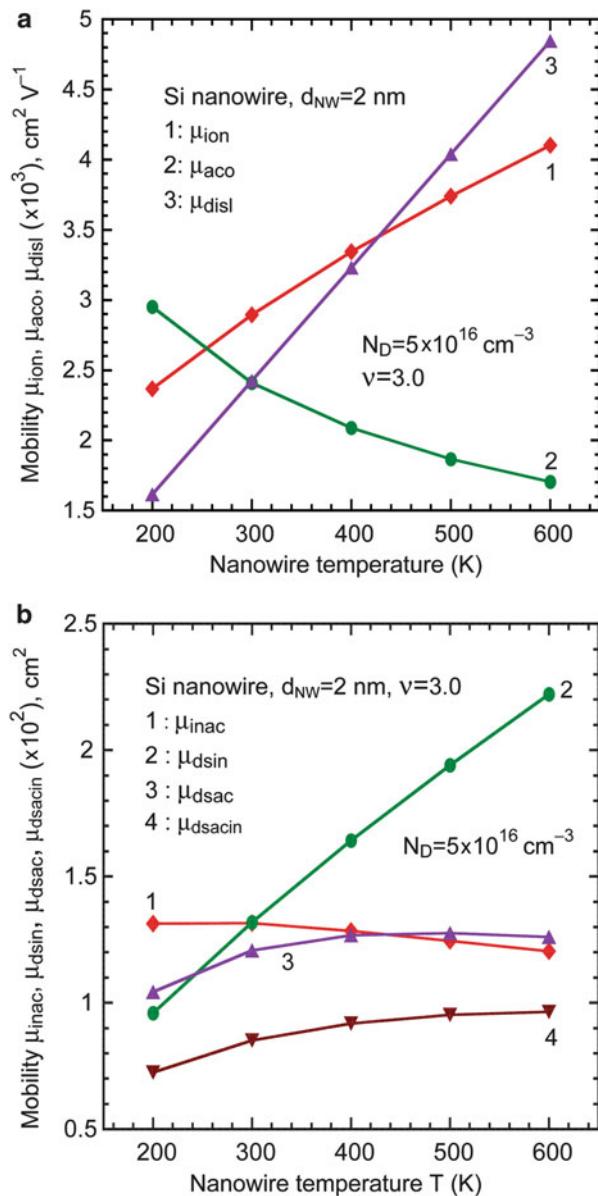
by these researchers are very similar to those of curve 2 of Fig. 2.9a, and this is due to the fact that the mobility by Ford et al. [85] was dominated by phonon scattering. The dependence of the mobilities  $\mu_{inac}$ ,  $\mu_{dsin}$ ,  $\mu_{dsac}$ ,  $\mu_{dsacm}$  on temperature is presented in Fig. 2.9b. This figure shows that mobilities  $\mu_{dsin}$ ,  $\mu_{dsac}$ , and  $\mu_{dsacm}$  increase with increasing temperature, but the mobility  $\mu_{inac}$  is almost independent of temperature. The near-temperature-independence of curve 1 of Fig. 2.9b is

**Fig. 2.8** Doping-dependent carrier mobilities taking both the quantum confinement and the dielectric confinement into account. (a) Comparison of  $\mu_{\text{ion}}$  and  $\mu_{\text{disl}}$  for nanowires of diameters  $d_{\text{NW}} = 2 \text{ nm}$  and  $5 \text{ nm}$ , respectively. (b) Comparison of  $\mu_{\text{inac}}$ ,  $\mu_{\text{dsin}}$ ,  $\mu_{\text{dsac}}$ , and  $\mu_{\text{dsacin}}$ , respectively, at 300 K



particularly noteworthy. With the mobility  $\mu_{\text{dsin}}$  disregarded, the mobility  $\mu_{\text{inac}}$  is not only high but also covers a wide temperature range. This is not unreasonable, as Cheng et al. [92] demonstrated that carrier mobility can indeed be almost independent of temperature over a temperature range of  $5 \text{ K} \leq T \leq 300 \text{ K}$ .

**Fig. 2.9** Temperature-dependent carrier mobilities taking both the quantum confinement and the dielectric confinement ( $v = 3$ ) into account. (a) Comparison of the carrier mobilities (a)  $\mu_{ion}$ ,  $\mu_{disl}$ , and  $\mu_{aco}$ , respectively; (b)  $\mu_{inac}$ ,  $\mu_{dsin}$ ,  $\mu_{dsac}$ , and  $\mu_{dsacin}$ , respectively



## 2.10 Conclusions

The SNM model for quantum confinement in nanowires and the applicability of this model to the electrical transport in semiconductor nanowires for FinFETs have been studied. Quantum confinement, as depicted by the SNM model has recently been observed by Rustagi et al. [45, 46] in Si nanowires. These authors noted that

the energy levels  $E_{\text{nm}}$  for different diameters of a nanowire in a FinFET-like environment are very discrete, and there is an increase in separation of energy levels with decrease in the nanowire diameter. It was found that the SNM model provides a distinctly new route for enhanced carrier transport in nanowires. This route is quite aligned with experiments. On-current is about  $700 \mu\text{A}/\mu\text{m}$  in bulk-Si FinFET, about  $850 \mu\text{A}/\mu\text{m}$  in strained-Si FinFET, but about  $4,000 \mu\text{A}/\mu\text{m}$  in Si nanowire FinFET [93]. Such an enormous increase in on-current in Si nanowire FinFET is due to enhanced carrier mobility, as demonstrated by the SNM model.

The SNM model considers  $b = 1$ , and yet the yields in carrier mobility in thin nanowires is extremely well. It also correctly predicts that the nanowire carrier mobility is dependent on impurity scattering [74, 75] and is a function of  $T^{3/2}$ . But the nanowire carrier mobility is dependent on phonon scattering [85] and is a function of  $\sim T^{-3/2}$ . The present investigation addresses the impact of both quantum confinement and dielectric confinement of nanowires on nanowire mobility. It predicts that carrier mobility may be vastly enhanced if influenced by both quantum confinement and dielectric confinement. It takes only three dominant scattering mechanisms into account. It is though generally enough to take other scattering mechanisms into account. Good agreement of calculated results with available experiments suggests that the three scattering mechanisms taken into account for the present calculations are dominant.

The present results are in line with also those by Khanal et al. [11], who predicted that both calculated and measured carrier mobilities decrease exponentially or near-exponentially with increasing doping concentration due to ionized impurity scattering. Monte Carlo calculations by Jin et al. [94] suggest that carrier mobility decreases monotonically with decrease in nanowire diameter due to surface roughness scattering. Wang et al. [6] however predicted an opposite trend in the mobility as function of nanowire diameter. Khanal et al. [11] argued that surface roughness scattering limits mobility only under very high internal electric fields, and the electric field decreases with decrease in nanowire diameter. This is probably the reason of why Neophytou and Kosina [95] found the surface roughness scattering in thin nanowires to be quite insignificant. Persson et al. [96] made use of tight binding approximations to study the effect of charge impurity scattering on carrier mobility in Si [110] nanowires in a gate-all-around geometry. They concluded that the carrier mobility increases with decreasing nanowire diameter. Koley et al. [97] also observed that the measured carrier mobility increases with decreasing nanowire diameter. Figure 2.6a demonstrates the impact of acoustic phonon scattering on carrier mobility. In shows that carrier mobility increases linearly with increasing nanowire diameter. Murphy-Armando et al. [98] and Zhang et al. [99] studied the effect of electron–phonon scattering on carrier mobility and also found that carrier mobility increases almost linearly with nanowire dimension. The investigations by Murphy-Armando et al. [98] and Zhang et al. [99] thus suggest that the carrier mobility due to electron–phonon scattering is similar to the carrier mobility due to acoustic phonon scattering. They all attest to the strength and novelty of the SNM model. The SNM model [13] predicts that a doping  $N_D < 5 \times 10^{16} \text{ cm}^{-3}$ , temperature  $300 \text{ K} \leq T \leq 600 \text{ K}$ , and nanowire diameter  $d_{\text{NW}} < 4 \text{ nm}$  are most suitable for achieving high carrier mobility in nanowires.

## References

1. Colinge, J.-P.: FinFETs and other multigate transistors. Springer, New York, NY (2007)
2. Fan, H.J., Werner, P., Zacharias, M.: *Small* **2**, 700 (2006)
3. Mohammad, S.N.: *J. Appl. Phys.* **110**, 084310 (2011)
4. Kotlyar, R., Obradovic, B., Matagne, P., Stettler, M., Giles, M.D.: *Appl. Phys. Lett.* **84**, 5270 (2004)
5. Barraud, S., Sarrazin, E., Bournel, A.: *Semicond. Sci. Technol.* **26**, 025001 (2011)
6. Wang, J., Polizzi, E., Ghosh, A., Datta, S., Lundstrom, M.S.: *Appl. Phys. Lett.* **87**, 043101 (2005)
7. Jena, D., Konor, A.: *Phys. Rev. Lett.* **98**, 136805 (2007)
8. Fonoberov, V.A., Balandin, A.A.: *Nano Lett.* **6**, 2442 (2006)
9. Sakaki, H.: *Jpn. J. Appl. Phys.* **19**, L735 (1980)
10. Ramayya, E.B., Vasileska, D., Goodnick, S.M., Knezevic, I.: *J. Appl. Phys.* **104**, 063711 (2008)
11. Khanal, D.R., Levander, A.X., Yu, K.M., Liliental-Weber, Z., Walukiewicz, W., Grandal, J., Sanchez-Garcia, M.A., Calleja, E., Wu, J.: *J. Appl. Phys.* **110**, 033705 (2011)
12. Jena, D., Konor, A.: *J. Appl. Phys.* **102**, 123705 (2007)
13. Mohammad, S.N.: *Nanotechnology* **23**, 285707 (2012)
14. Mohammad, S.N.: *J. Chem. Phys.* **125**, 094705 (2006)
15. Mohammad, S.N.: *J. Chem. Phys.* **127**, 244702 (2007)
16. Schmidt, M., Kusche, R., Issendorff, V.B., Heberland, H.: *Nature* **393**, 238 (1998)
17. Bertsch, G.: *Science* **277**, 1619 (1997)
18. Jiang, Q., Zhang, Z., Li, J.C.: *Acta Mater.* **48**, 4791 (2000)
19. Lamber, R., Wetjen, S., Jaeger, N.I.: *Phys. Rev. B* **51**, 10968 (1995)
20. Goldstein, A.N., Echer, C.M., Alivisatos, A.P.: *Science* **256**, 1425 (1992)
21. Janczuk, B., Jcik, W.W., Guindo, M.C., Chibowski, E., Gonzalez, F., Ballero, C.: *Mater. Chem. Phys.* **37**, 64 (1993)
22. Nanda, K.K., Maisels, A., Kruis, F.E., Fissan, H., Stappert, S.: *Phys. Rev. Lett.* **91**, 106102 (2003)
23. Mohammad, S.N.: *Adv. Mater.* **24**, 1262 (2012)
24. Jiang, Q., Lu, H.M., Zhao, M.: *J. Phys. Condens. Matter* **16**, 521 (2004)
25. Mehl, M.J., Papaconstantopoulos, D.A.: *Phys. Rev. B* **54**, 4519 (1996)
26. Lindemann, F.A.: *Z. Phys.* **11**, 609 (1910)
27. Mohammad, S.N.: *J. Appl. Phys.* **110**, 054311 (2011)
28. Coombes, C.J.: *J. Phys. F: Metal Phys.* **2**, 441 (1972)
29. Nanda, K.K.: *Appl. Phys. Lett.* **87**, 021909 (2005)
30. Zhang, Y., Suenaga, K., Collie, C., Iijima, S.: *Science* **281**, 973 (1998)
31. Mohammad, S.N.: *J. Appl. Phys.* **107**, 114304 (2010)
32. Wang, Y., Luo, Z., Li, B., Ho, P.S., Yao, Z., Shi, L., Bryan, E.N., Nemanich, R.J.: *J. Appl. Phys.* **101**, 124310 (2007)
33. Ho, G.H., Wong, A.S.W., Wee, A.T.S., Welland, M.E.: *Nano Lett.* **4**, 2023 (2004)
34. Cui, L.-F., Ruffo, R., Chan, C.K., Peng, H., Cui, Y.: *Nano Lett.* **9**, 491 (2009)
35. Dan, Y., Seo, K., Takei, K., Meza, J.H., Javey, A., Crozier, K.B.: *Nano Lett.* **11**, 2527 (2011)
36. Bogart, T. D., Lu, X. and Korgel, B. A.: *Dalton Trans.* **42**, 12675 (2013)
37. Zhang, Q., Bayliss, S.C.: *J. Appl. Phys.* **79**, 1351 (1996)
38. Bruno, M., Palummo, M., Marini, A., Del Sole, R., Ossicini, S.: *Phys. Rev. Lett.* **98**, 036807 (2007)
39. Wolkin, M.V., Jorne, J., Fauchet, P.M., Allan, G., Delerue, C.: *Phys. Rev. Lett.* **82**, 197 (1999)
40. Street, R.A.: *Adv. Phys.* **30**, 593 (1981)
41. Ma, D.D.D., Lee, C.S., Au, F.C.K., Tong, S.Y., Lee, S.T.: *Science* **299**, 1874 (2003)
42. Nolan, M., O'Callaghan, S., Fagas, G., Greer, J.C., Frauenheim, T.: *Nano Lett.* **7**, 34 (2007)
43. Scheel, H., Reich, S., Thomsen, C.: *Phys. Status Solidi (b)* **242**, 2474 (2005)

44. Lin, Y.-M., Sun, X., Dresselhaus, M.S.: Phys. Rev. B **62**, 4610 (2000)
45. Rustagi, S.C., Singh, N., Lim, Y.F., Zhang, G., Wang, S., Lo, G.Q., Balasubramanian, N., Kwong, D.-L.: IEEE Electron Devices Lett. **28**, 909 (2007)
46. Leonard, F., Talin, A.A.: Phys. Rev. Lett. **97**, 026804 (2006)
47. Park, N.-M., Kim, T.-S., Park, S.-J.: Appl. Phys. Lett. **78**, 2575 (2001)
48. Wang, W.C., Frietsche, H.: J. Non Cryst. Solids **97–98**, 919 (1987)
49. Gen, C., Yu, V., Kaznacheev, A., Yunovich, A.E.: Fiz. Tekh. Poluprovodn **25**, 1681 (1991)
50. Gen, C., Yu, V., Kaznacheev, A., Yunovich, A.E.: Sov. Phys. Semicond. **25**, 1011 (1991)
51. Yi, K.S., Trivedi, K., Floresca, H.C., Yuk, H., Hu, W., Kim, M.K.: Nano Lett. **11**, 5465 (2011)
52. Niquet, Y.M., Lherbier, A., Quang, N.H., Fernández-Serra, M.V., Blasé, X., Delerue, C.: Phys. Rev. B **73**, 165319 (2006)
53. Boultrip, F., Dunstain, D.J.: Phys. Rev. B **28**, 5923 (1983)
54. Searle, T.M., Jackson, W.A.: Philos. Mag. B **60**, 237 (1989)
55. Giorgis, F., Pirri, C.F., Vinegoni, C., Pavesi, L.: Phys. Rev. B **60**, 11572 (1999)
56. Yang, C.C., Li, S.: Phys. Rev. B **75**, 165413 (2007)
57. Liu, D., Zhu, Y.F., Jiang, Q.: J. Phys. Chem. C **113**, 10907 (2009)
58. Naher, U., Bjrnholm, S., Frauendorf, S., Garcias, F., Guet, C.: Phys. Rep. **285**, 245 (1997)
59. Nanda, K.K., Sahu, S.N., Behera, S.N.: Phys. Rev. A **66**, 013208 (2002)
60. Rose, J.R., Vary, J.P., Smith, J.R.: Phys. Rev. Lett. **53**, 344 (1984)
61. Rose, J.R., Smith, J.R., Ferrante, J.: Phys. Rev. B **28**, 1835 (1983)
62. Qi, W.H., Wang, M.P., Xu, G.Y.: Chem. Phys. Lett. **372**, 632 (2004)
63. Sun, C.Q., Shi, Y., Li, C.M., Li, S., Yeung, T.C.: Phys. Rev. B **73**, 075408 (2006)
64. Delerue, C., Allan, G., Lannoo, M.: Phys. Rev. B **48**, 11024 (1993)
65. Wu, Z., Neaton, J.B., Grossman, J.C.: Phys. Rev. Lett. **100**, 246804 (2008)
66. Li, S., Wang, G.W.: Appl. Phys. Lett. **95**, 073106 (2009)
67. Li, M., Li, J.C.: Mater. Lett. **60**, 2526 (2006)
68. Verma, A.S.: Phys. Status Solidi (b) **246**, 345 (2009)
69. Rotaru, C., Nastase, S., Tomozeiu, N.: Phys. Status Solidi (a) **171**, 365 (1999)
70. Chen, C.-W., Chen, K.-H., Shen, C.-H., Ganguly, A., Chen, L.-C., Wu, J.-J., Wen, H.-I., Pong, W.-F.: Appl. Phys. Lett. **88**, 241905 (2006)
71. Robertson, J.: Eur. Phys. J. Appl. Phys. **28**, 265 (2004)
72. Robertson, J.: Rep. Prog. Phys. **69**, 327 (2006)
73. Lee, J., Spector, H.N.: J. Appl. Phys. **54**, 3921 (1983)
74. Salfi, J., Philipose, U., Aouba, S., Nair, S.V., Ruda, H.E.: Appl. Phys. Lett. **90**, 032104 (2007)
75. Motayed, A., Vaudin, M., Davydov, A.V., Melngailis, J., He, M., Mohammad, S.N.: Appl. Phys. Lett. **90**, 043104 (2007)
76. Cimpoiasu, E., Stern, E., Cheng, G., Munden, R., Sanders, A., Reed, M.A.: Braz. J. Phys. **36**, 824 (2006)
77. Nag, B.R.: Electrical transport in compound semiconductors. Springer, New York, NY (1980)
78. Ramayya, E.B., Valiseska, D., Goodnick, S.M., Knezevic, I.: IEEE Trans. Nanotechnol. **6**, 113 (2007)
79. Read, W.T.: Philos. Mag. **45**, 775 (1954)
80. Read, W.T.: Philos. Mag. **46**, 111 (1954)
81. Podor, B.: Phys. Status Solidi **16**, K167 (1966)
82. Cui, Y., Zhong, Z., Wang, D., Wang, W.U., Lieber, C.M.: Nano Lett. **3**, 149 (2003)
83. Koo, S.M., Fujiwara Han, A.J.P., Vogel, E.M., Richter, C.A., Bonevich, J.E.: Nano Lett. **4**, 2197 (2004)
84. Takagi, S., Koga, J., Toriumi, A.: Jpn. J. Appl. Phys. **37**, 1289 (1998)
85. Ford, A.C., Ho, J.C., Chueh, Y.L., Tseng, Y.C., Fan, Z., Guo, J., Bokor, J., Javery, A.: Nano Lett. **9**, 360 (2009)
86. Trivedi, K., Yuk, H., Floresca, H.C., Kim, M.J., Hu, W.: Nano Lett. **11**, 1412 (2011)
87. Thelander, C., Agarwal, P., Brongersma, S., Eymer, J., Feiner, L.F., Forchel, A., Scheffer, M., Riess Ohlsson, W.B.J., Goesel, U., Samuelson, L.: Mater. Today **9**, 28 (2006)

88. Khanal, D.R., Yim, J.W.L., Walukiewicz, W., Wu, J.: *Nano Lett.* **7**, 1186 (2007)
89. Huang, Y., Duan, X., Cui, Y., Lauhon, L.J., Kim, K.H., Lieber, C.M.: *Nano Lett.* **2**, 101 (2002)
90. Chang, C.-Y., Chi, G.-C., Wang, W.-M., Chen, L.-C., Chen, K.-H., Ren, F., Pearton, S.J.: *Appl. Phys. Lett.* **87**, 093112 (2005)
91. Stern, E., Cheng, G., Cimpoiiasu, E., Klie, R., Guthrie, S., Klemic, J., Kretzschmar, I., Steinlauf, E., Turner-Evans, D., Broom-field, E., Hyland, J., Koudelka, R., Boone, T., Young, M., Sanders, A., Mundon, R., Lee, T., Routenberg, D., Reed, M.A.: *Nanotechnology* **16**, 2941 (2005)
92. Cheng, G., Stern, E., Turner-Evans, D., Reed, M.A.: *Appl. Phys. Lett.* **87**, 253103 (2005)
93. Jiang, Y., Liow, T.Y., Singh, N., Tan, L.H., Lo, G.Q., Chan, D.S.H., Kwong, D.L. In: IEEE Symposium on VLSI Technology Digest of Technical Papers, p. 34 (2008)
94. Jin, S., Fischetti, M.V., Tang, T.W.: *J. Appl. Phys.* **102**, 083715 (2007)
95. Neophytou, N., Kosina, H.: *Solid State Electron.* **70**, 81 (2012)
96. Persson, M.P., Mera, H., Niques, Y.-M., Delerue, C.: *Phys. Rev. B* **82**, 115318 (2010)
97. Koley, G., Cai, Z., Quddus, E.B., Liu, J., Qazi, M., Webb, R.A.: *Nanotechnology* **22**, 295701 (2011)
98. Murphy-Armando, F., Fagas, G., Greer, J.C.: *Nano Lett.* **10**, 869 (2010)
99. Zhang, W., Delerue, C., Niquet, Y.-M., Allan, G., Wang, E.: *Phys. Rev. B* **82**, 115319 (2010)

# Chapter 3

## Understanding the FinFET Mobility by Systematic Experiments

Kerem Akarvardar, Chadwin D. Young, Mehmet O. Baykan,  
and Christopher C. Hobbs

**Abstract** The impact of the surface orientation, strain, fin doping, and gate stack on SOI double-gate FinFET mobility is systematically investigated. Impact of channel material, temperature, and fin width were also touched upon to better understand the trends. For the unstrained case, the (110) sidewall electron mobility is very close to the (100) sidewall electron mobility irrespective of the fin doping level and gate stack. This weak dependence of electron mobility to surface orientation distinguishes the FinFETs from the bulk planar MOSFETs, where (100) electron mobility is systematically reported to be much higher than that of (110). On the other hand, the (110) sidewall hole mobility is substantially higher than the (100) sidewall hole mobility in FinFETs, as in the planar case. Both the (100)/ $<100>$  and (110)/ $<110>$  FinFET electron mobility can be improved with tensile strain. It is also confirmed that the (110)/ $<110>$  FinFET hole mobility can be significantly improved with compressive strain while the (100)/ $<100>$  hole mobility is sensitive to neither compressive nor tensile strain. Compared to Si, the use of a SiGe channel increases the hole mobility drastically, and even further improvement is achievable by external compressive stress. Overall, the experimental results in

---

K. Akarvardar (✉)

Technology Research Group, Globalfoundries Inc., Albany, NY, USA

e-mail: [Kerem.akarvardar@globalfoundries.com](mailto:Kerem.akarvardar@globalfoundries.com)

C.D. Young

Materials Science and Engineering Department, UT Dallas, Richardson, TX, USA

e-mail: [chadwin.young@utdallas.edu](mailto:chadwin.young@utdallas.edu)

M.O. Baykan

Department of Electrical and Computer Engineering, University of Florida,  
Gainesville, FL, USA

e-mail: [baykan@ufl.edu](mailto:baykan@ufl.edu)

C.C. Hobbs

Front End Processes, SEMATECH Inc., Albany, NY, USA

e-mail: [chris.hobbs@sematech.org](mailto:chris.hobbs@sematech.org)

this chapter suggest that the  $(110)/<110>$  Si FinFETs conventionally built on standard  $(100)$  wafers offer simultaneously high electron and hole mobility, which can be further improved by tensile and compressive stress, respectively.

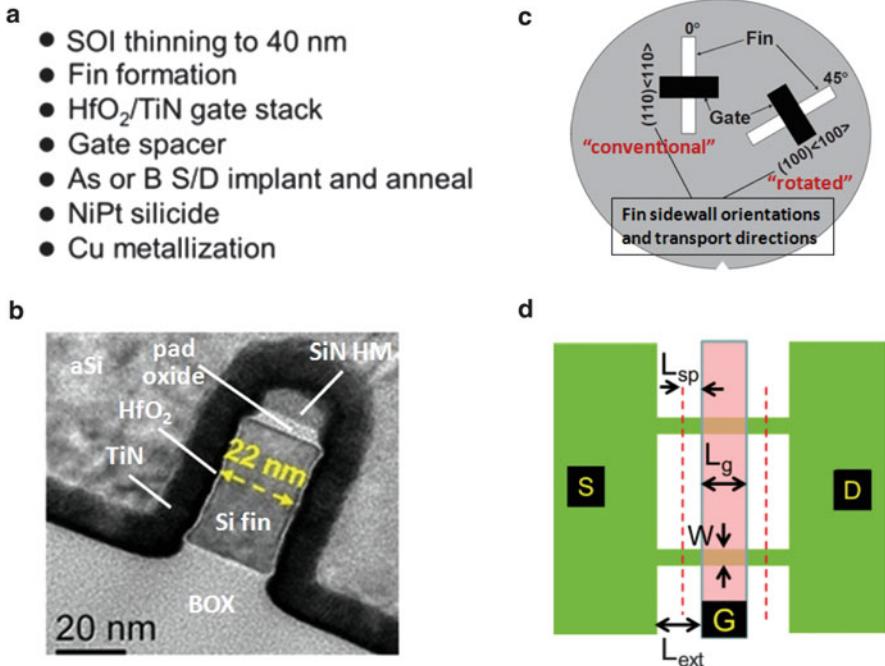
### 3.1 Introduction

Long channel MOSFET mobility remains a relevant measure of nanoscale transport efficiency due to its correlation with the short channel current drive and injection velocity [1–4]. Considering that FinFETs are in production from the 22 nm node onwards [5], a thorough understanding of mobility in these devices is critical. To this end, we provide systematic mobility data on double-gate (DG) silicon-on-insulator (SOI) FinFETs fabricated at SEMATECH Albany, NY. Our purpose is twofold: (1) Provide experimental reference data that can be used to verify or calibrate TCAD or custom mobility models, and (2) Guide the FinFET device design for transport optimization.

All of the experimental data presented in this chapter uses the same FinFET baseline (summarized in Fig. 3.1a), hence forms a coherent set in terms of sensitivity to device splits. The flow starts by thinning  $(100)$  SOI films down to 40 nm on top of 140 nm buried oxide. Then fins are patterned and etched using a nitride hard mask on top of a thin buffer oxide. This is followed by an  $H_2$  bake to smooth the fins. Next, an ALD  $HfO_2$  ( $\sim 2$  nm)/ALD TiN (10 nm) gate stack is formed. Subsequent steps are nitride spacer formation, source-drain (S/D) implantation (directly into the fins), and activation anneal. NiPt silicidation and Cu metallization finalize the gate-first flow. An exemplar fin cross-section featuring 22 nm fin width and 37 nm fin height is shown in Fig. 3.1b. Note that gate stack is deposited without the hard mask being removed, yielding a double-gate structure where the transport is through the fin sidewalls. In majority of our experiments, both the conventional fins with  $(110)$  sidewalls/ $<110>$  fin direction and the 45° rotated fins with  $(100)$  sidewalls/ $<100>$  fin direction were characterized (Fig. 3.1c).

In the following subchapters, we will start by the most basic configuration featuring undoped and unstrained fins with high-K/metal gate (HK/MG). The impact of surface orientation will be first characterized for this configuration. Then we will individually analyze the impact of strain, fin doping, and finally the gate stack. In the greater part of the cases, our electrical data will involve both conventional and rotated fins. Fins with alternative channel material will also be briefly touched upon when we analyze the impact of strain.

For the long-channel mobility extraction ( $L_g = 1$  or  $10$   $\mu m$ ), split CV technique [8] with capacitance–voltage measurements at 100 kHz and current–voltage measurements at  $V_{DS} \leq 50$  mV is used. FinFETs with 200 fins in parallel enabled the measurement of a high enough capacitance for parameter extraction. The fin widths used in this work were varied between 20 and 45 nm. Figure 3.1d shows the representative layout (with only two fins) that is used to generate the data in this chapter unless otherwise mentioned.

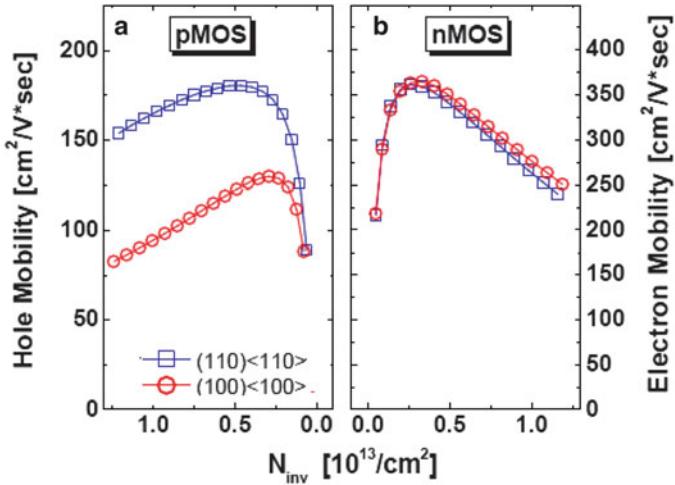


**Fig. 3.1** (a) SOI undoped DG FinFET gate-first flow summary [6], (b) typical Si fin cross-section under  $\text{HfO}_2/\text{TiN}/\text{aSi}$  gate stack [6], (c) sidewall surface orientations and fin channel directions for “conventional” and “rotated” fins on the standard (100) wafer with  $<110>$  notch [7]. (d) Conventional layout that is used to generate the data in this chapter unless otherwise mentioned:  $W$  is the fin width ( $\sim 20$  nm),  $L_g$  is the drawn gate length,  $L_{sp}$  is the spacer width (25 nm),  $L_{ext}$  is the gate to S/D pad distance (100 nm)

### 3.2 Impact of Surface Orientation

Figure 3.2 shows the overall trend in orientation-dependent mobility for undoped HK/MG CMOS FinFETs. Consistent with the prior experimental results [10–13] [23], the (110) hole mobility in Fig. 3.2a is substantially ( $\sim 2 \times$  at high field) better than the (100) hole mobility. Therefore, on standard (100) surface, DG FinFETs that are conventionally laid out along the  $<110>$  direction (hence having a (110) sidewall) present significantly higher unstrained hole mobility compared to their planar counterparts with a (100) surface and  $<110>$  channel direction (which is reported to be an even worse combination than (100)/ $<100>$  in terms of unstrained hole mobility [14, 15]).

The orientation dependence of electron mobility, on the other hand, exhibits a trend that contradicts with the well-established findings from the planar MOSFETs [10–13] [23] (Fig. 3.2b): the conventional (110)/ $<110>$  FinFET electron mobility is very high in absolute value and very close to the (100)/ $<100>$  electron mobility



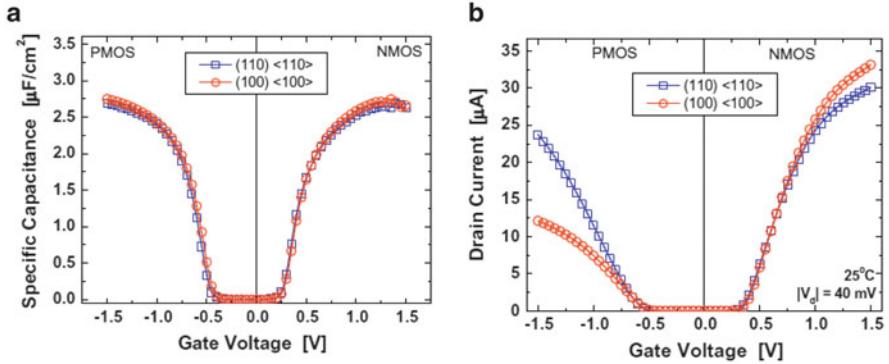
**Fig. 3.2** Undoped PFET (a) and NFET (b) mobility as a function of carrier density for conventional and rotated DG Si FinFETs [9]

measured on rotated FinFETs. Similar results on 11 nm-wide undoped FinFETs with a  $\text{SiO}_2/\text{TiSiN}$  gate stack were also reported in [16] where  $(110)/<110>$  NFET transconductance was measured even higher than that of  $(100)/<100>$  at high field. Our results also correlate with [17] where very high  $(110)/<110>$  electron mobility was reported on 26 nm-wide  $\text{SiO}_2/\text{polyFinFETs}$ .

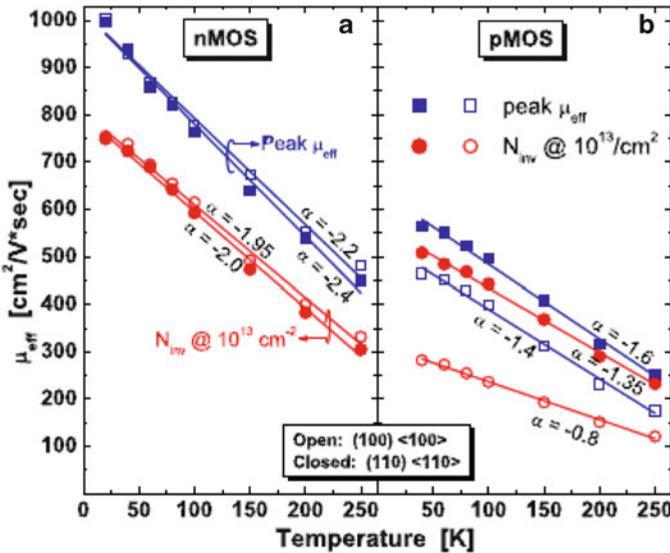
Since planar MOSFET  $(100)/<100>$  and  $(100)/<110>$  unstrained electron mobilities are comparable [14], it follows that conventional unstrained DG FinFET electron mobility is competitive with the conventional planar  $(100)/<110>$  electron mobility. The key result from the data in Fig. 3.2 is that for the unstrained case, conventional  $(110)/<110>$  DG FinFETs built on standard  $(100)/<110>$  wafers have simultaneously very high hole and electron mobility. This, combined with the possibility of undoped body, favors FinFETs in terms of transport for future technology nodes where the efficiency of stress enhancement techniques is reduced due to scaled gate pitch.

Figure 3.3 shows the capacitance–voltage ( $C-V$ ) and linear drain current ( $I-V$ ) characteristics that were used to extract the mobility in Fig. 3.2. Clearly the capacitance, hence equivalent insulator thickness,  $T_{\text{inv}}$ , is very similar for  $(110)$  and  $(100)$  sidewalls for the employed high-K deposition process, and the difference in drain current correlates directly with the difference in respective channel mobilities rather than any charge difference.

In order to evaluate the dominant scattering mechanisms for different orientations, we extracted the peak and the high field electron and hole mobilities at cryogenic temperatures (Fig. 3.4). Power-law exponents ( $\alpha$ ) for the peak electron mobility (Fig. 3.4a) were found to be very high in absolute value—as high as the phonon scattering limited *bulk* mobility exponents [18, 19] and well above the coefficients extracted on planar MOSFETs [12, 20–22]. This result is especially



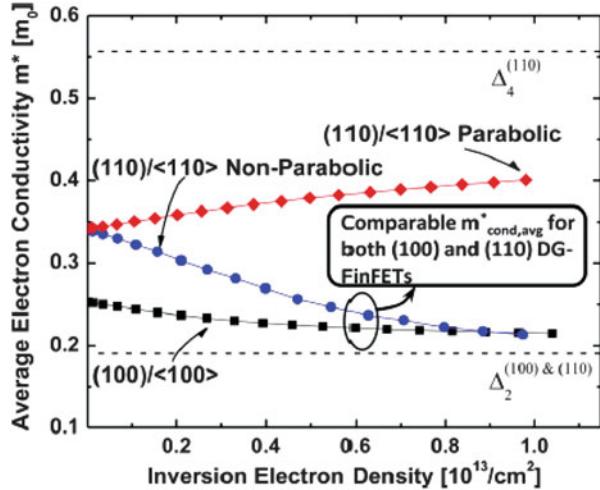
**Fig. 3.3** Undoped PFET and NFET  $C$ - $V$  (a), and  $I$ - $V$  (b) characteristics [7] that were used to extract the mobility vs. carrier density data in Fig. 3.2 using split CV method



**Fig. 3.4** Undoped NFET (a) and PFET (b) peak and high field mobility as a function temperature for conventional and rotated DG Si FinFETs [7]

surprising given the possibility of additional carrier scattering from the  $\text{HfO}_2$  [12, 23, 24]. Such a high  $\alpha$  suggests that phonon scattering is the dominant mechanism governing the low-field mobility in our FinFETs. Also, coefficients for (110)/<110> and (100)/<100> devices are very close to each other suggesting similar scattering rates for the two combinations. Temperature coefficients at high field are also very close for the conventional and rotated NFETs but slightly reduced compared to peak mobility due to the impact of surface roughness scattering, which reduces the overall mobility per Mathiessen's rule [8].

**Fig. 3.5** Average conductivity effective mass as a function of inversion electron density calculated with parabolic and non-parabolic confinement mass methods for conventional and rotated undoped DG FinFETs ( $W_{\text{fin}} = 20 \text{ nm}$ ) [25]

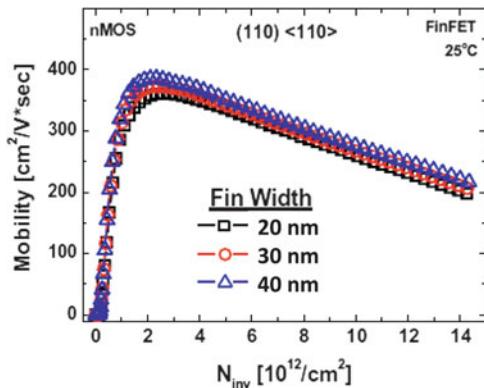


The fact that both the absolute value of electron mobility and the scattering rates are similar for (110)/<110> and (100)/<100> implies that the conductivity effective masses ( $m_{\text{cond,avg}}^*$ ) should be similar for these two combinations [7]. This is indeed confirmed by the  $\text{sp}^3\text{d}^5\text{s}^*$  tight-binding band structure calculations coupled with self-consistent Schrodinger–Poisson simulations on 20 nm-wide undoped FinFETs: when the non-parabolicity in confinement mass of  $\Delta_2$  valley of (110) conduction band ( $m_z^*$ ) is properly taken into account, (110)/<110>  $m_{\text{cond,avg}}^*$  gets closer to (and can even be lighter than) the (100)/<100>  $m_{\text{cond,avg}}^*$ , especially at high inversion carrier density (Fig. 3.5) [25]. By contrast, the parabolic confinement mass approach [26] results in majority of electrons to occupy  $\Delta_4$  valley for (110), leading to a significantly higher  $m_{\text{cond,avg}}^*$  for (110) compared to (100).

Low temperature trends for FinFET hole mobility are shown in Fig. 3.4b. Power-law exponents for (110) and (100) peak hole mobilities suggest again the dominance of phonon scattering for both orientations [20]. On the other hand, for (110)/<110>, there is a slight reduction from peak to high field hole mobility exponents, owing to the impact of increased surface roughness scattering. At high-field,  $\alpha$  for (100) is lower than the  $\alpha$  for (110), similar to what has been observed in [12] on planar bulk MOSFETs with  $\text{HfO}_2$  or  $\text{SiO}_2/\text{poly-Si}$  gate stack.

Given the similarity of measured electron mobility in (100) and (110) FinFETs and its theoretical foundation reported in [25], the repeatedly measured inferior electron mobility on bulk planar (110) MOSFETs [11–13] [Krishnan'08] is puzzling. In fact, the electron mobility disadvantage of bulk planar (110) surface is so widely acknowledged that it gave rise to the consideration of alternative CMOS solutions such as “Hybrid Orientation Technology,” where planar NFETs and PFETs are exclusively built on (100) and (110), respectively [11]. The empirical modeling closely matching the experimental temperature dependent mobility data on planar and FinFET devices in [27] shows that the discrepancy between the “low” planar and

**Fig. 3.6** (110)/<110> undoped FinFET electron mobility for various fin widths [7]



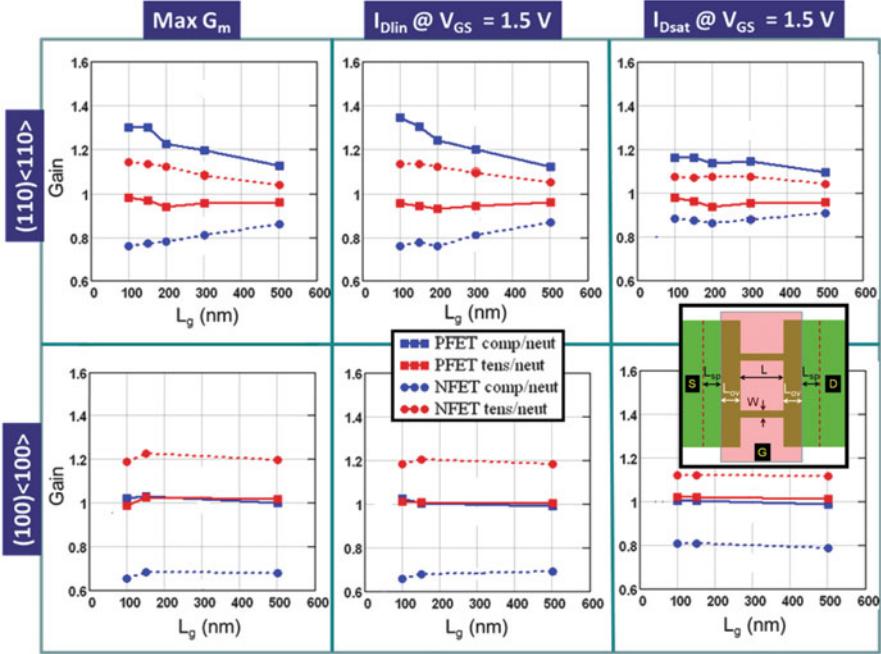
“high” FinFET (110) electron mobility could be explained as follows: for (110) bulk planar MOSFETs, Coulomb scattering (at low field) and surface roughness scattering (at high field) are significantly higher than in FinFETs and (100) planar MOSFETs where the phonon scattering is the dominant mechanism that limits mobility.

Before we complete this subsection, it is worth to mention that the high (110) electron mobility measured in FinFETs is essentially insensitive to fin width (Fig. 3.6). This, together with the fact that even the narrowest fin considered here is wider than 20 nm, suggests that quantum confinement or volume inversion [28] does not play a visible role in high (110) electron mobility measured on our devices.

### 3.3 Impact of Strain

In this subsection, we will first analyze the strain sensitivity of FinFETs using compressive and tensile contact etch stop liners (CESLs) that were deposited post-silicidation. Although this method is ineffective for the advanced technology nodes with tight gate pitches [29], it is still a good fundamental learning vehicle if transistor test structures with a single gate or wide enough gate pitch are available. The stress transfer from the CESL to the MOSFET channel is reported to occur from the fin extensions at the S/D regions [30]. Accordingly, this method gets more effective as the channel length scales. Usually no mobility change is observed in conventional long channel devices ( $L_g \geq \sim 1 \mu\text{m}$ ) and one should focus on short channel MOSFETs to observe the impact of CESL stress on transport.

Compressive (3 GPa) and tensile (1 GPa) CESL strain response of (110)/<110> and (100)/<100> FinFETs is investigated in Fig. 3.7, where peak transconductance ( $G_{mmax}$ ) and drain current ( $I_{Dlin}$  and  $I_{Dsat}$ ) gains compared to neutral CESL case are shown for various channel lengths. In (110)/<110> FinFETs, the impact of CESL stress gets more significant with  $L_g$  scaling as expected. However, this is surprisingly not the case for rotated FinFETs suggesting that the stress may directly be transferred through the gate, leading to  $L_g$ -insensitivity. Otherwise it is observed

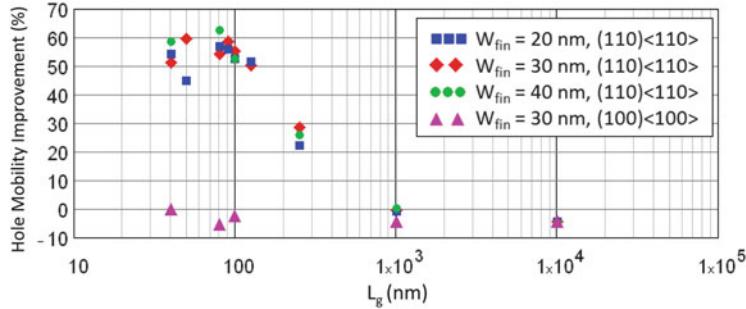


**Fig. 3.7** Gain in peak transconductance, linear current, and saturation current with tensile (1 GPa) and compressive (3 GPa) CESL as a function of  $L_g$  for conventional and rotated complementary undoped DG SOI FinFETs [31]. Data generated using the layout at the inset where the gate overlaps with S/D pads by  $L_{ov} = 30$  nm and  $L_g = L$  corresponds to the fin length between the S/D pads [32]

that  $I_{Dlin}$  and  $G_{mmax}$  trends are very similar and hence interchangeable for all combinations. Also, the impact of CESL stress is observed to be more remarkable on  $I_{Dlin}$  than on  $I_{Dsat}$ .

Starting with p-channel FinFETs, one clear trend is the appreciable strain sensitivity of  $(110)/<110>$  and insensitivity of  $(100)/<100>$  to external compressive stress. So, the  $(100)/<100>$  orientation not only presents a very low PFET mobility but also does not show any improvement with compressive stress. By contrast, the  $(110)/<110>$  orientation which has a high mobility to begin with, shows an even further improvement with compressive stress. On the other hand,  $(100)/<100>$  p-channel FinFETs are insensitive to tensile strain while  $(110)/<110>$  FinFETs show degradation with tensile CESL strain. Different than the p-channel FinFETs, n-channel FinFETs appear to show higher sensitivity to stress for  $(100)/<100>$  than for  $(110)/<110>$  [33]. The surface orientation dependent trends in Fig. 3.7 are essentially consistent with the CESL-induced stress simulation results on FinFETs using bulk-Si piezoresistance coefficients [34].

In order to monitor the change in short-channel mobility with CESL stress, either a substantial number of short-channel FETs in parallel with high capacitance are needed for split CV measurements (which is impractical), or a full DC method



**Fig. 3.8** Low-field hole mobility improvement with 3 GPa compressive CESL over neutral CESL case as a function of drawn gate length for conventional and rotated undoped DG p-channel Si FinFETs. Conventional devices have three different fin widths [31]

needs to be used. In what follows, we show how the “Y-function method” [35] can be used to assess to the first order the impact of stress on short-channel MOSFET mobility.

To obtain  $Y$  function we start with the basic linear drain current expression in strong inversion:

$$I_d = \frac{W_{device} \times C_{ox}}{L_g} \frac{\mu_0}{[1 + \theta(V_{GS} - V_T)]} (V_{GS} - V_T) V_d, \quad (3.1)$$

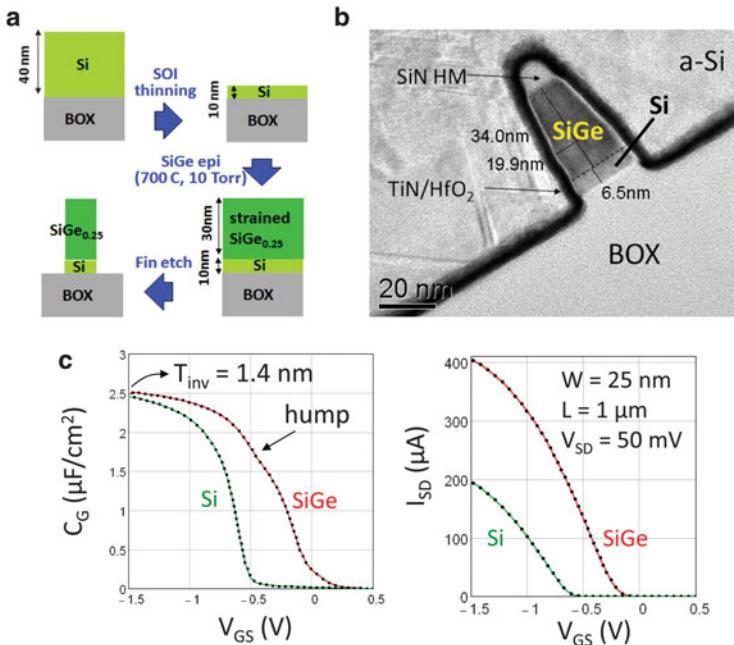
where  $W_{device}$  is the device width,  $C_{ox}$  is the oxide capacitance,  $\mu_0$  is the low-field mobility,  $V_{GS} - V_T$  is the gate overdrive voltage, and  $\theta$  is the mobility attenuation factor. Using (3.1),  $Y$  function is defined as:

$$\frac{I_d}{g_m^{1/2}} = \left( \frac{W_{device}}{L_g} C_{ox} \mu_0 V_d \right)^{1/2} (V_{GS} - V_T), \quad (3.2)$$

which is independent of  $\theta$  and presents a linear variation in strong inversion as a function of  $(V_{GS} - V_T)$  [35]. Assuming that  $W_{device}$ ,  $L_g$ , and  $C_{ox}$  remain unchanged with CESL stress, the strained to unstrained low field mobility ratio is given, from (3.2), by:

$$\frac{\mu_{0,\text{strained}}}{\mu_{0,\text{unstrained}}} = \left[ \frac{\text{Slope}(Y_{\text{strained}})}{\text{Slope}(Y_{\text{unstrained}})} \right]^2. \quad (3.3)$$

Note that (3.3) does not require the exact value of  $W_{device}$ ,  $L_g$ , and  $C_{ox}$  to study the dependence of mobility improvement on those parameters. Figure 3.8 exemplifies the use of proposed methodology to extract the  $\mu_0$  improvement by 3 GPa compressive CESL on p-channel FinFETs as a function drawn gate length. Again, consistent with the CESL stress transfer mechanism, no improvement is



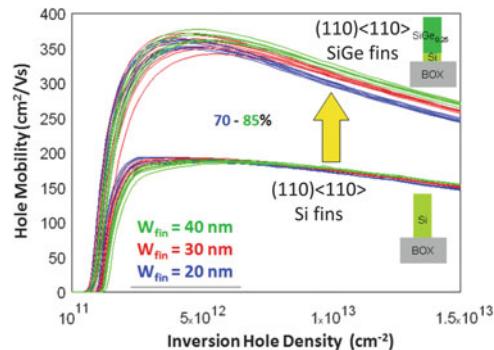
**Fig. 3.9** (a) Strained SiGe fin formation on SOI, (b) strained SiGe fin cross-section post gate stack formation, (c) typical long-channel  $I$ - $V$  and  $C$ - $V$  curves for p-channel strained SiGe PFETs on SOI [39]

noticed for long-channel (1 and 10  $\mu\text{m}$ ) devices. The ineffectiveness of CESL in long-channel devices is confirmed independently by split CV measurements. For shorter channel lengths, compressive CESL gets gradually effective and improves  $\mu_0$  up to 50–60 % for the shortest channel (100)/<110> FinFETs irrespective of the fin width. By contrast, the (100)/<100> FinFETs exhibit no improvement with cCESL, consistent with the data in Fig. 3.7.

Unlike the CESL or embedded SiGe S/D-induced stress, the substrate-induced stress is effective in long-channel as well as in short-channel FETs assuming no relaxation occurs during the downstream flow. In FinFETs, substrate induced biaxial stress transforms into uniaxial stress along the channel post-fin formation [36]. Hence, substrate-induced stress is more beneficial when fabricating uniaxially strained FinFETs than planar FETs under biaxial stress due to additional band bending induced transport mass reduction and greater suppression of phonon scattering rates. In the following, we will consider SiGe p-channel FinFETs built on thin SOI seed films [37, 38] as an example of 3D devices benefitting from the substrate-induced stress.

Figure 3.9a shows the fin formation flow for SiGe fins on SOI with 25 % Ge content [37]: after the SOI film is thinned down to ~5 nm, single crystal, 35 nm Si<sub>0.75</sub>Ge<sub>0.25</sub> is grown by epitaxy. Since at this point SiGe film has the same lattice constant as Si perpendicular to the growth direction, it is under biaxial compressive

**Fig. 3.10** Hole mobility comparison for conventional (110)/<110> strained SiGe and Si fins for various fin widths [39]

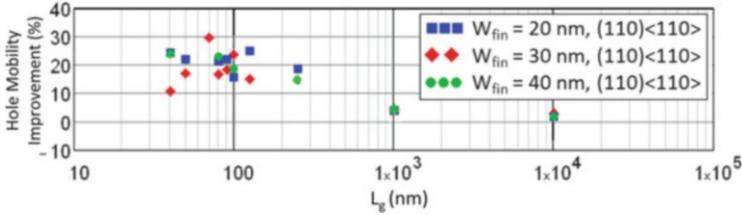


stress. Note that 35 nm is beyond the critical thickness for 25 % Ge [40] such that SiGe layer is in metastable state. Subsequent fin formation relaxes the stress along the fin width [41] and enables to uniaxial compressive stress along the SiGe fin channel (fin formation also reduces the total energy and improve robustness of SiGe against relaxation along the fin during downstream processes). The rest of the flow is similar to the Si FinFET flow described in Fig. 3.1a.

SiGe fin cross-section post full flow is shown in Fig. 3.9b, where a clear TEM contrast between the bottom seed Si and the upper channel SiGe is noticed. Figure 3.9c shows the typical long-channel  $I$ - $V$  and  $C$ - $V$  characteristics for SiGe as well as for the control Si FinFETs. The SiGe device has a  $\sim$ 350 mV lower  $V_T$  than the Si FinFET due to the Si-to-SiGe valence band offset and a possible contribution from additional charge at the high-K interface. The conduction of SiGe/Si stack FinFET in Fig. 3.9b is dominated by the top SiGe layer which has much lower  $V_T$  than the bottom Si. A high temperature anneal could be used to diffuse germanium to the bottom Si while still retaining the strain [42]. Note that the seed Si turn-on is noticeable on SiGe curve with a hump (which does not show up in  $I$ - $V$  curves plotted at logarithmic scale [37]).

Figure 3.10 shows the long-channel (110)/<110> hole mobility characteristics of  $\text{Si}_{0.75}\text{Ge}_{0.25}$  and control Si FinFETs for different fin widths. A substantial improvement of 70–85 % with SiGe fins is noticed at high carrier density of  $10^{13} \text{ cm}^{-2}$ . Such a remarkable enhancement would not be possible without the substrate-induced uniaxial compressive stress of  $\sim$ 1.2 GPa, since it is known that low Ge% relaxed SiGe suffers from the alloy scattering and not expected to provide significant benefit—if any—over Si [19]. One of the biggest integration challenges for SiGe devices lattice matched to Si is the retention of substrate-induced strain down to the shortest channel lengths [43].

MOSFETs that already benefit from the substrate-induced strain can be further stressed using process-induced techniques such as embedded SiGe S/D (in PFET case). The sensitivity of uniaxially strained SiGe FinFETs to “external” stress is investigated using again the compressive CESL layer and analyzed based on  $Y$  function methodology described above: Fig. 3.11 indicates an extra 20–30 % low field mobility boost at the shortest channel irrespective of the fin



**Fig. 3.11** Low-field hole mobility improvement with 3 GPa compressive CESL over neutral CESL case as a function of channel length for conventional (110)/<110> SiGe p-channel FinFETs, which are already under uniaxial compressive stress (>1 GPa) from the substrate [39]

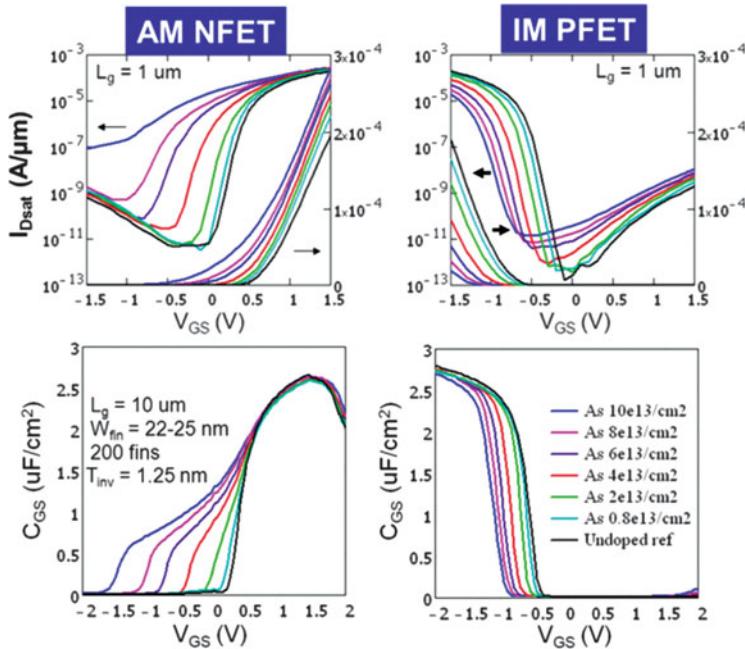
width. These numbers are consistent with the linear and saturation current improvements reported in [37]. Note that this improvement is less than what is reported for Si FinFETs in Fig. 3.8 since the SiGe FinFET already leverages a substantial amount of stress from the seed Si on insulator.

While the SiGe PFETs on Si perform well, the situation is the opposite for SiGe NFETs since, as in the Si FinFET case shown in Fig. 3.7, compressive stress degrades the NFET performance substantially. Even though some of the electron mobility can be recovered by process-induced techniques, compressively strained SiGe NFETs remain behind their Si counterparts in terms of current drive. Accordingly for full CMOS integration, combining compressively strained SiGe PFETs with tensile-strained Si NFETs is recommended [37].

### 3.4 Impact of Fin Doping

Besides being a valuable MOSFET physics study tool, the technological relevance of FinFET doping comes from its apparent simplicity for FinFET threshold voltage ( $V_T$ ) tuning. In undoped FinFETs,  $V_T$  adjustment requires modulation of the metal gate (MG) work function (WF) [44, 45], whose integration may be complex due to different N/PFET  $V_T$  requirements, 3D fin configuration, unconventional work function metals or cap materials, and challenging device reliability. Compared to WF tuning, doping the fins is in principle easier to integrate; however, it gets less manufacturable with scaled fin width due to reduced sensitivity of electrical characteristics to doping and random dopant fluctuation [6]. Basic analytical models show that for aggressively scaled (5 nm-wide) fins, the impact of single dopant atom on  $V_T$  can be as high as 25 mV, severely challenging the viability of the technique towards the end of roadmap, especially if ion implantation (rather than a conformal doping technique) continues to be used [6].

Doped fins considered in this study were obtained using a blank “doping-first” approach where the implant and anneal processes preceded the fin formation [6, 46]. For the same doping conditions, this scheme is confirmed to provide more heavily doped and uniform fins compared to the case where the implant is done after fins

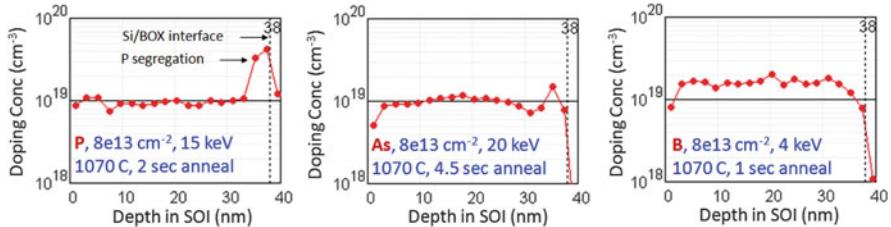


**Fig. 3.12** Current–voltage and capacitance–voltage characteristics of the As-doped Accumulation Mode (AM) n-channel FinFETs and inversion mode (IM) p-channel FinFETs for various As doses [6]

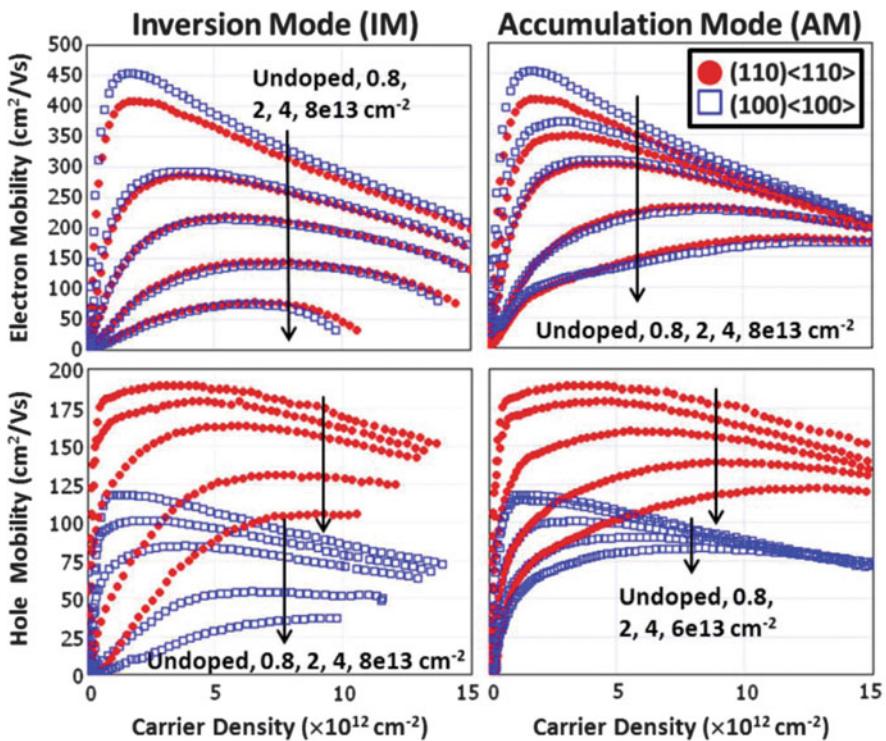
are formed. This is probably because in the former scenario, all the species were implanted into Si and then uniformly distributed via anneal whereas in the second scheme, the species that do not hit the Si fins to begin with have substantially reduced probability to get redistributed into Si post-anneal. The rest of the doped fin flow was as outlined in Fig. 3.1a.

The maskless doping at the beginning of flow enabled inversion-mode (IM) FinFETs of one type and accumulation mode (AM) FinFETs of the opposite type on the same wafer. In Fig. 3.12, the  $I$ – $V$  and  $C$ – $V$  curves for As-doped AM n-channel and IM p-channel FinFETs were shown as an example. As  $I$ – $V$  curves clearly show,  $V_T$  is lowered in NFETs and increased in PFETs with increased As doping. In IM devices, there is a corresponding shift in CV curves with doping. In AM FETs the CV behavior is more complicated than a bare shift of the characteristics and involves also a stretch-out associated with the modulation of the depletion charge inside the fin.

SIMS results from the planar samples that received an implant dose of  $8.10^{13} \text{ cm}^{-2}$  indicate in Fig. 3.13 nearly uniform P, As, and B profiles along the majority of SOI film as shown in Fig. 3.13. Dopant concentrations are in the  $10^{19} \text{ cm}^{-3}$  range, which is significantly higher than the doping extracted from the FinFET  $V_T$  values. This difference suggests dopant out-diffusion from the fins or deactivation during post-processing and/or incomplete activation [57].



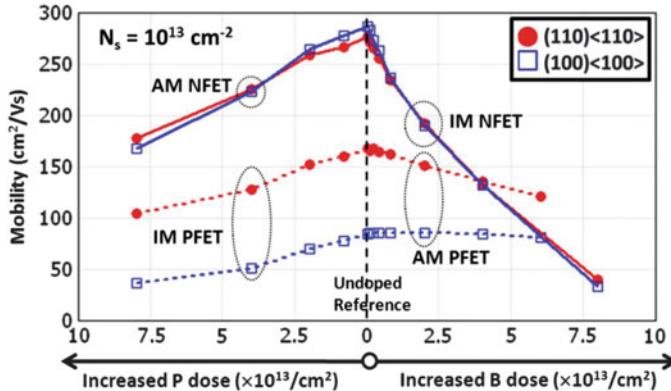
**Fig. 3.13** SIMS data from the planar SOI samples doped with various species at a dose of  $8.10^{13} \text{ cm}^{-2}$ . The energy and anneal conditions are the same as those used in P, As, or B-doped FinFETs [6]



**Fig. 3.14** Mobility versus carrier density ( $N_s$ ) plots for various B (top left and bottom right) and P (top right and bottom left) doses [46]

Comparing the n-type dopants, As is observed to modulate the  $V_T$  more effectively than P for the same implant dose. The reason for this is not clear; however, considering the profiles in Fig. 3.13, one potential reason is the dopant segregation at Si/BOX interface that is observed only in P-doping case.

The variation of FinFET accumulation/inversion electron and hole mobility as a function of carrier density is shown for various B and P doses and for (110) and (100) surface orientations in Fig. 3.14. Compared to the undoped case, mobility



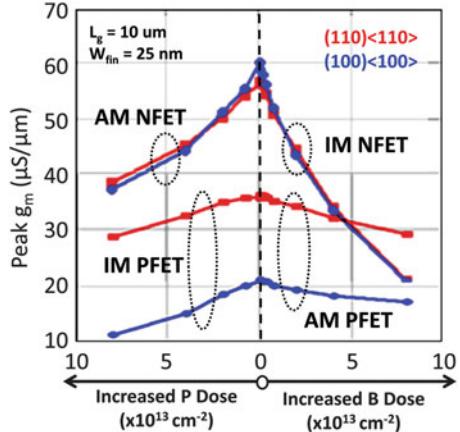
**Fig. 3.15** IM/AM P/NFET mobility at  $N_s = 10^{13} \text{ cm}^{-2}$  as a function of B and P implant doses, summarizing the mobility trends at high carrier density shown in Fig. 3.14 [46]

reduces with doping to a greater or lesser extent for all combinations considered. At low carrier density, mobility degradation with fin doping is due to Coulomb (ionized impurity) scattering in both IM and AM. In inversion-mode devices, mobility degradation with fin doping is visible even at high carrier density because the effective gate field,  $E_{\text{eff}}$ , needed to achieve a given carrier density ( $N_s$ ) increases with doping (i.e., with the depletion charge) and consequently leads to increased surface roughness scattering at that  $N_s$ . By contrast, in accumulation-mode devices, mobility curves tend to merge at high  $N_s$  since there is no depletion-charge at the on state, and  $E_{\text{eff}}$  (hence the intensity of surface roughness scattering) is similar for various doping densities. The absence of depletion charge and the consequent reduction in vertical field is one of the advantages of AM compared to IM [47].

It should be noted however that, in highly doped AM FETs, the sensitivity to channel doping may be significant even at high carrier density provided that the number of volume (body) carriers is comparable to the number of carriers accumulated at the interface. For example, the mobility in “junctionless” FETs [48] exhibits strong doping dependence for the entire range of operation since this category of (JFET-like) devices relies solely on volume (body) conduction.

The high field mobility trends in Fig. 3.14 are summarized in Fig. 3.15 by plotting the mobility at  $N_s = 10^{13} \text{ cm}^{-2}$  as a function of the P and B doses. MOSFET electron mobility clearly degrades more with doping than does the hole mobility. This is similar to the bulk mobility case [8]. Also, at fixed carrier density, inversion mode NFETs and PFETs show higher amount of degradation with doping compared to their accumulation mode counterparts due to the extra field required to deplete the surface. Otherwise, the similarity of (110) and (100) electron mobilities reported in Sect. 3.2 remains valid for all doping values from AM and IM. Hence, the discrepancy between the surface orientation sensitivities of electron mobility in planar and FinFET cases cannot be explained by doping.

**Fig. 3.16** IM/AM P/NFET peak transconductance as a function of B and P implant doses [31]

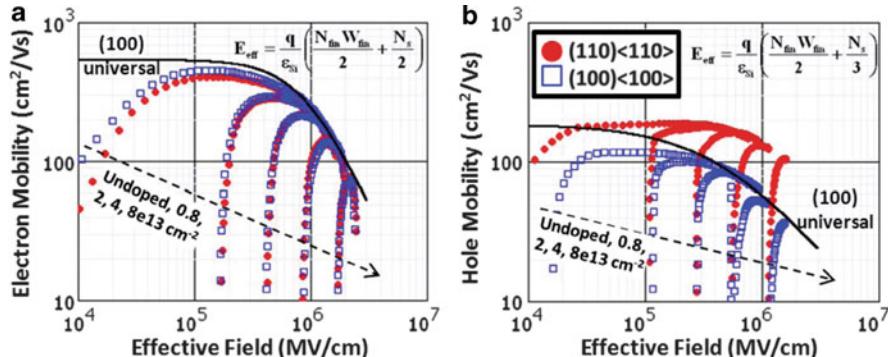


On the other hand, the strong surface orientation dependence shown in Fig. 3.2a for undoped PFET mobility remains valid for doped FinFETs from inversion to accumulation modes (Fig. 3.15). An interesting observation in Fig. 3.15 is that the majority carrier AM hole mobility degrades substantially less than does the minority carrier IM electron mobility with the fin doping such that, at high B doses, AM PFET mobility is equal to or better than that of IM NFET [49]. This is even true even for (100) surface, where the mobility difference between electrons and holes is substantial ( $\sim 3 \times$ ) for undoped fins. As a sanity check to our mobility extraction, the variation of peak transconductance (proportional to low-field mobility,  $\mu_0$ ) is shown as a function of implant doses for IM/AM n/p-channel FinFETs in Fig. 3.16, which confirmed that the general trends are similar to those of high field mobility.

Figure 3.17 shows the IM FinFET electron and hole mobilities for various implant doses as a function of effective field given by:

$$E_{\text{eff}} = \frac{|Q_d| + \eta |N_s|}{\epsilon_{\text{Si}}}, \quad (3.4)$$

where  $Q_d = (N_{\text{fin}}W_{\text{fin}})/2$  is the depletion charge,  $\eta$  is the empirical coefficient enabling a “universal” (e.g., doping, substrate bias, and gate oxide thickness-independent) mobility vs. effective field behavior, and  $\epsilon_{\text{Si}}$  is the Si permittivity [51]. At high enough  $E_{\text{eff}}$ , where surface roughness scattering dominates the IM FET mobility, (100) and (110) electron mobilities are in agreement with Takagi’s universal (100) electron mobility if  $\eta = 1/2$  is used for both orientations (Fig. 3.17a) ( $\eta = 1/3$ , the recommended value for (110) poly/SiO<sub>2</sub> planar NFETs [51], did not yield a doping-independent trend for (110) IM n-channel FinFETs hence not used; similar experimental findings were also reported in [1]). At high  $E_{\text{eff}}$ , there is also agreement between (100) hole mobility data and universal (100) hole mobility curve for  $\eta = 1/3$ , Fig. 3.17b [51]. Interestingly, for  $E_{\text{eff}} > 0.2$  MV/cm and  $\eta = 1/3$ , (110) hole mobility also exhibits a doping-independent trend.



**Fig. 3.17** IM FinFET (100) and (110) electron and hole mobilities versus  $E_{\text{eff}}$ .  $\eta = 1/2$  for electrons and  $\eta = 1/3$  for holes were assumed irrespective of the surface orientation [46]. (100) universal curves were plotted using compact models in [50]

**Fig. 3.18** Inversion mode electron and hole mobility versus fin doping (extracted from threshold voltage data from the same fins) at constant  $E_{\text{eff}}$  of 1 MV/cm for electrons and 0.7 MV/cm for holes.  $U(n)$  and  $U(p)$  designate the (100) universal mobility values at the respective  $E_{\text{eff}}$  [46]

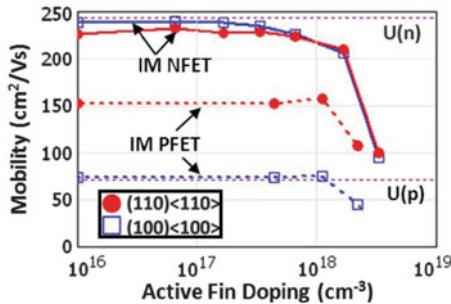
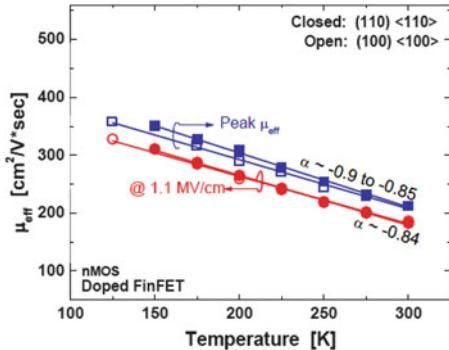


Figure 3.18 shows the minority carrier (inversion mode) mobility at fixed effective field as a function of fin doping. For doping values smaller than about  $10^{18} \text{ cm}^{-3}$ , the inversion carrier density  $N_s$  is high, surface roughness scattering dominates, and mobility is doping independent. For higher fin doping, the inversion carrier density is not high enough to screen the depletion charge, hence ionized impurity scattering dominates, lowering the mobility with increased fin doping. As in Fig. 3.15, fixed  $E_{\text{eff}}$  comparison in Fig. 3.18 also confirms that (110)/<110> minority carrier electron mobility is very close to that of (100)/<100> whereas (110)/<110> minority carrier hole mobility is about twice as high as that of (100)/<100>.

In order to elucidate the mobility behavior of doped FinFETs even further, low temperature measurements on IM n-channel FinFETs implanted with a B dose of  $4.10^{13} \text{ cm}^{-3}$  were performed. Figure 3.19 shows the variation of high field and peak mobility values down to 50 K for (100)/<100> and (110)/<110> devices. Compared to the undoped channel low temperature data in Fig. 3.4, it is observed that the exponents are decreased by more than  $2\times$ , confirming that in doped FinFETs the mobility should indeed be limited by a mechanism other than the phonon scattering. In this case, the limiting mechanism is ionized impurity

**Fig. 3.19** Doped IM n-channel DG Si FinFET peak and high field mobility as a function temperature for  $(110)<110>$  and  $(100)<100>$  combinations [9]

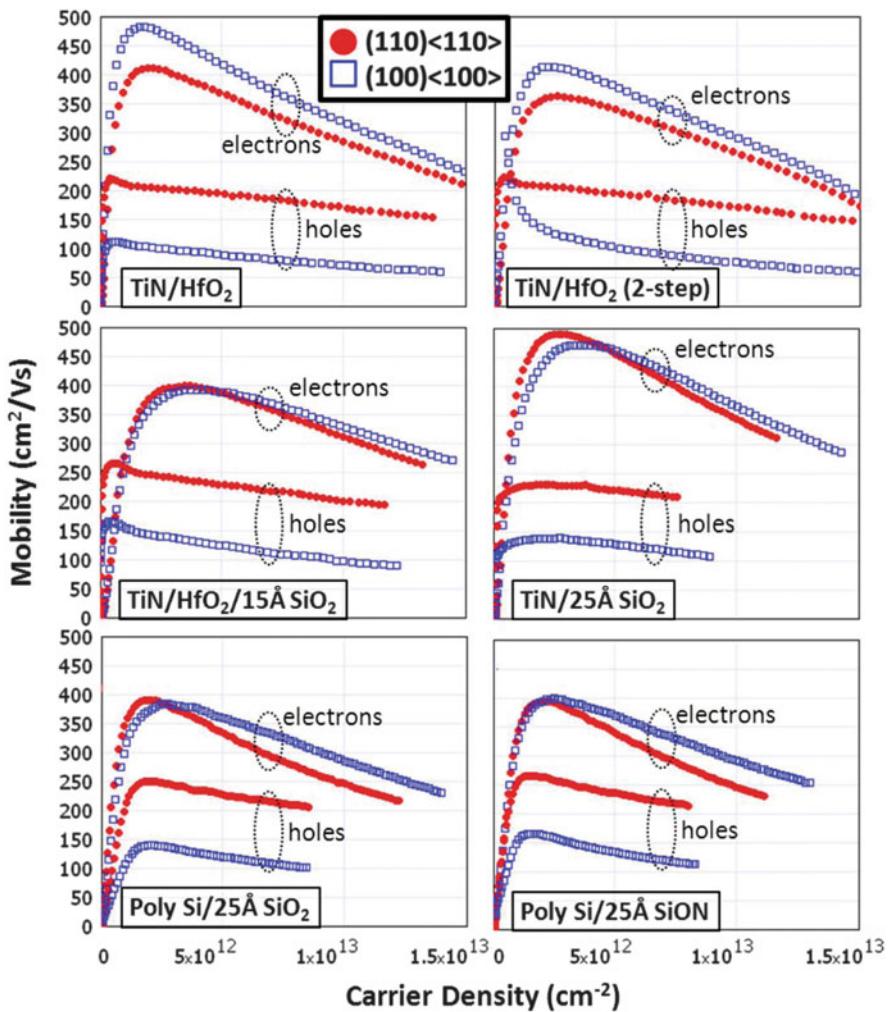


scattering. Accordingly a decreased sensitivity to surface orientation and vertical field is also visible irrespective of the temperature. Note, in particular, that at 1.1 MV/cm, the (100) and (110) doped FinFET electron mobilities are essentially the same across the entire temperature range under study.

### 3.5 Impact of Gate Stack

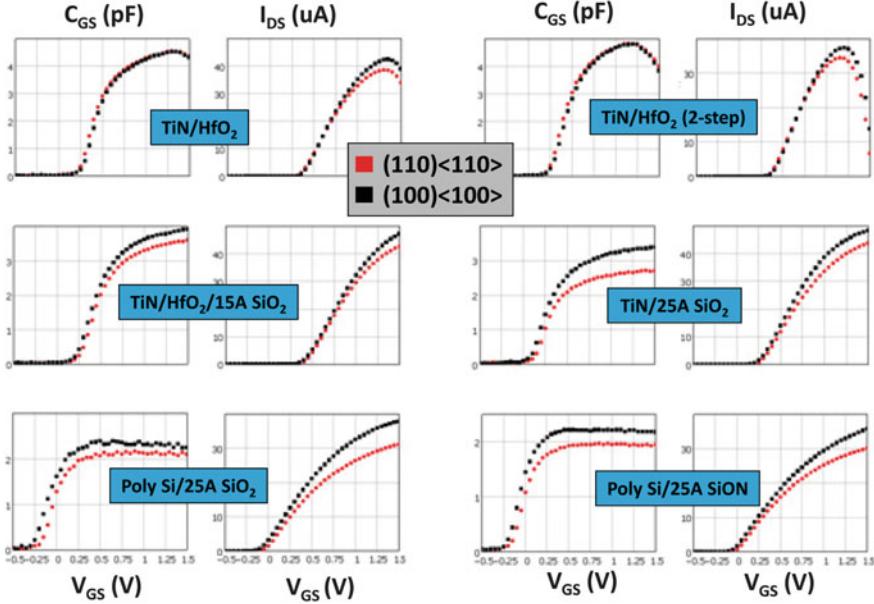
As a parallel experiment to the fin doping study, undoped SOI FinFETs with various gate stacks (TiN/HfO<sub>2</sub>, TiN/HfO<sub>2</sub>/SiO<sub>2</sub>, TiN/SiO<sub>2</sub>, poly/SiO<sub>2</sub>, and poly/SiON) were fabricated. Electron and hole mobilities for (110) and (100) sidewalls, along with the corresponding gate stack splits are shown in Fig. 3.20. From one split to another, TiN, poly-Si deposition and their etch processes were identical. Only the insulator etch had to be individually tuned for each split. For experiments using intentional thermal SiO<sub>2</sub> growth, the intended SiO<sub>2</sub> thicknesses are provided in Fig. 3.20. For splits involving HfO<sub>2</sub> without thermal oxidation, an interfacial SiO<sub>x</sub> layer (chemical oxide) in the range of 1 nm was present between Si and HfO<sub>2</sub>. The “two-step high-K” split mentioned in Fig. 3.20 enables a reduction of the equivalent insulator thickness in inversion ( $T_{\text{inv}}$ ) due to a combination of both compositional and structural changes as a result of the deposition/anneal/deposition/anneal sequence (rather than a single deposition and anneal) [52]. As mentioned in Introduction, the target physical thickness was 2 nm for HfO<sub>2</sub> and 10 nm for TiN.

For all stacks, surface orientation dependence of electron mobility is observed to be weak, and surface orientation dependence of hole mobility is observed to be very strong as in the case of doped fins reported in the previous subchapter. In particular, (110)/<110> electron mobility is very high ( $\geq 240 \text{ cm}^2/\text{V s}$  at  $N_s = 10^{13} \text{ cm}^{-2}$ ) and remains within 15 % of the (100) electron mobility irrespective of the gate stack. This result does not correlate with [53] where the closeness of (110)/<110> and (100)/<100> electron mobilities were attributed to metal gate (TiSiN) stress that only improves (110)/<110> electron mobility without having substantial impact on (100)/<100> electron mobility.



**Fig. 3.20** Conventional and rotated undoped DG FinFET electron and hole mobilities for various gate stacks [46]

When a thermal  $\text{SiO}_2$  layer is intentionally grown at the interface (bottom four splits in Fig. 3.20), it is observed that the peak electron mobility is essentially the same for the  $(110)/<110>$  and  $(100)/<100>$ . In the poly/ $\text{SiO}_2$  and poly/SiON devices, the  $(110)$  electron mobility degrades visibly faster with  $N_s$  than does the  $(100)$  electron mobility, suggesting that the surface roughness scattering is more significant for a  $(110)$  than for a  $(100)$  surface for those particular gate stacks [10]. Accordingly at high  $N_s$ , the difference between  $(100)$  and  $(110)$  electron mobility can be expected to increase with fin doping (compared to undoped fin case) in doped poly/ $\text{SiO}_2$  and poly/SiON FinFETs (which were not a part of this

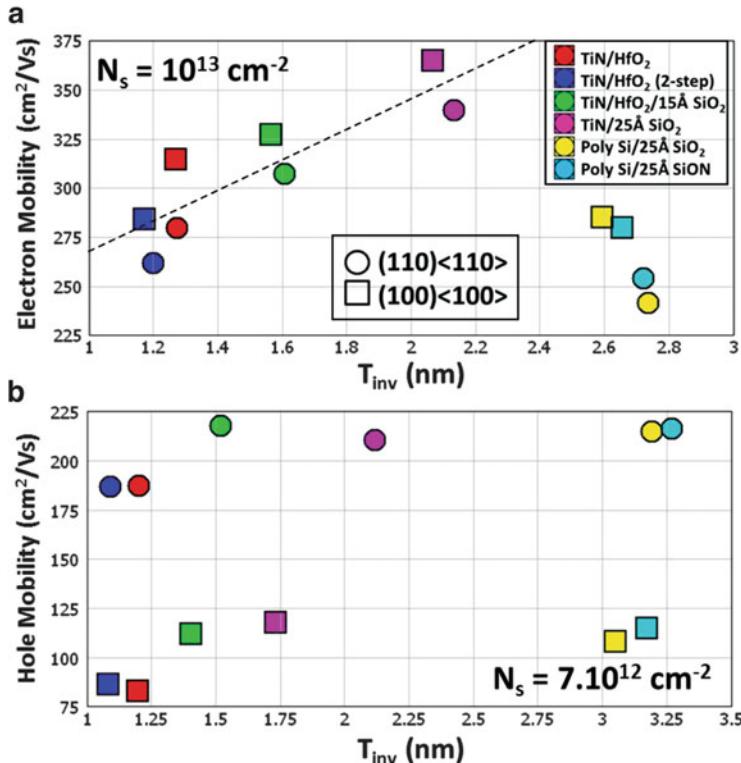


**Fig. 3.21** n-Channel conventional and rotated undoped DG FinFET long-channel  $C-V$  and  $I-V$  curves that were used to extract the electron mobility in Fig. 3.20 for various gate stacks [31]

study). Figure 3.21 shows the long-channel NFET  $I-V$  and  $C-V$  curves that were used to extract the electron mobility in Fig. 3.20.

Figure 3.22 summarizes the mobility trends in Fig. 3.20 as a function of  $T_{inv}$  at a constant carrier density (arbitrarily taken as  $N_s = 10^{13} \text{ cm}^{-2}$  for electrons and  $N_s = 7.10^{12} \text{ cm}^{-2}$  for holes). Again, Fig. 3.22a confirms that electron mobility values for (100) and (110) are very close to each other for all gate stack combinations. Polygate devices have the lowest mobility among all stacks. A possible yet speculative reason for reduced electron mobility in poly gate FinFETs compared to metal gate FinFETs is the absence of beneficial tensile metal-gate stress and/or additional scattering mechanisms. However, if the metal gate improvement hypothesis is correct, this would imply that the metal-gate stress improves the electron mobility for both (110)/<110> and (100)/<100>, which does not correlate with the orientation-dependent improvement assumption of [53, 54]. TiN/SiO<sub>2</sub> FinFETs exhibit the highest electron mobility possibly because they benefit from the metal-gate stress while being immune to additional scattering mechanisms from high-K [24]. Otherwise for all TiN gate FinFETs, a tradeoff between electron mobility and  $T_{inv}$  is observed [55].

With regard the PFET trends, Fig. 3.22b confirms again that the dependence to surface orientation is substantial for all gate stacks [56], while the dependence to  $T_{inv}$  is weak except the degradation when high-K is directly deposited on fins. Tensile metal-gate stress that we speculated to improve the electron mobility compared to polygate devices (Fig. 3.22a) does not degrade the hole mobility that



**Fig. 3.22** (a) Electron mobility versus  $T_{inv}$  at  $N_s = 10^{13} \text{ cm}^{-2}$  ( $T_{inv}$  is extracted from NFET CV curves at  $V_{GS} = 1.1 \text{ V}$ ). *Dashed line* is the best fitting line to TiN gate FinFET data points and illustrates the tradeoff between mobility and  $T_{inv}$ ). (b) Hole mobility versus  $T_{inv}$  at  $N_s = 7 \times 10^{12} \text{ cm}^{-2}$  ( $T_{inv}$  is extracted from PFET CV curves at  $V_{GS} = -1.5 \text{ V}$ ) [46]

is observed to be insensitive to the gate electrode if SiO<sub>2</sub> is used as gate insulator. This is consistent with the very low sensitivity of hole mobility to tensile CESL stress, shown in Fig. 3.7.

Another observation from Fig. 3.22 is that when a thermal SiO<sub>2</sub> or SiON layer is intentionally grown at the interface, (110) FinFETs tend to have a higher  $T_{inv}$  than do the (100) FinFETs due to higher oxidation rate on (110) [10, 11]. However, this trend breaks down for high-K/metal gate (HK/MG) FinFETs where  $T_{inv}$  is extracted to be very similar/the same for different surface orientations as confirmed by the CV data in Fig. 3.21. The effectiveness of the two-step ALD (“dep-anneal-dep-anneal process” [52]) HfO<sub>2</sub> in terms of  $T_{inv}$  scaling, compared to single-step (DA) process, is also worth noticing together with the indication of increased gate leakage in Fig. 3.21.

### 3.6 Conclusion

The impact of the surface orientation, strain, fin doping, and gate stack on SOI double-gate FinFET mobility is systematically investigated. Impact of channel material, temperature, and fin width were also touched upon to better understand the trends. For the unstrained case,  $(110)/<110>$  electron mobility is within 10–15 % of  $(100)/<100>$  electron mobility irrespective of the fin doping level and gate stack. In contrast,  $(110)/<110>$  hole mobility is  $\sim 2\times$  higher than the  $(100)/<100>$  hole mobility for all conditions. Both the  $(100)/<100>$  and  $(110)/<110>$  electron mobility can be improved with tensile strain. It is also confirmed that  $(110)/<110>$  FinFET hole mobility can be significantly improved with compressive stress while  $(100)/<100>$  hole mobility is sensitive to neither compressive nor tensile stress. Switching from Si to SiGe channel (leveraging the inherent substrate induced compressive stress) increases the hole mobility drastically and even further improvement is available by process-induced compressive stress. Overall, the experimental results in this chapter suggest that the  $(110)/<110>$  Si FinFETs conventionally built on standard  $(100)$  wafers offer simultaneously high electron and hole mobility, which can be further improved by tensile and compressive stress, respectively.

## References

- Lin, C.-H., Kambhampati, R., Miller, R.J., Hook, T.B., Bryant, A., Haensch, W. et al.: Channel doping impact on FinFETs for 22nm and beyond. In: Symposium on VLSI technology. Digest of technical papers, pp. 15–16 (2012)
- Lochtefeld, A., Djomehri, I.J., Samudra, G., Antoniadis, D.A.: New insights into carrier transport in n-MOSFETs. IBM J. Res. Dev. **46**(2/3), 347–357 (2002)
- Nayfeh, O.M., Yu, S., Antoniadis, D.A.: On the relationship between carrier mobility and velocity in sub-50 nm MOSFETs via calibrated Monte Carlo simulation. In: SISPAD, pp.117–120 (2004)
- Khakifirooz, A., Antoniadis, D.: Transistor performance scaling: the role of virtual source velocity and its mobility dependence. In: IEDM technical digest, pp. 667–670 (2006)
- Jan, C.-H., Bhattacharya, U., Brain, R., Choi, S.-J., Curello, G., Gupta, G. et al.: A 22nm SoC platform technology featuring 3-D tri-gate and high-k/metal gate, optimized for ultra low power, high performance and high density SoC applications. In: IEDM technical digest, pp. 44–47 (2012)
- Akarvardar, K., Young, C.D., Veksler, D., Ang, K.-W., Ok, I., Rodgers, M., et al.: Performance and variability in multi-VT FinFETs using fin doping. In: Proceedings of the VLSI TSA, pp. 1–2 (2012)
- Young, C.D., Akarvardar, K., Baykan, M.O., Matthews, K., Ok, I., Ngai, T., et al.:  $(110)$  and  $(100)$  Sidewall-oriented FinFETs: a performance and reliability investigation. Solid State Electron. **78**, 2–10 (2012)
- Taur, Y., Ning, T.: Fundamentals of modern VLSI devices. Cambridge University Press, Cambridge (1998)

9. Young, C.D., Akarvardar, K., Matthews, K., Baykan, M.O., Pater, J., Ok, I., et al.: Electrical characterization and reliability assessment of double-gate FinFETs. *ECS Trans.* **50**(4), 201–206 (2013)
10. Momose, H.S., Ohguro, T., Kojima, K., Nakamura, S.-I., Toyoshima, Y.: 1.5-nm gate oxide CMOS on (110) surface-oriented Si substrate. *IEEE Trans. Electron Dev.* **50**(4), 1001–1008 (2003)
11. Yang, M., Chan, V., Chan, K., Shi, L., Fried, D., Stathis, J., et al.: Hybrid-orientation technology (HOT): opportunities and challenges. *IEEE Trans. Electron Dev.* **53**(5), 965–978 (2006)
12. Mereu, B., Rossel, C., Gusev, E.P., Yang, M.: The role of Si orientation and temperature on the carrier mobility in metal oxide semiconductor field-effect transistors with ultrathin  $\text{HfO}_2$  gate dielectrics. *J. Appl. Phys.* **100**, 014504 (2006)
13. Trojman, L., Pantisano, L., Ferain, I., Severi, S., Maes, H.E., Groeseneken, G.: Mobility and dielectric quality of 1-nm EOT  $\text{HfSiON}$  on Si(110) and (100). *IEEE Trans. Electron Dev.* **55**(12), 3414–3420 (2008)
14. Sayama, H., Nishida, Y., Oda, H., Oishi, T., Shimizu, S., Kunikiyo, T., et al.: Effect of <100> channel direction for high performance SCE immune pMOSFET with less than  $0.15\mu\text{m}$  gate length. In: IEDM technical digest, pp. 657–660 (1999)
15. Matsumoto, T., Maeda, S., Dang, H., Uchida, T., Ota, K., Hirano, Y., et al.: Novel SOI wafer engineering using low stress and high mobility CMOSFET with <100>-channel for embedded RF/analog applications. In: IEDM technical digest, pp. 663–666 (2002)
16. Lee, C.-W., Afzalian, A., Ferain, I., Yan, R., Dehdashti, N., Byun, K.-Y., et al.: Comparison of different surface orientation in narrow fin MuGFETs. *Microelectron. Eng.* **86**, 2381–2384 (2009)
17. Chowdhury, M.M., Fossum, J.G.: Physical insights on electron mobility in contemporary FinFETs. *IEEE Electron Device Lett.* **27**(6), 482–485 (2006)
18. Arora, N.D., Hauser, J.R., Roulston, D.J.: Electron and hole mobilities in silicon as a function of concentration and temperature. *IEEE Trans. Electron Dev.* **ED-29**, 292–295 (1982)
19. Fischetti, M.V., Laux, S.E.: Band structure, deformation potentials, and carrier mobility in strained Si, Ge, and SiGe alloys. *J. Appl. Phys.* **80**, 2234 (1996)
20. Arora, N.D., Gildenblat, G.S.H.: A semi-empirical model of the MOSFET inversion layer mobility for low-temperature operation. *IEEE Trans. Electron Dev.* **ED-34**, 89–93 (1987)
21. Jeon, D.S., Burk, D.A.: MOSFET electron inversion layer mobilities – a physically based semi-empirical model for a wide temperature range. *IEEE Trans. Electron Dev.* **36**(8), 1453–1463 (1989)
22. Takagi, S., Toriumi, A., Iwase, M., Tango, H.: On the universality of inversion layer mobility in SiMOSFETs: part I—effects of substrate impurity concentration. *IEEE Trans. Electron Dev.* **41**(12), 2357–2368 (1994)
23. Krishnan, S.A., Rusty Harris, H., Kirsch, P.D., Krug, C., Quevedo-Lopez, M., Young, C. et al.: High performing pMOSFETs on Si(110) for application to hybrid orientation technologies – comparison of  $\text{HfO}_2$  and  $\text{HfSiON}$ . In: IEDM technical digest, pp. 1–4 (2006)
24. Tonutti, P., Palestri, P., Esseni, D., Driussi, F., De Michielis, M., et al.: On the origin of the mobility reduction in n- and p-metal–oxide–semiconductor field effect transistors with hafnium based metal gate stacks. *J. Appl. Phys.* **112**, 034502 (2012)
25. Baykan, M.O., Young, C.D., Akarvardar, K., Majhi, P., Hobbs, C., Kirsch, P., et al.: Physical insights on comparable electron transport in (100) and (110) double-gate fin field-effect transistors. *Appl. Phys. Lett.* **100**, 123502 (2012)
26. Irisawa, T., Numata, T., Tezuka, T., Usuda, K., Sugiyama, N., Takagi, S.-I.: Device design and electron transport properties of uniaxially strained-SOI tri-gate nMOSFETs. *IEEE Trans. Electron Dev.* **55**(2), 649–654 (1994)
27. Young, C.D., Baykan, M.O., Agrawal, A., Madan, H., Akarvardar, K., Hobbs, C., et al.: Critical discussion on (100) and (110) orientation dependent transport: nMOS Planar and FinFET. In: Symposium on VLSI technology. Digest of technical papers, pp. 18–19 (2011)

28. Balestra, F., Cristoloveanu, S., Benachir, M., Brini, J., Elewa, T.: Double-gate silicon-on-insulator transistor with volume inversion: a new device with greatly enhanced performance. *IEEE Electron Device Lett.* **8**(9), 410–412 (1987)
29. Auth, C., Cappellani, A., Chun, J.-S., Dalis, A., Davis, A., Ghani, T., et al.: 45nm high-k + metal gate strain-enhanced transistors. In: *Symposium on VLSI technology. Digest of technical papers*, pp. 128–129 (2008)
30. Suthram, S., Hussain, M.M., Harris, H.R., Smith, C., Tseng, H.-H., Jammy, R., et al.: Comparison of uniaxial wafer bending and contact-etch-stop-liner stress induced performance enhancement on double-gate FinFETs. *IEEE Electron Device Lett.* **29**(5), 480–482 (2008)
31. Akarvardar, K., et al.: Unpublished SEMATECH data (2011)
32. Akarvardar, K., Rodgers, M., Kaushik, V., Johnson, C.S., Chong, H., Ok, I., et al.: Impact of NiPt thickness scaling on contact resistance from thin-body FD SOI to trigate FETs. *IEEE Electron Device Lett.* **33**(5), 631–633 (2012)
33. Chu, M., Sun, Y., Aghoram, U., Thompson, S.E.: Strain: a solution for higher carrier mobility in nanoscale MOSFETs. *Annu. Rev. Mater. Res.* **39**, 203–229 (2009)
34. Shin, K., Chui, C.O., King, T.-J.: Dual stress capping layer enhancement study for hybrid orientation FinFET CMOS technology. In: *IEDM technical digest*, pp. 988–991 (2005)
35. Ghibaudo, G.: New method for the extraction of MOSFET parameters. *Electron. Lett.* **24**(9), 543–545 (1988)
36. Irisawa, T., Numata, T., Tezuka, T., Usuda, K., Nakaharai, S., Hirashita, N., et al.: High performance multi-gate pMOSFETs using uniaxially-strained SGOI channels. In: *IEDM technical digest*, pp. 709–712 (2005)
37. Ok, I., Akarvardar, K., Lin, S., Baykan, M.O., Young, C.D., Hung, P.Y., et al.: Strained SiGe and Si FinFETs for high performance logic with SiGe/Si stack on SOI. In: *IEDM technical digest*, pp. 776–779 (2010)
38. Smith, C.E., Adhikari, H., Lee, S.-H., Coss, B., Parthasarathy, S., Young, C., et al.: Dual channel FinFETs as a single high-k/metal gate solution beyond 22nm node. In: *IEDM technical digest*, pp. 1–4 (2009)
39. Akarvardar, K., Ok, I., et al.: Unpublished SEMATECH data (2011)
40. Matthews, J.W., Blakeslee, A.E.: Defects in epitaxial multilayers: I. Misfit dislocations. *J. Cryst. Growth* **27**, 118 (1974)
41. Nadeau, J., Deeb, C., Hung, P.Y., Ok, I., Hobbs, C.: Application of scanning transmission electron microscopy (STEM) – based techniques for development of novel Si/SiGe on SOI FinFET structures. In: *Proceedings of the 2011 Frontiers of characterization and metrology conference* (2011)
42. Cheng, K., Khakifirooz, A., Loubet, N., Luning, S., Nagumo, T., Vinet, M., et al.: High performance extremely thin SOI (ETSOI) hybrid CMOS with Si channel NFET and strained SiGe channel PFET. In: *IEDM technical digest*, pp. 419–422 (2012)
43. Andrieu, F., Ernst, T., Romanjek, K., Weber, O., Renard, C., Hartmann, J.-M., et al.: SiGe channel p-MOSFETs scaling-down. In: *Proceedings of the ESSDERC*, pp. 267–270 (2003)
44. Veloso, A., Witters, L., Demand, M., Feraim, I., Son, N.J., Kaczer, B., et al.: Flexible and robust capping-metal gate integration technology enabling multiple-VT CMOS in MuGFETs. In: *Symposium on VLSI technology. Digest of technical papers*, pp. 14–15 (2008)
45. Singanamalla, R., Boccardi, G., Tseng, J., Petry, J., Vellianitis, G., van Dal, M.J.H., et al.: Multi-VT engineering in highly scaled CMOS bulk and FinFET devices through ion implantation into the metal gate stack featuring a 1.0nm EOT High-K oxide. In: *Proceedings of the VLSI-TSA technical digest*, pp. 112–113 (2010)
46. Akarvardar, K., Young, C.D., Baykan, M.O., Ok, I., Ngai, T., Ang, K.-W., et al.: Impact of fin doping and gate stack on FinFET (110) and (100) electron and hole mobilities. *IEEE Electron Dev. Lett.* **33**(3), 351–353 (2012)
47. Cheng, W., Teramoto, A., Hirayama, M., Sugawa, S., Ohmi, T.: Impact of improved high-performance Si (110)-oriented metal-oxide-semiconductor field-effect transistors using accumulation-mode fully depleted silicon-on-insulator devices. *Jpn. J. Appl. Phys.* **45**, 3110–3116 (2006)

48. Colinge, J.-P., Lee, C.-W., Afzalian, A., Akhavan, N.D., Yan, R., Ferain, I., et al.: Nanowire transistors without junctions. *Nat. Nanotechnol.* **5**, 225–229 (2010)
49. Afzalian, A., Lederer, D., Lee, C.-W., Yan, R., Xiong, W., Cleavelin, C., et al.: Multi-gate MOSFETs: accumulation versus enhancement mode. In: Proceedings of the IEEE silicon nanoelectron workshop, pp. 1–6 (2008)
50. Chen, K., Wann, H.C., Ko, P.K., Hu, C.: The impact of device scaling and power supply change on CMOS gate performance. *IEEE Electron Device Lett.* **17**(5), 202–204 (1996)
51. Takagi, S., Toriumi, A., Iwase, M., Tango, H.: On the universality of inversion layer mobility in Si MOSFET's: part II-effects of surface orientation. *IEEE Trans. Electron Dev.* **41**(12), 2363–2368 (1994)
52. Clark, R.D., Aoyama, S., Consiglio, S., Nakamura, G., Leusink, G.J.: Physical and electrical effects of the Dep-Anneal-Dep-Anneal (DADA) process for HfO<sub>2</sub> in high K/metal gate stacks. *ECS Trans.* **35**(4), 815–834 (2011)
53. Xiong, W.W., Cleavelin, C.R., Hsu, C.-H., Ma, M., Schruefer, K., Von Arnim, K., et al.: Intrinsic advantages of SOI multiple-gate MOSFET (MuGFET) for low power applications. *ECS Trans.* **6**(4), 59–69 (2007)
54. Conzatti, F., Serra, N., Esseni, D., De Michielis, M., Paussa, A., Palestri, P., et al.: Investigation of strain engineering in FinFETs comprising experimental analysis and numerical simulations. *IEEE Trans. Electron Dev.* **58**(6), 1583–1593 (2011)
55. Muller, M., Duguay, S., Guillaumot, B., Garros, X., Leroux, C., Tavel, B., et al.: Towards a better EOT-mobility trade-off in high-K oxide/metal gate CMOS devices. In: Proceedings of the ESSDERC, pp. 367–370 (2003)
56. Yang, M., Gusev, E.P., Ieong, M., Gluschenkov, O., Boyd, D.C., Chan, K.K.: Performance dependence of CMOS on silicon substrate orientation for ultrathin oxynitride and HfO<sub>2</sub> gate dielectrics. *IEEE Electron Device Lett.* **24**(5), 339–341 (2003)
57. Endo K, Ishikawa Y, Liu Y, Masahara M, Matsukawa T, O'uchi S-I, Ishii K, Yamauchi H, Tsukada J, Suzuki E.: Experimental Evaluation of Effects of Channel Doping on Characteristics of FinFETs. *IEEE Electron Device Lett.* **28**(12), 1123–1125 (2007)

# Chapter 4

## Quantum Mechanical Potential Modeling of FinFET

Balwinder Raj

**Abstract** This chapter focus on a full Two-Dimensional (2D) Quantum Mechanical (QM) analytical modeling in order to evaluate the 2D potential profile within the active area of FinFET structure. Various potential profiles such as surface, back to front gate, and source to drain potential have been presented in order to appreciate the usefulness of the device for circuit simulation purposes. As we move from source end of the gate to the drain end of the gate, there is substantial increase in the potential at any point in the channel. This is attributed to the increased value of longitudinal electric field at the drain end on application of a drain to source voltage. Further, in this chapter, the detailed study of threshold voltage and its variation with the process parameters is presented. A threshold voltage roll-off with fin thickness is observed for both theoretical and experimental results. The fin thickness is varied from 10 to 60 nm. From the analysis of S/D resistance, it is observed that for a fixed fin width, as the channel length increases, there is an enhancement in the parasitic S/D resistance. This can be inferred from the fact as the channel length decreases, quantum confinement along the S/D direction becomes more extensive. For our proposed devices a close match is obtained with the results through analytical model and reported experimental results, thereby validating our proposed QM analytical model for DG FinFET device.

### 4.1 Introduction

The scaling of CMOS structure is approaching its limits; multiple gate architecture such as Double Gate FinFET structure presents significant advantages to fulfill long range International Technology Roadmap for Semiconductors (ITRS) [1] requirements. The Poisson's equation-based numerical modeling of Double Gate

---

B. Raj (✉)

Department of Electronics and Communication Engineering, National Institute of Technology Jalandhar, Jalandhar 144011, Punjab, India  
e-mail: [balwinderraj@gmail.com](mailto:balwinderraj@gmail.com)

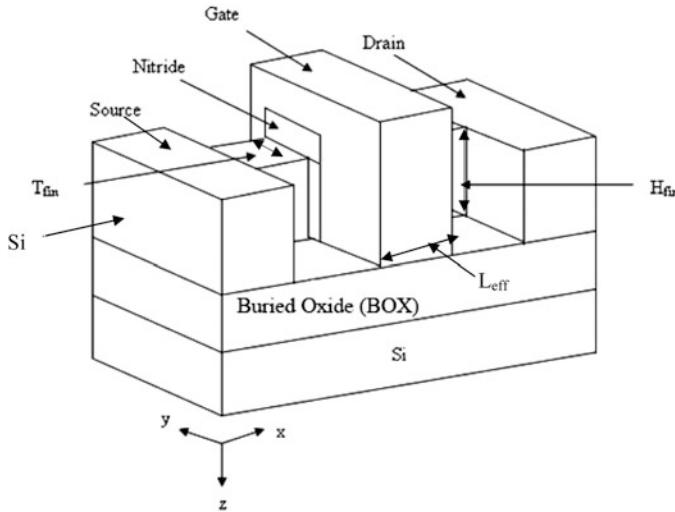
FinFET device has been carried out by many workers [2–6] which presents generic implicit surface potential solution for FinFET device. For nanoscale multi-gate devices, two-dimensional analytical approach would be required which will be valid under Quantum Mechanical (QM) domain of FinFET device under study. For this purpose, we present a fully quantum mechanical surface potential model for the channel region of FinFET device using analytical modeling.

For a CMOS technology to keep pace with downscaling, improved carrier transport and low parasitic source/drain resistance are required. Pei et al. [7] investigated FinFET simulation and analytical modeling. Double gate FinFET has been considered as one of the most promising candidate for sub-50 nm designs. But double gate structure suffers from possible misalignment between source/drain with gate region, thereby increasing the overlap capacitances as well as source to drain series resistance. This would result in a slower device, and hence high frequency operation of the device would be restricted. Fin height and Fin thickness are modified in order to achieve optimized operation of the device. Potential in the active area of FinFET device and threshold voltage have also been evaluated. Dixit et al. [8] used a 45 nm FinFET structure to understand the implication of source/drain resistance on the device characteristics. Sub-20 nm FinFET using SiGe as a gate material was developed by Hisamoto et al. [9]; they showed the ease of fabrication using planer MOSFET process technology. Double Gate FinFET structure offers higher driving capabilities and reduces SCE [10, 11]. To develop sub-50 nm MOSFETs, double gate FinFET structure has been widely studied [12–14]. For the double gate MOSFETs, the gate controls the energy barrier between source and drain effectively [1, 12, 15–17]. Further studies have shown [18–22] that controlling threshold voltage and parasitic for ultrathin body is a difficult task.

The threshold voltage of a transistor is one of the key parameters in the design of FinFET circuits. Katti et al. [23] have modeled fully depleted SOI MOSFETs using the solution of three-dimensional (3-D) Poisson's equation. As the device dimensions continue to scale down to deep sub-micrometer regime to obtain better performance, analytical modeling of these devices becomes even more challenging. Although Kedzierski et al. [24] have addressed this issue and proposed a technology solution, an analytical understanding of parasitic series resistance in the FinFET device is desirable. In this chapter, a full quantum mechanical analytical modeling for FinFET structure has been carried out. The subsequent section deals with 3D FinFET structure followed by quantum mechanical potential modeling, threshold voltage modeling, and source/drain (S/D) resistance modeling. The results obtained based on our model are compared and contrasted with reported, experimental, and simulated results for the purpose of validation and verification of our proposed analytical model.

## 4.2 FinFET Structure

Figure 4.1 shows 3D view of FinFET. The gate “wraps” over the thin Si Fin, yielding a quasi-planar symmetrical double-gate FinFET structure with two inversion channels that are charge coupled. Both the front and back gates might have the



**Fig. 4.1** Structure of FinFET

same work function. They are further tied to same applied potential. The key challenges in the fabrication of Double Gate FinFET devices are [25–28] self-alignment of the two gates and formation of an ultrathin silicon film. In FinFET device, the fin is a narrow channel of silicon patterned on an SOI wafer. The gate wraps around the fin on three faces. The top insulator is usually thicker than the side insulators; hence, the device has effectively two channels. The top insulator may be reduced in thickness in order to control the channel as well.

#### 4.2.1 FinFET Design Parameters

FinFET parameters are indicated in the Fig. 4.1. The definitions of the various parameters are:  $L_{\text{eff}}$ : effective channel length of FinFET, which is the actual distance between source and drain region,  $H_{\text{fin}}$ : height of silicon fin defined by the distance between top gate and buried oxide,  $T_{\text{fin}}$ : thickness of silicon fin defined as the distance between front and back gate oxides,  $W_{\text{fin}}$ : geometrical channel width defined as:  $W_{\text{fin}} = (2 \times H_{\text{fin}}) + T_{\text{fin}}$ . When the thickness of silicon film ( $T_{\text{fin}}$ ) is much larger than its height ( $H_{\text{fin}}$ ) or when top gate oxide is much thinner than the front and back oxides, FinFET can be approximately treated as single-gate fully depleted (FD) SOI MOSFET [7, 29, 30].

On the other hand, when height of the silicon film ( $H_{\text{fin}}$ ) is much larger than its thickness ( $T_{\text{fin}}$ ) or top gate oxide is much thicker than the front and back oxides, FinFET can be approximately treated as Double Gate FET device. The two limits of FinFET, namely, FD-FET and DG-FET have been widely studied and well understood [3, 20, 21, 31, 32]. To our understanding in the regime where both fin height and thickness have control over short channel effects (SCE), the dependence

of SCE on device dimensions is not well extracted or known. For the purpose of understanding the dependence of output characteristics of FinFET with respect to various device/process parameters, a full quantum mechanical analytical potential modeling is carried out in the next section.

### 4.3 Quantum Mechanical Potential Modeling

In order to extract full two-dimensional potential profile within the active area of the device, QM solution is carried out. For this purpose, several methods have been proposed [33–36], where the potential function is divided into two parts, the first one being the long channel solution and the second one, a short channel evaluation. But the evaluation of short channel term takes into account the functional dependence of device parameters, which is a complicated issue and takes large computational time. For the purpose of simplification and also to have a reduced complexity in time, we have assumed the following dependence of potential, where two-dimensional potential is broken down into 1D surface potential and a 2D function [37] as given below:

$$\psi(x, y) = \psi_s(x) \times A(x, y) \quad (4.1)$$

where  $\psi_s(x)$  is the surface potential and  $A(x, y)$  is the vertical distribution of the envelop function.

$A(x, y)$  as given in (4.1) can be written as [37]:

$$A(x, y) = \frac{Z(x, y)}{Z(x, y = 1)} \quad (4.2)$$

where  $Z(x, y)$  can be written as [37]

$$Z(x, y) = \psi_0(x) - \frac{2}{\beta} \ln \left\{ \cos \left[ \sqrt{\frac{q^2 n_i}{2kT\epsilon_{Si}}} e^{\frac{\beta(\psi_0(x) - V_F(x))}{2}} \left( y - \frac{T_{fin}}{2} \right) \right] \right\} \quad (4.3)$$

The behavior of center potential  $\psi_0(x)$  as a function of effective gate voltage is given as [3]:

$$\psi_0(x) = U - \sqrt{U^2 - (V_{gs} - V_{FB})\psi_{0max}(x)} \quad (4.4)$$

where  $\psi_{0max}(x)$  is the maximum potential that can be obtained at the center of the channel under a given bias at the terminal and  $U$  is given as

$$U = \frac{1}{2} [(V_{gs} - V_{FB}) + (1 + r)\psi_{0max}(x)] \quad (4.5)$$

$\psi_{0\max}(x)$  can be evaluated as:

$$\psi_{0\max}(x) = V_F(x) + \frac{1}{\beta} \ln \left( \frac{2\pi^2 \epsilon_{\text{Si}} kT}{q^2 n_i T_{\text{fin}}^2} \right) \quad (4.6)$$

where  $r$  in (4.5) is defined as smoothing parameter which weakly depends on oxide and silicon thickness and quasi-Fermi potential which is given by [37]:

$$r = (At_{\text{ox}} + B) \left( \frac{C}{T_{\text{fin}}} + D \right) e^{-EV_F(x)} \quad (4.7)$$

The optimized value of  $A$ ,  $B$ ,  $C$ ,  $D$ , and  $E$  are given as  $0.0267 \text{ nm}^{-1}$ ,  $0.0270$ ,  $0.4526 \text{ nm}$ ,  $0.0650$ , and  $3.2823 \text{ V}^{-1}$  respectively. The optimized values obtained are for the device dimensions of  $t_{\text{ox}} < 10 \text{ nm}$  and  $T_{\text{fin}} > 5 \text{ nm}$  [37]. Extensive numerical simulations show that quasi-Fermi potential also depends on gate voltage, effective channel length, and fin thickness and is given by a semiempirical relationship as

$$V_F(x) = \frac{2kT}{q} \frac{m}{n} \ln \left[ \left( \exp \left( -\frac{V_{\text{ds}}(m/n)^{-1}}{kT/q} \right) - 1 \right) \left( \frac{x}{L_{\text{eff}}} \right)^{\frac{c}{V_{\text{gs}} - V_{\text{FB}}}} + 1 \right]^{-1} \times (a \times T_{\text{fin}})^{\frac{V_{\text{ds}}}{2c}} \quad (4.8)$$

where  $m/n = 2 + b(V_{\text{gs}} - V_{\text{FB}})$ ,  $a = 0.2 \text{ nm}^{-1}$ ,  $b = 7.5 \text{ V}^{-1}$ ,  $c = 1 \text{ V}$ , and  $V_{\text{ds}}$  is the applied drain voltage. The quasi-Fermi potential given in (4.8) is a function of position along the channel length and drain voltage  $V_{\text{ds}}$ . Substituting the value of  $\psi_0(x)$  from (4.4) and  $V_F(x)$  from (4.8) in (4.3), we obtain  $Z(x, y)$  as:

$$Z(x, y) = \left( U - \sqrt{U^2 - (V_{\text{gs}} - V_{\text{FB}})\psi_{0\max}(x)} \right) - \frac{2}{\beta} \ln \left\{ \cos \left[ \sqrt{\frac{q^2 n_i}{2kT \epsilon_{\text{Si}}}} e^{\frac{\beta}{2} \left( U - \sqrt{U^2 - (V_{\text{gs}} - V_{\text{FB}})\psi_{0\max}(x)} \right) \left( \frac{2kT m}{q n} \ln \left[ \left( \exp \left( -\frac{V_{\text{ds}}(m/n)^{-1}}{kT/q} \right) - 1 \right) \left( \frac{x}{L_{\text{eff}}} \right)^{\frac{c}{V_{\text{gs}} - V_{\text{FB}}}} + 1 \right]^{-1} \times (a \times T_{\text{fin}})^{\frac{V_{\text{ds}}}{2c}}} \right) \left( y - \frac{T_{\text{fin}}}{2} \right)} \right] \right\} \quad (4.9)$$

An expression of  $Z(x, y)$  is used to obtain the analytical solution of the function  $A(x, y)$ . The solution of one-dimensional Poisson equation is:

$$\psi_s(x) = C_1 \exp(m_1 x) + C_2 \exp(-m_1 x) - \frac{R}{m_1^2} \quad (4.10)$$

where  $C_1$ ,  $C_2$ ,  $m_1$ , and  $R$  are calculated by putting the following boundary conditions based on the physics of the device as:

$$\psi_s(x = 0) = \phi_s \quad \text{and} \quad \psi_s(x = L_{\text{eff}}) = \phi_s + V_{\text{ds}}$$

We obtained the values of the parameters as:

$$C_1 = \frac{\phi_S[1 - \exp(-m_1 L_{\text{eff}})] + V_{\text{ds}} + \frac{R[1-\exp(-m_1 L_{\text{eff}})]}{m_1^2}}{2\sinh(m_1 L_{\text{eff}})} \quad (4.11)$$

$$C_2 = -\frac{\phi_S[1 - \exp(+m_1 L_{\text{eff}})] + V_{\text{ds}} + \frac{R[1-\exp(+m_1 L_{\text{eff}})]}{m_1^2}}{2\sinh(m_1 L_{\text{eff}})} \quad (4.12)$$

$$R = \frac{\eta}{\epsilon_{\text{Si}} T_{\text{fin}}} [qN_a T_{\text{fin}} - 2C_{\text{ox}}(V_{\text{gs}} - V_{\text{FB}} - \phi_F)] \quad (4.13)$$

$$m_1 = \sqrt{\frac{2\eta C_{\text{ox}}}{\epsilon_{\text{Si}} T_{\text{fin}}}}$$

where  $\eta$  is a fitting parameter which incorporates the effects of the variation of the lateral field on the depleted film under the channel. As demonstrated by Harrison et al. [38],  $\eta$  is lower than 1 for  $V_{\text{gs}} \leq V_{\text{th}}$  and depends on the channel doping concentration and thickness. Therefore, this parameter has to be calibrated for each technology.  $C_{\text{ox}} = \epsilon_{\text{ox}}/t_{\text{ox}}$  is the oxide capacitance per unit area, and  $N_a$  is the channel doping. Substituting the value of  $C_1$ ,  $C_2$ , and  $R$  in (4.10), surface potential,  $\psi_s(x)$ , is obtained as:

$$\begin{aligned} \psi_s(x) = & \left( \frac{\phi_S[1 - \exp(-m_1 L_{\text{eff}})] + V_{\text{ds}} + \frac{R[1-\exp(-m_1 L_{\text{eff}})]}{m_1^2}}{2\sinh(m_1 L_{\text{eff}})} \right) [\exp(m_1 x)] \\ & + \left( \frac{\phi_S[1 - \exp(+m_1 L_{\text{eff}})] + V_{\text{ds}} + \frac{R[1-\exp(+m_1 L_{\text{eff}})]}{m_1^2}}{2\sinh(m_1 L_{\text{eff}})} \right) [\exp(-m_1 x)] \\ & - \frac{\left( \frac{\eta}{\epsilon_{\text{Si}} T_{\text{fin}}} [qN_a T_{\text{fin}} - 2C_{\text{ox}}(V_{\text{gs}} - V_{\text{FB}} - \phi_F)] \right)}{m_1^2} \end{aligned} \quad (4.14)$$

From (4.2) and (4.14), we obtained the full QM two-dimensional surface potential as:

$$\begin{aligned} \psi(x, y) = & \left[ \left( \frac{\phi_S[1 - \exp(-m_1 L_{\text{eff}})] + V_{\text{ds}} + \frac{R[1-\exp(-m_1 L_{\text{eff}})]}{m_1^2}}{2\sinh(m_1 L_{\text{eff}})} \right) [\exp(m_1 x)] \right. \\ & + \left( \frac{\phi_S[1 - \exp(+m_1 L_{\text{eff}})] + V_{\text{ds}} + \frac{R[1-\exp(+m_1 L_{\text{eff}})]}{m_1^2}}{2\sinh(m_1 L_{\text{eff}})} \right) [\exp(-m_1 x)] \\ & \left. - \frac{\left( \frac{\eta}{\epsilon_{\text{Si}} T_{\text{fin}}} [qN_a T_{\text{fin}} - 2C_{\text{ox}}(V_{\text{gs}} - V_{\text{FB}} - \phi_F)] \right)}{m_1^2} \right] \times \frac{Z(x, y)}{Z(x, y = 1)}. \end{aligned} \quad (4.15)$$

## 4.4 Threshold Voltage Modeling

The threshold voltage of a FinFET can be defined as that voltage (gate) which would be able to invert all the channels within the Fin structure simultaneously. We can derive the QM threshold voltage,  $V_{\text{th,QM}}$ , of the DG FinFET as [39]:

$$V_{\text{th,QM}} = V_{\text{FB}} + \psi_{\text{s(inv)}} - \frac{Q_b}{2C_{\text{ox}}} + \Delta V_{\text{th,QM}} \quad (4.16)$$

where  $\psi_{\text{s(inv)}}$  is the surface potential at threshold, and  $\Delta V_{\text{th,QM}}$  is the threshold voltage change due to QME's, which can be approximated as a function of the ratio of the carrier effective mass in the direction of confinement to the free electron mass and silicon film thickness which is given as [40]:

$$\Delta V_{\text{th,QM}} \cong \frac{S}{(kT/q)\ln(10)} \times \frac{0.3763}{(m_x/m_o)T_{\text{fin}}^2} \quad (4.17)$$

where  $S$  is the subthreshold slope,  $T_{\text{fin}}$  is fin thickness, and  $m_x/m_o$  is the ratio of the carrier effective mass in the direction of confinement to the free electron mass (e.g., 0.92 for electrons and 0.29 for holes).

The bulk charge  $Q_b$  is given as:

$$Q_b = -qN_a T_{\text{fin}} \quad (4.18)$$

When considering the quantum-mechanical confinement of inversion-layer carriers,  $V_{\text{th,QM}}$  of (4.16) should be augmented with  $\Delta V_{\text{th,QM}}$ . The surface potential at threshold is given by:

$$\psi_{\text{s(inv)}} = 2\psi_b \quad (4.19)$$

$$\psi_b = \frac{kT}{q} \ln\left(\frac{N_a}{N_i}\right) \quad (4.20)$$

Substituting the value of  $\psi_b$  from (4.20) into (4.19), we obtained:

$$\psi_{\text{s(inv)}} = 2\left(\frac{kT}{q} \ln\left(\frac{N_a}{N_i}\right)\right) \quad (4.21)$$

Substituting the value of  $Q_b$  from (4.18),  $\psi_{\text{s(inv)}}$  from (4.21) and  $\Delta V_{\text{th,QM}}$  from (4.17) into threshold expression (4.16), the final expression for the threshold voltage with QM corrections is obtained as:

$$\begin{aligned} V_{\text{th,QM}} &= V_{\text{FB}} + 2\left(\frac{kT}{q} \ln\left(\frac{N_a}{N_i}\right)\right) - \frac{(-qN_a T_{\text{fin}})}{2C_{\text{ox}}} + \frac{S}{(kT/q)\ln(10)} \\ &\quad \times \frac{0.3763}{(m_x/m_o)T_{\text{fin}}^2} \end{aligned} \quad (4.22)$$

## 4.5 Source/Drain Resistance Modeling

The quasi-nonplanar devices suffer from a high parasitic resistance due to narrow width of their source/drain (S/D) regions. The series resistance in the S/D regions of a FinFET has contributions from its components arising from different parts of the S/D geometry. The enhanced total resistance of FinFET reduces the driving capability of the device at all applied biases. We analyzed the parasitic S/D resistance and the total resistance of FinFET device using analytical model. The sheet resistance  $R_{sh}$  in the S/D extension is given by [8]:

$$R_{sh} = \rho_{ext} \left( \frac{W_{sp}}{H_{fin} \times W_{fin}} \right) \quad (4.23)$$

where  $\rho_{ext}$  is the resistivity of the S/D extension, and  $W_{sp}$  is the length of S/D extension.

Spread resistance ( $R_{sp1}$ ) is due to the spread of current from the thin accumulation layer into the S/D extension which can be written as [8]:

$$R_{sp1} = \frac{1}{2} \times \left[ \frac{2\rho_{ext}}{\pi H_{fin}} \ln \left( 0.75 \frac{(W_{fin})}{x_c} \right) \right] \quad (4.24)$$

where  $x_c$  is the channel thickness.  $R_{sp2}$  is the resistance due to the spread of current from S/D extensions into the wider Heavily Doped S/D (HDD) region and is given as [8].

$$R_{sp2} = \frac{\rho_{hdd} \times [\ln(0.75) + \ln(W_{sd}) - \ln(W_{fin})]}{\pi(H_{fin} + T_{SEG} - T_{SIL})} \quad (4.25)$$

where  $T_{SIL}$  is the thickness of the S/D silicide and  $T_{SEG}$  is S/D SEG thickness.

$R_{sd}$  has been modeled as series combination of two resistances,  $R_1$  and  $R_2$ , and is given as:

$$R_{sd} = 2 \times (R_1 + R_2) \quad (4.26)$$

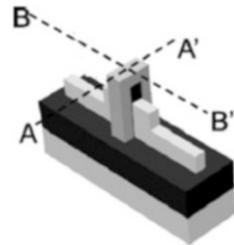
$R_1$  is the resistance between the gate and S/D spacer edge.

$$R_1 = R_{sp1} + R_{sh} \quad (4.27)$$

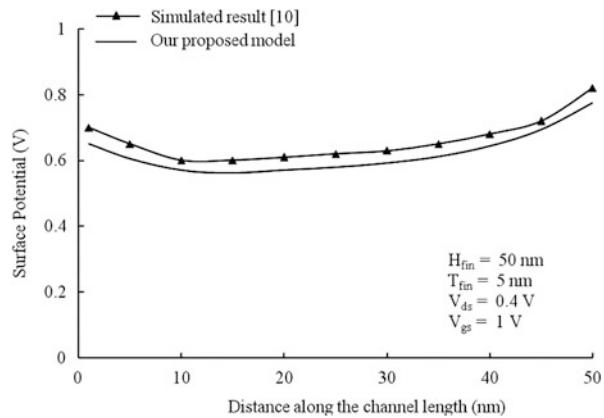
Substituting the value of  $R_{sh}$  and  $R_{sp1}$  from (4.23) and (4.24) in (4.27), we get the value of  $R_1$  as:

$$R_1 = \left( \frac{1}{2} \times \left[ \frac{2\rho_{ext}}{\pi H_{fin}} \ln \left( 0.75 \frac{(W_{fin})}{x_c} \right) \right] \right) + \left( \rho_{ext} \left( \frac{W_{sp}}{H_{fin} \times W_{fin}} \right) \right) \quad (4.28)$$

**Fig. 4.2** Geometry of the FinFET device [8]



**Fig. 4.3** Surface potential variations along the channel length for comparing our quantum result and through reported simulation result [37]



Also  $R_2$  is the parallel combination of resistance of plane A–A' and plane B–B' (Fig. 4.2) and is given by

$$R_2 = \frac{R_a \times R_b}{R_a + R_b} \quad (4.29)$$

$R_a = R_{\text{conA}}$  and  $R_b = R_{\text{sp2}} + R_{\text{conB}}$ ;  $R_{\text{conA}}$  is contact resistance of plane A–A' and  $R_{\text{conB}}$  is contact resistance of plane B–B'.

Contact resistance in plane A–A' (Fig. 4.2) is given:

$$R_{\text{conA}} = \frac{\rho_{\text{int}}}{W_{\text{fin}} \times T_{\text{SIL}}} \quad (4.30)$$

where  $\rho_{\text{int}}$  is the contact resistivity. Contact resistance in plane B–B' (Fig. 4.3) is given by:

$$R_{\text{conB}} = \frac{\rho_{\text{int}}}{(L_{\text{transfer}} \times W_{\text{sd}})} \coth\left(\frac{L_{\text{con}}}{L_{\text{transfer}}}\right) \quad (4.31)$$

where  $L_{\text{con}}$  is the physical length and  $W_{\text{sd}}$  is the width of HDD region. In case of the current conduction parallel to a semiconductor–metal interface, a minimum contact length exists before this conduction current is actually transferred from the semiconductor to the metal. This length is known as transfer length ( $L_{\text{transfer}}$ ).

Substituting the value of  $R_1$  and  $R_2$  from (4.28) and (4.29) in (4.26), we get the S/D resistance ( $R_{\text{sd}}$ ) as:

$$R_{\text{sd}} = 2 \times \left( \left( \left( \frac{1}{2} \times \left[ \frac{2\rho_{\text{ext}}}{\pi H_{\text{fin}}} \ln \left( 0.75 \frac{\left( \frac{W_{\text{fin}}}{2} \right)}{x_c} \right) \right] \right) + \left( \rho_{\text{ext}} \left( \frac{W_{\text{sp}}}{H_{\text{fin}} \times W_{\text{fin}}} \right) \right) \right) + \left( \frac{R_a \times R_b}{R_a + R_b} \right) \right) \quad (4.32)$$

The total resistance is obtained as:

$$R_{\text{total}} = \frac{V_{\text{ds}}}{I_s} = R_{\text{ch}} + R_{\text{sd}} \quad (4.33)$$

where  $R_{\text{ch}}$  is the resistance of channel region. From (4.32), substitute the value of  $R_{\text{sd}}$  in (4.33). The final expression for the total resistance is obtained as given below:

$$R_{\text{total}} = R_{\text{ch}} + \left( 2 \times \left( (R_{\text{sp1}} + R_{\text{sh}}) + \left( \frac{R_a \times R_b}{R_a + R_b} \right) \right) \right) \quad (4.34)$$

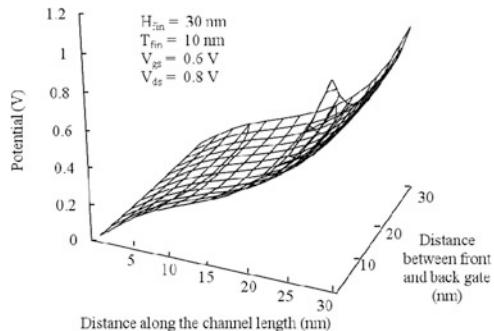
## 4.6 Results and Discussion

A full two-dimensional potential analytical modeling scheme taking into consideration various quantum mechanical effects has been presented for FinFET structure for a channel length of 30 nm, fin thickness of 10 nm, and fin height of 30 nm. For the purpose of validation of our analytical model, the results obtained have been compared and contrasted with reported simulated results as well as experimental results.

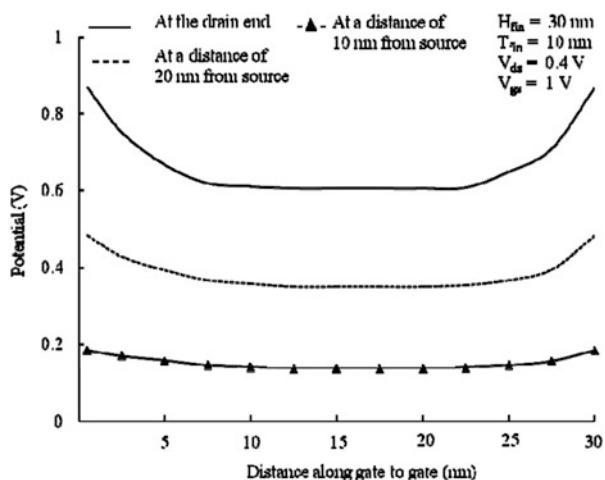
Figure 4.3 shows the variation of surface potential with distance along the channel length obtained on basis of our model and reported simulation results. The device parameters used for the analysis are shown within the figure. It can be seen from the figure that there is a good match between the reported result and result obtained through our modeling at any point along the channel length from source to drain. It can be observed that the potential initially falls to a minimum value at around the center of channel length and then monotonically increases at the drain end. At any point along the channel length, our model predicts a lower value of surface potential as compared to the simulated results. The small deviation seen in the results might be due to variation of carrier mass due to quantum confinement at an applied drain and gate voltage of 0.4 V and 1 V, respectively.

Figure 4.4 shows the variation of three-dimensional surface potential profile in the active region of the device. It can be seen from the figure that there is an increase in the potential along the channel length toward the drain end. It can be also be

**Fig. 4.4** 3D potential plot of FinFET for 30 nm channel length



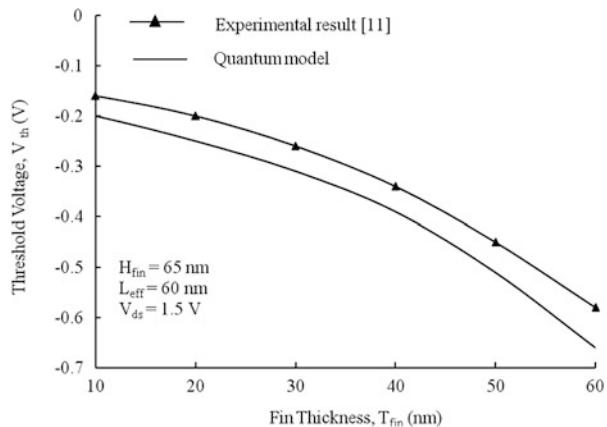
**Fig. 4.5** Potential variations from front gate to back gate at various positions along the channel length



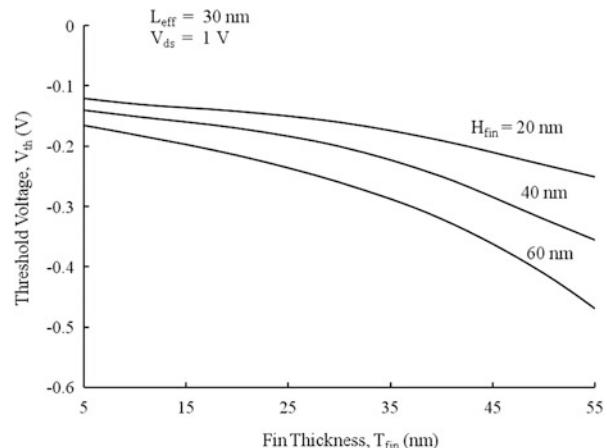
observed that the potential variation from the gate to gate at drain is more pronounced as compared to the variation at the source end. This is due to a large transverse as well as longitudinal direction electric field within the channel near the drain end as compared to source end.

The variation of channel potential from front gate to back gate at various distances from source side for fixed drain and gate bias is shown in Fig. 4.5. The gate length, fin height, and thickness have been taken as 30 nm, 30 nm, and 10 nm, respectively. It can be seen from the figure that as we move from source end of the gate to the drain end of the gate, there is substantial increase in the potential at any point in the channel. This is attributed to the increased value of longitudinal electric field at the drain end on application of a drain to source voltage. It can be further observed that near the source end, the potential is almost constant as one moves from front to back gate. But near the drain end, the variation of potential near either of the gates is very drastic. This is because of larger effective gate voltage at the drain end of the device as compared to the source end. This also implies that the

**Fig. 4.6** Variation of threshold voltage with Fin thickness for our proposed QM model, classical model, and experimental reported result [7]



**Fig. 4.7** Variation of threshold voltage with fin thickness for various fin height

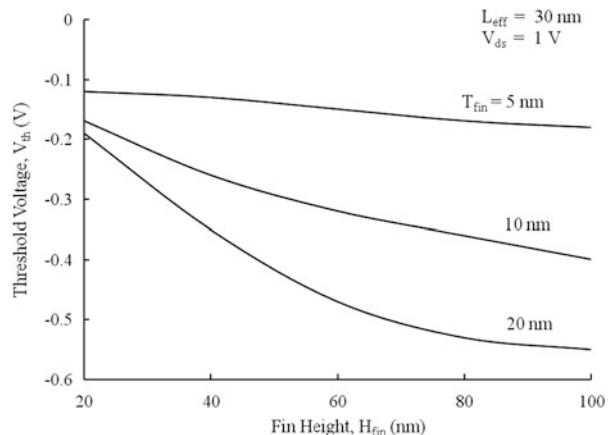


longitudinal electric field is enhanced near the Si–SiO<sub>2</sub> interface due to its proximity to metal gates.

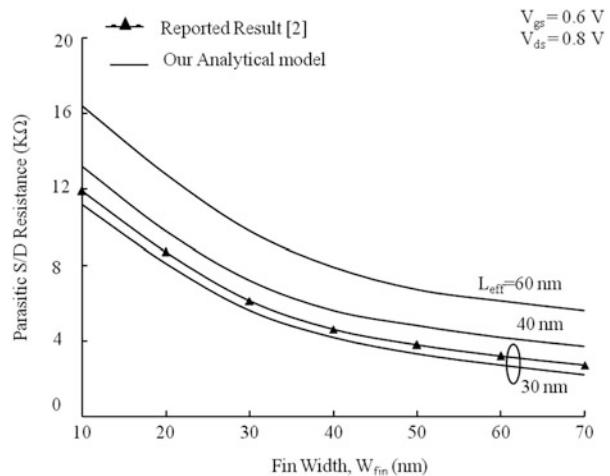
Variation of threshold voltage with fin thickness for our quantum mechanical model, classical model, and experimental results has been shown in Fig. 4.6 for the purpose of comparison. A threshold voltage roll-off with fin thickness is observed for both theoretical and experimental results. The fin thickness is varied from 5 to 55 nm. The percentage roll-off for our model is 77 % and that for experimental result it is 75 %. It can be inferred, therefore, that there is a close match of percentage variation between our results and experimental measurement, given the fact that the device process parameters undergo fluctuations at such low dimensions.

Figure 4.7 shows the variation of threshold voltage with fin thickness for varying fin height. It can be seen from the figure that as the fin height increases, the rate of reduction of threshold voltage with fin thickness also increases. Moreover, the

**Fig. 4.8** Variation of threshold voltage with fin height for various fin thickness



**Fig. 4.9** Variation of parasitic S/D resistance with varying fin width for proposed analytical model and reported result [8]

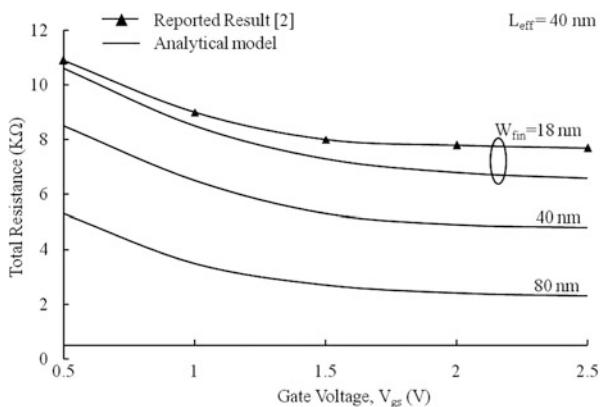


absolute value of threshold shows an enhancement with larger fin thickness. This is because at larger fin thickness, the top gate of the fin is able to control the channel charge to a lesser amount. Hence, there is an increase in threshold voltage. It can be further seen that as fin thickness increases, the transverse electric field reduces and hence a larger gate voltage is reduced in order to form channel, thereby increasing threshold voltage.

The Variation of threshold voltage with fin height for varying fin thickness is shown in Fig. 4.8. From this figure, it may be seen that as the fin thickness increases, the rate of reduction of threshold voltage with fin height also increases. This is because as the fin thickness increases, the effective area under the gate also increases, thus increasing threshold voltage.

Figure 4.9 shows the variation of parasitic S/D resistance with varying fin width for our proposed analytical model and reported numerical result [8] for the purpose

**Fig. 4.10** Variation of total resistance with variation of gate voltage for varying fin width



of validation for all fin width. A close match is found between the two results for channel length of 30 nm.

It can be seen from the figure that as the fin width increases, there is a decrease in the parasitic resistance for all values of channel length. As the fin width increases, the total area through which drain current flows also increases. This results in a decrease in the parasitic resistance. Further it is observed that for a fixed fin width, as the channel length increases, there is an enhancement in the parasitic S/D resistance. This can be inferred from the fact as the channel length decreases, quantum confinement along the S/D direction becomes more extensive. This results in an enhancement in the mobility of charge carriers which in turn increase the drain current and hence the parasitic S/D resistance decrease.

Figure 4.10 shows the variation of total resistance between S and D with variation of gate voltage. The results obtained by our analytical model have been compared with the reported numerical result [8] for  $W_{fin} = 18\text{ nm}$ . The variation is also shown for fin width of  $W_{fin} = 40\text{ nm}$  and  $80\text{ nm}$ . It is observed that as the fin width increases, there is almost linear decrease in the total resistance for a fixed applied gate voltage. Further for large gate voltage, the total resistance becomes almost independent of applied gate voltage.

## 4.7 Conclusion

In this chapter, a full 2D quantum mechanical analytical modeling has been presented in order to evaluate the 2D potential profile within the active area of FinFET structure. The key issues related to device parameters and structures are also shown in the chapter. The variation of potential from gate to gate is also reported in this chapter. For potential profile, there is close match between our results and reported experimental results. The results obtained would be useful to design device and for fabricating future nanoscale devices. Various potential

profiles such as surface, back to front gate, and source to drain potential have been presented in order to appreciate the usefulness of the device for circuit simulation purposes. Further, in this chapter, the detailed study of threshold voltage and its variation with the process parameters is presented for our proposed devices and a close match is obtained with the results through analytical model and reported experimental results.

## References

1. International Technology Roadmap for Semiconductors (ITRS), Semiconductor Industry Association, San Jose, CA (2007)
2. Chaudhry, A., Kumar, M.J.: Controlling short-channel effect in deep-submicron SOI MOSFETs for improved reliability: a review. *IEEE Trans. Dev. Mater. Reliab.* **4**(3), 99–109 (2004)
3. Ortiz-Conde, A., García-Sánchez, F.J., Malobabic, S.: Analytical solution of the channel potential in undoped symmetric dual-gate MOSFETs. *IEEE Trans. Electron Dev.* **52**(7), 1669–1672 (2005)
4. Frank, D.J.: Power-constrained CMOS scaling limits. *IBM J. Res. Dev.* **46**, 235–244 (2002)
5. Havaldar, D.S., Katti, G., DasGupta, N., DasGupta, A.: Subthreshold current model of FinFETs based on analytical solution of 3-D Poisson's equation. *IEEE Trans. Electron Dev.* **53**(4), 737–742 (2006)
6. Chen, Q., Harrell, E.M., Meindl, J.D.: A physical short-channel threshold voltage model for undoped symmetric double-gate MOSFETs. *IEEE Trans. Electron Dev.* **50**(7), 1631–2581 (2003)
7. Pei, G., Kedzierski, J., Oldiges, P., Ieong, M., Kan, E.C.-C.: FinFET design considerations based on 3-D simulation and analytical modeling. *IEEE Trans. Electron Dev.* **49**(8), 1411–1419 (2002)
8. Dixit, A., Kottantharayil, A., Collaert, N., Goodwin, M., Jurczak, M., De Meyer, K.: Analysis of the parasitic S/D resistance in multiple-gate FET. *IEEE Trans. Electron Dev.* **52**(6), 1132–1139 (2005)
9. Hisamoto, D., Lee, W.C., Kedzierski, J., Takeuchi, H., Asano, K., Kuo, C., Anderson, E., King, T.J., Bokor, J., Hu, C.: FinFET—a self-aligned double-gate MOSFET scalable to 20 nm. *IEEE Trans. Electron Dev.* **47**, 2320–2325 (2000)
10. Colinge, J.P.: Multiple-gate SOI MOSFET. *Solid State Electron.* **48**, 897–905 (2004)
11. Colinge, J.P.: Silicon-on-insulator technology: materials to VLSI. Kluwer Academic, Dordrecht (1991)
12. Tang, S.H., Chang, L., Lindert, N., Choi, Y.K., Lee, W.C., Huang, X., Subramanian, V., King, T.J., Bokor, J., Hu, C.: FinFET – a quasiplanar double gate MOSFET. In: Proceedings of international solid-state circuits conference (ISSCC), pp. 118–119 (2001)
13. Hodges, D.A.: Analysis and design of digital integrated circuits, 3rd edn. Tata McGraw-Hill, New Delhi (2003)
14. Moore, G.E.: Cramming more components onto integrated circuits. *Electronics* **38**(8), 114–117 (1965)
15. Chang, L.: Nanoscale thin-body CMOS devices. Ph.D. dissertation, University of California, Berkeley, Department of Electrical Engineering and Computer Sciences (2003)
16. Yeap, G.: Practical low power digital VLSI design. Kluwer Academic, Boston (1998)
17. Kang, S.M., Leblebici, Y.: CMOS digital integrated circuits, 3rd edn. Tata McGraw-Hill, New Delhi (2003)

18. Lázaro, A., Nae, B., Moldovan, O., Iñiguez, B.: A compact quantum model of nanoscale metal-oxide semiconductor field effect transistor for high frequency and noise simulations. *J. Appl. Phys.* **100**(8), 4320–4332 (2006)
19. Ortiz Conde, A., Sánchez, F.J.G., Schmidt, P.E., Sa Neto, A.: The non-equilibrium inversion layer charge of the thin-film SOI MOSFET. *IEEE Trans. Electron Dev.* **36**(9), 1651–1656 (1989)
20. Suzuki, K., Tanaka, T., Tosaka, Y., Horie, H., Arimoto, Y.: Scaling theory for double-gate SOI MOSFET's. *IEEE Trans. Electron Dev.* **40**, 2326–2329 (1993)
21. Oh, S.H., Monroe, D., Hergenrother, J.M.: Analytic description of short channel effects in fully-depleted Double-Gate and cylindrical, surrounding gate MOSFET. *IEEE Electron Device Lett.* **21**(9), 445–447 (2000)
22. Arora, N.: MOSFET models for VLSI circuits simulations theory and practice, 1st edn. Springer, New York (1993)
23. Katti, G., DasGupta, N., DasGupta, A.: Threshold voltage model for mesa-isolated small geometry fully depleted SOI MOSFETs based on analytical solution of 3-D Poisson's equation. *IEEE Trans. Electron Dev.* **51**(7), 1169–1177 (2004)
24. Kedzierski, J., Leong, M., Nowak, E., Kanarsky, T.S., Zhang, Y., Roy, R., Boyd, D., Fried, D., Wong, H.S.P.: Extension and source/drain design for high-performance FinFET devices. *IEEE Trans. Electron Dev.* **50**(4), 952–958 (2003)
25. Tsui, B.Y., Lin, C.P.: Process and characteristics of modified Schottky barrier (MSB) p-channel FinFET. *IEEE Trans. Electron Dev.* **52**(11), 2455–2462 (2005)
26. Hu, W., Chen, X., Zhou, X., Quan, Z., Lu, W.: Quantum-mechanical effects and gate leakage current of nanoscale n-type FinFETs: a 2D simulation study. *Microelectron. J.* **37**, 613–619 (2006)
27. Rabaeij, J.M., Chandrakasan, A., Nikolic, B.: Digital integrated circuits, 2nd edn. Prentice Hall, Upper Saddle River (2002)
28. Uyemura, J.P.: Introduction to VLSI circuit and systems. Wiley, New York (2002)
29. Roy, K., Prasad, S.: Low power CMOS VLSI circuit design. Wiley Interscience, New York (2000)
30. Itoh, K.: Review and future prospects of low-voltage memory circuits. *IBM J. Res. Dev.* **47**, 525–552 (2003)
31. Yan, R.H., Ourmazd, A., Lee, K.F.: Scaling the Si MOSFET: from bulk to SOI to bulk. *IEEE Trans. Electron Dev.* **39**, 1704–1710 (1992)
32. Verheyen, P., Collaert, N., Rooyackers, R., Loo, R., Shamiryan, D., De Keersgieter, A., Eneman, G., Leyns, F., Dixit, A., Goodwin, M., Yim, Y.S., Caymax, M., De Meyer, K., Absil, P., Jurczak, M., Biesemans, S.: 25% Drive current improvement for P-type multiple gate FET (MuGFET) devices by introduction of recessed  $\text{Si}_{0.8}\text{Ge}_{0.2}$  in the source and drain regions. In: Symposium on VLSI technology, pp. 194–195 (2005)
33. Granzner, R., Schwierz, F., Polyakov, V.M.: An analytical model for the threshold voltage shift caused by 2D quantum confinement in undoped multiple-gate MOSFETs. *IEEE Trans. Electron Dev.* **54**, 2562–2565 (2007)
34. Kolberg, S., Fjeldly, T.A.: 2D modeling of DG SOI MOSFETs and near the subthreshold regime. *J. Comput. Electron.* **52**(2), 217–222 (2006)
35. Taur, Y.: Analytic solutions of charge and capacitance in symmetric and asymmetric double-gate MOSFET. *IEEE Trans. Electron Dev.* **48**(12), 2861–2869 (2001)
36. Monfray, S., Skotnicki, T., Morand, Y., Descombes, S., Coronel, P., Mazoyer, P., Harrison, S., Ribot, P., Talbot, A., Dutartre, D., Haond, M., Palla, R., Le Friec, Y., Leverd, F., Nier, M.E., Vizioz, C., Louis, D.: 50 nm-gate all around (GAA)-silicon on nothing (SON) devices: a simple way to Co-integration of GAA transistors within bulk MOSFET process. In: Symposium on VLSI technology. Digest of technical papers, pp. 108–109 (2002)
37. Munteanu, D., Autran, J.L., Loussier, X., Harrison, S., Cerutti, R., Skotnicki, T.: Quantum short-channel compact modeling of drain-current in double-gate MOSFET. *Elsevier Solid State Electron.* **50**, 680–686 (2006)

38. Harrison, S., Munteanu, D., Autran, J.L., Cros, A., Cerutti, R., Skotnicki, T.: Electrical characterization and modeling of high-performance SON DG MOSFETs. In: Proceedings of European solid-state device research conference (ESSDERC), pp. 373–376 (2004)
39. Chiang, M.H., Lin, C.N., Lin, G.S.: Threshold voltage sensitivity to doping density in extremely scaled MOSFETs. *Semicond. Sci. Technol.* **21**, 190–193 (2006)
40. Trivedi, V.P., Fossum, J.G., Vandooren, A.: Non-classical CMOS device design. In: Proceedings of IEEE international SOI conference, pp. 155–157 (2003)

# Chapter 5

## Physical Insight and Correlation Analysis of Finshape Fluctuations and Work-Function Variability in FinFET Devices

Emanuele Baravelli

**Abstract** One of the major challenges that technology evolution has been facing in the last few years is the increasing severity of variability associated with the discrete nature of charge and the atomicity of matter, which become relevant in aggressively scaled devices. This is even more critical as device architecture has evolved from the conventional planar CMOS technology into three-dimensional multi-gate structures such as FinFETs. The 3D nature of FinFETs is reflected in an enhanced impact of geometry fluctuations in various dimensions: line-edge roughness (LER) in these devices affects both the top and sidewall gate profiles, as well as the fin thickness. Furthermore, different orientations of metal grains which appear in modern metal-gate electrodes result in undesired work-function variations (WFV). The impact of LER and WFV on FinFET electrical performance is studied in this chapter through extensive Monte Carlo (MC) ensemble simulations compared with simplified models for variability estimation. Relevant electrical parameters are correlated with representative descriptors of the various roughness or gate granularity configurations. The analysis provides insight on the physical phenomena that cause fluctuations as well as indication on critical device features to be optimized for improved variation tolerance. The presented investigation has general validity and its conclusions are expected to apply to both current and future generations of multi-gate devices.

### 5.1 Introduction

Scaling of the CMOS technology has seen a number of revolutionary milestones over the last few years. The 45nm technology generation has been characterized by a mainstream introduction of high- $\kappa$ /metal gate (HK/MG) stacks [1, 23] instead of

---

E. Baravelli (✉)  
University of Bologna, Viale Risorgimento, 2 40136 Bologna, Italy  
e-mail: [ebaravelli@arces.unibo.it](mailto:ebaravelli@arces.unibo.it)

the conventional SiO<sub>2</sub>/polysilicon-based structure. The second remarkable step has been the introduction of tri-gate architectures at the 22 nm node [8]. Continued employment of conventional CMOS technology had become increasingly difficult due to a number of limiting factors, including short-channel effects (SCEs), leakage and variability issues in devices approaching the atomic scale [29]. On the other hand, three-dimensional (3D) architectures offer a much tighter electrostatic control over the channel and hence reduced SCEs. When combined with a proper selection of high- $\kappa$  and metallic materials for the gate stack, these architectures help with resuming a healthy trend of channel length scaling, limiting gate leakage and setting the device threshold voltage without recurring to increased channel doping. Metal is needed in modern gate electrodes due to incompatibility of polysilicon with the HK stack. The adoption of metal gates has the fortunate implication of eliminating poly-depletion issues.

One of the major challenges that technology evolution has been facing in the last few years is the increasing severity of variability associated with the discrete nature of charge and the atomicity of matter, which become relevant in nanoscale devices [4]. Major sources of fluctuations in recent generations of planar CMOS devices are related to the intrinsic nature of the employed materials and interfaces (e.g., polysilicon granularity, PSG [4, 5, 12], as well as fixed and trapped charges at the channel to gate dielectric interface, FTC [3, 4]), or are introduced at various process steps (line-edge roughness, LER [7, 26], oxide thickness variation, OTV [6, 24], random dopant fluctuations, RD [2, 4, 24]). Some of these fluctuations have been suppressed (e.g., PSG), or significantly reduced with the adoption of HK/MG stacks [1]. The use of undoped channels in multiple-gate structures such as FinFETs has also significantly reduced the impact of RD. However, other variability sources have emerged with these new technologies. The 3D nature of FinFETs is reflected in an enhanced impact of geometry fluctuations in various dimensions: LER in these devices affects both the top- and sidewall-gate profiles, as well as the fin thickness [10, 31, 32, 34]. Moreover, different orientations of metal grains which appear in MG electrodes result in undesired work-function variations (WFV) [13].

The aim of this chapter is to study how LER and WFV affect the electrical performance of FinFET devices. Extensive Monte Carlo (MC) ensemble simulations will be compared with simplified models for variability impact estimation in order to assess their robustness and propose suitable alternatives to optimize the trade-off between computational effort and statistical confidence. Performance sensitivity to local variations in different device regions will be highlighted, thus providing insight on the physical phenomena involved as well as indication on critical device features to be optimized for improved variation tolerance. This will be done by correlating electrical parameters with representative descriptors of the various roughness or gate granularity configurations. Although conducted with reference to a specific FinFET structure, this analysis has general validity and its conclusions are expected to apply to future generations of multi-gate devices.

## 5.2 Modeling Approach

The reference device considered in this chapter for analysis of LER- and WFV-induced fluctuations is a transistor architecture that was originally designed for low-voltage/low-power applications at the 32 nm technology node, where mainstream employment of these devices was first expected [17]. Although technology has further evolved and multi-gate devices have been introduced later at the 22 nm generation, considerations stemming from the presented analysis still have general validity in assessing the relative importance of the investigated sources of fluctuations and the different physical mechanisms through which they affect device performance. Hence, conclusions of this study can be applied to current and future technology generations as well.

The LSTP-32 nm compatible FinFETs considered throughout this chapter have nominal geometry and doping specifications described in Table 5.1. Symbols in this table have the following meaning. Geometrical parameters are the gate length ( $L_{\text{gate}}$ ), fin width ( $W_{\text{fin}}$ ), fin height ( $H_{\text{fin}}$ ), and equivalent oxide thickness ( $EOT$ ). Doping concentrations are specified for the channel ( $N_{\text{ch}}$ ), source (S), and drain (D) pads ( $N_{\text{hdd}}$ ), as well as for the extension regions of the n-type ( $N_{\text{ext,n}}$ ) and p-type ( $N_{\text{ext,p}}$ ) device. Relevant electrical parameters are the drain current in the on ( $I_{\text{ON}}$ ) and off state ( $I_{\text{OFF}}$ ), the saturation threshold voltage ( $V_{\text{T,sat}}$ ), and the subthreshold slope ( $SSlope$ ). Values reported in the table have been extracted considering an n-channel device and a supply voltage  $V_{\text{DD}} = 1$  V.

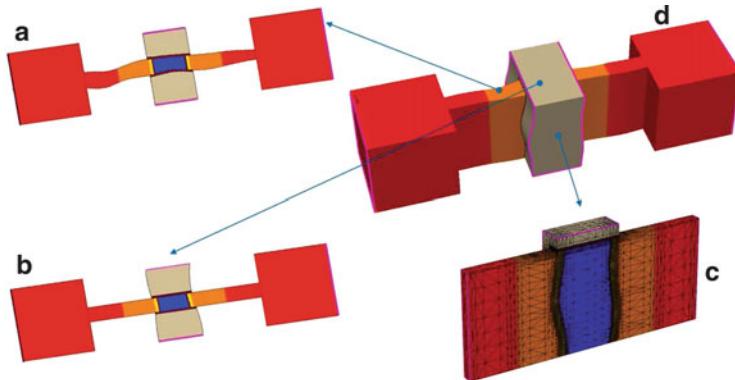
Technology Computer-Aided Design (TCAD) simulation [27] of the reference device and of devices affected by variability has been performed with a hydrodynamic transport model to provide a reasonably accurate description of the saturation regime. Quantum confinement associated with the small fin widths considered has been accounted for through a density gradient approximation (DGA). Mobility degradation and velocity saturation effects in the presence of high electric fields have also been included.

### 5.2.1 LER Modeling

A 3D representation of a considered FinFET device with superimposed LER is shown in Fig. 5.1d. Roughness contributions affecting the fin, top-, and sidewall-gate edges are, respectively, illustrated in Fig. 5.1a–c. Rough features have been

**Table 5.1** Reference device specifications

Geometry	Doping	Electrical parameters
$L_{\text{gate}} = 30$ nm	$N_{\text{ch}} = 1 \times 10^{15} \text{ cm}^{-3}$	$I_{\text{ON}} = 750 \mu\text{A}/\mu\text{m}$
$W_{\text{fin}} = 10$ nm	$N_{\text{hdd}} = 1 \times 10^{20} \text{ cm}^{-3}$	$I_{\text{OFF}} = 10 \text{ pA}/\mu\text{m}$
$H_{\text{fin}} = 50$ nm	$N_{\text{ext,n}} = 5 \times 10^{18} \text{ cm}^{-3}$	$V_{\text{T,sat}} = 0.36$ V
$EOT = 1.2$ nm	$N_{\text{ext,p}} = 2 \times 10^{19} \text{ cm}^{-3}$	$SSlope = 69 \text{ mV/dec.}$

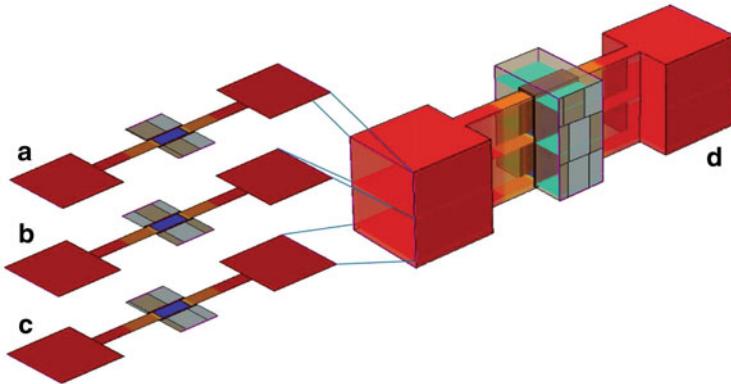


**Fig. 5.1** TCAD models of FinFETs with LER on the fin edges (a), top- (b), and sidewall-gates (c). (d) Overall representation of the 3D structure

generated using a Fourier synthesis technique [7, 10], whereby random sequences with desired statistical properties are produced and superimposed to straight edges to create LER patterns. Specifically, it has been shown that the power spectrum obtained from LER measurements can be modeled through a Gaussian autocorrelation process whose characteristic parameters are the rms amplitude  $\Delta$  and correlation length  $\Lambda$  of the roughness. Values considered in this work for these parameters are  $\Delta \in 1\text{--}2\text{ nm}$  and  $\Lambda \in 10\text{--}30\text{ nm}$ , which fall within typically measured ranges [7, 14]. The Fourier technique has been used to generate ensembles of FinFET instances with unique LER patterns, whose simulation provides distributions of electrical parameters to be analyzed statistically for variability assessment. The typical ensemble size considered in the following sections is 200 devices, which was found to provide sufficient accuracy for the purpose of this study [11].

### 5.2.2 WVF Modeling

Work-function variability should be modeled by splitting the gate area of a device according to realistic patterns of individual grains which form the metal gate electrode. In FinFETs with narrow fin, the top gate is basically uninfluential and the device width is set by the fin height  $H_{\text{fin}}$ . In our approach, the area of each sidewall gate is randomly split into rectangles representing individual grains, as depicted in Fig. 5.2d. Despite the rectangular shape approximation, this approach is more accurate than the way WVF is treated in most of the literature [19, 35], because the dimensions of each rectangle are not fixed in our simulations, but normally distributed around a given average size. A 2D slice methodology is adopted for device simulation to contain the computational cost. The approach is to model the 3D FinFET in Fig. 5.2d as a parallel composition of two-dimensional



**Fig. 5.2** WFV simulation approach: the 3D FinFET with random rectangular grains on the sidewall-gate electrodes (**d**) is modeled as a parallel composition of 2D slices (**a**)–(**c**)

**Table 5.2** TiN metal properties

Crystal orientation	WF value	Occurrence probability
$\langle 200 \rangle$	$\Phi_1 = 4.62 \text{ eV}$	$P_1 = 0.6$
$\langle 111 \rangle$	$\Phi_2 = 4.42 \text{ eV}$	$P_2 = 0.4$

structures like those shown in Fig. 5.2a–c. Large Monte Carlo ensembles of 2D slices with random grain widths and orientations can thus be simulated, and drain current vs. gate voltage ( $I_D$ - $V_{GS}$ ) curves of the composite 3D structures are then calculated as a weighted sum of randomly selected 2D slice characteristics. Weights represent the fraction of the fin height covered by each slice. The number of slices and the number of grains in each slice both depend on the selected mean value and spread of the grain size. The slice approximation is similar to the approach used in [26] for gate-LER investigation in planar MOSFETs, which was validated against 3D simulations, thus boosting confidence on its validity even in the WFV framework considered here.

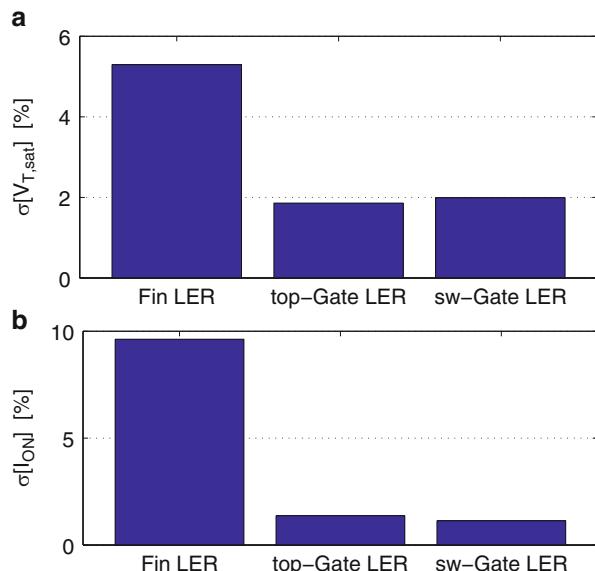
Work-function variations arise due to the fact that metal grains are characterized by different WF values depending on their crystal orientation, and the number of possible orientations is a property of the particular metal used. The associated occurrence probabilities as well as the average grain size are strongly influenced by the fabrication process, including deposition technique, temperature and duration, film thickness, incorporation of other materials, and underlying HK stack [13, 33, 35]. As a result, the average grain size, orientation probabilities, and respective WFs can vary significantly [25]. In the present study, TiN has been selected for the metal gate, as it is one of the most commonly used materials. The granular structure of this metal is characterized by two possible crystal orientations, whose respective occurrence probabilities  $P_1$  and  $P_2$  are summarized in Table 5.2, based on typically reported values [13]. The associated work functions, indicated as  $\Phi_1$  and  $\Phi_2$  in the table, have been slightly tuned to account for the dielectric stack and achieve the  $V_{T,\text{sat}}$  value reported in Table 5.1 for a device

with uniform orientation of gate grains along the  $\langle 200 \rangle$  direction. Statistics related to TiN metal granularity critically depends on process conditions. Average grain sizes in the 4 – 22 nm range have been reported [13, 16, 25], while the magnitude of deviations around these values is not well assessed yet. Simulations presented in this chapter will consider average grain sizes  $g_s$  varying between 10 and 20 nm, with a 10% standard deviation around the mean value.

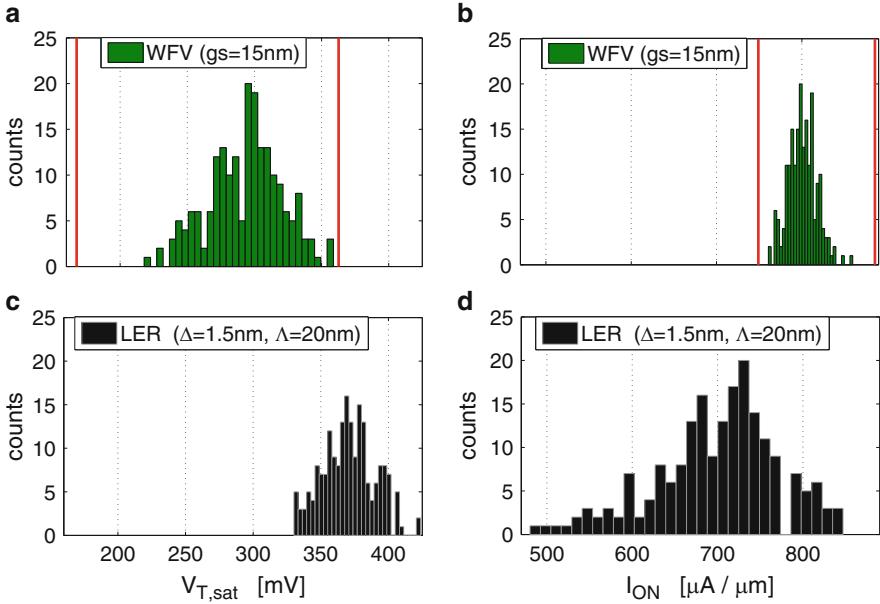
### 5.3 Statistical Analysis of LER- and WVF-Induced Fluctuations

Relative importance of the three main components of line-edge roughness in FinFETs, i.e. fin-LER, top-gate LER, and sw-gate LER, has been compared by simulating ensembles of devices similar to those in Fig. 5.1a, b, and c, respectively. The considered ensemble size has been reduced from 200 to 100 in the latter case to contain the computational cost of 3D simulations needed to fully account for the sidewall-gate roughness, without severely compromising accuracy [11]. LER parameters in this analysis are  $\Delta = 1.5$  nm,  $\Lambda = 20$  nm.

Standard deviations of saturation threshold voltage and on-current have been extracted from the obtained transcharacteristics and are reported in Fig. 5.3 as a percentage of nominal parameter values in Table 5.1. Comparison of different LER contributions in Fig. 5.3 clearly indicates that fin-LER is the most critical issue for both  $V_T$  and  $I_{ON}$  variability, while top- and sidewall-gate-LER produce similar fluctuations in electrical parameters. Fin-LER will therefore be studied more in



**Fig. 5.3** Comparison of LER contributions to (a) threshold voltage and (b) on-current variability. Corresponding standard deviations are expressed as a percentage of the nominal  $V_{T,sat}$  and  $I_{ON}$  values reported in Table 5.1. LER parameters are  $\Delta = 1.5$  nm,  $\Lambda = 20$  nm



**Fig. 5.4** Histograms of (a), (b) WVF and (c), (d) LER contributions to (a), (c) threshold voltage and (b), (d) on-current fluctuations. The different nature of the distributions can be observed. Theoretical bounds of WVF-related data are indicated by *solid vertical lines* in (a) and (b)

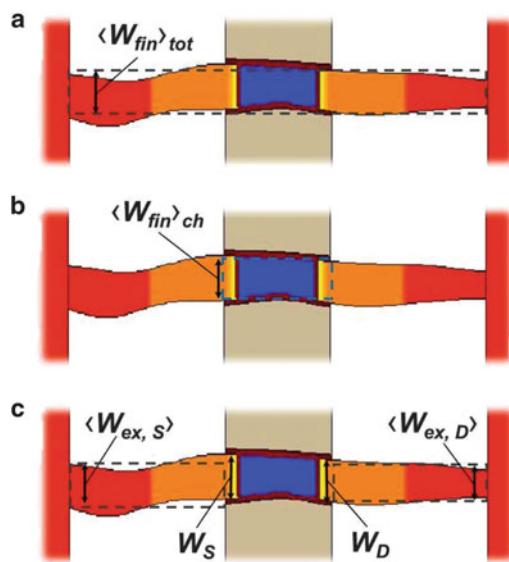
detail in the remainder of this chapter due to its dominant importance, while gate-LER issues will not be given further consideration. It should be added that the results in Fig. 5.3 have been obtained assuming random uncorrelated roughness of the two fin or gate edges, as in Fig. 5.1. Fin-LER could be mitigated by using a spacer-based process for fin definition, which results in correlated edge shapes [10, 14]. However, the worst-case situation of uncorrelated fin-LER is most commonly encountered in practice and will be considered throughout this chapter.

Work-function variability has been analyzed with the slice composition approach described in Sect. 5.2.1. A large ensemble of slice instances with random gate composition has been simulated, and individual slices have been grouped to form 200 pseudo-3D FinFETs. Resulting  $V_{T,\text{sat}}$  and  $I_{\text{ON}}$  histograms are shown in Fig. 5.4a,b, respectively, for an average grain size  $gs = 15$  nm. Distributions associated with the fin-LER simulations discussed above are reported in Fig. 5.4c,d for comparison. First, it can be observed that the mean value of  $V_{T,\text{sat}}$  is considerably reduced for the WVF distribution in Fig. 5.4a compared to the LER data of Fig. 5.4c. This is due to the presence of metal grains with a lower work-function  $\Phi_2$  in the WVF ensemble, while a uniform gate WF of value  $\Phi_1$  is considered for LER evaluation. Moreover, WVF has a stronger impact than LER on threshold voltage spread. An opposite situation is observed when comparing on-current data in Fig. 5.4b,d. WVF-induced fluctuations for this parameter are much less important than those produced by LER, although the mean value of  $I_{\text{ON}}$  is

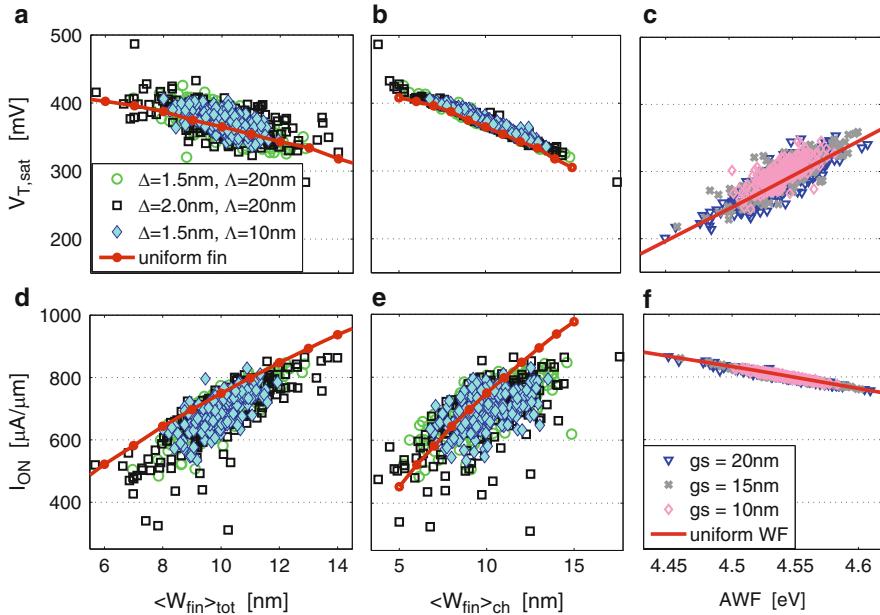
raised by the multiple grain orientations. Moreover, it is worth noting that the nature of LER- and WFV-related statistics is substantially different. In the latter case, fluctuations of electrical parameters cannot exhibit infinite tails as they are bounded by the two extreme situations of equally oriented grains with either WF value covering the whole gate area. Corresponding limits are indicated by solid vertical lines in Fig. 5.4a,b, while no theoretical bounds exist for LER-induced distributions in Fig. 5.4c, d. The presence of clearly identifiable corner cases is expected to become relevant for circuit applications like SRAM, where distribution tails need to be accurately monitored in order to achieve sufficient yield [18]. Finally, it can be observed that the peaks of WFV histograms in Fig. 5.4a,b are not centered with respect to the indicated boundaries due to the different orientation probabilities in Table 5.2.

## 5.4 Correlation-Based Approaches for Variability Estimation

Additional LER and WFV ensembles have been simulated using different values of roughness parameters  $\Delta$  and  $\Lambda$ , and of the average grain size  $g_s$ . Correlation of the resulting electrical parameters with characteristic geometrical features of the associated device realizations is investigated in this section. For LER datasets, the considered characteristic features are the average value of the fin width calculated either over the whole fin length ( $\langle W_{fin} \rangle_{tot}$ ) or in the channel region only ( $\langle W_{fin} \rangle_{ch}$ ). Definition of these parameters is illustrated in Fig. 5.5a,b, respectively.



**Fig. 5.5** Definition of (a) overall average fin width  $\langle W_{fin} \rangle_{tot}$ , (b) average fin width in the channel region  $\langle W_{fin} \rangle_{ch}$ , and (c) characteristic parameters used to compute an equivalent fin width  $W_{equiv}$  for improved  $I_{ON}$  correlation. These parameters include the source ( $W_S$ ) and drain end ( $W_D$ ) of the gate and the average extension widths  $\langle W_{ex,s} \rangle$ ,  $\langle W_{ex,d} \rangle$



**Fig. 5.6** Scatterplots of (a)–(c) threshold voltage, and (d)–(f) on-current associated with (a), (b), (d), (e) LER datasets with different  $\Delta$  and  $\Lambda$  values, and (c),(f) WFV datasets with different grain sizes. LER data are plotted against the  $\langle W_{\text{fin}} \rangle_{\text{tot}}$  or  $\langle W_{\text{fin}} \rangle_{\text{ch}}$  parameters defined in Fig. 5.5a,b, while WFV data are ordered according to the area-weighted average work-function AWF of each device instance

The representative parameter for WFV datasets is instead the average work-function AWF, calculated as an area-weighted sum of work-functions associated with individual grains present in each device realization.

#### 5.4.1 Correlations and Sensitivity Analysis

Scatterplots of  $V_{\text{T},\text{sat}}$  and  $I_{\text{ON}}$  associated with the various datasets are shown in Fig. 5.6. It can be seen from Fig. 5.6b that LER-induced threshold voltage fluctuations are strongly correlated to  $\langle W_{\text{fin}} \rangle_{\text{ch}}$ , i.e. LER mainly influences the device  $V_{\text{T}}$  by changing the average fin width in the channel, while roughness of the source and drain extension regions has little relevance to this parameter [11] (see Fig. 5.6a). Solid lines in Fig. 5.6a,b,d,e have been obtained by simulating straight-fin devices with different fin widths, in order to compare the Monte Carlo approach with a simplified sensitivity analysis (SA) of LER-induced fluctuations. Correlations highlighted in Fig. 5.6b suggest that a simple sensitivity-based estimation of  $\sigma[V_{\text{T}}]$  might be fairly accurate for relatively small values of the roughness, as indicated by the parallel trends of solid line and scatter data in the central portion of this figure.

However, it can be seen that corner situations are not properly captured through the simple linear trend provided by the sensitivity analysis. Moreover, this simplified approach tends to underestimate the mean value of  $V_T$ , as indicated by the slight negative shift between SA line and distributions centroid. A physical interpretation of this phenomenon is provided in Sect. 5.4.3.

LER-induced drive current fluctuations are much less correlated to the average fin width in the channel as compared to  $V_T$  fluctuations, as can be seen by comparing Fig. 5.6e,b. Correlation is increased by plotting  $I_{ON}$  data against  $\langle W_{fin} \rangle_{tot}$  in Fig. 5.6d, which indicates that FinFET on-current is also influenced by roughness of the extensions. Moreover, fin-LER decreases the average drive current compared with SA-based predictions represented by the solid lines.

As for the WFV datasets, the AWF parameter provides a good representation of on-current fluctuations in Fig. 5.6f, which are anyway much less significant than those associated with LER, even when the average grain size is varied. Instead, threshold voltage points are much more scattered around the “uniform WF” line in Fig. 5.6c, which describes the  $V_T$  of devices with uniform gate work-function of value equal to the AWF coordinate.

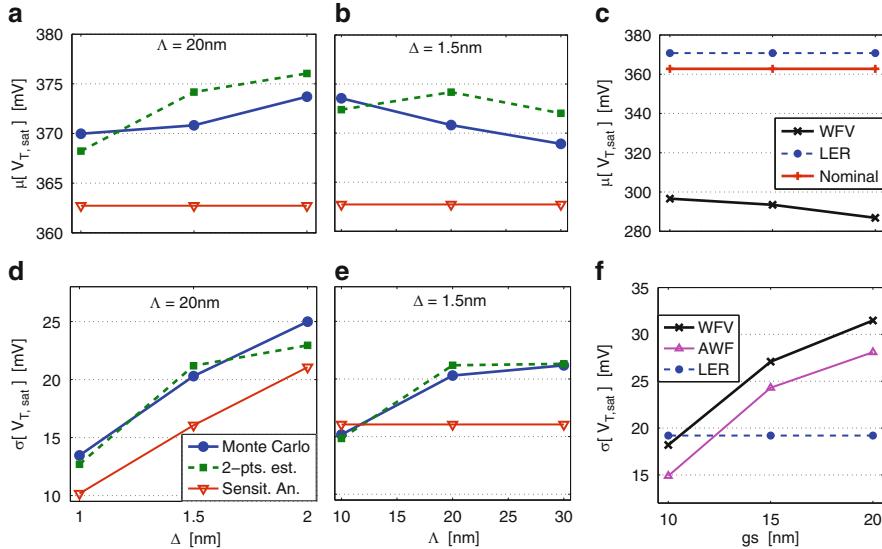
### 5.4.2 Simplified Approaches for Variability Estimation

Correlations highlighted in Fig. 5.6 can be exploited to achieve approximate estimations of LER- and WFV-induced variability at a reduced computational cost compared to ensemble Monte Carlo simulations.

#### 5.4.2.1 Threshold Voltage Variability

The strong correlation of  $V_T$  to  $\langle W_{fin} \rangle_{ch}$  in Fig. 5.6b suggests that the full distribution trend can be approximately represented through a linear interpolation between two properly selected “corner” points. Basically, the idea is to generate the full ensemble of devices affected by fin-LER, but choose a very small subset of these devices for simulation and “corner” data extraction, based on the  $\langle W_{fin} \rangle_{ch}$  parameter. The selection must be done carefully to avoid outliers that could affect the resulting statistical estimates. The whole procedure is described in detail in [11], where a selection approach is proposed, which involves four simulations only.

$V_T$  variability estimation based on the “corner” approach is compared with statistical analysis of full MC ensembles and with a conventional sensitivity analysis in Fig. 5.7. The latter procedure is unable to model variations of the mean  $V_T$ , as shown in Fig. 5.7a, b, nor it can predict the impact of different correlation lengths of the roughness, see Fig. 5.7e: the only effect sensitivity analysis can track is the impact of various rms amplitudes, as can be seen from Fig. 5.7d. At the same computational expense, the “corner” method fully accounts for fin-LER impact on device threshold with a fairly good accuracy (maximum



**Fig. 5.7** Impact of LER with different rms amplitudes (a), (d) and correlation lengths (b), (e) on (a), (b) the mean value, and (d), (e) standard deviation of saturation threshold voltage. Full Monte Carlo simulations are compared with simplified estimations of variability based on a conventional sensitivity analysis, and on a “2-point estimation” exploiting identification of corner cases. (c), (f) Impact of WVF with different grain sizes  $gs$  on the mean value and standard deviation of  $V_T$ . LER contribution for  $\Delta = 1.5\text{ nm}$ ,  $\Lambda = 20\text{ nm}$  is also reported for comparison, along with a simplified estimation based on the average work-function approach in (f)

error = 8% with respect to MC-based statistics obtained with a two orders of magnitude higher computational cost). The “ $V_T$ -raising” (i.e., higher mean value than the nominal  $V_T$ ) is seen to increase with increasing  $\Delta$  and with decreasing  $\Lambda$  in Fig. 5.7a,b, respectively. Of course higher values of  $\Delta$  also enhance the  $V_T$  spread, which is systematically underestimated by sensitivity analysis, as shown in Fig. 5.7d. Moreover, Fig. 5.7e indicates that  $\sigma[V_T]$  increases with increasing  $\Lambda$  as well, because of a lower self-averaging of fin thickness variations within the channel.

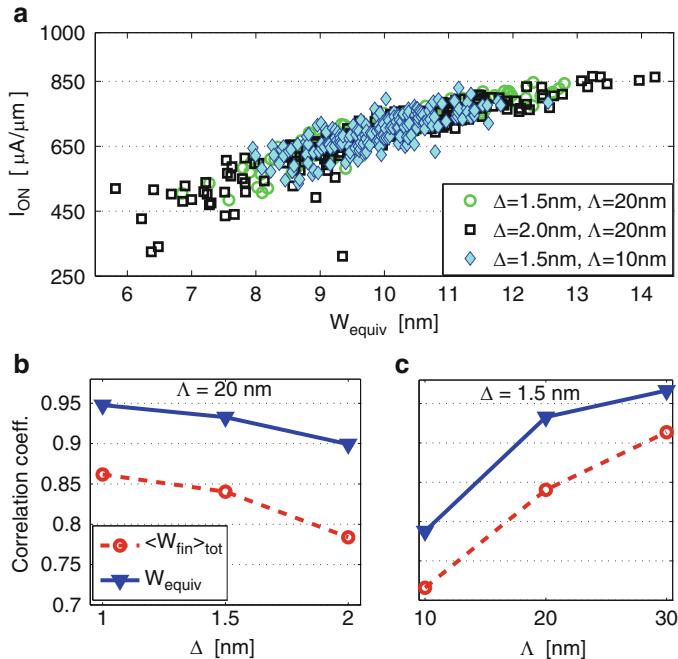
$V_{T,\text{sat}}$  statistics associated with work-function variations is quantitatively presented in Fig. 5.7c,f, where LER data related to the ensemble with  $\Delta = 1.5\text{ nm}$ ,  $\Lambda = 20\text{ nm}$  are also reported for easier comparison, since y scales in these plots are different from those in the rest of Fig. 5.7. It can be seen from Fig. 5.7c that the lowering of threshold voltage compared with the case of a uniform WF of value  $\Phi_1$  is not independent of the grain size, but becomes more pronounced as  $gs$  is increased. Although this  $V_T$ -lowering is partially contrasted by the  $V_T$ -raising effect produced by fin-LER, this is not sufficient to counterbalance the impact of metal granularity. While confirming that WVF has a stronger impact than LER on  $V_T$  spread, as already observed in Sect. 5.3, Fig. 5.7f indicates that this issue can be mitigated by enhanced self-averaging with grain size reduction through process

engineering. The most common technique to rapidly estimate the impact of WVF on device electrical performance is through an area-weighted average work-function (AWF) methodology [13]. In this simplified approach, the impact of metal granularity is lumped into an effective gate work-function  $\Phi_M$ , which is calculated as a weighted sum of individual work-functions associated with the possible crystal orientations. Weights are random numbers representing the fraction of gate area occupied by each grain orientation, thus resulting in a probabilistic distribution of  $\Phi_M$ . Figure 5.7f shows that the AWF model provides similar trends for  $\sigma[V_T]$  to those obtained with full MC simulations; however, threshold voltage spread is systematically underestimated by 10% or more with this simplified approach. Thus, the AWF-based technique is a valuable tool for coarse predictions of the impact of WVF, but extensive ensemble simulations are required when higher accuracy is required.

#### 5.4.2.2 Drive Current Variability

It was shown in Sect. 5.4.1 that FinFET drive current is scarcely correlated to the average fin width in the channel  $\langle W_{\text{fin}} \rangle_{\text{ch}}$ , while it exhibits a moderate correlation to the overall average fin width  $\langle W_{\text{fin}} \rangle_{\text{tot}}$ . Scatter can be further reduced by reordering  $I_{\text{ON}}$  data according to an equivalent fin width  $W_{\text{equiv}}$ . This parameter is defined as a weighted sum of widths  $W_j$  in a few key regions of the fin, chosen among the maximum, average and minimum fin thickness within the channel, source and drain extensions, as well as the widths at the two gate ends. The best correlation is obtained when selecting the two end points of the gate at the source ( $W_s$ ) and drain side ( $W_d$ ) and the average width of the extensions  $\langle W_{\text{ex,s}} \rangle, \langle W_{\text{ex,d}} \rangle$ , as depicted in Fig. 5.5c. Weights used to combine these characteristic parameters into  $W_{\text{equiv}}$  are estimated through a least mean square problem, as illustrated in [9]. Their values indicate the relative importance of LER in different fin regions: it is found that drive current is mainly influenced by the source side of the device. Similar weights are obtained when considering datasets with different LER parameters, thus confirming the physical foundation of this approach. Scatter plots of  $I_{\text{ON}}$  vs.  $W_{\text{equiv}}$  are shown in Fig. 5.8a: improved correlation can be clearly seen when comparing this figure to Fig. 5.6d. More specifically, correlation coefficients of  $I_{\text{ON}}$  to  $W_{\text{equiv}}$  and  $\langle W_{\text{fin}} \rangle_{\text{tot}}$  are reported in Fig. 5.8b,c as a function of  $\Delta$  and  $\Lambda$ , respectively. Correlation is improved by 10% or more with the  $W_{\text{equiv}}$ -based reordering.

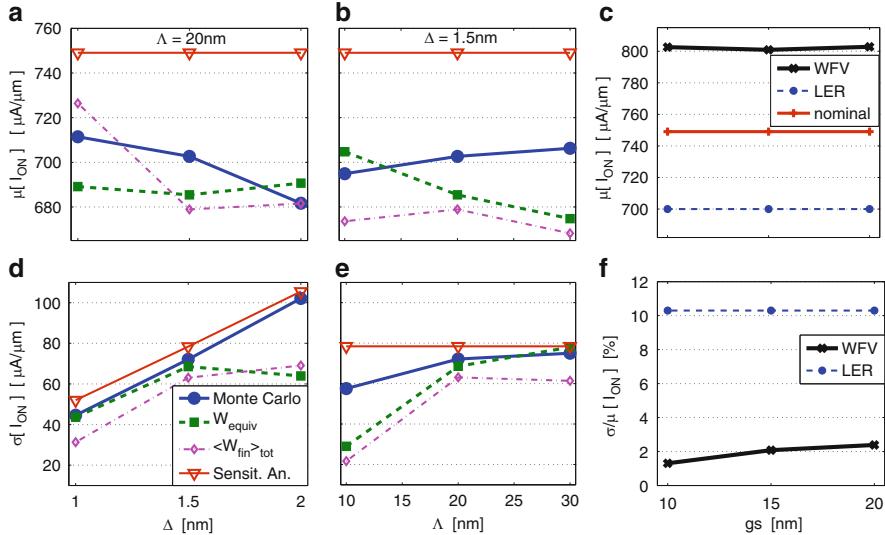
This increased correlation may be exploited to improve variability estimation of drive current through a reduced ensemble, based on a “corner” approach similar to that described above for  $V_T$ , although the accuracy is expected to be lower here, due to the still larger spread of  $I_{\text{ON}}$  data. Although  $W_{\text{equiv}}$  estimation would, in principle, require simulation of all the available device instances to determine the proper weights for a specific LER dataset, it has been verified that  $I_{\text{ON}}$  correlation is largely insensitive to small changes in the weights. Average values of these parameters have thus been calculated [9], which can be assumed as generally valid for the



**Fig. 5.8** (a)  $I_{ON}$  data for various LER datasets plotted as a function of the equivalent fin width  $W_{equiv}$ . Correlation of  $I_{ON}$  to  $W_{equiv}$  and to  $\langle W_{fin} \rangle_{tot}$  is compared in (b) and (c) as a function of  $\Delta$  and  $\Lambda$ , respectively, and is found to be at least 10% higher when  $W_{equiv}$  is considered

FinFET structure under investigation. Application of the “corner” method to both the  $I_{ON}$  vs.  $\langle W_{fin} \rangle_{tot}$  and the  $I_{ON}$  vs.  $W_{equiv}$  distributions produces the results in Fig. 5.9. It can be seen that the  $W_{equiv}$ -based reordering improves estimation of  $I_{ON}$  statistics in most of the considered cases, although “corner”-based predictions significantly deviate from MC results at high rms amplitudes and small correlation lengths of the roughness (see Fig. 5.9d,e), due to a higher data spread and to the presence of more outliers. Like in the  $V_T$  case, sensitivity analysis can only capture the  $\Delta$ -dependence of  $\sigma[I_{ON}]$ ; in particular, it does not account for the significant LER-induced lowering of the mean drive current observed in Fig. 5.9a,b.

On the other hand, WVF has an opposite effect on  $\mu[I_{ON}]$ : this is shown in Fig. 5.9c for various grain sizes. The amount of  $\mu[I_{ON}]$  increase produced by WVF is similar in magnitude to that of  $\mu[I_{ON}]$  reduction provided by LER with  $\Delta = 1.5\text{ nm}, \Lambda = 20\text{ nm}$ . Thus, the combined effect of the two fluctuation sources might result in almost zero net shift of the average drive current from its nominal value also reported in Fig. 5.9c. However, fluctuations of the drive current around its mean value are much stronger when LER is considered, while the WVF contribution to percentage  $I_{ON}$  variations is below 2.5% for  $g_s$  smaller than 20 nm, as shown in Fig. 5.9f. Estimation of drive current variability cannot be directly obtained with a simplified AWF model, which is only able to predict threshold voltage fluctuations.



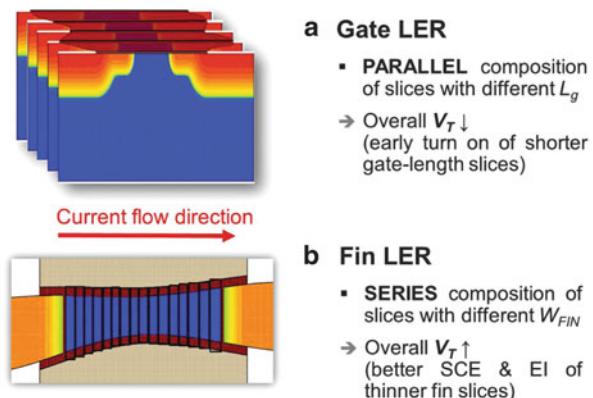
**Fig. 5.9** (a), (b) Mean value and (d), (e) standard deviation of drive current as a function of (a), (d) rms amplitude and (b), (e) correlation length of the roughness. Results of the Monte Carlo approach are compared with estimations based on sensitivity analysis and on corner points selected from the  $I_{ON}$  vs.  $W_{equiv}$  and  $I_{ON}$  vs.  $\langle W_{fin} \rangle_{tot}$  distributions. (c), (f) Impact of WFG with different grain sizes  $gs$  on the mean value and percentage ratio of standard deviation to mean value of  $I_{ON}$ . LER contribution for  $\Delta = 1.5 nm$ ,  $\Lambda = 20 nm$  is also reported for comparison

### 5.4.3 Physical Insight of Fin LER-Induced Threshold Voltage Increase

As shown in previous sections, the impact of LER and WFG is not limited to fluctuations of FinFET threshold voltage and on-current around the mean values of the respective distributions, but is also manifested in a shift of these values from the nominal  $V_T$  and  $I_{ON}$ . These shifts should be studied in detail as they provide physical insight into the way the analyzed fluctuation sources influence the device behavior. Specifically, shifts of the mean  $V_T$  are considered in this section, while the on-current behavior is analyzed more thoroughly in the next section.

It was shown in Fig. 5.6a,b that line-edge roughness causes the mean value of threshold voltage to increase compared with predictions based on a conventional sensitivity analysis. This was confirmed by  $\mu[V_T]$  extraction as a function of LER parameters in Fig. 5.7a,b. This “ $V_T$ -raising” phenomenon, which has been observed by several authors [34], is opposed to the  $V_T$ -lowering effect produced by gate-LER [7]. An explanation to it can be found by comparing the two situations [9]: as shown in Fig. 5.10a, a transistor with rough gate edges can be represented as a parallel combination of device slices with different gate lengths [26], where the early turn-on of the shorter gate-length slices lowers the overall threshold. Instead, fin-LER may be modeled through a series composition of fin slices with different

**Fig. 5.10** Interpretation of shifts in the mean  $V_T$  caused by (a) gate-LER and (b) fin-LER

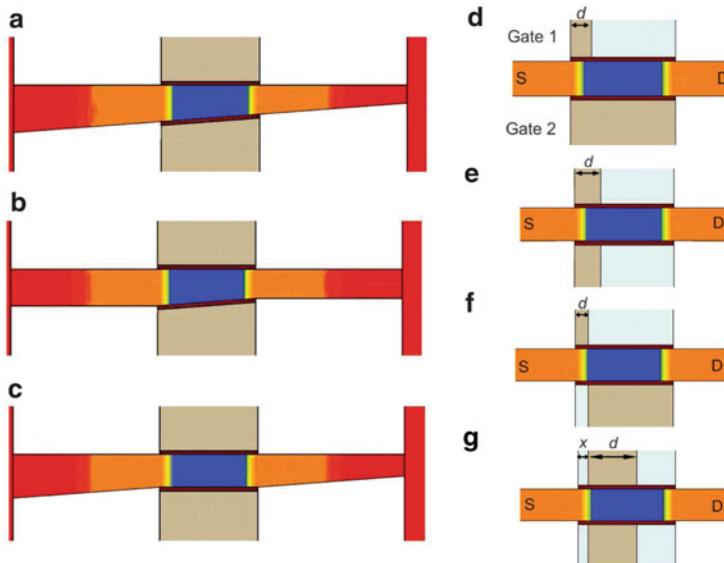


widths, as in Fig. 5.10b: in this case, the overall conduction is determined by the thinner slices, where higher electrostatic control and reduced SCEs contribute to raising the device  $V_T$ . This interpretation is able to explain trends of  $\mu[V_T]$  when LER characteristic parameters are varied. Specifically, a higher rms amplitude of the roughness decreases the minimum slice thickness, while a smaller correlation length increases the probability of a thin slice occurring within the channel. Both situations thus cause the average  $V_T$  to increase based on the slice model: this is confirmed by simulation results presented in Fig. 5.7a,b. Further investigation is required to check whether the opposite effects of the fin and gate roughness on FinFET threshold voltage can compensate each other.

It is also interesting to consider the impact of work-function variations on the mean  $V_T$ . Of course this parameter is significantly lower than the “nominal” case in Fig. 5.7c, which refers to a uniform gate with the high WF value  $\Phi_1$ . However,  $\mu[V_T]$  is not independent of the grain size, but is seen to increase with decreasing  $g_s$ . The parallel/series explanation can be applied to this situation as well: the 3D FinFET in Fig. 5.2 is modeled as a parallel composition of slices, but each slice contains a sequence of grains in a series configuration along the current flow direction. Grain size reduction increases the probability of having at least a grain with the high WF value within the gate, which tends to increase the overall device threshold.

## 5.5 Asymmetric Impact of Localized Fluctuations

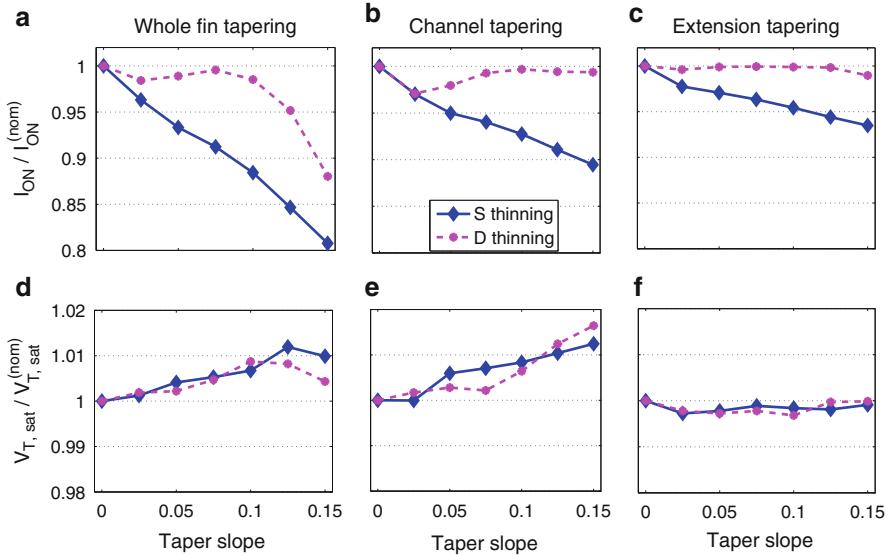
Fin-LER causes local thinning/thickening of the fin, and it was shown in Sect. 5.4 that the device behavior is sensitive to the locations where these width variations occur. For example, specific fin regions highlighted in Fig. 5.5c give the highest contribution to variations in the on-current, and asymmetric importance of the source (S) and drain (D) side of the device has been detected. Similarly, the size



**Fig. 5.11** Simplified structures for the analysis of local fluctuations. Local fin thinning produced by LER is studied by a device with tapering introduced in (a) the whole fin, (b) the channel region only, or (c) the extensions only. The impact of metal grain position and size is studied through gate configurations reported in (d)–(g), where  $d$  is the size of the  $\langle 200 \rangle$ -oriented grain ( $g_1$ ) in “Gate 1”, and  $x$  indicates its distance from the S gate end

and specific location of metal grains within the gate is expected to significantly influence the device performance.

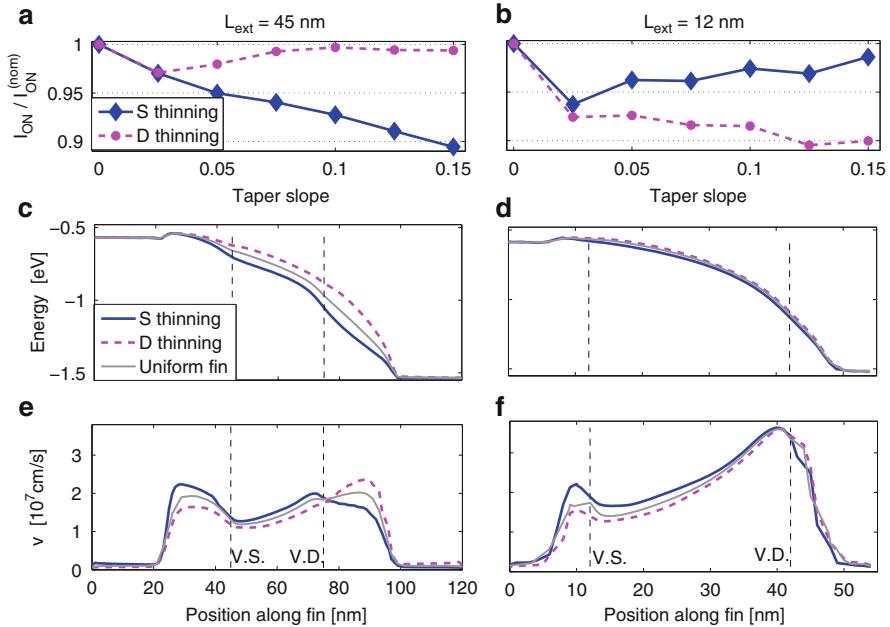
Simplified device structures shown in Fig. 5.11 are considered in this section to gain physical insight into the way LER and WFV influence FinFET behavior. In particular, a tapered structure is studied to better elucidate the local impact of fin-width fluctuations.  $W_{\text{fin}}$  is varied linearly over the whole fin length (Fig. 5.11a) or just in the channel (Fig. 5.11b) or extensions (Fig. 5.11c), while keeping the overall average fin width fixed to 10 nm. The tapering is applied either from S to D or vice versa. As for WFV, the impact of size and position of metal grains with different orientations is analyzed on the basis of four experiments conducted on a single device slice. In the first case (Fig. 5.11d) one of the gates is split into two grains whose size is varied, while a single WF is used for the second gate. The second and third configurations (Fig. 5.11e,f) account for the interaction of nonuniform orientations in both gates, respectively, considering the same or opposite area fractions for the two possible WFs in the two gates. The last considered case (Fig. 5.11g) investigates the impact of varying the position of a grain with fixed size.



**Fig. 5.12** Relative variation of saturation drain current ( $I_{ON}$ ) and threshold voltage ( $V_{T,sat}$ ) with respect to nominal values in Table 5.1, as a consequence of different fin tapering configurations described in Fig. 5.11a–c. “Taper slope” indicates the fin width variation per unit fin length, while “S thinning” and “D thinning” are the taper directions.  $I_{ON}$  is extracted at  $V_{DS} = V_{GS} = 1$  V

### 5.5.1 Impact of Local Fin Thinning

Changes in drive current and saturation threshold voltage as a consequence of the tapering configurations of Fig. 5.11a–c are shown in the three columns of Fig. 5.12, respectively. The  $x$  axis in these plots represents the fin thickness change per unit fin length associated with different slopes of the inclined fin edge. Increasing slopes correspond to a more pronounced constriction of the fin on one side of the device, and to a proportionate thickening on the opposite side. Figure 5.12a–c indicate that a local thinning of one fin end always degrades the  $I_{ON}$  with respect to the nominal case, although it is compensated by an equal thickening at the other end. This explains the LER-induced  $I_{ON}$ -lowering observed in Sect. 5.4.1 (see especially Fig. 5.6d): any local thinning of the fin produced by LER patterns contributes to reducing FinFET drive current. However, in the saturation regime where  $I_{ON}$  is extracted, this degradation is strongly dependent on the tapering direction: geometry constrictions towards the source cause a strong reduction of  $I_{ON}$ , while drain thinning is less critical, except for extremely small values of the minimum thickness close to the D pad (see Fig. 5.12a). Figure 5.12d–f show that  $V_{T,sat}$  is basically insensitive to the tapering, because it is mainly determined by the average fin width in the channel (as discussed in Sect. 5.4.1), and  $\langle W_{fin} \rangle_{ch}$  is not changed by the tapering. Therefore, the observed  $I_{ON}$  variability cannot be ascribed to threshold voltage fluctuations. The  $I_{ON}$ -lowering is instead attributed to increased parasitic



**Fig. 5.13** (a), (b) Relative variation of  $I_{ON}$  produced by channel tapering in a FinFET with long ( $L_{ext} = 45\text{ nm}$ ) and short extensions ( $L_{ext} = 12\text{ nm}$ ), respectively, in comparison with a device with uniform fin width (nominal case,  $I_{ON}^{(n o m)}$ ). (c), (d) Potential energy and (e), (f) electron velocity plotted for the two FinFET structures versus the position along the fin (V.S. = virtual source, V.D. = virtual drain). These cuts were taken inside the fin, 5 nm away from the axis-aligned edge (see Fig. 5.11b), but similar trends were observed at different y positions

resistance in both the channel ( $R_{ch}$ ) and S/D extension regions ( $R_{S/D}$ ). The  $R_{ch}$  contribution seems to be dominant when comparing Fig. 5.12b,c. However, channel tapering also results in different widths of the S and D extensions (see Fig. 5.11b), so the influence of  $R_{S/D}$  is not completely eliminated in the data of Fig. 5.12b.

To better decouple the role of channel and extension regions, the tapered structure of Fig. 5.11b is compared with a device whose extensions are shortened from  $L_{ext} = 45\text{ nm}$  to  $L_{ext} = 12\text{ nm}$ . Results in Fig. 5.13a,b reveal that the taper direction has opposite effects on the two configurations, i.e. the FinFET with short extensions experiences a more pronounced reduction of the on-current when the fin constriction occurs at the drain side. This can be interpreted according to Lundstrom's model for quasi-ballistic transport [20]. In fact, geometrical constrictions in a nanoscale device enhance quantization effects, thus raising the energy of subbands in the associated cross sections. When the constriction occurs towards the D side of the FinFET, the global effect on device electrostatics is a wider  $kT$ -layer, i.e. a less steep decay of the potential energy along the channel compared to a device with uniform fin. This results in a larger backscattering coefficient and therefore in a smaller injection velocity at the virtual source (V.S.). Swapping the thinning direction is expected to

produce an opposite effect and ideally improves drive current of the *intrinsic* (i.e., no extensions) device with respect to the straight-fin case [28]. Figure 5.13c,d indeed show a wider  $k$   $T$ -layer in the “D thinning” case for both devices with long and short extensions, which is expected to result in lower velocity at the V.S. compared to a device with uniform fin width. This is confirmed by Fig. 5.13e,f. Opposite trends are observed for the “S thinning” case. It can be deduced that conduction of the *intrinsic* transistor is degraded when the fin thickness is reduced from the virtual source towards the virtual drain (V.D.), and enhanced if the tapering direction is swapped. However, tapering has an opposite effect on the parasitic S/D resistance, so even very short extensions of length  $L_{\text{ext}} = 12$  nm are sufficient to impede  $I_{\text{ON}}$  improvement in the “S thinning” case (Fig. 5.13b), and longer extensions produce opposite effects of the thinning direction compared to the intrinsic device (Fig. 5.13a).

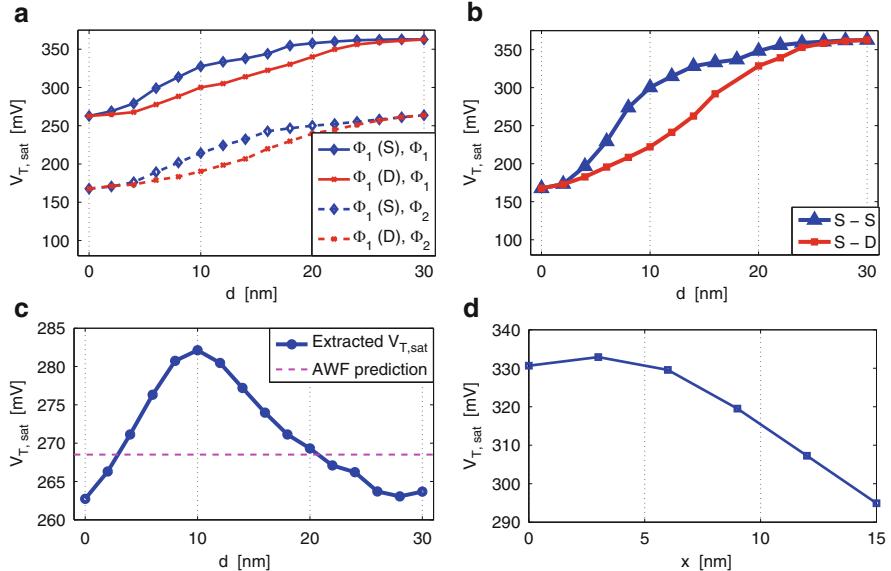
This asymmetric impact of fin width fluctuations on the parasitic S/D resistance can be explained by recalling that the sheet resistance  $R_{\text{sh}}$  is inversely proportional to  $W_{\text{fin}}$  [15]. Decreasing the fin width in the S extension thus results in a higher  $R_{\text{sh}}$  in this region, which in turn reduces the effective  $V_{\text{GS}}$  and hence the drive current [30]:

$$V_{\text{GS},\text{eff}} = V_{\text{GS}} - R_{\text{S}} I_{\text{D}} \quad (5.1)$$

The impact of  $R_{\text{D}}$  is much lower because  $V_{\text{GS},\text{eff}}$  is not affected by this parameter. The critical importance of extension resistances for FinFET variability is confirmed by experimental data [22]. The highlighted asymmetry in roles of  $R_{\text{S}}$  and  $R_{\text{D}}$  is in agreement with results of the correlation analysis in Sect. 5.4.1. The main conclusion is that electrical fluctuations are especially sensitive to LER in the source side of the device. Optimized extension engineering should be performed taking these asymmetries into account in order to enhance FinFET robustness to fin-LER.

### 5.5.2 Impact of Grain Location and Size

Changes in FinFET performance as a consequence of different metal grain configurations are analyzed with reference to the four situations illustrated in Fig. 5.11d–g. The first case is that of a single grain boundary located at various positions on “Gate 1,” while “Gate 2” is uniform with work-function equal to either  $\Phi_1$  or  $\Phi_2$  in Table 5.2. The grain with high WF  $\Phi_1$  in “Gate 1” is referred to as  $g_1$ , and its size  $d$  is increased from zero to the whole gate length  $L_g$ , starting from either the S or the D side of the device. The corresponding variation of saturation threshold voltage is plotted in Fig. 5.14a, where solid and dashed lines have been obtained by setting the WF of the second gate to  $\Phi_1$  or  $\Phi_2$ , respectively.  $V_{\text{T}}$  obviously increases with  $d$  in all cases because  $\Phi_1 > \Phi_2$ , but the trends are nonlinear. This indicates that the device threshold cannot be rigorously considered as proportional to an area-weighted effective WF, in agreement with inaccuracies of the AWF model discussed in Sect. 5.4.2.1. Figure 5.14a also reports higher  $V_{\text{T,sat}}$  values when the

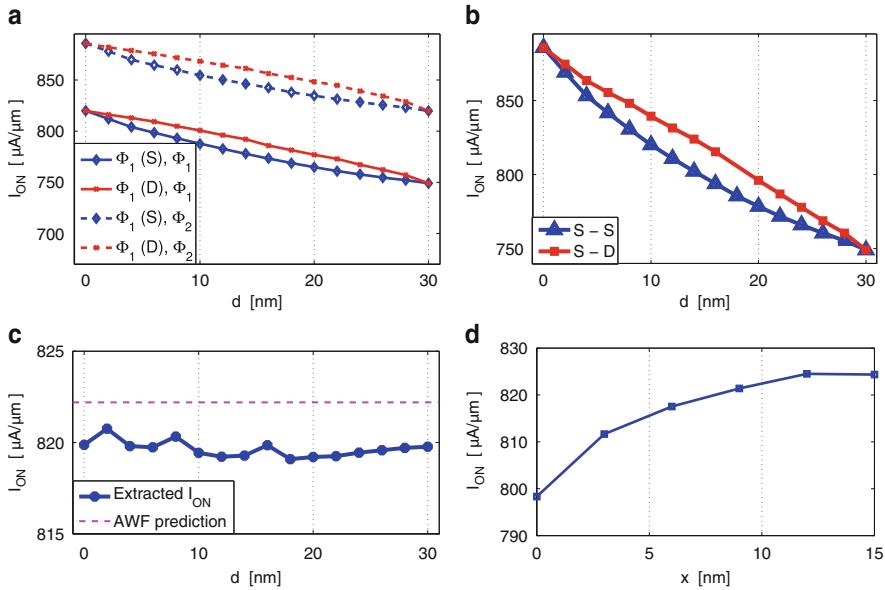


**Fig. 5.14** Changes in the saturation threshold voltage ( $V_{DS} = 1$  V) with varying metal grain size ( $d$ ) and position ( $x$ ) in the four configurations of Fig. 5.11d–g. (a)  $\Phi_1(S)$  and  $\Phi_1(D)$  in the legend indicate the position (source side or drain side) of the  $\langle 200 \rangle$ -oriented grain ( $g_1$ ) in “Gate 1” of Fig. 5.11d, and are followed by the WF of “Gate 2”. (b) Size  $d$  of  $g_1$  in the configuration of Fig. 5.11e is increased from S to D in both gates (S–S), or in opposite directions in “Gate 1” vs. “Gate 2” (S–D). (c) Impact of  $d$  in the configuration of Fig. 5.11f is compared with an AWF-based estimate of  $V_{T,\text{sat}}$  (dashed line). (d)  $V_T$  change when distance  $x$  of  $g_1$  from the S is varied while keeping its size fixed at  $d = 15$  nm

grain with high WF is located at the S (instead of at the D) side of the device, with a maximum difference of about 25 mV. This results in a lower on-current, as shown in Fig. 5.15a. Reducing the WF of the second gate from  $\Phi_1$  to  $\Phi_2$  shifts  $V_T$  and  $I_{ON}$  to lower and higher values, respectively, but does not change the discussed trends.

The second set of simulations considers the presence of a grain  $g_1$  with high WF  $\Phi_1$ , whose (variable) size  $d$  is the same in both “Gate 1” and “Gate 2”, and whose location with respect to the S is either the same (as illustrated in Fig. 5.11e) or opposite (i.e.,  $d$  increasing from S to D in “Gate 1” and from D to S in “Gate 2”). Different trends are observed for  $V_{T,\text{sat}}$  in Fig. 5.14b, depending on the relative position of  $g_1$  in the two gates: the threshold voltage increases more rapidly with  $d$  when  $g_1$  occupies the S end of both gates, while a more linear trend is observed when  $g_1$  is located on opposite sides in “Gate 1” and “Gate 2.” Consequently,  $I_{ON}$  in Fig. 5.15b has a roughly linear dependence on the grain size in the S–D configuration, while it decreases more rapidly as  $d$  increases in the S–S case.

The configuration illustrated in Fig. 5.11f presents one grain boundary with the same (varying) location in both electrodes, but opposite grain orientations in “Gate 1” vs. “Gate 2,” so that the total fraction of gate area with WF =  $\Phi_1$  is 0.5

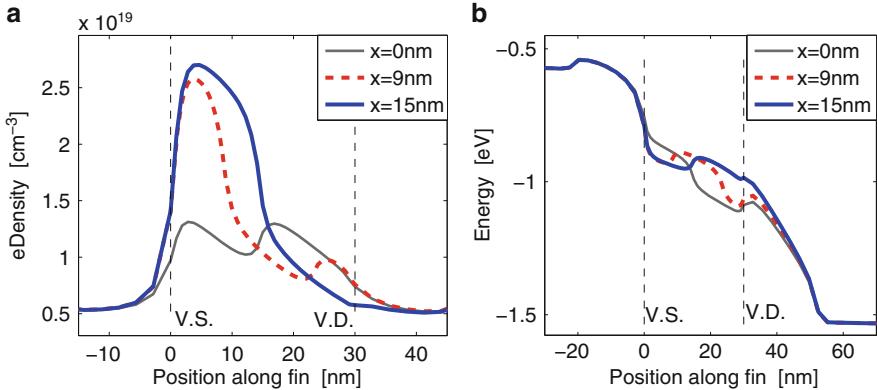


**Fig. 5.15** Changes in the on-state current ( $V_{GS} = V_{DS} = 1$  V) with varying metal grain size ( $d$ ) and position ( $x$ ) in the four configurations of Fig. 5.11d–g. (a)  $\Phi_1(S)$  and  $\Phi_1(D)$  in the legend indicate the position (source side or drain side) of the  $\langle 200 \rangle$ -oriented grain ( $g_1$ ) in “Gate 1” of Fig. 5.11d, and are followed by the WF of “Gate 2”. (b) Size  $d$  of  $g_1$  in the configuration of Fig. 5.11e is increased from S to D in both gates (S-S), or in opposite directions in “Gate 1” vs. “Gate 2” (S-D). (c) Impact of  $d$  in the configuration of Fig. 5.11f is compared with an AWF-based estimate of  $I_{ON}$  (dashed line). (d)  $I_{ON}$  change when distance  $x$  of  $g_1$  from the S is varied while keeping its size fixed at  $d = 15$  nm

regardless of  $d$ . Consequently, the AWF model predicts a constant value for each electrical parameter when  $d$  is varied, as indicated by dashed lines in Figs. 5.14c, and 5.15c. In contrast, simulations show a significant dependence of  $V_T$  on grain boundary position, with a threshold increase of up to 15–20 mV compared to AWF predictions when the boundary is located around the gate center (Fig. 5.14c). The strong inversion behavior is not severely affected by grain boundary location, probably due to compensation from the opposite configurations of the two gates, and  $I_{ON}$  is only slightly overestimated by the AWF model (Fig. 5.15c).

Finally, a worst-case situation is considered in Fig. 5.11g, where both gates present the same configuration, i.e. a grain  $g_1$  with fixed size  $d = 15$  nm and various positions along the gate ( $x$  = distance from the S). Results in Figs. 5.14d and 5.15d reveal that  $V_{T,sat}$  is reduced by about 10% as  $g_1$  is moved from S to D, while  $I_{ON}$  increases only slightly with  $x$ . Trends observed in the saturation regime may be interpreted through the MOSFET transport model in [20], starting from the expression of drain current:

$$I_D = WQ_i(0)\langle v(0) \rangle \quad (5.2)$$



**Fig. 5.16** (a) Electron density and (b) potential energy profiles at the Si/HK interface for different  $x$  values in the configuration of Fig. 5.11g ( $V_{GS} = V_{DS} = 1$  V). *Dashed vertical lines* indicate the gate endpoints

$\langle v(0) \rangle$  and  $Q_i(0)$  in Eq. (5.2) are the average carrier velocity and the inversion charge at the S end of the channel, respectively. The latter is related to the threshold voltage according to:

$$Q_i(0) = qn_s(0) \simeq C_{eff}(V_{GS} - V_T) \quad (5.3)$$

where  $n_s$  is the 2D inversion layer density and  $C_{eff}$  the effective dielectric capacitance. Moving grain  $g_1$  away from the S, i.e. inserting a grain  $g_2$  with lower WF  $\Phi_2$  at the S side, gives rise to a larger inversion charge in the channel portion beneath this grain, as indicated by electron density profiles at the Si/HK interface in Fig. 5.16a. Higher values of the peak carrier density produced by increasing the size of  $g_2$  correspond to an increase of  $Q_i(0)$  for equal  $V_{GS}$ , and hence to the observed  $V_T$  reduction and  $I_D$  increase, through Eqs. (5.3) and (5.2), respectively. Potential energy profiles at the Si/HK interface are shown in Fig. 5.16b and reflect the gate configurations associated with different values of  $x$ . The  $kT$ -layer tends to shrink as  $g_1$  is shifted away from the S, which should correspond to higher injections velocities  $\langle v(0) \rangle$  [20], thus further contributing to  $I_D$  increase according to Eq. (5.2). However,  $I_{ON}$  variation in Fig. 5.15d might be mitigated by mobility degradation due to higher transverse fields close to the V.S. for larger values of  $x$ .

The analysis presented in this section thus provides insight of WFV impact on FinFET physics and a better understanding of statistical data discussed in previous sections. Moreover, theoretical shortcomings of the AWF model are shown to provide awareness of the limits to its applicability and of the degree of accuracy that it can reach. Finally, the crucial role of grain orientation close to the source end

of each gate is highlighted. Optimization of metal gates for WFV mitigation should therefore seek ways to improve control over the polycrystalline structure of the electrode portion close to the source, unless granularity-related issues can be further suppressed by the use of amorphous metal gates [21].

## 5.6 Conclusions

The impact of line-edge roughness and work-function variability on FinFET electrical performance has been studied by comparing extensive Monte Carlo analysis based on TCAD simulation of large statistical ensembles with simplified models for variability estimation. Relevant electrical parameters have been monitored, including the saturation threshold voltage and on-state current. Simulation results indicate roughness of the fin edges as the major source of  $I_{ON}$  variability, while  $V_T$  fluctuations are dominated by WFV. Moreover, full-MC simulations reveal that fin-LER increases the average value of threshold voltage distributions and degrades the on-current compared with simplified predictions based on a conventional sensitivity analysis. Physical reasons for these effects have been investigated and correlation of electrical parameters to specific geometrical features has been highlighted, which provides approximate variability estimation at the same computational cost as sensitivity analysis, but with improved accuracy. TCAD simulations also reveal strong sensitivity of  $V_T$  to the position and size of individual grains that form the metal gate. This results in a nonlinear dependence of the threshold voltage on the area-weighted effective work function, thus highlighting limits to the applicability and accuracy of simplified WFV estimations based on the AWF model.

Asymmetries in the impact of local fluctuations at the source and drain side of the device have been investigated. Although any local thinning of the fin caused by LER has a negative impact on  $I_{ON}$ , the degradation is found to be stronger when the constriction occurs at the S end, due to the asymmetric role of parasitic S/D extension resistances. Similarly, electrical performance fluctuations due to WFV are mainly correlated to the orientation and size of grains at the source gate end. A physical interpretation of this phenomenon has been suggested. Device optimization for improved robustness to LER and WFV should therefore pay special attention to engineering the source extension and to control polycrystalline structure of the gate electrode portion close to the source. In circuit applications where yield depends on the far tails of device parameter distributions, the different nature of WFV- and LER-related statistics is expected to play a crucial role, since LER-induced distributions can have infinitely long tails, while WFV gives rise to bounded distributions.

## References

1. Arnaud, F., Liu, J., Lee, Y.M., Lim, S., Chen, J., Moon, B.K., Lai, C.W., Lipinski, M., Sang, L., Guarin, F., Hobbs, C., Ferreira, P., Ohuchi, K., Li, J., Zhuang, H., Mora, P., Zhang, Q., Nair, D.R., Lee, D.H., Chan, K.K., Satadru, S., Yang, S., Koshy, J., Hayter, W., Zaleski, M., Coolbaugh, D.V., Kim, H.W., Ee, Y.C., Sudijono, J., Thean, A., Sherony, M., Samavedam, S., Khare, M., Goldberg, C., A.Steegen: 32nm general purpose bulk CMOS technology for high performance applications at low voltage. In: IEDM Tech. Dig. (2008)
2. Asenov, A.: Random dopant induced threshold voltage lowering and fluctuations in sub 0.1 micron MOSFETs: a 3D ‘atomistic’ simulation study. *IEEE Trans. Electron Dev.* **45**, 2505–2513 (1998)
3. Asenov, A., Brown, A.R., Davies, J.H., Kaya, S., Slavcheva, G.: Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale MOSFETs. *IEEE Trans. Electron Dev.* **50**(9), 1837–1852 (2003)
4. Asenov, A., Brown, A.R., Roy, G., Cheng, B., Alexander, C., Ridder, C., Kovac, U., Martinez, A., Seoane, N., Roy, S.: Simulation of statistical variability in nano-CMOS transistors using drift-diffusion, Monte Carlo and non-equilibrium Greens function techniques. *J. Comput. Electron.* **8**, 349–373 (2009)
5. Asenov, A., Cathignol, A., Cheng, B., McKenna, K.P., Brown, A.R., Shluger, A.L., Chanemougame, D., Rochereau, K., Ghibaudo, G.: Origin of the asymmetry in the magnitude of the statistical variability of n- and p-channel Poly-Si gate bulk MOSFETs. *IEEE Electron. Dev. Lett.* **29**(8), 913–915 (2008)
6. Asenov, A., Kaya, S.: Effect of oxide interface roughness on the threshold voltage fluctuations in decanano MOSFETs with ultrathin gate oxides. In: Proc. SISPAD, pp. 135–138 (2000)
7. Asenov, A., Kaya, S., Brown, A.R.: Intrinsic parameter fluctuations in decananometer MOSFETs induced by gate line edge roughness. *IEEE Trans. Electron Dev.* **50**(5), 1254–60 (2003)
8. Auth, C., Allen, C., Blattner, A., Bergstrom, D., Brazier, M., Bost, M., Buehler, M., Chikarmane, V., Ghani, T., Glassman, T., Grover, R., Han, W., Hanken, D., Hattendorf, M., Hentges, P., Heussner, R., Hicks, J., Ingerly, D., Jain, P., Jaloviar, S., James, R., Jones, D., Jopling, J., Joshi, S., Kenyon, C., Liu, H., McFadden, R., McIntyre, B., Neirynck, J., Parker, C., Pipes, L., Post, I., Pradhan, S., Prince, M., Ramey, S., Reynolds, T., Roesler, J., Sandford, J., Seiple, J., Smith, P., Thomas, C., Towner, D., Troeger, T., Weber, C., Yashar, P., Zawadzki, K., Mistry, K.: A 22nm high performance and low-power cmos technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density mim capacitors. In: VLSI Tech. Dig., pp. 131–132 (2012)
9. Baravelli, E., De Marchi, L., Speciale, N.: Fin shape fluctuations in FinFET: correlation to electrical variability and impact on 6-T SRAM noise margins. *Solid State Electron.* **53**(12), 1303–1312 (2009)
10. Baravelli, E., Dixit, A., Rooyackers, R., Jurczak, M., Speciale, N., De Meyer, K.: Impact of line-edge roughness on FinFET matching performance. *IEEE Trans. Electron Dev.* **54**(9), 2466–2474 (2007)
11. Baravelli, E., Jurczak, M., Speciale, N., De Meyer, K., Dixit, A.: Impact of LER and random dopant fluctuations on FinFET matching performance. *IEEE Trans. Nanotechnology* **7**(3), 291–8 (2008)
12. Brown, A.R., Roy, G., Asenov, A.: Poly-Si-gate-related variability in decananometer MOSFETs with conventional architecture. *IEEE Trans. Electron Dev.* **54**(11), 3056–3063 (2007)
13. Dadgour, H., De, V., Banerjee, K.: Statistical modeling of metal-gate work-function variability in emerging device technologies and implications for circuit design. In: Proc. ICCAD, pp. 270–277 (2008)
14. Dixit, A., Anil, K.G., Baravelli, E., Roussel, P., Mercha, A., Gustin, C., Bamal, M., Grossar, E., Rooyackers, R., Augendre, E., Jurczak, M., Biesemans, S., De Meyer, K.: Impact of stochastic

- mismatch on measured SRAM performance of FinFETs with resist/spacer-defined fins: Role of line-edge-roughness. In: IEDM Tech. Dig. (2006)
15. Dixit, A., Anil, K.G., Collaert, N., Goodwin, M., Jurczak, M., De Meyer, K.: Analysis of the parasitic S/D resistance in multiple-gate FETs. IEEE Trans. Electron Dev. **52**(6), 1132–40 (2005)
16. He, J.L., Setsuhara, Y., Shimizu, I., Miyake, S.: Structure refinement and hardness enhancement of titanium nitride films by addition of copper
17. ITRS: <http://public.itrs.net>
18. Kovac, U., Reid, D., Millar, C., Roy, G., Roy, S., Asenov, A.: Statistical simulation of random dopant induced threshold voltage fluctuations for 35nm channel length MOSFET. Microelectron. Reliab. **48**(8-9), 1572–1575 (2008)
19. Li, Y., Hwang, C.H., Li, T.Y., Han, M.H.: Proces-variation effects, metal-gate work-function fluctuation, and random-dopant fluctuation in emerging CMOS technologies. IEEE Trans. Electron Dev. **57**(2), 437–447 (2010)
20. Lundstrom, M., Ren, Z.: Essential physics of carrier transport in nanoscale MOSFETs. IEEE Trans. Electron Dev. **49**(1), 133–41 (2002)
21. Matsukawa, T., Liu, Y., Mizubayashi, W., Tsukada, J., Yamauchi, H., Endo, K., Ishikawa, Y., O’uchi, S., Ota, H., Migita, S., Morita, Y., Masahara, M.: Suppressing  $V_t$  and  $G_m$  variability of FinFETs using amorphous metal gates for 14 nm and beyond. In: IEDM Tech. Dig., pp. 175–178 (2012)
22. Mérelle, T., Curatola, G., Nackaert, A., Collaert, N., van Dal, M., Doornbos, G., Doorn, T., Christie, P., Vellianitis, G., Duriez, B., Duffy, R., Pawlak, B., Voogt, F., Rooyackers, R., Witters, L., Jurczak, M., Lander, R.: First observation of FinFET specific mismatch behavior and optimization guidelines for SRAM scaling. In: IEDM Tech. Dig. (2008)
23. Mistry, K., Allen, C., Auth, C., Beattie, B., D. Bergstrom and, M.B., Brazier, M., Buehler, M., Cappellani, A., Chau, R., Choi, C.H., Ding, G., Fischer, K., Ghani, T., Grover, R., Han, W., Hanken, D., Hattendorf, M., He, J., Hicks, J., Huessner, R., Ingerly, D., Jain, P., James, R., Jong, L., Joshi, S., Kenyon, C., Kuhn, K., Lee, K., Liu, H., Maiz, J., McIntyre, B., Moon, P., Neirynck, J., Pae, S., Parker, C., Parsons, D., Prasad, C., Pipes, L., Prince, M., Ranade, P., Reynolds, T., Sandford, J., Shifren, L., Sebastian, J., Seiple, J., Simon, D., Sivakumar, S., Smith, P., Thomas, C., Troeger, T., Vandervoorn, P., Williams, S., Zawadzki, K.: A 45nm logic technology with High-k+Metal Gate transistors, strained silicon, 9 Cu interconnect layers, 193nm dry patterning, and 100% Pb-free packaging. In: IEDM Tech. Dig., pp. 247–50 (2007)
24. Mizuno, T., Okamura, J., Toriumi, A.: Experimental study of threshold voltage fluctuation due to statistical variation of channel dopant number in MOSFETs. IEEE Trans. Electron Dev. **41**(11), 22162221 (1994)
25. Ohmori, K., Matsuki, T., Ishikawa, D., Morooka, T., Aminaka, T., Sugita, Y., Chikyow, T., Shiraiishi, K., Nara, Y., Yamada, K.: Impact of additional factors in threshold voltage variability of Metal/High-k gate stacks and its reduction by controlling crystalline structure and grain size in the metal gates. In: IEDM Tech. Dig. (2008)
26. Oldiges, P., Lint, Q., Petrellot, K., Sanchez, M., Ieong, M., Hargrove, M.: Modeling line edge roughness effects in sub 100 nanometer gate length devices. In: Proc. SISPAD, pp. 131–4 (2000)
27. Sentaurus Device Manual: Synopsys, Inc.
28. Serra, N., Palestri, P., Smit, G., Selmi, L.: Multi-subband Monte Carlo simulations of  $I_{ON}$  degradation due to fin thickness fluctuations in FinFETs. Solid State Electron. **53**(4), 424–32 (2009)
29. Skotnicki, T., Fenouillet-Beranger, C., Gallon, C., Boeuf, F., Monfray, S., Payet, F., Pouydebasque, A., Szczap, M., Farcy, A., Arnaud, F., Clerc, S., Sellier, M., Cathignol, A., Schoellkopf, J.P., Perea, E., Mingam, R.F.H.: Innovative materials, devices, and CMOS technologies for low-power mobile multimedia. IEEE Trans. Electron Dev. **55**(1), 96–130 (2008)

30. Subramanian, V., Mercha, A., Parvais, B., Loo, J., Gustin, C., Dehan, M., Collaert, N., Jurczak, M., Groeseneken, G., Sansen, W., Decoutere, S.: Impact of fin width on digital and analog performances of n-FinFETs. *Solid State Electron.* **51**(4), 551–9 (2007)
31. Wang, X., Brown, A., Cheng, B., Asenov, A.: Statistical variability and reliability in nanoscale finfets. In: IEDM Tech. Dig., pp. 5.4.1–5.4.4 (2012)
32. Xiong, S., Bokor, J.: Sensitivity of double-gate and FinFET devices to process variations. *IEEE Trans. Electron Dev.* **50**(11), 2255–61 (2003)
33. Yagishita, A., Saito, T., Nakajima, K., Inumiya, S., Matsuo, K., Shibata, T., Tsunashima, Y., Suguro, K., Arikado, T.: Improvement of threshold voltage deviation in damascene metal gate transistors. *IEEE Trans. Electron Dev.* **48**(8), 1604–1611 (2001)
34. Yu, S., Zhao, Y., Song, Y., Du, G., Kang, J., Han, R., Liu, X.: 3-D simulation of geometrical variations impact on nanoscale FinFETs. In: Proc. ICSICT, pp. 408–11 (2008)
35. Zhang, X., Li, J., Grubbs, M., Deal, M., Magyari-Köpe, B., Clemens, B.M., Nishi, Y.: Physical model of the impact of metal grain work function variability on emerging dual metal gate MOSFETs and its implications for SRAM reliability. In: IEDM Tech. Dig., pp. 3.4.1–3.4.4 (2009)

# Chapter 6

## Characteristic and Fluctuation of Multi-fin FinFETs

Hui-Wen Cheng and Yiming Li

**Abstract** As the gate length of a metal oxide semiconductor field effect transistor (MOSFET) decreases, vertical channel transistors, such as the fin-typed field effect transistors (FinFETs) have attracted much attention because of their promising characteristics. In this chapter, we explore the electrical characteristics of 16 nm multi-fin FinFETs with different fin aspect ratios [AR = fin height ( $H_{\text{fin}}$ )/fin width ( $W_{\text{fin}}$ )]. The 16-nm multi-fin FinFET device and circuits' characteristics are simulated by solving a set of 3D quantum-mechanically corrected transport equations coupling with circuit nodal equations self-consistently. Device's electrical characteristics and their fluctuation are discussed with respect to the AR varying from 0.5 to 2 including the number of silicon channel fins. Dynamic and transfer characteristics of static random access memory, inverter, and analog circuits using single-/multiple-fin FinFETs are further discussed, respectively.

### 6.1 Introduction

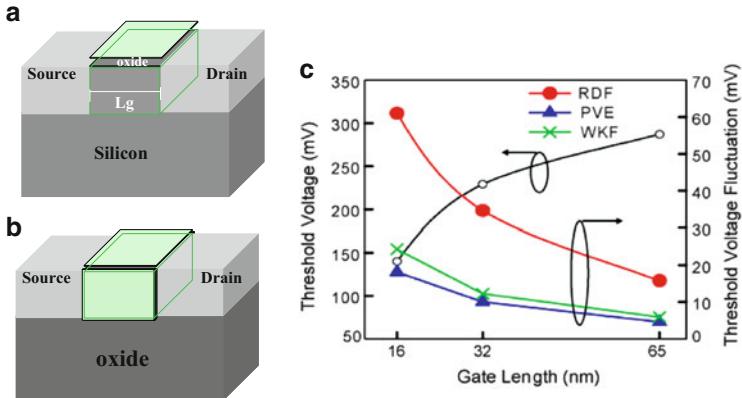
Shifting from a standard planar metal-oxide-semiconductor (MOS) device to a multi-gate field effect transistor (FET) design provides a promising alternative in semiconductor manufacturing. Intel® has first implemented 22-nm 3D MOSFET technology in 2011 and successfully discloses third generation core processors, codenamed Ivy Bridge [1] in 2012 due to having ten times less leakage and allowing chips with higher performance to operate at lower voltages [2–7]. Paying attention to what's happening with the technology, our studies of 16-nm 3D transistors benefit the advanced transistor design. An introduction of this chapter begins from the various methods of fluctuation suppression in nanoscale transistors,

---

H.-W. Cheng • Y. Li (✉)

Parallel and Scientific Computing Laboratory, Department of Electrical and Computer Engineering, National Chiao Tung University, 1001 Ta-Hsueh Road, Hsinchu 300, Taiwan, ROC

e-mail: [hwcheng@mail.ymlab.org](mailto:hwcheng@mail.ymlab.org); [yml@faculty.nctu.edu.tw](mailto:yml@faculty.nctu.edu.tw)



**Fig. 6.1** Schematic (a) planar and (b) SOI single-fin FinFET structures. (c) Plot of  $V_{th}$  and  $V_{th}$ 's fluctuation induced by RDF, PVE, and WKF for different gate lengths of n-type MOSFETs

where the literature review of current research status and motivation are presented briefly. The inevitable random dopant (RD)-induced threshold voltage ( $V_{th}$ ) fluctuation is one of the most critical challenges for the future of planar MOS devices [8]. The random dopants along the channel regions are considered to be one of the major sources of local threshold voltage fluctuation. Random dopants effects have been recently studied [9–14]. These studies that showed characteristic fluctuations caused by implantation processing are determined not only by a variation in an average doping density, which is associated with a fluctuation in the number of impurities, but also with a particular process-dependent random distribution of impurities in the channel region. Among various fluctuation reductions, such as doping profile engineering, vertical channel structure, and circuit topology, the vertical channel structure is a promising and effective way to reduce the fluctuation. The details will be discussed in this section.

### 6.1.1 Random Dopant Fluctuation

For estimating fluctuation, the planar and SOI single-fin FinFET structures we used here are based on previous study [9, 10, 13, 14], as shown in Fig. 6.1a, b. Figure 6.1c shows the RD, process variation effect (PVE), and work-function fluctuation (WKF)-induced  $V_{th}$  fluctuation ( $\sigma V_{th}$ ) of the 65-, 32-, and 16-nm-gate n-type MOSFETs. The roll-off of the nominal  $V_{th}$  follows the gate length decreased shown in black solid line with white dots, where the nominal  $V_{th}$  for the 16-, 32-, and 65-nm-gate devices are 140 mV, 220 mV, and 280 mV, respectively.

Compared with these fluctuations, the RD fluctuation (RDF) dominates the  $\sigma V_{th}$ , which may result in critical issue of stability, as shown in red line of Fig. 6.1c. The  $\sigma V_{th}$  of 16-nm MOSFET is around four times larger than that of 65-nm, which follows the trend of the analytical model

$$(\sigma V_{\text{th,RDF}})^2 = 3.19 \times 10^{-8} \frac{t_{\text{ox}} N_A^{0.401}}{\sqrt{WL}}, \quad (6.1)$$

where  $t_{\text{ox}}$  is the thickness of gate oxide, and  $W$  and  $L$  are the width and length of transistor [14]. To find the most efficient way to reduce the RDF, we discuss methods from device engineering points of view to achieve it. Here, we propose those successful studies and then select one of them to continue the future fluctuation study.

### 6.1.2 Reduction Techniques of Random Dopant Fluctuation

About implantation techniques, we have developed a Kinetic Monte Carlo (KMC) model to simulate impurity implantation along the channel region of the devices. Then the 3D physical positions of activated random discrete dopants within the channel regions are extracted from the KMC model. Figure 6.2a shows a 3D discrete random dopants distribution along the channel regions of a 15-nm device with  $3 \times 10^{20} \text{ cm}^{-3}$  as the source/drain doping concentration and  $1 \times 10^{15} \text{ cm}^{-3}$  as the substrate concentration, where the drain voltage ( $V_D$ ) = 0.8 V, the gate length ( $L_{\text{gate}}$ ) = 15 nm, and the effective oxide thickness ( $T_{\text{ox}}$ ) = 0.8 nm. To study the fluctuations caused by the dopant number and the dopant position in the channel region, the discrete impurities generated from KMC in a prescribed large cube are mapped into device channel region for atomistic device simulation including discrete dopants [9, 10]. The number of dopants varies from 3 to 17, as shown in Fig. 6.2b, c. The device characteristic is simulated by solving a set of 3D density-gradient equations coupled with Poisson equation and electron–hole current continuity equations [15–17], as shown below:

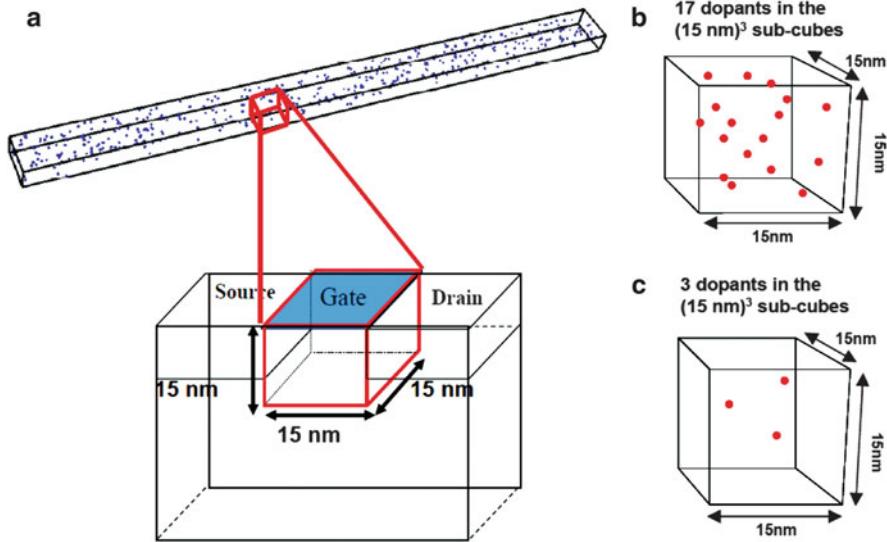
$$\Delta\phi = \frac{q}{\epsilon_s} (n - p + D), \quad (6.2)$$

$$\frac{1}{q} \nabla \cdot J_n = R(n, p), \quad (6.3)$$

and

$$\frac{1}{q} \nabla \cdot J_p = -R(n, p), \quad (6.4)$$

where  $\phi$  is the electrostatic potential and its unit is volt.  $n$  and  $p$  are classical electron and hole concentrations,  $q$  is the elementary charge, and the net doping concentration is  $D(x, y, z) = N_D^+(x, y, z) - N_A^-(x, y, z)$ . The  $R(n, p)$  is the net recombination rate. The carrier's currents densities are given by



**Fig. 6.2** (a) Discrete active dopants distribution along the channel regions of a 15-nm MOSFET device. The number of dopants varies from (b) 17 to (c) 3

$$\vec{J}_n = -q\mu_n n \nabla \phi + qD_n \nabla n - qnu_n \nabla \gamma_n, \quad (6.5)$$

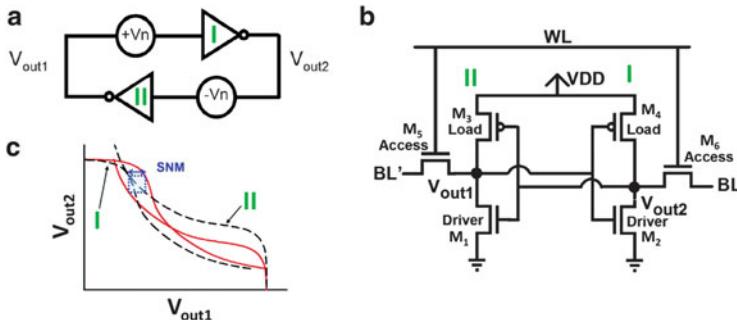
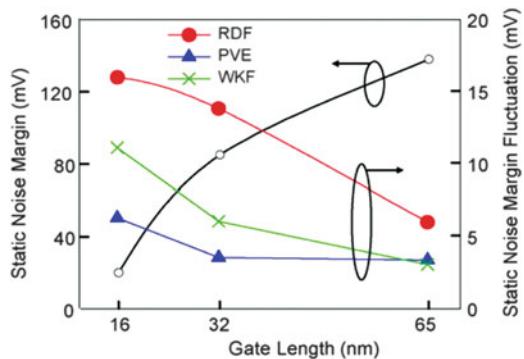
$$\vec{J}_p = -q\mu_p p \nabla \phi + qD_p \nabla p - qnu_p \nabla \gamma_p, \quad (6.6)$$

where  $\mu_n$  and  $\mu_p$  are the electron mobility and hole mobility. The diffusion coefficients,  $D_n$  and  $D_p$ , satisfy the Einstein relation.  $\gamma_n$  and  $\gamma_p$  are the quantum potentials for electrons and holes.  $\gamma_n = 2b_n \nabla^2 n^{1/2}/n^{1/2}$ ,  $\gamma_p = 2b_p \nabla^2 p^{1/2}/p^{1/2}$ ,  $b_n$  and  $b_p$  are density-gradient coefficients for electrons and holes.  $b_n = \hbar^2/(12qm_n^*)$  and  $b_p = \hbar^2/(12qm_p^*)$ .  $m_n^*$  and  $m_p^*$  are effective masses for the electrons and holes.  $\hbar$  is the Planck constant.  $b_n$  and  $b_p$  in (6.5) and (6.6) are the density gradient coefficient which determines the strength of the gradient effect in the electron and hole gas. The last terms in the right-hand side of (6.5) and (6.6) are referred to as “quantum diffusion,” which makes the electron continuity equation a fourth-order partial differential equation. The large-scale statistical device simulation is performed in our parallel computing system [18]. The density-gradient approximation is used to smooth the singularities of Coulomb potential by properly introducing related quantum-mechanical effects [10]. The statistical results are summarized in Table 6.1. For the n-type MOSFETs, the  $\sigma V_{th}$  of FLA is reduced from 74 to 71 mV by using LSA. The LSA is more attractive in p-type MOSFETs, in which the  $V_{th}$  is reduced from 110 to 70 mV. However, the reduced  $\sigma V_{th}$  by using LSA is insufficient. The more reduction on the characteristic fluctuation in circuit is necessary [19–21]. For example, the fluctuation in static random access memory

**Table 6.1** Summaries of mean ( $\mu$ ) and standard deviation ( $\sigma$ ) of the estimated saturated  $V_{th}$  for the n- and p-type MOSFETs with FLA and LSA processes

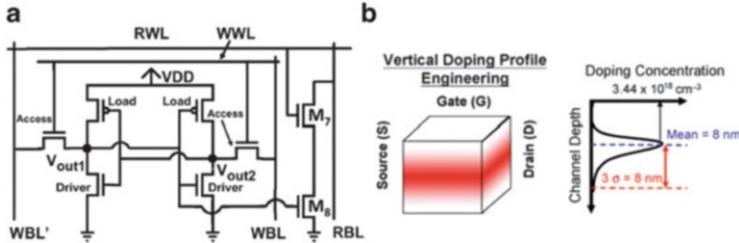
$V_{th,sat}$	n-type MOSFET		p-type MOSFET	
	FLA	LSA	FLA	LSA
$\mu$	0.355	0.462	0.493	0.488
$\sigma$	0.074	0.071	0.110	0.070
$\mu/\sigma$	20.8	15.3	22.4	14.3

**Fig. 6.3** Plot of SNM and SNM fluctuations that are induced by RDF, PVE, and WKF for different gate lengths of SRAM cells



**Fig. 6.4** (a) and (b) are the operation block diagrams and architecture of a 6T-SRAM Cell. (c) Plot of moving the static characteristics vertically or horizontally along the side of the maximum nested square until the curves intersect at only one point [22]

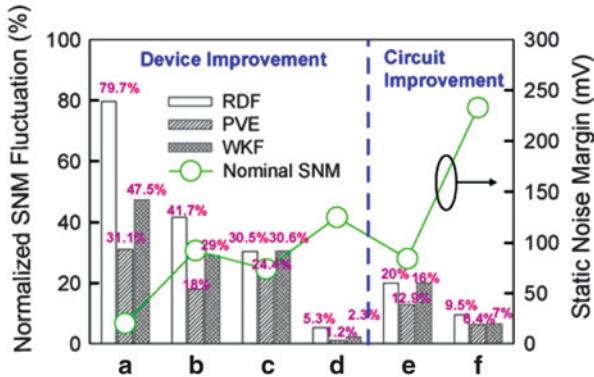
(SRAM) is saturated as the gate length is decreased, as shown in Fig. 6.3. Among fluctuations, the RDF is the most critical issue in improving stability of SRAM, where an SRAM cell is seen as two equivalent inverters with the noise sources insert between the corresponding inputs and outputs, as shown in Fig. 6.4a. Both series voltage noise sources ( $V_n$ ) have the same value and act together to upset the state of the cell. For example, while a noise signal which magnitude equal to  $+V_n$ , is applied on node  $V_{out1}$ , the  $V_{out2}$  will correspond a negative noise signal. Therefore, it shifts the transfer characteristic. Then it can be operated in write, read, or hold mode. Here, the read operation of SRAM is briefly introduced. As we know, each SRAM cell is composed by 4 n-type MOSFETs and 2 p-type MOSFETs, where  $M_1$  and  $M_2$  are



**Fig. 6.5** (a) Architecture of an 8T-SRAM cell, where the additional MOSFETs M<sub>7</sub> and M<sub>8</sub> in 8T-SRAM are used to avoid the impact; thus, they increase the read stability. (b) Illustration of the vertical doping profile from the surface to substrate which follows a normal distribution

driver MOSFETs, M<sub>3</sub> and M<sub>4</sub> are load MOSFETs, and M<sub>5</sub> and M<sub>6</sub> are access MOSFETs shown in Fig. 6.4b. We assume that the content of the memory is a logical 1 and stored at V<sub>out1</sub>, in this condition, M<sub>2</sub> and M<sub>3</sub> are turned off. The first step is to charge both bit lines to a logical 1, and then assert the word line. It will enable the access MOSFETs (M<sub>5</sub> and M<sub>6</sub>). Then the MOSFETs M<sub>4</sub> and M<sub>6</sub> will remain the bit line voltage at the precharged level (logical 1). Also the load capacitance of BL' will be discharged and pulled BL' toward logical 0. One of the important factors of a SRAM cell is the cell ratio (CR), which is defined as the ratio of width over length between driver and access MOSFETs, and will affect the performance of read operation. The stability of SRAM cell is often related to the static noise margin (SNM), due to the cell is most vulnerable to noise during a read access since the “0” storage node rises to a voltage higher than ground due to a voltage division along the access and inverter pull-down n-type MOSFET devices between the precharged bit line (BL) and the ground terminal of the cell [22]. There are several definitions of the SNM; the commonly used approach to estimating SNM is first proposed by Hill [23], where the SNM is defined as the minimum noise voltage present at each of the cell storage nodes necessary to flip the state of the cell shown in Fig. 6.4c. Developed from circuit design viewpoint, we have found that the planar 6T-SRAM with CR = 2 and planar 8T-SRAM require 30 % extra chip area, where the architecture of an 8T-SRAM Cell is shown in Fig. 6.5a. Unlike the 6T-SRAM, the access devices of 8T-SRAM can be turned off when data reading and the data will be gathered through M<sub>7</sub> and M<sub>8</sub> which can avoid direct flow between the bit line and the stored data and then increase the noise margin. The planar 8T-SRAM exhibits interesting SNM and  $\sigma_{\text{SNM}}$  because of the turnoff of the access devices when data reading. For planar devices with  $V_{\text{th}} = 140$  mV, the SNM of 8T-SRAM could be enlarged to 233 mV, and the  $\sigma_{\text{SNM}}$  is reduced to 9.5 %, as shown in Fig. 6.6f.

The circuit improvement approach can provide large SNM and is without changing the fabrication process. Therefore, 8T architecture could be considered for SRAM design with a compromise between performance and chip density. To keep a minimal chip area comparable with the conventional planar 6T-SRAM, the design approach from device engineering has to be developed. By adjusting the  $V_{\text{th}}$  to 350 mV, the SNM of the planar 6T-SRAM can be enhanced to 92 mV with 41.7 %  $\sigma_{\text{SNM}}$ , as shown in Fig. 6.6b. Using doping profile engineering can further



**Fig. 6.6** Summary of intrinsic-parameter-induced normalized SNM fluctuation and nominal SNM for different improvement techniques. (a)–(c) Planar 6T-SRAM with (a) CR = 1, (b) raised  $V_{th}$ , and (c) both raised  $V_{th}$  and doping-profile engineering. (d) 6T-SOI single-fin FinFET SRAM. (e) Planar 6T-SRAM with CR = 2. (f) Planar 8T-SRAM

reduce the RDF-induced  $\sigma_{SNM}$  to 30.5 %, as shown in Fig. 6.6c. The key technique of vertical doping profile engineering is to control fewer dopants near the current conducting path, as shown in Fig. 6.5b. However, the SNM is reduced to 71 mV because of the serious short channel effect. To have a large SNM with sufficient small  $\sigma_{SNM}$ , the explored SOI single-fin FinFETs 6T-SRAM exhibits a sufficiently large SNM (125 mV) with 5.3 %  $\sigma_{SNM}$ , as shown in Fig. 6.6d, where the structure of SOI single-fin FinFET is shown in Fig. 6.1b. From this study, we find that the SOI single-fin FinFET structure is not only without a cost of 30 % extra chip area but also suppresses RDF to 5.3 % significantly. Hence, the SOI FinFET structure is adopted in this chapter for further fluctuation estimation. The device simulation tasks are conducted using in-house device simulation programs [18, 24].

## 6.2 Effect of Channel Fin Aspect Ratio

The silicon on insulator (SOI) FinFET is studied owing to its high driving current, good suppression of SCEs, high transconductance, ideal subthreshold swing, better control over leakage, and manufacturing compatibility of planar MOSFETs. The high performance of SOI FinFET is strongly depended on fin aspect ratio (AR). Devices with multiple channel fin (multi-fin) were fabricated [25, 26]; computer simulation of electrical characteristics depending upon the channel fin AR of multi-fin devices will benefit the technology and design of multi-fin structures. A experimentally validated three-dimensional quantum drift-diffusion device simulation [27–30] coupled with device circuit simulation [30, 31] is simultaneously conducted to explore the device and digital circuit characteristics of single- and triple-fin structures with different fin aspect ratios (AR = 2, 1, 0.5). The electrical

characteristics of the devices are studied in terms of the threshold voltage roll-off, driving current, transconductance, gate capacitance, and intrinsic gate delay characteristics. In addition, the static noise margin of the state-of-the-art SRAM using six triple-fin FETs is investigated to examine the transfer characteristics. The transient characteristics are also examined through the estimation of the delay time of the inverter circuit with triple-fin structures.

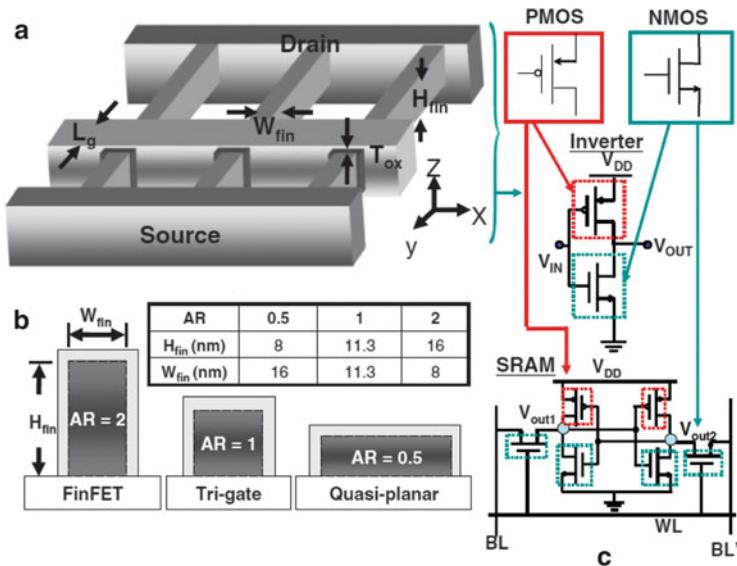
We first discuss the DC and AC characteristics of the n-type single- and triple-fin structures. Based on the results of n- and p-type MOSFETs, we examine the transfer characteristics and SNM of the 16-nm gate single- and triple-fin SRAM circuits. We further calculate the device capacitance for the analysis of the dynamic behavior of the CMOS inverter using 16-nm gate single- and triple-fin structures. We notice that the simulation time of DC and AC characteristics of triple fin are 2.5 and 5 h; the simulation time of inverter and SRAM with triple-fin devices are about 8 and 11 h, respectively.

### 6.2.1 Triple-Fin Devices

Figure 6.7a illustrates the structure of studied 3D triple-fin MOSFETs, where the MOSFETs are with different AR. The ARs of FinFETs, tri-gate, and quasi-planar MOSFETs are defined as 2, 1, and 0.5, respectively, as shown in Fig. 6.7b. The dimensions of three fin shapes and adopted device parameters of n-type MOSFETs are summarized in Tables 6.2 and 6.3; for p-type MOSFETs, not shown here, we have similar parameters. Except estimating roll-off characteristics, the cross-sectional areas of explored fins are fixed at about  $128 \mu\text{m}^2$  for a fair basis, as shown in Table 6.2. For devices' gate length starting from 32 nm, all threshold voltages of the explored transistors are first calibrated to 200 mV in order to examine  $V_{\text{th}}$  roll-off. The similar cross-sectional area and  $V_{\text{th}}$  indicate having the same control volume of the device channel under the same operation condition. The device transport characteristics are then calculated by solving a set of 3D density-gradient equations coupled with Poisson equations as well as electron–hole current continuity equations [27–30] under our parallel computing system [18, 32]. Figure 6.7c shows that the simulated 6T SRAM and inverter circuit using the 16-nm gate triple-fin structures are investigated for transfer and transient characteristics. There is no well-established compact model for 16-nm gate triple-fin structures, so the dynamic characteristic of the digital circuit is directly estimated using a coupled device-circuit simulation approach [30, 31]. It is worth noting that the physical models adopted in the 3D quantum-mechanically corrected device equations were calibrated with the fabricated and measured samples for the best accuracy [28].

### 6.2.2 Roll-Off Characteristics

Figure 6.8a, b shows the  $V_{\text{th}}$  roll-off for the examined single- and triple-fin n-type devices with respect to different ARs. The gate length varies from 32 to 16 nm. The triple-fin FinFET structure with AR = 2 is less sensitive to the gate length



**Fig. 6.7** (a) A schematic and (b) cross-sectional view of the triple-fin structure. Three different AR of channel fins are studied; they are FinFET (i.e., device with AR = 2), tri-gate (AR = 1), and quasi-planar (AR = 0.5) MOSFETs, respectively. (c) Inverter and SRAM are used as the tested digital circuits, where BL and BL' are bit lines; WL is word line

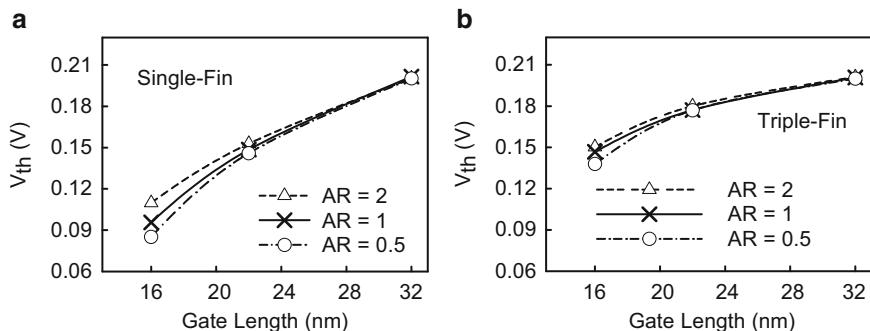
**Table 6.2** The dimension of channel fin height and channel fin width corresponding to three fin shapes: FinFET (AR = 2), tri-gate (AR = 1), and quasi-planar (AR = 0.5)

	AR = $H_{fin}/W_{fin} = 0.5$	AR = $H_{fin}/W_{fin} = 1$	AR = $H_{fin}/W_{fin} = 2$
$H_{fin}$ (nm)	8	11.3	16
$W_{fin}$ (nm)	16	11.3	8

**Table 6.3** Summary of device parameters for the explored n-type MOSFET

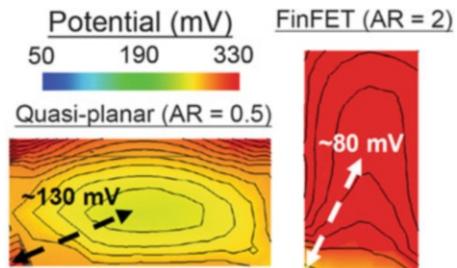
Parameter	Range
$L_g$ (nm)	16–32
$T_{ox}$ (nm)	1.2
Channel doping ( $\text{cm}^{-3}$ )	$1.48 \times 10^{18}$
S/D doping ( $\text{cm}^{-3}$ )	$\sim 3 \times 10^{20}$
Gate work function (eV)	~4.4

scaling. The alleviation of  $V_{th}$  roll-off implies that the FinFET possesses the better channel controllability and resistance to the intrinsic device parameter variations. Figure 6.9 shows the potential profile of the n-type triple-fin FinFET and quasi-planar device. According to Coulomb's law, the FinFET has larger perimeter between two lateral gates. Therefore, it exhibits a more uniform potential profile than quasi-planar structure. In addition, the longitudinal electric field within the fin will increase as the fin height is scaled down; consequently, it degrades the carrier

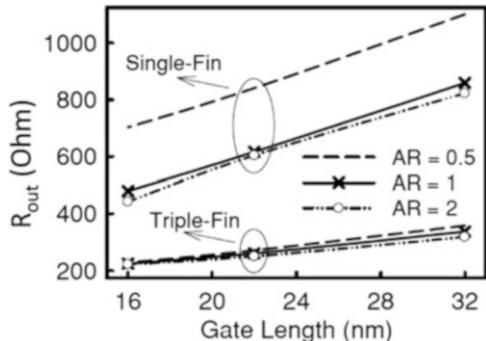


**Fig. 6.8** Plots of  $V_{th}$  roll-off characteristics for the 16-nm gate n-type (a) single-fin and (b) triple-fin MOSFETs with different fin ARs

**Fig. 6.9** Plots of on-state potential distributions for FinFET (AR = 2) and quasi-planar (AR = 0.5), where the gate length of the device is 16 nm



**Fig. 6.10** Plot of the total output resistance ( $R_{out}$ ) versus the gate length with respect to AR



mobility and limits the device saturation current. For output resistance of triple-fin structures with different ARs, they are not sensitive to gate length scaling compared with that of single-fin structures, as shown in Fig. 6.10. If the AR of single-fin structure is smaller than 1, the output resistance is increased rapidly.

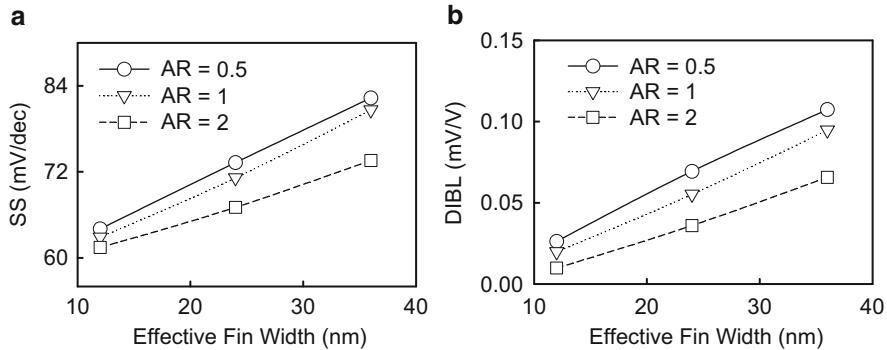
### 6.2.3 DC/AC Characteristics

For the considered n-type triple-fin MOSFETs, Fig. 6.11a, b shows the subthreshold swing and the drain-induced-barrier-lowering (DIBL) of single-fin MOSFETs as a function of effective fin width [=2 ×  $H_{\text{fin}}$  +  $W_{\text{fin}}$ ]. The DIBL is defined as the difference in threshold voltage when the drain voltage is increased from 0.05 to 1 V. From the layout efficiency viewpoint, to design a multiple-gate device with sufficiently suppressed SCEs, the FinFET exhibits the best layout area efficiency and its better electrostatic integrity. For FinFET structure, the requirement on the height of the fin to obtain a competitive layout density is achieved; for example, to design a device with the SS < 70 mV/dec, the layout area of FinFETs is 1.67 and 1.33 times smaller than those of quasi-planar and tri-gate structures. The FinFET may possess a better channel controllability, higher driving current, and better layout efficiency than structures with smaller AR. Figure 6.12a shows the on-state current ( $I_{\text{on}}$ , at the bias condition:  $V_D = V_G = 1$  V) of the explored 16-nm gate n-type MOSFETs, whose  $V_{\text{th}}$  are calibrated. Equation (6.8) estimates the dependence of  $I_{\text{on}}$  on the fin number and fin geometry:

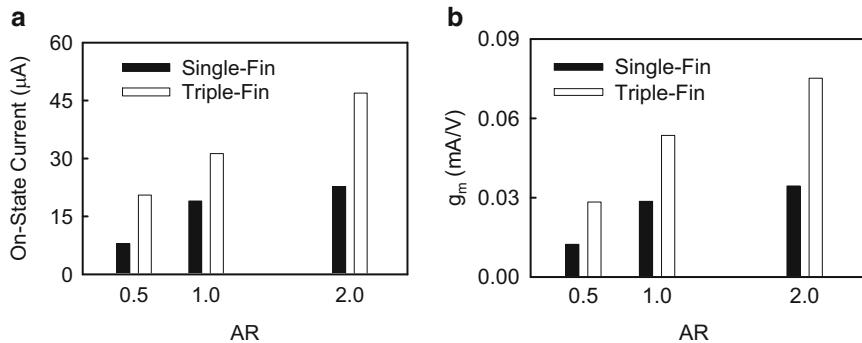
$$I_{\text{on}} = \frac{n(2H_{\text{fin}} + W_{\text{fin}}) \cdot \mu \cdot C_{\text{ox}}}{L} (V_G - V_{\text{th}})^2, \quad (6.7)$$

$$\Rightarrow I_{\text{on}} = \frac{n(2 \cdot \text{AR} + 1) \cdot W_{\text{fin}} \cdot \mu \cdot C_{\text{ox}}}{L} (V_G - V_{\text{th}})^2, \quad (6.8)$$

where  $n$  is the number of fins,  $H_{\text{fin}}$  is the fin height,  $W_{\text{fin}}$  is the fin width, as indicated in Fig. 6.7b;  $C_{\text{ox}}$  is the capacitance of per unit gate area,  $L$  is the gate length,  $\mu$  is the mobility, and  $V_G$  is the gate voltage. The device with a higher AR and more number of fins possesses a large on-state current. The FinFET structure provides 1.2 and 2.9 times larger  $I_{\text{on}}$  than the tri-gate and quasi-planar MOSFETs. Moreover, the triple-fin FinFETs offer a 2.1 times better driving capability than that with a single-fin structure. Figure 6.12b shows that the FinFET structure has the maximum transconductance ( $g_{\text{m,max}}$ ) among these explored structures, especially for the triple-fin structure. Generally speaking, the triple-fin device with a larger AR results in a larger on-state current as well as a smaller output resistance of the MOSFET which is smaller than that of a single-fin device, as shown in Fig. 6.10. Likewise, not shown here, we also have the simulated p-type MOSFETs. We note that theoretically, the on-state current obtained from equation is the approximation only in a gradual channel. It can be used only for qualitative discussion when the problem was solved using computer simulation. Figure 6.13a plots the gate capacitance ( $C_g$ ) of the explored 16-nm gate single- and triple-fin MOSFETs versus different ARs, where  $C_g$  is one of the important indexes for channel controllability. The  $C_g$  of 16-nm gate triple-fin MOSFETs is about three times larger than that of single-fin MOSFETs. It shows that the increase of AR and fin number increases the intrinsic  $C_g$  of MOSFETs. Additionally, compared with the triple-fin quasi-planar structures, the  $C_g$  of the triple-fin FinFETs is increased by a factor of 1.13. Though the large  $C_g$

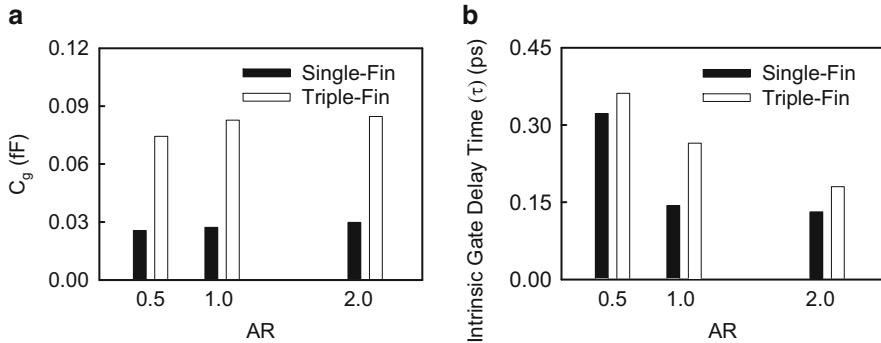


**Fig. 6.11** Plots of the (a) SS and (b) DIBL for the single-fin MOSFET with the quasi-planar, tri-gate, and FinFET structures, where the effective fin width ( $W_{\text{eff}}$ ) is calculated by  $2 \times H_{\text{fin}} + W_{\text{fin}}$

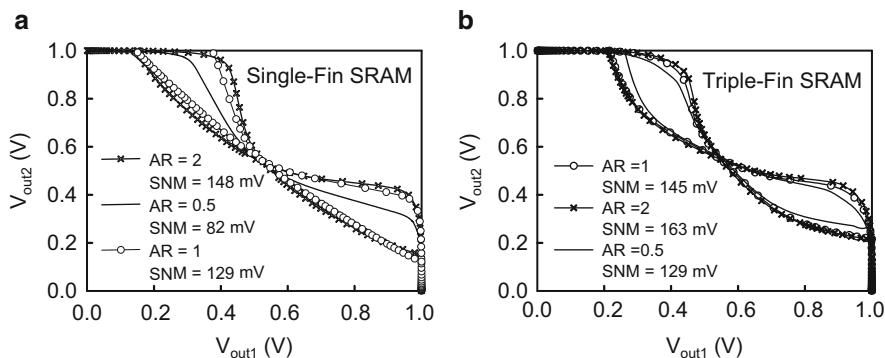


**Fig. 6.12** Plots of the (a) on-state current and (b) maximum transconductance for the studied 16-nm gate n-type single- and triple-fin MOSFETs with respect to different ARs

of the triple-fin transistor with a large AR can enhance the channel controllability, the increased  $C_g$  impacts the speed of the transistor. The trade-off between  $I_{\text{on}}$  and  $C_g$  is very important. Therefore, the intrinsic gate delay of the transistor ( $\tau = C_g V_{DD} / I_{\text{on}}$ ) is calculated, as shown in Fig. 6.13b. We find that the intrinsic gate delay is dominated by  $C_g$ , and the single-fin transistor exhibits a smaller delay than the triple-fin MOSFETs. For example, the intrinsic gate delay of the triple-fin FinFET is 1.4 times slower than that of the single-fin FinFET structure. The degraded intrinsic gate delay of the transistor may also impact the cutoff frequency in high frequency applications.



**Fig. 6.13** Plots of the (a) device gate capacitance and (b) intrinsic gate delay for the studied 16-nm gate n-type single- and triple-fin MOSFETs with different AR

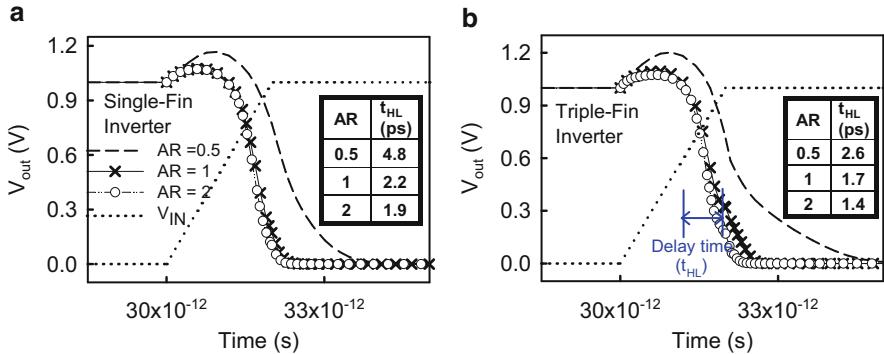


**Fig. 6.14** Plots of the SNM of the 6T-SRAM with the 16-nm gate (a) single-fin and (b) triple-fin transistors, where  $V_{out1}$  and  $V_{out2}$  refer to the symbols, where SRAM circuit is as shown in Fig. 6.7c

#### 6.2.4 Characteristics of SRAM and Inverter

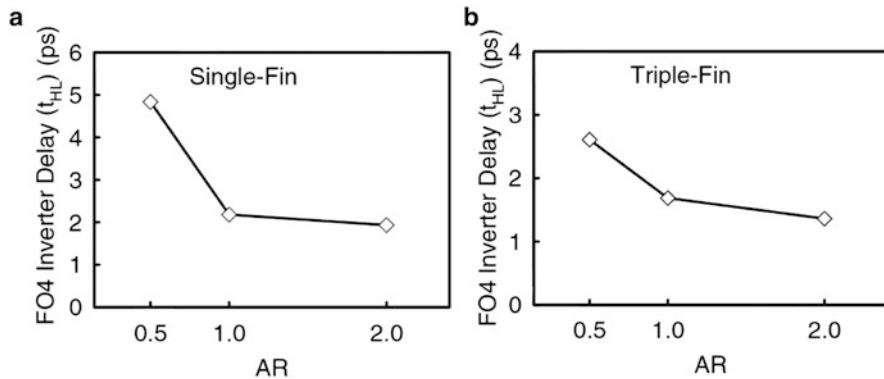
The calculated static noise margins of the 6T-SRAM with 16-nm gate single- and triple-fin FinFETs are shown in Fig. 6.14a, b. Here, the  $V_{th}$  of all 16-nm gate transistors has been calibrated to 200 mV. Based upon the 3D device simulation transfer curves, we can obtain SNMs which are numerically calculated by drawing and mirroring the inverter characteristics and find the maximum possible square between them [33]. The relation between the device transconductance and SNM of SRAM could be expressed as [22]

$$\text{SNM} \propto \sqrt{1 - \frac{I_{nx}}{g_{m,\text{pmos}}} - \frac{I_{ax}}{g_{m,\text{nmos}}}}, \quad (6.9)$$

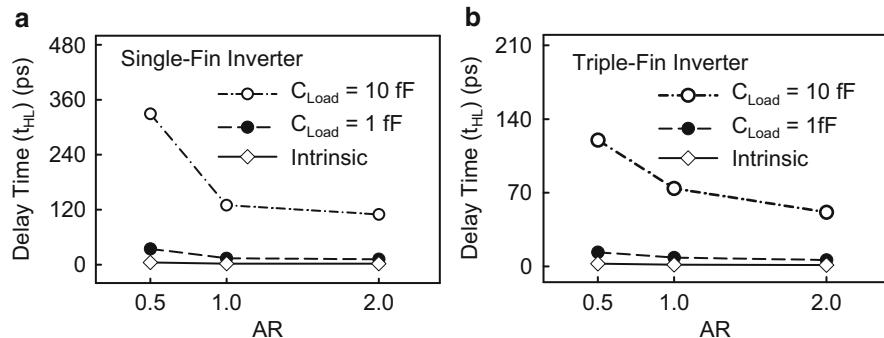


**Fig. 6.15** Plots of transient characteristics of (a) single- and (b) triple-fin inverters, where the extracted rise time, fall time, and hold time of the input signal are 2, 2, and 30 ps, respectively

where  $I_{nx}$  is the saturation drain current of the driver transistor of SRAM and  $I_{ax}$  is the saturation drain current of the access transistor. The large  $g_m$  of triple-fin FinFETs enlarges the SNM, as shown in Fig. 6.14b. The 6T-SRAM with 16-nm gate triple-fin FinFETs has a relatively larger SNM (about 163 mV from the inset of Fig. 6.14b) than that of single-fin FinFETs. Therefore, for the SRAM circuit, using a triple-fin FinFET structure is an effective way to improve performance. As described hereinbefore, the triple-fin FinFET possesses a better channel controllability, layout efficiency, driving current, and SNM than the transistors with smaller AR, such as tri-gate and quasi-planar MOSFETs. It is known that the large  $I_{on}$  of triple-fin FinFETs provides fast charge and discharge capabilities; however, it has larger intrinsic  $C_g$ . From device viewpoint, the  $\tau$  dominated by larger  $C_g$  has a lot of compensation from  $I_{on}$ , as shown in Fig. 6.13b. Hence, it is important to understand the impact of structure characteristics with different ARs on inverter behaviors. Therefore, the inverters with single- and triple-fin structures with different ARs are explored. Figure 6.15a, b is the high-to-low transition characteristics for single- and triple-fin inverters with respect to different ARs, in which the transistor's intrinsic capacitance is used as the load capacitance ( $C_{load}$ ). The solid lines are the output signal of devices with different fin shapes; the dotted line is the input signal. The rise time, fall time, and hold time of the input signal are 2 ps, 2 ps, and 30 ps, respectively. The high-to-low delay time ( $t_{HL}$ ) is defined as the difference between the times of the 50 % points of the input and output signals during the falling of the output signal. The delay times of the studied single- and triple-fin MOSFETs are summarized in the inset of Figs. 6.15a and 6.15b, respectively. As expected, both single- and triple-fin FinFET inverters present smallest  $t_{HL}$  among different ARs, and show the benefits of the FinFET structure in both DC and transient characteristics. Figure 6.16 shows a comparison of the single-fin and triple-fin FinFETs with AR = 0.5, 1, and 2 in terms of the fan-out of 4 (FO4) inverter delay. For the single-fin MOSFET, as shown in Fig. 6.16a, delay time decreases substantially as AR is increased. Increasing the number of fins and AR enhance



**Fig. 6.16** Plots of gate delay for inverters with fan-out of 4 (FO4) using (a) single- and (b) triple-fin structures



**Fig. 6.17** Plots of delay time for (a) single- and (b) triple-fin inverters with different load capacitances

driving ability and allow delay time to be further reduced, as shown in Fig. 6.16b. The delay time of the triple-fin FinFET inverter is about 1.4 times smaller than that of the single-fin FinFET inverter, for example.

In order to examine the associated delay time of the digital circuit, a load capacitance ( $C_{load}$ ) is further added on the inverter circuits which is composed of the  $C_g$  of the n-type and p-type MOSFETs, as illustrated in Fig. 6.13b. For the increased load capacitance, the required charge and discharge time are increased. Completely different from Fig. 6.13b, the delay times of the triple-fin inverters are smaller than those of the single-fin inverters, since the delay time is dominated by the driving capability of structures. As shown in Fig. 6.17, the solid lines are the inverter with merely the intrinsic device gate capacitance; the dashed and dashed-dotted lines are the inverter with the capacitive load of 1 fF and 10 fF, respectively. The delay time increases significantly as the load capacitance increases. For the triple-fin

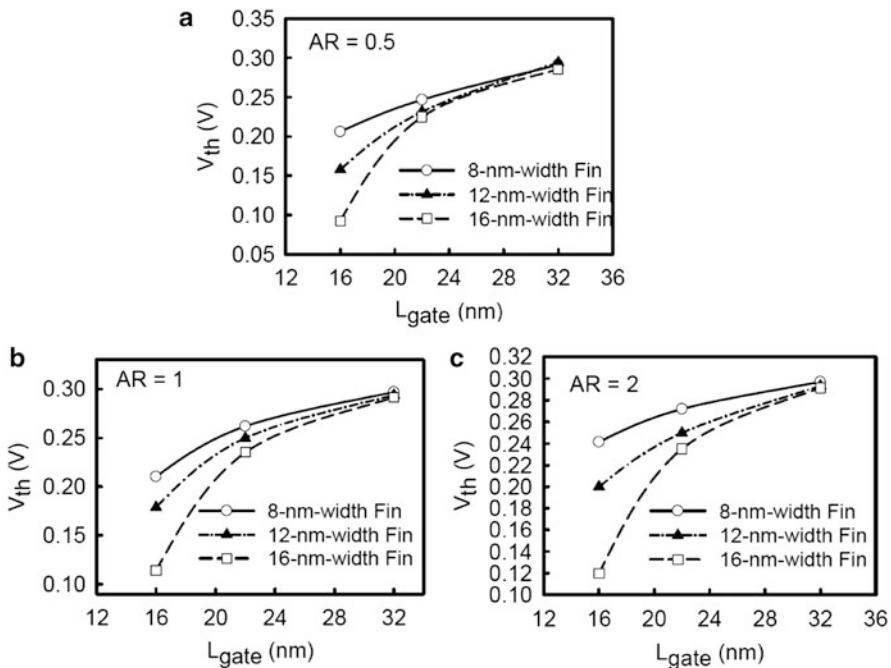
FinFET inverter with  $C_{\text{load}} = 10 \text{ fF}$ , the delay time approaches 50 ps, which is 30 times larger than that of the inverter only with the intrinsic device gate capacitance. Because the load capacitance dominates the overall capacitive load, the difference of the device's intrinsic gate capacitance resulting from the fin structure becomes negligible. For the triple-fin FinFET inverter with  $C_{\text{load}} = 10 \text{ fF}$ , the delay time is about two times smaller than that of the single-fin FinFETs. The triple-fin structure provides a smaller transition delay than that of the single-fin structure due to the increase of the driving current. In addition, the result shows that the triple-fin transistor with the FinFET structure exhibits a smallest delay time to benefit the timing of digital circuits. It is worth noting that the delay time for triple-fin FinFET inverters increases by a factor of 50, as the load capacitance increases from the device's intrinsic capacitance to 10 fF. Nevertheless, the delay time increases 70 times for triple-fin quasi-planar inverters. The increased difference of the delay time for the quasi-planar structure indicates that the high driving current of the FinFET mitigates the impact of load capacitance variation from passive components.

## 6.3 Channel Fin Aspect Ratio of Triple-Fin Structures

The 22-nm triple gate structure wraps the gate electrode around three sides of tall and narrow silicon fin [26], which is of particular interest as it combines excellent SCE immunity with high drivability per unit chip area. The vertical structure [43, 44] theoretically improves channel control and results in a steeper subthreshold slope [34] which can provide up to a  $10\times$  off-state leakage reduction [26]. Based on our triple-fin SOI FinFET structure, we explore the electrical characteristics of 16-nm tri-gate MOSFETs with different geometric aspect ratio through the dependence of SCEs of transistors on the device dimensions. To ensure the best accuracy, the used physical models of the explored device have been carefully calibrated with experimental data [27, 30, 35–37]. In this section, the electrical characteristics of the studied 3D devices are estimated, discussed, and compared between the single- and triple-fin structures with respect to different AR.

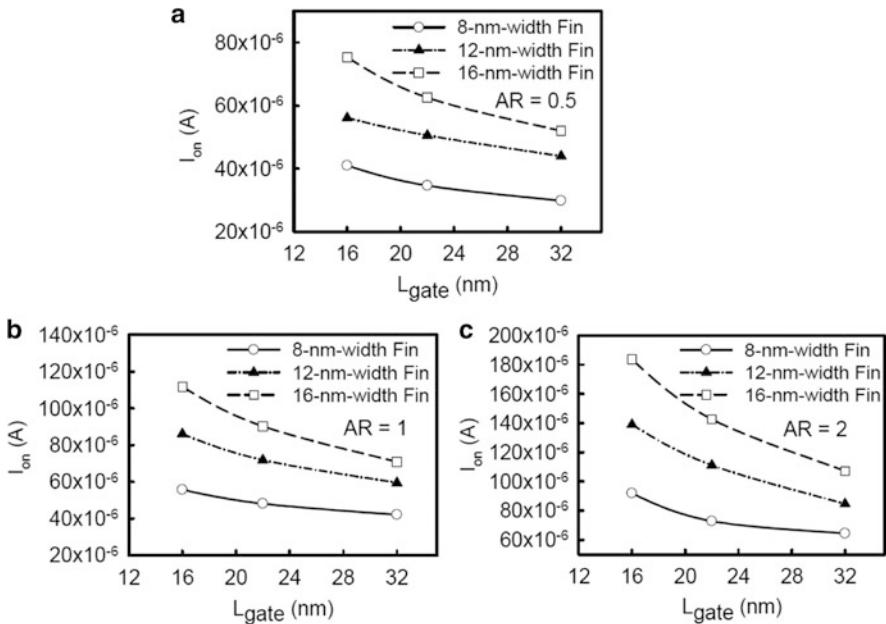
### 6.3.1 Roll-Off Characteristics of Triple-Fin Structure

In order to investigate the fin width impact on channel fin AR of triple-fin structure characteristics, several major index of transistor performance, such as the  $V_{\text{th}}$  roll-off, the on-state current ( $I_{\text{on}}$ ), the  $I_{\text{on}}/I_{\text{off}}$  ratio, the subthreshold swing, the maximum transconductance, the gate capacitance, the intrinsic gate delay ( $\tau$ ), and the cutoff frequency ( $F_T$ ), are examined comprehensively. For the 16-nm FinFET and the quasi-planar structures with a similar effective fin width, the potential difference inside the FinFET's channel is about 1.5 times smaller than that of quasi-planar device. The  $V_{\text{th}}$  roll-off characteristic for the triple-fin structure is then



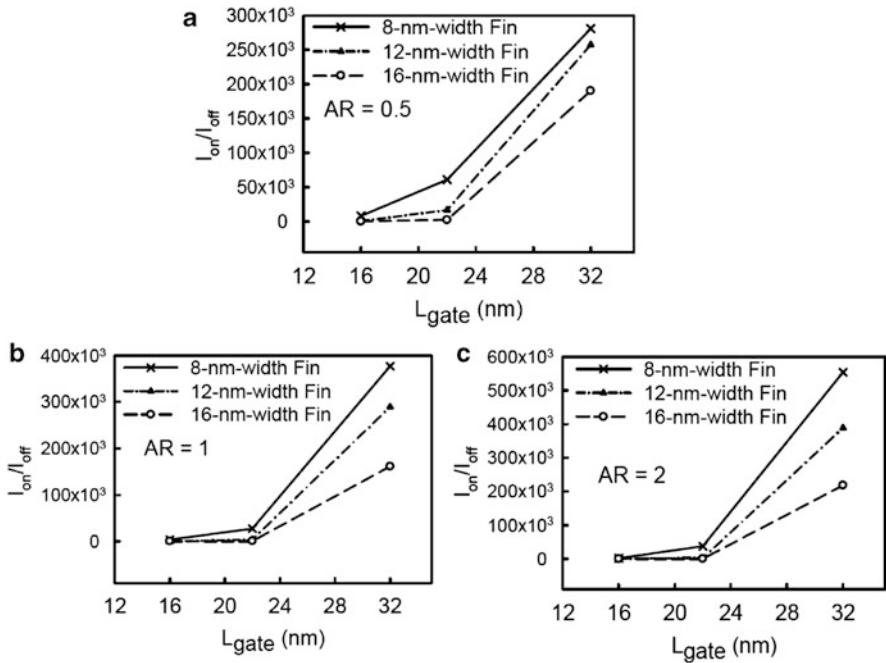
**Fig. 6.18** Plots of the  $V_{th}$  roll-off for the triple-fin (a) quasi-planar, (b) tri-gate, and (c) FinFET structures

estimated from the different AR's viewpoint, as shown in Fig. 6.18, in which the fin width varies from 8 to 16 nm, and the device gate length varies from 16 to 32 nm, where the  $V_{th}$  is determined exactly by a constant current ( $1 \times 10^{-7}$  A) at the intersecting point of tangent line of  $I_D - V_G$ . For triple-fin quasi-planar with 16-nm-width fin (the dash line with the square open symbols), as the gate length scales from 32 to 16 nm, the  $V_{th}$  decrease from 0.3 to 0.093 V, as shown in Fig. 6.18a. And the  $V_{th}$  difference is about 96 %. The  $V_{th}$  of triple-fin tri-gate with 8-nm-width fin decreases from 0.3 to 0.206 V as the gate length scales from 32 to 16 nm. The reduction of  $V_{th}$  is about 45 %. The tri-gate and FinFET exhibit similar characteristics. The results reveal that the structure with a thinner fin width may show less  $V_{th}$  roll-off characteristics due to the well gate control of fully depleted channel. The  $V_{th}$  roll-off differences between the wide width (the case of 16 nm) and the narrow one (the case of 8 nm) are increased with decreasing AR, which indicates the requirement of narrow width structure for MOSFETs with small AR design. However, we would like to mention that for devices with narrow width, the narrow-width effect will be enhanced and introduce another source of fluctuation in the threshold voltage. Besides the fin-width effect, for FinFETs and tri-gate MOSFETs with 16-nm-width fin (the dash line with the solid triangles and the dash-dot line with the solid triangles), as shown in Fig. 6.18b, c. The  $V_{th}$  differences are about 60 % and 65 %, respectively. Comparing triple-fin structures with



**Fig. 6.19** Plots of the on-state current versus the gate length for the triple-fin (a) quasi-planar, (b) tri-gate, and (c) FinFET structures

different aspect ratio, the FinFET structure possesses a smallest  $V_{th}$  variation. The  $I_{on}$  ( $V_G = 1$  V and  $V_D = 1$  V), the on/off current ratio ( $V_G = 0$  V and  $V_D = 1$  V), and the SS =  $d[\log(I_D)]/dV_G$  of the explored devices are then investigated, shown in Figs. 6.19, 6.20, and 6.21, respectively, where the structures with a thick fin width exhibit a larger on-state current due to the decrease of channel resistance. The on-state current of the 16-nm-fin width MOSFET is about two times larger than that of the 8-nm-fin width device. The relation between driving current and AR could be estimated by (6.8). The device with a higher aspect ratio may have a larger on-state current. The conducted  $I$ - $V$  curve simulation also verifies (6.8), where FinFET has a higher on-state current. Though the increase of transistor's fin width may enhance the driving capability, the wider fin width of transistors also may loss the channel controllability and then degrades the device performance, as shown in Figs. 6.20 and 6.21. The on/off current ratio and the SS for the wide transistors are degraded significantly, which forms a trade-off with driving capability and should be designed carefully. Moreover, for structures with 16-nm-gate length and 8-nm-fin width, the on-state currents of the FinFET is about 1.65 and 2.24 times larger than those of tri-gate and quasi-planar structures, respectively, as shown in Fig. 6.19. As shown in Fig. 6.20, the triple-fin FinFET with 16-nm-gate-length and 8-nm-fin width shows a larger ratio of the on/off current, which is about 1.94 and 4.43 times larger than those of tri-gate and quasi-planar structures. As for the SS

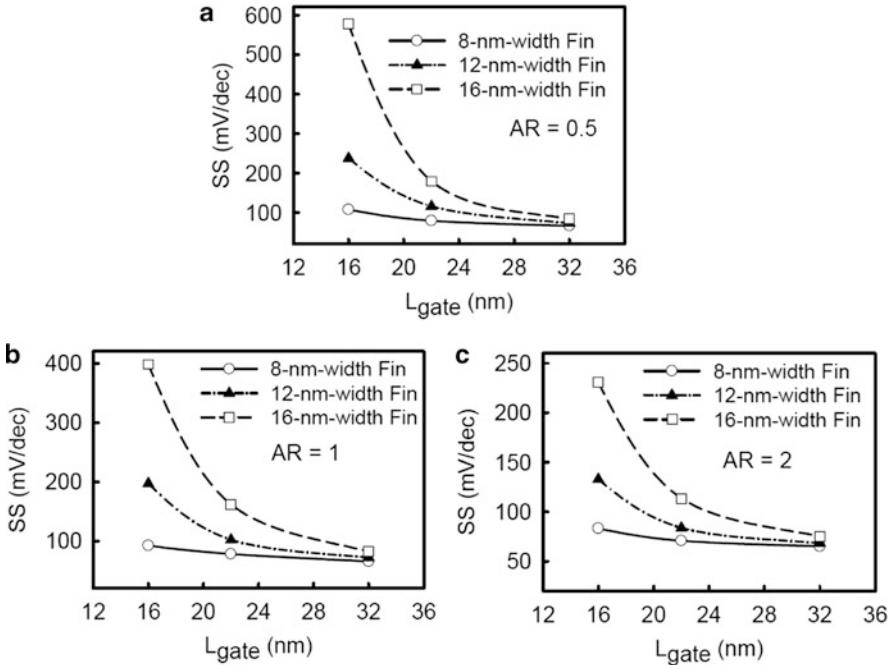


**Fig. 6.20** Plots of ratio of  $I_{on}/I_{off}$  versus the gate length for the triple-fin (a) quasi-planar, (b) tri-gate, and (c) FinFET structures

characteristics, the triple-fin FinFET shows 1.73 and 2.5 times smaller than those of tri-gate and quasi-planar devices, as shown in Fig. 6.21. Figure 6.22 shows the  $g_m = \partial I_D / \partial V_G$  of the explored devices which increases as the gate length scales and the increasing driving current. The thinner device exhibits a smaller  $g_m$  due to the smaller driving current. For devices with different AR, the  $g_m$  of triple-fin FinFET of 16-nm-gate-length and 8-nm-fin width shows 1.36 and 2.28 times larger than those of tri-gate and quasi-planar structures, as depicted in Fig. 6.22, according to the estimation of (6.8).

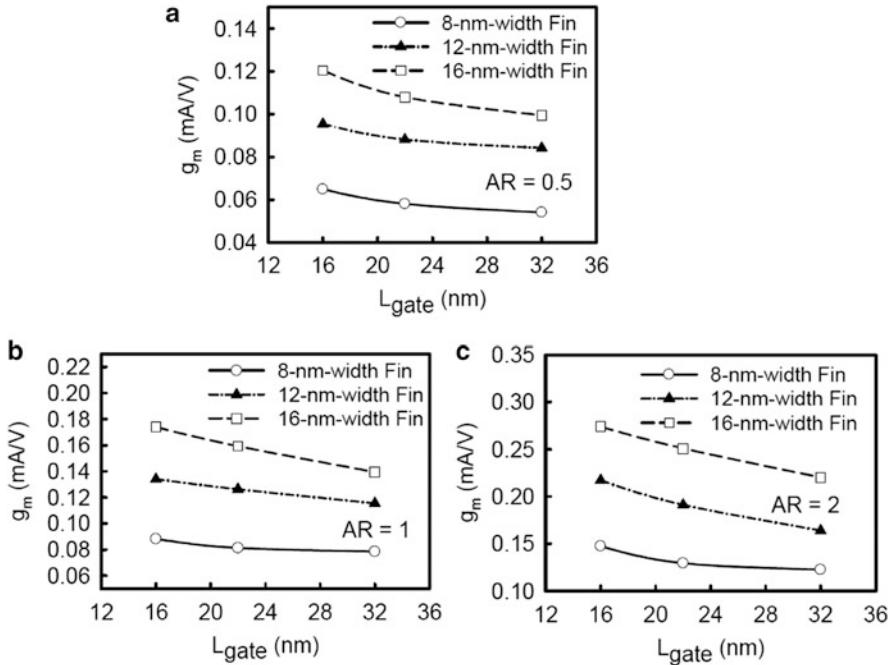
### 6.3.2 AC Characteristics of Triple-Fin Structure

The  $C_g$  is another important index for device characteristics. Figure 6.23 shows the  $C_g$  for the explored triple-fin structures, where the  $C_g$  with thicker fin width is larger than that of thinner fin width due to having more induced charges inside the channel region. The  $C_g$  of 16-nm-fin width structure is about two times larger than 8-nm-fin width one. For structures with different AR, the FinFETs possess the largest  $C_g$ , where the  $C_g$  of triple-fin FinFET is about 1.53 and 2.07 times larger than those of tri-gate and quasi-planar structures. The intrinsic gate delay and the cutoff



**Fig. 6.21** Plots of ratio of SS versus the gate length for the triple-fin (a) quasi-planar, (b) tri-gate, and (c) FinFET structures

frequency characteristics are further investigated, as shown in Figs. 6.24 and 6.25, to explore the effect of geometry aspect ratio on the 16-nm MOSFETs' AC characteristics. The definition of intrinsic gate delay is:  $\tau = C_g V_{DD} / I_{on}$  and the cutoff frequency  $F_T = g_m / 2\pi C_g$ . As the gate length scales, the device intrinsic gate delay is decreased due to the smaller  $C_g$  and the larger on-state current. Similarly, the cutoff frequency is increased due to smaller  $C_g$  and the larger  $g_m$ . Since the delay time and the cutoff frequency are dependent on the driving current and  $C_g$ , as the device-fin-width decreases, the delay time is increased and the cutoff frequency is decreased due to the significant decrease of driving current and  $g_m$ , as shown in Figs. 6.19 and 6.22. Comparing the structures with different AR, the FinFET exhibits a smaller delay time and higher cutoff frequency than those of tri-gate and quasi-planar structures because of its better driving capability. For devices with 16-nm-gate-length and 8-nm-fin width, the intrinsic gate delay of FinFET is 1.04 and 1.17 times smaller than those of tri-gate and quasi-planar structures, respectively. As for the cutoff frequency, the FinFET is 1.05 and 1.11 times larger than tri-gate and quasi-planar structures, respectively. However, we have to notice that in design of triple-fin structures, although the device with narrower channel width exhibits an even smaller  $V_{th}$  fluctuation because of the more uniform potential

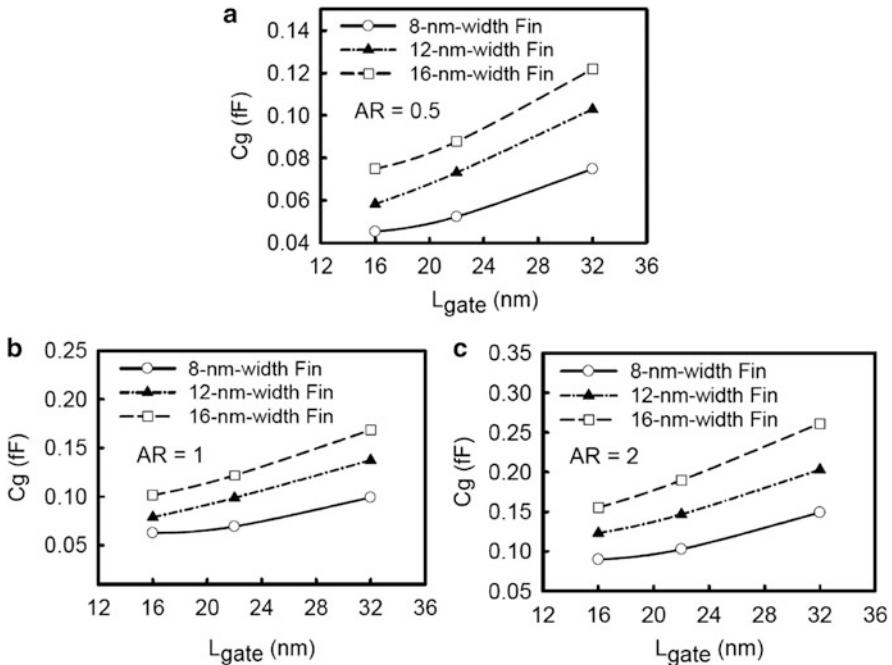


**Fig. 6.22** Plot of ratio of transconductance versus the gate length for the triple-fin (a) quasi-planar, (b) tri-gate, and (c) FinFET structures

distribution inside the device channel, the worse driving capability may degrade the intrinsic gate delay and cutoff frequency of MOSFETs. Moreover, the narrow-width effect will be enhanced and introduce another source of fluctuation in  $V_{th}$ .

## 6.4 Characteristic Fluctuation of FinFET Devices

From the studies above, we know the triple-fin FinFET shows the best performance than tri-gate and quasi-planar FETs. We now study the electrical characteristic fluctuation of FinFET devices induced by random sources. In this section, a coupled device-circuit simulation [37, 39–42] is performed to study the device and circuit characteristics of single- and triple-fin devices with different fin shapes (FinFET, tri-gate, and quasi-planar MOSFETs). The estimated electrical characteristics include threshold voltage, gate capacitance, delay time of the inverter, and static noise margin of a six-transistor static random access memory. Random-dopant-induced fluctuations in the aforementioned characteristics are further discussed with respect to the different ARs. The results of this study indicate that the structures with triple fins and a large AR may exhibit excellent characteristics

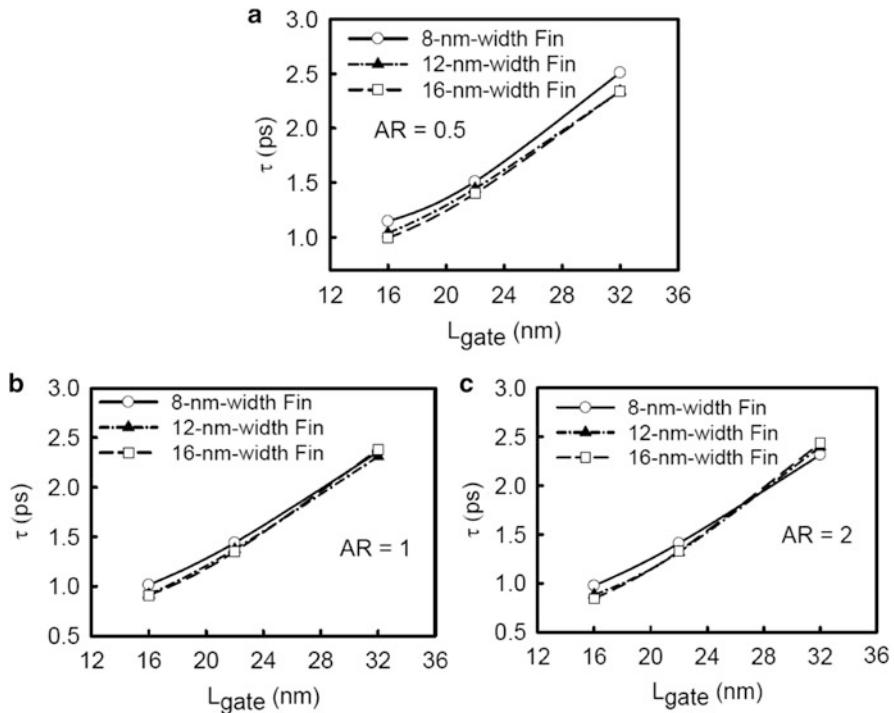


**Fig. 6.23** Plots of gate capacitance versus the gate length for the triple-fin (a) quasi-planar, (b) tri-gate, and (c) FinFET structures

and fluctuation suppression. The accuracy of the three-dimensional quantum mechanically corrected drift-diffusion device simulation performed was experimentally verified [37].

#### 6.4.1 Process Variation Effect

Figure 6.26a, b shows the gate-length-variation-induced  $V_{th}$  deviation ( $\Delta V_{th}$ ) for the single- and triple-fin structures with respect to different AR. The  $V_{th}$  deviation is defined as the difference of  $V_{th}$  between the nominal case (i.e., the 16-nm-gate length) and process variation altered cases (they could be 14.5-nm- or 17.5-nm-gate length), as shown in the inset of Fig. 6.26a. The results show that the  $V_{th}$  deviation of the triple-fin structure is significantly smaller than that of single-fin one. For example, the  $V_{th}$  variations of FinFET are 1.42 and 1.78 times smaller than those of tri-gate and quasi-planar FETs due to having larger gate capacitance. For FinFET structure, the  $V_{th}$  variations of triple-fin FinFET are much smaller than those of single-fin one, in which the three sigma of the gate-length-variation including the line-edge-roughness and the gate-length-deviation is about 1.5 nm according to the International Technology Roadmap for Semiconductors (ITRS) [38]. The normalized  $V_{th}$  variation for

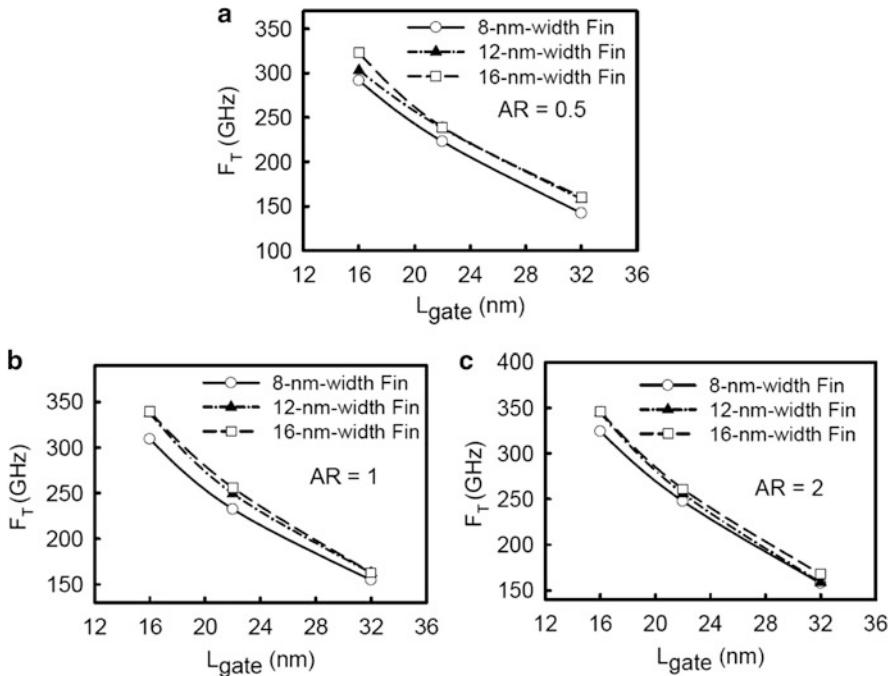


**Fig. 6.24** Plots of  $\tau$  versus the gate length for the triple-fin (a) quasi-planar, (b) tri-gate, and (c) FinFET structures

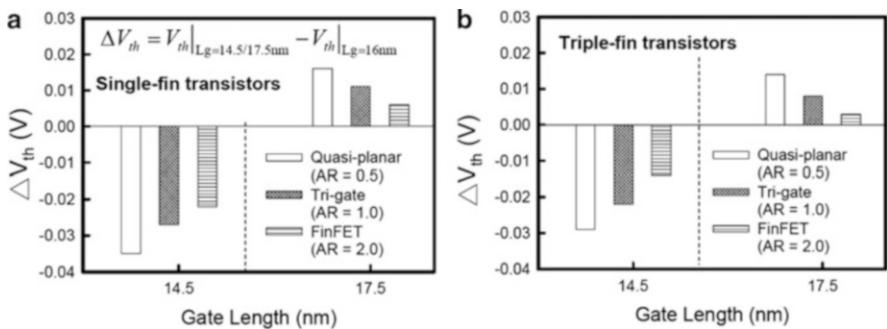
the single- and triple-fin FETs is summarized in Table 6.4, where the normalized  $V_{\text{th}}$  variation is defined as the difference of  $V_{\text{th}}$  between the 14.5-nm- and the 17.5-nm-gate FETs. From the results of Fig. 6.9, the FinFET can possess the smallest  $V_{\text{th}}$  variations among the explored devices due to a more uniform potential distribution and well channel controllability.

#### 6.4.2 Random Dopant Fluctuation in Digital Circuits

As aforementioned, the FinFETs exhibits excellent channel controllability and high resistance to intrinsic parameter variations. The random dopant-induced fluctuation will be investigated in this section. For example, the method to generate random dopant in triple-fin FinFET is presented. A total of 375 dopants are randomly generated in a  $80 \times 40 \times 80 \text{ nm}^3$  cube, yielding an equivalent doping concentration of  $1.48 \times 10^{18} \text{ cm}^{-3}$ , as shown in Fig. 6.27a. The  $80 \times 40 \times 80 \text{ nm}^3$  cube is then partitioned into 125 subcubes of  $16 \times 8 \times 16 \text{ nm}^3$ . The number of dopants in the subcubes varies from zero to nine with an average of three. The 125 subcubes



**Fig. 6.25** Plots of frequency  $F_T$  versus the gate length for the triple-fin (a) quasi-planar, (b) tri-gate, and (c) FinFET structures

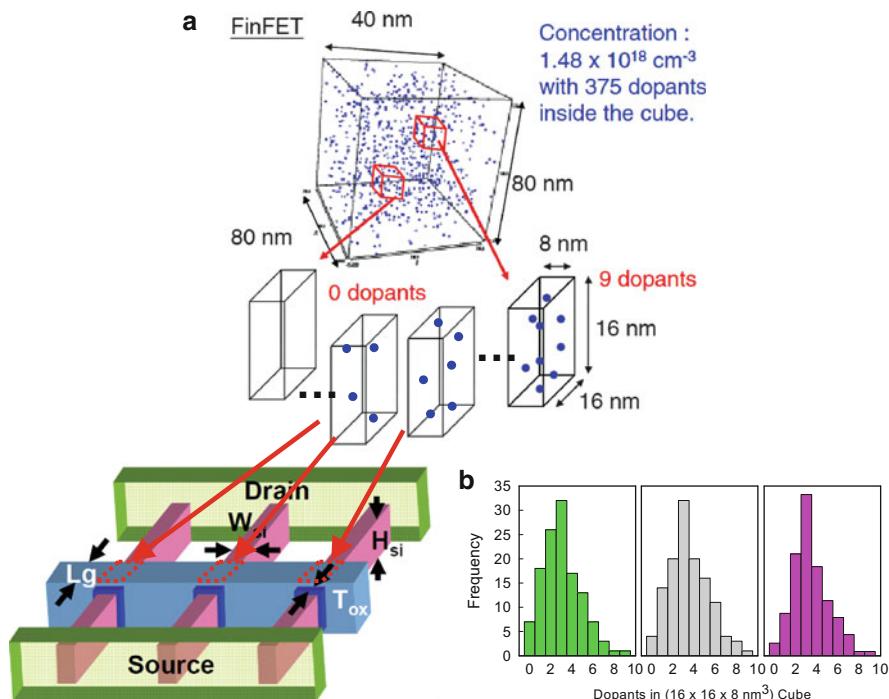


**Fig. 6.26** Plots of threshold voltage deviation of (a) single-fin and (b) triple-fin MOSFETs with the 14.5-nm- and the 17.5-nm-gate length FinFET, tri-gate, and quasi-planar structures, where the  $V_{\text{th}}$  deviation is defined as the  $V_{\text{th}}$  difference between the nominal  $V_{\text{th}}$  (it is 150 mV for the 16-nm-gate length) and  $V_{\text{th}}$  for the 14.5-nm- or 17.5-nm-gate length

are equivalently mapped into the channel region to simulate the sensitivity of the device to the position and number of dopants, where the distribution of each fin of triple-FinFET is independently shown in Fig. 6.27b. Similarly, the 125 subcubes of  $11.3 \times 11.3 \times 16$  and  $8 \times 16 \times 16 \text{ nm}^3$  are mapped into the channel region for

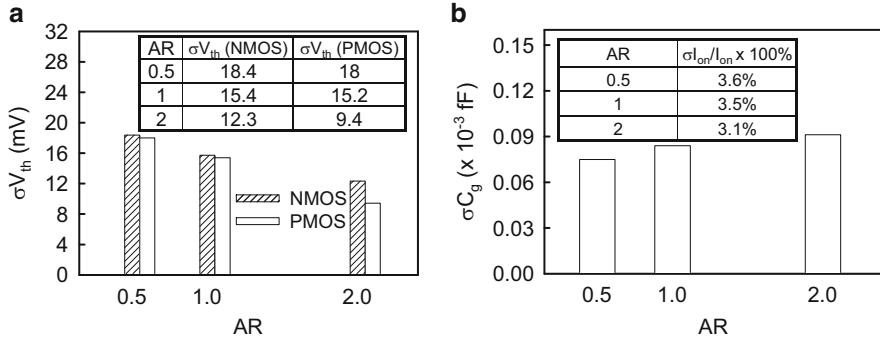
**Table 6.4** Summary of normalized  $V_{th}$  variation for single- and triple-fin quasi-planar, tri-gate, and FinFET structures, where the  $V_{th}$  variation is defined as the difference of  $V_{th}$  between the 14.5-nm- and 17.5-nm-gate lengths, and the  $V_{th}$  variation is then normalized with respect to the nominal  $V_{th}$  of 150 mV

	AR = 0.5	AR = 1	AR = 2
Normalized $V_{th}$ variation of single-fin structure	33.3 %	26.7 %	18.7 %
Normalized $V_{th}$ variation of triple-fin structure	27.3 %	20.0 %	11.3 %



**Fig. 6.27** (a) Illustration of the generated discrete models of FinFET, which follow a Gaussian distribution and range from 0 to 9, with an average of value 3, then mapped into channel region. (b) Plot of distribution of random dopant for each channel region of triple-fin FinFET

the tri-gate (AR = 1) and the quasi-planar (AR = 0.5) structures. Figures 6.15a and 6.15b, respectively, show plots of the high-to-low transition characteristics of both single- and triple-fin inverters at various ARs, with a power supply voltage of 1 V. The three solid lines represent the output signals of the devices with different fin structures and the dotted line represents the input signal. The high-to-low delay time ( $t_{HL}$ ) is defined as the time difference between 50 % points of the input and output signals during the falling of the output signals. The insets in Figs. 6.15a and 6.15b, respectively, show plots of the high-to-low delay times ( $t_{HL}$ ) of the studied single- and triple-fin structures, which are affected by the shape of the fins.



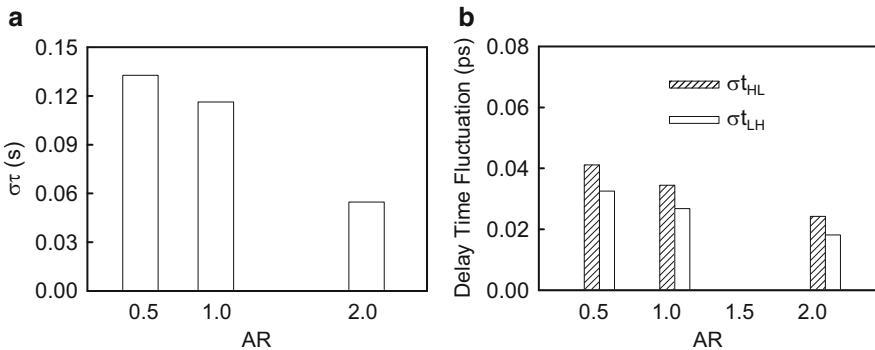
**Fig. 6.28** (a)  $\sigma V_{th}$  and (b)  $\sigma C_g$  induced by random dopants versus AR for the triple-fin structure. The two insets show the summarized  $\sigma V_{th}$  and normalized on-state current fluctuation ( $= \sigma I_{on}/I_{on} \times 100\%$ )

As expected, both single- and triple-fin FinFET inverters have the smallest  $t_{HL}$  for various ARs, indicating the advantages afforded by FinFET in terms of both DC and dynamic characteristics. Although the gate capacitance of the triple-fin structure is larger than that of the single-fin structure, it provides a smaller transition delay because the increase in drive current is larger. Figure 6.28 shows the random-dopant-induced threshold voltage fluctuation ( $\sigma V_{th}$ ), and the gate capacitance fluctuation ( $\sigma C_g$ ) of the studied triple-fin structures.  $\sigma V_{th}$  is derived as

$$\sigma V_{th} = \frac{q}{C_{ox}} \sqrt{\frac{N_a W_{dm}}{3LW}}, \quad (6.10)$$

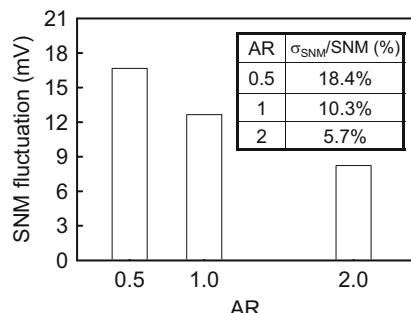
where  $W_{dm}$  denotes the maximum depletion width,  $N_a$  denotes the background doping concentration,  $L$  and  $W$  are the gate length and width, respectively, and  $C_{ox}$  is the oxide capacitance.

Because  $\sigma V_{th}$  is proportional to depletion width, the  $V_{th}$  of p-type MOSFET is lower than that of n-type MOSFET because the depletion depth is small. The  $\sigma V_{th}$  values of n- and p-type FinFETs are 1.5 and 1.9 times, respectively, smaller than that of quasi-planar structures, as shown in Fig. 6.28a, suggesting that, for the same channel volume, FinFET has a more uniform surface potential. The  $\sigma C_g$  of triple-fin FinFETs is slightly higher than triple-fin tri-gate and quasi-planar MOSFETs because its gate area is larger, as shown in Fig. 6.28b; the inset in Fig. 6.28b shows a plot of the normalized on-state current fluctuation ( $\sigma I_{on}/I_{on} \times 100\%$ ). Although the  $\sigma C_g$  of the triple-fin FinFETs is slightly higher than triple-fin tri-gate and quasi-planar MOSFETs, the large on-state current reduces the  $\sigma t$  of triple-fin FinFETs, as shown in Fig. 6.29a. Figure 6.29b shows the  $\sigma t_{HL}$  and  $\sigma t_{LH}$  of triple-fin device inverters.  $\sigma t_{HL}$  and  $\sigma t_{LH}$  are dominated by n-type MOSFET and p-type MOSFET, respectively. Thus,  $\sigma t_{HL}$  exceeds  $\sigma t_{LH}$  because n-type MOSFETs have a large  $\sigma V_{th}$ . Figure 6.14b shows a plot of the SNM of the triple-fin device SRAM cells, where cell ratio and pull-up ratio are assumed to be unity in this determination.



**Fig. 6.29** (a)  $\tau$  of triple-fin structures. (b)  $\sigma t_{HL}$  and  $\sigma t_{LH}$  values of the tested inverter with AR = 2, 1, and 0.5

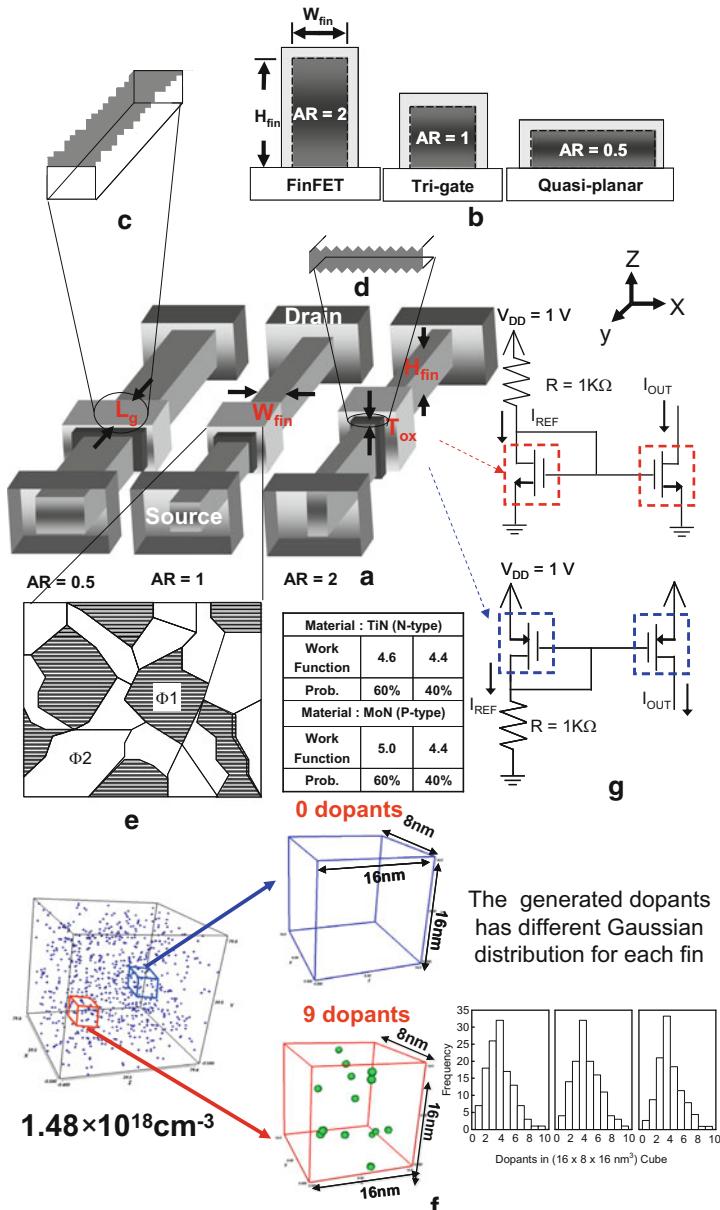
**Fig. 6.30** Plot of SNM fluctuations of the examined SRAM with AR = 2, 1, and 0.5. The normalized  $\sigma_{SNM}\%$  ( $= \sigma_{SNM}/SNM \times 100\%$ ) is summarized in the *inset*



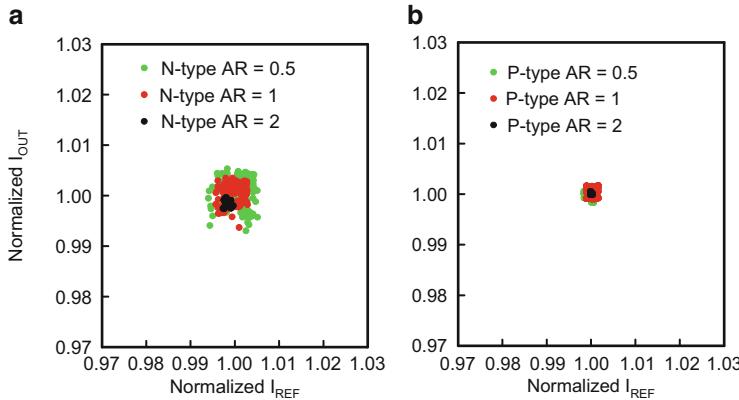
The relation between the device transconductance and SNM of SRAM could be expressed as (6.9). The calculated transconductances for AR = 0.5, 1, and 2 are 0.0284, 0.0536, and 0.0752 mA/V, respectively. Consequently, among the explored three structures, FinFET has the largest SNM owing to having the largest transconductance, as shown in Fig. 6.30a. Figure 6.30b shows the random-dopant-induced SNM fluctuation ( $\sigma_{SNM}$ ) of the triple-fin device SRAM cells. Triple-fin FinFETs have the smallest  $\sigma_{SNM}$  because they have the smallest  $\sigma V_{th}$ . The table in the inset of Fig. 6.30b shows the normalized  $\sigma_{SNM}$ .

#### 6.4.3 Intrinsic Parameter Fluctuation in Current Mirror Circuit

Figure 6.31a shows the structure of studied 3D transistors with different AR of channel fin, where the gate oxide is 1.2 nm and the channel doping concentration is  $1.48 \times 10^{18} \text{ cm}^{-3}$ . The definition of fin shape is the ratio of fin height and fin width



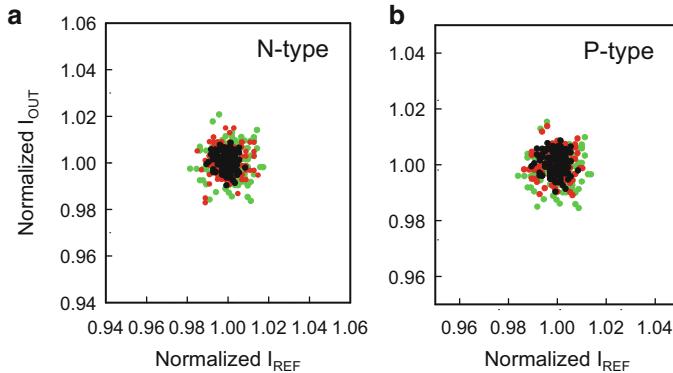
**Fig. 6.31** (a) Schematic diagram of transistor with different AR. (b) Three fin types: quasi-planar (AR = 0.5), tri-gate (AR = 1), and FinFET (AR = 2). Variation sources of metal/high- $\kappa$  gate stack COMS: (c) process variation effect (PVE), (d) oxide thickness fluctuation (OTF), (e) work-function fluctuation (WKF), and (f) random dopant fluctuation (RDF) are considered, where the control devices are TiN/HfSiON gate stack with 1.2-nm oxide thickness. (g) The explored n-type and p-type current mirrors of analog circuit with the variation sources which induced current ( $I_{REF}/I_{OUT}$ ) mismatch



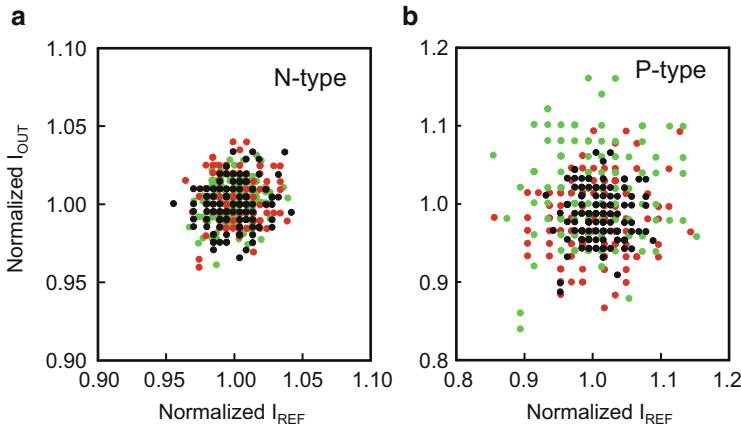
**Fig. 6.32** (a) The plot of normalized  $I_{OUT}$  versus  $I_{REF}$  in (a) n-type and (b) p-type current mirror with AR = 0.5, 1, and 2, respectively, where the fluctuation is induced by oxide thickness. The green, red, and black solid circles indicate device with various ARs, respectively

which are FinFETs (AR = 2), tri-gate (AR = 1), and quasi-planar (AR = 0.5) FETs, as shown in Fig. 6.31b, respectively. The different AR of fin reflects the different charge-control-capability and affects design of FETs with vertical channels. To estimate the PVE (Fig. 6.31c) and OTF (Fig. 6.31d), the variation is Gaussian distribution. To investigate the WKF (Fig. 6.31e) of explored current mirror (Fig. 6.31g) with different fin shape transistor, we consider the metal grain size and device gate area using a statistical-sound Monte Carlo approach based on metal property. According to different AR of fin, the cube is partitioned into 125 sub-cubes which are then equivalently mapped into the 3D device channel region for fluctuation analysis. For example, we take FinFET fin shape transistor; the number of dopants inside the channel is ranged from 0 to 9, as shown in Fig. 6.31f, where the volume of channel fin is  $16 \times 8 \times 16$  (nm)<sup>3</sup>. The analysis of fluctuation induced by OTF, PVE, WKF, and RDF in current mirror circuit is presented. And, the simulation results are compared under different AR of fin. According to our previous work [3], the FinFET fin shape offers better performance due to having larger driving current and more uniform potential in channel. In order to explain the effect of gate oxide thickness on the performance of n- and p-type current mirror, the computed variations of  $I_{OUT}$  and  $I_{REF}$  are plotted as shown in Fig. 6.32a, b. The results show that the current mismatch of n-type current mirror is larger than that of p-type due to higher mobility of electron. It is known that the current mismatch can be decreased by increasing the AR. As the gate length deviation is considered, the current mismatch of n-type current mirror is slightly larger fluctuation than that of p-type, as shown in Fig. 6.33a, b.

Compared with OTF, the PVE-induced current mismatch is larger due to larger fluctuation source. Due to being comparable with the grain size of the metal gate electrode, WKF effects on device characteristic are interesting studies for device. The gate material is TiN in n-type current mirror, whose grain orientations are



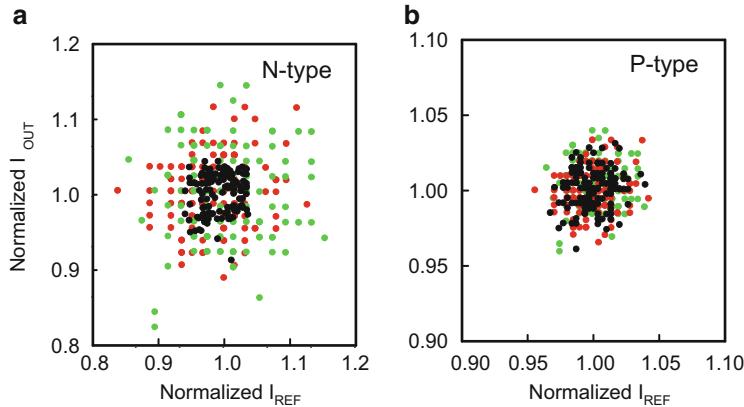
**Fig. 6.33** The plot of normalized  $I_{OUT}$  versus  $I_{REF}$  in (a) n-type and (b) p-type current mirror with AR = 0.5, 1, and 2, respectively, where the fluctuation is induced by process variation effect



**Fig. 6.34** The plot of normalized  $I_{OUT}$  versus  $I_{REF}$  in (a) n-type and (b) p-type current mirror with AR = 0.5, 1, and 2, respectively, where the fluctuation is induced by work-function fluctuation

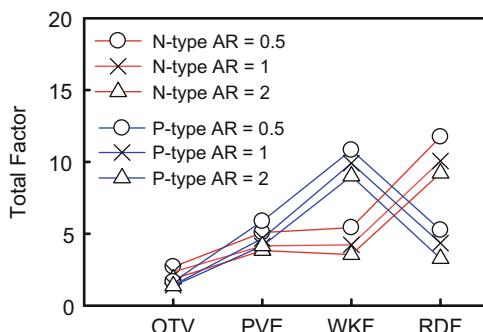
<200> and <111> and the associated work functions are 4.6 eV and 4.4 eV, respectively. The possibility of grain orientations for <200> and <111> are 60 % and 40 %, respectively, as shown in the table of Fig. 6.31e. Similarly, the probability of grain orientations and associated work function is summarized in the table.

Figure 6.34a, b shows that the work-function fluctuation resulting in current mismatch in p-type current mirror is quite larger than that in n-type one due to larger deviation of grain size and work function. There are 758 dopants are first randomly generated in a cube of side 80 nm, in which the equivalent doping concentration is  $1.48 \times 10^{18} \text{ cm}^{-3}$ , as shown in Fig. 6.31a. The random dopant effect in the 16-nm-gate CMOS circuits is significant. The 378 dopants in  $80 \times 40 \times 80 (\text{nm})^3$  cube is partitioned into  $16 \times 8 \times 16 (\text{nm})^3$  sub-cubes and



**Fig. 6.35** The plot of normalized  $I_{\text{OUT}}$  versus  $I_{\text{REF}}$  in (a) n-type and (b) p-type current mirror with  $\text{AR} = 0.5, 1$ , and  $2$ , respectively, where the fluctuation is induced by random dopant fluctuation

**Fig. 6.36** The summary of current mismatch induced by OTV, PVE, WKF, and RDF in n-type and p-type current mirror circuit with various ARs



then equivalently mapped into the channel region of the device channel for current mirror circuit, as show in Fig. 6.31f. The results show that the RDF dominates current mismatch in n-type current mirror and is reduced by increasing AR, as shown in Fig. 6.35. All the results are summarized in Fig. 6.36, where the current factor is calculated by  $(|I_{\text{REF}}/I_{\text{nominal}}|^2 + |I_{\text{OUT}}/I_{\text{nominal}}|^2)^{0.5}/2$ .

## 6.5 Conclusions

In this chapter, a coupled device-circuit simulation has been performed to study the device and circuit characteristics of single- and triple-fin devices with different shapes of fins (FinFET, tri-gate, and quasi-planar MOSFETs). The accuracy of the three-dimensional quantum-mechanically corrected drift-diffusion device simulation

performed was experimentally verified. The estimated electrical characteristics include threshold voltage, gate capacitance, the delay time of the inverter, and the static noise margin of a 6T SRAM. The results of this study indicate that the triple-fin FinFET ( $AR = 2$ ) has better channel controllability than the triple-fin tri-gate ( $AR = 1$ ) and triple-fin quasi-planar ( $AR = 0.5$ ) FETs. A six-transistor static random access memory using triple-fin FinFETs also provides the largest static noise margin because it supports the highest transconductance in FinFETs. Although FinFETs have a large effective fin width and driving current, their large gate capacitance limits gate delay. The transient characteristics of an inverter with triple-fin transistors have further been examined and compared with those of an inverter with single-fin transistors. The triple-fin inverter has a shorter delay because it is dominated by the driving current of the transistor. With respect to random-dopant-induced fluctuations, the triple-fin FinFET suppresses not only the surface potential but also its variation because it has a more uniform surface potential than the triple-fin tri-gate and quasi-planar FET, and so the effects of random dopants on the circuits are attenuated. In summary, the triple-fin structure has relatively smaller  $R_{out}$ , SS, DIBL,  $t_{HL}$ , and larger  $I_{on}$ ,  $g_m$ , SNM, compared with single-fin structures, especially for FinFET structure. And for 8 nm  $W_{fin}$  FinFET, it has larger  $I_{on}/I_{off}$  and smaller SS and  $C_g$ , compared with that of thicker  $W_{fin}$ . Therefore, the  $I_{on}$  and  $g_m$  of triple-fin structure can be enhanced by thinner fin design. The characteristics of current mirror circuit with different fin shape have been explored. For 16-nm-gate n- and p-type current mirror, the fluctuations dominated by RDF and WKF are successfully suppressed by increasing the AR, respectively.

**Acknowledgment** This work was supported in part by the Taiwan National Science Council (NSC), under Contracts NSC-101-2221-E-009-092, NSC-100-2221-E-009-018, and NSC-99-2221-E-009-175, and by Taiwan Semiconductor Manufacturing Company, Hsinchu, Taiwan, under a 2011-2013 grant.

## References

1. James, D.: Intel Ivy Bridge unveiled – the first commercial tri-gate, high-k, metal-gate CPU. In: IEEE international custom integrated circuits conference, pp. 1–4 (2012)
2. Damaraju, S., George, V., Jahagirdar, S., Khondker, Milstrey, T.R., Sarkar, S., Siers, S., Stolero, I., Subbiah, A.: A 22nm IA multi-CPU and GPU system-on-chip. In: IEEE international solid-state circuits conference digest of technical papers, pp. 56–57 (2012)
3. Auth, C.: 22-nm fully-depleted tri-gate CMOS transistors. In: IEEE international custom integrated circuits conference, pp. 1–6 (2012)
4. Karl, E., Yih, W., Ng, Y.-G., Guo, Z., Hamzaoglu, F., Bhattacharya, U., Zhang, K., Mistry, K., Bohr, M.: A 4.6 GHZ 162Mb SRAM design in 22nm tri-gate CMOS technology with integrated active VMIN-enhancing assist circuitry. In: IEEE international solid-state circuits conference digest of technical papers, pp. 230–232 (2012)
5. Kulkarni, J., Geuskens, B., Karnik, T., Khellah, M., Tschanz, J., De, V.: Capacitive-coupling wordline boosting with self-induced VCC collapse for write VMIN reduction in 22-nm 8T SRAM. In: IEEE international solid-state circuits conference digest of technical papers, pp. 234–236 (2012)

6. Seifert, N., Gill, B., Jahinuzzaman, S., Basile, J., Ambrose, V., Shi, Q., Allmon, R., Bramnik, A.: Soft error susceptibilities of 22 nm tri-gate devices. *IEEE Trans. Nucl. Sci.* **59**, 2666–2673 (2012)
7. Auth, C., Allen, C., Blattner, A., Bergstrom, D., Brazier, M., Bost, M., Buehler, M., Chikarmane, V., Ghani, T., Glassman, T., Grover, R., Han, W., Hanken, D., Hattendorf, M., Hentges, P., Heussner, R., Hicks, J., Ingerly, D., Jain, P., Jaloviar, S., James, R., Jones, D., Jopling, J., Joshi, S., Kenyon, C., Liu, H., McFadden, R., McIntyre, B., Neirynck, J., Parker, C., Pipes, L., Post, I., Pradhan, S., Prince, M., Ramey, S., Reynolds, T., Roesler, J., Sandford, J., Seiple, J., Smith, P., Thomas, C., Towner, D., Troeger, T., Weber, C., Yashar, P., Zawadzki, K., Mistry, K.: A 22nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors. In: *Symposium on VLSI technology. Digest of technical papers*, pp. 131–132 (2012)
8. Kuhn, K.J., Giles, M.D., Becher, D., Kolar, P., Kornfeld, A., Kotlyar, R., Ma, S., Maheshwari, T.A., Mudanai, S.: Process technology variation. *IEEE Trans. Electron Dev.* **58**, 2197–2208 (2011)
9. Li, Y., Hwang, C.-H., Li, T.-Y.: Random-dopant-induced variability in nano-CMOS devices and digital circuits. *IEEE Trans. Electron Dev.* **56**, 1588–1597 (2009)
10. Li, Y., Hwang, C.-H., Li, T.-Y., Han, M.-H.: Process-variation effect, metal-gate work-function fluctuation, and random-dopant fluctuation in emerging CMOS technologies. *IEEE Trans. Electron Dev.* **57**, 437–447 (2010)
11. Baravelli, E., Jurczak, M., Speciale, N., De Meyer, K., Dixit, A.: Impact of LER and random dopant fluctuations on FinFET matching performance. *IEEE Trans. Nanotechnol.* **7**, 291–298 (2008)
12. Weber, O., Faynot, O., Andrieu, F., Buj-Dufournet, C., Allain, F., Scheiblin, P., Foucher, J., Daval, N., Lafond, D., Tosti, L., Brevard, L., Rozeau, O., Fenouillet-Beranger, C., Marin, M., Boeuf, F., Delprat, D., Bourdelle, K., Nguyen, B.Y., Deleonibus, S.: High immunity to threshold voltage variability in undoped ultra-thin FDSoI MOSFETs and its physical understanding. In: *International electron device meeting. Technical digest*, pp. 245–248 (2008)
13. Li, Y., Hwang, C.-H., Han, M.-H.: Simulation of characteristic variation in 16 nm gate FinFET devices due to intrinsic parameter fluctuations. *Nanotechnology* **21**, 095203 (2010)
14. Li, Y., Hwang, C.-H., Li, T.-Y.: Discrete-dopant-induced timing fluctuation and suppression in nanoscale CMOS circuit. *IEEE Trans. Circ. Syst. II Express Briefs* **56**, 379–383 (2009)
15. Ancona, M.G., Tiersten, H.F.: Macroscopic physics of the silicon inversion layer. *Phys. Rev. B* **35**, 7959–7965 (1987)
16. Odanaka, S.: Multidimensional discretization of the stationary quantum drift-diffusion model for ultrasmall MOSFET structures. *IEEE Trans. Comput. Aided Des. Integr. Circ. Sys.* **23**, 837–842 (2004)
17. Tang, T.-W., Wang, X., Li, Y.: Discretization scheme for the density-gradient equation and effect of boundary conditions. *J. Comput. Electron.* **1**, 389–393 (2002)
18. Li, Y., Sze, S.M., Chao, T.S.: A practical implementation of parallel dynamic load balancing for adaptive computing in VLSI device simulation. *Eng. Comput.* **18**, 124–137 (2002)
19. Cheng, B., Roy, S., Asenov, A.: Impact of intrinsic parameter fluctuations on SRAM cell design. In: *Proceedings of international solid-state and integrated circuit technology conference*, pp. 1290–1292 (2006)
20. Cheng, B., Roy, S., Asenov, A.: The scalability of 8T-SRAM cells under the influence of intrinsic parameter fluctuations. In: *Proceedings of European solid-state circuits conference*, pp. 93–96 (2007)
21. Cheng, B., Roy, S., Asenov, A.: CMOS 6T SRAM cell design subject to “atomistic” fluctuations. *Solid State Electron.* **51**, 565–571 (2007)
22. Bhavnagarwala, A.J., Tang, X., Meindl, J.D.: The impact of intrinsic device fluctuations on CMOS SRAM cell stability. *IEEE J. Solid State Circ.* **36**, 658–665 (2001)
23. Hill, C.: Definitions of noise margin in logic systems. *Mullard Tech. Commun.* **89**, 239–245 (1967)

24. Li, Y., Liu, J.-L., Chao, T.-S., Sze, S.M.: A new parallel adaptive finite volume method for the numerical simulation of semiconductor devices. *Comput. Phys. Commun.* **142**, 285–289 (2001)
25. Kim, S.M., Yoon, E.J., Jo, H.J., Li, M., Oh, C.W., Lee, S.Y., Yeo, K.H., Kim, M.S., Kim, S.H., Choe, D.U., Choe, J.D., Suk, S.D., Kim, D., Park, D., Kim, K., Ryu, B.: A novel multi-channel field effect transistor (McFET) on bulk Si for high performance Sub-80 nm application. In: International electron device meeting. Technical digest, pp. 639–642 (2004)
26. Bohr, M.: The evolution of scaling from the homogeneous era to the heterogeneous era. In: International electron device meeting. Technical digest, pp. 1–6 (2011)
27. Li, Y., Hwang, C.-H.: Discrete-dopant-induced characteristic fluctuations in 16 nm multiple-gate silicon-on-insulator devices. *J. Appl. Phys.* **102**, 084509 (2007)
28. Li, Y., Hwang, C.-H.: Effect of fin angle on electrical characteristics of nanoscale round-top-gate bulk FinFETs. *IEEE Trans. Electron Dev.* **54**, 3426–3429 (2007)
29. Li, Y., Yu, S.-M., Lee, J.-W.: Quantum mechanical corrected simulation program with integrated circuit emphasis model for simulation of ultrathin oxide metal-oxide-semiconductor field effect transistor gate tunneling current. *Jpn. J. Appl. Phys.* **44**, 2132–2136 (2005)
30. Li, Y., Hwang, C.-H.: High-frequency characteristic fluctuations of nano-MOSFET circuit induced by random dopants. *IEEE Trans. Microw. Theor. Tech.* **56**, 2726–2733 (2008)
31. Grasser, T., Selberherr, S.: Mixed-mode device simulation. *Microelectron. J.* **31**, 873–881 (2000)
32. Li, Y., Yu, S.-M.: A parallel adaptive finite volume method for nanoscale double-gate MOSFETs simulation. *J. Comput. Appl. Math.* **175**, 87–99 (2005)
33. Seevinck, E., List, F.J., Lohstroh, J.: Static-noise margin analysis of MOS SRAM cells. *IEEE J. Solid State Circ.* **22**, 748–754 (1987)
34. Frank, D.J., Laux, S.E., Fischetti, M.V.: Monte Carlo simulation of a 30 nm dual-gate MOSFET: how short can Si go. In: International electron device meeting. Technical digest, pp. 553–556 (1992)
35. Yang, F.L., Lee, D.H., Chen, H.Y., Chang, C.Y., Liu, S.D., Huang, C.C., Chung, T.X., Chen, H.W., Huang, C.C., Liu, Y.H., Wu, C.C., Chen, C.C., Chen, S.C., Chen, Y.T., Chen, Y.H., Chen, C.J., Chan, B.W., Hsu, P.F., Shieh, J.H., Tao, H.J., Yeo, Y.C., Li, Y., Lee, J., Chen, W.P., Liang, M.S., Hu, C.: 5nm-gate nanowire FinFET. In: Symposium on VLSI technology. Digest of technical papers, pp. 196–197 (2004)
36. Li, Y., Chou, H.-M., Lee, J.-W.: Investigation of electrical characteristics on surrounding-gate and omega-shaped-gate nanowire Fin-FETs. *IEEE Trans. Nanotechnol.* **4**, 510–516 (2005)
37. Li, Y., Yu, S.-M., Hwang, J.-R., Yang, F.-L.: Discrete dopant fluctuated 20nm/15nm-gate planar CMOS. *IEEE Trans. Electron Dev.* **55**, 1449–1455 (2008)
38. International Technology Roadmap for Semiconductors (ITRS) 2007 Edition
39. Li, Y., Hwang, C.-H., Huang, H.-M.: Large-scale atomistic approach to discrete-dopant-induced characteristic fluctuations in silicon nanowire transistors. *Phys. Status Solidi A Appl. Mater. Sci.* **205**, 1505–1510 (2008)
40. Li, Y., Yu, S.-M.: Comparison of random-dopant-induced threshold voltage fluctuation in nanoscale single-, double-, and surrounding-gate field-effect transistors. *Jpn. J. Appl. Phys.* **45**, 6860–6865 (2006)
41. Li, Y., Hwang, C.-H., Cheng, H.-W.: Discrete-dopant-fluctuated transient behavior and variability suppression in 16-nm-gate complementary metal-oxide-semiconductor field-effect transistors. *Jpn. J. Appl. Phys.* **48**, 04C051 (2009)
42. Li, Y., Hwang, C.-H.: Discrete-dopant-fluctuated threshold voltage roll-off in sub-16nm bulk FinFETs. *Jpn. J. Appl. Phys.* **47**, 2580–2584 (2008)
43. Endo, K., Ishikawa, Y., Liu, Y., Masahara, M., Matsukawa, T., O'uchi, S., Ishii, K., Yamauchi, H., Tsukada, J., Suzuki, E.: Experimental evaluation of effects of channel doping on characteristics of FinFETs. *Electron Dev. Lett.* **28**, 1123–1125 (2007)
44. Matsukawa, T., Endo, K., Ishikawa, Y., Yamauchi, H., O'uchi, S., Liu, Y., Tsukada, J., Ishii, K., Sakamoto, K., Suzuki, E., Masahara, M.: Fluctuation analysis of parasitic resistance in FinFETs with scaled fin thickness. *Electron Dev. Lett.* **30**, 407–409 (2009)

# Chapter 7

## Variability in Nanoscale FinFET Technologies

Greg Leung and Chi On Chui

**Abstract** Stochastic process variability is a major obstacle to scaling field-effect transistor (FET) dimensions toward the nanometer regime. The impacts of line edge roughness (LER), random dopant fluctuation (RDF), oxide thickness fluctuation (OTF), and work function variation (WFV) can significantly affect the performance of individual transistors through random variations in device metrics such as threshold voltage ( $V_T$ ), on-state drive current ( $I_{on}$ ), off-state leakage current ( $I_{off}$ ), subthreshold swing (SS), and drain-induced barrier lowering (DIBL). In this chapter, we provide an introductory overview to process variability in modern fabrication and investigate in detail how LER and RDF affect inversion-mode (IM) and junctionless (JL) FinFET performance variability for sub-32nm technology nodes. The insights gained in this study will enable us to understand how different variability mechanisms affect the fundamental operation of IM- and JL-FinFETs, and also allow us to evaluate the viability of JL versus IM technology to meet future design challenges in the nanoscale era.

### 7.1 Introduction

Continued scaling of modern field-effect transistor (FET) dimensions from 10  $\mu\text{m}$  feature sizes in 1970 to 22nm in 2012 has fueled the exponential growth of the semiconductor industry over the past four decades. A reduction in transistor size

---

G. Leung (✉)

Department of Electrical Engineering, University of California, Los Angeles, 68-118  
Engineering IV Building, Box 951594, Los Angeles, CA 90095-1594, USA  
e-mail: [ggleung1@ucla.edu](mailto:ggleung1@ucla.edu)

C.O. Chui

Department of Electrical Engineering, University of California, Los Angeles, 6730B Boelter  
Hall, Box 951594, Los Angeles, CA 90095-1594, USA  
e-mail: [chui@ee.ucla.edu](mailto:chui@ee.ucla.edu)

increases device density, enables faster switching speeds, and reduces energy consumption for the same wafer cost and is the primary motivator for technology scaling. As individual devices are shrunk toward the nanometer regime and beyond, however, the impact of random processing variations becomes more and more significant to the point where device variability now plays a major obstacle to further scaling. The reason for this is that processing variations typically do not scale at the same rate as feature sizes do, so that any random fluctuations in transistor size/shape/composition comprise a larger fraction of the original design. Since most fluctuations are typically random in nature, a circuit designer will not be able to predict the exact performance of any given device within the circuit, and he/she will have to make larger allowances in the design margin to account for these variations, especially at finer technologies. Thus, proper assessment of variability effects in deeply scaled technologies will play a critical role in determining whether the demands of future generations can be met using current state-of-the-art FET designs and fabrication tools.

The effects of *process variability* have been studied extensively for planar metal-oxide-semiconductor field-effect transistors (MOSFETs), but only recently have variability effects been investigated for nonplanar devices such as multi-gate Fin field-effect transistors (FinFETs). These sources of variability can be manifested in the forms of *line edge roughness* (LER), *random dopant fluctuation* (RDF), *oxide thickness fluctuation* (OTF), *work function variation* (WFV), and others. Due to the shared process flows between traditional complementary metal-oxide-semiconductor (CMOS) technology and FinFET technology, FinFETs are expected to be susceptible to the same forms of variability as planar FETs, albeit with some differences which are related to the specific structure and design of nonplanar FinFETs. Moreover, some types of FinFETs may inherently be more vulnerable to LER, RDF, etc. than other types (again, depending on the device design)—this will be explored later in the chapter when we compare inversion-mode (IM) and junctionless (JL) FinFETs.

In this chapter, we will explore the effects of process variations on the variability of nanoscale FinFETs implemented in IM and JL varieties. In Sect. 7.2, we provide an introduction to some of the most common process variability mechanisms encountered today, including LER, RDF, OTV, and WFV, and explain their physical origins and various modeling approaches for use in statistical variability studies. In Sect. 7.3, we specifically investigate the impacts of LER and RDF in sub-32nm FinFET technologies (both IM- and JL-based) using technology computer-aided design (TCAD) simulations in order to understand and assess the vulnerability of these different FinFET technologies to random processing variations. Finally, we summarize our main conclusions regarding the significance that process and device variability effects can have in determining the feasibility of pursuing FinFET technology in the nanoscale era.

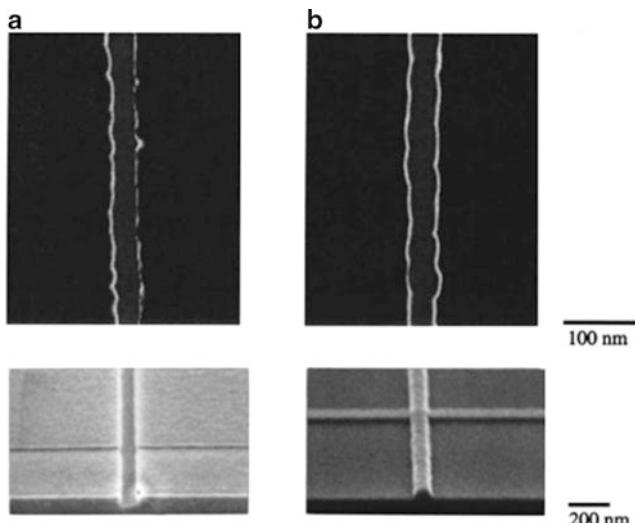
## 7.2 Process Variability Mechanisms

### 7.2.1 Line Edge Roughness

When a printed feature contains sidewalls that are not perfectly smooth after the fabrication process, it is said to exhibit *line edge roughness* or LER. An example of this is illustrated in Fig. 7.1 for two different types of photoresist. Today, LER is one of the most prevalent forms of process variation and is a major concern for device designers because it directly alters the size and shape of important active regions, such as the metallurgical gate length  $L_g$  and transistor width  $W$  in planar transistors. In nonplanar transistors, LER can also result in random fluctuation of the body thickness, which can result in a significant contribution to total device variability in deeply scaled technologies.

#### 7.2.1.1 Origins of LER

While there are many factors that can contribute to LER, the primary origins can be traced back to the lithography and etching processes during fabrication. During the lithography process, factors such as exposure contrast, shot noise, mask roughness, and resist material properties can directly affect the quality of resist patterns.



**Fig. 7.1** (Top) Atomic force microscope and (bottom) scanning electron microscope images of (a) ZEP-520 and (b) SAL-601 resist patterns exhibiting LER along the sidewalls [1]. Copyright © 1998 American Vacuum Society

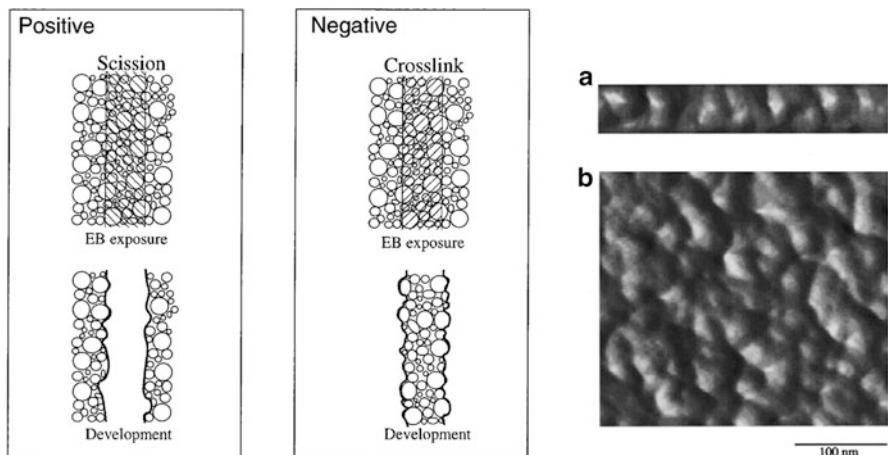
During the etching process, edge roughness on the resist sidewalls is partially transferred to the etched feature, and the etching process may induce additional roughness from random kinetics of the etching chemistry. In this section, we will briefly review and discuss some of these factors.

In real world exposure systems, even unexposed photoresist areas receive a finite exposure dose resulting from optical diffraction effects, thereby reducing the aerial image contrast ( $AIC < 100\%$ ). Shin et al. [2] investigated the correlation between AIC and LER produced by X-ray, extreme ultraviolet (EUV), and e-beam exposure on UV6 and poly(methyl methacrylate) (PMMA) photoresists, and determined that AIC values below 50% resulted in progressively higher LER. Above 50% AIC, however, there was no significant correlation between AIC and LER. Sanchez et al. [3] performed interferometric lithography on a variety of resists and reached a similar conclusion that higher AIC resulted in lower LER. They also observed that material properties of the resist such as polymer weight had influenced the dependence of LER on AIC, with heavier resists producing even worse LER in the presence of low AIC.

Shot noise has also been identified as a possible mechanism for LER when the exposure dose is sufficiently low, in which statistical fluctuations of electron or photon arrival rate can cause dosing variations along the feature profile. Leunissen et al. [4] simulated shot noise effects in e-beam exposure on chemically amplified resists and showed that the 3-sigma LER amplitude rises at low exposure doses ( $<1,000 \mu\text{C}/\text{cm}^2$ ) and remains constant at higher doses, such that resist material properties became the dominant factor.

Roughness or defects in masks can be directly transferred to photoresist layers during exposure and may also be a source of LER. Reynolds and Taylor [5] investigated the amount of LER produced in chemically amplified resists using X-ray exposure masks with various amounts of sidewall roughness. They indeed saw that rougher masks produced line patterns with higher LER (up to several nm root-mean-square) compared to reference masks.

Resist material properties such as polymer weight/size, aggregate formation, and photoacid diffusion play a critical role in the overall LER of resist patterns. The discretization of resist polymers into chains of repeating monomer blocks becomes significant as LER requirements approach the dimensions (1–2nm) of the monomers themselves. Therefore, using polymers with reduced molecular size and/or weight should reduce the LER effect from the finite resolution of discrete monomer elements. Patsis et al. [6] used simulations based on a fast 2D/3D dissolution algorithm to show that smaller polymer weight and size led to reduced LER in resist patterns. Namatsu et al. [1] pointed to an “aggregate extraction model” as the main cause of LER in resist patterns, in which large (20–30nm) entangled polymer chains (aggregates) are not fully deprotected upon exposure and cannot be removed during development. These large aggregates remain along the edge profiles resulting in LER in the form of rolling bumps and valleys as shown in Fig. 7.2; only until the surrounding resist matrix is dissolved can the aggregates then be “extracted” or removed.



**Fig. 7.2** (Left) Schematic illustrations of the polymer “aggregate extraction model” leading to LER in positive and negative tone resists. (Right) AFM images showing LER on resist sidewalls after electron beam exposure and development [1]. Copyright © 1998 American Vacuum Society

Additionally, the effects of photoacid diffusion in chemically amplified resists during postexposure baking (PEB) has also been studied by Yoshizawa and Moriya [7]. In a chemically amplified resist, photon or e-beam exposure stimulates the release of photoacids from photoacid generators (PAGs) which diffuse and catalytically deprotect nearby resist molecules. They found that enhanced acid diffusion had two competing effects: one is a smoothing effect which reduces LER by averaging out the acid concentration along exposed/unexposed edges, and the other is a loss in aerial image quality which increases LER and degrades critical dimension control.

Until recently, there has been less attention on the etching process as a major source of LER despite the fact that the roughness on the final *etched* layer—typically a polysilicon gate line—is what dictates the overall device performance and matching quality. Namatsu et al. [1] showed that roughness profiles on the photoresist mask layer generally do transfer to the underlying etched layer in anisotropic etching systems such as reactive ion etching (RIE), although some attenuation in the amount of LER in the final etched layer may occur when compared to the parent resist layer.

When high energy reactive ions and neutrals impinge on a surface to be etched, morphological changes can take place resulting in surface and sidewall roughness due to random collision events. Etched sidewalls tend to exhibit striations along the vertical direction in which reactive species are accelerated, resulting in anisotropic roughness along the feature sidewalls. Wet etching, on the other hand, produces isotropic roughness typified by a globular morphology along the sidewalls.

Yahata et al. [8] used atomic force microscopy (AFM) to characterize sidewall roughness on silicon trenches etched by RIE and observed the presence of these vertical striations. They also noted the correlation between roughness in the resist pattern and the final roughness of the trenches after RIE; rougher resists produced rougher features after etching.

### 7.2.1.2 Modeling LER

LER is a stochastic (random) process, and hence, requires a statistical description. It is typically characterized by a root-mean-square (RMS) amplitude  $\sigma_{\text{LER}}$ , a correlation length  $\lambda$ , and a characteristic power spectrum shape (e.g., Gaussian, exponential, or other). In almost all cases, the most important LER metric is  $\sigma_{\text{LER}}$  since it describes how large or small the roughness magnitude is. Typical values for  $\sigma_{\text{LER}}$  are on the order of a few nanometers, with smaller values being better. When discussing LER magnitude, some authors will refer to LER by its 3-sigma value rather than its 1-sigma value; however, this is not a universally adopted convention. In this chapter, we will always refer to LER amplitude by its 1-sigma RMS value (i.e.,  $\sigma_{\text{LER}}$ ) unless otherwise stated.

At this point, it is important to distinguish between line *edge* roughness (LER) and line *width* roughness (LWR): when LER exists along both sidewalls of a printed line, the fluctuation of both line edges results in a net fluctuation of the printed line width or *critical dimension* (CD), as depicted in Fig. 7.3.

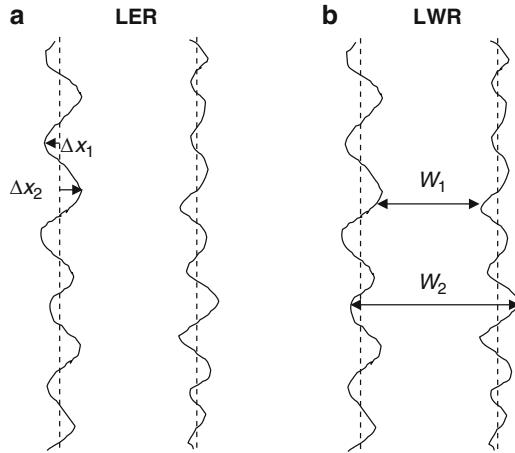
Assuming the same amount of LER on both sides of a line pattern, the magnitudes of LER and LWR will be related by

$$\sigma_{\text{LWR}} = \sqrt{2}\sigma_{\text{LER}}(1 - \rho_X), \quad (7.1)$$

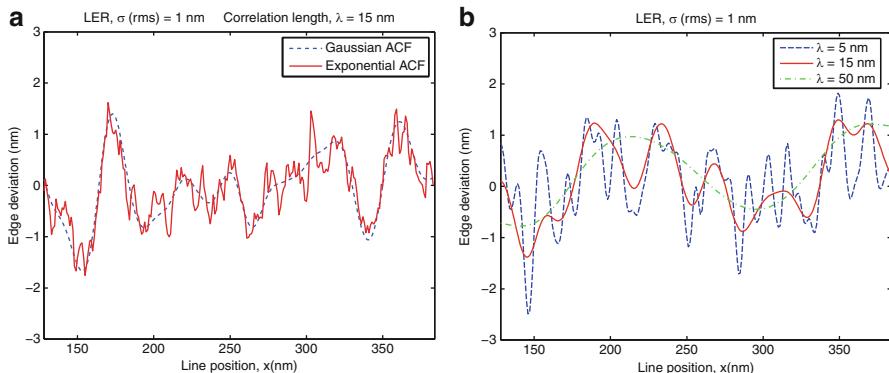
where  $\rho_X$  is the cross-correlation coefficient and ranges between 0 (completely uncorrelated) and 1 (completely correlated). Thus, if  $\rho_X = 1$ , then  $\sigma_{\text{LWR}} = 0$  even if  $\sigma_{\text{LER}} \neq 0$ . This can be achieved using spacer lithography in which sacrificial spacers having correlated sidewall features are used as etch masks rather than standard photoresist as used in resist lithography.

As device dimensions shrink to smaller scales, the LER magnitude must reduce in proportion to ensure that the resulting amount of device variability is kept under control. However, this is not an easy task. The 2011 International Technology Roadmap for Semiconductors (ITRS) [9] calls for  $3\sigma_{\text{LWR}} \leq 2.5\text{nm}$  for 32nm technology (Year 2012), which equates to an aggressive  $\sigma_{\text{LER}} < 0.6\text{nm}$  assuming standard resist patterning. Keeping  $\sigma_{\text{LER}}$  to such sufficiently low values represents one of the major challenges of scaling state-of-the-art front-end processes.

While  $\sigma_{\text{LER}}$  simply describes how large in magnitude the roughness is, the correlation length  $\lambda$  and power spectrum shape essentially describe how “wavy” or “jagged” the roughness is. More specifically, these metrics tell us how the roughness is distributed in spatial frequency when LER is modeled in analytical

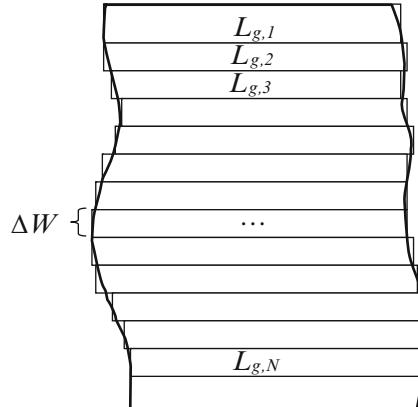


**Fig. 7.3** Definition of LER and LWR in a line pattern. (a) LER is the deviation of the actual edge position (solid line) from the mean or ideal edge position (dotted line). (b) LWR is the deviation of the line width along the length of the line pattern. Note that a rough line structure can have both a nonzero LER and zero LWR, simultaneously



**Fig. 7.4** (a) Representative Gaussian and exponential LER patterns with  $\sigma_{\text{LER}} = 1\text{ nm}$  and  $\lambda = 15$ . (b) Comparison of Gaussian LER patterns with  $\sigma_{\text{LER}} = 1\text{ nm}$  and  $\lambda = 5, 10, 15\text{ nm}$

form via its autocorrelation function (ACF). LER profiles modeled after a Gaussian ACF are predominantly composed of low frequency roughness, while those modeled after an exponential ACF possess more high frequency components, as shown in Fig. 7.4a. Examples of LER patterns from a Gaussian ACF with different  $\lambda$  are shown in Fig. 7.4b, with longer correlation lengths corresponding to slow-varying LER and shorter correlation lengths leading to abruptly varying LER. Typical values of  $\lambda$  are on the order of tens of nanometers, usually between 10 and 50 nm depending on processing.



**Fig. 7.5** Modeling of gate LER using a 2D channel “slicing” approach. Each segment has a characteristic width  $\Delta W$  chosen to be smaller than  $\lambda$ . The gate length  $L_{g,1}, L_{g,2}, L_{g,3}, \dots, L_{g,N}$  is constant within each of  $N$  segment but varies from segment to segment. The total channel current is given by the sum of all  $N$  segments

LER patterns can be experimentally characterized using metrology tools such as CD-SEM or AFM wherein the positions of a line edge are measured at regular intervals and automated software is used to calculate  $\sigma_{\text{LER}}$  and  $\lambda$  from the measured ACF. From that point, a simple Gaussian or exponential model for the ACF may be fitted to represent the measured LER characteristics for use in subsequent process and device modeling.

Once a suitable LER model is constructed with the appropriate  $\sigma_{\text{LER}}$ ,  $\lambda$ , and ACF, random representations of the LER can be generated using the method of “Fourier synthesis” proposed by Asenov et al. [10] and used in device simulations. In this approach, random phases are applied to each frequency component of the power spectrum (obtained by Fourier transforming the ACF), and upon taking the inverse Fourier transform of the spectrum, random instances of LER corresponding to the desired  $\sigma_{\text{LER}}$  and  $\lambda$  are obtained which can then be incorporated into simulated device structures in a statistical ensemble study. Theoretically, any form of LER can be modeled this way, including gate LER and fin or body LER in nonplanar FETs. As such, this modeling approach is very flexible and has been adopted by many research groups for LER simulation studies, including bulk MOSFETs and FinFETs.

A simpler approach [11] which is computationally faster, but less physically rigorous, has been used to study gate LER effects in bulk MOSFETs and is based on the idea of “gate slicing” (Fig. 7.5). In this approach, the channel of a planar MOSFET is divided into thin slices across the transistor width  $W$ , with each slice having a width of  $\Delta W$  such that  $\Delta W \ll \lambda$  and  $W$ . The gate length in each slice is randomly chosen from a Gaussian distribution and the total MOSFET current is calculated as the sum of the currents contributed from each individual slice. A key assumption is that current flow is strictly parallel to each slice and no transverse

current exists. This assumption only holds if  $\sigma_{\text{LER}} \ll L_g$  and the roughness is not too spatially abrupt (i.e.,  $\lambda \gg \sigma_{\text{LER}}$ ), otherwise the current flow may show a sizeable perpendicular component.

### 7.2.2 Random Dopant Fluctuation

*Random dopant fluctuation* or RDF represents the statistically random probability of impurity ions (dopants) occupying certain atomic sites within a crystal lattice. Both the number and the position of dopants can vary in a semiconductor region, and these can have major implications in the specific behavior of individual devices.

Generally speaking, RDF becomes problematic when the expected total number of dopants in a semiconductor device approaches a countable number, usually less than a hundred or so. This frequently occurs when device sizes are reduced without a proportional increase in the doping concentration, such as in the channel of a bulk MOSFET. When the total number of dopants reaches a countable number, any integer fluctuation in that amount constitutes a larger relative variation in the net doping concentration. For example, a MOSFET channel with  $L_g = 100\text{nm}$ ,  $W = 150\text{nm}$ , and  $t_{\text{ch}} = 10\text{nm}$  and nominal doping concentration of  $N = 1.33 \times 10^{18} \text{ cm}^{-3}$  has a total of 200 dopants on average, and the addition or subtraction of a single impurity ion results in just a  $\pm 0.5\%$  change in the resulting doping concentration. On the other hand, a MOSFET channel with  $L_g = 30\text{nm}$ ,  $W = 50 \text{ nm}$ ,  $t_{\text{ch}} = 5\text{nm}$ , and  $N = 1.33 \times 10^{18} \text{ cm}^{-3}$  has only ten dopants on average, and the addition or subtraction of a single impurity ion already results in a  $\pm 10\%$  change in net doping concentration.

Moreover, the spatial inhomogeneity in dopant distribution has a significant effect when the total number of dopants is strongly discretized. In large semiconductor regions with a high number of dopants, the electrostatic potential due to individual impurities will average out in a macroscopic scale and appear more or less homogeneous. In very small devices with a low number of dopants, however, the potential will not average out and microscopic fluctuations will become more pronounced. In this case, RDF effects in the channel of a MOSFET may result in sizeable variation of device parameters such as threshold voltage, drive current, and leakage current [12].

#### 7.2.2.1 Origins of RDF

Impurities may be added to a semiconductor either through deposition or implantation steps during processing; however, the precise location and number of dopants are always somewhat random by nature. While the macroscopic doping profile can be controlled by adjusting the implant dose and energy in an ion implantation process, for example, the exact number and location of added dopants is always

uncertain due to randomness in ionic collisions and diffusion kinetics. During implantation, not all impurity ions will impinge the semiconductor with the same kinetic energy or trajectory, and the resulting distribution can be modeled as a Monte Carlo process. In addition, subsequent annealing steps will cause further redistribution of the microscopic dopant arrangements via diffusion, resulting in further spatial randomization. Since these effects are well known from basic semiconductor processing, we will not spend any further details to understand the exact origins of RDF beyond what we have discussed here. Instead, we will devote the remainder of this section to cover various RDF modeling approaches.

### 7.2.2.2 Modeling RDF

When considering the effects of RDF in semiconductor devices, we must somehow describe the variation in *total number of dopants* as well as the *spatial distribution of those dopants*. Unfortunately, neither of these quantities can be easily measured in actual experiments to great accuracy, unlike the case for LER profiling. This is because the presence of individual dopants cannot be readily located in experimental settings due to their atomic size. Metrology tools such as secondary ion mass spectroscopy (SIMS) which are used to experimentally measure doping profiles simply do not have the required atomic resolution to pinpoint individual ions from the host lattice and can only give crude estimates for the net doping concentration over nanometer scales. Other techniques such as AFM may theoretically provide the necessary resolution to identify and locate individual dopants, but are still highly impractical for RDF mapping through an entire volume of crystal [13]. Contrast these limitations with the ease of visualizing and extracting LER profiles using CD-SEM where variation is on the scale of several to tens of nanometers, and one quickly sees the difficulty of implementing truly accurate RDF metrology.

Because of the high difficulty in directly measuring experimental doping fluctuations, we must determine the amount of RDF in a semiconductor using indirect methods. Let us first try to determine the fluctuation in the number of dopants for a doped semiconductor sample. We cannot precisely count how many total dopants  $N_T$  reside in the sample nor can we measure the variation in total number of dopants  $\sigma N_T$  from sample to sample. However, we can measure the macroscopic doping concentration  $N$  (per  $\text{cm}^3$ ) in the sample along with the volume  $V$  itself, thus allowing us to calculate the average total number of dopants in the sample  $N_T = N \times V$ . Now, integrating the occupational probability of a lattice site by a dopant ion with the sample volume will give us the total number of dopants, meaning that the occupational probability can be determined since  $V$  is known and  $N_T$  can be determined from  $N$ . If completely random occupation is assumed, then the occupation probability follows a Poisson distribution, which results in

$$\sigma N_T = \sqrt{N_T}. \quad (7.2)$$

We can normalize this variation by the average total number of dopants which gives us

$$\frac{\sigma N_T}{N_T} = \frac{1}{\sqrt{N_T}} = \frac{1}{\sqrt{N \times V}}. \quad (7.3)$$

Equation (7.3) captures how RDF gets progressively worse for smaller devices with lower nominal doping concentrations.

We have quantified the fluctuation in number of dopants, and now we must describe how those dopants are spatially distributed. The most rigorous approach is to perform Kinetic Monte Carlo (KMC) simulations to track the motion of discrete dopants through the semiconductor during implantation and annealing, for example. This yields a physically accurate picture of how dopants will be spatially distributed but is computationally expensive, time consuming, and requires details of the processing stage(s). A much simpler, but less accurate, approach is to begin with a known (or assumed) macroscopic doping profile based on a continuum model, and later “atomize” the structure by randomly placing individual dopants at discrete mesh points in the structure, such that the total number of added dopants equals (on average) the integrated value  $N_T$ , and the effective local doping concentration around each mesh point roughly matches the original continuum profile. The latter approach is very fast and does not require any process simulation but instead relies on an accurate continuum model as an input.

In conventional drift-diffusion (DD) simulators, the doping concentration is treated as a *macroscopic* quantity which averages out any microscopic nonuniformities due to the random placement and number of dopants. For devices with a large number of dopants, this continuum (or “jellium”) model is physically reasonable since the electrostatic potential appears spatially homogenous and is sufficiently described by the averaged charge density. In devices where only a few dopants exist and RDF effects are pronounced, we must resolve the potential due to individual dopants which will not spatially average out. Several methods have been proposed to tackle this problem for application in DD simulators, with many based on the so-called “atomistic” approach. In the atomistic approach, dopants are first assigned to random mesh points in the structure either by KMC simulation or by “atomizing” the structure as discussed previously. Once the dopants are distributed, the resulting charge density at each mesh point is calculated by dividing the dopant charge by the volume of the mesh cube which contains it—this is known as the “nearest grid point” (NGP) method [14]. The calculated charge density is then used in the Poisson solver to obtain the electrostatics as usual. A similar method to NGP is called “cloud-in-cell” (CIC), in which the charge of an individual dopant is equally divided among the nearest mesh points surrounding the element wherein the dopant is placed. The resulting charge density is calculated in a similar way for use in the Poisson solver.

In cases where the mesh spacing is sufficiently small (<2nm or so) so that mesh points contain either one or zero dopants only, both the NGP and CIC methods yield potential distributions which exhibit abrupt spikes which coincide with the

Coulomb potential at mesh points where discrete dopants have been placed. In DD simulators, this becomes a problem because mobile carriers become artificially trapped by the potential singularities which results in heavy screening of the impurity charge; this can result in an underestimation of the space charge in a MOSFET channel, for example, leading to an underestimation of  $V_T$ . A further complication arises when the mesh spacing is varied since the electrostatic potential is directly tied to the length of the mesh spacing in the atomistic approach—this leads to an unphysical scenario in which the device electrostatics depends on the mesh size. Several methods have been attempted to rectify this problem, such as (1) artificially increasing carrier mobility in source/drain regions to compensate for trapping, (2) using quantum correction models such as the density gradient approximation (DGA) to smooth out the carrier distributions, or (3) splitting the Coulomb potential into short-range and long-range components according to an appropriate screening length. Among these methods, the splitting of the Coulomb potential is an attractive choice because of its (mostly) physically meaningful basis within the context of the DD framework, despite some shortcomings, while the other solutions lack physical justification for various reasons.

In a study by Asenov et al. [15] the mobility within the source and drain regions of a MOSFET were boosted in order to compensate for the reduction in carrier concentration due to trapping in the bare Coulomb potentials; however, this technique is inherently arbitrary and empirical and does not address the  $V_T$  lowering due to discrete dopants in the channel. Roy et al. [16] invoked the density gradient quantization model to prevent significant trapping of carriers within the Coulomb wells of discrete dopants in a resistor. This resulted in a smoother band profile which was virtually insensitive to mesh spacing. However, the authors still observed some deviation in average resistance with respect to the continuous doping model, suggesting that charge trapping effects were still present, despite avoiding the mesh size dependence. Physically speaking, we should point out that the main purpose of the density gradient model was never intended to resolve issues related to atomistic doping but rather to account for quantum confinement pushing the peak carrier density away from the oxide–channel interface in a thin-body MOSFET within a semiclassical framework.

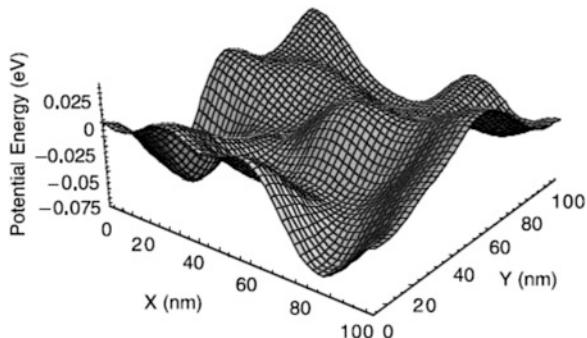
Sano et al. [17] proposed a different method to treat the atomistic dopants in which the Coulomb potential due to an impurity (e.g., acceptor ion) located at  $r = 0$  is split into long-range and short-range components according to:

$$\phi_{\text{long}}(r) = -\frac{2k_c}{\pi} \frac{\text{Si}(k_cr)}{k_cr}, \quad (7.4)$$

$$\phi_{\text{short}}(r) \approx -\frac{e^{-k_cr}}{r}, \quad (7.5)$$

where  $k_c$  is a screening parameter and  $\text{Si}(x)$  is the sine integral. Only the long-range component is used in the DD simulations, and the resulting charge density can be expressed as

**Fig. 7.6** Long-range potential fluctuations due to random discrete dopants under the gate oxide of an MOS capacitor [17]. Copyright © 2002 Elsevier Science



$$\rho_{\text{long}}(r) = -\frac{k_c^3}{2\pi^2} \frac{\sin(k_c r) - (k_c r) \cos(k_c r)}{(k_c r)^3}. \quad (7.6)$$

This results in smoothly varying (and mesh insensitive) charge density and band profiles even when random dopant configurations are taken into account, which is consistent with the assumptions of a semiclassical treatment. An example of the smooth potential obtained from this approach is shown in Fig. 7.6 for an MOS capacitor with RDF. The short-range component, which is the screened Coulomb potential of (7.5), is responsible for carrier scattering and is already taken into account under the assumptions of quasi-equilibrium in a drift-diffusion framework [17]. In other words, the short-range potential is implicitly present in the DD simulations, despite the fact that we have not explicitly added it. The remaining issue, then, is how to choose the screening parameter  $k_c$ . Several possibilities include using (1) the Debye or Thomas–Fermi screening length, (2) the inverse of the mean separation of dopants ( $k_c = 2 \times N^{1/3}$ ), or (3) leaving it as a fitting parameter. The exact choice is somewhat ambiguous and no universal consensus on the best option exists—this remains one of the biggest shortcomings of the Sano approach.

Another approach proposed by Mayergoyz and Andrei [18], and later by Wettstein et al. [19], has been used to simulate the effects of RDF in DD simulations and is known as the impedance field method (IFM). IFM as an RDF model is not physically rigorous and is akin to a small-signal treatment where fluctuations in the number of dopants and carrier densities are assumed to be small enough to justify the effects as a linear perturbation. The doping perturbation is modeled as a noise source, and the resulting variations in potential and terminal currents can be very quickly evaluated without resorting to large ensemble simulations. The major limitations of this approach are its lack of physical validity and the inability to account for spatial doping fluctuations; however, the computational efficiency of the IFM can be exploited in cases where brute force ensemble simulations are infeasible.

At present, there is no ideal RDF model which is completely suitable for use in traditional DD simulators, and all the ones discussed so far have some limitations. So far, however, many researchers have employed at least one of these techniques to study device variability from RDF in various types of FETs, including planar MOSFETs, FinFETs, and even tunnel FETs [20–23]. In the remaining sections of this chapter, we will see how these RDF models (particularly the Sano model) can be used to estimate the impact of RDF in various FinFET technologies.

A simple scaling law which describes performance mismatch due to stochastic variability in planar MOSFETs was formulated by Pelgrom [24], which states that the variation in a given parameter, such as  $V_T$ , can be expressed as

$$\sigma V_T = \frac{A_{VT}}{\sqrt{W \times L}}, \quad (7.7)$$

with  $W$  and  $L$  representing the dimensions of the active channel area and  $A_{VT}$  as the Pelgrom coefficient which quantifies the amount of device variability in a given process. In general,  $A_{VT}$  can be modeled as a function of various device parameters (e.g., oxide thickness  $t_{ox}$ , channel doping  $N$ , etc.) and is typically expressed in units of mV  $\mu\text{m}$  for comparison between different processes. Typical values for  $A_{VT}$  are on the order of several mV  $\mu\text{m}$  with lower values being better.

A frequently cited formula [25, 26] for  $\sigma V_T$  in the case of RDF for planar bulk MOSFETs is

$$\sigma V_T = \left( \frac{\sqrt[4]{4q^3\epsilon_s\phi_B}}{2} \right) \frac{t_{ox}}{\epsilon_{ox}} \left( \frac{\sqrt[4]{N}}{\sqrt{W \times L}} \right), \quad (7.8)$$

where  $\phi_B = 2k_B T \times \ln(N/n_i)$  with  $k_B$  the Boltzmann constant,  $T$  the absolute temperature,  $n_i$  the intrinsic carrier density,  $q$  the elementary charge, and  $\epsilon_s$  and  $\epsilon_{ox}$  the permittivities of the semiconductor and oxide, respectively. Equation (7.8) has been verified against experimental data and simplifies to the more general form of (7.7) where  $A_{VT}$  becomes a function of  $N$  and  $t_{ox}$ . Based on Pelgrom's law, we expect that stochastic device variability will grow worse when the channel area is reduced (i.e., the device is scaled) following an inverse square root dependence, which is intimately related to a reduction in statistical averaging of random, uncorrelated events (essentially the manifestations of random process variations) along the active device area.

### 7.2.3 Other Variability Mechanisms

#### 7.2.3.1 Oxide Thickness Fluctuation

*Oxide thickness fluctuation* or OTF presents another source of device variability in highly scaled MOSFET technologies and can be especially problematic when the nominal oxide thickness  $t_{ox}$  is scaled to less than a couple nanometers. Up until the

65nm node where silicon oxide (or oxynitride) was still predominantly used for the gate oxide material,  $t_{\text{ox}}$  was scaled to as low as 1.2nm which is less than five atomic layers thick. For such a process, fluctuation in just a single atomic layer already represents more than 20% variation in  $t_{\text{ox}}$ ! This can cause significant variation in device performance between similar devices, just like the effects of LER and RDF. It is easy to imagine how further scaling of silicon oxide-based gate dielectrics to 1nm or thinner would become problematic not just from a gate leakage standpoint but also from OTF.

Modeling OTF can be accomplished using the same metrology-based approaches as those used for LER. For example, OTF profiles can be characterized from TEM cross sections of a gate oxide–channel interface (i.e., Si– $\text{SiO}_2$ ) and an ACF can be computed to generate a statistical description of the interface roughness, ultimately yielding an RMS amplitude and correlation length. These can be used to construct random structures with OTF for use in device simulations, just as was described for LER modeling in Sect. 7.2.1.2. Such an experiment was conducted by Asenov et al. [27] using the same Fourier synthesis technique as described previously to determine the impact of OTF (using  $\sigma_{\text{OTF}} = 0.2\text{nm}$  and  $\lambda = 2\text{nm}$ ) on threshold voltage variation for bulk  $30 \times 30\text{nm}$  MOSFETs.

Beginning at the 45nm node, however, the industry began replacing  $\text{SiO}_2$  with high- $\kappa$  dielectrics (e.g.,  $\text{HfO}_2$  with  $\kappa \approx 25$ ) which can be made thicker while still providing the same equivalent oxide thickness (EOT) as  $\text{SiO}_2$  ( $\kappa = 3.9$ ). In this case, not only does the nominal  $t_{\text{ox}}$  become thicker to the point where single- or even multilayer fluctuations comprise a smaller relative variation in  $t_{\text{ox}}$ , but newer deposition processes such as atomic layer deposition (ALD) can significantly reduce the likelihood of OTF occurring in the first place. This is a result of the highly conformal and precise monolayer degree of control which ALD offers. As a result of the industry’s migration towards high- $\kappa$  dielectrics, concerns over OTF have greatly waned since the last decade when planar MOSFETs with low- $\kappa$  dielectrics were still the norm. For modern FinFET technologies utilizing hafnium-based dielectrics, OTF is of limited concern especially when compared to the potential impacts from LER (and to a degree, RDF) as we will see in Sect. 7.3.

### 7.2.3.2 Work Function Variation

The gate material of a MOSFET, whether polysilicon or metal-based, usually consists of multiple randomly oriented grains as opposed to a single grain configuration, thus giving rise to variation of the gate work function between different transistors. This is called *work function variation* (WFV) and presents another possible source of device variability. Like LER and RDF, the significance of WFV is expected to become higher and higher for finer technologies. For transistors with sufficiently large gate area compared to the average grain size, the gate will be composed of many randomly oriented grains whose individual contributions will average out across the entire device, assuming the grain distributions are independent. However, when the gate size approaches that of a single average grain, the

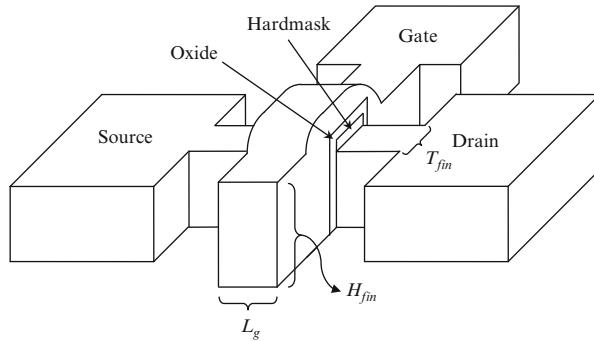
work function distribution can become highly modal and self-averaging effects will not occur in each transistor, thereby resulting in a larger WFV impact. For example, in a typical FinFET using TiN as the gate material, two possible grain orientations are most likely to occur:  $<200>$  ( $\psi_{<200>} = 4.6$  eV) or  $<111>$  ( $\psi_{<111>} = 4.4$  eV), with a 60–40% probability, respectively [28]. When the average grain size (~20nm) becomes comparable to that of the gate area, the resulting variation in threshold voltage reaches a maximum and more or less saturates as the gate size is reduced further [29].

Numerous WFV studies [28–34] have appeared in the last decade with many predicting that WFV effects may become as significant as LER and RDF for gate lengths of ~30nm or lower. As transistors continue shrink to nanoscale dimensions, WFV may become a dominant source of variability and may provide further incentive to develop alternative systems based on amorphous gate materials [35, 36] (e.g., TiN injected with carbon) which are also compatible with state-of-the-art hafnium-based dielectrics. Very recently, the use of amorphous TaSiN metal gates in DG FinFETs has shown up to a 2 $\times$  reduction in  $V_T$  variation compared to polycrystalline TiN metal gates [37]. For fully depleted silicon-on-insulator (FD-SOI) and/or FinFET technologies which rely on metal gates with tunable work functions to set the threshold voltage, issues over WFV will unquestionably become an important topic.

### 7.3 Device Variability in FinFET Technologies

With their successful commercial introduction at the 22nm node in 2011 [38], nonplanar multi-gate FETs will likely become the standard transistor architecture for years to come. Whether in a double-gate (DG), tri-gate (TG), or gate-all-around (GAA) configuration, the fundamental advantages of a multi-gate device (i.e., lower power consumption and higher performance) over a traditional planar FET are realized by the greatly improved electrostatic gate control over the channel region. Despite their improved performance, however, multi-gate FETs are still susceptible to the same process variations that trouble planar CMOS at deeply scaled technologies, due in large part to their shared manufacturing steps.

The question to ask, then, is whether multi-gate FETs such as FinFETs are any more or less vulnerable to device variability from LER, RDF, etc., as a consequence of their physically different architecture—this will be our focus for the remainder of the chapter. Throughout the following sections, we will investigate in detail the impact of random process variations, specifically LER and RDF, on device variability for various FinFET technologies designed for sub-32nm generations to get a sense of (1) how significant variability effects may become for FinFETs heading into the nanoscale regime, and (2) how each variability mechanism specifically impacts device variation from a physical standpoint. Most of the results presented in this section are from our own work in this research area; however, we will also refer to related variability work from other research groups whenever appropriate.



**Fig. 7.7** Schematic of the standard FinFET. A vertical semiconducting “fin” is straddled by the gate electrode running perpendicular to the fin length. The metallurgical gate length  $L_g$  and the fin body thickness  $T_{fin}$  are indicated in the diagram. For a TG FinFET, the effective channel width is the sum of  $T_{fin}$  and  $2 \times H_{fin}$ , where  $H_{fin}$  is the fin height. For a DG FinFET, the effective channel width is just  $2 \times H_{fin}$

Based on these findings, we will develop some basic insights for how different forms of variability can fundamentally affect transistor operation, and how different FET types (e.g., inversion-mode versus junctionless) may be inherently more or less sensitive to fluctuations in geometry or composition due to their underlying technology.

### 7.3.1 Inversion-Mode FinFET Technology

#### 7.3.1.1 IM-FinFET Overview and Design

The standard FinFET, as illustrated in Fig. 7.7, is a vertical extension of the classic MOSFET structure into the third dimension and consists of a vertical stripe or “fin” (forming the channel) which bridges the source and drain regions with the gate electrode wrapped around the fin. Since the gate covers both sidewalls of the channel as well as the top surface of the channel, the FinFET effectively becomes a TG structure with gate-channel coupling arriving from all three sides and current flow parallel to each surface of the fin. However, a DG FinFET may be formed when the dielectric thickness on the upper surface is made much larger than on the sidewalls—this can occur when an additional hardmask layer is deposited on the upper fin surface, thereby only allowing gate-channel coupling along the left and right sidewalls.

In fully depleted FinFETs, the fin body is left intrinsic with the source and drain heavily  $n+$  ( $p+$ ) doped for  $n$ -type ( $p$ -type) conduction. When a gate voltage  $V_G \geq V_T$  is applied an inversion layer forms along each fin sidewall and current flows through the channel. When  $V_G < V_T$ , no inversion layer is formed and no drain current flows. This is characteristic of a standard inversion-mode FET, and such a

**Table 7.1** Nominal design and performance values for sub-32nm IM-FinFETs

Quantity	Technology node			Description
	32nm	21nm	15nm	
$L_g$ (nm)	22	17	13	Physical gate length
EOT (nm)	0.90	0.77	0.64	Equivalent oxide thickness
$N$ ( $\text{cm}^{-3}$ )	$10^{15}$	$10^{15}$	$10^{15}$	Body/fin doping
$T_{\text{fin}}$ (nm)	9.6	8	6.4	Fin thickness
$L_{\text{sp}}$ (nm)	10	8	6	Spacer width
$\Psi_M$ (eV)	4.47	4.47	4.47	Gate work function
$V_{\text{DD}}$ (V)	0.9	0.81	0.73	Power supply voltage
$V_{T,\text{lin}}$ (mV)	272	282	298	Lin. threshold voltage (max $g_m$ method with $V_{DS} = 50$ mV)
$V_{T,\text{sat}}$ (mV)	201	203	208	Sat. threshold voltage (const. $I = W/L_g \times 10^{-7}$ A with $V_{DS} = V_{DD}$ )
$I_{\text{on}}$ ( $\mu\text{A}/\mu\text{m}$ )	1,432	1,527	1,734	On-state drive current with $V_{GS} = V_{DS} = V_{DD}$
$I_{\text{off}}$ ( $\text{nA}/\mu\text{m}$ )	6.7	9.7	13.3	Off-state leakage current with $V_{GS} = 0$ and $V_{DS} = V_{DD}$
SS (mV/dec)	67.9	69.8	71.6	Subthreshold swing
DIBL (mV/V)	24.0	32.0	39.7	Drain-induced barrier lowering

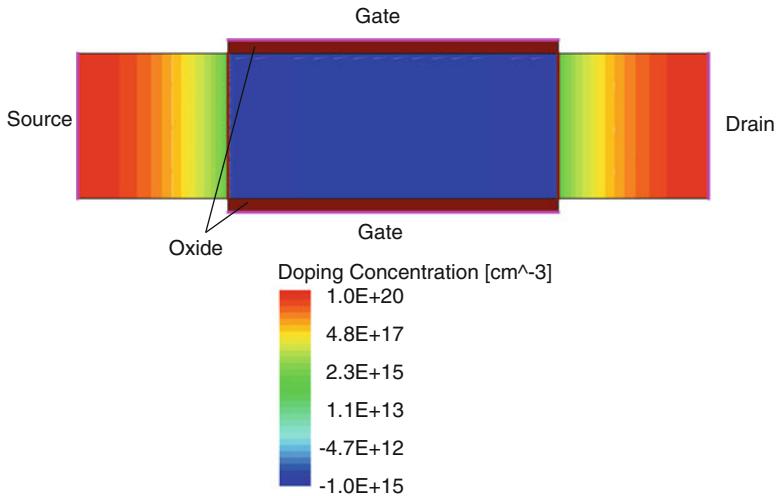
[40] Copyright © 2011 IEEE

FinFET is called an *inversion-mode* FinFET (IM-FinFET) to distinguish it from another type of FinFET which will be discussed in Sect. 7.3.2.

To assess the impacts of LER and RDF on FinFET variability, we often rely on statistical TCAD simulations—rather than direct experimental measurements<sup>1</sup>—using one or more of the variability models discussed in Sects. 7.2.1 and 7.2.2. Before proceeding to investigate the impacts of LER and RDF, however, we must first design/construct the baseline FinFET structure which will be used in those simulations. In our study, we will consider DG FinFET structures designed to meet the 32, 21, and 15nm ITRS [9] high-performance logic nodes. These devices correspond to the IM-FinFETs investigated in our related works [39] and [40]. For IM devices, the design parameters and nominal performance values are given in the upper and lower portions of Table 7.1, respectively.

Figure 7.8 shows the nominal structure for our simulated IM-FinFETs without any process variations. Only  $n$ -type silicon devices are used throughout this work with the assumption that the majority carrier type will not significantly change the variability trends to follow, other than a shift in baseline performance. The structure shown is implemented in Sentaurus Device [41] using 2D simulations only and represents a true DG SOI-based IM-FinFET where we assume the hardmask in Fig. 7.7 and the buried oxide are both infinitely thick so that no current flows in the vertical direction. The use of 2D simulations greatly reduces

<sup>1</sup>This is due to the enormous burden involved with measuring a sufficiently large number (hundreds to thousands) of experimental devices, along with the inability to completely separate the contributions of each variability source from one another in real devices.

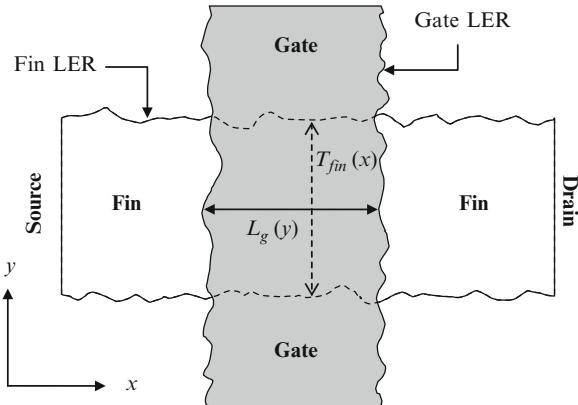


**Fig. 7.8** Simulated *n*-type DG IM-FinFET structure (32nm case shown) used for our device simulations. The structure represents a planar cut across the fin height and parallel to the wafer plane

computation time which is crucial for ensemble variability studies employing thousands of simulations. The source/drain extension doping is Gaussian and terminates at the gate edges, forming no overlap or underlap. The channel is assigned a residual doping of  $10^{15} \text{ cm}^{-3}$  corresponding to a background wafer doping, although at the dimensions of interest the corresponding depletion charge has a negligible contribution (on the order of microvolts) to the threshold voltage. A hydrodynamic (HD) transport model with modified energy relaxation time ( $\tau_n = 1.4 \text{ ps}$ ) and flux ( $r_n = 0.3$ ) parameters is used to capture velocity overshoot effects in the short channel devices [42], and the DGA is used to model carrier quantization in the narrow fins, an effect which leads to volume inversion in the channel. These models give a reasonable balance between accuracy and simulation time versus more sophisticated models such as Monte Carlo or Schrodinger solvers. Other mobility models include surface scattering, impurity scattering, and carrier temperature-dependent components. Gate leakage is neglected with the assumption that the ultrathin  $t_{\text{ox}}$  represents the EOT of a high- $\kappa$  dielectric (e.g.,  $\text{HfO}_2$ ).

In the next section, we will see how LER affects the resulting variability of our IM-FinFETs in terms of the six performance metrics listed in the lower portion of Table 7.1.

**Fig. 7.9** Illustration of gate LER and fin LER in a FinFET. Fin LER modulates the fin thickness  $T_{fin}$  along the channel. Gate LER modulates the gate length  $L_g$  along the fin body

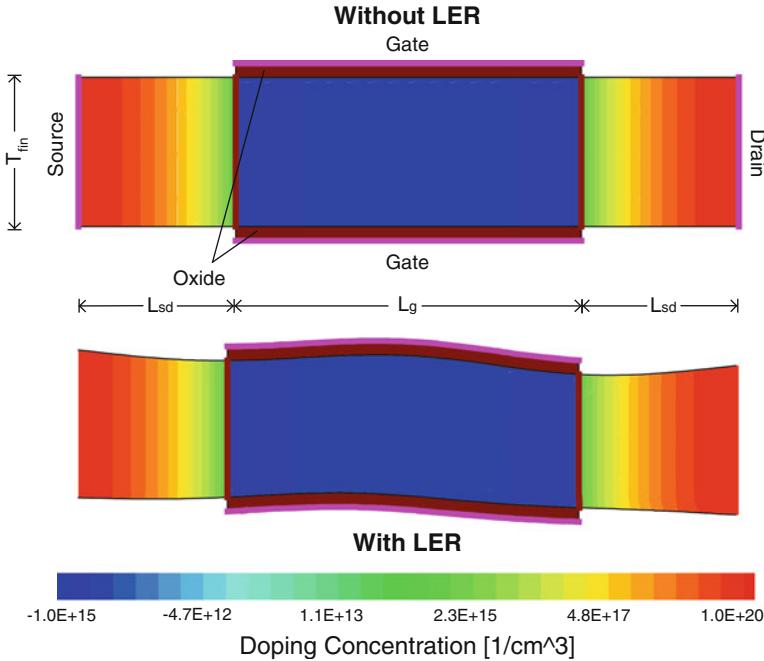


### 7.3.1.2 LER Impact on IM-FinFET Variability

LER can exist along the gate or fin of a FinFET, leading to “gate LER” and “fin LER”, respectively. As illustrated in Fig. 7.9, gate LER leads to fluctuation of the gate length  $L_g$  across the channel while (lateral) fin LER<sup>2</sup> leads to fluctuation of the fin body thickness  $T_{fin}$  along the channel. Gate LER in a FinFET is similar to gate LER in a planar MOSFET, an effect which has already been extensively studied. Fin LER, however, is somewhat unique to nonplanar devices and results in an additional variability contribution not seen in planar MOSFETs. Usually,  $T_{fin}$  is kept  $\leq 0.6L_g$  in a well-designed DG FinFET in order to maintain good electrostatics and proper leakage control. When fin LER is present, however, the fluctuation in body thickness directly alters the FinFET’s short channel effect (SCE) control and may have a significant impact on performance variation, even more so than gate LER. Previous studies [43–46] have shown that fin LER is the more dominant contributor to threshold voltage variation in 32nm IM-FinFETs compared to gate LER, and will likely remain so for smaller generations. In this section, we will focus primarily on fin LER for this reason, and henceforth any references to LER will be taken to mean fin LER unless otherwise indicated. The effects of gate LER will undoubtedly contribute to the overall device variability in real-world settings, but an explicitly treatment of gate LER is outside of the scope of our work.

In our study, we generated a large number of random LER patterns (similar to Fig. 7.4) from a Gaussian ACF using the method of “Fourier synthesis” described in Sect. 7.2.1.2 and [10], corresponding to various  $\sigma_{LER}$  and  $\lambda$  combinations. These random LER patterns are then used to augment the fin edges of our simulated

<sup>2</sup> We neglect vertical fin LER (along the fin height) in this study for two reasons (1) to reduce the simulation complexity and (2) fins formed by anisotropic etching typically consist of mostly lateral roughness along with correlated vertical striations (yielding less vertical roughness).



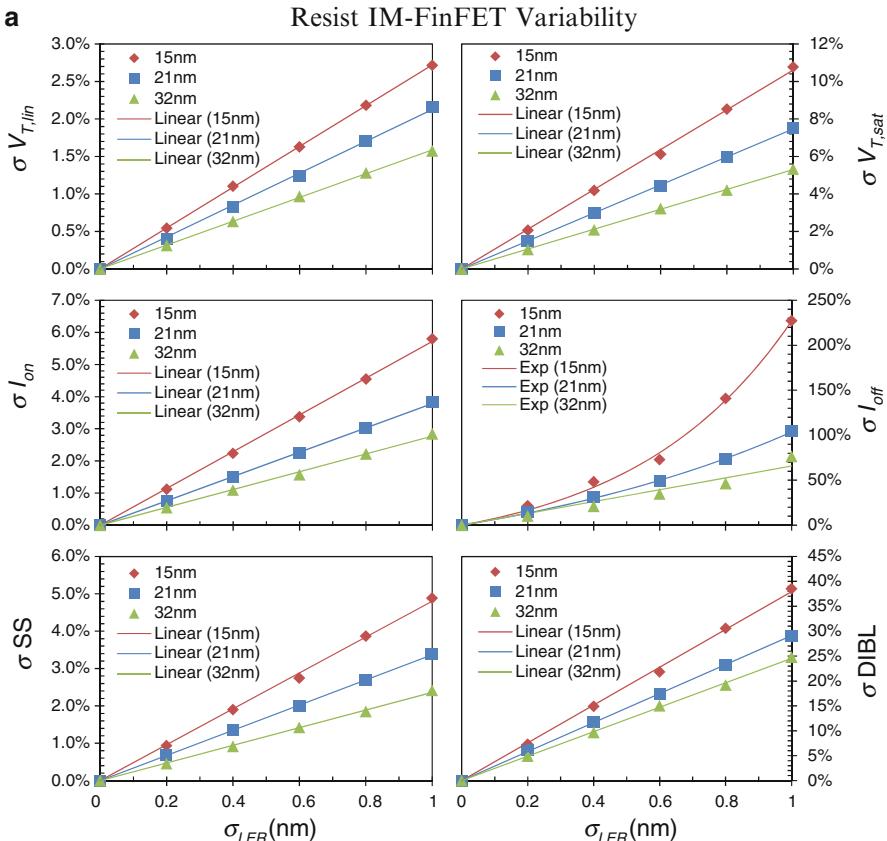
**Fig. 7.10** Simulated IM-FinFET structures with and without 1nm LER along the fin sidewalls [40]. Copyright © 2012 IEEE

devices, yielding IM-FinFETs with fin LER. We considered the range  $0 \leq \sigma_{LER} \leq 1\text{nm}$  to represent typical LER values which may be required by industry heading beyond 32nm technology, based on the 2011 ITRS forecast [9] and experimental data [10]. We fixed the correlation length at  $\lambda = 15\text{nm}$  in our work for several reasons (1) to limit the permutations of  $\sigma_{LER}$ ,  $\lambda$ , and technology node to a reasonable number in our study; (2) previous studies [44, 45] have shown that the effect of  $\lambda$  diminishes as  $\lambda > 15\text{--}20\text{nm}$ ; and (3) some experimental data has shown that current values of  $\lambda$  are estimated between 20 and 30nm [10] and generally reduces with technology, suggesting  $\lambda = 15\text{nm}$  as a reasonable estimate for sub-32nm generation lithography. Examples of IM-FinFETs with and without 1nm LER applied to the fin edges are depicted in Fig. 7.10.

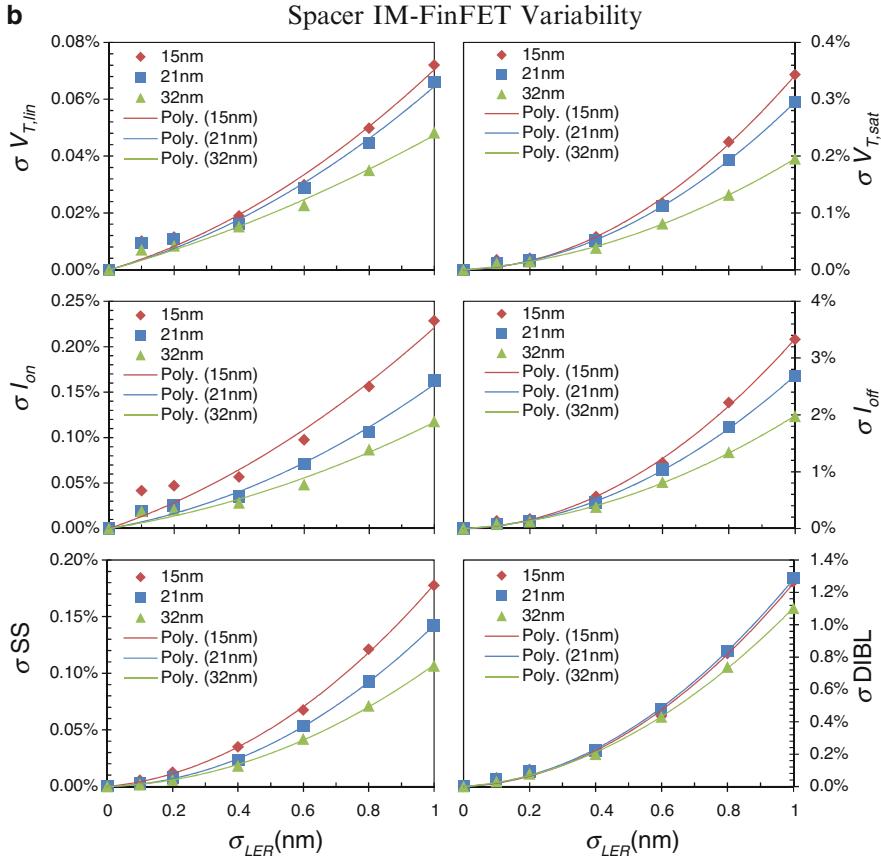
The Gaussian LER model was chosen for reasons which will be explained next. Surface smoothing treatments such as thermal annealing [47, 48], sacrificial oxidation [49], and resist trimming [50] are capable of eliminating the majority of high frequency roughness, leaving mostly low frequency roughness in etched features. Moreover, it has been shown [44] that low frequency roughness is the more significant source of intra-die variability characteristic of LER. With this in mind, we desired a simple analytical form for the ACF in order to reduce the simulation complexity, and one whose power spectrum consisted of mostly low frequency roughness and negligible contribution from higher frequencies, leading us to

consider the Gaussian model over the exponential model which retains nonnegligible high frequency components.

The LER impact on lithographically defined (“resist”) IM-FinFETs is shown in Fig. 7.11 as a function of  $\sigma_{LER}$ , where all performance figure variations are expressed as percentages of their nominal values from Table 7.1. Each data marker in Fig. 7.11 represents 200 unique simulations, where 200 is our ensemble sample size for each pair of  $\sigma_{LER}$  and technology node. Moderate to large variation of  $V_{T,lin}$  and  $V_{T,sat}$  with LER is evident, especially in the latter case where  $\sigma V_{T,sat}$  can exceed 10%. As expected, the 15nm devices show the most variation, while the 32nm devices show the least. The threshold voltage variation depends linearly on  $\sigma_{LER}$  since the total depletion charge in a fully depleted FinFET is directly impacted by fin thickness fluctuations, i.e., fin LER. This amount of LER-induced  $V_T$  variation



**Fig. 7.11** (a) Resist and (b) spacer IM-FinFET device variability as a function of LER amplitude and technology node. *Markers* indicate actual simulated data while *solid lines* indicate best fits. Note the zoomed scale for spacer IM-FinFET data compared to resist IM-FinFET data [40]. Copyright © 2012 IEEE



**Fig. 7.11** (continued)

may be troublesome in circuits requiring precise threshold voltage matching. Similar levels of  $V_T$  variation due to fin LER have also been found in [44] and [45].

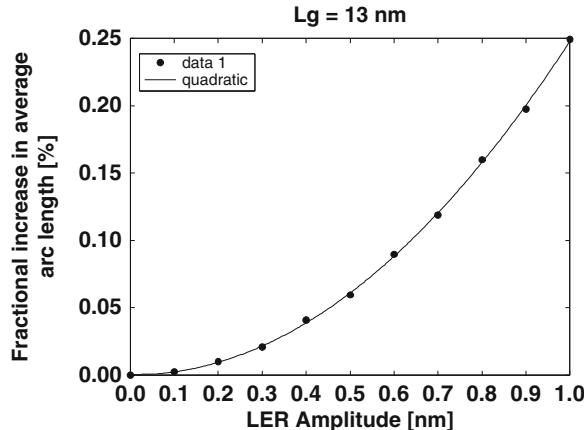
$I_{on}$  variation exhibits a similar but weaker dependence considering that  $\sigma_{I_{on}}$  can easily be kept within 10% of the nominal value in each technology node up to  $\sigma_{LER} = 1\text{nm}$ ; similar findings have also been reached in [44] and [45]. Note that  $\sigma_{I_{on}}$  is also linear with  $\sigma_{LER}$  since drive current is linearly proportional to  $V_{DD} - V_T$  in velocity saturated FETs. The variation of  $I_{off}$  is much more pronounced, however, where  $\sigma_{I_{off}}$  varies exponentially with  $\sigma_{LER}$  (since  $I_{off}$  is an exponential function of  $V_T$ ) and reaches more than 200% of the nominal value for 15nm devices. Such wild fluctuations in  $I_{off}$  may be detrimental to circuit performance if the power dissipation of individual devices and circuit blocks cannot be kept within acceptable margins. In light of these results, it appears that the drastic variation of  $I_{off}$  due to fin LER may be a critical obstacle toward further scaling of FinFETs beyond 32nm.

The effect of LER on SS is somewhat low on the order of a few percent and is also linear since the fluctuation of  $T_{\text{fin}}$  due to  $\sigma_{\text{LER}} \leq 1\text{nm}$  can be treated as a linear perturbation in  $C_D$ , i.e.,  $C_D = \epsilon_{\text{Si}}/(T_{\text{fin}} + \Delta T_{\text{fin}}) \approx \epsilon_{\text{Si}}/T_{\text{fin}}(1 - \Delta T_{\text{fin}}/T_{\text{fin}})$  where  $\Delta T_{\text{fin}}$  is roughly given by  $\sigma_{\text{LER}}$ . DIBL variation is more considerable— $\sigma_{\text{DIBL}}$  easily exceeds 10% in each generation over the LER range—as opposed to the SS variation, which can be kept under 10% for the entire LER range.

For spacer-defined (“spacer”) IM-FinFETs having 100% correlated LER on both fin sidewalls, the impact of LER is drastically reduced in terms of parameter fluctuations for all three technology generations. Note the zoomed vertical scales used in Fig. 7.11 for spacer IM-FinFETs compared to those for resist IM-FinFETs. From the data, the elimination of LWR by spacer lithography (due to sidewall correlation) offers substantial improvement in minimizing device variation. These results compare well to the findings in [44] which demonstrate a significant reduction in the saturation threshold voltage mismatch and current factor mismatch to less than 1% of the nominal values over a similar LER range. We also observe that in most cases the variability curves show less dependence on the actual technology node for spacer IM-FinFETs. In other words, there is little difference between the 32, 21, and 15nm cases here. From this, we see that the presence (absence) of LWR is responsible for the observed variability trends in the resist (spacer) IM-FinFETs, rather than the actual LER itself.

Interestingly, every parameter investigated appears to vary quadratically, rather than linearly, with  $\sigma_{\text{LER}}$ . To explain why, we first observe that because of the correlated fin edges in a spacer IM-FinFET the body thickness does not change along the length of the fin, i.e.,  $\sigma_{\text{LWR}} = 0$ . However, the presence of LER causes the body/channel region to bend and curve in shape which results in a curved potential profile compared to an ideal device, and hence, the path for current should roughly follow the curvature of the fin geometry. Mathematically, the total arc length from source to drain can only lengthen due to random vertical displacement of the fin edge, i.e., LER, and the fractional increase in arc length tends to increase quadratically with the root-mean square vertical deviation. This was confirmed by directly analyzing the LER patterns in MATLAB and determining the relationship between average arc length and roughness amplitude as shown in Fig. 7.12. Variation in the arc length due to LER can thus be treated as variation in the effective channel length of the device which is subsequently manifested in the trends of Fig. 7.11b.

Note that we have assumed perfectly correlated fin sidewalls, i.e., zero LWR, in this analysis. In reality, spacer lithography may not generate 100% correlated edges on both sides due to variations in the deposition and etch processes, or subsequent annealing steps. Experimentally, it has been shown that the actual LWR can be nonzero in spacer-defined FinFETs [43] so that a more realistic estimate of spacer FinFET variability would likely involve a weighted average of the resist and spacer FinFET results, where the emphasis on each depends on the magnitude of the cross-correlation coefficient  $\rho_X$ . However, systematically generating random LER patterns where each top–bottom pair represents a deterministic  $\rho_X$  is nontrivial and impractical here.

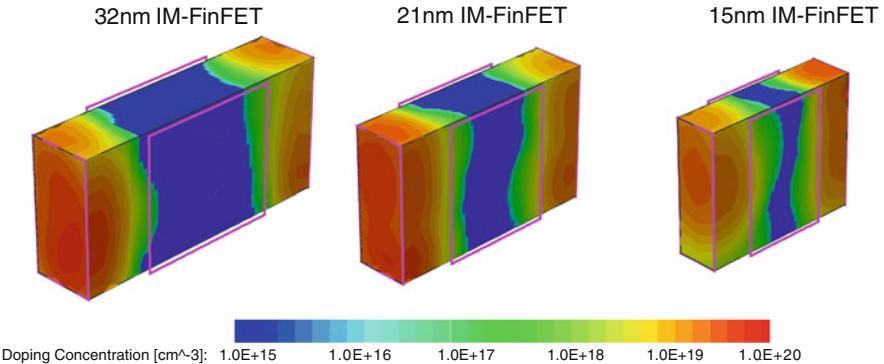


**Fig. 7.12** Quadratic rise in average arc length for spacer FinFETs due to LER as a function of root-mean-square amplitude. The nominal arc length corresponds to a 13nm channel length for the data shown [40]. Copyright © 2012 IEEE

Ultimately, the impact of LER does not appear to pose a major obstacle for IM-FinFETs to meet the demands of future generations (15nm and smaller) given current and projected lithography capabilities (i.e.,  $\sigma_{\text{LER}} \leq 1\text{nm}$ ). This finding is primarily attributed to the robustness of IM-based technology whose fundamental mode of operation is not jeopardized by geometric fluctuations arising from LER. In IM-FETs, switching is predicated on the existence of opposing  $p-n$  junctions at the source-channel and drain-channel interfaces to block current flow in the “off” state, and bridging those junctions by means of electrostatically generating an inversion layer which enables current flow in the “on” state. Fundamentally, this action has no outright dependence on the geometry (i.e., thickness) of the channel, meaning any geometric fluctuations within the channel do not directly prevent the switching operation from happening. Only at short channel lengths does the body thickness matter, yet it remains a secondary effect only. It is for this reason that IM-FinFETs remain viable at small geometries even in the presence of LER, whereas other FET technologies (e.g., junctionless) may not, which we will discover in Sect. 7.3.2.

### 7.3.1.3 RDF Impact on IM-FinFET Variability

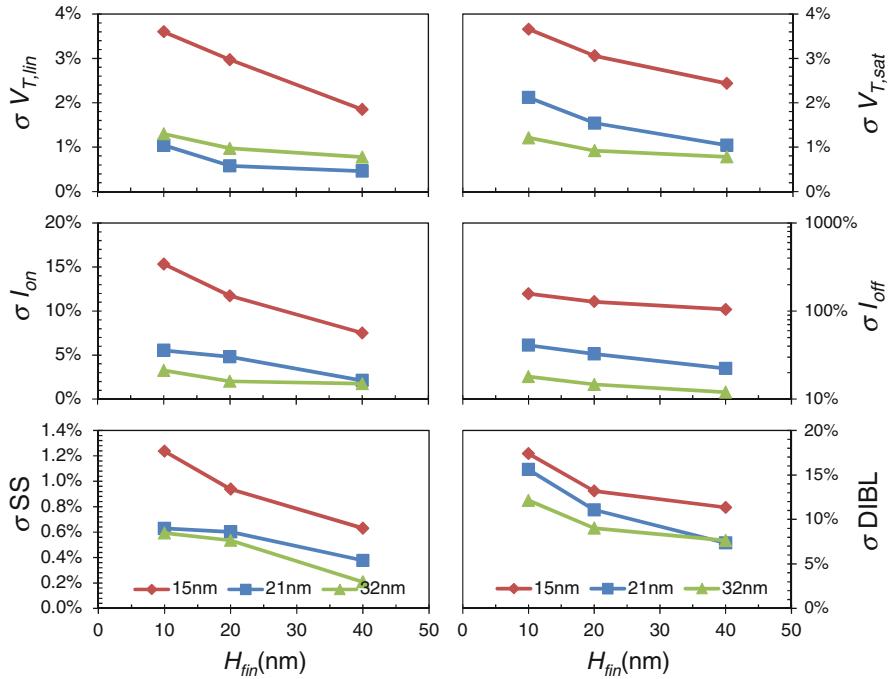
One of the desirable features of an IM-FinFET is the absence of channel doping which would normally contribute to RDF-induced threshold voltage and current variations, as in the case of planar bulk MOSFETs. As a result, IM-FinFETs have gained a reputation for being relatively immune to RDF compared to their traditional planar counterparts. In reality, however, IM-FinFETs are not as immune to



**Fig. 7.13** Random doping profiles in 32, 21, and 15nm IM-FinFET devices. The effective channel length becomes nonuniform and reduces on average at smaller nodes

RDF as one might think due to the presence of dopants in the source and drain regions. When scaled to very small dimensions, the source and drain extensions in an IM-FinFET can exhibit a sizeable amount of RDF resulting in nonnegligible fluctuation of source/drain resistance and junction overlap length. These effects will become more significant for devices with shorter  $L_g$ ,  $L_{sd}$ , and  $T_{fin}$  (all of which occur at finer technologies) and thus warrants our attention.

To account for RDF in our IM-FinFETs within our simulation framework, we adopted the modeling approach pioneered by Sano which uses the long-range component of the Coulomb potential in DD simulations as described earlier in Sect. 7.2.2.2. For each IM-FinFET, we first extrude the structure of Fig. 7.8 into the  $z$ -dimension to create a 3D structure which is necessary for RDF modeling. The resulting structure is still a DG device, however, since the top gate is still absent. Second, we take the continuous doping profile and randomize the total number and position of each dopant according to the “atomization” technique of Sect. 7.2.2.2. Finally, once the dopants have been randomly distributed within the simulated structure, the resulting charge density is calculated using (7.6) for each discrete dopant allocated in the structure with a dynamic screening parameter given by  $k_c = 2 \times N(x_o, y_o, z_o)^{1/3}$  where  $N(x_o, y_o, z_o)$  is the local dopant density at the discrete dopant’s location. Effectively, this randomizes the local doping profile for each IM-FinFET, but does so in a way that what results is still a continuous (but randomized) doping profile which the device simulator can utilize as before. Examples of IM-FinFETs with RDF incorporated are shown in Fig. 7.13. We see that the lateral influence of source/drain dopants encroaches into the intrinsic channel and results in a shortened and nonuniform effective channel length  $L_{eff}$  compared to the metallurgical gate length  $L_g$ . For smaller technology nodes, the amount of encroachment becomes a larger fraction of the overall gate length, resulting in higher unintentional overlap. In a moment, we will see the implications of this effect on the performance variability of our IM-FinFETs.



**Fig. 7.14** RDF-induced variability in IM-FinFETs as a function of fin height and technology node [22]. Copyright © 2012 IEEE

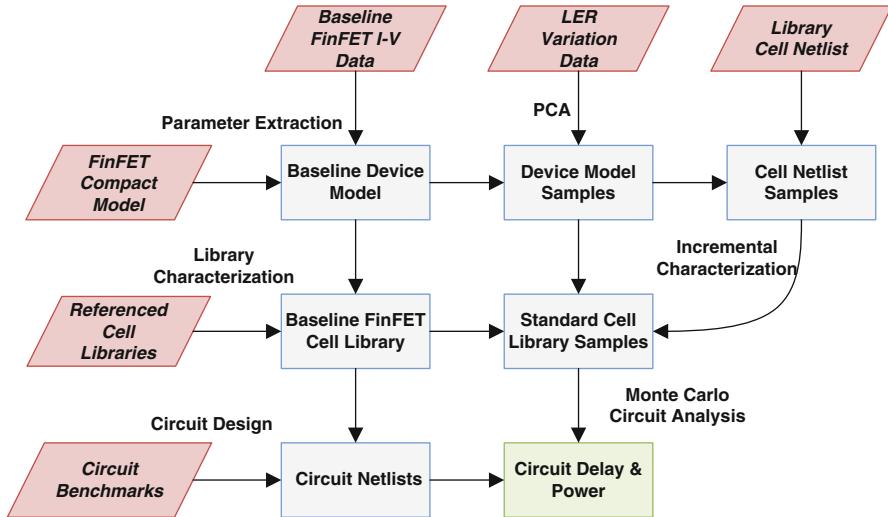
The RDF impact on IM-FinFETs is presented in Fig. 7.14 which shows that Pelgrom's scaling law also applies to IM-FinFETs as well where  $H_{\text{fin}}$  replaces the channel width  $W$  in (7.7). As  $H_{\text{fin}}$  increases, performance variation generally reduces with an inverse relationship appearing, and smaller technologies such as 15nm tend to exhibit more RDF-induced variability than larger technologies such as 32nm. We also observe that  $\sigma V_{T,\text{lin}}$  and  $\sigma V_{T,\text{sat}}$  are kept below 5% in all cases,  $\sigma SS$  is kept below 2%, and  $\sigma DIBL$  below 20%, all of which are good results and demonstrate the advantage of having a completely intrinsic channel to suppress RDF. Leakage current variation is also relatively well controlled with  $\sigma I_{\text{off}} \leq 100\%$ , compared to roughly 200% with 1nm LER (Fig. 7.11). However,  $\sigma I_{\text{on}}$  is still somewhat large reaching up to 15% for 15nm IM-FinFETs; this is a direct result of the shortened, and highly variable,  $L_{\text{eff}}$  resulting from RDF illustrated in Fig. 7.13. In this respect, the major concern for RDF in IM-FinFETs will likely be variation in drive current, especially as the gate length is scaled toward nanometer dimensions. We should note, however, that RDF does not jeopardize the intrinsic switching capability of IM-FinFETs since the electrostatic generation and removal of an inversion layer in the channel occurs regardless of whether RDF exists in the source and drain. RDF only changes  $L_{\text{eff}}$  and as such it has a secondary impact on transistor performance only—again, this will be in contrast to the situation for

JL-FETs. Overall, however, IM-FinFET technology demonstrates good resistance to RDF by virtue of its intrinsic channel with minimal impact on threshold voltage and SCE control. Although we have only considered DG FinFETs so far, we expect the same trends to hold true for TG or GAA structures as well which operate as IM devices having intrinsic channels.

### 7.3.1.4 Circuit-Level IM-FinFET Variability

Besides focusing on transistor-level performance variations, it is equally important to consider their resulting impact on circuit-level variability when judging the strengths and weaknesses of a given process technology. In other words, we should map variations in device performance metrics, i.e.,  $\sigma V_T$ ,  $\sigma I_{on}$ ,  $\sigma I_{off}$ , etc. to variations in circuit performance metrics such as gate delay, active and standby power, and noise margin so that we may systematically evaluate how variability affects performance and design strategies across multiple hierarchies. The challenge lies in connecting each level of abstraction in a meaningful way, allowing us to start from process variability, translate that to device variability, then to circuit variability, and finally to the system and/or architecture level. This type of cross-layer technology evaluation is highly ambitious and often difficult to realize fully in practice, especially for those technologies which remain in research and developmental stages. In this section, we will describe our recent work aimed to systematically evolve the device-level variability results from LER for IM-FinFETs (Sect. 7.3.1.2) to circuit-level variability results obtained through large-scale digital circuit simulations.

A simple framework was developed to transform the device-level FinFET behavior, including  $I$ - $V$  characteristics and performance variations due to LER, into fluctuations in digital circuit delay and power consumption as illustrated in Fig. 7.15. Here, a reference FinFET compact model [51] is fitted to match the nominal  $I$ - $V$  characteristics obtained from TCAD simulation using parameter extraction to generate a baseline compact model. The baseline compact model is then used to characterize a baseline cell library that contains the timing and power information of each logic gate, which will later be used for circuit synthesis, placement and routing (SPR), and further incremental characterizations. SPR is performed for two processor benchmarks: MIPS [52] and ARM Cortex-M0 [53], clocked at different periods (fast, medium, and slow). Variability is modeled by varying the compact model parameters such that device metric sample variations match with those obtained from device-level TCAD simulation. The method of principal component analysis (PCA) [54] is used to translate device-level variations to compact model parameter variations [55]. Using the compact model samples, the cell library samples are then generated from our baseline library and incrementally characterized to simulate their resulting circuit performance by conventional tools. Finally, Monte Carlo circuit analysis is performed on the cell library samples to extract variations in delay and power.



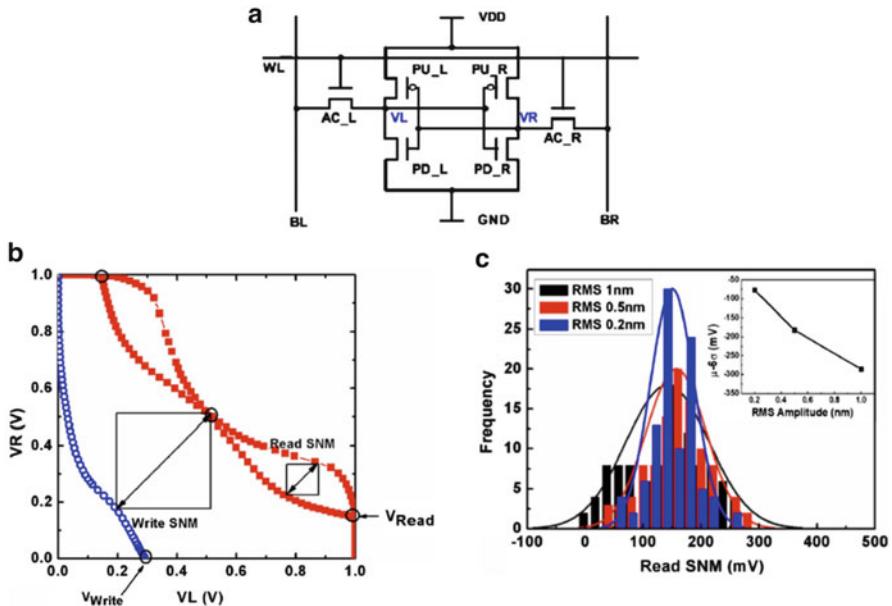
**Fig. 7.15** Overall flow of the circuit benchmark evaluation process. Parallelograms represent input data while rectangles represent output data [40]. Copyright © 2012 IEEE

**Table 7.2** Mean and variations in microprocessor delay and leakage power for different IM-FinFET technologies

Node	Delay			Leakage		
	Baseline w/o LER	Mean w/LER	Sigma w/LER	Baseline w/o LER	Mean w/LER	Sigma w/LER
32-S	952 ps	100%	0.00%	14.65 $\mu$ W	100%	0.0%
32-R		101%	0.15%		114%	0.1%
21-S	635 ps	100%	0.00%	11.47 $\mu$ W	100%	0.0%
21-R		106%	0.30%		125%	0.1%
15-S	381 ps	100%	0.01%	6.59 $\mu$ W	100%	0.0%
15-R		102%	0.04%		149%	0.2%

[40] Copyright © 2012 IEEE

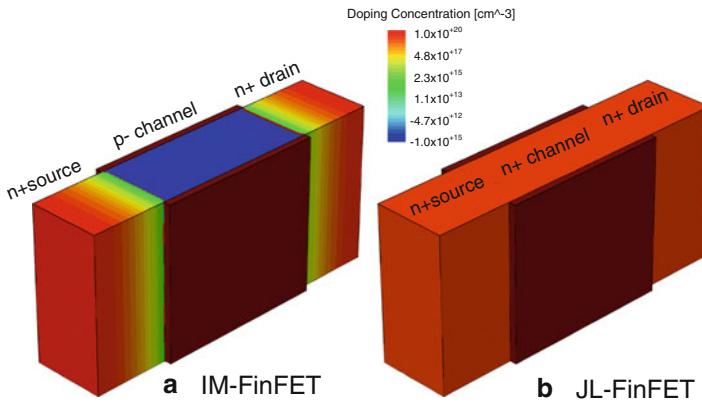
A summary of the findings for the LER impact on IM-FinFET circuit performance is given in Table 7.2 which shows the normalized mean increase and variation for circuit delay and leakage power in 32, 21, and 15nm resist and spacer IM-FinFETs with 1nm LER. Despite the nonnegligible device-level variations from 1nm LER in Fig. 7.11, negligible (<1%) variation is observed for delay and leakage across all benchmarks in all IM-FinFET technologies considered. This conclusion is attributed to the stochastic nature of LER, whose effects average out between different cells along a critical path. We also observe that there is negligible increase in mean delay due to LER, while up to 49% increase in mean leakage is obtained for 15nm resist IM-FinFETs. The increase in mean leakage power is caused by a mean increase in  $I_{off}$  due to LER from the device level, while the lack of mean delay



**Fig. 7.16** (a) Schematic of 6T FinFET SRAM cell. (b) DC curves in read and write operation with no process variations. Read and write SNM are indicated by the largest sized squares nested in the read and write curves. (c) Distribution of read SNM versus  $\sigma_{LER}$  and calculated values for “ $\mu - 6\sigma$ ” versus  $\sigma_{LER}$  in the inset [56]. Copyright © 2012 IOP Publishing

change results from no discernible increase in  $I_{on}$  from LER at the device level. Overall, the impact of LER on large-scale digital microprocessors is minimal, except for a moderate increase in mean leakage power for resist IM-FinFET technologies. Additionally, spacer lithography eliminates all LER impacts at the circuit level. These conclusions will likely be repeated even for RDF considering the purely stochastic nature of both variability mechanisms.

While LER is shown to have little impact for large-scale IM-FinFET digital circuits, the same cannot be said for circuit topologies which are more sensitive to device mismatches such as static random access memory (SRAM) cells shown in Fig. 7.16a. For SRAMs, the effects of transistor mismatch due to stochastic variability will not average out when a cell only contains six transistors (6T) for example, and as a result the stability of the cell—measured by the static noise margin or SNM in Fig. 7.16b—can easily be jeopardized even for well-controlled amounts of process variation. Yu et al. [56] performed mixed-mode simulations on 6T IM-FinFET SRAMs with  $L_g = 20\text{nm}$ ,  $T_{fin} = 5\text{nm}$ , and  $\sigma_{LER}$  up to 1 nm and showed in Fig. 7.16c that extracted values for SNM during read operation (“read SNM”) can fail the “ $\mu - 6\sigma$ ” test even for  $\sigma_{LER} = 0.2\text{nm}$  and  $V_{DD}$  values ranging from 0.6 to 1 V. SNM during write operation (“write SNM”), however, was found to be higher than read SNM for the same LER amplitude, indicating better stability to process variations for write operations compared to read operations.



**Fig. 7.17** Examples of (a) inversion-mode and (b) junctionless FinFETs. The IM device has oppositely doped source/drain and channel regions, while the JL device has uniform doping in all regions

Baravelli et al. [44] also used mixed-mode simulations and experimental measurements to analyze 6T SRAMs for low standby power 32nm IM-FinFETs in the presence of LER, and showed that SNM variation can be reduced using multiple fin designs and at the cost of lower cell density. Spacer lithography can be exploited to generate multiple fins without significantly increasing cell area, but measured results showed no benefit in terms of SNM variation, possibly due to fabrication issues related to their spacer patterning.

### 7.3.2 Junctionless FinFET Technology

#### 7.3.2.1 JL-FinFET Overview and Design

*Junctionless* transistors [57–63] have quickly become a popular topic in recent years as a possible replacement for standard IM transistors due to their simplified processing and comparable performance. Fundamentally, the only defining characteristic of JL-FETs is the absence of any *p*–*n* junctions between the source, channel, and drain regions, hence the name “junctionless.” Since JL-FETs lack any junctions, the nominal doping concentration is typically designed to be uniform and homogeneous throughout the source, channel, and drain regions, making JL-FETs resemble gated resistors. A crucial benefit of this is the ability to bypass processing steps which normally plague IM devices related to ultra-shallow junction formation and downstream thermal budget management. JL-FETs may be conceived in any standard configuration based on planar or nonplanar architectures, including SOI, DG or TG FinFETs, nanowire (NW) FETs, etc. An example of a JL-FinFET is illustrated in Fig. 7.17 next to an IM-FinFET.

**Table 7.3** Nominal design and performance values for sub-32nm JL-FinFETs

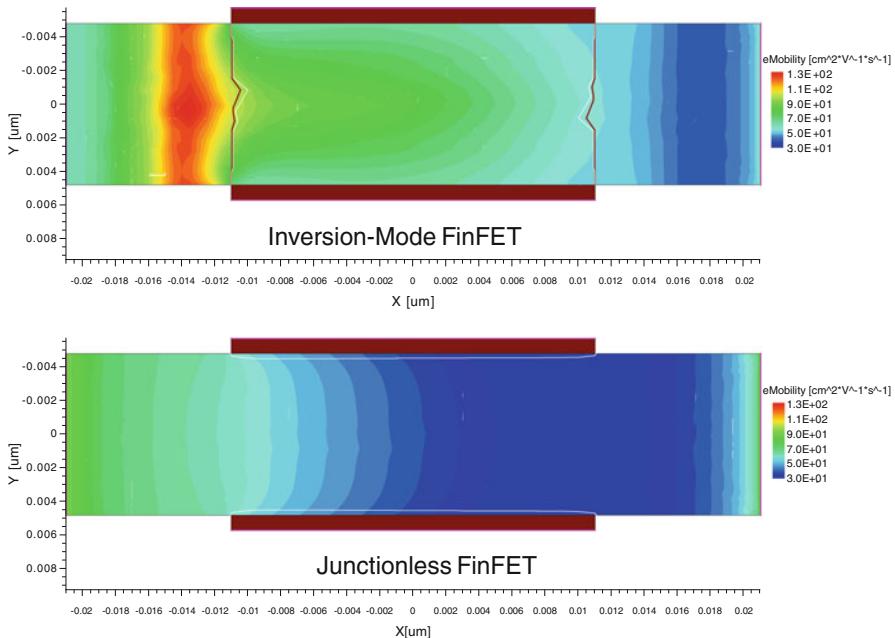
Quantity	Technology node			Description
	32nm	21nm	15nm	
$L_g$ (nm)	22	17	13	Physical gate length
EOT (nm)	0.90	0.77	0.64	Equivalent oxide thickness
$N$ ( $\text{cm}^{-3}$ )	$2 \times 10^{19}$	$2 \times 10^{19}$	$2 \times 10^{19}$	Body/fin doping
$T_{\text{fin}}$ (nm)	9.6	8	6.4	Fin thickness
$L_{\text{sp}}$ (nm)	10	8	6	Spacer width
$\Psi_M$ (eV)	5.25	5.02	4.82	Gate work function
$V_{\text{DD}}$ (V)	0.9	0.81	0.73	Power supply voltage
$V_{T,\text{lin}}$ (mV)	306	306	300	Lin. threshold voltage (max $g_m$ method with $V_{DS} = 50$ mV)
$V_{T,\text{sat}}$ (mV)	200	192	185	Sat. threshold voltage (const. $I = W/L_g \times 10^{-7}$ A with $V_{DS} = V_{DD}$ )
$I_{\text{on}}$ ( $\mu\text{A}/\mu\text{m}$ )	1,144	1,225	1,330	On-state drive current with $V_{GS} = V_{DS} = V_{DD}$
$I_{\text{off}}$ ( $\text{nA}/\mu\text{m}$ )	11.3	21.3	36.4	Off-state leakage current with $V_{GS} = 0$ and $V_{DS} = V_{DD}$
SS (mV/dec)	72.5	74.2	75.3	Subthreshold swing
DIBL (mV/V)	77.3	89.8	95.6	Drain-induced barrier lowering

[39] Copyright © 2011 IEEE

By applying different gate voltage values, the depletion region under the gate either pinches off the channel in the “off” state or opens up a buried channel in the “on” state. For an  $n$ -type JL-FET, the combination of a low gate voltage and an appropriate gate work function results in total depletion of the channel and an energy barrier for carriers between the source and drain. When a high gate voltage is applied, the depletion regions retract and the energy barrier vanishes, resembling flat-band (i.e., resistor-like) conditions in the channel. As a result, the JL-FET normally operates as a depletion-mode device rather than an IM device. The buried channel nature of JL-FETs is different from the surface channel nature of IM-FETs—this will have important consequences that will become apparent in later sections, especially when we discuss their performance vulnerability to LER and RDF.

The JL-FET structure modeled in this work resembles the same FinFET structure in Fig. 7.8, except for a different doping profile in the fin as shown in Fig. 7.17. In essence, the FET technology considered here will be JL-FinFETs designed to meet the same sub-32nm ITRS nodes for high-performance logic as before. This allows us to draw fair comparisons between the inherent advantages and disadvantages of JL and IM technologies when designed for the same physical layouts and operational targets.

Table 7.3 lists the nominal parameters and performance metrics for the JL-FinFETs considered in this work. As was the case in Sect. 7.3.1, only  $n$ -type devices will be simulated here. Comparing these values with those of Table 7.1 for



**Fig. 7.18** Electron mobility plots in 32nm IM- and JL-FinFETs at  $V_{GS} = V_{DS} = V_{DD} = 0.9$  V. The mobility is consistently higher in IM-FinFETs compared to JL-FinFETs due to reduced impurity and surface roughness scattering at these geometries

IM-FinFETs, the only design differences lie with the nominal doping concentration ( $N = 2 \times 10^{19} \text{ cm}^{-3}$  for JL-FinFETs) which is the same for the source, drain, and channel regions, and the gate work functions which are adjusted for each node to obtain  $V_{T,\text{sat}} \cong 0.2$  V with  $I_{\text{off}} < 100 \text{ nA}/\mu\text{m}$  according to the ITRS definition. The baseline performance values for JL-FinFETs remain comparable to those of IM-FinFETs, despite being marginally worse in all regards.  $I_{\text{on}}$  is about 20% worse and  $I_{\text{off}}$  is about 50% worse for JL-FinFETs compared to IM-FinFETs at the same generation, and JL-FinFETs consistently have higher SS and DIBL. These conclusions are well explained by the nature of buried channel formation in JL devices, resulting in weaker gate-channel capacitive coupling leading to worse SCE control. In a JL device, the highly doped channel results in significant mobility degradation due to impurity scattering despite the reduction in surface roughness scattering from the reduced transverse electric field above threshold. This is a necessary tradeoff to prevent excessive current loss from parasitic resistance if a lower channel doping were used. For IM-FinFETs, mobility degradation due to surface roughness scattering at maximum gate voltage is mitigated at the geometries considered due to volume inversion in the channel, resulting in overall higher channel mobility (Fig. 7.18) compared to JL-FinFETs; this may explain the higher  $I_{\text{on}}$  compared to JL-FinFETs.



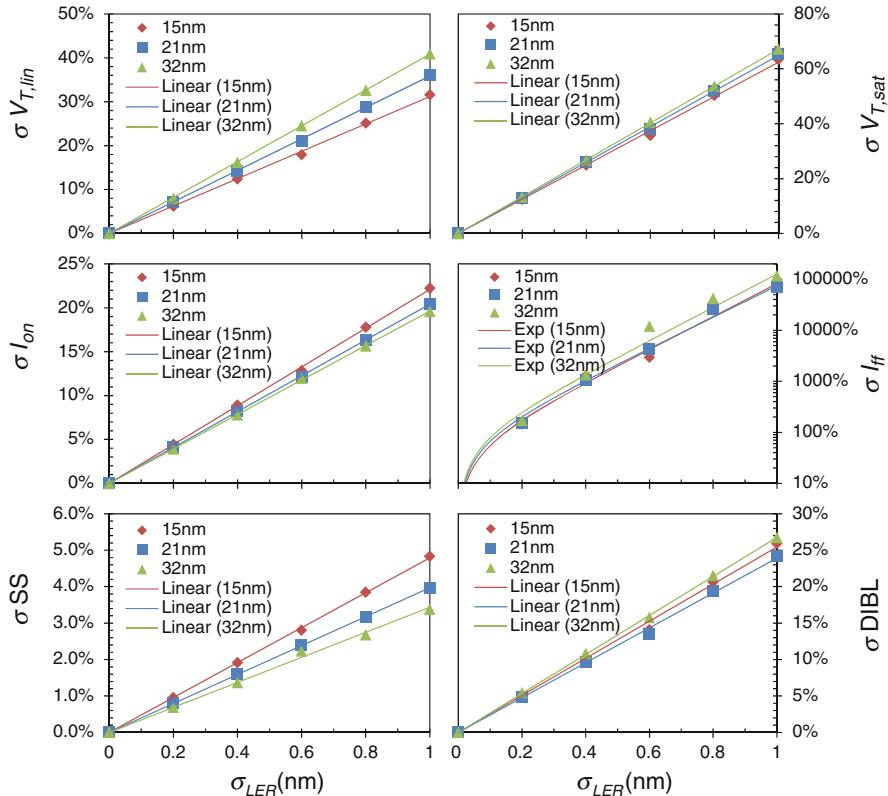
**Fig. 7.19** Representative 32nm JL-FinFET with 1nm LER applied to the fin edges

Nevertheless, the comparable level of performance between JL- and IM-FinFETs should hold regardless of the actual FET architecture, especially if the operation of JL-FETs is extended beyond depletion to accumulation, thereby blurring the distinction between a strictly buried channel JL device and a hybrid buried-surface channel device. Recent experimental evidence from Intel [62] supports our conclusions—however, we will still restrict the scope of our analysis in this section to depletion-mode JL devices only. The same set of device simulation models are used for our JL-FinFET simulations as they were for IM-FinFETs, namely the HD transport model, DGA for quantum corrections, and mobility models accounting for surface roughness scattering, doping dependence, and HD transport.

### 7.3.2.2 LER Impact on JL-FinFET Variability

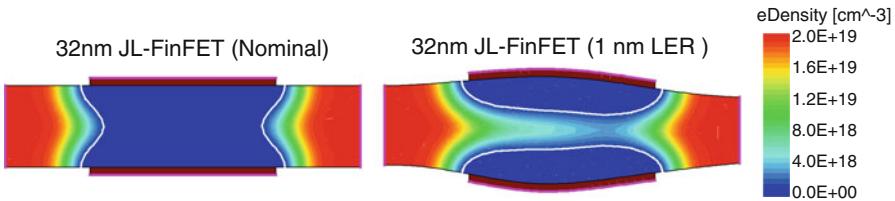
Using the same set of design guidelines and LER models/parameters as before for JL-FinFETs as we did for IM-FinFETs, we can make an apples-to-apples comparison to determine whether JL-FinFETs are any more or less vulnerable to LER than IM-FinFETs are under the same process technology constraints. To model LER in JL devices, we follow the exact same approach described earlier in Sect. 7.3.1.2 and obtain ensembles of JL-FinFETs with random (fin) LER profiles for various roughness amplitudes of  $\sigma_{\text{LER}} \leq 1\text{nm}$  and  $\lambda = 15\text{nm}$  for each technology node (32, 21, and 15nm). An example of a 32nm JL-FinFET with LER is shown in Fig. 7.19.

The extracted performance variations (as percentages of their nominal values from Table 7.3) due to LER for resist-defined JL-FinFETs are shown in Fig. 7.20 as a function of  $\sigma_{\text{LER}}$  and technology node. The functional trends remain identical as for the case of IM-FinFETs: linear versus  $\sigma_{\text{LER}}$  for all metrics except for  $\sigma I_{\text{off}}$  which is exponential. Since JL-FinFETs, like their IM counterparts, are MOSFET-inherited designs, this is not surprising. We find, however, that JL-FinFET variability from LER (Fig. 7.20) is substantially worse than for IM-FinFETs (Fig. 7.11), especially in terms of  $V_{\text{T,lin}}$ ,  $V_{\text{T,sat}}$ ,  $I_{\text{on}}$ , and  $I_{\text{off}}$ . We see that  $\sigma V_{\text{T,sat}}$  already exceeds



**Fig. 7.20** Resist JL-FinFET device variability as a function of LER amplitude and technology node. Markers indicate actual simulated data while solid lines indicate best fits [39]. Copyright © 2011 IEEE

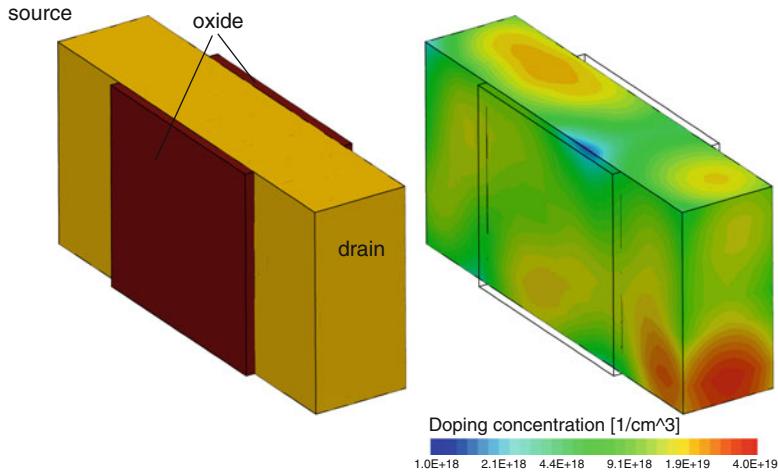
60% for  $\sigma_{LER} = 1\text{nm}$ ; by comparison this value was only 5–10% for IM-FinFETs. At the same LER amplitude, literature values for gate LER-induced  $\sigma V_{T,\text{sat}}$  are in the range of 2–8% for  $L_g = 30\text{nm}$  planar MOSFETs [10] and IM-FinFETs [44]. If one were to operate within a  $\sigma V_{T,\text{sat}} \leq 20\%$  limit as suggested by the ITRS, then  $\sigma_{LER}$  would need to be kept at or below  $0.2\text{nm}$ —a major burden on state-of-the-art lithography. This partially agrees with data in [57] where JL-FETs fabricated on SOI wafers exhibiting surface roughness  $\sigma T_{Si} \leq 0.2\text{nm}$  could expect  $\sigma V_T = 20\text{ mV}$  ( $V_T \simeq 0.2\text{ V}$ ). Additionally, with such high  $\sigma V_{T,\text{sat}}$  values for JL-FinFETs, it becomes very probable that some devices will have a negative threshold voltage. If only positive voltages are available, in some cases the extracted  $I_{off}$  no longer represents an “off-state” current. With this in mind, the actual leakage current values in Fig. 7.20 may be better interpreted as a minimum attainable current rather than a subthreshold current. Regardless, the key point is that LER has a significant impact on JL-FinFET variability, especially when compared against equivalent IM-FinFETs.



**Fig. 7.21** Electron density plots for two representative 32nm JL-FinFETs showing the inadvertent formation of a conducting channel due to fin LER at  $V_{GS} = 0.1$  V and  $V_{DS} = 0$ . White lines indicate depletion region boundaries [39]. Copyright © 2011 IEEE

At this point, we may wonder why JL-FinFETs exhibit much higher variability from LER compared to IM-FinFETs. In fact, the answer is intuitively simple. In Sect. 7.3.1.2, we described how an IM-FinFET fundamentally operates as a switch when the gate voltage switches from a low value ( $V_G < V_T$ ) to a high value ( $V_G \geq V_T$ ), resulting in the formation of a conducting inversion layer to bridge the source–channel and drain–channel  $p$ – $n$  junctions. Ignoring SCE for the moment, a potential barrier between source and drain is guaranteed in subthreshold by the presence of the back-to-back  $p$ – $n$  junctions between the source, channel, and drain. In this case, the presence of a subthreshold barrier is not threatened by body or fin thickness variations due to LER, since the subthreshold barrier is ensured by the oppositely doped source and channel regions. For the short channel IM-FinFETs studied in Sect. 7.3.1, DIBL will cause some sensitivity to LER in subthreshold and is the agent responsible for the LER-induced variability presented in Fig. 7.11. To reiterate, however, the existence of a subthreshold barrier is still ensured by the presence of a junction, thus making it robust to LER.

On the other hand, for JL-FinFETs (which operate as depletion-mode devices), the existence of a subthreshold barrier is stipulated on the body/channel being fully depleted—a condition which cannot always be guaranteed. Figure 7.21 reveals how variations in the body thickness from fin LER may inadvertently cause a conducting channel to form near the midsection to overcome the barrier, thereby driving the transistor out of subthreshold. For long-channel devices, this conduit will be entirely responsible for the resulting LER-induced variability (which will be significant). For short-channel devices, the LER contribution from SCE will also add to the net variability. However, the similarity between the 32, 21, and 15nm JL-FinFET curves implies that the dominant mechanism is not acting through SCE as it was for IM-FinFETs, but instead through the inadvertent opening of a conducting channel. Furthermore, the fact that peak variations in SS and DIBL due to LER remain fairly similar between JL- and IM-FinFETs further indicates that SCE degradation is not the major issue in JL devices. Rather, the (inadvertent) direct opening/closing of a conducting channel due to LER is reminiscent of a primary weakness in depletion-mode FETs: inherent dependence on a delicate balance between electrostatic and geometrical control. For this reason, LER has a primary effect on JL-FinFET variability.



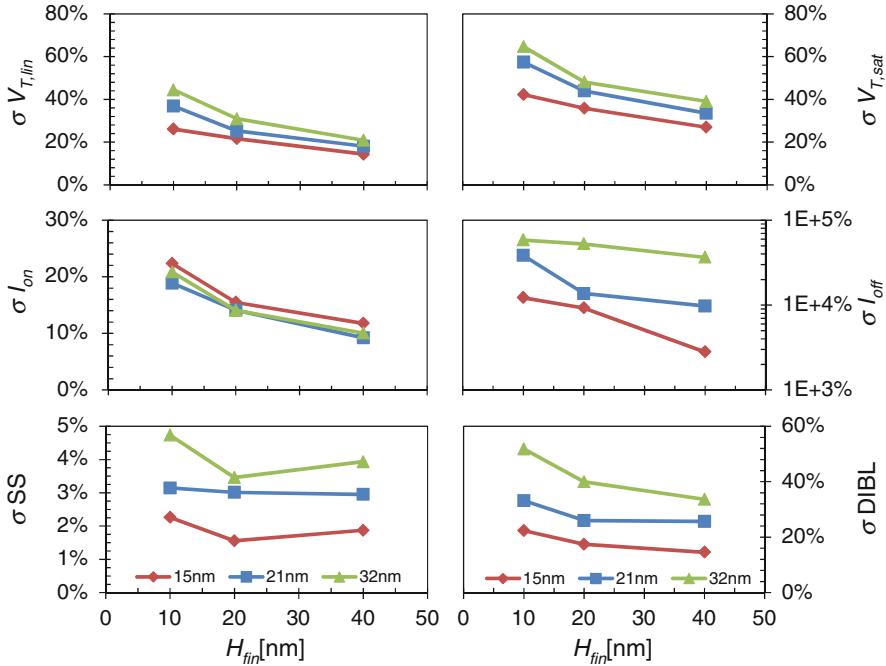
**Fig. 7.22** 32nm JL-FinFETs with and without RDF applied. The effective doping concentrations in both cases are shown with the same legend

because it directly jeopardizes transistor operation by negating proper geometrical control. This is in direct contrast with IM-FinFETs which do not depend on geometrical control to ensure correct operation and for which LER remains a secondary effect only. With this in mind, we see that by virtue of its characteristic nature, JL devices are not robust against LER whereas IM devices are, and its implications can be directly observed from Figs. 7.11 and 7.20. Despite focusing on double-gate JL-FinFETs, the same behavior should occur for TG JL-FETs and JL-NWFETs operating as depletion-mode devices, based on similar conclusions [60] drawn from experimental data.

### 7.3.2.3 RDF Impact on JL-FinFET Variability

Junctionless transistors are normally designed with a high doping concentration and a small channel volume, meaning RDF effects can become significant especially at sub-32nm nodes.<sup>3</sup> Using the same RDF modeling approach as was done for IM-FinFETs in Sect. 7.3.1.3, we will again perform an apples-to-apples comparison with JL-FinFETs to determine just how significant a role RDF will play in dictating overall device variability for JL-FinFETs compared to IM-FinFETs. Figure 7.22 depicts how RDF can be observed in the source, drain, and channel regions of a JL-FinFET compared to the situation for IM-FinFETs in which RDF only occurs in the source and drain.

<sup>3</sup>To illustrate this, the nominal dopant count in a JL-FinFET can range from ~300 total ions (32 nm node with  $H_{\text{fin}} = 40 \text{ nm}$ ) to only 3 ions (15 nm node with  $H_{\text{fin}} = 10 \text{ nm}$ ).



**Fig. 7.23** RDF-induced variability in JL-FinFETs as a function of fin height and technology node [22]. Copyright © 2012 IEEE

The RDF impact on JL-FinFETs is presented in Fig. 7.23 in the same manner as was done for IM-FinFETs in Fig. 7.14 for a direct comparison. We again see that increasing the fin height results in more self-averaging of RDF effects in each device in accordance with Pelgrom's law (7.7). A direct comparison of Figs. 7.14 and 7.23 also shows that RDF-induced variability is much more pronounced in JL-FinFETs compared to IM-FinFETs. Peak values of  $\sigma V_{T,sat}$  reach up to 60% for JL-FinFETs but are kept under 5% for IM-FinFETs. For  $\sigma I_{off}$ , we observe values on the order of 1,000–50,000% in JL-FinFETs compared to <100% for IM-FinFETs. In fact, for JL-FinFETs the variability magnitudes from RDF (Fig. 7.23) are very similar to the peak values from LER (Fig. 7.20). This is a clear indication that the operational integrity of JL devices is being threatened by RDF as well as LER.

Interestingly, we find that RDF-induced variability is actually reduced for most performance figures (except  $I_{on}$ ) in smaller technology nodes (15nm) compared to larger technology nodes (32nm); this is somewhat counterintuitive since variability normally becomes worse with technology scaling. This reverse scaling trend has been identified in other junctionless studies as well and has to do with the enhanced gate control over the channel when the body thickness  $T_{fin}$  is reduced in smaller technologies. Essentially, the effects of channel doping fluctuations are suppressed

by the improved gate control and smaller depletion regions surrounding the buried channel, allowing for more direct gate-to-channel control with fewer dopants “in the way.” This phenomenon has been observed in sensitivity analyses of parameters such as  $V_T$  to  $N$  (nominal) at different fin/body dimensions, with smaller body dimensions resulting in a flatter  $V_T$  vs.  $N$  profile (i.e., less sensitivity to RDF at smaller technology nodes) [63]. For  $\sigma I_{on}$ , there does not appear to be any meaningful difference between the 32, 21, and 15nm nodes; this may be due to added mobility variation due to RDF in smaller technologies to compensate for the lower amount of variation in  $V_T$ .

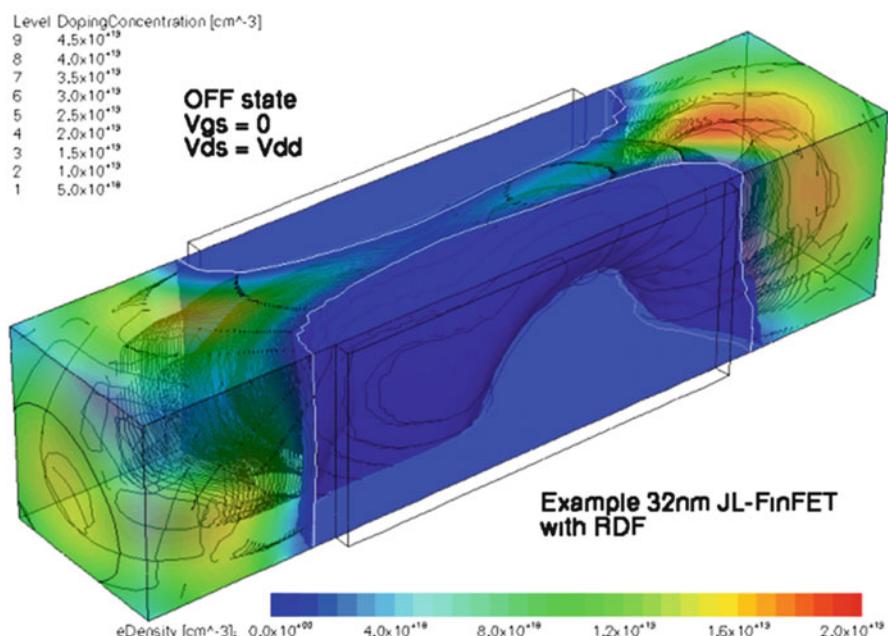
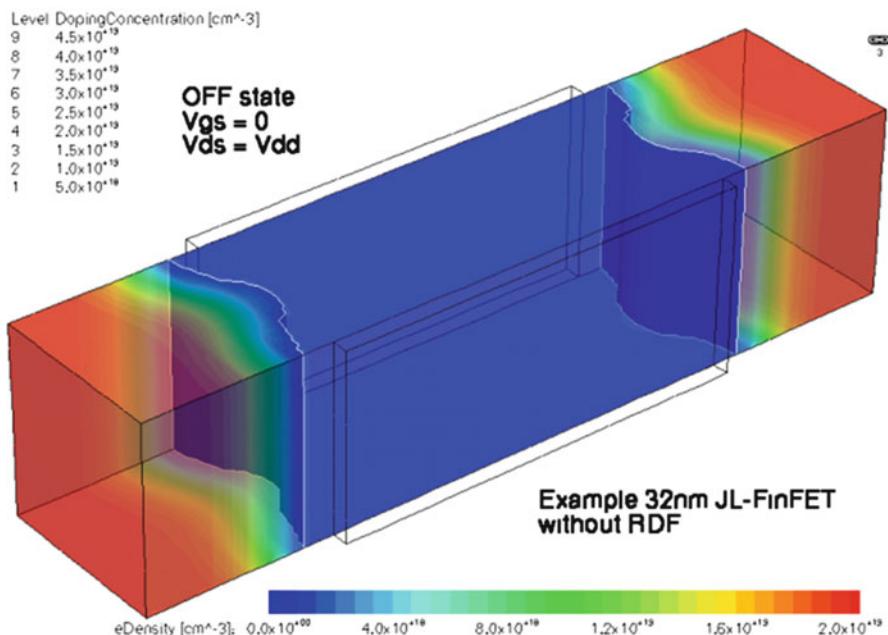
To understand why RDF has such a large impact on JL-FinFET performance, we must remember how depletion-mode JL transistors fundamentally operate. Recall from the previous section that LER potentially caused unwanted opening/closing of a conducting channel (Fig. 7.21), depending on the exact LER profiles along the fin sidewalls. In similar fashion, the exact number and positioning of ionized impurities can also result in fluctuation of the size and shape of the depletion region inside the fin. This, again, causes the buried channel to undulate with the topography of the randomized dopant profile (Fig. 7.22), and in some cases the buried channel may be undesirably opened or closed as a result. For example, a JL-FinFET with too many dopants won’t be fully depleted at zero gate bias and hence will remain “on” instead of “off” as shown in Fig. 7.24. Conversely, a JL-FinFET with too few dopants could remain fully depleted even at max gate bias and hence will remain “off” instead of “on.” This is another characteristic weakness of depletion-mode JL transistors which rely on a delicate balance between electrostatic and geometric control for correct operation.

### 7.3.3 Summary and Outlook

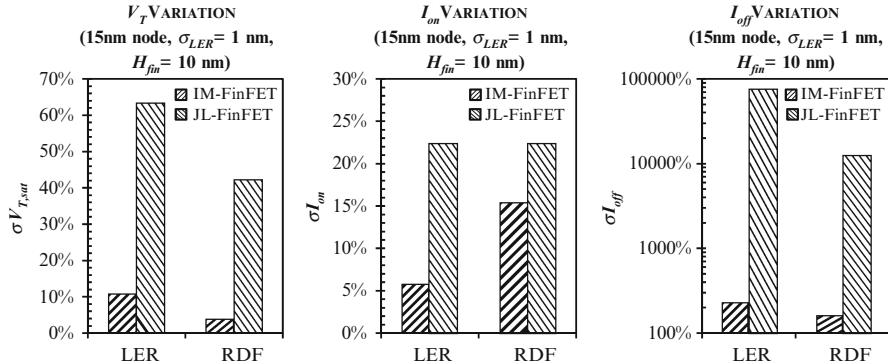
Up to this point, we have discussed in great length how sources of process variation including LER and RDF can cause random device variability in different FinFET technologies designed for near-future generations. A small subset of these results is shown in Fig. 7.25. In this final section we will briefly summarize the key points of the chapter and present an outlook on the viability of inversion-mode and junctionless FinFET technologies with considerations of technology scaling, manufacturability, and variability effects.

#### 7.3.3.1 Inversion-Mode FinFETs

The impact of LER on IM-FinFET variability is well managed at the 32, 21, and 15nm nodes, with smaller nodes (e.g., 15nm) exhibiting more variation than larger nodes (e.g., 32nm). (Fin) LER in IM-FinFETs results in nonuniform fluctuation of the fin/body thickness in individual devices, which primarily affects transistor SCE control. Resist-defined IM-FinFETs exhibit linear performance variation (except



**Fig. 7.24** Electron density plots in a representative 32nm JL-FinFET ( $H_{\text{fin}} = 10\text{nm}$ ) with and without RDF, showing the inadvertent formation of a conducting channel in the off state due to a surplus of dopants in the channel for the device with RDF



**Fig. 7.25** Comparison of device variability in IM- and JL-FinFETs due to LER and RDF at the 15nm technology node

for  $I_{off}$ ) versus LER amplitude, and fluctuation in most performance metrics are kept to a reasonable level. A maximum of 10%  $\sigma V_{T,sat}$  is obtained for 15nm resist IM-FinFETs at  $\sigma_{LER} = 1\text{nm}$ , and  $\sigma I_{on}$  is kept below 10% for all technology generations, both of which may be considered acceptable for logic applications. Variation in leakage current increases exponentially with  $\sigma_{LER}$ , with up to 250%  $\sigma I_{off}$  obtained for 15nm resist IM-FinFETs at  $\sigma_{LER} = 1\text{nm}$ . This is accompanied by an increase in mean leakage current which must be considered in a circuit's power budget design. For all performance metrics, the adoption of spacer lithography significantly alleviates the variability impact by eliminating LWR in individual devices and results in quadratic variability trends with  $\sigma_{LER}$ . From a standpoint of variability management, spacer lithography will be an indispensable manufacturing option for future generations.

The impact of RDF on IM-FinFET variability is also well managed, with smaller nodes again exhibiting more variation compared to larger nodes. Variability scaling in accordance with Pelgrom's law is observed with inverse dependencies appearing as a function of  $H_{fin}$ . Less than 5%  $\sigma V_{T,sat}$  is demonstrated while up to 15%  $\sigma I_{on}$  is obtained for 15nm IM-FinFETs with  $H_{fin} = 10\text{nm}$ , highlighting the different impacts from LER and RDF. RDF in the source and drain causes fluctuation in  $L_{eff}$  and becomes more significant for highly scaled generations with smaller nominal  $L_g$ .

Both LER- and RDF-induced variability impacts are considered secondary for IM-FinFETs and should not pose significant problems for near-term (15nm and below) technology adoption given current and projected manufacturing capabilities. On this basis, IM-FinFETs remain a viable option for continued scaling in the presence of manufacturing variability.

### 7.3.3.2 Junctionless FinFETs

The impact of LER on depletion-mode JL-FinFET variability is dangerously high at the 32, 21, and 15nm nodes under current and projected lithography capabilities ( $\sigma_{LER} \leq 1\text{nm}$ ). Fluctuation in body thickness from LER results in direct modulation of the size and shape of the buried channel in JL-FinFETs, leading to unwanted opening/closing of a conducting channel which destroys proper switching functionality. Little distinction is found between the variability magnitudes between different technology generations; this fact is related to the primary agent responsible for LER-induced variability in JL-FinFETs. A maximum of 60%  $\sigma V_{T,sat}$  is obtained for all JL-FinFET technologies at  $\sigma_{LER} = 1\text{nm}$  which is 5× higher than for 15nm IM-FinFETs.  $\sigma I_{on}$  also reaches up to 20% for all JL-FinFET technologies and is at least 3× higher than for IM-FinFETs. Leakage variation is exceptionally high with  $\sigma I_{off}$  approaching 100,000% along with a hefty increase in mean  $I_{off}$  when LER is present. These results indicate that JL-FETs may have great difficulty in meeting circuit requirements, especially those requiring precise matching of individual transistor characteristics. While not explicitly investigated in this work, the adoption of spacer lithography will likely alleviate the LER burden to levels closer to (but larger than) spacer IM-FinFETs.

The impact of RDF on JL-FinFET variability is also dangerously high for the technology nodes under consideration. For minimum height devices ( $H_{fin} = 10\text{ nm}$ ), performance variation from RDF is comparable to that from LER with 1nm amplitude:  $\sigma V_{T,sat}$  reaches up to 60% and  $\sigma I_{on}$  up to 20%. Such high levels of RDF variability arise from the dependency of JL-FinFETs on the depletion region profile, which can undulate strongly due to random placement and number of dopant ions. Counter intuitively, we find that 32nm JL-FinFETs exhibit more RDF-induced variability than 15nm JL-FinFETs, despite having a larger active volume and dopant count. The results are explained by noting that the closer gate-to-channel proximity in thin body devices (15nm) reduces its sensitivity to RDF by means of stronger gate-channel coupling.

Both LER- and RDF-induced variability impacts are considered primary for JL-FinFETs and will likely pose significant challenges for the near-term adoption of JL technology, unless significant improvements in lithography (i.e., reducing  $\sigma_{LER}$  or implementing spacer lithography) can be achieved. The key benefits of JL technology (e.g., manufacturing ease and scalability) must be weighed against the potential challenges associated with variability and lithography requirements to meet acceptable yields.

## References

1. Namatsu, H., Nagase, M., Yamaguchi, T., Yamazaki, K., Kurihara, K.: Influence of edge roughness in resist patterns on etched patterns. *J. Vac. Sci. Technol. B* **16**, 3315–3321 (1998)
2. Shin, J., Han, G., Ma, Y., Moloni, K., Cerrina, F.: Resist line edge roughness and aerial image contrast. *J. Vac. Sci. Technol. B* **19**, 2890–2895 (2001)

3. Sanchez, M.I., Hinsberg, W.D., Houle, F.A., Hoffnagle, J.A., Ito, H., Nguyen, C.V.: Aerial image contrast using interferometric lithography: effect on line-edge roughness. In: Proceedings of SPIE, pp. 160–171 (1999)
4. Leunissen, L.H.A., Ercken, M., Patsis, G.P.: Determining the impact of statistical fluctuations on resist line edge roughness. *Microelectron. Eng.* **78–79**, 2–10 (2005)
5. Reynolds, G.W., Taylor, J.W.: Factors contributing to sidewall roughness in a positive-tone, chemically amplified resist exposed by x-ray lithography. *J. Vac. Sci. Technol. B* **17**, 334–344 (1999)
6. Patsis, G.P., Constantoudis, V., Gogolides, E.: Effects of photoresist polymer molecular weight on line-edge roughness and its metrology probed with Monte Carlo simulations. *Microelectron. Eng.* **75**, 297–308 (2004)
7. Yoshizawa, M., Moriya, S.: Study of the acid-diffusion effect on line edge roughness using the edge roughness evaluation method. *J. Vac. Sci. Technol. B* **20**, 1342–1347 (2002)
8. Yahata, A., Urano, S., Inoue, T.: Striations on Si trench sidewalls observed by atomic force microscopy. *Jpn. J. Appl. Phys.* **36**, 6722 (1997)
9. International Technology Roadmap for Semiconductors. <http://www.itrs.net> (2011)
10. Asenov, A., Kaya, S., Brown, A.R.: Intrinsic parameter fluctuations in decanometer MOSFETs introduced by gate line edge roughness. *IEEE Trans. Electron Dev.* **50**, 1254–1260 (2003)
11. Diaz, C.H., Tao, H.-J., Ku, Y.-C., Yen, A., Young, K.: An experimentally validated analytical model for gate line-edge roughness (LER) effects on technology scaling. *IEEE Electron Device Lett.* **22**, 287–289 (2001)
12. Asenov, A., Brown, A.R., Davies, J.H., Kaya, S., Slavcheva, G.: Simulation of intrinsic parameter fluctuations in decanometer and nanometer-scale MOSFETs. *IEEE Trans. Electron Dev.* **50**, 1837–1852 (2003)
13. Sugimoto, Y., Pou, P., Abe, M., Jelinek, P., Perez, R., Morita, S., Custance, O.: Chemical identification of individual surface atoms by atomic force microscopy. *Nature* **446**, 64–67 (2007)
14. Hockney, R.W., Eastwood, J.W.: Computer simulation using particles. Taylor & Francis, New York (1998)
15. Asenov, A., Jaraiz, M., Roy, S., Roy, G., Adamu-Lema, F., Brown, A.R., Moroz, V., Gafiteanu, R.: Integrated atomistic process and device simulation of decanometre MOSFETs. In: Proceedings of SISPAD, pp. 87–90 (2002)
16. Roy, G., Brown, A.R., Adamu-Lema, F., Roy, S., Asenov, A.: Simulation study of individual and combined sources of intrinsic parameter fluctuations in conventional nano-MOSFETs. *IEEE Trans. Electron Dev.* **53**, 3063–3070 (2006)
17. Sano, N., Matsuzawa, K., Mukai, M., Nakayama, N.: On discrete random dopant modeling in drift-diffusion simulations: physical meaning of ‘atomistic’ dopants. *Microelectron. Reliab.* **42**, 189–199 (2002)
18. Mayergoyz, I.D., Andrei, P.: Statistical analysis of semiconductor devices. *J. Appl. Phys.* **90**, 3019–3029 (2001)
19. Wettstein, A., Penzin, O., Lyumkis, E., Fichtner, W.: Random dopant fluctuation modelling with the impedance field method. In: Proceedings of SISPAD, pp. 91–94 (2003)
20. Asenov, A.: Random dopant induced threshold voltage lowering and fluctuations in sub-0.1  $\mu\text{m}$  MOSFET’s: a 3-D ‘atomistic’ simulation study. *IEEE Trans. Electron Dev.* **45**, 2505–2513 (1998)
21. Damrongplasit, N., Shin, C., Kim, S.H., Vega, R.A., Liu, T.-J.K.: Study of random dopant fluctuation effects in germanium-source tunnel FETs. *IEEE Trans. Electron Dev.* **58**, 3541–3548 (2011)
22. Leung, G., Chui, C.O.: Variability impact of random dopant fluctuation on nanoscale junctionless FinFETs. *IEEE Electron Device Lett.* **33**, 767–769 (2012)
23. Leung, G., Chui, C.O.: Stochastic variability in silicon double-gate lateral tunnel field-effect transistors. *IEEE Trans. Electron Dev.* **60**, 84–91 (2013)

24. Pelgrom, M.J.M., Duinmaijer, A.C.J., Welbers, A.P.G.: Matching properties of MOS transistors. *IEEE J. Solid State Circ.* **24**, 1433–1439 (1989)
25. Mizuno, T., Okamura, J., Toriumi, A.: Experimental study of threshold voltage fluctuation due to statistical variation of channel dopant number in MOSFET's. *IEEE Trans. Electron Dev.* **41**, 2216–2221 (1994)
26. Stolk, P.A., Widdershoven, F.P., Klaassen, D.B.M.: Modeling statistical dopant fluctuations in MOS transistors. *IEEE Trans. Electron Dev.* **45**, 1960–1971 (1998)
27. Asenov, A., Kaya, S., Davies, J.H.: Intrinsic threshold voltage fluctuations in decanano MOSFETs due to local oxide thickness variations. *IEEE Trans. Electron Dev.* **49**, 112–119 (2002)
28. Dadgour, H., Endo, K., De, V., Banerjee, K.: Modeling and analysis of grain-orientation effects in emerging metal-gate devices and implications for SRAM reliability. In: IEDM technical digest, pp. 29.6.1–29.6.4 (2008)
29. Brown, A.R., Idris, N.M., Watling, J.R., Asenov, A.: Impact of metal gate granularity on threshold voltage variability: a full-scale three-dimensional statistical simulation study. *IEEE Electron Device Lett.* **31**, 1199–1201 (2010)
30. Brown, A.R., Roy, G., Asenov, A.: Poly-Si-gate-related variability in decanometer MOSFETs with conventional architecture. *IEEE Trans. Electron Dev.* **54**, 3056–3063 (2007)
31. Momose, H.S., Nakamura, S., Ohguro, T., Yoshitomi, T., Morifuji, E., Morimoto, T., Katsumata, Y., Iwai, H.: Study of the manufacturing feasibility of 1.5-nm direct-tunneling gate oxide MOSFET's: uniformity, reliability, and dopant penetration of the gate oxide. *IEEE Trans. Electron Dev.* **45**, 691–700 (1998)
32. Dadgour, H.F., Endo, K., De, V.K., Banerjee, K.: Grain-orientation induced work function variation in nanoscale metal-gate transistors—part I: modeling, analysis, and experimental validation. *IEEE Trans. Electron Dev.* **57**, 2504–2514 (2010)
33. Dadgour, H.F., Endo, K., De, V.K., Banerjee, K.: Grain-orientation induced work function variation in nanoscale metal-gate transistors—part II: implications for process, device, and circuit design. *IEEE Trans. Electron Dev.* **57**, 2515–2525 (2010)
34. Li, Y., Cheng, H.-W.: Random work-function-induced threshold voltage fluctuation in metal-gate MOS devices by Monte Carlo simulation. *IEEE Trans. Semicond. Manuf.* **25**, 266–271 (2012)
35. Ohmori, K., Matsuki, T., Ishikawa, D., Morooka, T., Aminaka, T., Sugita, Y., Chikyow, T., Shirashi, K., Nara, Y., Yamada, K.: Impact of additional factors in threshold voltage variability of metal/high-k gate stacks and its reduction by controlling crystalline structure and grain size in the metal gates. In: IEDM technical digest, pp. 1–4 (2008)
36. Grubbs, M.E., Zhang, X., Deal, M., Nishi, Y., Clemens, B.: Development and characterization of high temperature stable Ta-W-Si-C amorphous metal gates. *Appl. Phys. Lett.* **97**, 223505-1–223505-3 (2010)
37. Matsukawa, T., Liu, Y.X., Mizubayashi, W., Tsukada, J., Yamauchi, H., Endo, K., Ishikawa, Y., O'uchi, S., Ota, H., Migita, S., Morita, Y., Masahara, M.: Suppression of threshold voltage variability of double-gate fin field-effect transistors using amorphous metal gate with uniform work function. *Appl. Phys. Lett.* **102**, 162104 (2013)
38. Intel 22nm 3-D tri-gate transistor technology. Intel Press Release (2011)
39. Leung, G., Chui, C.O.: Variability of inversion-mode and junctionless FinFETs due to line edge roughness. *IEEE Electron Device Lett.* **32**, 1489–1491 (2011)
40. Leung, G., Lai, L., Gupta, P., Chui, C.O.: Device- and circuit-level variability caused by line edge roughness for sub-32-nm FinFET technologies. *IEEE Trans. Electron Dev.* **59**, 2057–2063 (2012)
41. Sentaurus TCAD User Manual. Synopsys, Inc., Mountain View, CA, Version F-2011.09
42. Nayfeh, O.M., Antoniadis, D.A.: Calibrated hydrodynamic simulation of deeply-scaled well-tempered nanowire field effect transistors. In: Proceedings of SISPAD, pp. 305–308 (2007)
43. Dixit, A., Anil, K.G., Baravelli, E., Roussel, P., Mercha, A., Gustin, C., Bamal, M., Grossar, E., Rooyackers, R., Augendre, E., Jureczak, M., Biesemans, S., De Meyer, K.: Impact of

- stochastic mismatch on measured SRAM performance of FinFETs with resist/spacer-defined fins: role of line-edge-roughness. In: IEDM technical digest, pp. 1–4 (2006)
44. Baravelli, E., Dixit, A., Rooyackers, R., Jurczak, M., Speciale, N., De Meyer, K.: Impact of line-edge roughness on FinFET matching performance. *IEEE Trans. Electron Dev.* **54**, 2466–2474 (2007)
45. Yu, S., Zhao, Y., Song, Y., Du, G., Kang, J., Han, R., Liu, X.: 3-D simulation of geometrical variations impact on nanoscale FinFETs. In: Proceedings of ICSICT, pp. 408–411 (2008)
46. Patel, K., King Liu, T.-J., Spanos, C.J.: Gate line edge roughness model for estimation of FinFET performance variability. *IEEE Trans. Electron Dev.* **56**, 3055–3063 (2009)
47. Xiong, W., Gebara, G., Zaman, J., Gostkowski, M., Nguyen, B., Smith, G., Lewis, D., Cleavelin, C.R., Wise, R., Shaofeng, Y., Pas, M., King, T.-J., Colinge, J.P.: Improvement of FinFET electrical characteristics by hydrogen annealing. *IEEE Electron Device Lett.* **25**, 541–543 (2004)
48. Choi, Y.-K., Chang, L., Ranade, P., Lee, J.-S., Ha, D., Balasubramanian, S., Agarwal, A., Ameen, M., King, T.-J., Bokor, J.: FinFET process refinements for improved mobility and gate work function engineering. In: IEDM technical digest, pp. 259–262 (2002)
49. Yahata, A., Urano, S., Inoue, T., Shinohe, T.: Smoothing of Si trench sidewall surface by chemical dry etching and sacrificial oxidation. *Jpn. J. Appl. Phys.* **37**, 3954 (1998)
50. Rauf, S., Stout, P.J., Cobb, J.: Modeling the impact of photoresist trim etch process on photoresist surface roughness. *J. Vac. Sci. Technol. B* **21**, 655–659 (2003)
51. Predictive Technology Model. <http://ptm.asu.edu>
52. <http://opencores.org>
53. ARM Cortex-M0 processor. <http://www.arm.com/products/processors/cortex-m/cortex-m0.php>
54. Power, J.A., Mathewson, A., Lane, W.A.: MOSFET statistical parameter extraction using multivariate statistics. In: Proceedings of ICMTS, pp. 209–214 (1991)
55. Takeuchi, K., Hane, M.: Statistical compact model parameter extraction by direct fitting to variations. *IEEE Trans. Electron Dev.* **55**, 1487–1493 (2008)
56. Yu, S., Zhao, Y., Du, G., Kang, J., Han, R., Liu, X.: The impact of line edge roughness on the stability of a FinFET SRAM. *Semicond. Sci. Technol.* **24**, 025005 (2009)
57. Colinge, J.-P., Lee, C.-W., Afzalian, A., Akhavan, N.D., Yan, R., Ferain, I., Razavi, P., O'Neill, B., Blake, A., White, M., Kelleher, A.-M., McCarthy, B., Murphy, R.: Nanowire transistors without junctions. *Nat. Nanotechnol.* **5**, 225–229 (2010)
58. Choi, S.-J., Moon, D.-I., Kim, S., Duarte, J.P., Choi, Y.-K.: Sensitivity of threshold voltage to nanowire width variation in junctionless transistors. *IEEE Electron Device Lett.* **32**, 125–127 (2011)
59. Kranti, A., Lee, C.-W., Ferain, I., Yu, R., Akhavan, N.D., Razavi, P., Colinge, J.-P.: Junctionless nanowire transistor (JNT): properties and design guidelines. In: Proceedings of ESSDERC, pp. 357–360 (2010)
60. Lee, C.-W., Ferain, I., Afzalian, A., Yan, R., Akhavan, N.D., Razavi, P., Colinge, J.-P.: Performance estimation of junctionless multigate transistors. *Solid State Electron.* **54**, 97–103 (2010)
61. Gundapaneni, S., Ganguly, S., Kottantharayil, A.: Bulk planar junctionless transistor (BPJLT): an attractive device alternative for scaling. *IEEE Electron Device Lett.* **32**, 261–263 (2011)
62. Rios, R., Cappellani, A., Armstrong, M., Budrevich, A., Gomez, H., Pai, R., Rahhal-orabi, N., Kuhn, K.: Comparison of junctionless and conventional trigate transistors with  $L_g$  down to 26 nm. *IEEE Electron Device Lett.* **32**, 1170–1172 (2011)
63. Trevisoli, R., Doria, R., de Souza, M., Pavanello, M.: Threshold voltage in junctionless nanowire transistors. *Semicond. Sci. Technol.* **26**, 105009-1–105009-8 (2011)

# Chapter 8

## Random Telegraph Noise in Multi-gate FinFET/Nanowire Devices and the Impact of Quantum Confinement

Runsheng Wang, Changze Liu, and Ru Huang

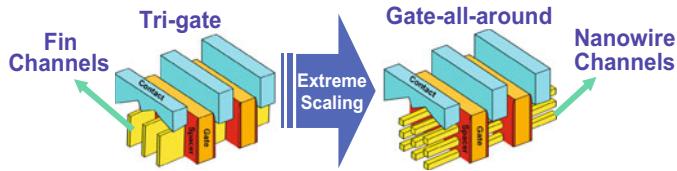
**Abstract** The multi-gate device with Fin- or nanowire-shaped channel is considered as the mainstream device structure towards the end of the CMOS technology roadmap. The performance of such ultra-scaled devices is strongly influenced by the localized oxide traps. One of the important effects is the random telegraph noise (RTN) in the device drain current, which is becoming worse with device scaling. The main part of this chapter discusses the impacts of nonnegligible quantum confinement in multi-gate devices on the RTN characteristics. Taking gate-all-around (GAA) Si nanowire FET (SNWT) as an example, the RTN statistical properties are found to be strongly impacted by the quantum confinement effects. The experimental results indicate that the strong quantum confinement in SNWTs enhances the bias dependence but alleviates the temperature dependence of RTN capture/emission time constants, which cannot be fully explained by classical RTN model based on Shockley–Reed–Hall (SRH) theory and should take into account the lattice relaxation and multiphonon processes. Thus, a full quantum RTN model for SNWTs is developed in this chapter, for fundamental understanding of the impacts of quantum confinement on RTN time constants, which agrees well the experimental results.

### 8.1 Introduction

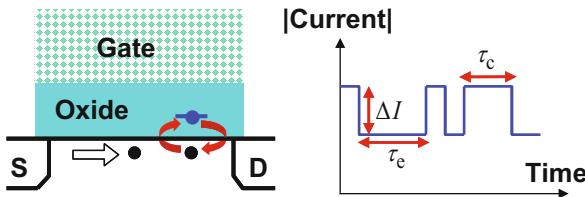
As the integrated circuit (IC) industry moving into the nanometer regime, the new structural multi-gate FET has attracted much attention for ultra-scaled CMOS technology, due to its superior gate control capability than conventional planar devices [1–15], as also discussed in some other chapters of this book. Recently, Intel has already commercialized tri-gate structure with Fin-shaped channel (i.e., tri-gate FinFET) in its advanced 22 nm products [16, 17]. Other major companies, such as

---

R. Wang (✉) • C. Liu • R. Huang  
Institute of Microelectronics, Peking University, Beijing 100871, China  
e-mail: [r.wang@pku.edu.cn](mailto:r.wang@pku.edu.cn)



**Fig. 8.1** From tri-gate Fin-FET to gate-all-around nanowire-FET



**Fig. 8.2** The illustration of drain current RTN in a MOSFET. The RTN parameters include amplitude ( $\Delta I$  or  $\Delta I/I$ ) and statistics in time: the average capture time (time to capture), or named capture time constant  $\langle \tau_c \rangle$ , and emission time constant  $\langle \tau_e \rangle$  (actually, most references omit the symbol  $\langle \rangle$  when indicating time constants  $\tau_c$  and  $\tau_e$ . In some of the contents of this chapter, we may also omit the  $\langle \rangle$  symbol for simplicity)

TSMC and IBM, have also announced their technology roadmaps to use multi-gate FinFET structure for their 16/14 nm node and beyond [18]. Therefore, the multi-gate device will be the mainstream technology until ultimately scaling.

Towards the ultimate scale, one can improve the gate architecture from double-gate or tri-gate to gate-all-around (GAA) for getting the strongest gate controllability [1–8, 19, 20], as shown in Fig. 8.1. Furthermore, with the feature size continuously shrinking, the Fin-channel thickness should be reduced accordingly, and will eventually become a nanowire-like geometry [21], which thus can be called as the Si nanowire FET (SNWT). In addition, the GAA SNWT with quasi-1D nanowire channel can achieve improved transport properties from volume inversion and quasi-ballistic transport [19, 22]. Therefore, in the relay race for scaling, the SNWT has been considered as the successor of FinFET, being the representation of extremely scaled multi-gate device at the end of the CMOS roadmap [1].

However, performance of ultra-scaled devices is strongly influenced by the localized effects such as the process-induced variation and random trap behavior [23–36]. The random telegraph noise (RTN) [37–55], as shown in Fig. 8.2, which manifests itself as a random switching of the current between two or several states due to capture/emission of carriers by localized oxide traps (named border traps or switching oxide trap and other names in the history), has been reported to severely impact the performance of flash memories, SRAM, logic circuits, and CMOS image sensor pixels [56–63]. And what is worse, the RTN amplitude increase rapidly with device scaling [37–40, 44–49], which will strongly impact the device reliability and circuit stability. Therefore, it is crucial to study the RTN in multi-gate FinFETs and SNWTs [55, 64–69].

For FinFET/nanowire devices, the channel is very thin and thus is strongly confined. Therefore, how the quantum confinement would affect the RTN characteristics in nano-scaled FinFET/nanowire devices should be paid special attention. In this chapter, we will take GAA nanowire devices as the extreme case of multi-gate devices to study this important issue [68].

In following parts of this chapter, at first a short review of the top-down fabrication method of GAA SNWTs will be given, followed by the typical characterization of RTN in SNWTs. Then the experimental results of the impacts of quantum confinement on RTN statistics will be presented and discussed. And a multiphonon-based quantum model for RTN in SNWTs will be developed for explaining the experimental results. Finally, the summary will be given.

## 8.2 From Multi-gate Fin-FETs to GAA Nanowire-FETs: A Brief Review on the Top-Down Fabrication Method

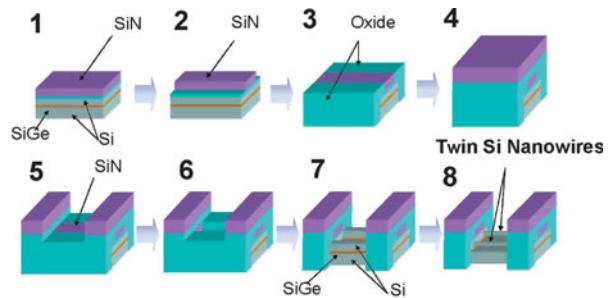
As feature size downscaling towards sub-10 nm, the multi-gate device will evolve into GAA nanowire structure type from FinFET as discussed above. Due to the surrounding gate structure, the most difficult step in fabricating a GAA SNWT is to form a suspending nanowire channel on the substrate. By using wet etching of oxide, it is relatively easy to form a nanowire on SOI substrate [70–72]. 25-Stage ring oscillator SNWT circuits have been demonstrated on SOI substrate with nanowire diameter down to 3 nm [73]. On bulk Si substrate, however, this process becomes difficult. The two typical fabrication methods of GAA SNWTs on bulk Si substrate are listed below.

One method was reported to adopt sacrificial layer method to form Si nanowire on bulk substrate [74, 75]. This method grows SiGe/Si multilayers on the bulk Si substrate, and then the SiGe under Si channel is selectively removed to release the nanowire, as shown in Fig. 8.3 [74]. SRAM circuits have been demonstrated by this method [76]. It has also been demonstrated that this method can be extended to sub-10 nm gate length [77], which shows great potential for manufacturing SNWTs at 11 nm technology node and beyond.

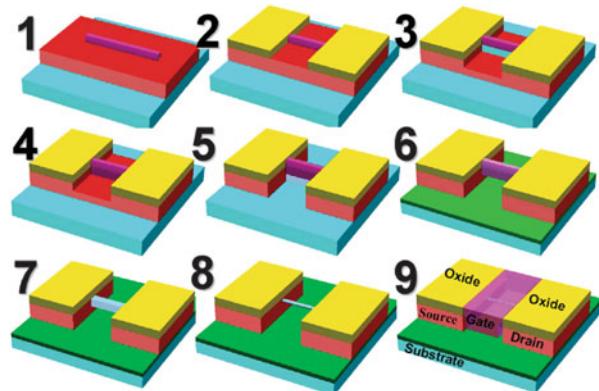
Another method was proposed to use self-limiting oxidation, which has saturation characteristics due to stress/temperature-dependent oxidation, to form the self-limited nanowire on the prepared silicon pillar [78]. As shown in Fig. 8.4, a vertical Si Fin is firstly formed, and the bottom of Fin is then isotropically etched with Fin sidewall protected by nitride spacer. After the nitride spacer is striped, a self-limiting oxidation process is then performed to trim and round the Fin body. Due to retardation effect of the self-limiting oxidation, the profile of released Fin will be changed into triangle firstly then cylinder to finally become a nanowire [78, 79]. Current mirror circuits have also been demonstrated based on this method [80].

Compared to the sacrificial layer method, the latter method shows higher compatibility to conventional CMOS process and can effectively avoid the junction

**Fig. 8.3** Schematic process sequence of GAA SNWT by SiGe scarification method on bulk Si substrate [74]



**Fig. 8.4** Schematic process sequence of GAA SNWT by self-limiting oxidation on bulk Si substrate [78]

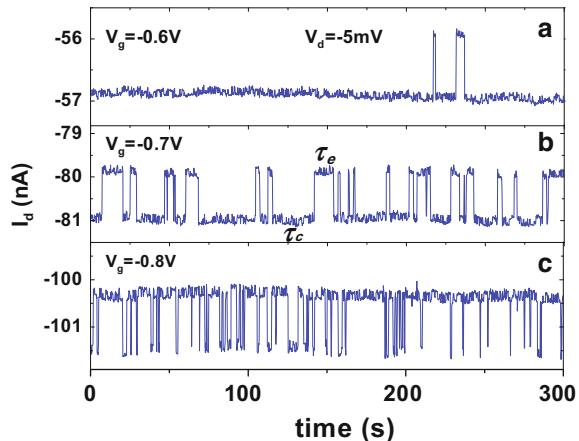


leakage issue resulting from the interface of the remaining SiGe and Si substrate in the sacrificial method.

### 8.3 Characterization of RTN in GAA SNWTs

The characterization techniques of RTN in MOS devices is essentially the previously measurement of small-current fluctuations. Conventional RTN characterization method is based on the low-noise  $I$ - $V$  preamplifier. The gate of the device under test (DUT) is biased by a DC voltage source; the drain is connected to a low-noise amplifier (e.g., SR 570) for drain biasing, amplifying, and converting drain current to voltage; then the voltage fluctuation is recorded by a dynamic signal analyzer (e.g., Agilent 35670A), which can also be sampled by a high-speed high-resolution digital multimeter (e.g., Agilent 3458A) alternatively or by a digital storage oscilloscope (with a large memory depth). The new technique for characterizing RTN can use an advanced fast measurement unit (e.g., Agilent B1530A WGFMU module), which has a wide bandwidth of up to 1 MHz and is easily set-up for biasing. Both techniques have their own advantages [40]: the

**Fig. 8.5** Typical measured RTN in a p-type SNWT with  $d_{\text{NW}} = 10 \text{ nm}$ ,  $L_g = 58 \text{ nm}$ , and  $V_d = 5 \text{ mV}$ . (a)  $V_g = -0.6 \text{ V}$ , (b)  $V_g = -0.7 \text{ V}$ , and (c)  $V_g = -0.8 \text{ V}$  [66]



conventional method has high precision and less drift, but the new method is ultrafast and with simple setup.

The extraction and analysis of the measured RTN raw data are based on statistical histogram, which set a criterion voltage to decide the state of RTN, or fit the data to Gaussian distribution [81]. This conventional method holds for most measurements in devices SiO<sub>2</sub> or SiON gate dielectrics. If the amplitude of background “ghost” noise is large, e.g., in some cases of measured RTN in high- $\kappa$  devices [49, 51, 52, 82], this method faces challenges. In that case, the extraction method can be improved based on clustering [83] and Hidden Markov Model [84], as in [51, 52, 82]. In this chapter, since all the samples are SiO<sub>2</sub> devices, the RTN data are very clear and can be extracted and analyzed by conventional method.

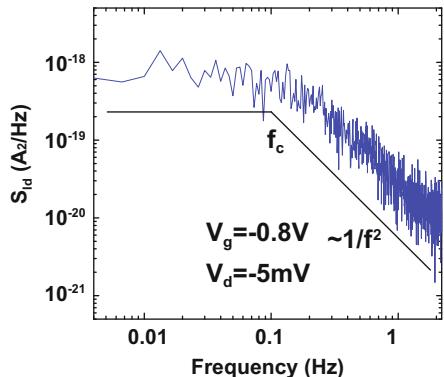
Figure 8.5 shows typical measured current fluctuations of a p-type SNWT in time domain [66], with nanowire diameter ( $d_{\text{NW}}$ ) of 10 nm and gate length ( $L_g$ ) of 58 nm, biased at drain voltage ( $V_d$ ) of -5 mV and different gate voltages ( $V_g$ ). Distinct switching between two states is observed, which can be ascribed to capture and emission behaviors of a single oxide trap. It is found that, with increasing the gate bias, the state with lower amplitude current dominates, indicating the repulsive nature of the charged oxide trap.

The corresponding current power spectral density is present in Fig. 8.6 by Fourier transformation, showing distinct Lorentzian spectrum:

$$S_{\text{Id}} = \frac{A^2}{1 + (f/f_c)^2}, \quad (8.1)$$

where the  $f_c$  is the corner frequency, and the parameter  $A$  is independent of frequency. These two parameters of  $S_{\text{Id}}$  in frequency domain can be expressed by the basic parameters of RTN in time domain as the following.

**Fig. 8.6** Current noise spectral density of RTN measured in Fig. 8.5c, indicating a Lorentzian spectrum



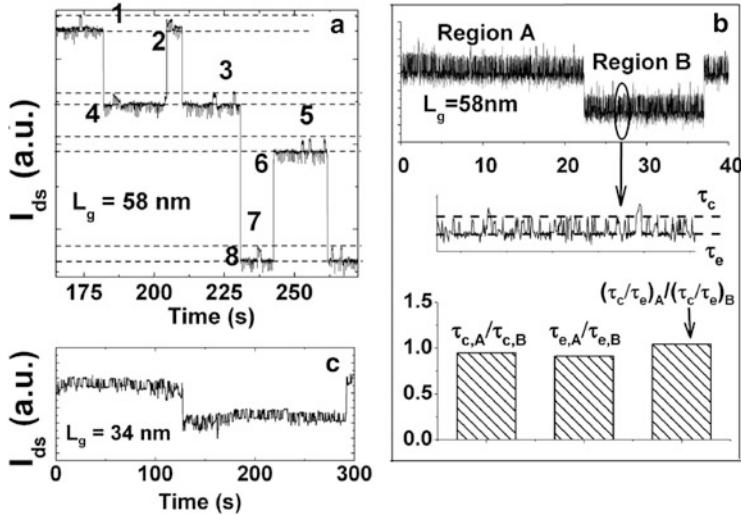
$$f_c = \frac{\tau_c + \tau_e}{2\pi\tau_c\tau_e}, \quad (8.2)$$

$$A^2 = \frac{(2\Delta I\tau_c\tau_e)^2}{(\tau_c + \tau_e)^3}. \quad (8.3)$$

The typical two-level RTN in Fig. 8.5 is caused by switching of a single trap. The multilevel RTN (also named as complex RTN) with several switching states are also observed, which is caused by the switching of multi-traps in one device, since the trap number in one device is randomly distributed in space. Figure 8.7a shows a multilevel RTN with eight switching states due to the capture/emission behaviors of three traps. Figure 8.7b shows four-state RTN in an n-type SNWT with  $L_g = 58$  nm, which contains a “fast” RTN with time constants on the order of milliseconds and a “slow” RTN with large time constants due to the slow trap. Region A and region B stand for the current states with the slow trap being neutral or charged, respectively. By extracting and comparing the time constants of the fast RTN between region A and region B, little difference can be found, indicating that the behavior of the fast trap is barely impacted by the state of the slow trap in this case.

The time constants of the oxide traps are found to widely spread from microseconds to hundreds of seconds. Figure 8.7c shows an example of a single switching event during about 300 s in a p-type SNWT with  $L_g = 34$  nm. As a result, special considerations should be made to take into account the long-time switching events in circuit designs, because the omitting of these events may lead to underestimation of the current fluctuation and give rise to unexpected functional error in circuits.

The complex RTN in Fig. 8.7b shows large amplitude up to 34 % at inversion region. And other groups also reported a giant complex RTN with huge slow RTN amplitude larger than 70 % in an n-type SNWT at near-threshold region [65], as shown in Fig. 8.8.



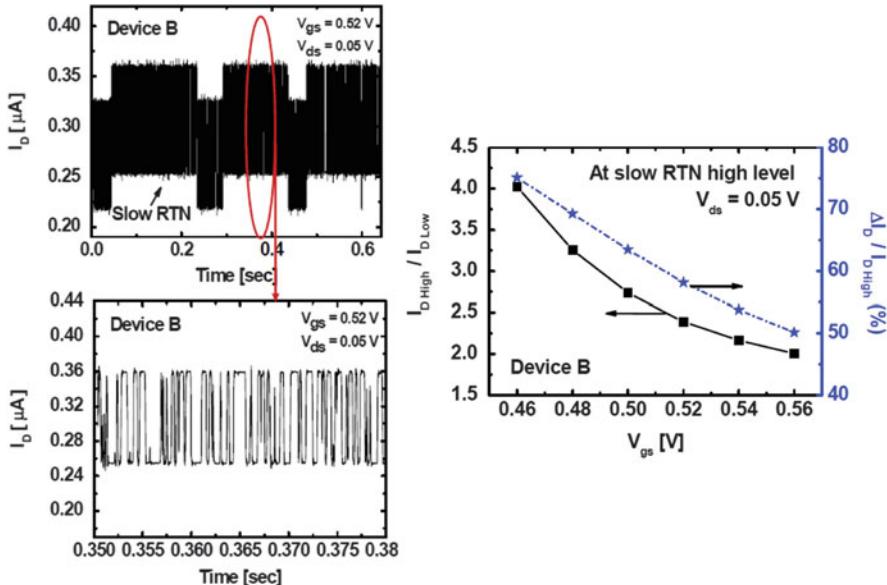
**Fig. 8.7** (a) Multilevel RTN caused by 3 traps in a p-type SNWT with  $L_g = 58$  nm,  $V_g = -0.6$  V,  $V_d = -5$  mV. (b) Both “fast” and “slow” RTN observed in an n-type SNWT with  $L_g = 58$  nm,  $V_g = 0.6$  V,  $V_d = 5$  mV (*upper figure*), and the time constants of the fast trap in region A and region B (*bottom figure*). (c) Slow RTN with time constant on the order of hundreds of seconds is observed [66]

## 8.4 Impacts of Quantum Confinement on RTN in GAA SNWTs

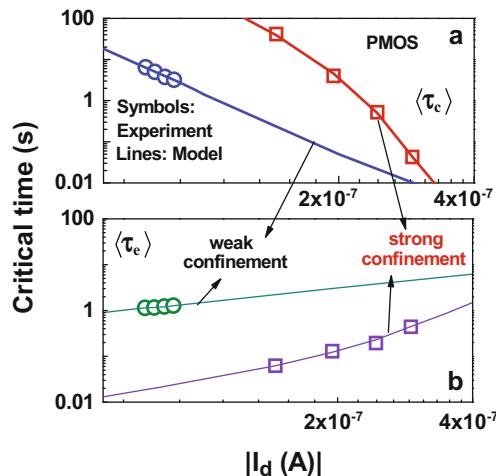
Since RTN is a statistical trapping/detrapping event in time, it is important to gain more understanding of the statistics of RTN (capture/emission probability) in multi-gate devices, which is the key information for robust circuit design against RTN. In this section, we will discuss the impacts of quantum confinement on RTN time constants in GAA SNWTs.

By comparing the RTN in SNWTs with both large diameter (weak confinement) and small diameter (strong confinement), one can directly observe the impacts of structural confinement in SNWTs [68].

Figure 8.9 shows the typical results of the measured RTN capture and emission time constants ( $\tau_c$  and  $\tau_e$ ) with varying drain current (i.e., gate bias) in p-type SNWTs with both large diameter (~80 nm for weak confinement) and small diameter (~10 nm for strong confinement) [68]. As shown in the figure, for both cases, the  $\tau_c$  reduces with increasing gate bias, and  $\tau_e$  increases with increasing gate bias, which is consistent with general observations in most SNWTs. However, it can be observed that the RTN in strongly confined SNWTs has stronger gate-bias dependence (i.e., larger slope in the figure) than those in weakly confined SNWTs. If using the relationship of  $\tau \sim I_d^m$ , one can compare the parameter  $m$  for RTN in SNWTs with different confinement levels. We further extracted the  $m$  values from all the measurement results of RTN in Table 8.1. It can be found that for both capture and emission, the bias dependence is enhanced by quantum confinement.



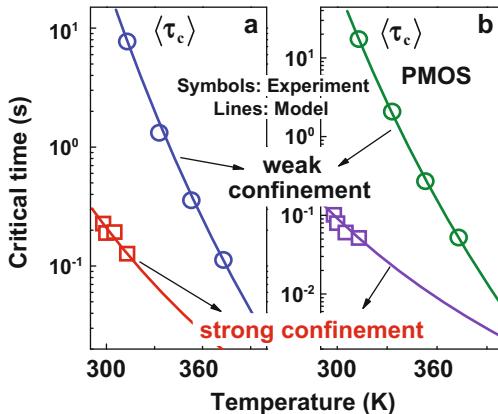
**Fig. 8.8** Four-level RTN observed in a n-type SNWT with  $V_d = 50 \text{ mV}$  and  $L_g = 40 \text{ nm}$  (*left*). And the slow RTN amplitude as the function of gate bias (*right*). Here (in *right figure*),  $I_{D \text{ Low}}$  and  $I_{D \text{ High}}$  are the current levels when fast trap is occupied and all traps are unoccupied, respectively [65]



**Fig. 8.9** Typical experimental results (*symbols*) of the measured (**a**) capture time constant  $\tau_c$  and (**b**) emission time constant  $\tau_e$  with varying drain current (i.e., gate bias) of RTN in SNWTs with large diameter ( $\sim 80 \text{ nm}$  for weak confinement) and small diameter ( $\sim 10 \text{ nm}$  for strong confinement) [68]. The *lines* are the proposed quantum RTN model for SNWTs in Sect. 8.5

**Table 8.1** The comparisons of the gate-bias dependence of RTN time constants in SNWTs with strong or weak confinement

<i>m</i>	Capture	Emission
Strongly confined	-9 to -10	2.7–4
Weakly confined	-4.5 to -6.5	0.8–1



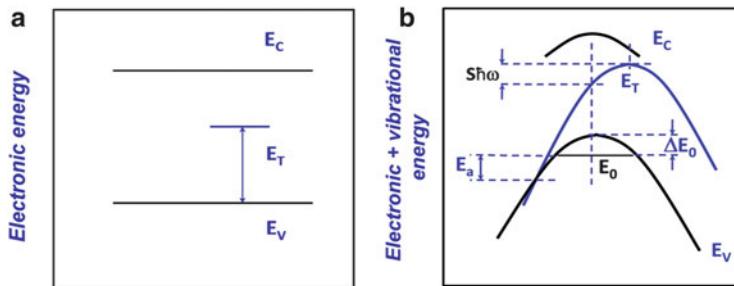
**Fig. 8.10** Typical experimental results (symbols) of the measured (a)  $\tau_c$  and (b)  $\tau_e$  with varying temperature of RTN in SNWTs with large diameter (~80 nm for weak confinement) and small diameter (~10 nm for strong confinement) [68]. The lines are the proposed full quantum RTN model for SNWTs in Sect. 8.5

**Table 8.2** The comparisons of the temperature dependence of RTN time constants in SNWTs with strong or weak confinement

Slope of $\text{Log}(\tau) \sim T$ curve	Capture or emission
Strongly confined	-0.015 to -0.018
Weakly confined	-0.03 to -0.04

Figure 8.10 shows the typical results of the measured RTN capture and emission time constants with varying temperature in p-type SNWTs with both large diameter (~80 nm for weak confinement) and small diameter (~10 nm for strong confinement) [68]. As shown in the figure, the time constants reduce with increasing temperatures, which is consistent with RTN in planar devices and indicates the trap switching behavior is a thermal activated process [37, 39]. It can also be observed that, the RTN in strongly confined SNWT has weaker temperature dependence (i.e., smaller slope in the figure) than that in weakly confined SNWT. We further extracted the slopes of the  $\text{Log}(\tau) \sim T$  curves from all the measurement results of RTN in Table 8.2. It can be found that for both capture and emission, the temperature dependence is weakened by stronger quantum confinement. The weak temperature dependence indicates a smaller thermal activation energy ( $E_a$ ) of trap switching.

To summarize the above observations, the experimental results indicate that the strong quantum confinement in GAA SNWTs enhances the bias dependence but alleviates the temperature dependence of RTN statistical properties.



**Fig. 8.11** (a) The conventional SRH generation-recombination framework for the trap-center assisted carrier transition. (b) The multiphonon-based framework for the trap-assisted carrier transition: in this example, the energy diagram is produced by an oxide hole trap; the total energy is the sum of hole energy and vibrational energy in p-type SNWTs. The  $x$ -axis is the configuration (or reaction) coordinate.  $E_a$  is the capture activation energy,  $S$  is the Huang–Rhys factor [87],  $S\hbar\omega$  is the semi-Franck–Condon energy shift [87] which reflects the lattice relaxation energy, and  $E_T$  is the oxide trap energy

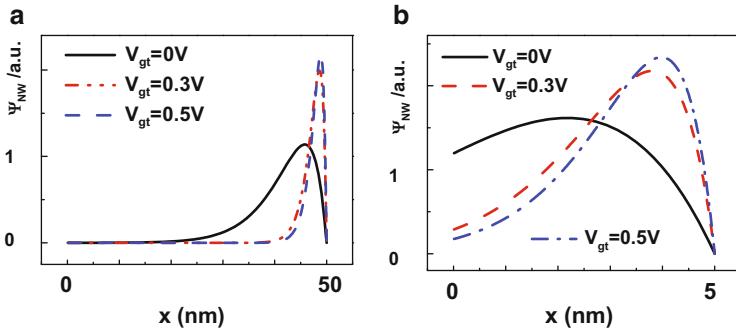
This observation cannot be explained by the classical RTN model based on Shockley–Reed–Hall (SRH) theory [85, 86], which is thus not suitable for understanding detailed physics of RTN in ultra-scaled multi-gate devices. As shown in Fig. 8.11a, in SRH framework, it only considers the electronic energy of the carrier (i.e., hole energy in this experiment for PMOS) and ignores the deformation energy. That is to say, the SRH theory considers the capture/emission behavior in RTN as the elastic tunneling process between the oxide trap energy level and valance band (for hole). However, recent studies [38, 39] found that the SRH theory no longer holds for scaled devices with ultrathin gate oxides, due to the fact that the tunneling process in oxides is actually inelastic tunneling.

To consider the inelastic process, one needs to take into account the lattice distortion after the trap capturing or emitting a carrier, which needs lattice relaxation energy. Since the lattice is evolved in actual capture process, phonons can assist in the inelastic tunneling process, which can be modeled by nonradiative multiphonon transition theory [39, 87–92]. Figure 8.11b gives an example of the energy diagram for multiphonon framework, in which the total energy, i.e., the sum of electronic energy of hole and the vibrational energy of lattice, is considered. For more information on the SRH and multiphonon models for RTN, one can consult a detailed review written by Grasser recently [39].

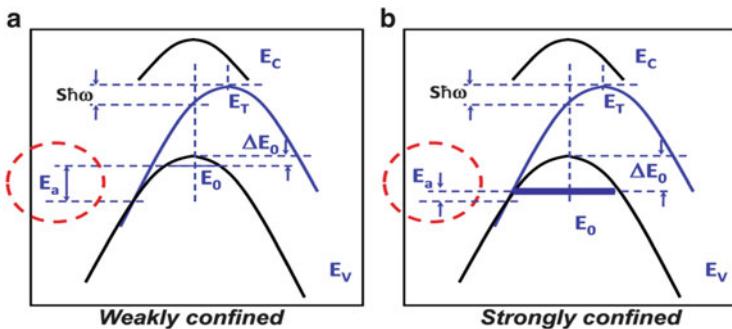
It should be noticing that, here since we are discussing hole (instead of electron), in order to better compare with Fig. 8.11a, the lattice energy curve bends downward in Fig. 8.11b, indicating the higher total energy downwards the  $y$ -axis. And the quantum confinement effect is reflected in Fig. 8.11b as well, which results in ground energy lifting and sub-band split, as the  $\Delta E_0$  in the figure.

Based on the above knowledge of the simple physical picture of the multiphonon-assisted capture/emission transitions, now we can qualitatively discuss the impacts of quantum confinement on RTN statistics observed above.

Let us first look at the gate-bias dependence of the RTN time constants. Since the physical process of RTN is the nonradiative transition of channel carrier to oxide



**Fig. 8.12** The calculated Si nanowire channel wave functions under different gate biases: (a) in a weakly confined SNWT with  $d_{NW} = 100$  nm and (b) in a strongly confined SNWT with  $d_{NW} = 10$  nm. Note:  $x$  is along the nanowire radius direction with  $x = 0$  locates at the nanowire center, and the gate overdrive voltage  $V_{gt} = V_g - V_t$



**Fig. 8.13** The illustrations of multiphonon processes in (a) weakly confined and (b) strongly confined SNWTs

trap, it is fundamentally determined by the wave functions of the channel and the trap. The quantum confinement of the channel will strongly affect the channel wave function. Figure 8.12 compares the calculated nanowire channel wave functions at different gate biases by solving 2D Schrodinger equations, in two extreme cases of strongly confined and weakly confined SNWTs. As shown in Fig. 8.12a, in the weak confinement case, the channel wave function only changes near the Si nanowire surface with varying gate bias. However, as shown in Fig. 8.12b, the nanowire wave function changes within the entire channel and thus exhibit stronger gate-bias dependence. That is why the RTN time constants are observed having stronger gate-bias dependence in more confined SNWTs with smaller nanowire diameters.

Now look at the temperature dependence of RTN time constants. To the first-order understanding from macroscopic perspective, since the temperature dependence reflects a thermal activation energy  $E_a$ , we can analyze  $E_a$  in the energy diagram of multiphonon processes for SNWTs with different confinements. As shown in Fig. 8.13, weak confinement results in smaller ground band lifting

(i.e., smaller  $\Delta E_0$ ), thus having larger  $E_a$  than strong confinement, which can cause larger temperature dependence observed above. It should be noted that, this analysis here is only a qualitative approach from macroscopic perspective. As will be discussed in the next section, the impact of quantum confinement on the temperature dependence fundamentally originates from the characteristics of the channel wave function and its interaction with the trap wave function.

## 8.5 A Multiphonon-Based Quantum Model for RTN Statistics in GAA SNWTs

As discussed in the last section, due to the blockade of phonon transport along gate stack direction and the lattice relaxation, the multiphonon process is nonnegligible in SNWTs, which assists the trap capture/emission and results in RTN, as illustrated in Fig. 8.14. Therefore, in order to understand the additional impacts of quantum confinement on RTN in SNWTs, a full quantum model based on multiphonon framework is developed in this section. We will take PMOS for example, so the hole trap is considered in the model, in which the empty trap is neutral and the charged trap is positive.

In order to accurately take into account the impacts of quantum confinement, we start from the fundamental wave function assumptions to build a full quantum RTN model in SNWTs.

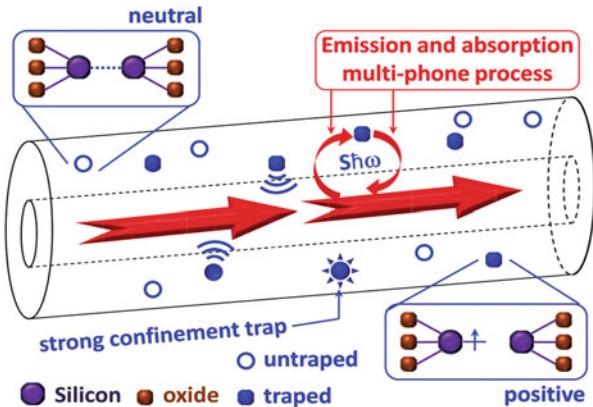
Figure 8.15 gives the basic modeling flow in this section. In general, to naturally incorporate the quantum effects, we introduce the wave functions of the nanowire channel and the oxide trap in calculating the nonradiative multiphonon transition rate. The nanowire wave function is also used for the calculation of the quantization-induced ground sub-band energy lifting in SNWTs. Note that, the lattice relaxation energy is introduced by the Huang–Rhys factor  $S$  [87] in the multiphonon model, as schematically shown in Fig. 8.11b. In addition, since the Coulomb blockade effect is important for RTN in nanoscaled devices, we also include the Coulomb energy ( $\Delta E$ ) [93]. The details of the model derivation are as the following.

### 8.5.1 Wave Functions Assumption

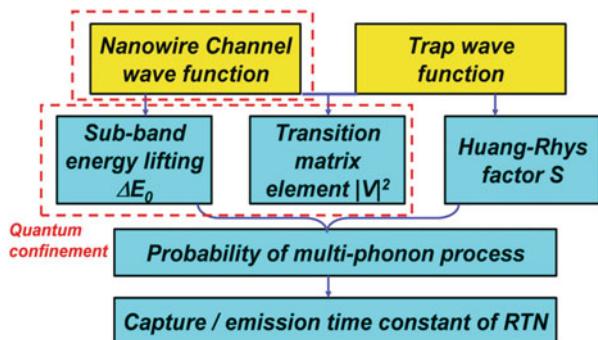
To discuss the hole tunneling process between the nanowire channel and the oxide trap state, the wave functions of them ( $|\psi_{\text{NW}}\rangle$  and  $|\psi_T\rangle$ ) should be defined separately, considering the different conditions of quantum confinement.

Considering the cylindrical symmetry of the nanowire channel in SNWTs, the problem can be simplified into 1D condition. The 1D channel wave function in SNWT for discussing the tunneling process is proposed based on variational approach [94], which can be described as:

**Fig. 8.14** The illustration for the major effects of oxide hole traps in p-type SNWTs: lattice relaxation, multiphonon process, quantum confinement, etc. [68] These effects will impact the trap capture/emission processes and thus alter the RTN properties in SNWTs



**Fig. 8.15** The basic modeling flow in this section



$$|\psi_{\text{NW}}(x)\rangle = a \sqrt{\frac{1}{R_{\text{eff}}}} \sin \left[ \frac{\pi(x + R_{\text{eff}})}{2R_{\text{eff}}} \right] \exp \left[ \frac{-b(-x + R_{\text{eff}})}{2R_{\text{eff}}} \right], \quad (8.4)$$

where  $x$  is along the nanowire radius direction with  $x = 0$  locates at the nanowire center. The normalization parameter is

$$a = \left( \frac{2b(b^2 + \pi^2)}{-2b^2 + (e^{-b} - 1)\pi^2} \right)^{1/2} \quad (8.5)$$

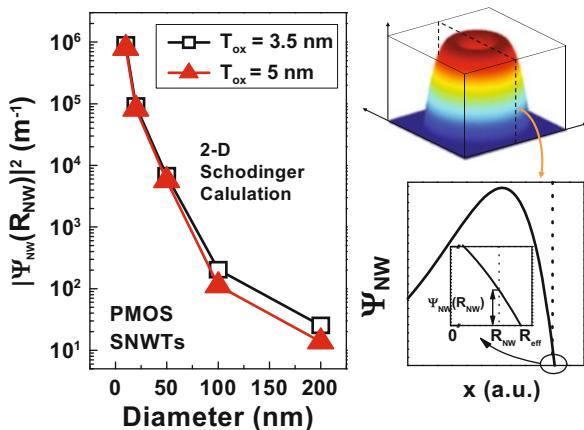
and the variational parameter is

$$b = b_0 R_{\text{NW}} (V_g - V_t)^{1/3}, \quad (8.6)$$

where  $R_{\text{NW}}$  is the nanowire radius,  $V_t$  is the threshold voltage of the device, and  $b_0$  is a fitting parameter.

If one assumes that the confinement potential is infinite, the  $R_{\text{eff}}$  in (8.4) should be equal to  $R_{\text{NW}}$ , which leads to  $\Psi_{\text{NW}}(R_{\text{NW}}) = 0$ . However, considering the real

**Fig. 8.16** The probability density at the Si surface of p-type SNWTs (*left*), calculated by 2D Schrodinger equation. The parameter  $R_{\text{eff}}$  in nanowire channel wave function can be obtained from the calculation (*right*) [68]



confinement potential is finite, the channel wave can penetrate into the oxide. As a result,  $R_{\text{eff}}$  can be larger than  $R_{\text{NW}}$  in this case, which leads to  $\Psi_{\text{NW}}(R_{\text{NW}}) > 0$ . In order to obtain the precise value of  $R_{\text{eff}}$ , we solve the 2D Schrodinger equation. As shown in Fig. 8.16, the results of probability density at Si channel surface can be derived with varying nanowire diameter. Due to the stronger confinement in narrower nanowire, the ground energy can be lifted leading to the increase of the surface probability density. Based on the results of Fig. 8.16, the  $R_{\text{eff}}$  can be determined and the gate bias induced wave function penetration can be estimated by the variational parameter  $b$  in (8.4).

For the oxide trap state, we define it as a solid box to consider the trap core effects [88]. Therefore, the trap wave function can be simply described as:

$$|\psi_T\rangle = \left(\frac{1}{z_T^3}\right)^{1/2} = \left(4\pi\left(\hbar/[2m_{\text{ox}}|E_{\text{vox}} - E_T|]^{1/2}\right)^3/3\right)^{-\frac{1}{2}}, \quad (8.7)$$

where  $m_{\text{ox}}$  is the hole effective mass in the gate oxide,  $E_{\text{vox}}$  is the oxide valence-band maximum, and  $E_T$  is the trap energy level.

Note that, considering the structure relaxation effects, the total energy (the sum of hole electronic energy and lattice vibrational energy) of valence band in nanowire channel or trap state in oxide is not a single horizontal level but a curve as a function of the configuration coordinate, as shown in Fig. 8.11b.

### 8.5.2 Key Parameters Calculations

As mentioned above, the strong confinement in small-diameter nanowire devices can result in nonnegligible ground energy lifting and sub-band split. For simplicity, we only consider the ground sub-band in the following calculation. The energy difference between the ground sub-band edge and valence-band of nanowire channel can be estimated as:

$$\Delta E_0 = |E_0 - E_V| = \left| \frac{\hbar^2}{2m_x} \int_0^{R_{\text{NW}}} \psi_{\text{NW}}(x) \frac{d^2 \psi_{\text{NW}}(x)}{dx^2} dx + q \int_0^{R_{\text{NW}}} \psi_{\text{NW}}^2(x) \phi(x) dx \right|, \quad (8.8)$$

where  $\phi(x)$  is the electrostatic potential in nanowire channel which can be estimated according to [25]. The calculated result of  $\Delta E_0$  is about 0.1 eV in the case of 10 nm diameter nanowire devices.

The transition matrix element, which is needed to characterize the carrier–phonon interaction, can be described as:

$$|V|^2 = |\langle \psi_{\text{NW}} | U | \psi_{\text{T}} \rangle|^2 \approx 2\pi S(\hbar\omega)^2 \frac{z_T^2}{A} \int_{d-z_T/2}^{d+z_T/2} |\psi_{\text{NW}}(z)|^2 dz = V_0 \begin{bmatrix} \exp\left(\frac{-2(d-z_T/2)}{\hbar} \sqrt{2m_{\text{ox}}|E_V - E_{\text{vox}}|}\right) \\ -\exp\left(\frac{-2(d+z_T/2)}{\hbar} \sqrt{2m_{\text{ox}}|E_V - E_{\text{vox}}|}\right) \end{bmatrix}. \quad (8.9)$$

To describe the interaction between carrier and phonon, the well-known Huang–Rhys factor  $S$  [87] is introduced here, which can be expressed as:

$$S = \frac{1}{2(\hbar\omega)^2} |\langle \psi_{\text{T}} | U | \psi_{\text{T}} \rangle|^2. \quad (8.10)$$

Normally,  $S$  is within the range of  $1.5 \text{ eV} < S\hbar\omega < 2.5 \text{ eV}$ . One can also treat this  $S$  factor as a semi-empirical parameter (within that range) for simplicity.

### 8.5.3 Multiphonon-Assisted Nonradiative Transition Rate

Based on the multiphonon theory [87–92], the transition rate can be described as:

$$W_{\text{mp}} = \frac{\pi}{\hbar} S |V|^2 \left(1 - \frac{n_p}{S}\right)^2 \frac{1}{(2\pi)^{1/2} \hbar \omega} l^{-1/4} \left(\frac{2S(n(n+1))^{1/2}}{n_p + l^{1/2}}\right)^{n_p} \times \exp\left(l^{1/2} - (2n+1)S + \frac{n_p \hbar \omega}{kT}\right), \quad (8.11)$$

where the number of emitted phonons with energy is

$$n_p = \frac{|E_0 - E_{\text{T}}| + kT}{\hbar \omega}, \quad (8.12)$$

The phonon occupation factor is

$$n = \frac{1}{\exp(\hbar\omega/kT) - 1} \quad (8.13)$$

and

$$l = \left( n_p^2 + 2S(n(n+1))^{1/2} \right). \quad (8.14)$$

#### 8.5.4 Modeling RTN Statistics

Finally, the capture and emission time constants of RTN in SNWTs can be expressed as:

$$\begin{cases} \frac{1}{\langle \tau_c \rangle} = W_{mp} A N_i \exp\left(\frac{-\Delta E}{kT}\right) \\ \frac{1}{\langle \tau_e \rangle} = W_{mp} A N_i \exp\left(\frac{|E_T - E_F|}{kT}\right), \end{cases} \quad (8.15)$$

where  $A$  is the gate area, and  $N_i$  is carrier concentration at the nanowire channel surface.

It is worth noting that the impact of Coulomb energy on carrier capture process is also considered as the parameter  $\Delta E$  in the equation, which can be obtained according to [89].

Now, we have obtained a full quantum RTN model for SNWTs. The modeling results agree well with experimental results, as shown in Figs. 8.9 and 8.10, which indicates that the developed model can successfully capture the impacts of quantum confinement on the RTN time constants in GAA SNWTs. This modeling method can also be applied to other types of multi-gate devices, e.g., double-gate or tri-gate FinFETs, if carefully get their channel wave functions.

## 8.6 Summary

Taking the GAA SNWT as the representative for extremely scaled multi-gate devices, the impacts of quantum confinement on RTN is discussed in this chapter based on both experimental and theoretical studies. It is found that the RTN capture/emission time constants are strongly impacted by the quantum confinement in multi-gate devices with thin channel. The experimental results indicate that the strong quantum confinement in SNWTs enhances the bias dependence but alleviates the temperature dependence of RTN time constants, which cannot be explained by classical RTN model based on SRH theory. By taking into account the lattice relaxation and multiphonon processes, the experiments can be successfully understood. A full quantum model for RTN time constants in SNWTs is also developed, which agrees well with the experimental results.

**Acknowledgments** The work in this chapter was supported in part by NSFC (61106085), 973 Projects (2011CBA00601), National S&T Major Project (2009ZX02035-001) and the Collaborative Project between Peking University and Samsung Electronics. The authors would like to thank the staff of Peking University Fabrication Lab for their assistance in device fabrication.

## References

1. International Technology Roadmap for Semiconductors (ITRS). <http://public.itrs.net/>
2. Colinge, J.-P. (ed.): FinFETs and other multi-gate transistors. Springer, New York (2008)
3. Huang, R., Wu, H.M., Kang, J.F., Xiao, D.Y., Shi, X.L., An, X., Tian, Y., Wang, R.S., Zhang, L.L., Zhang, X., Wang, Y.: Challenges of 22 nm and beyond CMOS technology. *Sci. China Ser. F Inform. Sci.* **52**, 1491 (2009)
4. Wang, Y.Y.: The driving force for development of IC and system in future: reducing the power consumption and improving the ratio of performance to power consumption. *Sci. China Ser. F Inform. Sci.* **54**, 915 (2011)
5. Li, M.: Review of advanced CMOS technology for post-Moore era. *Sci. China Phys. Mech. Astron.* **55**, 2316 (2012)
6. Kuhn, K.J.: Considerations for ultimate CMOS scaling. *IEEE Trans. Electron Dev.* **59**, 1813 (2012)
7. Jurczak, M., Collaert, N., Veloso, A., Hoffmann, T., Biesemans, S.: Review of FinFET technology. In: Proceedings of the IEEE international SOI conference, p. 1 (2009)
8. Ferain, I., Colinge, C.A., Colinge, J.P.: Multigate transistors as the future of classical metal-oxide-semiconductor field-effect transistors. *Nature* **479**, 310 (2011)
9. Huang, X., Lee, W.-C., Kuo, C., Hisamoto, D., Chang, L., Kedzierski, J., Anderson, E., Takeuchi, H., Choi, Y.-K., Asano, K., Subramanian, V., King, T.-J., Bokor, J., Hu, C.: Sub 50-nm FinFET: PMOS. In: IEDM technical digest, December, p. 67 (1999)
10. Park, J.-T., Colinge, J.P., Diaz, C.H.: Pi-gate SOI MOSFET. *IEEE Electron Device Lett.* **22**, 405 (2001)
11. Yang, F.-L., Chen, H.-Y., Chen, F.-C., Huang, C.-C., Chang, C.-Y., Chiu, H.-K., Lee, C.-C., Chen, C.-C., Huang, H.-T., Chen, C.-J., Tao, H.-J., Yeo, Y.-C., Liang, M.-S., Hu, C.: 25 nm CMOS Omega FETs. In: IEDM technical digest, December, p. 255 (2002)
12. Kavalieros, J., Doyle, B., Datta, S., Dewey, G., Doczy, M., Jin, B., Lionberger, D., Metz, M., Rachmady, W., Radosavljevic, M., Shah, U., Zelick, N., Chau, R.: Tri-gate transistor architecture with high-k gate dielectrics, metal gates and strain engineering. In: Symposium on VLSI technology. Digest of technical papers, p. 50 (2006)
13. Xu, X., Wang, R., Huang, R., Zhuge, J., Chen, G., Zhang, X., Wang, Y.: High-performance BOI FinFETs based on bulk-silicon substrate. *IEEE Trans. Electron Dev.* **55**(11), 3246 (2008)
14. Wu, C.C., Lin, D.W., Keshavarzi, A., Huang, C.H., Chan, C.T., Tseng, C.H., Chen, C.L., Hsieh, C.Y., et al.: High performance 22/20 nm FinFET CMOS devices with advanced high-k/metal gate scheme. In: IEDM technical digest, December, p. 600 (2010)
15. Yeh, C.C., Chang, C.-S., Lin, H.-N., Tseng, W.-H., Lai, L.-S., Perng, T.-H., Lee, T.-L., Chang, C.-Y., et al.: A low operating power FinFET transistor module featuring scaled gate stack and strain engineering for 32/28nm SoC technology. In: IEDM technical digest, December, p. 772 (2010)
16. Auth, C., Allen, C., Blattner, A., Bergstrom, D., Brazier, M., Bost, M., Buehler, M., Chikarmane, V., Ghani, T., et al.: A 22 nm high performance and low-power CMOS technology featuring fully-depleted trigate transistors, self-aligned contacts and high density MIM capacitors. In: Symposium on VLSI technology. Digest of technical papers, p. 131 (2012)
17. Jan, C.-H., Bhattacharya, U., Brain, R., Choi, S.-J., Curello, G., Gupta, G., Hafez, W., Jang, M., Kang, M., et al.: A 22nm SoC platform technology featuring 3-D tri-gate and high-k/metal gate, optimized for ultra low power, high performance and high density SoC applications. In: IEDM technical digest, December, p. 44 (2012)

18. EE Times. <http://www.eetimes.com/>
19. Huang, R., Wang, R.S.: Investigation of gate-all-around silicon nanowire transistors for ultimately scaled CMOS technology from top-down approach. *Front. Phys. China* **5**, 414 (2010)
20. Huang, R., Wang, R., Zhuge, J., Liu, C., Yu, T., Zhang, L., Huang, X., Ai, Y., Zou, J., Liu, Y., Fan, J., Liao, H., Wang, Y.: Characterization and analysis of gate-all-around Si nanowire transistors for extreme scaling. In: IEEE custom integrated circuits conference (CICC), pp. 2–1 (2011)
21. Yang, F.L., Lee, D.H., Chen, H.Y., Chang, C.Y., Liu, S.D., Huang, C.C., Chung, T.X., Chen, H.W., et al.: 5 nm-gate nanowire finFET. In: Symposium on VLSI technology. Digest of technical papers, p. 196 (2004)
22. Wang, R., Liu, H., Huang, R., Zhuge, J., Zhang, L., Kim, D.-W., Zhang, X., Park, D., Wang, Y.: Experimental investigations on carrier transport in Si nanowire transistors: Ballistic efficiency and apparent mobility. *IEEE Trans. Electron Dev.* **55**, 2960 (2008)
23. Asenov, A., Brown, A.R., Davies, J.H., Kaya, S., Slavcheva, G.: Simulation of intrinsic parameter fluctuations in decanometer and nanometer-scale MOSFETs. *IEEE Trans. Electron Dev.* **50**, 1837 (2003)
24. Bernstein, K., Frank, D.J., Gattiker, A.E., Haensch, W., Ji, B.L., Nassif, S.R., Nowak, E.J., Pearson, D.J., Rohrer, N.J.: High-performance CMOS variability in the 65-nm regime and beyond. *IBM J. Res. Dev.* **50**, 433 (2006)
25. Kuhn, K., Giles, M.D., Becher, D., Kolar, P., Kornfeld, A., Kotlyar, R., Ma, S.T., Maheshwari, A., Mudanai, S.: Process technology variation. *IEEE Trans. Electron Dev.* **58**, 2197 (2011)
26. Takeuchi, K.: Impact of discrete-charge-induced variability on scaled MOS devices. *IEICE Trans. Electron.* **E95-C**, 414 (2012)
27. O'uchi, S., Matsukawa, T., Nakagawa, T., Endo, K., Liu, Y.X., Sekigawa, T., Tsukada, J., Ishikawa, Y., Yamauchi, H., Ishii, K., Suzuki, E., Koike, H., Sakamoto, K., Masahara, M.: Characterization of metal-gate FinFET variability based on measurements and compact model analyses. In: IEDM technical digest, p. 709 (2008)
28. Wang, R., Zhuge, J., Huang, R., Yu, T., Zou, J., Kim, D.-W., Park, D., Wang, Y.: Investigation on variability in metal-gate Si nanowire MOSFETs: analysis of variation sources and experimental characterization. *IEEE Trans. Electron Dev.* **58**, 2317 (2011)
29. Asenov, A., Balasubramaniam, R., Brown, A.R., Davies, J.H.: RTS amplitudes in decanometer MOSFETs: 3-D simulation study. *IEEE Trans. Electron Dev.* **50**(839) (2003)
30. Ielmini, D., Compagnoni, C.M., Spinelli, A.S., Lacaia, A.L., Gerardi, C.: A new channel percolation model for VT shift in discrete trap memories. In: International reliability physics symposium (IRPS), p. 515 (2004)
31. Huard, V., Parthasarathy, C., Guerin, C., Valentin, T., Pion, E., Mammasse, M., Planes, N., Camus, L.: NBTI degradation: From transistor to SRAM arrays. In: International reliability physics symposium (IRPS), p. 289 (2008)
32. Reisinger, H., Grasser, T., Hofmann, K., Gustin, W., Schlünder, C.: A study of NBTI by the statistical analysis of the properties of individual defects in pMOSFETs. In: International integrated reliability workshop. Final report (IIRW), p. 30 (2009)
33. Grasser, T., Reisinger, H., Goes, W., Aichinger, Th., Hehenberger, Ph., Wagner, P.-J., Nelhiebel, M., Franco, J., Kaczer, B.: Switching oxide traps as the missing link between negative bias temperature instability and random telegraph noise. In: International electron devices meeting (IEDM), p. 729 (2009)
34. Kaczer, B., Grasser, T., Roussel, P.J., Franco, J., Degraeve, R., Ragnarsson, L., Simoen, E., Groeseneken, G., Reisinger, H.: Origin of NBTI variability in deeply scaled pFETs. In: International reliability physics symposium (IRPS), p. 26 (2010)
35. Liu, C., Zou, J., Wang, R., Huang, R., Xu, X., Liu, J., Wu, H., Wang, Y.: Towards the systematic study of aging induced dynamic variability in nano-MOSFETs: adding the missing cycle-to-cycle variation effects into device-to-device variation. In: IEDM technical digest, p. 571 (2011)
36. Liu, C., Ren, P., Wang, R., Huang, R., Ou, J., Wang, J., Wu, J., Yu, S., Lee, S.-W., Wang, Y.: New observations on AC NBTI induced dynamic variability in scaled high-k/metal-gate MOSFETs: characterization, origin of frequency dependence, and impacts on circuits. In: IEDM technical digest, p. 466 (2012)

37. Kirton, M., Uren, M.: Noise in solid-state microstructure: a new perspective on individual defects, interface states and low-frequency (1/f) noise. *Adv. Phys.* **38**, 367 (1989)
38. Campbell, J.P., Qin, J., Cheung, K.P., Yu, L.C., Suehle, J.S., Oates, A., Sheng, K.: Random telegraph noise in highly scaled nMOSFETs. In: Proceedings of the IRPS, p. 382 (2009)
39. Grasser, T.: Stochastic charge trapping in oxides: from random telegraph noise to bias temperature instabilities. *Microelectron. Reliab.* **52**, 39 (2012)
40. Frank, D.J.: Random telegraph noise – measurement, analysis, and consequences. In: IRPS tutorial (2012)
41. Hung, K.K., Ko, P.K., Hu, C., Cheng, Y.C.: Random telegraph noise of deep-submicrometer MOSFETs. *IEEE Electron Device Lett.* **11**, 90 (1990)
42. Tsai, M.H., Ma, T.P.: The impact of device scaling on the current fluctuations in MOSFET's. *IEEE Trans. Electron Dev.* **41**, 2061 (1994)
43. Shi, Z., Mieville, J.-P., Dutoit, M.: Random telegraph signals in deep submicron n-MOSFET's. *IEEE Trans. Electron Dev.* **41**, 1161 (1994)
44. Ghibaudo, G., Bouthachacha, T.: Electrical noise and RTS fluctuations in advanced CMOS devices. *Microelectron. Reliab.* **42**, 573 (2002)
45. Simoen, E., Claeys, C.: Random telegraph signal: a local probe for single point defect studies solid-state devices. *Mater. Sci. Eng.* **B91-92**, 136 (2002)
46. Tega, N., Miki, H., Pagette, F., Frank, D., Ray, A., Rooks, M., Haensch, W., Torii, K.: Increasing threshold voltage variation due to random telegraph noise in FETs as gate length scale to 20 nm. In: Symposium on VLSI technology. Digest of technical papers, p. 50 (2009)
47. Takeuchi, K., Nagumo, T., Yokogawa, S., Imai, K., Hayashi, Y.: Single-charge-based modeling of transistor characteristics fluctuations based on statistical measurement of RTN amplitude. In: Symposium on VLSI technology. Digest of technical papers, p. 54 (2009)
48. Nagumo, T., Takeuchi, K., Yokogawa, S., Imai, K., Hayashi, Y.: New analysis methods for comprehensive understanding of random telegraph noise. In: IEDM technical digest, p. 759 (2009)
49. Tega, N., Miki, H., Ren, Z., D'Emic, C.P., Zhu, Y., Frank, D.J., Cai, J., Guillorn, M.A., Park, D.-G., Haensch, W., Torii, K.: Reduction of random telegraph noise in high-k/metal-gate stacks for 22 nm generation FETs. In: IEDM technical digest, p. 771 (2009)
50. Miki, H., Tega, N., Ren, Z., D'Emic, C.P., Zhu, Y., Frank, D.J., Guillorn, M.A., Park, D.-G., Haensch, W., Torii, K.: Hysteretic drain-current behavior due to random telegraph noise in scaled-down FETs with high-k/metal-gate stacks. In: IEDM technical digest, p. 620 (2010)
51. Realov, S., Shepard, K.L.: Random telegraph noise in 45-nm CMOS: analysis using an on-chip test and measurement system. In: IEDM technical digest, p. 624 (2010)
52. Miki, H., Yamaoka, M., Tega, N., Ren, Z., Kobayashi, M., D'Emic, C.P., Zhu, Y., Frank, D.J., Guillorn, M.A., Park, D.-G., Haensch, W., Torii, K.: Understanding short-term BTI behavior through comprehensive observation of gate-voltage dependence of RTN in highly scaled high-k/metal-gate pFETs. In: Symposium on VLSI technology. Digest of technical papers, p. 148 (2011)
53. Miki, H., Yamaoka, M., Frank, D.J., Cheng, K., Park, D.-G., Leobandung, E., Torii, K.: Voltage and temperature dependence of random telegraph noise in highly scaled HKMG ETSOI nFETs and its impact on logic delay uncertainty. In: Symposium on VLSI technology. Digest of technical papers, p. 137 (2012)
54. Zou, J., Wang, R., Gong, N., Huang, R., Xu, X., Ou, J., Liu, C., Wang, J., Liu, J., Wu, J., Yu, S., Ren, P., Wu, H., Lee, S.-W., Wang, Y.: New insights into AC RTN in scaled high-k/metal-gate MOSFETs under digital circuit operations. In: Symposium on VLSI technology. Digest of technical papers, p. 139 (2012)
55. Zou, J., Wang, R., Luo, M., Huang, R., Xu, N., Ren, P., Liu, C., Xiong, W., Wang, J., Liu, J., Wu, J., Wong, W., Yu, S., Wu, H., Lee, S.-W., Wang, Y.: Deep understanding of AC RTN in MuGFETs through new characterization method and impacts on logic circuits. In: Symposium on VLSI technology. Digest of technical papers, accepted (2013)
56. Tega, N., Miki, H., Yamaoka, M., Kume, H., Toshiyuki, M., Ishida, T., Yuki, M., Yamada, R., Kazuyoshi, T.: Impact of threshold voltage fluctuation due to random telegraph noise on scaled-down SRAM. In: IEEE international reliability physics symposium, p. 541 (2008)

57. Ghetti, A., Compagnoni, C., Spinelli, A., Visconti, A.: Comprehensive analysis of random telegraph noise instability and its scaling in deca Cnanometer flash memories. *IEEE Trans. Electron Dev.* **56**, 1746 (2009)
58. Sakoda, T., Tamura, N., Xiao, S., Minakata, H., Morisaki, Y., Nishigaya, K., Saiki, T., Uetake, T., Iwasaki, T., Ehara, H., Matsuyama, H., Shimizu, H., Hashimoto, K., Kimoto, M., Kase, M., Ikeda, K.: 1st quantitative failure-rate calculation for the actual large-scale SRAM using ultra-thin gate-dielectric with measured probability of the gate-current fluctuation and simulated circuit failure-rate. In: *Symposium on VLSI technology. Digest of technical papers*, p. 26 (2007)
59. Wang, X., Rao, P.R., Mierop, A., Theuwissen, A.J.P.: Random telegraph signal in CMOS image sensor pixels. In: *IEDM technical digest*, p. 1 (2006)
60. Toh, S., Tsukamoto, Y., Guo, Z., Jones, L., King Liu, T.-J., Nikolic, B.: Impact of random telegraph signals on Vmin in 45nm SRAM. In: *IEDM technical digest*, p. 767 (2009)
61. Toh, S.O., Liu, T.-J.K., Nikolic, B.: Impact of random telegraph signaling noise on SRAM stability. In: *Symposium on VLSI technology. Digest of technical papers*, p. 204 (2011)
62. Yamaoka, M., Miki, H., Bansal, A., Wu, S., Frank, D.J., Leobandung, E., Torii, K.: Evaluation methodology for random telegraph noise effects in SRAM arrays. In: *IEDM technical digest*, p. 745 (2011)
63. Matsumoto, T., Kobayashi, K., Onodera, H.: Impact of random telegraph noise on CMOS logic delay uncertainty under low voltage operation. In: *IEDM technical digest*, p. 581 (2012)
64. Lim, Y.F., Xiong, Y.Z., Singh, N., Yang, R., Jiang, Y., et al.: Random telegraph signal noise in gate-all-around Si-FinFET with ultranarrow body. *IEEE Electron Device Lett.* **27**, 765 (2006)
65. Yang, S., Yeo, K.H., Kim, D.W., Seo, K.I., Park, D., et al.: Random telegraph noise in N-type and P-type Silicon nanowire transistors. In: *IEDM technical digest*, p. 1 (2008)
66. Zhuge, J., Zhang, L., Wang, R., Huang, R., Kim, D.-W., Park, D., Wang, Y.: Random telegraph signal noise in gate-all-around silicon nanowire transistors featuring Coulomb-blockade characteristics. *Appl. Phys. Lett.* **94**, 083503 (2009)
67. Zhang, L., Zhuge, J., Wang, R., Huang, R., et al.: New insights into oxide traps characterization in gate-all-around nanowire transistors with TiN metal gates based on combined Ig-Id RTS technique. In: *Symposium on VLSI technology. Digest of technical papers*, p. 46 (2009)
68. Liu, C., Wang, R., Zou, J., Huang, R., Fan, C., Zhang, L., Fan, J., Ai, Y., Wang, Y.: New understanding of the statistics of random telegraph noise in Si nanowire transistors – the role of quantum confinement and non-stationary effects. In: *IEDM technical digest*, p. 521 (2011)
69. Fan, M.L., Hu, V.P.H., Chen, Y.N., Su, P., Chuang, C.T.: Analysis of single-trap-induced random telegraph noise on FinFET devices, 6T SRAM cell, and logic circuits. *IEEE Trans. Electron Dev.* **59**, 2227 (2012)
70. Singh, N., Agarwal, A., Bera, L.K., Liow, T.Y., Yang, R., Rustagi, S.C., Tung, C.H., Kumar, R., Lo, G.Q., Balasubramanian, N., Kwong, D.-L.: High-performance fully depleted silicon nanowire (diameter  $\leq 5$  nm) gate-all-around CMOS devices. *IEEE Electron Device Lett.* **27**, 383 (2006)
71. Jiang, Y., Liow, T.Y., Singh, N., Tan, L.H., Lo, G.Q., Chan, D.S.H., Kwong, D.L.: Nanowires FETs for low power CMOS applications featuring novel gate-all-around single metal FUSI gates with dual  $\mu$ m and VT tune-ability. In: *IEDM technical digest*, p. 869 (2008)
72. Bangsaruntip, S., Cohen, G.M., Majumdar, A., Zhang, Y., Engelmann, S.U., Fuller, N., Gignac, L.M., Mittal, S., Newbury, J.S., Guillorn, M., Barwicz, T., Sekaric, L., Frank, M. M., Sleight, J. W.: High performance and highly uniform gate-all-around silicon nanowire MOSFETs with wire size dependent scaling. In: *IEDM technical digest*, p. 297 (2009)
73. Bangsaruntip, S., Majumdar, A., Cohen, G.M., Engelmann, S.U., Zhang, Y., Guillorn, M., Gignac, L.M., Mittal, S., Graham, W.S., Joseph, E.A., Klaus, D.P., Chang, J., Cartier, E.A., Sleight, J.W.: Gate-all-around silicon nanowire 25-stage CMOS ring oscillators with diameter down to 3 nm. In: *Symposium on VLSI technology. Digest of technical papers*, p. 21 (2010)
74. Suk, S.D., Lee, S.-Y., Kim, S.-M., Yoon, E.-J., Kim, M.-S., Li, M., Oh, C.W., Yeo, K.H., Kim, S. H., Shin, D.-S., Lee, K.-H., Park, H.S., Han, J.N., Park, C.J., Park, J.-B., Kim, D.-W., Park, D., Ryu, B.-I.: High performance 5nm radius twin silicon nanowire MOSFET (TSNWFET): fabrication on bulk Si wafer, characteristics, and reliability. In: *IEDM technical digest*, p. 717 (2005)
75. Yeo, K.H., Suk, S.D., Li, M., Yeoh, Y.-Y., Cho, K.H., Hong, K.-H., Yun, S., Lee, M.S., Cho, N., Lee, K., Hwang, D., Park, B., Kim, D.-W., Park, D., Ryu, B.-I.: Gate-all-around (GAA)

- twin silicon nanowire MOSFET (TSNWFET) with 15 nm length gate and 4 nm radius nanowires. In: IEDM technical digest, p. 539 (2006)
76. Suk, S.D., Yeoh, Y.Y., Li, M., Yeo, K.H., Kim, S.-H., Kim, D.-W., Park, D., Lee, W.-S.: TSNWFET for SRAM cell application: performance variation and process dependency. In: Symposium on VLSI technology. Digest of technical papers, p. 38 (2008)
77. Li, M., Yeo, K.H., Suk, S.D., Yeoh, Y.Y., Kim, D.-W., Chung, T.Y., Oh, K.S., Lee, W.-S.: Sub-10 nm gate-all-around CMOS nanowire transistors on bulk Si substrate. In: Symposium on VLSI technology. Digest of technical papers, p. 94 (2009)
78. Tian, Y., Huang, R., Wang, Y., Zhuge, J., Wang, R., Liu, J., Zhang, X., Wang, Y.: 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. In: IEDM technical digest, p. 895 (2007)
79. Ai, Y., Huang, R., Hao, Z., Fan, C., Wang, R., Pu, S., Wang, Y.: Top-down fabrication of shape controllable Si nanowires based on conventional CMOS process. *Phys. E* **43**, 102 (2010)
80. Huang, R., Zou, J., Wang, R., Fan, C., Ai, Y., Zhuge, J., Wang, Y.: Experimental demonstration of current mirrors based on silicon nanowire transistors for inversion and subthreshold operations. *IEEE Trans. Electron Dev.* **58**, 3639 (2011)
81. Yuzhelevski, Y., Yuzhelevski, M., et al.: Random telegraph noise analysis in time domain. *Rev. Sci. Instrum.* **71**, 1681 (2000)
82. Li, M., Wang, R., Zou, J., Huang, R.: Characterization of random telegraph noise in scaled high-k/metal-gate MOSFETs with SiO<sub>2</sub>/HfO<sub>2</sub> Gate Dielectrics. *ECS Trans.* **52**, 941 (2013)
83. Brook, D., Evans, D.A.: An approach to the probability distribution of cusum run length. *Life Sci. Math. Phys. Sci.* **59**, 539 (1972)
84. Rabiner, L.R.: A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **77**, 257 (1989)
85. Shockley, W., Read, W.T.: Statistics of recombinations of holes and electrons. *Phys. Rev.* **87**, 835 (1952)
86. Hall, R.N.: Electron-hole recombination in germanium. *Phys. Rev.* **87**, 387 (1952)
87. Huang, K.: Adiabatic approximation-theory and static coupling theory of nonradiative transition. *Sci. Sin.* **24**, 27 (1981)
88. Zheng, J., Tan, H., Ng, S.: Theory of non-radiative capture of carriers by multiphonon processes for deep centres in semiconductors. *J. Phys. Condens. Matter.* **6**, 1695 (1994)
89. Palma, A., Godoy, A., Jimenz-Tejada, J., Carceller, J., Lopez-Villanueva, J.A.: Quantum two-dimensional calculation of time constants of random telegraph signals in metal-oxide-semiconductor structures. *Phys. Rev. B* **56**, 9565 (1997)
90. Henry, C., Lang, D.: Nonradiative capture and recombination by multiphonon emission in GaAs and GaP. *Phys. Rev. B* **15**, 989 (1977)
91. Andersson, M.O., Xiao, Z., Norrman, S., Engstrom, O.: Model based on trap-assisted tunneling for two-level current fluctuations in submicrometer metal-silicon-dioxide-silicon diodes. *Phys. Rev. B* **41**, 9836 (1990)
92. Ielmini, D., Manigrasso, M., Gattel, F., Valentini, M.G.: A new NBTI model based on hole trapping and structural relaxation in MOS dielectrics. *IEEE Trans. Electron Dev.* **56**, 1943 (2009)
93. Shulz, M.: Coulomb energy of traps in semiconductor space-charge regions. *J. Appl. Phys.* **74**, 2649 (1993)
94. Cousin, B., et al.: Compact modeling of quantization effects for cylindrical gate-all-around MOSFETs. In: IEEE international conference of ultimate integration of silicon (ULIS), p. 269 (2009)

## **Chapter 9**

# **Investigations on Transport Properties of Poly-silicon Nanowire Transistors Featuring Independent Double-Gated Configuration Under Cryogenic Ambient**

**Wei-Chen Chen and Horng-Chih Lin**

**Abstract** Transport properties of poly-Si nanowire transistors, which were fabricated by a simple and low-cost method, are examined in this chapter. The proposed device features two independent gates and thus allows more flexible operation. Electrical measurements performed under cryogenic ambient displayed intriguing characteristics in terms of length-dependent abrupt switching behavior for one of the single-gated modes that is in obvious conflict with the conventional theory concerning short channel effects. Through simulation and experimental verification, it was found that such phenomenon is related to a number of structural parameters, and the root cause was identified to be the nonuniformly distributed dopants introduced by ion implantation.

### **9.1 Introduction**

To maintain the momentum of CMOS scaling, multiple-gated nanowire (NW) devices have been proposed as one of the most promising future transistor structures due to the better gate controllability to suppress the short channel effects (SCE). NW, basically, can be defined as a narrow stripe material with its cross-sectional feature size less than 100 nm. It has been conceived as an ideal building block for nanoelectronics and optoelectronics. One unique feature of NW lies in its very tiny volume and large surface-to-volume ratio, thus making it suitable for a

---

W.-C. Chen (✉)

Emerging Central Laboratory, Macronix International Co., Ltd, Hsinchu 300, Taiwan, ROC  
e-mail: [weichenchen@mxic.com.tw](mailto:weichenchen@mxic.com.tw)

H.-C. Lin

Department of Electronics Engineering and Institute of Electronics,  
National Chiao Tung University, Hsinchu 300, Taiwan, ROC

National Nano Device Laboratories, Hsinchu 300, Taiwan, ROC  
e-mail: [hclin@faculty.nctu.edu.tw](mailto:hclin@faculty.nctu.edu.tw)

wide array of applications. For memory applications, fast programming/erasing efficiency and low voltage operation can be facilitated and the large surface-to-volume ratio feature promotes high sensitivity for sensor applications.

Though most of the previous works regarding NW research focused on single-crystalline Si-based technology for realizing high performance CMOS transistors, recently a new concept of NW-based thin film transistors (TFTs) has started to gain attention [1, 2]. Compared to single-crystalline Si wafer technology, TFT devices completely eliminate the need for a high-cost substrate. This merit is especially profound when prohibitive silicon-on-insulator (SOI) architecture starts to emerge as another substrate alternative in order to promote the fabrication of multiple-gated device such as FinFETs [3]. For the macroelectronics industry where it is preferred to construct electronic devices and components over a large area substrate, TFT architecture is a promising option since it merely requires a thin active layer to be deposited onto an insulating substrate. Therefore, glass and plastic substrates become the primary choice as they are flexible, light weight, and most importantly, cheap. Poly-Si TFTs are also the main architectures used in 3D memory technology [4] to avoid the complexity and difficulty in high temperature deposition of single crystalline Si layers. Though grain boundary effects are a concern in poly-Si, poly-Si-based TFTs with NW channel and multi-gated scheme have been shown to perform comparably to their bulk-Si counterparts [5–7]. In addition, poly-Si TFTs greatly facilitate the integration of a wide array of circuit components, making possible system-on-panel (SOP) applications [8].

As a result of the tiny cross section of NW body, confinement of carriers in the plane perpendicular to the channel direction can occur. Already a plethora of intriguing phenomena are associated with this nanoscale structure such as Coulomb blockade [9], quantized conductance [10], and reduced phonon scattering [11], to name just a few. Given that most of the intriguing quantum-mechanical effects tend to occur in low temperature regime, in this chapter to investigate the underlying transport mechanism of the proposed independent double-gated NW poly-Si TFTs, devices with channel lengths ( $L$ ) ranging from 39 nm to 5  $\mu\text{m}$  were fabricated using mix-and-match lithography of I-line stepper and e-beam direct writing method and characterized under cryogenic ambient. Electrical characterization performed in the temperature ( $T$ ) ranging from 300 to 78 K shows abnormal switching phenomena for one of the single-gated modes with a very steep subthreshold swing (SS), which completely disappears when altering the gate-doping technique. It is also found that such phenomenon is related to a number of structural parameters. A simple model based on the process of electron trapping and detrapping in the channel is proposed.

The content of this chapter is arranged as follows. In Sect. 9.2, brief background information of multiple-gated devices is given, followed by Sect. 9.3 that describes some of the important transport properties found in NW devices. Device structure and process flow of the proposed poly-Si NW transistors are elucidated in Sect. 9.4. Basic electrical characteristics at room temperature are analyzed in Sect. 9.5. Section 9.6 then investigates the underlying transport mechanism leading to the abrupt switching behavior under cryogenic ambient and suggests one model to interpret the experimental findings. Finally, a brief conclusion is presented in Sect. 9.7.

## 9.2 Background of Multiple-Gated Devices

Faced with an ever greater challenge of fabricating a “well-behaved” transistor for which SCEs are minimized in the era of 22 nm technology node and beyond, a number of new processing schemes have been embraced to address this issue. For the purpose of resolving poly-gate depletion and alleviating nonnegligible gate dielectric tunneling current, high- $\kappa$ /metal gate is being pursued actively and is already adopted in commercial 45-nm microprocessors of Intel [12]. Another option of relieving SCEs is to resort to multiple-gated configuration. As the name suggests, in a multiple-gated transistor, the active layer is controlled by more than one gate as opposed to planar single-gated counterparts. This kind of structure is capable of effectively preventing the drain field penetration that weakens the gate controllability over channels thereby leading to SCEs. On account of the better electrostatic control of channels, for a given channel width, larger driving current can be provided by multiple-gated configuration. Of all the multiple-gated structures ever proposed so far, FinFET is considered as the one most appropriate for practical applications because it is actually quasi-planar and is most compatible with modern ULSI (ultra large-scale integration) technology [13]. Moreover, the conduction width in FinFET, which is determined by the fin height, is along a direction that is normal to the substrate plane. This vertical feature offers an improvement in density that is not restricted by the resolution of lithographic tools. Evolving from the original double-gated structure, several variations of FinFET have been reported, including tri-gate [14] and  $\Omega$ -gate [15]. In particular, the gate-all-around type device where the channel is fully wrapped by a surrounding gate is considered the most ideal structure to proffer the best gate controllability over the channel [16].

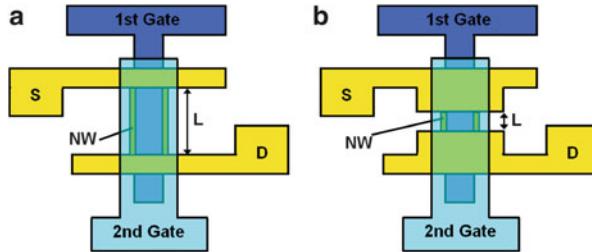
## 9.3 Transport Behavior of Nanowire Devices

The paradigm shift caused by NW creates whole new concepts and perspectives on device physics and possibilities that conventional planar counterparts have yet to offer. From a microscopic point of view, this kind of low-dimensional structure is suitable for studying quantum-mechanical effects. Many reports focused on the carrier transport properties in NW devices for which well-established theories of three- or two-dimensional materials are no longer appropriate. Energy bands are split into sub-bands, and the energy levels become discrete by the tiny volume of NW [17]. Because of the nanoscale cross section that confines the wave functions of sub-bands, carriers in NW devices must transport through a large number of one-dimensional sub-bands. Quantum confinement, sub-band splitting, and surface and interface relaxation [18–20] are among a plethora of effects that must be taken into account in order to correctly interpret the NW characteristics, including

unexpected increase of threshold voltage ( $V_{TH}$ ) with reduced NW width [21], oscillation of drain current and mobility [22], and reduced Stark effect [23]. The essence of quantum confinement lies in the fact that the energy levels are no longer continuous in this scenario but become discrete sub-bands instead. For an N-channel device, electrons tend to populate the lowest sub-band whose energy difference between the conduction band edge increases with the scaling of NW dimensions. Consequently,  $V_{TH}$  becomes higher as NW is made narrower. The presence of discrete energy levels also leads to valleys in the drain current as the gate voltage is increased under low temperature regime because the sub-band must lie within the range of thermal energy from the previous sub-band that is already populated by carriers before further increase of the number of carriers is possible. Another intriguing effect exclusive to NWs is that the drain current of linear regime (low  $V_D$ ) in strong inversion is found to decrease with decreasing temperature, which is shown to be caused by the inter-sub-band scattering induced by quantum confinement [22]. Yet, in saturation regime (high  $V_D$ ), this effect is diminished and the mobility–temperature relationship again resorts to what is dictated by phonon scattering. In addition, differential conductance fluctuations are observed in output curves as the series resistance of the drain extension is changed by the interplay between inter-sub-band transitions and quasi-ballistic transport [24].

Simulation results have shown that NW transistors may approach ballistic transport [25], i.e., a carrier does not experience any collision with other carriers or elastic centers during its traverse from the source toward the drain. Drain current under ballistic transport has been shown to depend only on the carrier concentration near the source and the injection velocity at the peak of the barrier at the source side [26, 27], which is determined solely by the thermal velocity in the case of non-degeneracy. Thus, the concept of mobility becomes meaningless, leading to a profound change of mind-set on how the carrier transport properties should be examined. For typical short channel planar devices, scattering defects associated with pocket implant near source/drain junctions result in mobility degradation as the channel length is made shorter [28, 29]. Even though NW exhibits similar behavior that the “extracted” mobility tends to decrease with reduced channel length, this is essentially an artifact and can be explained by the “ballistic mobility” model [30]. As a matter of fact, since NW has entered into ballistic regime, the measured mobility no longer determines the transport property, and the actual transport behavior will not degrade with channel length scaling [31]. In other words, not only does NW device possess better immunity against SCE, it is also promising to provide ballistic transport when being downsized to ultrashort channel lengths.

One concept called “quantum capacitance limit” introduced recently further highlights the performance advantage of 1D NW device over bulk counterparts in terms of the power delay product improvement that can be achieved from scaling [32]. The inversion layer or quantum capacitance of a bulk device is proportional to the density of states in the channel [33], which increases with the gate voltage; thus, the modulation of the surface potential is less efficient in the on-state, explaining

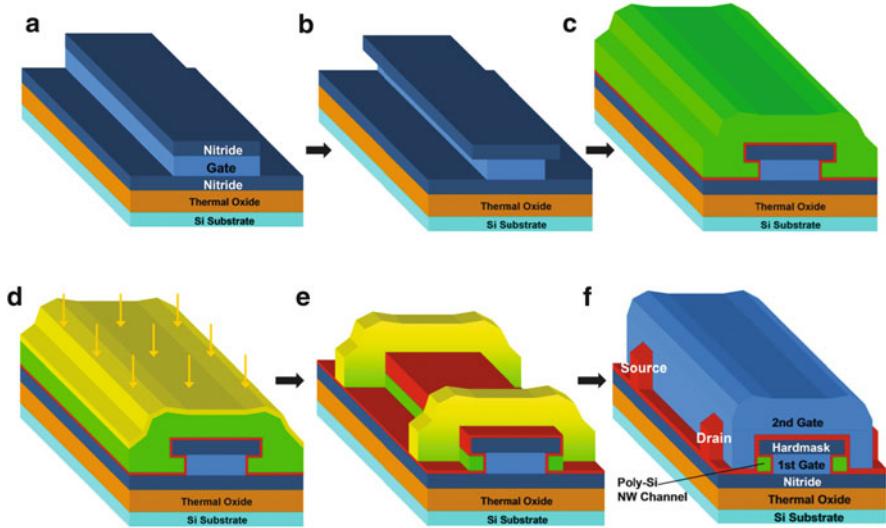


**Fig. 9.1** Layout of the device fabricated using (a) I-line only and (b) mix-and-match between e-beam and I-line.  $L$  defines the channel length, which is greater than  $0.4\text{ }\mu\text{m}$  in (a) and smaller than  $100\text{ nm}$  in (b). Reproduced with permission of IOP Publishing from [36]

the gradual degradation of SS with a larger gate voltage. The scenario is vastly different in the case of 1D NW device for which the density of states is inversely proportional to the square root of the difference between the carrier energy and surface potential. Consequently, it is easier for NW to reach the so-called quantum capacitance limit, where the gate dielectric capacitance readily exceeds the quantum capacitance, and the gate electrode is still able to provide ideal control of the surface potential in the on-state. In other words, the thickness of the gate dielectric in a 1D NW transistor required to meet the criterion of quantum capacitance limit can be thicker and is much more technologically feasible than the bulk devices.

## 9.4 Device Structures and Fabrication

By cleverly forming a cavity at two sides of a nitride/Si/nitride stack through selective etching of the sandwiched Si, followed by refilling with the active layer, we had demonstrated a novel and simple method for fabricating devices with decanometer NW dimensions using an optical I-line stepper [34, 35], whose layout and process flow are shown in Figs. 9.1a and 9.2, respectively. To overcome the minimum channel length limit set by the resolution of I-line stepper while maintaining a low cost approach, a slightly modified fabrication process by adding e-beam direct writing was adopted in this study to achieve sub-100 nm channel length with layout given in Fig. 9.1b. It is worth mentioning that this new approach requires only one additional lithographic step and can be easily integrated with the original version on the same wafer. In short, the new version is only different in the manner how channel length is defined. That is, instead of directly forming source/drain (S/D) and channels in a single step, in the newly proposed method, mesa-isolation of active layer was first performed followed by another patterning to simultaneously define S/D and channels using e-beam writing with positive resist, which in turn also determined channel length.

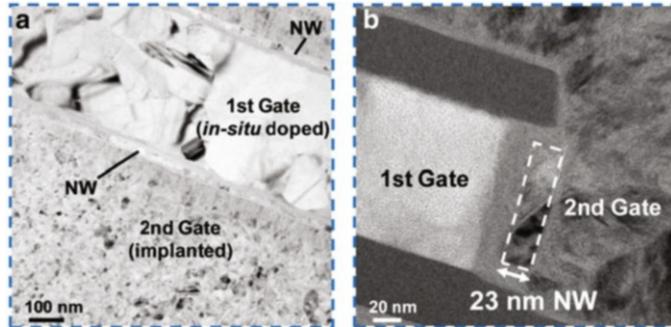


**Fig. 9.2** Schematic process flow for the double-gated NW devices with long and short  $L$ . (a) First gate stack (100 nm in-situ doped n<sup>+</sup> poly) patterning. (b) Selective plasma etching of the first gate using chlorine (Cl<sub>2</sub>) and sulfur hexafluoride (SF<sub>6</sub>) gases. (c) First gate dielectric (15 nm TEOS oxide) and 100-nm amorphous Si layer deposition by LPCVD. (d) Recrystallization of amorphous Si performed at 600 °C for 24 h in N<sub>2</sub> ambient and S/D implantation by phosphorus with 5 × 15 cm<sup>-2</sup> dosage and 15 keV energy. (e) Definition of S/D and NW channels by Cl<sub>2</sub> dry etching. For long channel devices patterned by I-line,  $L$  is larger than 0.4 μm; for short channel devices patterned by e-beam,  $L$  is smaller than 100 nm. (f) Second gate stack deposition (5-nm TEOS oxide and 100-nm implanted n<sup>+</sup> poly) by LPCVD and patterning. Reproduced with permission of IOP Publishing from [36]

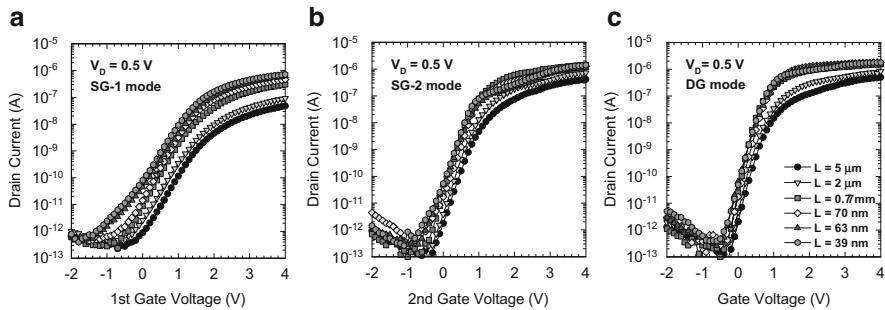
## 9.5 Electrical Characteristics at Room Temperature

Plane-view and cross-sectional transmission electron microscopic (TEM) images of the fabricated device are shown in Figs. 9.3a and 9.3b, respectively. It can be seen from the plane-view picture that the grain sizes of first and second gates are quite different which can be ascribed to the difference in doping methods [37]. Namely, the first gate of larger grain size is of in situ-doped N<sup>+</sup> poly-Si, while the second gate with smaller grain size is of implanted N<sup>+</sup> poly-Si. Note that the dielectrics for the first and second gates are 15-nm and 5-nm tetraethyl orthosilicate (TEOS) oxide, respectively, while the thickness of NW (or the width between the two gate dielectrics) is 23 nm.

Channel-length-dependent transfer characteristics at 25 °C are shown in Fig. 9.4. Because of the independent double-gated feature, there are three feasible operation modes, i.e., SG-1, SG-2, and DG modes. Specifically, SG-1 and SG-2 modes refer to, respectively, the scheme when the first or second gate serves as the driving gate



**Fig. 9.3** (a) Plane- and (b) cross-sectional view images of a fabricated device showing a 23-nm-thick NW channel surrounded by the first and second gates. The drastic grain size discrepancy between the first and second gates is a consequence of the different doping method implemented: the first gate is by in-situ doping while the second gate by implantation. Reproduced with permission of IOP Publishing from [36]



**Fig. 9.4** Length-dependent transfer curves at 25 °C under (a) SG-1, (b) SG-2, and (c) DG modes

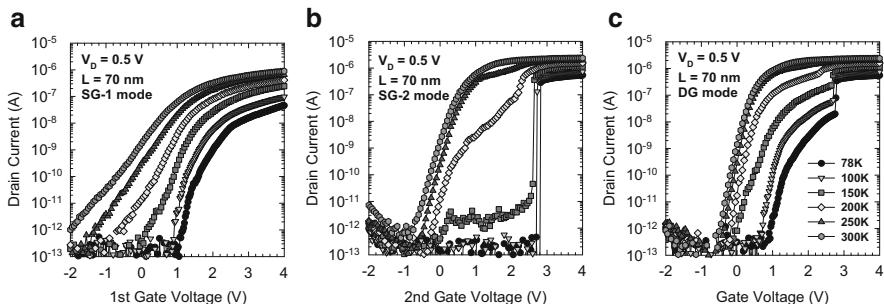
with the other gate electrode grounded. While for DG mode, the two gates are tied together for driving the device simultaneously. Among the three operation modes, SG-1 mode is the one most vulnerable to SCEs and thereby shows more severe SS degradation as  $L$  is reduced, an inherent feature of this structure [34]. For SG-2 and DG modes, only minor SS enlargement is observed even when  $L$  is scaled to less than 100 nm. One point worth pointing out is that S/D regions of devices measured are formed by low energy ion implantation, so instead of increasing with decreased  $L$ , the drive current is seen to saturate as  $L$  is below 0.7  $\mu$ m, as can be clearly identified in SG-2 and DG modes of Fig. 9.4. This issue is expected to be relieved by silicided or in situ doped S/D [35].

## 9.6 Electrical Characteristics at Cryogenic Ambient

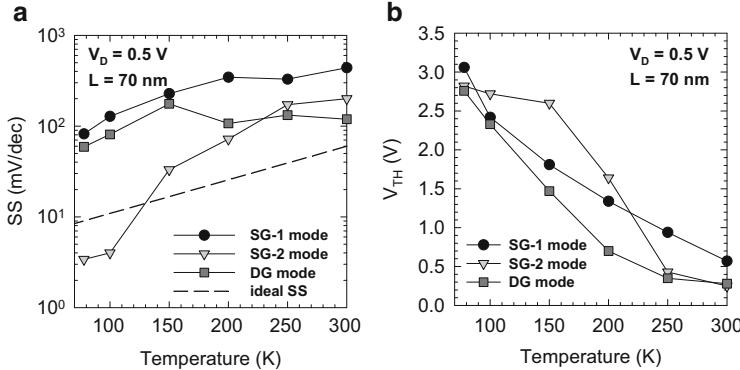
### 9.6.1 Evolution of Transfer Curves Under Different Operation Modes with Respect to Temperature

Transfer characteristics measured at various  $T$  for a device with  $L = 70$  nm are shown in Fig. 9.5. As compared with single-crystalline Si, carrier transport in poly-Si tends to experience additional scattering from trapping centers in grain boundaries, resulting in carrier depletion and the formation of potential barriers impeding carrier motion from one grain to another [38]. Therefore, thermionic emission over these energy barriers is usually considered as the major conduction mechanism when dealing with poly-Si, and simulation results have shown that thermionic conduction indeed well describes the transfer and output characteristics of the proposed device at room  $T$  [39]. This kind of thermally activated process also manifests itself in a way that the mobility and  $V_{TH}$  would decrease and increase, respectively, as  $T$  is reduced [40, 41]. Transfer curves under SG-1 mode in Fig. 9.5a are clearly consistent with the above statements. More importantly, an intriguing phenomenon is found in SG-2 mode. As shown in Fig. 9.5b,  $V_{TH}$  depicts an unexpected drastic increase for  $T$  below 200 K. Moreover, SG-2 mode displays a very abrupt turn-on phenomenon when  $T$  is lower than 150 K. Specifically, SS under SG-2 mode is 3.4 mV/dec at 78 K and 4 mV/dec at 100 K. Finally for DG mode in Fig. 9.5c, the contribution from the channel controlled by the second gate also causes a sudden increase of  $I_D$  at 100 and 78 K.

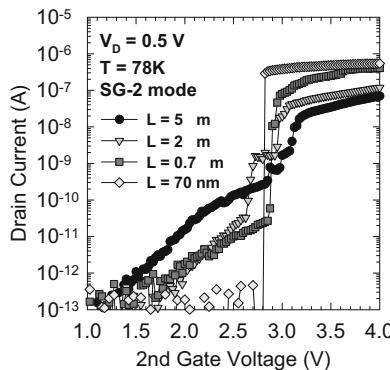
Extracted SS and  $V_{TH}$  as a function of  $T$  are shown in Figs. 9.6a and 9.6b, respectively. Ideal value of SS, which is equal to  $\ln 10 \times kT/q$  where  $k$  is Boltzmann constant and  $q$  the elemental charge [42], is also plotted for comparison. From



**Fig. 9.5** Transfer characteristics at various temperatures for a device with 70-nm channel length under (a) SG-1, (b) SG-2, and (c) DG modes. The behavior of SG-1 mode is in accordance with thermionic emission model while SG-2 mode starts to exhibit very steep SS when  $T$  is below 150 K. Abrupt increase of  $I_D$  for DG mode at 100 and 78 K is caused by the channel controlled by the second gate

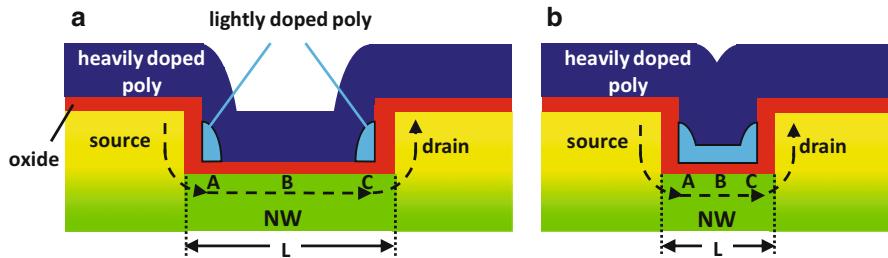


**Fig. 9.6** (a) SS and (b)  $V_{TH}$  as a function of temperature for three operation modes extracted from Fig. 9.5. Ideal value of SS equal to  $\ln 10 \times kT/q$  is also included for comparison in (a)



**Fig. 9.7** Transfer characteristics at 78 K showing impacts of  $L$  in determining the occurrence of steep SS. In conflict with the conventional SCE theory, the subthreshold current is reduced as  $L$  is shortened

Fig. 9.6a, under 100 and 78 K, SS of SG-2 mode is already smaller than the ideal value, an indicator that the conventional drift-diffusion principle [43] is no longer the dominant mechanism governing the transport behavior in these cases. On the other hand, as explained in [34], operation under SG-1 mode tends to exhibit larger  $V_{TH}$  than SG-2 mode due to additional non-gated routes in SG-1 mode *at room temperature*; however, the larger  $V_{TH}$  under SG-2 mode over SG-1 mode within the range of 200–100 K in Fig. 9.6b implies that some implicit effects in SG-2 mode may start to become prominent at low  $T$  regime. Transfer curves at 78 K in Fig. 9.7 reveal that the occurrence of this abrupt turn-on phenomenon has a strong dependency on  $L$  and in stark contrast to the widely accepted theory concerning SCE, the

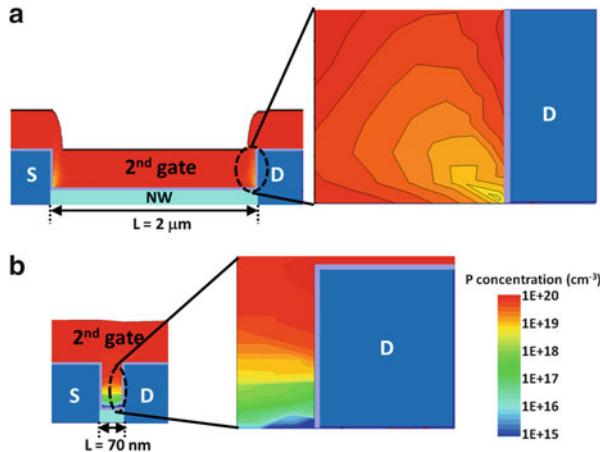


**Fig. 9.8** Schematic structures along the channel length direction showing the second gate-controlled side of channel for a device whose  $L$  is (a) long ( $>100$  nm) and (b) short ( $<100$  nm). Points A and C correspond to regions of the channel adjacent to S/D controlled by the locally thickened gate and point B to the middle of the channel. Because the thickness of the second gate is 100 nm, the part of the second gate between S/D is thicker in (b) than in (a). Reproduced with permission of IOP Publishing from [36]

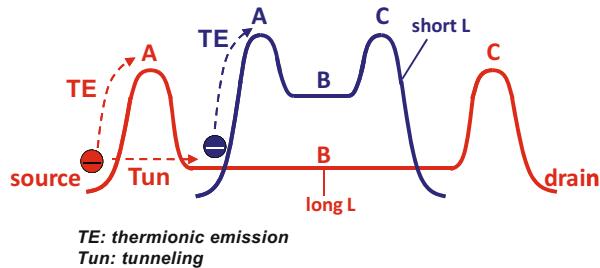
subthreshold current is seen to decrease with reducing  $L$ . Since all of the aforementioned unconventional features are exclusive only in SG-2 mode, the following discussion will focus on SG-2 mode unless otherwise specified.

### 9.6.2 Model Establishment and Discussion

A closer look at the process flow suggests this exclusivity is most likely caused by the different doping methods of the two gates. Namely, doping of the first gate was done by in situ technique that adds phosphine to silane during low pressure chemical vapor deposition (LPCVD) of poly-Si, where very uniform and heavy doping concentration throughout the whole first gate electrode is achieved. As for the second gate, dopants were introduced by phosphorus ion implantation with 15 keV at  $5 \times 10^{15} \text{ cm}^{-2}$  dose. Taking into account the unique device structure measured, i.e., one with raised S/D, a schematic device structure along channel length direction showing the second-gate-controlled side of channel is depicted in Fig. 9.8a for a long  $L$  ( $>100$  nm) device and Fig. 9.8b for a short  $L$  ( $<100$  nm) device. For devices with  $L$  shorter than the thickness of the second gate (100 nm), the central part of trench between S/D is filled with thicker poly-Si, compared with those with  $L$  longer than 100 nm, based on the nature of LPCVD deposition. As a result, for a given implantation energy, the doping concentration is distributed in a manner such that the portion of the second gate farthest from the top surface (i.e., closest to the channel) is only lightly doped, as in Fig. 9.8b, compared with the much higher doping concentration, as shown in Fig. 9.8a. In Fig. 9.8a, b, points A and C correspond to the regions of the channel adjacent to S/D controlled by the locally thickened gate and point B to the middle of the channel. Since the second gate is the thickest in the regions abutting S/D, the lightly doped area of the second



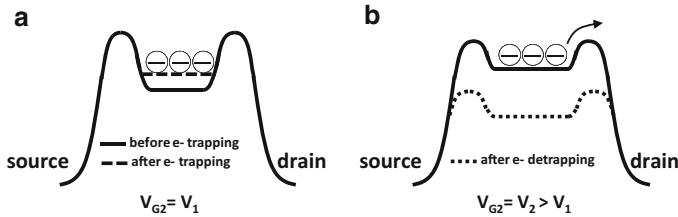
**Fig. 9.9** Simulated dopant distribution in an implanted gate corresponding to a device with (a)  $L = 2 \mu\text{m}$  and (b)  $L = 70 \text{ nm}$ . It is observed that the bottom portion of the second gate has much lower concentration in (b) than in (a). Reproduced with permission of IOP Publishing from [36]



**Fig. 9.10** Qualitative band diagrams for devices with long and short  $L$ . Points A, B, and C correspond to the regions labeled in Fig. 9.8. For a short  $L$  device, thermionic emission is the dominant transport mechanism while both thermionic emission and tunneling should be considered for a long  $L$  device. Reproduced with permission of IOP Publishing from [36]

gate spreads to a larger extent in the channel edge than the middle of the channel. The above surmise is clearly confirmed by simulation results in Fig. 9.9.

This nonuniform dopant distribution then leads to variation in gate controllability along the channel, i.e., gate to channel capacitance is the largest in the middle of the channel and decreases toward the channel edge. In this aspect, akin to [44] where the reported non-overlapped device leads to potential dip in the middle of the channel, electrostatic potential along the channel exhibits two humps near S/D, as qualitatively depicted in Fig. 9.10. Due to the good step coverage offered by LPCVD, those segments of the second gate of a long  $L$  device making contact with the central channel are still heavily doped (Fig. 9.8a), explaining the reason the potential of the central channel is much lower in a long  $L$  device than that of a short

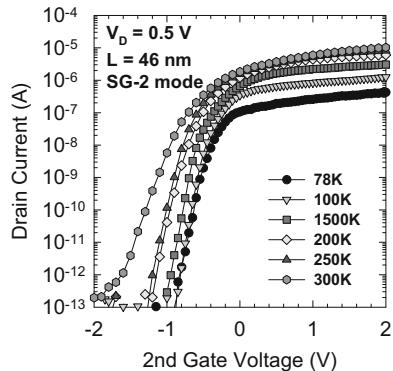


**Fig. 9.11** Proposed model for the origin of steep SS. (a) For a low second gate voltage ( $V_{G2} = V_1$ ), after electron trapping in the channel (dashed line), the potential is raised from its original level (solid line), which deters further injection from the source. (b) For a larger second gate voltage ( $V_{G2} = V_2$ ) that considerably reduces the barrier height at drain side, electrons are detrapped and flow to the drain, leading to channel potential drop (dotted line) and an abrupt increase of drain current. Reproduced with permission of IOP Publishing from [36]

one, as shown in Fig. 9.10. Meanwhile, even though the potential barriers from the two humps are always present for all the characterized  $L$ , the barrier height of the humps is smaller in a long  $L$  device owing to the stronger gate fringing fields from the more heavily doped gate for a given second gate voltage. With this picture in mind, the behavior of transfer characteristics with strong  $L$  dependency shown in Fig. 9.7 becomes reasonable based on the model shown in Fig. 9.11. Before further discussion, it should be noted that the tunneling current in poly-Si is often expected to have minor contribution because in practical situations thermionic emission current is always the preponderant component [38]. Nonetheless, tunneling should become appreciable at cryogenic ambient and in the scenario when the tunneling barrier is narrow enough.

For a short  $L$  device, the large barrier height of the humps and the high channel potential in virtue of lower second gate doping concentration render tunneling less likely to happen, and thermionic emission then assumes a major role even at low  $T$ . In this regard, as  $T$  is lowered, the thermionic emission current is gradually reduced, which in turn increases  $V_{TH}$ , and because the barrier heights of the two humps are essentially independent of  $T$ , it is rational that below a certain  $T$  when thermal energy is far smaller than the barrier height, the second gate voltage must be high enough ( $V_{G2} = V_1$ ) to reduce the barrier height at point A in Fig. 9.10 to a extent that carriers are able to be thermally emitted from the source. Yet instead of directly transporting to drain, carriers coming from the source to the channel will get trapped therein (in point B of Fig. 9.10), raising the channel potential level and preventing further injection from the source, as illustrated in Fig. 9.11a. This scenario will persist until the barrier height present at point C is reduced significantly by a higher applied gate voltage ( $V_{G2} = V_2 > V_1$ ) for the trapped electrons to overcome the barrier with ease and be released to the drain. As a consequence, the channel potential is lowered suddenly and an abrupt increase of drain current is observed (Fig. 9.11b). To further validate our argument, the elimination of the barriers is achieved by in situ doping both in the first and second gates, and the

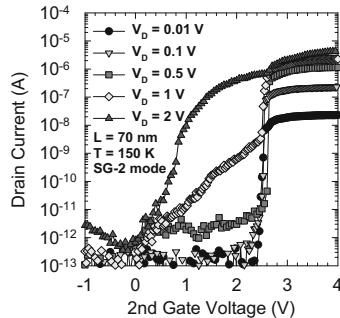
**Fig. 9.12** Transfer characteristics as a function of temperature for a device whose first and second gates are both in situ doped. Compared with Fig. 9.5b, steep SS completely vanishes owing to the removal of barriers when very uniform doping concentration throughout the whole second gate electrode is achieved by in situ doping



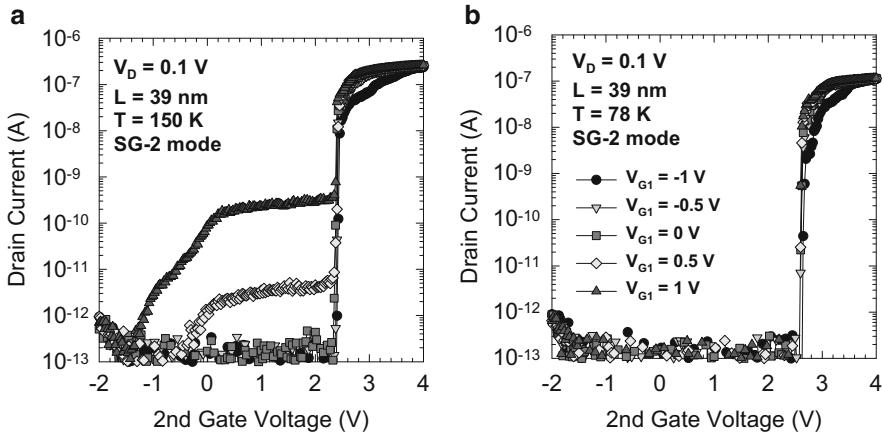
characteristics in Fig. 9.12 evidently corroborate the previous model that implanted gate is indeed the culprit for the peculiar abrupt switching phenomenon.

With a long  $L$  device, the same theory still applies. As mentioned previously with respect to the channel potential profile, the potential difference between the central channel and the source is sufficiently small so that tunneling through the barrier is possible besides thermionic emission. The critical  $T$  below which thermionic emission is no longer possible without strong gate-induced barrier reduction is supposed to be lower for a long  $L$  device thanks to its smaller magnitude of barrier height. It is then expected that thermionic emission and tunneling are both the participating conduction mechanisms, explaining the larger subthreshold current for a long  $L$  device. Further lowering of the measurement  $T$  can help verify this statement by examining the evolution of mobility and  $V_{TH}$  with respect to  $T$  based on the fact that the degree of  $T$  dependence of thermionic emission and tunneling is drastically different; that is, tunneling should have little temperature dependence as opposed to thermionic emission, which as its name suggests is strongly temperature dependent. Accurate modeling of the shape and magnitude of barriers is of significance as well in distinguishing between the individual contribution from thermionic emission and tunneling processes and is still under investigation owing to the complex structure of our proposed device. Nevertheless, the electron-trapping effect can also occur for a long channel device if  $T$  is low enough, so that an abrupt switching (e.g., see  $L = 0.7 \mu\text{m}$  in Fig. 9.7), though with a larger SS compared with a device with short  $L$ , is attained. And the  $L = 5 \mu\text{m}$  device in Fig. 9.7 exhibits a change of SS at  $V_{G2} = 3.1 \text{ V}$  as well on account of the onset of a sudden barrier lowering.

Given that the dominant factor determining the occurrence of abrupt transition lies in the barrier height present at S/D, the impact of drain-induced barrier lowering (DIBL) is investigated in Fig. 9.13. Consistent with the proposed model, the profound influence exerted by DIBL which lessens the barrier height at the source side with increasing drain bias is apparently demonstrated. Making use of the separate-gated property, Fig. 9.14a plots the transfer characteristics at 150 K as a function of the first gate bias ( $V_{G1}$ ).  $V_{G1}$  here ranges from -1 to 1 V in 0.5 V step.

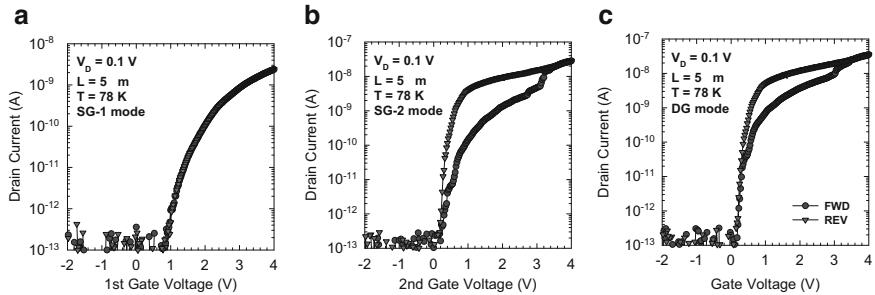


**Fig. 9.13** Drain-voltage-dependent transfer characteristics suggest DIBL could eliminate the occurrence of steep SS, in agreement with the proposed model

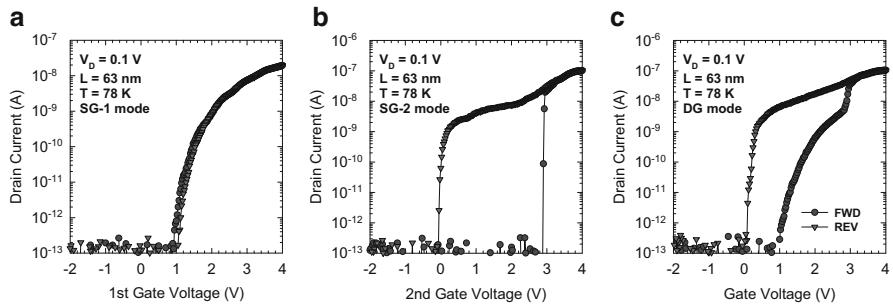


**Fig. 9.14** Transfer characteristics under SG-2 mode for various  $V_{G1}$  at (a) 150 K and (b) 78 K.  $V_{G1}$  here ranges from  $-1$  to  $1$  V in  $0.5$  V step

Subthreshold current becomes appreciable only when  $V_{G1}$  is higher than 0.5 V because under these conditions the channel controlled by the first gate starts to make contribution to the overall drain current, as can be seen from Fig. 9.5a. Nevertheless, SS remains the same irrespective of  $V_{G1}$  since the second gate still provides the preponderant control over the channel, and the maximum  $V_{G1}$  applied is less than what is required to invert the first-gate-controlled channel. Similar behaviors can be noted in 78 K in Fig. 9.14b except that now even  $V_{G1} = 1$  V only corresponds to accumulation for the channel on its side, and negligible IV curve shift is obtained as a consequence.



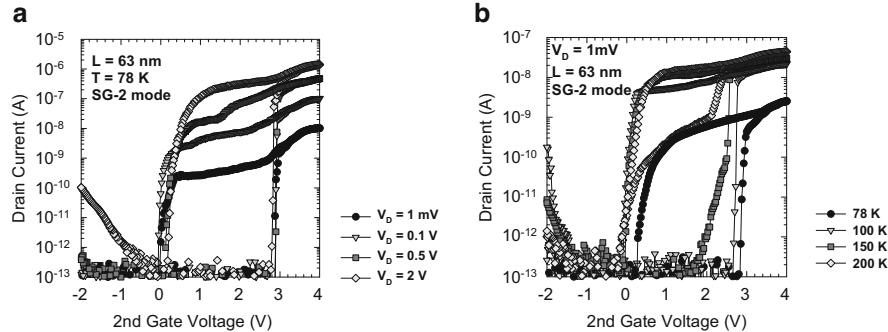
**Fig. 9.15** Transfer characteristics for a device with  $L = 5 \mu\text{m}$  under forward and reverse sweeping of the gate voltage for (a) SG-1, (b) SG-2, and (c) DG modes



**Fig. 9.16** Transfer characteristics for a device with  $L = 63 \text{ nm}$  under forward and reverse sweeping of the gate voltage for (a) SG-1, (b) SG-2, and (c) DG modes

### 9.6.3 Observation of Hysteresis and Single Electron Effects

Since the proposed model implies that the conduction mechanism is similar to a feedback process, which has been utilized in achieving steep SS devices [45], the forward and reverse sweeping of the gate voltage is expected to result in hysteresis of transfer characteristics, as evidenced in Figs. 9.15 and 9.16 for  $L = 5 \mu\text{m}$  and  $63 \text{ nm}$  at  $78 \text{ K}$ , respectively. Here the applied drain voltage is  $0.1 \text{ V}$ . Negligible hysteresis in SG-1 mode indicates that it still obeys the drift-diffusion principle so that the turn-on and turn-off processes are reversible, and the effect of trapping in grain boundaries of poly-Si can be ruled out since plasma treatment was performed on the devices prior to characterization [46]. In contrast, SG-2 mode exhibits apparent hysteresis, with  $V_{\text{TH}}$  window at  $V_D = 0.1 \text{ V}$  being  $0.7$  and  $3 \text{ V}$  for  $L = 5 \mu\text{m}$  and  $63 \text{ nm}$ , respectively. This is because of the voluminous mobile charges existing in the on-state, so during the reverse sweeping, the device cannot be turned off at the same voltage as it is turned on. In Fig. 9.15b, reverse sweeping displays a steeper SS than the forward one probably due to the additional effect of interface charges. A more detailed interpretation will be given later. It is interesting



**Fig. 9.17** Forward and reverse sweeping of transfer characteristics with (a) the drain voltage and (b) temperature as a parameter

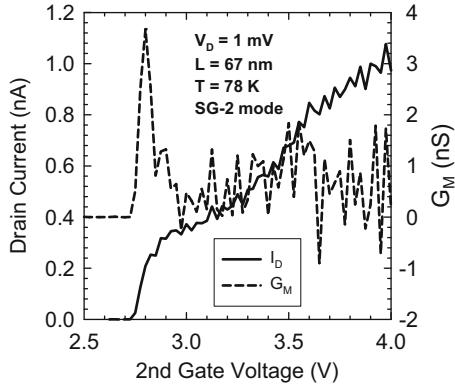
to note that SS of the reverse sweeping is larger than that of the forward one in Fig. 9.16b, which is related to their dissimilar effective barrier profiles. At the outset of the reverse sweeping when the second gate voltage is sufficiently large, the barrier heights present at both the middle and edges of the channel are small compared with the thermal energy. As the second gate voltage is gradually decreased, the conduction band edge of the central channel is supposed to be raised at a more rapid rate over that of the portion that is adjacent to S/D. In this regard, the major conduction barrier profile as seen by carriers is similar to that of a conventional MOSFET, and the depth of the potential well is reduced with decreasing gate voltage. For the same reason, the trapping of carriers is less severe in the reverse sweeping process, and a smaller  $V_{TH}$  is obtained as a result. Even though the effect of trapping/detrapping is now dramatically diminished compared with the turn-on process, this mechanism still needs to be taken into account given that during the turn-off process, SS of a short channel device in Fig. 9.16b is steeper than that of a longer one in Fig. 9.15b.

The above statement also provides another perspective on the characteristics in Fig. 9.15b. As noted in Sect. 9.6.2, the effect of trapping/detrapping is actually not negligible for a long  $L$  device; hence, in the forward sweeping curve from  $V_{G2} = 0.1\text{--}3 \text{ V}$ , the subthreshold current suffers carrier trapping and beyond  $V_{G2} = 3 \text{ V}$ , detrapping sets in along with a sudden jump of the drain current. Since electron trapping in this case does not completely suppress the drain current as in a short  $L$  device, SS of forward sweeping is reasonably larger than that of the reverse one.

$V_D$ -dependent  $I_D-V_G$  curves plotted in Fig. 9.17a imply that at 78 K hysteresis is observable at least up to  $V_D = 2 \text{ V}$ , and the hysteresis window is monotonically decreasing with  $T$ , as shown in Fig. 9.17b, which is consistent with the fact that the probability of trapping/detrapping process occurrence is reduced with higher  $T$ .

As a matter of fact, some of the measured devices exhibit drain current oscillation as the gate voltage is varied. Figure 9.18 (left vertical axis) depicts such a behavior by magnifying  $I_D-V_G$  curves in the strong inversion regime. The corresponding  $G_M$  characteristic is shown in Fig. 9.18 as well in the right vertical

**Fig. 9.18** Oscillation of the drain current and resultant  $G_M$  at 1 mV  $V_D$



axis. Although our proposed device is not intended to operate as a single electron transistor (SET), the potential valley in Fig. 9.11 assumes the role of an island, and the two humps to its side are essentially tunneling junctions. Another factor that may contribute to the drain current oscillation is inter-sub-band scattering whose influence becomes greater as the carrier concentration is increased due to the larger number of occupied sub-bands [19].

## 9.7 Conclusion

Making use of the mix-and-match of I-line stepper and e-beam direct writing, independent double-gated poly-Si NW transistor with  $L$  ranging from 39 nm to 5  $\mu\text{m}$  are successfully fabricated and demonstrate excellent SCE immunity in terms of insignificant SS roll-off. And the origin of abrupt turn-on characteristics observed in SG-2 mode at cryogenic ambient is comprehensively studied in this chapter. It is found that the occurrence of this behavior is greatly influenced by  $L$ ,  $T$ , and drain bias. A model taking into account the dopant distribution of an implanted gate is proposed to interpret our findings. It suggests that the non-intentionally formed barriers at the channel edge give rise to carrier trapping effects until an adequately large gate voltage is applied to lower the magnitude of the barriers for the trapped electrons to flow to the drain. Furthermore, since the channel potential profile resembles that of an island confined by two tunneling junctions, single-electron effects are also observed. This kind of process-induced barrier is an attractive scheme to build SETs with simple CMOS-compatible process flow and with the aid of further surface engineering should help realize steep SS transistors at room temperature.

**Acknowledgments** The authors would like to thank the staff at National Nano Device Laboratories (NDL) and Nano Facility Center (NFC) of NCTU for assistance in device fabrication, Prof. Pei-Wen Li of National Central University for support of the cryogenic measurement utilities, and Prof. Bing-Yue Tsui of NCTU for assistance in TEM characterization.

## References

1. Im, M., Han, J.W., Lee, H., Yu, L.E., Kim, S., Kim, C.H., Jeon, S.C., Kim, K.H., Lee, G.S., Oh, J.S., Park, Y.C., Lee, H.M., Choi, Y.K.: Multiple-gate CMOS thin-film transistor with polysilicon nanowire. *IEEE Electron Device Lett.* **29**, 102–105 (2008)
2. Duan, X.: Nanowire thin-film transistors: a new avenue to high-performance macroelectronics. *IEEE Trans. Electron Dev.* **55**, 3056–3062 (2008)
3. Hisamoto, D., Lee, W.C., Kedzierski, J., Takeuchi, H., Asano, K., Kuo, C., Anderson, E., King, T.J., Bokor, J., Hu, C.: FinFET-a self-aligned double-gate MOSFET scalable to 20 nm. *IEEE Trans. Electron Dev.* **47**, 2320–2325 (2000)
4. Lue, H.T., Hsu, T.H., Hsiao, Y.H., Hong, S.P., Wu, M.T., Hsu, F.H., Lien, N.Z., Wang, S.Y., Hsieh, J.Y., Yang, L.W., Yang, T., Chen, K.C., Hsieh, K.Y., Lu, C.Y.: A highly scalable 8-layer 3D vertical-gate (VG) TFT NAND flash using junction-free buried channel BE-SONOS device. In: Symposium on VLSI technology. Digest of technical papers, pp. 131–132 (2010)
5. Hsu, T.H., Lue, H.T., Hsieh, C.C., Lai, E.K., Lu, C.P., Hong, S.P., Wu, M.T., Hsu, F.H., Lien, N.Z., Hsieh, J.Y., Yang, L.W., Yang, T., Chen, K.C., Hsieh, K.Y., Liu, R., Lu, C.Y.: Study of sub-30nm thin film transistor (TFT) charge-trapping (CT) devices for 3D NAND flash application. In: IEDM technical digest, pp. 629–632 (2009)
6. Suk, S.D., Li, M., Yeoh, Y.Y., Yeo, K.H., Ha, J.K., Lim, H., Park, H.W., Kim, D.W., Chung, T. Y., Oh, K.S., Lee, W.S.: Characteristics of sub 5nm tri-gate nanowire MOSFETs with single and poly Si channels in SOI structure. In: Symposium on VLSI technology. Digest of technical papers, pp. 142–143 (2009)
7. Hsiao, Y.H., Lue, H.T., Chen, W.C., Chen, C.P., Chang, K.P., Shih, Y.H., Tsui, B.Y., Lu, C.Y.: Modeling the variability caused by random grain boundary and trap-location induced asymmetrical read behavior for a tight-pitch vertical gate 3D NAND Flash memory using double-gate thin-film transistor (TFT) device. In: IEDM technical digest, pp. 609–612 (2012)
8. Uchikoga, S.: Low-temperature polycrystalline silicon thin-film transistor technologies for system-on-glass displays. *MRS Bull.* **27**, 881–886 (2002)
9. Saitoh, M., Murakami, T., Hiramoto, T.: Large Coulomb blockade oscillations at room temperature in ultranarrow wire channel MOSFET formed by slight oxidation process. *IEEE Trans. Nanotechnol.* **2**, 241–245 (2003)
10. Akhavan, N.D., Afzalian, A., Lee, C.W., Yan, R., Ferain, I., Razavi, P., Fagas, G., Colinge, J. P.: Simulation of quantum current oscillations in trigate SOI MOSFETs. *IEEE Trans. Electron Dev.* **57**, 1102–1109 (2010)
11. Buin, A.K., Verma, A., Anantram, M.P.: Carrier-phonon interaction in small cross-sectional silicon nanowires. *J. Appl. Phys.* **104**, 053716-1–053716-9 (2008)
12. Chau, R., Datta, S., Doczy, M., Doyle, B., Kavalieros, J., Metz, M.: High- $\kappa$ /metal-gate stack and its MOSFET characteristics. *IEEE Electron Device Lett.* **25**, 408–410 (2004)
13. Wong, H.S.P.: Beyond the conventional transistor. *IBM J. Res. Dev.* **46**, 133–168 (2002)
14. Doyle, B., Boyanov, B., Datta, S., Doczy, M., Harelund, S., Jin, B., Kavalieros, J., Linton, T., Rios, R., Chau, R.: Tri-gate fully-depleted CMOS transistors: fabrication, design and layout. In: Symposium on VLSI technology. Digest of technical papers, pp. 133–134 (2003)
15. Yang, F.L., Chen, H.Y., Chen, F.C., Huang, C.C., Chang, C.Y., Chiu, H.K., Lee, C.C., Chen, C.C., Huang, H.T., Chen, C.J., Tao, H.J., Yeo, Y.C., Liang, M.S., Hu, C.: 25 nm CMOS omega FETs. In: IEDM technical digest, pp. 255–258 (2002)

16. Singh, N., Agarwal, A., Bera, L.K., Liow, T.Y., Yang, R., Rustagi, S.C., Tung, C.H., Kumar, R., Lo, G.Q., Balasubramanian, N., Kwong, D.L.: High-performance fully depleted silicon nanowire (diameter  $\leq 5$  nm) gate-all-around CMOS devices. *IEEE Electron Device Lett.* **27**, 383–386 (2006)
17. Natori, K.: Compact modeling of ballistic nanowire MOSFETs. *IEEE Trans. Electron Dev.* **55**, 2877–2885 (2008)
18. Colinge, J.P., Alderman, J.C., Xiong, W., Cleavelin, C.R.: Quantum-mechanical effects in trigate SOI MOSFETs. *IEEE Trans. Electron Dev.* **53**, 1131–1136 (2006)
19. Colinge, J.P., Quinn, A.J., Floyd, L., Redmond, G., Alderman, J.C., Xiong, W., Cleavelin, C. R., Schulz, T., Schruefer, K., Knoblinger, G., Patruno, P.: Low-temperature electron mobility in trigate SOI MOSFETs. *IEEE Electron Device Lett.* **27**, 120–122 (2006)
20. Thelander, C., Agarwal, P., Brongersma, S., Eymery, J., Feiner, L.F., Forchel, A., Scheffler, M., Riess, W., Ohlsson, B.J., Gosele, U., Samuelson, L.: Nanowire-based one-dimensional electronics. *Mater. Today* **9**, 28–35 (2006)
21. Majima, H., Ishikuro, H., Hiramoto, T.: Experimental evidence for quantum mechanical narrow channel effect in ultra-narrow MOSFETs. *IEEE Electron Device Lett.* **21**, 396–398 (2000)
22. Rustagi, S.C., Singh, N., Lim, Y.F., Zhang, G., Wang, S., Lo, G.Q., Balasubramanian, N., Kwong, D.L.: Low-temperature transport characteristics and quantum-confinement effects in gate-all-around Si-nanowire n-MOSFET. *IEEE Electron Device Lett.* **28**, 909–921 (2007)
23. Mojica, D.C., Niquet, Y.M.: Stark effect in GaN/AlN nanowire heterostructures: influence of strain relaxation and surface states. *Phys. Rev. B* **81**, 195313-1–195313-10 (2010)
24. Jin, S., Fischetti, M.V., Tang, T.W.: Differential conductance fluctuations in silicon nanowire transistors caused by quasiballistic transport and scattering induced intersubband transitions. *Appl. Phys. Lett.* **92**, 082103-1–082103-3 (2009)
25. Wang, J., Polizzi, E., Lundstrom, M.: A computational study of ballistic silicon nanowire transistors. In: IEDM technical digest, pp. 695–698 (2003)
26. Natori, K.: Ballistic metal-oxide-semiconductor field effect transistor. *J. Appl. Phys.* **76**, 4879–4890 (1994)
27. Natori, K.: Scaling limit of the MOS transistor—a ballistic MOSFET. *IEICE Trans. Electron.* **E84-C**, 1029–1036 (2001)
28. Barral, V., Poiroux, T., Barraud, S., Andrieu, F., Faynot, O., Munteanu, D., Autran, J.L., Deleonibus, S.: Evidences on the physical origin of the unexpected transport degradation in ultimate n-FDSOI devices. *IEEE Trans. Nanotechnol.* **8**, 167–173 (2009)
29. Cros, A., Romanjek, K., Fleury, D., Harrison, S., Cerutti, R., Coronel, P., Dumont, B., Pouydebasque, A., Wacquez, R., Duriez, B., Gwoziecki, R., Boeuf, F., Brut, H., Ghibaudo, G., Skotnicki, T.: Unexpected mobility degradation for very short devices: a new challenge for CMOS scaling. In: IEDM technical digest, pp. 663–666 (2006)
30. Shur, M.S.: Low ballistic mobility in submicron HEMTs. *IEEE Electron Device Lett.* **23**, 511–513 (2002)
31. Wang, R., Liu, H., Huang, R., Zhuge, J., Zhang, L., Kim, D.W., Zhang, X., Park, D., Wang, Y.: Experimental investigations on carrier transport in Si nanowire transistors: ballistic efficiency and apparent mobility. *IEEE Trans. Electron Dev.* **55** (2008)
32. Knoch, J., Riess, W., Appenzeller, J.: Outperforming the conventional scaling rules in the quantum-capacitance limit. *IEEE Electron Device Lett.* **29**, 372–374 (2008)
33. Appenzeller, J., Knoch, J., Björk, M.T., Riel, H., Schmid, H., Riess, W.: Toward nanowire electronics. *IEEE Trans. Electron Dev.* **55**, 2827–2845 (2008)
34. Lin, H.C., Chen, W.C., Lin, C.D., Huang, T.Y.: Performance enhancement in double-gated poly-Si nanowire transistors with reduced nanowire channel thickness. *IEEE Electron Device Lett.* **30**, 644–646 (2009)
35. Chen, W.C., Lin, H.C., Chang, Y.C., Lin, C.D., Huang, T.Y.: In situ doped source/drain for performance enhancement of double-gated poly-Si nanowire transistors. *IEEE Trans. Electron Dev.* **57**, 1608–1615 (2010)

36. Chen, W.C., Lin, H.C., Lin, Z.M., Hsu, C.T., Huang, T.Y.: A study on low temperature transport properties of independent double-gated poly-Si nanowire transistors. *Nanotechnology* **21**, 435201-1–435201-7 (2010)
37. Falckenberg, R., Doering, E., Oppolzer, H.: Surface roughness and grain growth of thin p-doped polycrystalline Si-films. In: Proceedings of the Electrochemical Society Meeting, p. 1492 (1979)
38. Seto, J.Y.W.: The electrical properties of polycrystalline silicon films. *J. Appl. Phys.* **46**, 5247–5254 (1975)
39. Lin, Z.M., Lin, H.C., Chen, W.C., Huang, T.Y.: Insight into the performance enhancement of double-gated polycrystalline silicon thin-film transistors with ultrathin channel. *Appl. Phys. Lett.* **96**, 072108-1–072108-3 (2010)
40. Puigdollers, J., Dosev, D., Orpella, A., Voz, C., Peiro, D., Bertomeu, J., Marsal, L.F., Pallares, J., Andreu, J., Alcubilla, R.: Microcrystalline silicon thin film transistors obtained by hot-wire CVD. *Mater. Sci. Eng. B* **69–70**, 526–529 (2000)
41. Puigdollers, J., Orpella, A., Dosev, D., Voz, C., Peiro, D., Pallares, J., Marsal, L.F., Bertomeu, J., Andreu, J., Alcubilla, R.: Thin film transistors obtained by hot wire CVD. *J. Non Cryst. Solids* **266–269**(part 2), 1304–1309 (2000)
42. Sze, S.M., Ng, K.K.: Physics of semiconductor devices, 3rd edn. Wiley-Interscience, Hoboken (2007)
43. Taur, Y., Ning, T.H.: Fundamentals of modern VLSI devices. Cambridge University Press, Cambridge (1998)
44. Boeuf, F., Jehl, X., Sanquer, M., Skotnicki, T.: Controlled single-electron effects in nonoverlapped ultra-short silicon field effect transistors. *IEEE Trans. Nanotechnol.* **2**, 144–148 (2003)
45. Padilla, A., Yeung, C.W., Shin, C., Hu, C., Liu, T.J.K.: Feedback FET: a novel transistor exhibiting steep switching behavior at low bias voltages. In: IEDM technical digest, pp. 171–174 (2008)
46. Lin, H.C., Hung, C.H., Chen, W.C., Lin, Z.M., Hsu, H.H., Huang, T.Y.: Origin of hysteresis in current–voltage characteristics of polycrystalline silicon thin-film transistors. *J. Appl. Phys.* **105**, 054502-1–054502-6 (2009)

# Chapter 10

## Towards Drain Extended FinFETs for SoC Applications

Mayank Shrivastava, Harald Gossner, and V. Ramgopal Rao

**Abstract** This chapter highlights the importance of drain extended class of high voltage MOS devices for advanced implementations of System on Chip (SoC) applications using FinFET or other tri-gate technologies. The working principle of drain extended MOS device is explained and an understanding is built in order to bridge existing drain extended planar devices and future drain extended, FinFETs or tri-gate devices. In this connection, a recently proposed Drain extended FinFET device is discussed for high voltage as well as high speed applications. This shows a better  $R_{ON}$  vs.  $V_{BD}$  trade-off when compared to a conventional device option. Finally, device design and optimization guidelines have been discussed for the new drain extended FinFET device.

### 10.1 Introduction

A System on Chip (SoC) concept offers exciting opportunities for a wide range of hand-held and mobile applications, including smart phones, high-end gaming consoles, and tablet PCs [1–3]. However, SoC designs come with unique challenges

---

Mayank Shrivastava was formerly with Intel Corp., Mobile and Communications Group, Munich, Germany.

M. Shrivastava (✉)  
Department of Electronic Systems Engineering, Indian Institute of Science,  
Bangalore, Karnataka, India  
e-mail: [mayank@dese.iisc.ernet.in](mailto:mayank@dese.iisc.ernet.in)

H. Gossner  
Intel Mobile Communication, Neubiberg, Germany  
e-mail: [harald.gossner@intel.com](mailto:harald.gossner@intel.com)

V.R. Rao  
Electrical Engineering Department, IIT Bombay, Mumbai, Maharashtra, India  
e-mail: [rao@ee.iitb.ac.in](mailto:rao@ee.iitb.ac.in)

when compared to processor designs, which include very low standby power, high packing density, high precision analog circuits, mixed signal design requirements, and high voltage circuitry. This requires specific device enablement. Also, circuit stress factors like electrostatic discharge (ESD) and self-heating become much more severe in SoC designs. An integration of RF power amplifiers will further enhance these complexities in SoC designs.

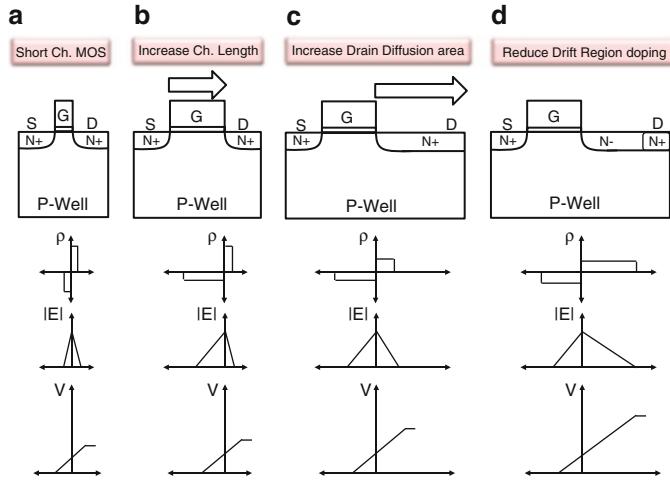
As the planar bulk MOS device reaches its scaling limits [4], FinFETs [5] or trigate FETs are gaining importance as the technology options for sub-20 nm gate lengths [6]. FinFET like devices are also found to be a suitable option for SoC applications [7]. Along with logic and memory blocks, which cover most of the area in a CPU chip, a wireless SoC design also requires RF and mixed signal features, which include Bluetooth, GSM and Wi-Fi transceiver modules, integrated power amplifiers, and power management blocks. Moreover, SoC designs usually require a high number of diverse I/O interfaces. In terms of transistor specifications and voltage handling capability, a technology for CPU like products covers a limited spectrum, where the devices are designed as per the digital specs, focusing on operating voltages less than 1 V. However, when we consider the technology requirements for SoC products, we need to account for the performance requirements of analog and RF components. Further, high voltage devices need to be integrated on the chip, without sacrificing on the overall performance of the particular technology node.

An SoC chip in advanced CMOS, consisting of various functional blocks, falls into three major voltage classes—(a) low voltages: 0.8–1.2 V, (b) medium voltage (HV): 1.8–2.5 V, and (c) high voltages: 3.3–10 V. In conventional/planar bulk CMOS technology, low and medium voltage blocks are traditionally implemented by the use of thin and thick gate-oxide devices with different channel lengths, while Drain extended MOS devices offer a beneficial implementation of high voltage functionalities [8]. Due to cost-effective availability of devices in various voltage classes within planar CMOS technologies, SoC implementation using planar CMOS is feasible. However, SoC implementation and integration of high voltage functionalities are still considered to be challenges for FinFET technologies, as implementation of high voltage devices in these technologies tends to be extremely difficult.

## 10.2 Planar Drain Extended MOS Devices

### 10.2.1 *From Low to High Voltage Devices*

We present in this section high voltage or high power MOS devices suitable for SoC applications. The very first power MOS device was proposed in early 1970s [9]. D-MOS and VMOS [9–11] were some of the very first inventions in this area. While these devices have been around for more than 40 years, their implementation in SoCs is still very challenging.



**Fig. 10.1** Step-by-step explanation of reduced surface field (RESURF) effect and guidelines for implementing a high voltage MOS device in a low voltage technology. Note that the modification depicted in (c) does not significantly increase the junction breakdown voltage in advanced CMOS technologies with highly doped (N++) source/drain regions

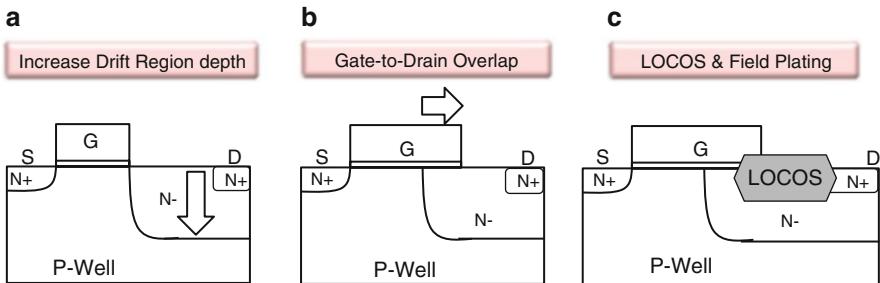
Figure 10.1 shows how a low voltage technology can be modified in order to integrate a high voltage device. The basic concept is to reduce surface electric field and to suppress early avalanche generation, which increases the junction breakdown voltage. Poisson equation provides guidance for the reduction of the electrical field for an applied voltage  $V$ :

$$\frac{\partial^2 V}{\partial x^2} = -\frac{\partial E}{\partial x} = -\frac{q\rho}{\epsilon},$$

where  $E$  is electric field,  $q$  is charge, and  $\rho$  is space charge density.

As a first step, channel length is increased (as depicted in Fig. 10.1b), which reduces the channel electric field in lateral direction. Gate oxide thickness can also be increased in order to reduce the drain-to-gate electric field. However, both measures degrade MOS performance. Furthermore, drain diffusion area is increased (as depicted in Fig. 10.1c), relaxing the space charge distribution in case of a moderately doped ( $< 5 \times 10^{19} \text{ cm}^{-3}$ ) source/drain region, thereby, suppressing the electric field further. As a final step, doping in the drain drift region (except drain contact) is lowered, which widens the depletion region reducing the peak electric field for a given drain bias by spreading the space charge region in a reduced surface field (RESURF) device (as depicted in Fig. 10.1d). Note that modification depicted in Fig. 10.1c does not significantly increase junction breakdown voltage in advanced CMOS technologies with highly doped ( $> 1 \times 10^{21} \text{ cm}^{-3}$ ) source/drain regions.

On the other hand, RESURF concept has some limitations. It inevitably leads to a trade-off between ON resistance and junction breakdown voltage. In order to



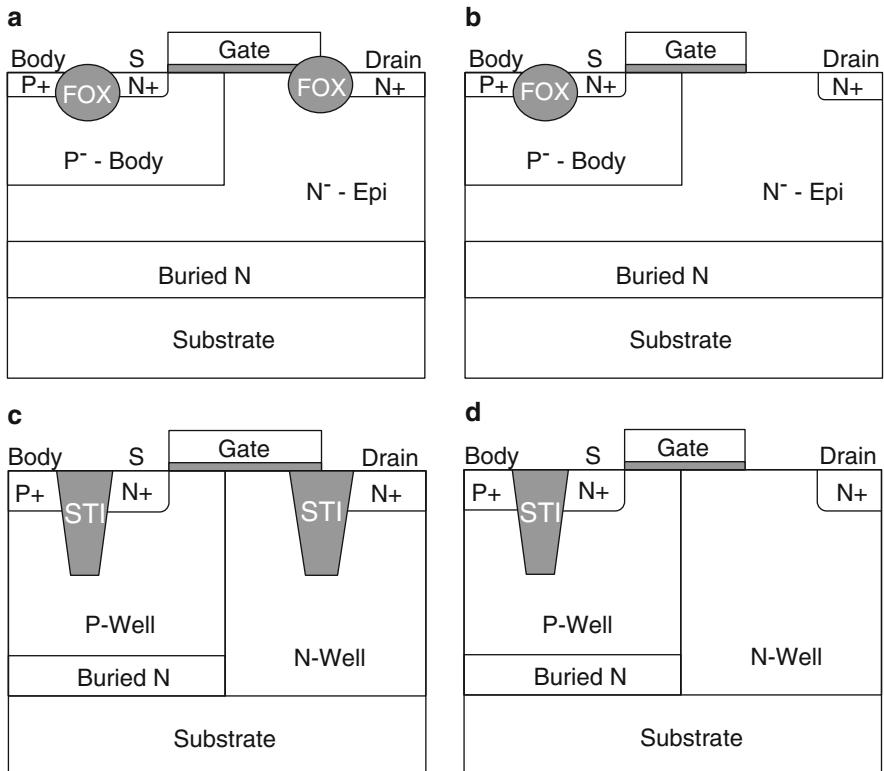
**Fig. 10.2** Implementation of drain extended MOS device by (a) increasing the depth of lightly doped drain region, (b) increasing gate-to-drain overlap, and (c) incorporating a field oxide in the drift region and underneath gate-drain overlap

increase breakdown voltage of a RESURF MOS device, length of drift region must be increased, which also increases ON resistance. Beyond this, in advanced CMOS technologies featuring high well doping, breakdown voltage of RESURF MOS is finally limited by well doping concentration, regardless of drift/RESURF region length.

To mitigate this problem, the concept of drain extended MOS (DeMOS) was proposed. By increasing the depth of lightly doped drain of DeMOS or LDMOS devices, a more extended well junction area is exploited, which relaxes space charge density (as depicted in Fig. 10.2a). Reduced space charge density leads to lower peak electric field at the well junction and therefore improves the junction breakdown voltage. If a high electric field at the gate-drain edge dominates the breakdown, a gate-to-drain overlap helps to mitigate this effect (as depicted in Fig. 10.2b). As this increases the parasitic capacitance and nonlinearity, field oxide is introduced in the drift region, which in addition protects the drain-to-gate breakdown under overvoltage stress (see Fig. 10.2c).

Figure 10.3 shows some of the high voltage device architectures in advance BiCMOS and CMOS technology nodes. In BiCMOS or automotive BCD technologies, these devices are known as LDMOS devices. In CMOS technologies they are referred to as Drain extended MOS devices. There are minor structural differences. However, the fundamental mechanism is the same.

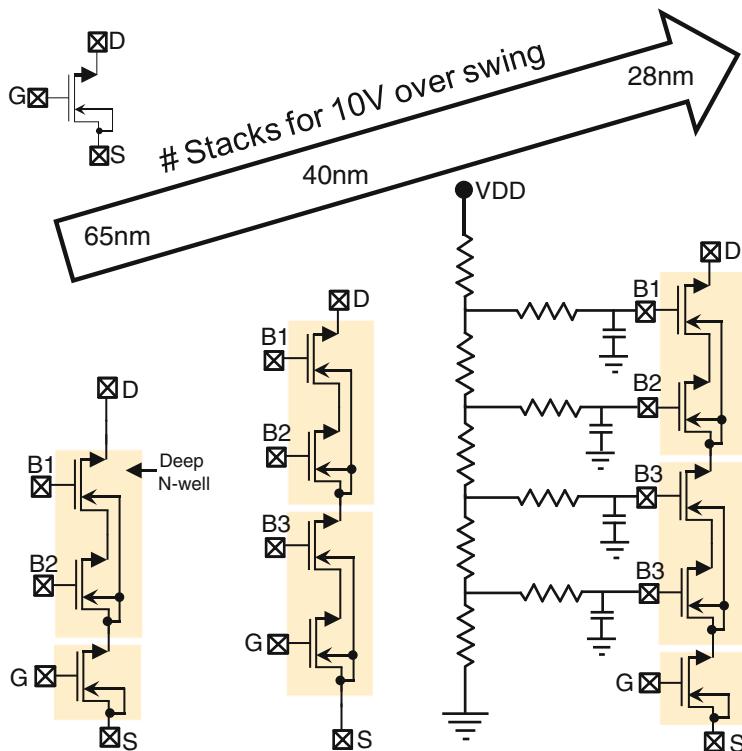
Another approach to realizing high voltage equivalent device or making low voltage devices capable of handling high drain-to-source voltage is by a stacking of low voltage transistors. Let's first look into why stacking of low voltage devices is no more a good option for high voltage circuit implementation. Ultimate goal of technology scaling is to reduce effective oxide thickness (EOT) and channel length in order to improve the performance. The same trend also applies for thick oxide devices, typically offered in low power (LP) or SoC processes. Hence, maximum allowed voltage across thick oxide device also scales with technology scaling. Figure 10.4 shows the number of devices required in a stacked configuration for different CMOS technologies, in order to achieve a swing of 10 V. For example a typical 65 nm technology provides a thick oxide device with a maximum voltage



**Fig. 10.3** Various drain extended MOS devices in advanced BiCMOS and CMOS technologies. (a, b) LDMOS devices typically used in automotive technologies and (c, d) DeMOS devices used in CMOS technologies

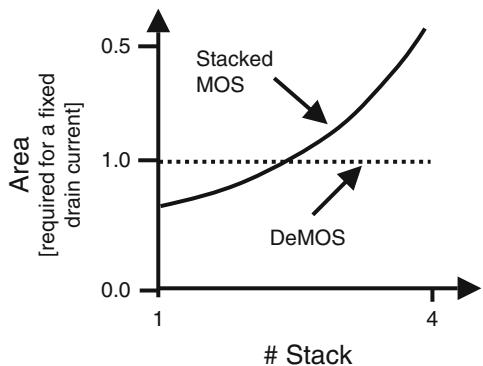
of 3.6 V. Therefore, three devices in stack can meet the purpose. However, the stack increases to four and five devices in 40 and 28 nm node, respectively, where the typical maximum voltage across thick oxide device is reduced down to less than 2.5 V. A similar trend is expected in future SoC technologies and would probably become even critical in FinFET or other nanoscale technologies.

Stacking reduces the MOS performance and costs additional area when compared to a drain extended MOS device, as depicted in Fig. 10.5. Stacking also gives rise to nonlinear MOS behavior, which, as a consequence, leads to additional circuit design efforts. Moreover, stacking requires an additional biasing circuitry and multiple guard rings, which cost additional area (Fig. 10.4). Depending on the application, design of biasing circuitry can cause additional design effort. In conclusion, a drain extended MOS as a single device solution provides a viable, more cost, and performance efficient option for high voltage circuits in advanced technology nodes.

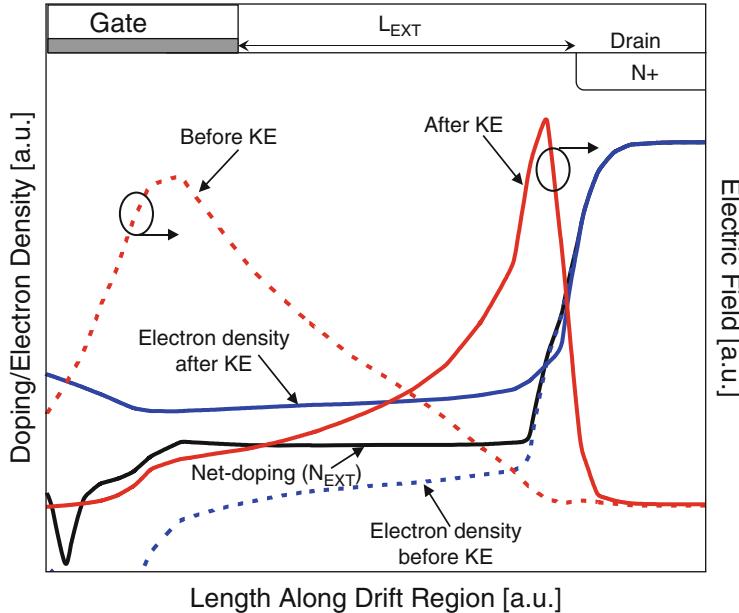


**Fig. 10.4** Various stacked transistor topologies supporting 10 V RF swing for downscaled technologies

**Fig. 10.5** Area benefit of drain extended MOS device over stack of low voltage MOS devices



Still, the design of a drain extended MOS device meeting high power RF or high speed IO performance targets is not trivial. The right balance of ON-resistance, breakdown voltage, and reliability has to be found depending on the application. Significant device level understanding and modeling efforts are necessary to address this trade-off.



**Fig. 10.6** Electron density (blue) and electric field (red) along the drain extension (drift region). Shift in the electric field profile from channel edge under the gate towards N<sup>+</sup> drain occurs during space charge modulation (SCM), commonly known as Kirk effect (KE) [15]

### 10.2.2 Fundamental Design Limit: Space Charge Modulation

Space charge modulation is a fundamental effect associated with drain extended MOS device design with lowly doped drift region. This situation exists under high current injection conditions, when majority carrier concentration in the drift region exceeds background doping density of the drift region, leading to a shift of electric field peak from the well junction to the highly doped drain region, as depicted in Fig. 10.6. The charge modulation of the lowly doped drain extension region and localization of peak electric field at the highly doped drain region is also referred as “Kirk-effect” [12], which was investigated in more detail using TCAD simulations by [13]. The onset of Kirk effect can approximately be described by the equation below:

$$J_{\text{KIRK}} = q \cdot N_{\text{DRIFT}} \cdot v_{\text{SAT}},$$

where  $J_{\text{KIRK}}$  is the majority carrier current density required for the onset of charge modulation,  $q$  is charge,  $N_{\text{DRIFT}}$  is the net doping density inside the lowly doped drift region and  $v_{\text{SAT}}$  is the majority carrier saturation velocity in drift region. Under space charge modulation condition gate loses its control on channel. Loss of gate control under space charge modulated condition may lead to early quasi-saturation in drain extended MOS devices [14]. Kirk effect was also found to be a root cause for peculiar impact ionization behavior [14] and early filament formations [15].

### 10.3 FinFET Technology

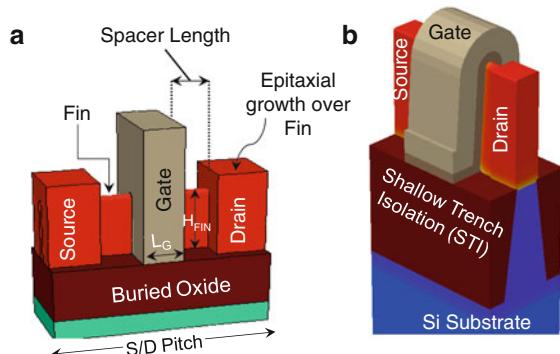
FinFET is a double- or tri-gate metal oxide semiconductor (MOS) device initially proposed to be used for technologies below 20 nm node CMOS [16]. In general FinFETs have a Fin region (Fig. 10.7), encapsulated by gate dielectric/electrode and spacers, which is connected to source and drain contacts at its two ends respectively. In the presence of an appropriate bias at the gate electrode, channel forms on the surface (and partially inside) of Fin region.

Due to its 3D geometry (Fig. 10.7), where the channel region is tightly enclosed by gate dielectric and gate electrode, channel electrostatic outperforms its planar bulk FET counterpart. Improved electrostatic control results into better short channel performance and hence higher scalability, when compared to a planar MOSFET device.

FinFETs like planar MOSFET devices can be manufactured on bulk as well as silicon-on-insulator (SOI) wafers [17, 18]. When manufactured on a SOI wafer (Fig. 10.7a), Source/Drain (S/D), Fin, and Gate electrode of a FinFET device are completely isolated with the silicon substrate by an insulator material in between. However, when manufactured on a bulk wafer (Fig. 10.7b), the fin is in electrical and thermal contact with the bulk material (e.g., silicon substrate) while the gate electrode and contact regions are isolated from silicon substrate by shallow trench isolation (STI).

Following are various design measures associated with FinFET technology and devices, which must be considered carefully:

- (a) *S/D doping profiles*: S/D doping profiles are critical for FinFET devices. A highly doped S/D with an abrupt junction to the channel region is beneficial for short channel performance. However, getting towards such a profile is difficult when using an ion-implant process for source and drain region. Recently, an implant free process was proposed, which can achieve a very abrupt profile [5].
- (b) *Channel or Fin doping*: Like in traditional CMOS processes, threshold voltage ( $V_T$ ) adjustment can be achieved by channel doping. However, channel doping in FinFET technology leads to random dopant fluctuations and  $V_T$  variations,



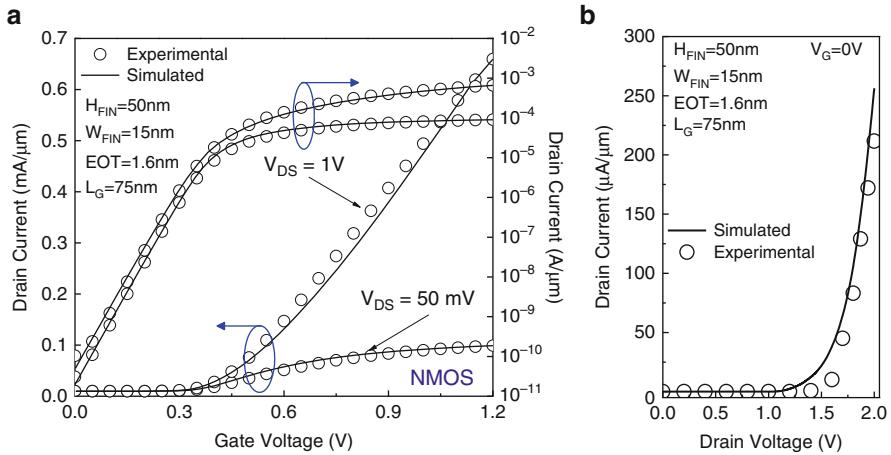
**Fig. 10.7** 3D view of a FinFET device realized over (a) SOI and (b) bulk substrates.  $L_G$  is the gate length and  $H_{FIN}$  is the fin height

which is attributed to a very small volume of fin region. Thus, an undoped channel is preferred for FinFET devices. In this case the threshold voltage  $V_T$  can be determined by the work-function engineering of the gate metal.

- (c) *Stress engineering*: Stress engineering in the channel region plays a key role in the advanced CMOS technology and is counted among the various performance boosters. There are two ways of applying stress over the channel region—(1) through external stress liners and (2) epitaxial growth of SiGe (for PMOS) or SiC (for NMOS) inside the S/D region. In advanced planar CMOS both of these options are available; however, the same cannot easily transferred to FinFET technology. Due to 3D nature of FinFET devices, epitaxial growth of SiGe/SiC inside the S/D region is technologically more challenging, and the built-in stress will relax along the fin. Therefore, mostly external stress liners are used for performance boosting. Recently, it was found that the impact of stress liners is larger on the bulk FinFET devices than that on the FinFET devices realized over SOI substrate. It was found that the induced stress on the bulk FinFETs is stronger, which was attributed to fin's connectivity to the Si substrate leading to decrease of stress relaxation in the Fin region [19].
- (d) *S/D underlap/overlap*: Overlap of highly doped S/D with gate region can give rise to inner fringe capacitance and reduce short channel performance. It was recently proposed that S/D with an underlap with gate edge can be beneficial for short channel performance [20]. However, underlap design is only effective with abrupt S/D profile [5].
- (e) *Raised S/D*: Raised S/D geometry, which is usually realized by selective epitaxial growth on top of S/D Fin region, is desired to reduce S/D contact and parasitic resistance. However, raised S/D geometry also leads to higher gate-S/D capacitance.
- (f) *Substrate/well doping*: Substrate or well doping does not play any role in SOI FinFET design; however, it is an important design parameter for bulk FinFETs. Substrate/well doping in bulk FinFET devices suppresses the drain-to-source punch through current in the lower part of the fin. There is a delicate balance when optimizing the doping profile of the well region. While lower doping levels give rise to a higher punch through current, a too high doping level increases the leakage current due to unwanted band-to-band tunneling.

## 10.4 TCAD Calibration and Framework

Undoped tri-gate SOI FinFET device, fabricated with a mid-gap metal (TiN) gate, SiON dielectric (EOT = 1.6 nm), channel length ( $L_G$ ) of 75 nm, and target fin width ( $W_{FIN}$ ) of 15 nm were used for TCAD calibration. Figure 10.8a shows the calibration of TCAD model parameters for drift–diffusion transport considering quantum corrections at the oxide–silicon channel interface, which are carefully matched with the experiments. Moreover, mobility degradation due to thin body/fin was accounted into simulations. Also *New University of Bologna* (UniBo2) model



**Fig. 10.8** Calibration of TCAD models for (a) drift diffusion transport (with quantum corrections) and (b) p–n junction breakdown with experimental data (figure modified from [21])

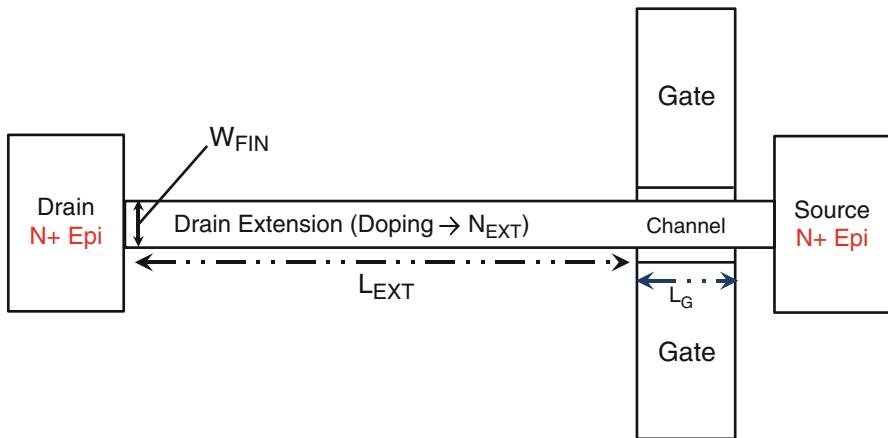
is used for p–n junction breakdown, which was calibrated with experimental data as shown in Fig. 10.8b.

For the investigation of proposed as well as conventional device, we further reduced  $W_{\text{FIN}}$  and EOT down to 10 nm and 1 nm respectively, which were predicted as the target  $W_{\text{FIN}}$  and EOT for sub-20 nm node FinFET technologies [4, 5].

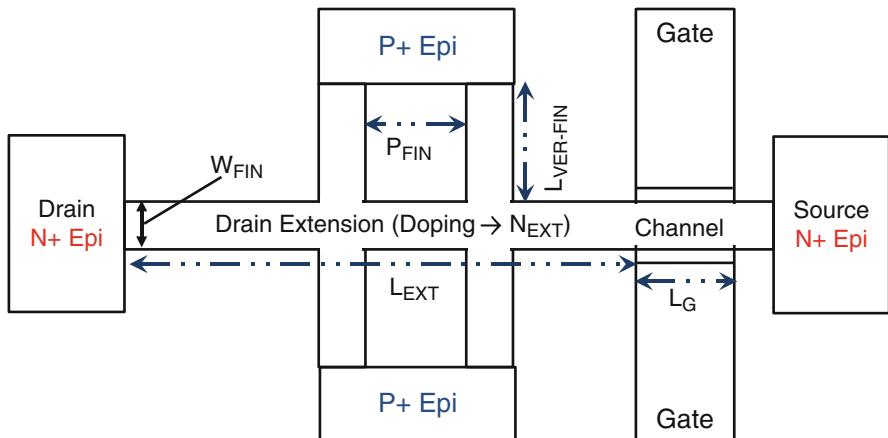
## 10.5 Drain Extended FinFET Devices

Before exploring new drain extended FinFET device concept [21], it's worth discussing the possible HV device option for FinFET technology by using the conventional understanding of RESURF LDMOS device for planar fully depleted SOI technology [22, 23]. Figure 10.9 shows the top view of conventional HV FinFET SOI device with a lightly doped extended fin region. Lightly doped drain extension increases the junction breakdown voltage, however, with the following limitations (1) width of extended drain region is the same as the fin width ( $W_{\text{FIN}}$ ), which is not a design parameter for a *spacer defined fin* process [24], (2) increasing channel length ( $L_G$ ) does not help much for increasing junction breakdown voltage ( $V_{BD}$ ), and (3) reducing extension region doping ( $N_{\text{EXT}}$ ) and increasing the extension region length ( $L_{\text{EXT}}$ ) at the same time leads to a severe increase in  $R_{\text{ON}}$ .

Figure 10.10 shows the top view of recently proposed drain extended FinFET (SOI) device [21], which—in addition to the conventional drain extended FinFET architecture (Fig. 10.9), i.e., transverse drain extended region—adds longitudinal fins, where the number of longitudinal fins is represented as  $N_{\text{LONG-FIN}}$ . The longitudinal fins consist of a P<sup>+</sup> Epi contact (connected to the same potential as source terminal, i.e., ground or 0 V for NMOS) and extends into a N<sup>-</sup> region



**Fig. 10.9** Top view of a (conventional) drain extended FinFET device (figure modified from [21])

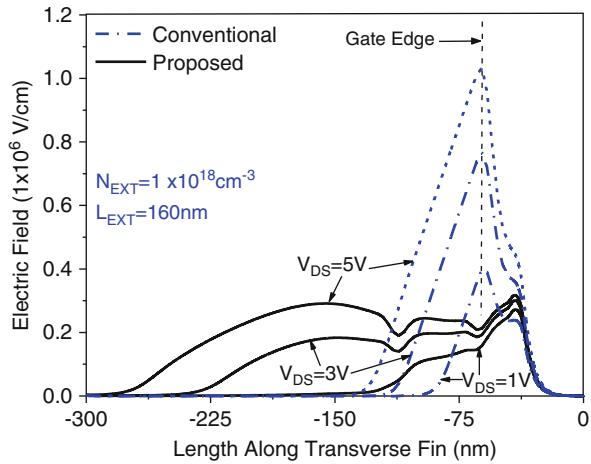


**Fig. 10.10** Top view of the proposed drain extended FinFET device. P+ Epi was connected to the same potential as source, i.e., ground. The additional p-n junctions are formed at P+ Epi and longitudinal Fin interface (figure modified from [21])

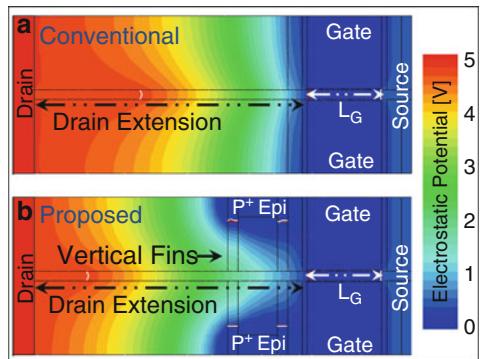
(with length =  $L_{LONG-FIN}$ ) attached to the transverse  $N^-$  region. The following discussion of drain extended FinFET is based on the authors' recently published research article [21].

The transverse and longitudinal  $N^-$  regions (1) have identical  $N^-$  doping concentration ( $N_{EXT}$ ), (2) they form additional p-n junctions, and (3) longitudinal fins are not along the direction of MOS current. Additional p-n junctions attribute to extension of space charge region along the longitudinal fins, which reduces space charge density for an applied potential at drain. This eventually relaxes electric field at the  $N^-$  drain and channel junction. An identical doping concentration in the longitudinal and transverse fin regions simplifies the device fabrication. Due to

**Fig. 10.11** Physical insight behind the  $V_{BD}/R_{ON}$  performance trade-off improvement. Comparison of electric field distribution, along the transverse fin, in conventional and newly proposed drain extended FinFET devices (figure modified from [21])



**Fig. 10.12** Physical insight behind the  $V_{BD}/R_{ON}$  performance trade-off improvement. 2D electrostatic potential contour across (a) conventional and (b) newly proposed drain extended FinFET devices, respectively (figure modified from [21])

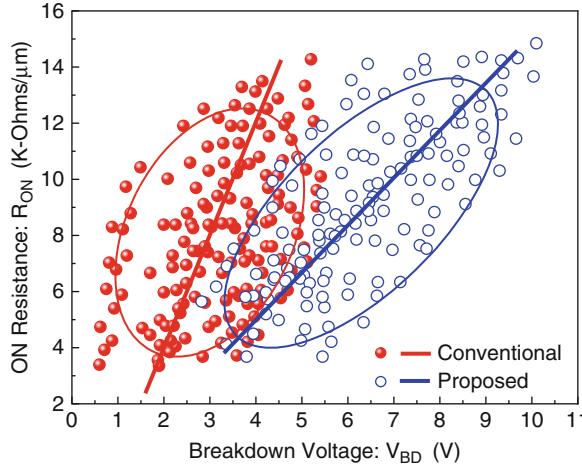


the fact that longitudinal fins are not along the current conduction path, they do not give rise to on-resistance [21].

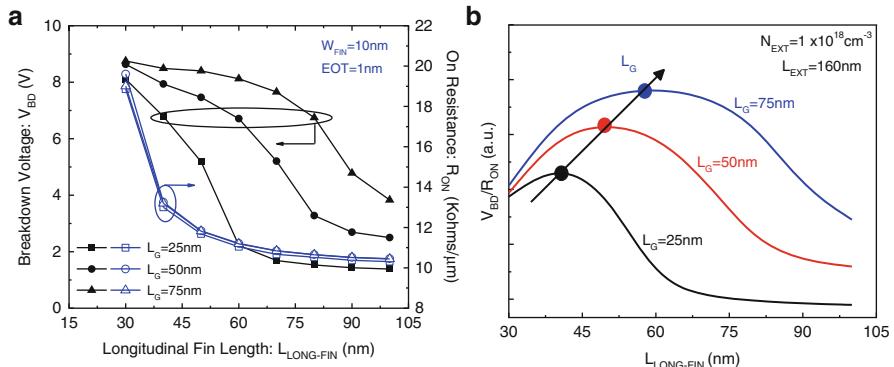
Figure 10.11 shows that the proposed device has a  $3\times$  reduced peak electric field compared to conventional device, when 5 V was applied at drain (gate, source, and body grounded). Additional space charge regions are formed along the p-n junctions of the longitudinal fin regions. Thus, as shown in Fig. 10.12, the proposed device has a significantly relaxed potential distribution near gate edge as compared to a conventional device.

## 10.6 Device Design Guidelines and Discussion

Since (a) longitudinal fins do not influence intrinsic MOS operation—as pointed out in previous section and (b) the proposed device has a  $3\times$  reduced electric field, one can predict that the proposed device will have a similar  $R_{ON}$  with improved  $V_{BD}$  as compared to a conventional device. As shown in Fig. 10.13 design of experiment



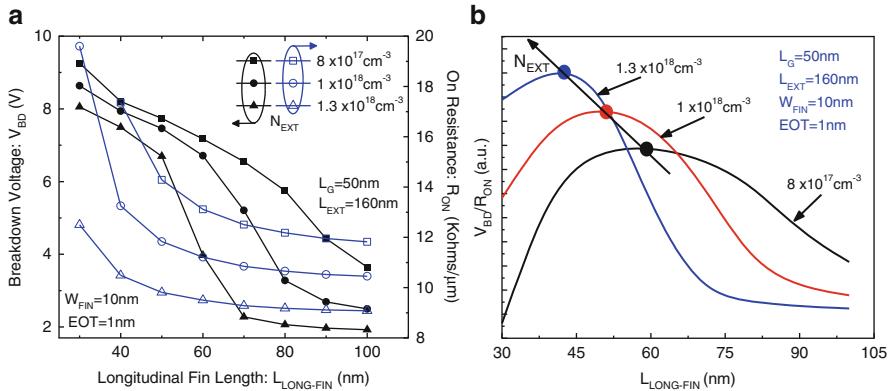
**Fig. 10.13** DOE simulated for conventional and proposed device. Design of experiment was performed by varying  $L_G$  from 25 to 100 nm,  $L_{EXT}$  from 120 to 280 nm, and  $N_{EXT-DOP}$   $7 \times 10^{17}$  to  $2 \times 10^{18} \text{ cm}^{-3}$ , while keeping  $W_{FIN} = 10 \text{ nm}$ , EOT = 1 nm,  $L_{LONG-FIN} = 50 \text{ nm}$ , and  $N_{LONG-FIN} = 2$  (figure modified from [21])



**Fig. 10.14** Optimization of longitudinal Fin length as a function of gate length ( $N_{LONG-FIN} = 2$  was used). Junction breakdown voltage (a) and ON-resistance (b) as a function of longitudinal fin length ( $L_{LONG-FIN}$ ), while varying channel length and fixing extension region doping as well as extension region length (figure modified from [21])

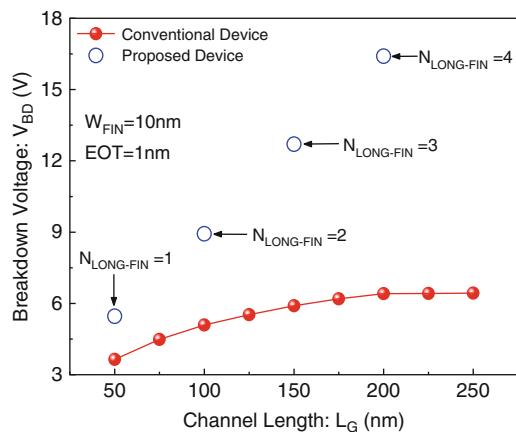
(D.O.E) results obtained through simulations for conventional as well as proposed device validate our prediction furthermore and show that proposed device achieved  $\sim 2 \times$  better  $R_{ON}$  vs.  $V_{BD}$  trade-off.

Figures 10.14 and 10.15 derive the optimization criterion for longitudinal fin length ( $L_{LONG-FIN}$ ) and its relation with  $L_G$  and  $N_{EXT}$ , respectively, while keeping  $L_{EXT}$  fixed. Figures 10.14a and 10.15a show that  $V_{BD}$  and  $R_{ON}$  increase with different rates when  $L_{LONG-FIN}$  was reduced below a certain value depending on



**Fig. 10.15** Optimization of longitudinal Fin length as a function of drain extension doping ( $N_{\text{LONG-FIN}} = 2$  was used). Junction breakdown voltage (a) and ON-resistance (b) as a function of longitudinal fin length ( $L_{\text{LONG-FIN}}$ ), while varying extension region doping and fixing channel length as well as extension region length (figure modified from [21])

**Fig. 10.16** Impact of channel length on the breakdown voltage of conventional as well as proposed device and its relation with number of longitudinal fins in the proposed device (figure modified from [21])



the respective  $L_G$  and  $N_{\text{EXT}}$  value. In addition to this, Figs. 10.14b and 10.15b show the optimum value of  $L_{\text{LONG-FIN}}$ , i.e.,  $V_{\text{BD}}$  should be much higher than the  $R_{\text{ON}}$  cost adder (maximum  $V_{\text{BD}}/R_{\text{ON}}$ ), as a function of  $L_G$  and  $N_{\text{EXT}}$ .

Overall, Figs. 10.14 and 10.15 show how  $V_{\text{BD}}/R_{\text{ON}}$  trade-off can be improved by keeping  $L_{\text{LONG-FIN}}$  approximately equal to 50 nm.

The impact of channel length discussed in Fig. 10.16 is extracted at optimum  $N_{\text{EXT}}$  ( $1 \times 10^{18} \text{ cm}^{-3}$ ),  $L_{\text{EXT}}$  (160 nm), and  $L_{\text{LONG-FIN}}$  (50 nm). Note that the symbol representing the proposed device shows only the optimum  $V_{\text{BD}}/R_{\text{ON}}$  points corresponding to different  $N_{\text{LONG-FIN}}$ . The proposed device can achieve significantly higher breakdown voltages without increasing the gate oxide thickness.

So far the optimization of  $L_{\text{LONG-FIN}}$  and its relation with  $L_G$ ,  $L_{\text{EXT}}$ , and  $N_{\text{EXT}}$  is well understood and the only unrevealed parameter remaining is number of

longitudinal fins ( $N_{\text{LONG-FIN}}$ ). Figure 10.16 also shows that  $V_{\text{BD}}$  increases with channel length, however saturates after a certain value of  $L_G$ , which depends on  $N_{\text{EXT}}$  and  $L_{\text{EXT}}$ . However, the proposed device behaves differently when  $N_{\text{LONG-FIN}}$  and  $L_G$  are increased at the same time. Increasing  $N_{\text{LONG-FIN}}$  and  $L_G$  simultaneously leads to a much higher breakdown voltage for a fixed  $N_{\text{EXT}}$  and  $L_{\text{EXT}}$ . This is due to the fact that increasing  $N_{\text{LONG-FIN}}$  increases the region of the fin which is affected by the reverse p–n junction area and eventually relaxes the space charge density (and electric field) for an applied voltage at the drain. Higher  $N_{\text{LONG-FIN}}$  for smaller  $L_G$  does not improve  $V_{\text{BD}}$  because, for smaller  $L_G$ , channel-to-“drain extension” junction electric field dominates the  $V_{\text{BD}}$ . This can be attributed to the insufficient amount of space charge contribution by longitudinal fins located far away from the gate edge.

## 10.7 Conclusion

Drain extended MOS devices and their beneficial use for high voltage circuits can be carried over from planar CMOS to FINFET technologies. A newly proposed drain extended FinFET device is presented using additional longitudinal fins to improve the breakdown voltage of the device without influencing the MOS behavior. Because of this the proposed drain extended FinFET device shows much lower  $R_{\text{ON}}$  for high  $V_{\text{BD}}$ . Further optimization of  $V_{\text{BD}}$  vs.  $R_{\text{ON}}$  trade-off in terms of device design parameters like  $L_G$  and  $N_{\text{LONG-FIN}}$  for a given  $N_{\text{EXT}}$  is discussed. Such an optimization cannot be achieved for conventional drain extended FinFET device. As the proposed device can achieve a significantly higher breakdown voltage at a lower gate oxide thickness, it opens up new opportunities for high voltage, high speed applications. The proposed device is also expected to show improved gate-oxide reliability and ESD hardness because of the relaxed electric fields and increased silicon volume.

## References

1. Fulde, M., Schmitt-Landsiedel, D., Knoblinger, G.: Analog and RF design issues in high-k & multi-gate CMOS technologies. In: International electron devices meeting, p. 447, Dec 2009
2. Borkar, S.: Design challenges for 22nm CMOS and beyond. In: International electron devices meeting, p. 446, Dec 2009
3. Jan, C.-H., Agostinelli, M., Buehler, M., Chen, Z.-P., Choi, S.-J., Curello, G., Deshpande, H., Gannavaram, S., Hafez, W., Jalan, U., Kang, M., Kolar, P., Komeyli, K., Lake, A., Lazo, N., Lee, S.-H., Leo, T., Lin, J., Lindert, N., Ma, S., McGill, L., Meining, C., Paliwal, A., Park, J., Phoa, K., Post, I., Pradhan, N., Prince, M., Rahman, A., Rizk, J., Rockford, L., Sacks, G., Tashiro, H., Tsai, C., Vandervoorn, P., Xu, J., Yang, L., Yeh, J.-Y., Yip, J., Zhang, K., Bai, P.: A 32nm SoC platform technology with 2nd generation high-k/metal gate transistors optimized for ultra low power, high performance, and high density product applications. In: International electron devices meeting, pp. 647–650, Dec 2009

4. International Technology Roadmap for Semiconductors (ITRS) (2009)
5. Shrivastava, M., et al.: Toward system on chip (SoC) development using FinFET technology: challenges, solutions, process co-development & optimization guidelines. *IEEE Trans. Electron Dev.* **58**(6), 1597–1607 (2011)
6. Various Publishers: Intel increases transistor speed by building upward, May 2011
7. Yeh, C.C., et al.: A low operating power FinFET transistor module featuring scaled gate stack and strain engineering for 32/28nm SoC technology. In: IEEE international electron devices meeting, pp. 772–775, Dec 2010
8. Shrivastava, M., et al.: Part I: mixed-signal performance of various high-voltage drain-extended MOS devices. *IEEE Trans. Electron Dev.* **57**(2), 448–457 (2010)
9. Sigg, H.J., Vendelin, G.D., Cauge, T.P., Kocsis, J.: D-MOS transistor for microwave applications. *IEEE Trans. Electron Dev.* **19**(1), 45–53 (1972)
10. Wickstrom, R.A.: US Patent 4,070,690
11. Jambotkar, C.G., et al.: US Patent 4,145,700
12. Kirk, C.T.: A theory of transistor cutoff frequency ( $f_T$ ) falloff at high current densities. *IEEE Trans. Electron Dev.* **9**(9), 164–174 (1962)
13. Pendharkar, S., Hower, P., Steinhoff, R., Brodsky, J., Devore, J., Grose, B.: Using two-dimensional filament modeling to predict LDMOS and SCR-LDMOS behavior under high current ESD conditions. In: IEEE ISPSD 2001, pp. 161–164, June 2001
14. Wang, L., Wang, J., Gao, C., Hu, J., Li, P., Li, W., Yang, S.H.Y.: Physical description of quasi-saturation and impact-ionization effects in high-voltage drain-extended MOSFETs. *IEEE Trans. Electron Dev.* **56**(3), 492–498 (2009)
15. Shrivastava, M., Gossner, H.: A review on the ESD robustness of drain extended MOS devices. *IEEE Trans. Device Mater. Reliab.* **12**(4), 615–625 (2012)
16. Hisamoto, D., Lee, W.-C., Kedzierski, J., Bokor, J., Hu, C.: FinFET—a self-aligned double-gate MOSFET scalable to 20 nm. *IEEE Trans. Electron Dev.* **47**(12), 2320–2325 (2000)
17. Park, T.-S., Choi, S., Lee, D.-H., Chung, U.-I., Moon, J.T., Yoon, E., Lee, J.-H.: Body-tied triple-gate NMOSFET fabrication using bulk Si wafer. *Solid State Electron.* **49**(3), 377–383 (2005)
18. Manoj, C.R., Nagpal, M., Varghese, D., Ramgopal Rao, V.: Device design and optimization considerations for bulk FinFETs. *IEEE Trans. Electron Dev.* **55**(2), 609–615 (2008)
19. Kanemura, T., Izumida, T., Aoki, N., Kondo, M., Ito, S., Enda, T., Okano, K., Kawasaki, H., Yagishita, A., Kaneko, A., Inaba, S., Nakamura, M., Ishimaru, K., Suguro, K., Eguchi, K., Ishiuchi, H.: Improvement of drive current in bulk-FinFET using full 3D process/device simulations. In: IEEE SISPAD 2006, pp. 131–134, Sept 2006
20. Sachid, A.B., Francis, R., Baghini, M.S., Sharma, D.K., Bach, K.-H., Mahnkopf, R., Rao, V. R.: Sub-20 nm gate length FinFET design: can high-k spacers make a difference? In: IEDM technical digest, pp. 1–4, Dec 2008
21. Shrivastava, M., Gossner, H., Ramgopal Rao, V.: A novel drain extended FinFET device for high voltage high speed applications. *IEEE Electron Device Lett.* **33**(10), 1432–1434 (2012)
22. Sharbati, S., Fathipour, M.: Investigation of vertical scaling on breakdown voltage and presentation of analytical model for electric field distribution in SOI RESURF LDMOSFET. In: International semiconductor device research symposium, pp. 1–2, 9–11 (2009)
23. Merchant, S., et al.: Dependence of breakdown voltage on drift length and buried oxide thickness in SOI RESURF LDMOS transistors. In: International symposium on power semiconductor devices and ICs, pp. 124–128, May 1993
24. Choi, Y.-K., King, T.-J., Hu, C.: A spacer patterning technology for nanoscale CMOS. *IEEE Trans. Electron Dev.* **49**(3), 436–441 (2002)

# Chapter 11

## Modeling FinFETs for CMOS Applications

Lining Zhang, Chenyue Ma, Xinnan Lin, Jin He, and Mansun Chan

**Abstract** As FinFETs are being under intense research explorations today, the corresponding models are essential for understanding their electronic properties and also for future developments of the technology itself. A compact model for FinFETs with double-gate configuration is developed to assist FinFET-based integrated circuit design. The core model includes descriptions of both the current–voltage and terminal charge–voltage characteristics of FinFETs and is suitable for their circuit simulations. A physics-based hot carrier effect model for prediction of FinFETs performance degradation due to the interface state is reported further based on the core model. For future generations of FinFETs, the quantum confinement is becoming more important and is coupled with the widely used strain engineering. Tight binding modeling shows that the effects of a certain amount of uniaxial strain will be less effective for the on-current improvement.

---

L. Zhang (✉) • M. Chan

Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China  
e-mail: [lnzhang@ust.hk](mailto:lnzhang@ust.hk); [mchan@ust.hk](mailto:mchan@ust.hk)

C. Ma

Graduate School of Advanced Sciences of Matter, Hiroshima University, Hiroshima, Japan  
e-mail: [machenye@hiroshima-u.ac.jp](mailto:machenye@hiroshima-u.ac.jp)

X. Lin

School of Electronic and Computer Engineering, Peking University, Shenzhen, China  
e-mail: [xnlin@pkusz.edu.cn](mailto:xnlin@pkusz.edu.cn)

J. He

Peking University SOC Key Laboratory, PKU-HKUST Shenzhen-HongKong Institute, Shenzhen, China  
e-mail: [frankhe@pku.edu.cn](mailto:frankhe@pku.edu.cn)

## 11.1 Introduction

The relentless scaling of CMOS technology pushes itself into the nanometer scale. The traditional structure and operation voltage scaling still with the bulk technology is not enough to fulfill Moore's law in improving the transistor performance in a subsequent generation. Uniaxial strain engineering was utilized at the 90 nm technology, and high-K/metal gate was incorporated at the 45 nm node, contributing to the transistor performance enhancements in terms of on-current increase and short channel effects (SCE) control (gate leakage current reduction) [1]. However, revolution in transistor structures is an inevitable next step to maintain the electrostatic integrity in further scaled devices. Various non-classical MOSFET devices have been presented to replace the planar bulk MOS structures. Among them, the FinFETs (with different configurations like the double-gate or triple-gate) have been and are being explored extensively from the fabrication process to device physics aiming for their applications in the nanoscale CMOS integrated circuits [2–5].

Device compact models are needed for the device technology application. Basically, a compact model is comprised of explicit equations to describe the nonlinear electric characteristics of the device. The model is implemented in the circuit simulator like SPICE to perform circuit design simulations, for example, the DC, transient, and AC simulations. Plenty of research work has been devoted to compact modeling of FinFETs [5–7]. The available models mainly refine them to the FinFETs with a undoped body. For double-gate (DG) FinFETs with heavily doped channels, there are also some models available [8]. While from the circuit simulation perspective, a unified model to describe the performance of the DG FinFETs with various levels of channel doping concentrations is highly desirable for different design considerations. In this chapter, such a unified model will be reported. Its implementation into SPICE and utilization in the circuit simulations will also be discussed.

Practically, MOSFETs have their reliability issues, like the hot carrier effect (HCE) and negative bias temperature instability (NBTI). In the scaled FinFETs, HCE is one important reliability issues [9]. Device performance is affected by the hot carrier-induced interface states close to the drain region. It is necessary to propose a characterization methodology to extract the spatial distribution of the interface states, and develop a corresponding model for aging simulations of FinFET circuits. For the characterization, some conventional methods, such as the capacitance–voltage (CV) or charge-pumping (CP) method, are impractical due to the short channel and relatively small area of a single FinFET [10, 11]. The forward gated-diode method is a simple direct-current (DC) method and is easily adapted to small devices. With the gated-diode configuration, different drain voltages lead to different generation-recombination (G-R) currents, which further reflect the interface states distribution along the FinFET channel [12]. With the knowledge of HCE in a single FinFET, predicting the circuit performance aging is made possible by incorporating the HCE model into the aforementioned compact

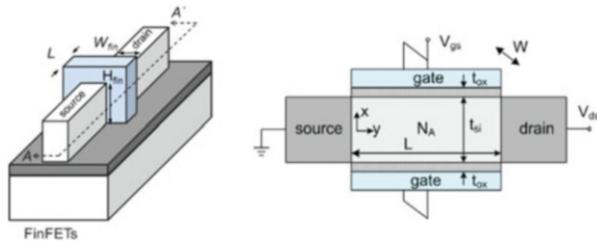
model. The third part of this chapter is devoted to the characterization of interface state distribution in FinFETs with the gated-diode method, also the degradation model development considering the effect of interface state distribution.

At the 22 nm technology node, FinFETs have their fin width smaller than 10 nm. In the coming generations, the fin width is expected to shrink further. With such a small channel, quantum confinement is significant to lead to the material bandstructure change [13, 14] and is coupled with the strain engineering. Whether the uniaxial strain engineering is still beneficial for extremely scaled FinFETs has not been carefully considered. There is some speculation that band splitting caused by structural quantum confinement makes the uniaxial strain effect insignificant [15]. On the other hand, experimental results showed that the [110] uniaxial strain effects are still observable in narrow nanowire n-type MOSFETs [16]. A careful investigation of the uniaxial strain effects in deeply scaled FinFETs is desirable. The fourth part of this chapter is devoted to the modeling of future FinFETs, from the bandstructure effects to the uniaxial strain, aiming to get the answer for the above-mentioned question.

## 11.2 Compact Modeling of DG FinFETs

A DG FinFET is a three-terminal device, and its schematic structure is shown in Fig. 11.1. The compact model associates its terminal charge and current (due to the lateral drift-diffusion transport) with the terminal bias. In the vertical direction of the cross section view, the gate terminal controls the channel properties and tunes the lateral transport. As a result, a compact model first describes the gate voltage ( $V_{gs}$ ) control over the channel potential or charge concentration, usually with the so-called gate control equation. A gradual channel approximation (GCA) [17] is often utilized, which assumes that the change of the lateral electric field is less than that of the vertical electric field. The application of GCA simplifies the gate control equation to one dimensional.

With the solution of channel potentials or charge densities at both source and drain sides, the drift-diffusion transport (in terms of the Pao-Sah double integral) is solved to get the current. Without a drain-source voltage ( $V_{ds}$ ), the total channel charge is equally contributed by the source and drain. With a positive  $V_{ds}$ , the source side of a MOSFET contributes more to the device channel charge. To describe the percentage contribution from source and drain, Ward-Dutton's channel charge partition theme [18] is usually utilized. This charge-orientated theme assumes a linear grading of the channel charge and guarantees the charge conservation in the transistor transient simulations. With the solution of the channel charge at both source and drain sides, the terminal charge of a DG FinFET is also obtained. The terminal current and charge characteristics [19] are described in detail in the following two sections.



**Fig. 11.1** Schematic of a FinFET with the double-gate configuration and its cross-sectional view along the A–A' cut.  $L$  is the channel length,  $W = H_{\text{fin}}$  is the channel width,  $t_{\text{ox}}$  is the gate oxide thickness,  $t_{\text{si}} = W_{\text{fin}}$  is the silicon body thickness, and  $N_A$  is the channel doping concentration

### 11.2.1 Current Model Development

Figure 11.1 shows the schematic diagram of the symmetric DG FinFETs and the modeling coordinates. Taking n-type MOSFETs as an example, the gate control equation is derived from the one-dimensional Poisson–Boltzmann equation including the inversion electron charge and the ionized dopant charge,

$$\frac{d^2\varphi}{dx^2} = \frac{qN_A}{\epsilon_{\text{si}}} \left[ 1 + \exp\left(\frac{\varphi - 2\phi_f - V_{\text{ch}}}{kT/q}\right) \right], \quad (11.1)$$

where  $\varphi$  is the electrostatic potential across the silicon body,  $\phi_f$  is the Fermi potential,  $\epsilon_{\text{si}}$  is the dielectric constant of silicon, and  $V_{\text{ch}}$  is the electron quasi-Fermi potential which varies along the channel from the source voltage to drain voltage.

In the following derivations, all the physical quantities are normalized. Considering the zero vertical electric field condition at the channel center, the normalized electric field profile in the channel is expressed as a function of the normalized center potential  $\phi_c$  ( $x = 0$ ) and the body factor  $\gamma$ :

$$|e| = \gamma \sqrt{\phi - \phi_c + \exp(\phi - 2\phi_f - v_{\text{ch}})[1 - \exp(\phi_c - \phi)]}. \quad (11.2)$$

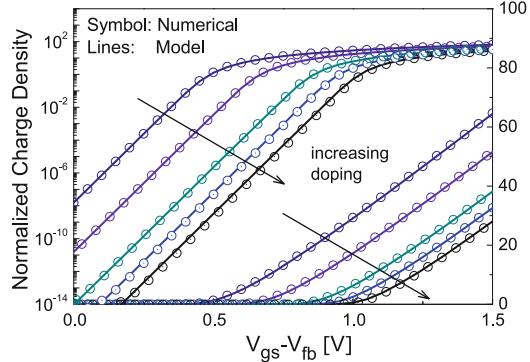
Applying Gauss's law at the channel and gate oxide interface ( $x = t_{\text{si}}/2$ ), the gate control equation is derived with the center potential and surface potential  $\phi_s$  as the unknown variables:

$$v_{\text{gs}} - v_{\text{fb}} - \phi_s = \gamma \sqrt{[\phi_s - \phi_c + \exp(\phi_s - 2\phi_f - v_{\text{ch}})[1 - \exp(\phi_c - \phi_s)]]}. \quad (11.3)$$

At the same time, the normalized channel charge density is also given by Gauss's law with  $q_{\text{inv}}$  as the inversion charge and  $q_{\text{dep}}$  as the depletion charge:

$$v_{\text{gs}} - v_{\text{fb}} - \phi_s = q_{\text{inv}} + q_{\text{dep}}. \quad (11.4)$$

**Fig. 11.2** Charge density comparisons between the modeling results and the numerical simulation results, and good agreements are observed. Five different channel doping levels of  $10^{11}$ ,  $10^{14}$ ,  $10^{17}$ ,  $5 \times 10^{17}$ , and  $10^{18} \text{ cm}^{-3}$  in the DG FinFETs are investigated



The full-depletion approximation gives the difference between the surface potential and center potential:

$$\phi_s - \phi_c = q^2 N_A t_{\text{si}}^2 / (8\epsilon_{\text{si}} kT). \quad (11.5)$$

By combining (11.3), (11.4), and (11.5), we can get the gate control equation which expresses the inversion charge density as an implicit function of the external bias:

$$q_{\text{inv}} + \ln q_{\text{inv}} + \ln(1 + \lambda \cdot q_{\text{inv}}) = v_{\text{gs}} - v_{\text{th}} - v_{\text{ch}}. \quad (11.6)$$

It is easy to see the third term of the LHS mainly plays roles in the transition from weak inversion to strong inversion. The factor  $\lambda = 1/2/q_{\text{dep}}$ . To be compatible with cases of undoped channels, a unified factor is proposed as:

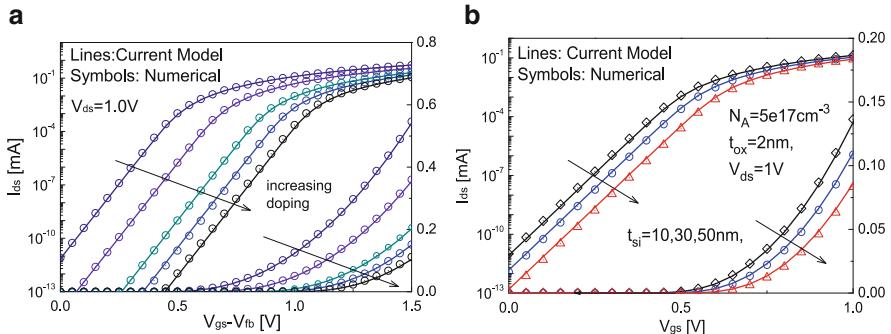
$$\lambda = [1 - \exp(-q^2 N_A t_{\text{si}}^2 / (4\epsilon_{\text{si}} kT))] / (2q_{\text{dep}}) \quad (11.7)$$

And the threshold voltage  $v_{\text{th}}$  is calculated as:

$$v_{\text{th}} = v_{\text{fb}} + 2\phi_f + q_{\text{dep}} - \ln \left\{ \frac{2\epsilon_{\text{si}} t_{\text{ox}}}{t_{\text{si}} \epsilon_{\text{ox}}} \left[ 1 - \exp \left( -\frac{q^2 N_A t_{\text{si}}^2}{4\epsilon_{\text{si}} kT} \right) \right] \right\}. \quad (11.8)$$

For DG structures, there is a unique “volume inversion effect,” which means that the channel center is similarly inverted as the channel surface, contributing partly to the total inversion charge. The volume inversion becomes stronger with lower channel dopings. It can be imagined that with decreasing channel dopings, the “volume inversion” becomes more significant, and the channel inversion charge does not change much even when the flat band voltage  $v_{\text{fb}}$  changes with the dopings. This means that for DG FinFETs with lower doping concentrations, the threshold voltage is almost a constant, which is described in its model (11.8).

Figure 11.2 compares the normalized inversion charge density in the channel of the DG MOS structure as a function of gate voltages under  $V_{\text{ds}} = 0$ . Lines are



**Fig. 11.3** The transfer characteristics of DG FinFETs with (a) different channel doping concentrations and (b) different body thickness but heavy channel doping. The current model agrees well with the numerical simulation results

results calculated from the unified gate control equation (11.6), and symbols represent the exact numerical simulations of the Poisson–Boltzmann equation. Different channel doping concentrations are included. If the charge density is plotted versus gate voltage, the lower doping cases cannot be distinguished, which verifies the above discussions of the threshold voltage. Good agreements of the unified inversion charge density from (11.6) and the numerical simulation are observed from the weak inversion through moderate inversion to strong inversion regions for all the doping levels.

As discussed at the beginning of this section, the drift–diffusion transport should be solved after the gate control equation. The channel charge density at the source and drain side is available from (11.6) by setting the quasi-Fermi level as 0 and  $V_{ds}$ . The current is derived based on the Pao–Sah double integrals [17]:

$$i_{ds} = \int_0^{V_{ds}} q_{inv} dv_{ch} = \int_{q_{inv,s}}^{q_{inv,d}} q_{inv} \cdot \frac{dv_{ch}}{dq_{inv}} dq_{inv}. \quad (11.9)$$

Note that in the above expression the current is normalized, also the integral variable is changed from quasi-Fermi level to the channel inversion charge density. Together with (11.6), the integral in (11.9) is easily performed, leading to the final drain current (normalized to  $2\mu C_{ox}WV_f^2/L$ ):

$$i_{ds} = F(q_{inv})|_{q_{inv,s}}^{q_{inv,d}}, \quad F(q) = 2q + q^2/2 - \ln[1 + \lambda \cdot q]/\lambda. \quad (11.10)$$

To get the charge density from (11.6), we can use the algorithm of W-Lambert function to obtain the initial guess and perform two or three Newton corrections to have the precise enough solutions.

Figure 11.3a compares the current modeling results with those from the numerical simulations of the coupled Poisson and current continuity equations. Different channel doping concentrations are included to show the current model is valid from

the intrinsic doping to heavy doping (full depletion). Similar to Fig. 11.2, the curves corresponding to the three lower doping cannot be distinguished if current is plotted versus gate voltage. Good agreements between the current model (11.10) and the numerical simulations validate the above derivations.

Due to its physics nature, the current model is valid for different geometry parameters like the channel thickness and gate oxide thickness. For DG FinFETs with lightly doped channels, increasing the body thickness means larger conducting area due to the “volume inversion”; hence, the threshold voltage is reduced. For DG FinFETs with heavily doped channels, larger body thickness means more dopants should be depleted before the inversion charge formation; hence, the threshold voltages are increased. Figure 11.3b shows the current characteristics of the doped DG FinFETs with different channel thickness where larger threshold voltages are observed in DG FinFETs with thicker channel. On the other hand, the “volume inversion” is not related to the gate oxide thickness. Thus the subthreshold conduction of undoped or lightly doped DG FinFETs is independent of the gate oxide thickness. However, small gate oxide thickness leads to larger surface charge density in the strong inversion operation region, hence larger on current. The current model (11.10) captures the “volume inversion” in the undoped or lightly doped DG FinFETs and also the depletion charge effect in heavily doped DG FinFETs.

### 11.2.2 Charge Model Development

With the inversion charge density along the channel known from the above (11.6), it is possible to make the partition of the total channel charge to the source and drain side according to some scheme and to develop the charge model.

The total channel inversion charge after normalization ( $2WLC_{\text{ox}}V_t$ ) is given by:

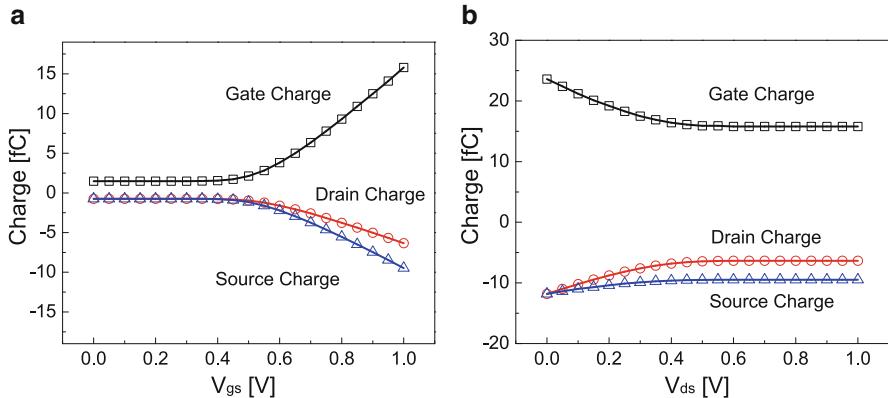
$$q_{\text{tot}} = \int_0^1 q_{\text{inv}} dy = \left( \int_0^{v_{\text{ds}}} q_{\text{inv}}^2 dv_{\text{ch}} \right) / i_{\text{ds}}. \quad (11.11)$$

The integral is performed with (11.6) and (11.11), leading to a closed form solution:

$$q_{\text{tot}} = -\frac{1}{i_{\text{ds}}} M(q_{\text{inv}}) \Big|_{q_{\text{inv},s}}^{q_{\text{inv},d}}, \quad M(q) = \frac{q^3}{3} + q^2 - \frac{q}{\lambda} + \frac{1}{\lambda^2} \ln(1 + \lambda \cdot q). \quad (11.12)$$

According to the Ward and Dutton’s charge partition scheme [18], the normalized drain side charge is calculated as follows:

$$q_d = \int_0^1 y q_{\text{inv}} dy = \int_{q_{\text{inv},s}}^{q_{\text{inv},d}} y q_{\text{inv}} \frac{dx}{dv_{\text{ch}}} \frac{dv_{\text{ch}}}{dq_{\text{inv}}} dq_{\text{inv}}. \quad (11.13)$$



**Fig. 11.4** The gate, source, and drain terminal charges in an example DG FinFET with varying (a) the gate voltage and (b) the drain voltage. Source terminal contributes more to the channel charge, and the percentage is about 60 % in the saturation region

The integral variable is changed from the position to the charge density, like what is done in (11.9). After a lengthy derivation, the final expression for the drain side charge is given [19]:

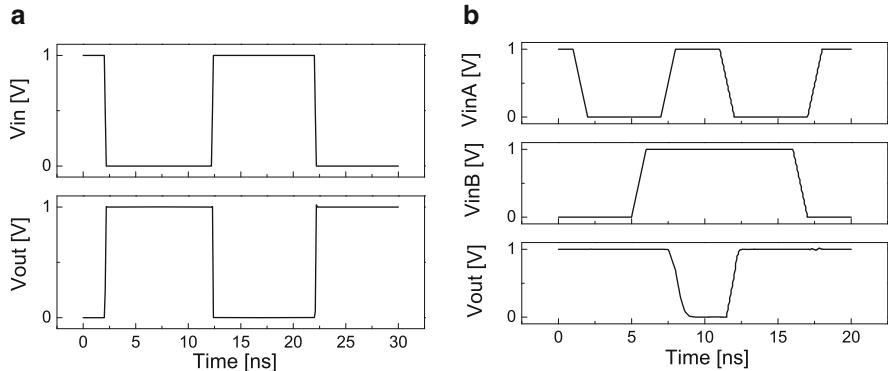
$$q_d = -[M(q_{\text{inv}})F(q_{\text{inv},s}) + N(q_{\text{inv}})] \Big|_{q_{\text{inv},s}}^{q_{\text{inv},d}} / \left[ F(q_{\text{inv}}) \Big|_{q_{\text{inv},s}}^{q_{\text{inv},d}} \right]^2. \quad (11.14)$$

The source side charge can be obtained following a similar routine or by the channel charge conservation.

Figure 11.4 shows the source/drain side charge and the total channel charge as a function of the gate voltage and drain voltage in an example DG FinFET. They will be used in calculating the charging and discharging current in the circuit transient simulations.

### 11.2.3 Circuit Simulations of DG FinFETs

With both the current and charge expressions, the DG FinFETs compact model is implemented into SPICE with Verilog-A [20] and utilized in the circuit simulations to demonstrate its flexibility. In the implementation, the floating point overflow and convergence issues are carefully taken care of. During the Newton iterations in SPICE simulations, the drain node voltage of a DG FinFET may be smaller than the source node voltage. In this case, the source node defined in the circuit netlist is actually the drain side since the drain of a MOSFET is always defined as the region where channel carriers are “drained” away. In the Verilog-A descriptions of the compact model, this issue is handled by swapping the source and drain. The possible floating point overflow is also due to the Newton iteration. The node voltage may be far larger than the practically expected values, for example, the



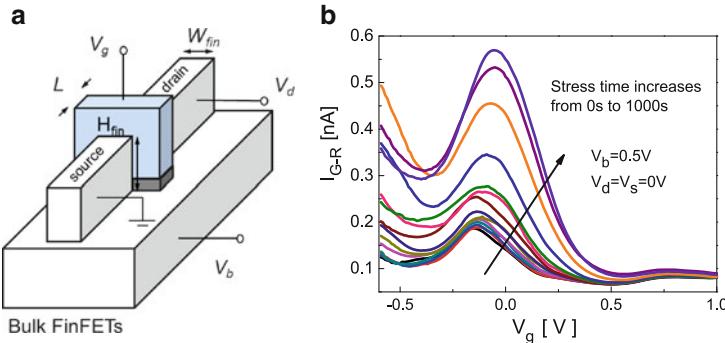
**Fig. 11.5** The transient behaviors of a DG FinFETs-based (a) CMOS inverter and (b) 2-input NAND simulated with the developed model

quasi-Fermi potential  $v_{ch}$  in (11.6) may be temporarily 100 V even for the circuits with  $V_{dd} = 1.0$  V in the iterations. The model implementation should be robust enough to go through these abnormal situations and make the iterations go back to the normal operation conditions.

With the developed compact model, an inverter and a 2-input NAND based on DG FinFETs are simulated, and their transient behaviors are summarized in Fig. 11.5. With more advanced effects incorporated, the developed compact model will be available for circuit simulation applications.

### 11.3 Hot Carrier Effect Modeling for FinFETs

Hot carrier effect (HCE) is due to the large channel electric field-induced interface state or trapped charge. Channel carriers accelerated with a high field acquire large energy to break the surface silicon/oxygen bonds or be injected into the gate oxide, which cause the degradation of the transistor performance, like the on-state current. Here, only the interface state from HCE will be considered. In the MOSFET channel, the high electric field exists near the drain side, which means that the interface state spatial distribution is not uniform. At the same time, it is not necessary that the interface state distribution in the energy band gap is uniform. For simplicity, all the interface states are assumed to center at the middle of the band gap. To model the effects of interface states on the FinFET characteristics, first their distribution along the channel should be characterized and modeled [21], then this spatial distribution information can be incorporated into the above-mentioned compact model for DG FinFETs. For the interface state characterization, the forward gated-diode method assumes that the current is mainly contributed by the carrier recombination process in the depletion region when the external bias is not that large. Figure 11.6a shows the schematic view of the bulk FinFETs and the



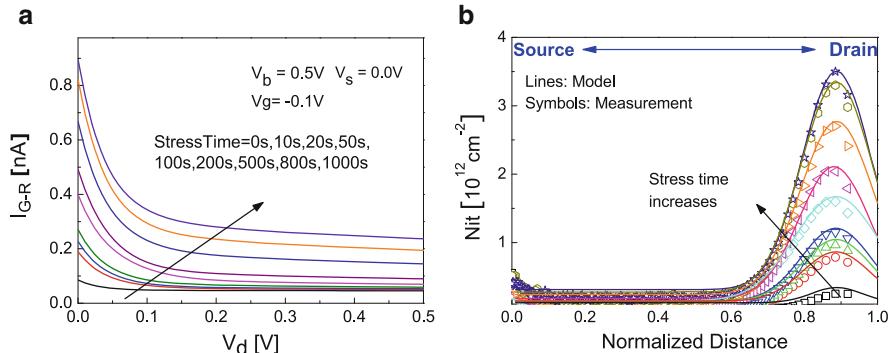
**Fig. 11.6** (a) The configuration of the forward gated-diode characterization method on the bulk FinFETs and (b) the substrate current with source/drain grounded and substrate biased at 0.5 V. The peaks of the generation current appear around  $V_g = -0.1$  V with different stress duration

configuration of the gated diode. With different drain bias, the effective depletion of the drain/substrate junction is changing, so is the recombination current. The distribution of the interface states can finally be extracted from the dependence of diode current on the external bias. For the interface states model development, their effects on the carrier mobility and threshold voltage will be considered in (11.6) and (11.10). After that, the impacts of hot carrier effect on circuits can be evaluated. These two aspects will be covered in the following two sections.

### 11.3.1 Interface States Characterizations

The measured FinFETs have the following parameters: the effective gate length  $L = 15$  nm, the fin height  $H_{\text{fin}} = 25$  nm, and width  $W_{\text{fin}} = 7$  nm. The gate oxide on the fin side wall is 1.2 nm. The fin top surface is covered by thicker oxide; thus, this FinFET is with a DG configuration. For this bulk FinFET, its channel part is connected with the substrate. To perform the acceleration test, we apply the stress bias  $V_{ds} = V_{gs} = 2.5$  V [9], and the stress time is increased from 0 to 1,000 s. After removal of the stress bias, the forward gated-diode connection is configured on the device. The substrate voltage ( $V_b$ ) is fixed at 0.5 V to make the substrate–drain junction operate in a small forward bias condition with  $V_{ds} = 0$ . When the interface state energy level (assumed to be at the intrinsic Fermi level) is bent towards the middle of the minority carrier quasi-Fermi level and the majority carrier quasi-Fermi level, the Shockley–Read–Hall (SRH) recombination rate arrives at its maximum so that one peak of the recombination current ( $I_{G-R}$ ) will appear. The recombination rate at the interface is also modulated by the drain voltage ( $V_{ds}$ ) since it changes the minority carrier quasi-Fermi level directly.

The  $I_{G-R}$  of the DG FinFET with different stress duration is measured by sweeping the gate voltage  $V_{gs}$  from  $-0.6$  to  $1.0$  V, and the  $I_{G-R}$  peak is covered in this bias window. As shown in Fig. 11.6b, the peaks of  $I_{G-R}$  appear around  $V_{gs} = -0.1$  V.



**Fig. 11.7** (a) The modulation effect of drain voltage on the recombination current measured with the forward gated-diode method and (b) the extracted interface state profile along the channel according to the proposed model. Reprinted from [21], with permission from Elsevier

With increasing the stress time, this peak current also increases. The peak  $I_{G-R}$  is modeled as the recombination current in a diode [12]:

$$I_{G-R_{peak}} = \frac{1}{2} W q n_i (c_n c_p)^{1/2} \exp\left(\frac{q V_b}{2kT}\right) \left( \int_{y_1}^{y_2} N_{it}(y) dy \right). \quad (11.15)$$

Here  $N_{it}(y)$  is the interface state density profile,  $y_1$  and  $y_2$  are start and end position of the  $N_{it}$  distribution along channel, and  $c_n = c_p = 6.7 \times 10^{-6} \text{ cm}^{-3} \text{ s}^{-1}$  are the capture cross sections of electrons and holes.

Figure 11.7a shows the modulation effect of  $V_{ds}$  on  $I_{G-R}$  with  $V_{gs}$  fixed at  $-0.1 \text{ V}$ , which corresponds to the maximum interface G-R rate.  $V_{ds}$  is swept from 0 to  $0.5 \text{ V}$ . The applied drain voltage alters the minority quasi-Fermi level profile in the channel near the drain side, and the recombination rate is reduced there since the interface state energy level is not at the middle of the quasi-Fermi levels any more. We assume that in a distance of the depletion region between drain and substrate the recombination rate is reduced to zero. Beyond this depletion region, the quasi-Fermi levels are not changed, so is the recombination rate.

The above (11.15) is rewritten as:

$$N_{it}(y) = \frac{dI_{G-R_{peak}}}{dV_d} \frac{dV_d}{dy} \frac{1}{\frac{1}{2} q n_i (c_n c_p)^{1/2} W \exp\left(\frac{q V_b}{2kT}\right)} \quad (11.16)$$

And the changing rate of the spatial range of effective surface states with increasing the drain voltage is found by:

$$y = \sqrt{\frac{2e_{si}V_d}{qn_i \exp\left(\frac{qV_b}{2kT}\right)}}, \quad \frac{dV_d}{dy} = \sqrt{\frac{2qn_i \exp\left(\frac{qV_b}{2kT}\right)V_d/e_{si}}{}}. \quad (11.17)$$

By combining the above (11.16) and (11.17), the spatial profile of the interface states is extracted. The distribution of  $N_{it}$  along the channel is shown in Fig. 11.7b, which is consistent with the previous results of bulk MOSFETs [22, 23]. In order to meet the need of circuit performance prediction, an empirical expression of  $N_{it}$  is proposed as:

$$N_{it}(y) = N_{it0} + \frac{A}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y/L - \eta)^2}{2\sigma^2}\right). \quad (11.18)$$

where  $N_{it0}$  is the interface states density away from the drain,  $\sigma$  and  $\eta$  are parameters which is related to the width and peak position of the interface state distribution,  $A$  is a parameter to account for the maximum interface states density due to HCE, and  $y/L$  is the normalized distance. The interface state density profile model (11.18) is ready to be incorporated into the DG FinFET compact model to predict the circuit level degradation with even longer stress time.

### 11.3.2 Interface States Modeling

In this section, the empirical interface state distribution model (11.18) is combined with the compact model in Sect. 11.2, leading to a model capable of predicting the circuit performances with HCE considered.

With the interface state distributed along the channel, the inversion charge concentration  $Q_{inv}(y)$  is expressed as:

$$Q_{inv}(y) = C_{ox} [V_{gs} - V_{th} - V_{ch}(y)] - qN_{it}(y), \quad (11.19)$$

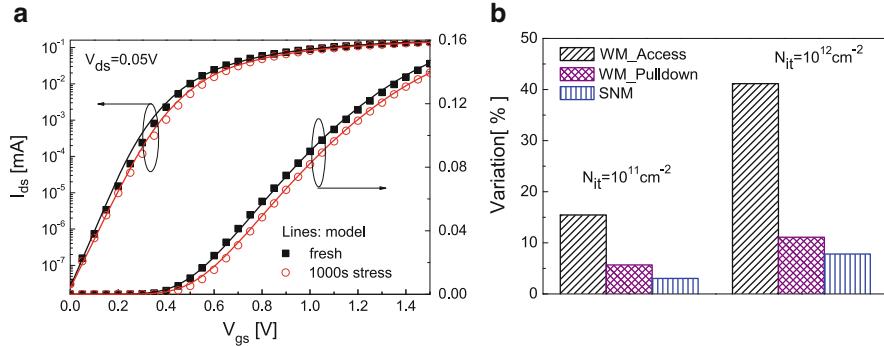
where  $C_{ox}$  is the gate oxide capacitance, and  $V_{th}$  and  $V_{ch}(y)$  are threshold voltage and quasi-Fermi level, respectively. The channel current is obtained from the inversion charge as

$$I_{ds} = W\mu Q_{inv}(y) \frac{dV_{ch}}{dy}, \quad (11.20)$$

where  $\mu = \mu_0/(1 + kN_{it})$  is the effective carrier mobility in the inversion layer.

Combining (11.18) and (11.20) and the compact model in the previous section, the drain current considering the interface states due to HCE is given by:

$$I_{ds} = \frac{L}{L + k \int_0^L N_{it}(y) dy} I_{ds0} - \frac{W\mu_0 q V_{ds} \int_0^L N_{it}(y) dy}{L \left[ L + k \int_0^L N_{it}(y) dy \right]}. \quad (11.21)$$



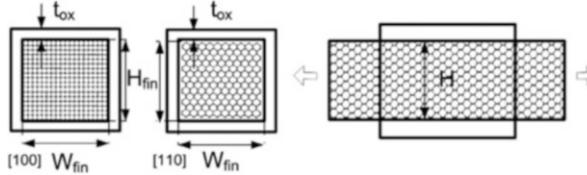
**Fig. 11.8** The impacts of the interface state profile on (a) the single transistor performances and (b) the SRAM performances. The device degradation is accounted for with (11.21). The HCE in access transistors lead to larger write margin degradations. Reprinted from [21], with permission from Elsevier

In (11.21),  $I_{ds0}$  is the drain current without any degradation, and the quasi-Fermi level gradient is approximated as  $V_{ds}/L$  [24]. After parameters extractions, the current model (11.21) reproduces the measured data well, as shown in Fig. 11.8a.

To demonstrate the impact of HCE on circuit performances, a 6T-SRAM is simulated as an example. The static noise margin (SNM) and write margin (WM) are used to monitor the circuit degradation. Figure 11.8b compares the variation of SNM and WM with HCE-induced interface states in the access and pull-down transistors. The WM is more sensitive to the access transistor degradation.

## 11.4 Quantum Effects Modeling for FinFETs

In the previous two sections, quantum confinement effects in the MOS channels are not considered. The developed models are more suitable for the current FinFETs technology; however, they should be revisited when the fin width is scaled to sub-5 nm, where quantum effects are significant. Traditionally, the quantum effects are modeled by approximately solving the effective mass Schrodinger equation in the MOSFET channel with the bulk material parameters. However, these bulk material parameters, like the electron mass, may be not valid any more for the extremely small transistors. The same question rises for the strain engineered FinFETs. In the literatures, the uniaxial strain engineering in nanoscale CMOS is explained by the changes of material parameters, like the conduction band edges and effective electron masses [25]. For example, the uniaxial tensile strain brings downwards the conduction band edge and reduces the threshold voltage of a [110]-oriented transistor [26]. Understanding the interactions between quantum confinement effects and the strain engineering in the extremely small transistors is important to guide the technology developments. One possible way is: firstly using the pseudopotential theory to get the material parameters like the effective



**Fig. 11.9** Schematics of the FinFETs with gate-all-around configurations. Devices with different channel orientations are to be discussed, with their atomistic cross sections shown

masses in different energy valleys and then solving the effective mass equation in the MOS channel to obtain the quantized energy levels and the channel charge. However, the effects of quantum confinement on material parameters are eliminated. Another way is using the empirical atomistic tight-binding method [27] to model the quantum effects and strain effects simultaneously. The coupling between these two effects is reserved.

In this section, the uniaxial strain effects in extremely scaled FinFETs are studied, aiming to answer the question whether strain engineering is still applicable for future MOSFETs [28]. For simplicity, the gate-all-around configured FinFETs are used as examples. Figure 11.9 shows the device cross sections along and perpendicular to the channel direction.  $H_{\text{fin}}$ ,  $W_{\text{fin}}$ , and  $t_{\text{ox}}$  are the fin height, width, and gate oxide thickness. The surfaces of [100] channels are all (110) facets while the top surface and sidewall of [110] channels are (001) and (110) facets.  $H_{\text{fin}}$  and  $W_{\text{fin}}$  are obtained with bulk lattice constant. For channels with  $H_{\text{fin}}/W_{\text{fin}} \sim 1$ , an effective width  $D = (H_{\text{fin}}W_{\text{fin}})^{1/2}$  is defined. The hollow arrows show the uniaxial strain.

#### 11.4.1 Bandstructures of Quantum FinFETs with Strain

The empirical tight-binding (TB) model reproduces the bulk Si bandstructures very well and is also suitable for the nanostructures where the crystal periodicity is reserved only in two dimensions or one. The unit cell that is the smallest repeating cell in the nanostructures is defined as the supercell. The Bloch theory is applied to construct the device channel Hamiltonian with the basis as:

$$\varphi_{ajk}(r) = \frac{1}{\sqrt{N}} \sum_l e^{i\vec{k} \cdot \vec{R}_l} \psi_j^{al}(r - R_l - t_a), \quad (11.22)$$

where  $R_l$  is the position of the  $l$ th supercell,  $t_a$  is the position of the  $a$ th atom in the supercell, and  $j$  is the  $j$ th atomic orbital of the Si atoms. For quantum FinFETs, it is assumed that the quantum confinements in both the fin height and width direction are significant, so the supercell only repeats itself along the channel direction. Electronic states are expressed as the linear combination of the above basis:

$$\varphi_{nk}(r) = \sum_{a,j} A_{naj} \varphi_{ajk}(r), \quad (11.23)$$

where  $a$  runs over all atoms in the supercell and  $j$  runs over the necessary atomic orbital. The coefficients  $A_{naj}$  are found by the variation method, and the device Hamiltonian is derived with the knowledge of the integral energy [29] between different orbitals. Bandstructures are obtained by finding the eigenvalues of the Hamiltonian. To account for the Si material, a ten-band nearest-neighbor  $sp^3d^5s^*$  TB model is used in this work with the empirical parameters provided in [29, 30]. The on-site and two-center integral energy parameters in [30] are used due to their applicability for [110] uniaxial strains. A hydrogen passivation scheme [31] is used to address the surface dangling bonds. With the TB method, quantum effects in the device channel are accounted for.

Upon strains, the atomic positions in (11.22) are changed according to [28]:

$$\vec{r} = (1 + \epsilon^{CCS}) \cdot \vec{r}_0, \quad (11.24)$$

where  $\epsilon^{CCS}$  is the strain tensor in the crystal coordinate system. With applied uniaxial strains along the [100] direction, the tensor is straightforwardly given by:

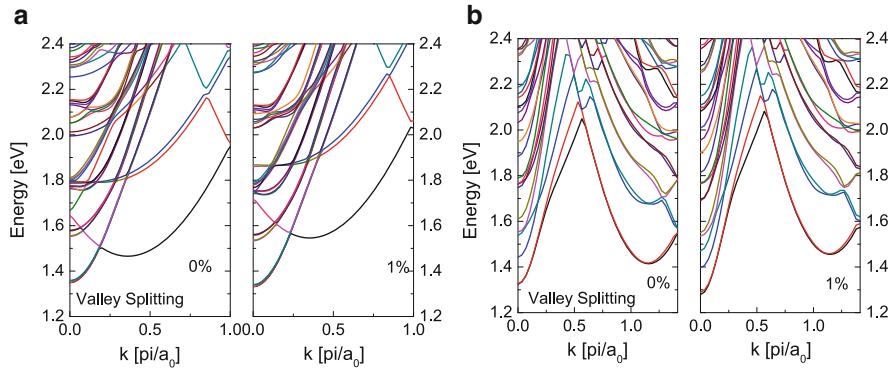
$$\epsilon_{[100]}^{CCS} = \begin{bmatrix} \epsilon_{\parallel} & 0 & 0 \\ 0 & \epsilon_{\perp} & 0 \\ 0 & 0 & \epsilon_{\perp} \end{bmatrix}, \quad (11.25)$$

where  $\epsilon_{\parallel}$  is the strain corresponding to the applied uniaxial stress. The normal strain  $\epsilon_{\perp} = R\epsilon_{\parallel}$  depends on the Poisson ratio  $R$ , which is a material parameter.  $R_{100} = -0.278$  is obtained from the elastic strain theory [32] of bulk silicon. With the uniaxial strains along the [110] direction, the strain tensor is derived to be:

$$\epsilon_{[110]}^{CCS} = \begin{pmatrix} (\epsilon_{\parallel} + \epsilon_{\perp,3})/2 & (\epsilon_{\parallel} - \epsilon_{\perp,3})/2 & 0 \\ (\epsilon_{\parallel} - \epsilon_{\perp,3})/2 & (\epsilon_{\parallel} + \epsilon_{\perp,3})/2 & 0 \\ 0 & 0 & \epsilon_{\perp,1} \end{pmatrix}, \quad (11.26)$$

where  $\epsilon_{\parallel}$  is the strain corresponding to the applied uniaxial stress. The normal strain component  $\epsilon_{\perp,1} = R_{\perp,1}\epsilon_{\parallel}$  is along the [001] direction and the shear component is given by  $\epsilon_{\perp,3} = R_{\perp,3}\epsilon_{\parallel}$ . Based on the strain theory, we can derive the Poisson ratios:  $R_{\perp,1} = -0.354$  and  $R_{\perp,3} = -0.054$ . These two Poisson ratios here are similar to those obtained through the first principle simulation. An internal displacement parameter is used to calculate the displacement of the center atom in the strained diamond crystal to get the final atomic positions.

With changes of atomic positions in the supercell, both bond length and angles are changed, and these changes modify the off-diagonal, same-atom matrix elements [30] and the two-center integral parameters in the Hamiltonian. These parameters are incorporated into the Hamiltonian and then the bandstructures of quantum FinFETs with applied strains are obtained.

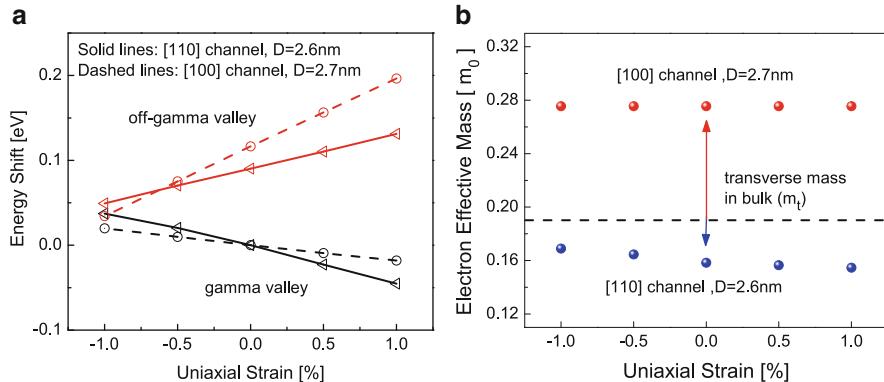


**Fig. 11.10** The bandstructures of (a) a [100]-oriented FinFET channel and (b) a [110]-oriented FinFET channel, without and with 1 % uniaxial tensile strain. Valley splitting is observed in these nanoscale silicon channels. The uniaxial tensile strain increases the splitting between energy valleys differently in the simulated two FinFET channels

Figure 11.10 shows the energy bandstructures of a [100]-oriented FinFET channel ( $H_{\text{fin}} = 2.68 \text{ nm}$ ,  $W_{\text{fin}} = 2.87 \text{ nm}$ ) and a [110]-oriented FinFET channel ( $H_{\text{fin}} = 2.58 \text{ nm}$ ,  $W_{\text{fin}} = 2.69 \text{ nm}$ ) without strain and with the 1 % uniaxial tensile strain. The [100]-oriented channel has a fourfold degenerated  $\Gamma$  (or “gamma” in Fig. 11.11a) valley and a twofold degenerated off- $\Gamma$  valley, and the  $\Gamma$  valleys are lower than the off- $\Gamma$  valleys. These results quantitatively agree with the effective mass physical picture. However, the valley splitting behavior at the  $\Gamma$  valley is beyond the capability of an effective mass method. The [110]-oriented channel has a twofold degenerated  $\Gamma$  valley, and a pair of twofold degenerated off- $\Gamma$  valleys. Electrons in the  $\Gamma$  valley also have a lower energy state. The  $\Gamma$  valley splitting is also confirmed with the more fundamental density functional theory (DFT) simulations like the VASP simulations [33]. The validity of the TB method in modeling the quantum confinements in the FinFET channels is verified.

The tensile strain-induced conduction band lowering is observed in both [100] and [110]-oriented channels in Fig. 11.10. The  $\Gamma$  valley and off- $\Gamma$  valley ground states energies are extracted and shown in Fig. 11.11a where the conduction band edge at the  $\Gamma$  valley is taken as the reference. This conduction band lowering qualitatively agrees with the experimental results in [8, 9] and explains the strain-induced negative threshold voltage shift. The conduction band edge lowering of a [110]-oriented channel ( $D \sim 0.85 \text{ nm}$ ) with the 1 % tensile strain is 70 meV, very close to the VASP simulation result [33]. The quantum confinement and uniaxial tensile strain play similar roles in inducing the band splitting, however, opposite roles in changing the channel material band gap.

With the parabolic dispersion approximation, the electron effective masses in  $\Gamma$  valleys of both [100] and [110]-oriented channels are extracted and plotted in

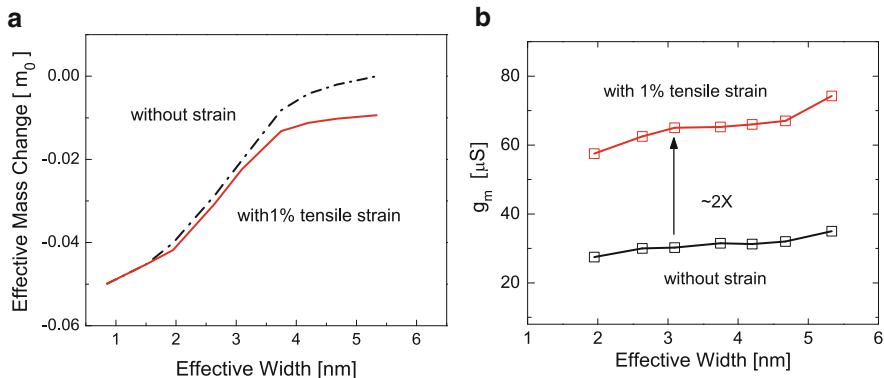


**Fig. 11.11** The uniaxial strain effects on (a) the conduction band edge shifts and (b) the electron effective masses in the [100] and [110]-oriented channels. The tensile strain increases the band splitting in both channels and reduces the electron masses in [110] channels

Fig. 11.11b. Without any strain, the electron mass in [100] channels ( $\sim 0.28m_0$ ) are larger than the mass ( $\sim 0.19m_0$ ) in bulk Si. Uniaxial strain does not change it. Without any strain, the effective mass in [110] channels ( $\sim 0.16m_0$ ) are smaller than the bulk one. With a tensile strain, the mass is further reduced. Quantum confinement and uniaxial tensile strain play similar roles in altering the electron effective mass of [110] channels. For both [100] and [110] channels, the tensile strain increases the conduction band valley splitting and lifts the valley with higher transport mass. As a result, the uniaxial tensile strain reduces the average electron transport mass for both channels.

Since the quantum confinement and tensile strain effects are similar for [110] channels, strain-induced effective mass changes in extremely scaled channels are expected to diminish due to strong confinements [33]. Figure 11.12a shows the electron effective masses changes in scaled FinFET channels without and with strain compared to the bulk one. When the effective channel width  $D$  is larger than 5 nm, the effective mass is about  $0.189m_0$ , close to the bulk value. A 1 % uniaxial tensile strain changes it by about  $0.01m_0$ . With smaller  $D$ , the effective mass decreases and the strain effect is also less significant. As the channel size decreases to 1.5 nm, 1 % tensile strain does not cause changes in the electron mass any more.

As seen, for the band splitting in both [100] and [110] channels and electron masses in [110] channels, the quantum confinement and uniaxial strain have similar effects. Therefore, the uniaxial tensile strain effect in improving the on current of further scaled FinFETs is expected to be less effective, which will be proved in the next section.



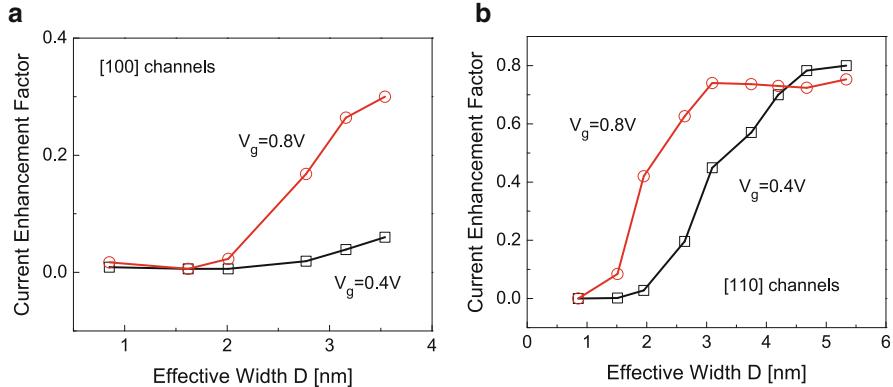
**Fig. 11.12** Impacts of the uniaxial tensile strain on (a) the electron effective masses and (b) the maximum transconductances  $g_m$  of scaled FinFETs. With smaller channels, effects of the same amount of strain become less significant in reducing the effective mass, but the  $g_m$  increase is valid for FinFETs with their width down to 2 nm

#### 11.4.2 Transports of Quantum FinFETs with Strain

With the knowledge of bandstructures of quantum FinFETs, the strain effects on their ballistic transport properties are studied with a well-developed semiclassical top-of-the-barrier ballistic model [34]. The simulated channel dispersions without and with strain are first used to get the FinFET channel charge for the given source Fermi energy and drain bias. Then the obtained mobile charge density at the barrier peak  $Q_{\text{top}}$  is fed back to the electrostatic coupling equation to get the self-consistent potential  $U_{\text{scf}}$  which further affects the  $Q_{\text{top}}$ . The iterations are repeated to achieve self-consistency between the  $Q_{\text{top}}$  and  $U_{\text{scf}}$ . Ballistic current through the FinFETs is obtained by substituting the self-consistent potential  $U_{\text{scf}}$  into the ballistic current equation [34].

Experimental results in [35, 36] demonstrated that [110] channels with around 1 % uniaxial tensile strain have a  $2\times$  peak transconductance ( $g_m$ ) compared to the channels without any strain. Figure 11.12b shows the simulation results of the maximum  $g_m$  in [110]-oriented channels of different sizes without and with strain. The ballistic  $g_m$  enhancement due to 1 % uniaxial tensile strain is also about  $2\times$  despite the smaller channel sizes. This peak  $g_m$  enhancement is still observable in FinFETs with the channel width down to 2 nm.

Figure 11.13 shows the current enhancement factors in the [100] and [110]-oriented FinFETs along their dimensional scaling. The source Fermi energies are tuned to have the same off current (0.1 nA) in all the FinFETs. Two gate voltages,  $V_g = 0.4$  V where the channels just enter the degenerate condition, and  $V_g = 0.8$  V where they operate in the degenerate mode, are considered. As shown, the strain-induced performance boost is different for different operation voltages.



**Fig. 11.13** The 1 % uniaxial tensile strain-induced on-current enhancements in (a) the [100] FinFETs and (b) the [110] FinFETs. In scaled FinFETs, the same amount of strain becomes less effective in improving the on currents due to the significant quantum confinements

For [100] channels with  $V_g = 0.4$  V, their current enhancement factors are only about several percent and decrease with reducing the channel size. This is because quantum confinement has already led to large valley splitting as discussed in the previous section. For channels with  $V_g = 0.8$  V, more inversion electrons that reside in higher energy states are still affected by the uniaxial tensile strain-induced band splitting. However, the enhancement becomes negligible for channels smaller than 2 nm [28]. For [110] channels, the uniaxial strain is more prominent. When the channel size is about 5 nm, the enhancement factor is roughly 80 % with both gate voltages, close to the experimental observations [35]. Under  $V_g = 0.8$  V, this enhancement factor extends to 3 nm wide FinFETs. Under  $V_g = 0.4$  V, the uniaxial tensile strain effects begin to decrease in FinFETs with a larger size.

## 11.5 Conclusions

In this chapter, modeling of FinFETs for their electronic applications is discussed, from the compact core model development to the hot carrier effects modeling and finally the quantum coupled strain effects modeling. The compact core model, including the descriptions of current and charge characteristics, for DG FinFETs with undoped or fully depleted channels, is reported and implemented into SPICE for circuit simulations. The essential device physics are captured in the core model. Further extensions to include more advanced effects will make the model suitable for FinFETs integrated circuit designs. To account for the aging of the FinFET circuits, a forward gated-diode methodology is applied to extract the hot carrier induced interface states, and the interface state profile is incorporated into the above compact model. The interface states near the drain side follow a Gaussian type

distribution and affect the threshold voltage and current of the DG FinFETs. In the further scaled FinFETs, the quantum confinement effects and strain engineering are coupled and are accounted for with the tight binding model. The same amount of strain becomes less effective in improving the on current of FinFETs with lower operation voltages. The modeling efforts here provide understandings of DG FinFETs operations, as well as guidelines for their future developments.

## References

1. International Technology Roadmap for Semiconductors. <http://www.itrs.net> (2011)
2. Auth, C., Allen, C., Blattner, A., Bergstrom, D., Brazier, M., Bost, M., Buehler, M., Chikarmane, V., Ghani, T., Glassman, T., Grover, R., Han, W., Hanken, D., Hattendorf, M., Hentges, P., Heussner, R., Hicks, J., Ingerly, D., Jain, P., Jaloviar, S., James, R., Jones, D., Jopling, J., Joshi, S., Kenyon, C., Liu, H., McFadden, R., McIntyre, B., Neirynck, J., Parker, C., Pipes, L., Post, I., Pradhan, S., Prince, M., Ramey, S., Reynolds, T., Roesler, J., Sandford, J., Seiple, J., Smith, P., Thomas, C., Towner, D., Troeger, T., Weber, C., Yashar, P., Zawadzki, K., Mistry, K.: A 22nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors. In: IEEE symposium on VLSI technology, Honolulu, HI, USA, 12–14 June 2012
3. Lin, C.H., Haensch, W., Oldiges, P., Wang, H., Williams, R., Chang, J., Guillorn, M., Bryant, A., Yamashita, T., Standaert, T., Bu, H., Leobandung, E., Khare, M.: Modeling of width-quantization-induced variations in logic FinFETs for 22nm and beyond. In: IEEE symposium on VLSI technology, Hyoto, Japan, 13–17 June 2011
4. Lu, D.D., Dunga, M.V., Lin, C.H., Niknejad, A.M., Hu, C.: A multi-gate MOSFET compact model featuring independent-gate operation. In: IEEE international electron device meeting, Washington, DC, USA, 10–12 Dec 2007
5. Iniguez, B., Fjeldly Tor, A., Lazaro, A., Danneville, F., Deen, M.J.: Compact-modeling solutions for nanoscale double-gate and gate-all-around MOSFETs. *IEEE Trans. Electron Dev.* **53**, 2128–2142 (2006)
6. Taur, Y., Liang, X., Wang, W., Lu, H.: A continuous, analytical drain-current model for double-gate MOSFETs. *IEEE Electron Device Lett.* **25**, 107–109 (2004)
7. Ortiz-Conde, A., Garcia-Sanchez, F.J., Muci, J.: Rigorous analytic solution for the drain current of undoped symmetric dual-gate MOSFETs. *Solid State Electron.* **49**, 640–647 (2005)
8. Sallese, J.M., Krümmenacher, F., Prégaldiny, F., Lallement, C., Roy, A., Enz, C.: A design oriented charge-based current model for symmetric DG MOSFET and its correlation with the EKV formalism. *Solid State Electron.* **49**, 485–489 (2005)
9. Kim, S., Lee, J.: Hot carrier-induced degradation in bulk FinFETs. *IEEE Electron Device Lett.* **26**, 566–568 (2005)
10. Yu, B., Chang, L., Ahmed, S., Wang, H., Bell, S., Yang, C., Tabery, C., Ho, C., Xiang, Q., King, T.-J., Bokor, J., Hu, C., Lin, M., Kysar, D.: FinFET scaling to 10nm channel length. In: IEEE international electron device meeting, San Francisco, CA, USA, 8–11 Dec 2002
11. Li, M.-F., Huang, D., Shen, C., Yang, T., Liu, W.J., Liu, Z.: Understand NBTI mechanism by developing novel measurement techniques. *IEEE Trans. Device Mater. Reliab.* **8**, 62–71 (2008)
12. He, J., Zhang, X., Huang, R., Wang, Y.: Application of forward gated-diode R–G current method in extracting F–N stress-induced interface traps in SOI NMOSFETs. *Microelectron. Reliab.* **42**, 145–148 (2002)

13. Gnani, E., Gnudi, A., Reggiani, S., Rudan, M., Baccarani, G.: Band structure effects on the current–voltage characteristics of SNW-FETs. In: IEEE international electron device meeting, Washington, DC, USA, 10–12 Dec 2007
14. Neophytou, N., Paul, A., Lundstrom, M.S., Klimeck, G.: Bandstructure effects in silicon nanowire electron transport. *IEEE Trans. Electron Dev.* **55**, 1286–1297 (2008)
15. Baykan, M.O., Thompson, S.E., Nishida, T.: Strain effect on three dimensional, two dimensional, and one dimensional silicon logic devices: predicting the future of strained silicon. *J. Appl. Phys.* **108**, 093716 (2010)
16. Jeong, Y., Miyaji, K., Saraya, T., Hiramoto, T.: Silicon nanowire n-type metal-oxide-semiconductor field-effect transistors and single-electron transistors at room temperature under uniaxial tensile strain. *J. Appl. Phys.* **105**, 084514 (2009)
17. Pao, H.C., Sah, C.T.: Effects of diffusion current on the characteristics of metal-oxide (insulator)-semiconductor transistors. *Solid State Electron.* **9**, 927–937 (1966)
18. Ward, D., Dutton, R.: A charge-oriented model for MOS transistor capacitances. *IEEE J. Solid State Circ.* **13**, 703–708 (1978)
19. Zhang, L., Zhang, J., Song, Y., Lin, X., He, J., Chan, M.: Charge-based model for symmetric double-gate MOSFETs with inclusion of channel doping effect. *Microelectron. Reliab.* **50**, 1070–1062 (2010)
20. Accellera Organization: Verilog-AMS language reference manual, ver. 2.3.1 (2009)
21. Ma, C., Zhang, L., Zhang, C., Zhang, X., He, J., Zhang, X.: A physical based model to predict performance degradation of FinFET accounting for interface state distribution effect due to hot carrier injection. *Microelectron. Reliab.* **51**, 337–341 (2011)
22. Ancona, M.G., Saks, N.S., McCarthy, D.: Lateral distribution of hot-carrier-induced interface traps in MOSFET's. *IEEE Trans. Electron Dev.* **35**, 2221–2228 (1988)
23. Park, T., Cho, H.J., Choe, J.D., Cho, I.H., Park, D., Yoon, E., Lee, J.-H.: Characteristics of body-tied triple-gate pMOSFETs. *IEEE Electron Device Lett.* **25**, 798–800 (2004)
24. Kang, H., Han, J.-W., Choi, Y.-K.: Analytical threshold voltage model for double-gate MOSFETs with localized charges. *IEEE Electron Device Lett.* **29**, 927–930 (2008)
25. Uchida, K., Krishnamohan, T., Saraswat, K.C., Nishi, Y.: Physical mechanisms of electron mobility enhancement in uniaxial stressed MOSFETs and impact of uniaxial stress engineering in ballistic regime. In: IEEE international electron device meeting, Washington, DC, USA, 5–7 Dec 2005
26. Thompson, S.E., Sun, G., Wu, K., Lim, J., Nishida, T.: Key differences for process-induced uniaxial vs. substrate-induced biaxial stressed Si and Ge channel MOSFETs. In: IEEE international electron device meeting, San Francisco, USA, 13–15 Dec 2004
27. Slater, J.C., Koster, G.F.: Simplified LCAO method for the periodic potential problem. *Phys. Rev.* **94**, 1498–1524 (1954)
28. Zhang, L., Lou, H., He, J., Chan, M.: Uniaxial strain effects on electron ballistic transport in gate-all-around silicon nanowire MOSFETs. *IEEE Trans. Electron Dev.* **58**, 3829–3836 (2011)
29. Boykin, T.B., Klimeck, G., Oyafuso, F.: Valence band effective-mass expressions in the  $sp^3d^5s^*$  empirical tight-binding model applied to a Si and Ge parametrization. *Phys. Rev. B* **69**, 115201 (2004)
30. Boykin, T.B., Luisier, M., Salmani-Jelodar, M., Klimeck, G.: Strain-induced, off-diagonal, same-atom parameters in empirical tight-binding theory suitable for [110] uniaxial strain applied to a silicon parametrization. *Phys. Rev. B* **81**, 125202 (2010)
31. Lee, S., Oyafuso, F., Allmen, P., Klimeck, G.: Boundary conditions for the electronic structure of finite-extent embedded semiconductor nanostructures. *Phys. Rev. B* **69**, 045316 (2004)
32. Nye, J.F.: Physical properties of crystals: their representation by tensors and matrices. Oxford University Press, Oxford (1985)
33. Maegawa, T., Yamauchi, T., Hara, T., Tsuchiya, H., Ogawa, M.: Strain effects on electronic bandstructure in nanoscaled silicon: from bulk to nanowire. *IEEE Trans. Electron Dev.* **56**, 553–559 (2009)

34. Lundstrom, M.S., Guo, J.: Nanoscale transistors: device physics, modeling and simulation. Springer, New York (2006)
35. Hashemi, P., Gomez, L., Hoyt, J.L.: Gate-all-around n-MOSFETs with uniaxial tensile strain-induced performance enhancement scalable to sub-10nm nanowire diameter. IEEE Electron Device Lett. **30**, 401–403 (2009)
36. Feste, S.F., Knoch, J., Habicht, S., Buca, D., Zhao, Q.T., Mantl, S.: Silicon nanowire FETs with uniaxial tensile strain. Solid State Electron. **53**, 1257–1262 (2009)

# Chapter 12

## Enhanced Quantum Effects in Room-Temperature Coulomb Blockade Devices Based on Ultrascaled finFET Structure

Jung B. Choi

**Abstract** The availability of an ultrasmall dot-based single-electron device could provide new operating regimes where Coulomb blockade and quantum confinement are two distinct effects, competing to control electron transport. Most recently, we have made a successful implementation of CMOS-compatible room-temperature single-electron transistor by ultrascaling a finFET structure down to an ultimate limiting form, resulting in the reliable formation of a sub-5-nm silicon Coulomb island. The charge stability of the device features, for the first time, three and a half clear multiple Coulomb diamonds at 300 K, showing high PVCRs. The device dot size is sufficiently small that Coulomb blockade and other quantum effects persist up to room temperature. The charge stability at 300 K with additional fine structures of low-temperature Coulomb peaks are successfully modeled by including the interplay between Coulomb interaction, valley splitting, and strong quantum confinement that become enhanced in ultrasmall scale. This supports that for a sub-5 nm device even small number of electron occupation ensures that quantum many-body interactions strongly influence the room-temperature electron transport characteristics. Under this condition, quantum effect can be used as an additional state variable to provide another multi-switching functionality.

### 12.1 Introduction

Single-electron transistor (SET), considered as an ultimate limit of the field-effect transistor (FET) where the conduction channel is replaced by a nanoscale quantum dot, has been intensively studied because of its unique multi-functionality with ultralow power and scalability down to sub-nm regime [1]. The basic philosophy of the SET-based electronics is a manipulation of the individual electron charge and how

---

J.B. Choi (✉)

Department of Physics & Research Institute for Nano Science & Technology,

Chungbuk National University, Cheongju 361-763, Korea

e-mail: [jungchoi@chungbuk.ac.kr](mailto:jungchoi@chungbuk.ac.kr)

to apply it to the digital electronics. The single-electron-based multiple-valued (MV) memory, or MV logic, reflects such capability of controlling discrete charges of an individual electron. The MV memory can be implemented by utilizing the threshold voltage quantization in nanoscale floating dot-based memory [2–4], while multiple switching on/off of the SET enables us to realize the MV logic scheme [5–12], which has higher functionality with lower hardware complexity. To be specific, incorporating SETs with FETs, implementation of a two-input NAND/NOR [10], exclusive-OR (XOR) [6, 11], and half-adder [12] have been reported. Their voltage transfer characteristics display typical flexible MV logic gate function depending on the two-input SET voltages. However, the reliability of room-temperature (RT) characteristics and requirement of CMOS-compatible processes have been the main bottlenecks for implementing further practical device applications. Significant earlier works [13–20] have been aimed at implementing a RT-operating SET, using various schemes, device structures, and fabrication processes, but reliable processes for the controlled fabrication of ultrasmall size Coulomb islands of less than 5 nm have not been firmly established yet. Most of the gate-dependent drain currents  $I-V_g$  have exhibited only a single negative differential resistance(NDR)-like peak [18], or weak ripples with very poor peak-to-valley current ratios (PVCRs)  $\ll 1$  [17, 19, 20], limiting practical device applications at room temperature.

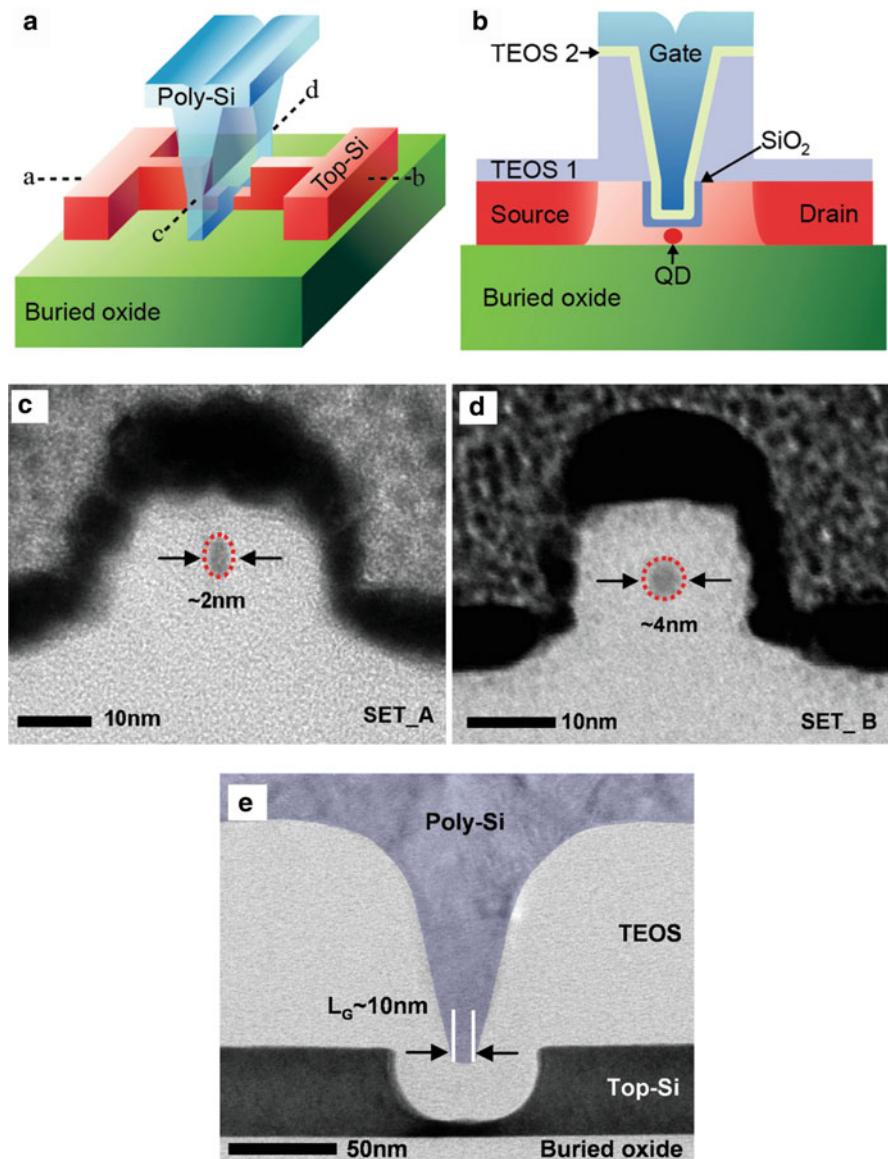
Most recently, we have made a successful implementation of CMOS-compatible room-temperature SET by ultrascaling a state-of-the-art fin field-effect transistor (finFET) structure down to an ultimate limiting form, resulting in the reliable formation of a sub-5-nm silicon Coulomb island [21, 22]. The charge stability of the device features, for the first time, three and a half clear multiple Coulomb diamonds at 300 K, showing high PVCRs. The performance and CMOS-compatible manufacturability of the device certainly moves SETs closer to semiconductor industrial feasibility. Moreover, the availability of such ultrasmall dot-based Coulomb blockade devices could provide new operating regimes for the single-electron tunneling where Coulomb blockade and quantum confinement are two distinct effects, competing to control electron transport through a device. As the island size is reduced below 5 nm, the energy scale for quantized level spacing might become dominant over the Coulomb charging as well as thermal fluctuation. This is right the case that occurs in our device whose dot size is sufficiently small that Coulomb blockade, and other quantum effects, can persist up to even room temperature [22]. Even small number of electron occupations on the Coulomb island ensures that quantum many-body interactions strongly influence the room-temperature electron transport characteristics. Under this condition, quantum effect may be used as an additional state variable to control electron transport and could provide an additional resource scheme to be exploited in multi-valued nanoelectronics [23], in contrast to conventional CMOS, where quantum effects usually contribute only problems such as noise and unwanted threshold voltage shifts.

In this chapter, concentrating on the CMOS-compatible ultrasmall SETs based on the most recent pioneering works of [21, 22], we will discuss their device structure, fabrication process, and  $I-V_g$  characteristics. The observed room-temperature charge stabilities, exhibiting multiple Coulomb diamonds, are

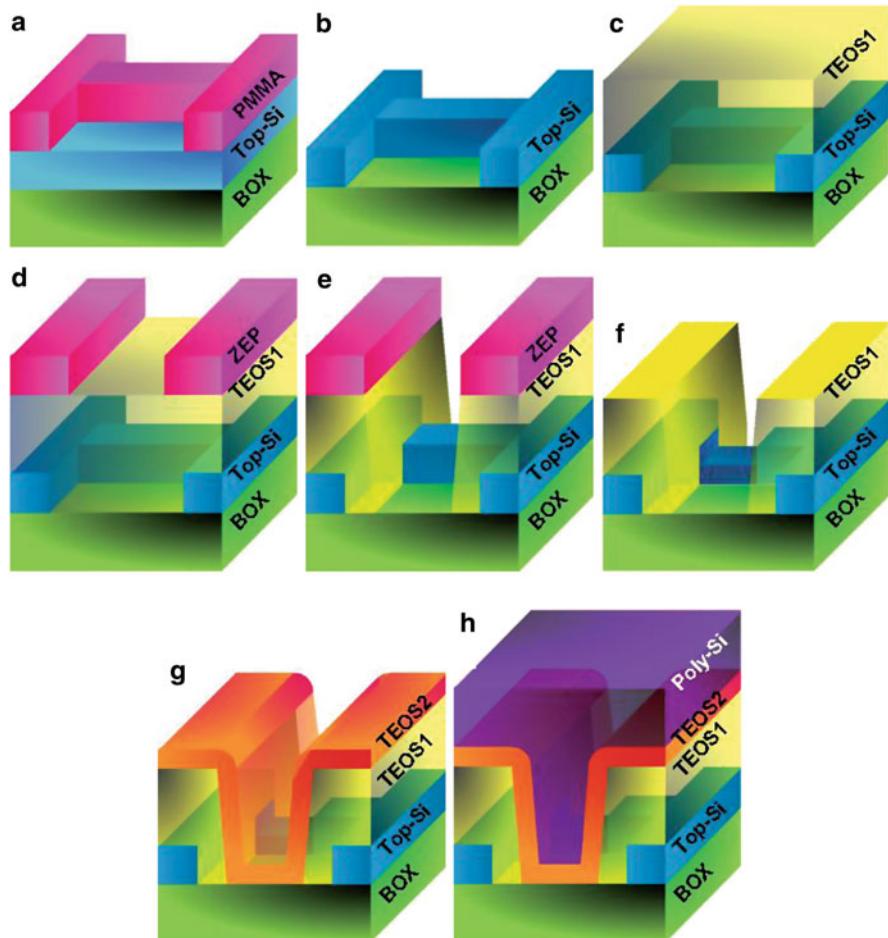
presented and analyzed by strong interplay of the Coulomb interaction and additional quantum effects associated with very low electron occupation on the ultrasmall silicon dot of sub-5-nm size.

## 12.2 Room-Temperature Coulomb Blockade Devices Based on Ultrascaled finFET: Device Structure and Fabrication Process

A state-of-the-art finFET structure has been modified and ultra-downscaled to an ultimate form by using deep-trench and pattern-dependent oxidation-induced strains, resulting in a SET device with a Coulomb island size of less than 5 nm [21]. By wrapping a gate almost completely around this island, good control of the local electron potential is maintained. Figure 12.1a shows a schematic 3-D layout of the device. A cross-sectional schematic view along the channel (cutline a–b) is seen in Fig. 12.1b. The fabrication process is summarized as follows. Devices were fabricated using a 50-nm-thick undoped silicon-on-insulator substrate. First, a 20-nm-wide nanowire, connecting the S/D electrodes, was formed using electron-beam lithography (EBL) with PMMA resist and subsequent reactive ion etching (RIE) using  $SF_6/CF_4/O_2$ . After deposition of a 100 nm-thick TEOS1 layer, an 80-nm-wide trench was etched across the nanowire using EBL with ZEP520A resist and subsequent RIE etching using  $CHF_3/O_2$ . The exposed silicon area under the trench was etched by a further 30 nm in depth by dry etching using  $SF_6/CF_4/O_2$ , and followed by gate oxidation at 900 °C for 50 min (SET\_A) and 40 min (SET\_B). See Fig. 12.1c, d for TEM images of the cross-sectional view of the etched wires (along the cutline c–d) after oxidation, showing the island size reduced to ~2-nm (SET\_A) and ~4-nm (SET\_B) respectively. Good control over the island size was achieved through this oxidation process. Next, a 20-nm-thick TEOS2 layer was deposited into the trench to form spacers on the sidewalls of the source and drain (S/D) electrodes. Doped poly-silicon was then deposited into the trench to form a self-aligned gate very close to the Coulomb island. The fabrication was completed with conventional S/D ion implants, annealing, and contact processes. Figure 12.1e shows a TEM image of the top-Si nanowire active channel, the poly-Si gate, and the TEOS spacer profiles along the cutline a–b. The gate above the center of the etched region of the active channel is about 10-nm wide and wraps most of the way around the active channel. The schematic 3D layout for the fabrication process flow is illustrated in Fig. 12.2. Note how the top-Si nanowire, exposed by the gap between the source and the drain, is further etched down to 30 nm in depth by dry etching and subsequent gate oxidation. In this process, most of nanowire is masked by the first spacer TEOS1, while only the exposed part by the gap is subject to the oxidation-induced strain. This key process step, different from the conventional finFETs [24], enables an ultrasmall size Coulomb island to be formed with nearly identical tunnel barriers in a self-aligned manner.

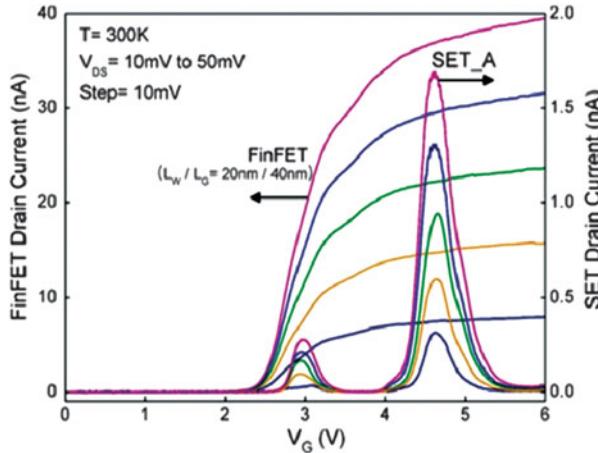


**Fig. 12.1** (a) Schematic 3-D layout of the SET device. (b) Cross-sectional view along the channel (the *cutline a-b*). (c, d) Cross-sectional TEM image of the etched Si wires (along the *cutline c-d*) after oxidation at 900 °C for 50 and 40 min, showing Coulomb island sizes of  $\sim 2$  nm for SET\_A and  $\sim 4$  nm for SET\_B. (e) Cross-sectional TEM images along the *cutline a-b* after deposition of the poly-Si fin-gate. From [21]



**Fig. 12.2** Schematic 3-D layout for process flow. (a) A nanowire was patterned into the 50-nm-thick top-Si layer of the SOI using electron-beam lithography (EBL) with PMMA resist. (b) The pattern transferred by RIE etching with  $\text{SF}_6/\text{CF}_4/\text{O}_2$  to realize a 20-nm-wide nanowire. (c) Deposition of a 100-nm-thick TEOS layer (TEOS1). (d) Trench patterning using EBL with ZEP520A resist. (e) Etching of the TEOS1 layer using RIE with  $\text{CHF}_3/\text{O}_2$  to form a nanogap. (f) The top-Si nanowire, exposed under the gap, was further etched by 30 nm in depth using RIE with  $\text{SF}_6/\text{CF}_4/\text{O}_2$ . (g) Gate oxidation at  $900^\circ\text{C}$ , followed by deposition of a 20-nm-thick TEOS layer (TEOS2) to form spacers on the sidewalls of the source and drain. (h) Deposition of doped poly-Si into the gap to form a self-aligned fin-gate of  $\sim 10$  nm width, which wraps almost completely around the etched Si-nanowire

Figure 12.3 shows the drain current measured at 300 K as a function of the gate voltage  $V_G$  for SET\_A, which is compared with that of the conventional finFET that was fabricated by similar process in the same wafer, but without the deep trench and pattern-dependent oxidation process on the silicon wire channel [21, 22]. As expected, the finFET exhibits typical  $I-V_g$  MOSFET characteristics with its



**Fig. 12.3** Comparison of the  $I-V_g$  characteristics of SET with those of the conventional finFET for drain bias up to 50 mV at 300 K. The finFET that was fabricated by similar process in the same wafer but without the deep-trench and pattern-dependent oxidation process on the silicon wire channel. Typical  $I-V_g$  MOSFET characteristics are observed in the finFET, with its threshold at  $V_g \sim 2.5$  V at 300 K, while well-defined Coulomb oscillations with high PVCRs are observed in the SET. Note that the 1st Coulomb peak of the SET appears just above the threshold of the finFET of  $V_g \sim 2.5$  V for drain bias voltages up to 50 mV. From [22]

threshold at  $V_g \sim 2.5$  V at 300 K, while well-defined Coulomb oscillations with high PVCRs are observed in the SET. This demonstrates that the ultrasmall Coulomb island of the SET is formed by deep-trench of the wire and subsequent oxidation process. Note that the 1st Coulomb peak of the SET appears just above the threshold of the finFET of  $V_g \sim 2.5$  V for drain bias voltages up to 50mV. The onset of the Coulomb oscillations was not detected below the threshold voltage, indicating the 1st peak to be associated with the first electron tunneling.

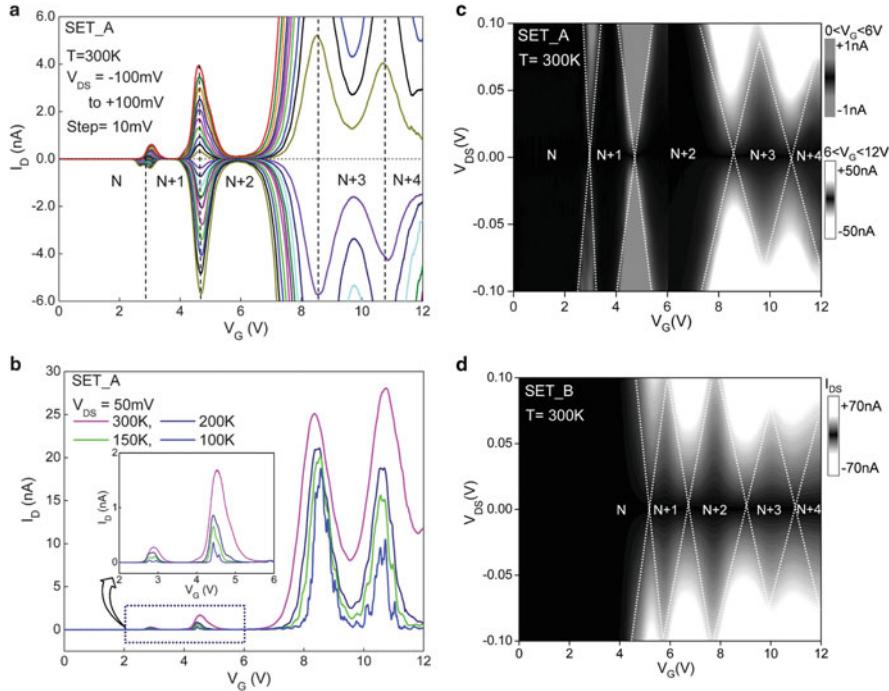
### 12.3 Room-Temperature Multi-switching $I-V_g$ Characteristics: Rough Estimation of the Charging Energy and Quantized Level Spacing

Figure 12.4a shows typical Coulomb oscillations from the SET\_A measured at room temperature [21]; these characteristics show high and nearly symmetric PVCRs, indicating that the tunnel barriers at both sides of a Coulomb island are practically identical. The temperature dependence of these characteristics is shown in Fig. 12.4b; as the temperature is reduced down to 100 K the Coulomb peak positions barely change, but the PVCRs increase. Room temperature charge stability plots are shown in Fig. 12.4c, d for SET\_A and SET\_B, where three and a half successive Coulomb diamonds are clearly seen; each diamond corresponds to a

stable charge configuration on the Coulomb island with the number of occupied electrons  $N$ . Note that Coulomb diamonds for each  $N$  are very symmetric with respect to the positive and negative drain biases, strongly indicating that a single ultrasmall Coulomb island is formed at middle point of the channel and that its tunnel barriers with source and drain are nearly identical. This rules out a possibility of that the observed Coulomb oscillations may be related to the possible dopant or defect which must be randomly formed. The charge stability data also exhibit that as the gate voltage is made less positive, the slope of each Coulomb diamond steeply increases, and the Coulomb diamond (for  $V_G < 3$  V) does not close. This feature is consistent with the lack of any Coulomb peaks below the threshold (Fig. 12.3), indicating that the island is unpopulated by electrons for  $V_G \leq 3$  V, allowing the dot occupancy  $N = 1$  for the 1st diamond. To more convince the assignment of the dot occupancy, a charge sensing device by means of a separated circuit such as an additional quantum point contact [25–27], or RF-SET [28–31], can be installed next to the Coulomb island. In this case, however, an additional sensing bridge gate should be designed to be located very close to the dot to maximize mutual charge coupling. This, without doubt, yields a substantial increase of the total capacitance of the Coulomb island, leading to the SET operating only at ultralow temperatures. Any kinds of room-temperature features will vanish. This is why most of experiments based on these devices including charge sensor have shown Coulomb oscillation behavior only at dilution refrigerator temperatures. We, therefore, addressed the assignment of the dot occupancy by somewhat indirect ways (mentioned above) without using additional charge sensor.

It is noted that substantial change in slopes and diagonal sizes of each successive diamond (Fig. 12.4c, d) implies that the charging energy is not constant over the gate voltage range studied, and the orthodox Coulomb blockade theory may break down in this case. This behavior may be caused by strong interplay of the Coulomb interaction and additional quantum effects associated with very low electron number on the island. Nevertheless, we can roughly estimate the size of the island and calculate the Coulomb charging energy because the 1st diamond associated with the lowest dot occupancy  $N = 1$  is determined mainly by the Coulomb charging energy. Values for the gate and junction capacitances can be directly obtained from the Coulomb peak spacing  $\Delta V_G$  and the slopes of the 1st diamond [1]. We find that, for SET\_A,  $C_G \sim 0.094$  aF,  $C_S \sim 0.16$  aF, and  $C_D \sim 0.17$  aF, yielding the total capacitance  $C_{\Sigma} \sim 0.42$  aF, which corresponds to a  $\sim 1.94$  nm diameter spherical dot, are in good agreement with the TEM image of SET\_A in Fig. 12.1c. Similarly, the total capacitance of SET\_B is estimated to be  $C_{\Sigma} \sim 1.04$  aF, which corresponds to a  $\sim 4.6$  nm diameter dot. The charging energy of the SET\_A (SET\_B) is thus  $e^2/C_{\Sigma} \sim 0.377$  eV ( $\sim 0.154$  eV) which is more than one order magnitude larger than the thermal energy at room temperature, yielding the clear Coulomb oscillations with high PVCRs even at RT.

The level spacing due to the quantum confinement can be also roughly estimated. The quantum effect is expected to be enhanced in such a small dot of  $< 5$  nm, which must be stronger in the SET\_A whose size is much smaller than the



**Fig. 12.4** (a) Drain current of SET\_A measured at 300 K as a function of fin-gate voltage  $V_G$  using drain voltages from  $-100$  to  $100$  mV with a step of  $10$  mV. (b) Temperature dependence of SET\_A at a fixed drain voltage of  $V_{DS} = 50$  mV. The inset shows detail from the first two Coulomb peaks in the current regime marked with dotted line. (c, d) Charge stability plot at room temperature for SET\_A and SET\_B; each diamond corresponds to a stable charge configuration state with fixed electron occupancy  $N$ . From [21]

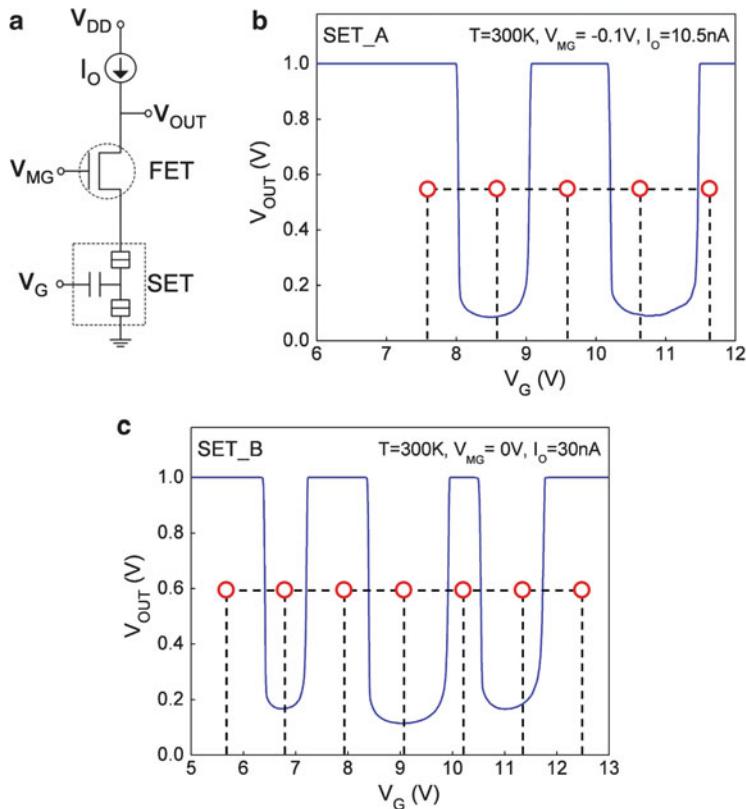
SET\_B. If we associate the 2nd diamond with the dot occupancy  $N = 2$ , the unusual large shift of the 3rd Coulomb peak of the SET\_A could be related to the electron filling in a spin-up-spin-down sequence, i.e., spin-filling of the next orbital state for the 3rd electron number following Pauli's exclusion rule. (Similarly, the 3rd Coulomb peak of the SET\_B is observed to have an additional gate voltage shift of  $\sim 0.6$  V, which is  $\sim 40\%$  larger than the other two  $V_G$  spacings.) The level spacing due to the quantum confinement can be roughly estimated based on the effective mass approach. For a 2-nm silicon dot modeled by a 3D hard-wall potential, the level spacing is estimated to be  $\sim 0.493$  eV, which is comparable to the Coulomb energy  $e^2/C_\Sigma \sim 0.377$  eV. More complete analysis considering quantum effects needs extensive transport studies including  $I-V_g$  measurements for ultralow temperatures, which will be discussed in Sect. 12.5.

## 12.4 Room-Temperature SET/FET Hybrid Circuits: Multi-valued Literal Gate Functionality

For practical circuit applications, we integrated the SET with a FET and measured the multiple-valued literal gate functionality at room temperature [21]. A schematic layout of the SET/FET hybrid circuit is seen in Fig. 12.5a, showing that it comprises of a SET, a MOSFET, and a constant-current load. The FET connected in series with the SET was to attain the SET bias voltage under a range of Coulomb blockade. This circuit, called universal literal gate consisting of a FET and a SET connected in series [7], operates under a constant current mode with a drain voltage limit (given by setting a compliance voltage of the HP4155). In the SET-OFF mode, the current is off, and the output of the circuit is same as the drain voltage limit. On the other hand, in the SET-ON mode, the circuit operates under a constant current mode, but also with a drain voltage limit. The voltage transfer characteristics of the resultant SET/FET literal gate measured at 300 K are seen in Fig. 12.5b, c, displaying five (seven) periodic voltage outputs of high/low level multiple switching with a sharp swing as high as  $\sim 1$  V for SET\_A (SET\_B). This enables multi-bit ( $>5$  bit) functionality at room temperature and can provide a basic block for the practical implementation of the SET-based MV logic architectures with ultralow power consumption.

## 12.5 Room-Temperature Charge Stability: Enhanced Coulomb Blockade and Quantum Effects in a Ultrasmall Coulomb Island with Few Electrons

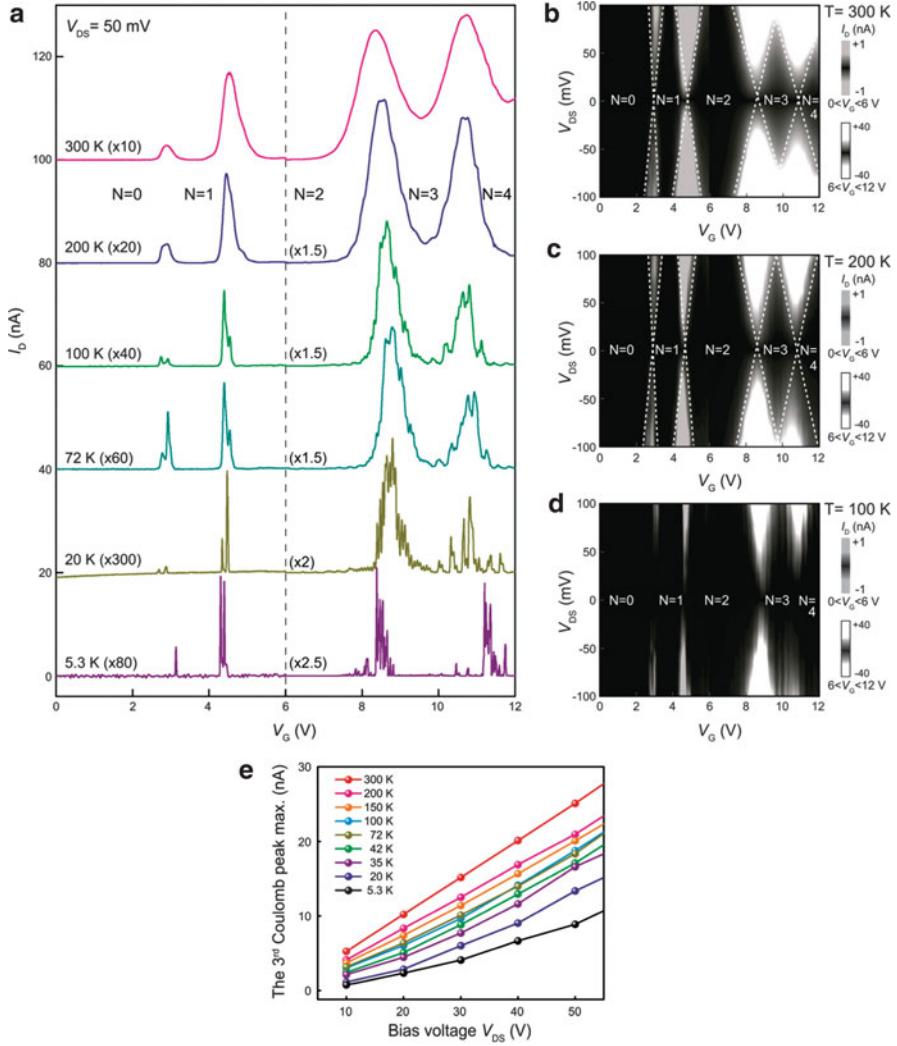
Apart from the nanoelectronic applications, the ultrasmall SET can be regarded as a mesoscopic artificial atom system. Tunneling through it can provide a rich experimental environment for studying quantum transport phenomena (for reviews, see [32–36]). Previously, these quantum effects have been investigated using relatively large devices at low temperatures, where they give rise to additional fine structures on the Coulomb oscillations [14, 37–43]. However, as temperature increases up to 300 K, such fine structures observed at low temperature normally vanish together with Coulomb peaks themselves because of the weak Coulomb charging energy due to the relatively large dot size. The size of dot should thus be ultrasmall in order to observe the quantum effect as well as the Coulomb blockade at room temperature. As the dot size is reduced below 5 nm, the very small number of electrons on the dot is expected, even at 300 K, to ensure that electron–electron interactions with Pauli spin exclusion strongly influence the electron transport characteristics. Regarding this issue, we reported on an extensive transport measurement performed on a room-temperature-operating Coulomb blockade device with an ultrasmall silicon island of sub-5-nm size [22]. Transport data exhibit a striking feature of that the main room-temperature characteristics of the Coulomb peaks persist even at ultra-low temperature down to 5.3 K. Substantial change in slopes and diagonal size of



**Fig. 12.5** (a) Circuit layout of the SET/FET literal gate that comprises of a SET, a MOSFET, and a constant-current load. (b, c) The input-output voltage transfer of the resultant SET/FET literal gate measured at 300 K, displaying more than 5 bits multiple switching with a sharp swing as high as  $\sim 1$  V. The constant-current  $I_0$  is 10.5 nA (30 nA) and the  $V_{MG}$  for the MOSFET is set to  $-0.1$  V (0 V) for SET\_A (SET\_B). From [21]

the room-temperature Coulomb diamond and bias-dependent peak splitting must reflect low energy many-body excited states associated with total spin for each dot occupancy  $N$ . As expected, the quantum effects become enhanced in the ultrasmall Coulomb island of sub-5 nm and persist even at room temperature, leading to the substantial modulation of the charge stability for finite bias window.

Figure 12.6 shows the extensive  $I-V_g$  characteristics of the SET\_A measured over the temperature range from 300 K down to 5.3 K for a bias 50 mV [22]. As seen in Fig. 12.6a, the main feature of the Coulomb oscillations of 300 K persists even at low temperature down to 5.3 K, except for additional splitting observed in each Coulomb peak. We point out that this temperature-dependent feature is quite new and remarkable because the peak splitting so far observed at low-temperatures have been reported to vanish together with Coulomb peaks themselves with increasing temperature. This strongly indicates another evidence that the Coulomb island of the SET



**Fig. 12.6** (a) Temperature dependence of the  $I$ - $V_g$  characteristics of the SET measured for various temperatures down to 5.3 K for a bias  $V_d = 50$  mV. Note that the main feature of 300 K persists even at low temperature down to 5.3 K, but a striking temperature-dependent splitting is observed in each Coulomb peak. (b–d) Charge stability plot for temperatures of 300 K, 200 K, and 100 K, respectively. Each Coulomb diamond corresponds to a stable charge configuration state with fixed electron occupancy  $N$ . Peak splitting are clearly seen for diamonds of  $N = 3$  and 4 even at 100 K. (e) Temperature-dependent magnitude of the 3rd Coulomb peak for each bias up to 50 mV. From [22]

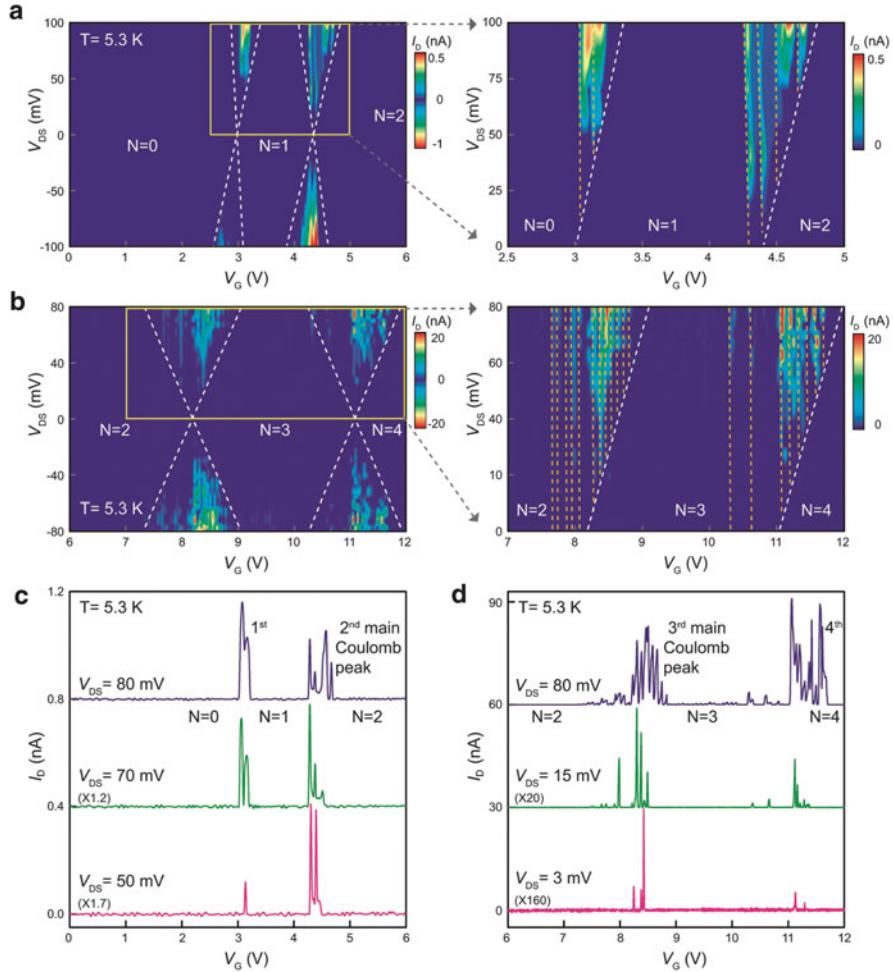
device is a quite small well-defined single dot (of sub-5 nm size), and its charging energy is large enough to get over the room-temperature thermal energy. If the main four Coulomb peaks were due to some multiple defects at Si/SiO<sub>2</sub> interface, some of the peaks should be randomly created or disappear depending on their thermal

activation energies as temperature changes. Note also that the magnitude of the Coulomb peak decreases for low temperatures, as seen in Fig. 12.6e which shows the temperature-dependent magnitude of the 3rd Coulomb peak for each bias up to 50 mV. This can be attributed to the possible decrease of the carrier concentration (or, carrier freeze-out) in the S/D wires as temperature decreases down to 5.3 K.

Charge stability plots (displayed for down to 100 K) of the SET are seen in Fig. 12.6b-d, respectively, where successive Coulomb diamonds are clearly seen. Each diamond corresponds to a stable charge configuration state with fixed electron occupancy  $N$ . Coulomb peak splitting are seen for diamonds of  $N = 3$  and 4 in the charge stability even at 100 K. It is noted that substantial change in slopes and diagonal size of each successive diamond is observed, implying that the charging energy is not constant over the gate voltage range studied. As mentioned in the previous section, this behavior could be accounted for by strong interplay of the Coulomb interaction and additional quantum effects associated with very low electron number on the island. The fine structure with decreasing temperature must reflect low energy excited levels associated with each dot occupancy  $N$  and can be explored more in detail with increasing bias window. Figure 12.7a, b is charge stability data measured at 5.3 K [22]. They illustrate the fine structure of the bias dependence of the Coulomb oscillations, showing typical behavior of increasing splitting with bias window. For more clarity, the charge stability data are illustrated in Fig. 12.7c, d, reproducing  $I_d-V_g$  for some specific bias voltages. As seen in Fig. 12.7a with Fig. 12.7c, when bias voltage increases up to 100 mV, the 1st peak starts to split into two sub-peaks and persists even at high bias, while the 2nd peak splits into four sub-peaks. Note that for the 1st main peak, the valley between two sub-peaks is raised up with bias voltage, indicating the increase in tunneling current as bias window becomes wide. This is not due to a peak broadening effect because the heating energy by increasing bias is only about  $\sim 0.05$  nW, negligible compared to the thermal energy of 5.3 K. Similarly, Fig. 12.7b with Fig. 12.7d illustrate that the number of splitting of the 3rd Coulomb peaks rapidly increase from three to more than 12, while that of the 4th peak increases from two to more than 8. This strong bias dependence of peak splitting demonstrates the evident transition of the transport behavior of our device from linear to nonlinear transport regime where single-electron tunneling can be made through many excited levels lying within the bias window [44, 45].

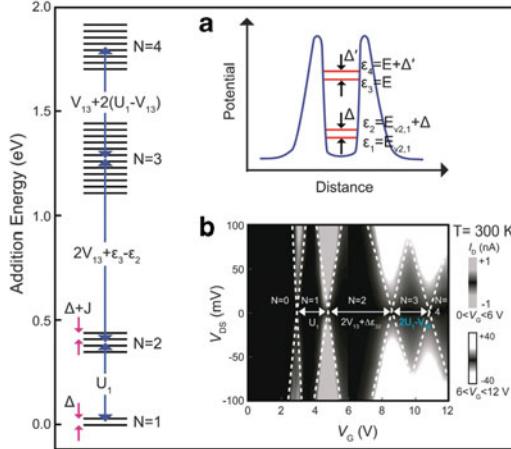
## 12.6 Model Analysis for Interplay of the Coulomb Interaction and Additional Quantum Effects Associated with Few Electron Dot Occupations

To explain the observed fine structure in each main Coulomb peak of the ultrasmall SET device in the previous section, we need to know the low energy level spectrum associated with each dot occupancy  $N$ . Here we will follow the model analysis developed in [22].



**Fig. 12.7** Charge stability plot at 5.3 K and specific bias dependence of each main Coulomb peak; (a, b) Charge stability plot at 5.3 K, showing typical behavior of increasing splitting with bias window. (c, d)  $I-V_g$  characteristics for some specific bias voltages, which are reproduced from the charge stability data. Strong bias dependences of peak splitting are clearly seen, which can be accounted for by the nonlinear transport made through many excited levels associated with each dot occupancy  $N$ . From [22]

The Coulomb channel in the SET fabricated on (100) Si-2D system is surrounded by SiO<sub>2</sub> insulator and is along <110> direction that is parallel to the notch orientation axis of our SOI wafer. In such a wire valley splitting lifts the two- and fourfold degeneracies into two, and the energy levels of  $\Gamma$  valleys are lower than those of off- $\Gamma$  valleys [46, 47]. In Si valley splitting is a main source of spin decoherence. It has been observed even in zero external electric field in strongly confined nanostructures and could be further enhanced by the application of bias,



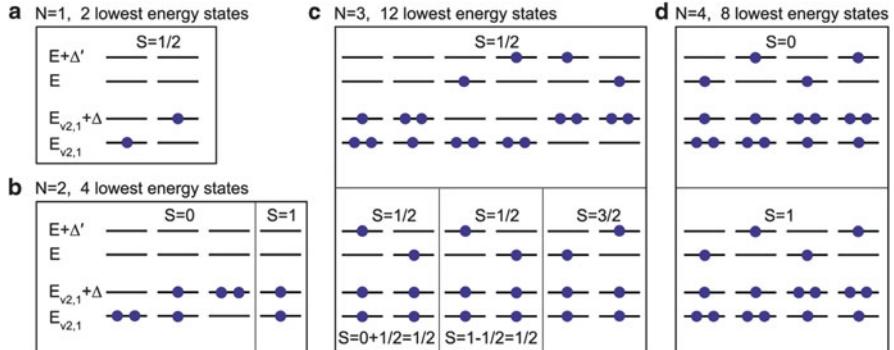
**Fig. 12.8** Addition charging energies of  $N = 1 \rightarrow 2$ ,  $2 \rightarrow 3$ , and  $3 \rightarrow 4$ , estimated from the many-body Hamiltonian. The low energy level spectrum associated with each dot occupancy  $N$  are illustrated. Inset (a) illustrates a confinement potential along the wire where the energy levels of valleys are quantized. The calculated addition charging energies, approximately  $U_1 \approx 0.38$  eV,  $2V_{13} + \epsilon_3 - \epsilon_2 \approx 0.9$  eV, and  $V_{13} + 2(U_1 - V_{13}) \approx 0.46$  eV (for  $N = 1 \rightarrow 2$ ,  $2 \rightarrow 3$ , and  $3 \rightarrow 4$ , respectively) are denoted by arrows in inset (b), which are in the same range as those of the charge stability data observed at 300 K. From [22]

strain, or magnetic field [48, 49]. In the presence of a confinement potential along the wire, the energy levels of valleys are quantized, as shown schematically in the inset (a) of Fig. 12.8. With the inclusion of the valley, splitting the four lowest energy levels of the dot are  $\epsilon_1 = E_{v2,1}$ ,  $\epsilon_2 = E_{v2,1} + \Delta$ ,  $\epsilon_3 = E$ , and  $\epsilon_4 = E + \Delta'$ , where  $E_{v2,1}$  is the lowest energy quantized with valley splitting  $\Delta$ , and  $E$  is the second lowest energy quantized with valley splitting  $\Delta'$ . While  $E_{v2,1}$  and  $E_{v2,1} + \Delta$  originate from  $\Gamma$  valley, the energy  $E$  may originate either from  $\Gamma$  or off- $\Gamma$  valleys.

Based on the above information on single electron levels, we model the many-body Hamiltonian (for reviews, see [50, 51]) of the dot by

$$H = \sum_i \epsilon_i n_i + \sum_{i < j} V_{ij} n_i n_j + \sum_i U_i n_{i\uparrow} n_{i\downarrow} - \sum_{i < j} J_{ij} \left( \vec{S}_i \cdot \vec{S}_j + \frac{1}{4} n_i n_j \right). \quad (12.1)$$

Here, the label  $i = 1, 2, 3, 4$  denotes the four single electron states. For each single electron level  $i$  we define, respectively, the quantities  $\vec{S}_i$ ,  $n_{i\sigma}$ ,  $n_i$ , and  $U_i$  as the spin operator, number operator of electrons with spin  $\sigma$ , number operator of occupied electrons, and the intra-level Coulomb repulsion. The Coulomb repulsion (exchange) energy between an electron in the  $i$ 'th and an electron in  $j$ 'th levels is  $V_{ij}$  ( $J_{ij}$ ). Each many-body eigenstate can be represented by a ket state  $| \{n_i\}, S, S_z \rangle$ . Level occupation numbers  $\{n_i\}$  and the total spin quantum number  $S$  of some of the lowest energy many-body states are analyzed and displayed in Fig. 12.9 (note that since the Hamiltonian is spin rotationally invariant, eigenvalues are independent of



**Fig. 12.9** Electronic occupation configurations illustrating the number of lowest energy states for (a)  $N = 1$ , (b)  $N = 2$ , (c)  $N = 3$ , and (d)  $N = 4$ , respectively. Note that due to many-body exchange interactions states with the same  $\{n_i\}$  but with different  $S$  do not have the same energy. From [22]

the  $z$ -component total spin  $S_z$ ). In this classification of eigenstates it is useful to exploit the fact that when the  $i$ 'th level is doubly occupied, its spin state is necessarily a singlet state with  $\vec{S}_i = 0$ . In addition, according to the quantum rules of spin addition, the total spin state  $S = 1/2$  of 3 electrons can be constructed by adding spin 0 and 1/2 or by subtracting spin 1/2 from 1. This implies that, when adding 3 electron spins, there are two different spin wavefunctions for the same  $S = 1/2$  and  $\{n_i\}$ . Due to many-body exchange interactions, states with the same  $\{n_i\}$  but with different  $S$  do not have the same energy. For example, for  $N = 2$ , the  $S = 0$  singlet and  $S = 1$  triplet states are split, see Fig. 12.9b. From our observed data, Fig. 12.7c, the magnitude of the exchange interaction is estimated to be about 0.01 eV. Using these rules we find, respectively, 2, 4, 12, and 8 number of lowest energy states for  $N = 1, 2, 3$ , and 4 that can be formed from the energies  $\epsilon_1, \epsilon_2, \epsilon_3$ , and  $\epsilon_4$ , see Fig. 12.9a-d. We stress that these numbers are independent of model parameters. For a given  $N$ , estimate of energies shows that there is an energy gap to the next excited states from the group of lowest energy states mentioned above.

For  $N = 1$  and 2 the theoretical number of lowest energy levels agree with experimentally observed peak values of 2 and 4. According to our model for  $N = 2$  there are 3 singlet and 1 triplet levels, see Fig. 12.9b. The three singlet states have all different energies because their occupation number configurations  $\{n_i\}$  are different. For  $N = 3$  we displayed 12 lowest energy states in Fig. 12.9c. In near agreement with this value, the observed number low energy excited states is about 12–14 (when small noise-like peaks are included there are 14 peaks). For  $N = 4$  some of the lowest excited energy states have the same  $\{n_i\}$  but different spin values  $S = 0$  and 1, see Fig. 12.9d. Their energies are again different because these states have different spin wavefunctions and, therefore, have different exchange energies. For  $N = 4$  the observed number of excited states is between 8 and 10, which is close to predicted value of 8, see Fig. 12.9d. Since there may be

other single electron energy levels close to  $\varepsilon_3$  and  $\varepsilon_4$  different many-body excited energies may be present near those 12 and 8 states as shown in Fig. 12.9c, d. Moreover, the effect of quantum fluctuations of occupation numbers, which is absent in Hartree–Fock approximations, may give rise to additional excited states [52]. These factors suggest a possible explanation for why more than 12 and 8 peaks may have been observed for  $N = 3$  and 4, respectively, when small noise-like peaks are included.

The Coulomb interaction energy between two electrons in the level 1 is  $U_1 \approx \frac{\epsilon^2}{\epsilon R} \approx 0.38 \text{ eV}$ , where  $\epsilon$  and  $R$  are the dielectric constant and the radius of the dot. The energy separation between different Coulomb peaks can be fitted by choosing two parameters  $V_{13}$  and  $\varepsilon_3 - \varepsilon_2$  judiciously:  $V_{13} \approx 0.3 \text{ eV}$  is the Coulomb interaction energy between two electrons in the levels 1 and 3 and  $\varepsilon_3 - \varepsilon_1 \approx 0.3 \text{ eV}$  is the energy separation between them. The fit value of  $V_{13}$  is reasonable because the inter-level Coulomb interaction is comparable to the intra-level  $U_i$  but must be smaller than it. The fit value of the quantum confinement of  $\varepsilon_3 - \varepsilon_1$  is also in the range of expected value for our dot size of 2 nm by theoretical calculations [46, 47]. Using these values we find that the addition charging energies of  $N = 1 \rightarrow 2$ ,  $2 \rightarrow 3$ , and  $3 \rightarrow 4$  are approximately  $U_1 \approx 0.38 \text{ eV}$ ,  $2V_{13} + \varepsilon_3 - \varepsilon_2 \approx 0.9 \text{ eV}$ , and  $V_{13} + 2(U_1 - V_{13}) \approx 0.46 \text{ eV}$ , which are illustrated in Fig. 12.8. Using the energy coupling factor, defined by  $\alpha_G = C_G/C_\Sigma \sim 0.22$ , the corresponding observed values 0.37, 0.84, and 0.50 eV (see the inset b in Fig. 12.8) are in the same range. For  $N = 1$  the observed Coulomb oscillations exhibit splitting of the 1st peak is due to the valley splitting, see Fig. 12.9a. The measured value of valley splitting  $\Delta$  is 16 meV, whose order of magnitude is consistent with recently reported theoretical values of sub-3 nm Si nanostructures [46, 47]. Note that its value is much smaller than Coulomb charging and quantum confinement energies. These approximate agreements between experimental and theoretical values suggest that our model can account consistently for several features of excited states in the ultrasmall Si dot formed along  $<110>$  direction.

## 12.7 Conclusions

In summary, we report an extensive transport measurements carried out on a sub-5-nm ultrasmall dot-based single-electron device; this size is sufficiently small that Coulomb blockade, and other quantum effects, persist up to room temperature. These devices were made by ultrascaling a state-of-the-art finFET structure down to an ultimate limiting form, resulting in the reliable formation of a sub-5-nm silicon Coulomb island. The charge stability data feature the first exhibition of three and a half clear Coulomb diamonds at 300 K, each showing high PVCRs. Its charging energy and quantized level spacing are estimated to be more than one order magnitude larger than the thermal energy at room temperature. The room-temperature feature of  $I-V_g$  persists even at low temperature down to 5.3 K,

where additional fine structures of Coulomb peaks appear. The unusual energy separation between Coulomb diamonds and the fine splitting of each Coulomb peak are accounted for by including quantum many-body interactions, leading to the substantial modulation of the room-temperature charge stability. This supports that for a sub-5-nm device even small number of electron occupations on the Coulomb island ensures that quantum many-body interactions strongly influence the room-temperature electron transport characteristics. Under this condition, quantum effect may be used as an additional state variable to provide another multi-switching functionality. Besides reducing the power consumption and multi-switching functionality, the successful CMOS-compatible implementation of the room-temperature multi-switching SETs could find use in a variety of other applications such as ultrasensitive single-molecule biological sensors, brain-inspired neuromorphic circuits, and scalable solid-state quantum computing/read-out systems.

**Acknowledgments** This work was supported by the National Research Foundation through the Frontier 21 National Program for Tera-level NanoDevices, Global Partnership Research Program with University of Cambridge, and in part by a research grant of the Chungbuk National University in 2011.

## References

1. Averin, D.V., Likarev, K.K.: In: Grabert, H., Devoret, M.H. (eds.) Single charge tunneling, p. 311. Plenum, New York (1992)
2. Yano, K., Ishii, T., Hashimoto, T., Kobayashi, T., Murai, F., Seki, K.: IEEE Trans. Electron Dev. **41**, 1628 (1994)
3. Tiwari, S., Rana, F., Hanafi, H., Hartstein, A., Crabbé, E.F., Chan, K.: Appl. Phys. Lett. **68**, 1377 (1996)
4. Guo, L., Leobandung, E., Chou, S.Y.: Science **275**, 649 (1997)
5. Akazawa, M., Kanami, K., Yamada, T., Amemiya, Y.: IEICE Trans. Electron. E **82-C**(9), 1607 (1999)
6. Takahashi, Y., Fujiwara, A., Yamazaki, K., Namatsu, H., Kurihara, K., Murase, K.: Appl. Phys. Lett. **76**(5), 637–639 (2000)
7. Inokawa, H., Fujiwara, A., Takahashi, Y.: IEEE Trans. Electron Dev. **50**, 462 (2003)
8. Degawa, K., Aoki, T., Higuchi, T., Inokawa, H., Takahashi, Y.: IEICE Trans. Electron. E **87-C** (11), 1827 (2004)
9. Mahapatra, S., Ionescu, A.M.: IEEE Trans. Nanotechnol. **4**(6), 705 (2005)
10. Lee, C.K., Kim, S.J., Shin, S.J., Choi, J.B., Takahashi, Y.: Appl. Phys. Lett. **92**, 093101 (2008)
11. Kim, S.J., Lee, C.K., Chung, R.S., Park, E.S., Shin, S.J., Choi, J.B., Yu, Y.S., Kim, N.S., Lee, H.G., Park, K.H.: IEEE Trans. Electron Dev. **56**, 1048 (2009)
12. Kim, S.J., Lee, J.J., Kang, H.J., Choi, J.B., Yu, Y.-S., Takahashi, Y., Hasko, D.G.: One Electron-based Smallest Flexible Logic Cell. Appl. Phys. Lett. **101**, 183101 (2012)
13. Takahashi, Y., Naease, M., Namatsu, H., Kurihara, K., Iwdate, K., NakajiGa, Y., Horiguchi, S., Murase, K., Tabe, M.: Electron. Lett. **31**, 136 (1995)
14. Leobandung, E., Guo, L., Wang, Y., Chou, Y.S.: Appl. Phys. Lett. **67**, 938 (1995)
15. Reed, M.A., Zhou, C., Muller, C.J., Burgin, T.P., Tour, J.M.: Science **278**, 252 (1997)

16. Ono, Y., Takahashi, Y., Yamazaki, K., Nagase, M., Namatsu, H., Kurihara, K., Murase, K.: IEEE Trans. Electron Dev. **47**, 147 (2000)
17. Tan, Y.T., Kamiya, T., Durrani, Z.A.K., Ahmed, H.: Room temperature nanocrystalline silicon single-electron transistors. J. Appl. Phys. **94**, 633–637 (2003)
18. Saitoh, M., Hiramoto, T.: Appl. Phys. Lett. **84**, 3172 (2004)
19. Li, P.W., Kuo, D.M.T., Liao, W.M., Lai, W.T.: Appl. Phys. Lett. **88**, 213117 (2006)
20. Kimura, Y., Itoh, K., Yamaguchi, R., Ishibashi, K., Itaya, K., Niwano, M.: Appl. Phys. Lett. **90**, 093119 (2007)
21. Shin, S.J., Jung, C.S., Park, B.J., Yoon, T.K., Lee, J.J., Kim, S.J., Choi, J.B., Takahashi, Y., Hasko, D.G.: Appl. Phys. Lett. **97**, 103101 (2010); patent US 8158538; NPG Asia Materials Featured Highlight, “Transistors: One at a time” (8 Nov 2010); doi:[10.1038/asiamat.2010.176](https://doi.org/10.1038/asiamat.2010.176)
22. Shin, S.J., Lee, J.J., Kang, H.J., Choi, J.B., Yang, S.-R.E., Takahashi, Y., Hasko, D.G.: Nano Lett. **11**(4), 1591 (2011)
23. Mol, J.A., Verduin, J., Levine, R.D., Remacle, F., Rogge, S.: Proc. Natl. Acad. Sci. USA **108**, 13969 (2011)
24. Hisamoto, D., Lee, W.C., Kedzierski, J., Takeuchi, H., Asano, K., Kuo, C., Anderson, E., King, T.J., Bokor, J., Hu, C.: IEEE Trans. Electron Dev. **47**, 2320 (2000)
25. DiCarlo, L., Lynch, H.J., Johnson, A.C., Childress, L.I., Crockett, K., Marcus, C.M., Hanson, M.P., Gossard, A.C.: Phys. Rev. Lett. **92**, 226801 (2004)
26. Hu, Y., Churchill, H.O.H., Reilly, D.J., Xiang, J., Lieber, C.M., Marcus, C.M.: Nat. Nanotechnol. **2**, 622 (2007)
27. Nordberg, E.P., Stalford, H.L., Young, R., Ten Eyck, G.A., Eng, K., Tracy, L.A., Childs, K.D., Wendt, J.R., Grubbs, R.K., Stevens, J., Lilly, M.P., Eriksson, M.A., Carroll, M.S.: Appl. Phys. Lett. **95**, 202102 (2009)
28. Brenning, H., Kafanov, S., Duty, T., Kubatkin, S., Delsing, P.: J. Appl. Phys. **100**, 114321 (2006)
29. Angus, S.J., Ferguson, A.J., Dzurak, A.S., Clark, R.G.: Appl. Phys. Lett. **92**, 112103 (2008)
30. Podd, G.J., Angus, S.J., Williams, D.A., Ferguson, A.: J. Appl. Phys. Lett. **96**, 082104 (2010)
31. Villis, B.J., Orlov, A.O., Jehl, X., Snider, G.L., Fay, P., Sanquer, M.: Defect detection in nanoscale transistors based on radio-frequency reflectometry. Appl. Phys. Lett. **99**, 152106 (2011)
32. Averin, D.V., Likharev, K.K.: Mesoscopic phenomena in solids. North-Holland, Amsterdam (1991)
33. Kastner, M.A.: Rev. Mod. Phys. **64**, 849 (1992)
34. Van Houten, H., Beenakker, C.W., Staring, A.A.M.: Single Charge Tunneling. Plenum, New York (1992)
35. Kouwenhoven, L.P., Austing, D.G., Tarucha, S.: Rep. Pro. Phys. **64**, 701 (2001)
36. Van der Wiel, W.G., De Franceschi, S., Elzerman, J.M., Fujisawa, T., Tarucha, S., Kouwenhoven, L.P.: Rev. Mod. Phys. **75**, 1 (2003)
37. McEuen, P.L., Foxman, E.B., Kinaret, J., Meirav, U., Kastner, M.A., Wingreen, N.S., Wind, S. J.: Phys. Rev. B **45**, 11419 (1992)
38. Tarucha, S., Austing, D.G., Honda, T., van der Hage, R.J., Kouwenhoven, L.P.: Phys. Rev. Lett. **77**, 3613 (1996)
39. Kouwenhoven, L.P., Oosterkamp, T.H., Danoesastro, M.W.S., Eto, M., Austing, D.G., Honda, T., Tarucha, S.: Science **278**, 1788 (1997)
40. Simmel, F., Abusch-Magder, D., Wharam, D.A., Kastner, M.A., Kotthaus, J.P.: Phys. Rev. B **59**, 10441 (1999)
41. Khoury, M., Rack, M.J., Gunther, A., Ferry, D.K.: Appl. Phys. Lett. **74**, 1576 (1999)
42. Lee, S.D., Park, K.S., Park, J.W., Choi, J.B., Yang, S.-R.E., Yoo, K.-H., Kim, J., Park, S.I., Kim, K.T.: Phys. Rev. B **62**, R7735 (2000)
43. Rokhinson, L.P., Guo, L.J., Chou, S.Y., Tsui, D.C.: Phys. Rev. B **63**, 035321 (2001)
44. Johnson, A.T., Kouwenhoven, L.P., de Jong, W., van der Vaart, N.C., Harmans, C.J.P.M., Foxon, C.T.: Phys. Rev. Lett. **69**, 1592 (1992)

45. Lee, S.D., Shin, S.J., Choi, S.J., Lee, J.J., Choi, J.B., Park, S., Yang, S.-R.E., Lee, S.J., Zyung, T.H.: *Appl. Phys. Lett.* **89**, 023111 (2006)
46. Neophytou, N., Paul, A., Lundstrom, M.S., Klimeck, G.: *IEEE Trans. Electron Dev.* **55**, 1286 (2008)
47. Maegawa, T., Yamauchi, T., Hara, T., Tsuchiya, H., Ogawa, M.: *IEEE Trans. Electron Dev.* **56**, 553 (2009)
48. Goswami, S., Slinker, K.A., Friesen, M., McGuire, L.M., Truitt, J.L., Tahan, C., Klein, L.J., Chu, J.O., Mooney, P.M., Van der weide, D.W., Joynt, R., Coppersmith, S.N., Eriksson, M.A.: *Nat. Phys.* **3**, 41 (2007)
49. Friesen, M., Eriksson, M.A., Coppersmith, S.N.: *Appl. Phys. Lett.* **89**, 202106 (2006)
50. Fazekas, P.: Lecture notes on electron correlation and magnetism. World Scientific, Singapore (1963)
51. Anderson, P.W.: Concepts in solids. Benjamin, London (1963)
52. Park, S., Yang, S.-R.E.: *Phys. Rev. B* **72**, 125410 (2005)

# Chapter 13

## Single-Electron Tunneling Transistors Utilizing Individual Dopant Potentials

Daniel Moraru and Michiharu Tabe

**Abstract** For many decades since the invention of the transistor in 1947, the dimensions of transistor channels have been continuously downscaled so that more and more functionality can be incorporated into one chip. However, nowadays, critical dimensions of transistors enter into the real nanoscale and fundamental limitations, in physics and technology, raise serious challenges in front of further miniaturization. A conceptually different operation mechanism for the next-generation transistors must be considered. In this framework, we focus on transport characteristics arising from single-electron tunneling via individual dopant atoms, the basic operation mode of *single-dopant transistors*. Single-dopant transistors are devices that make use of individual dopant potentials as natural, ultrasmall quantum dots. In this chapter, we outline basic results related to our research on single-dopant transistors, after briefly introducing the concept and fundamental physics of their operation. First, Kelvin probe force microscopy is used for direct observation of individual dopants in the channel of transistors under normal operation. Next, focus falls on the electrical characteristics indicating single-electron tunneling via individual donor atoms in different temperature ranges. Finally, a domain of dopant-based applications is outlined at the end of the chapter, opening the door for the development of atomic-level electronics.

### 13.1 Introduction

Since the invention of the transistor in 1947, fundamental research on silicon-based transistors allowed a continuous reduction of their critical dimensions, closely following the trend predicted by the famous Moore's law [1]. In spite of predictions

---

D. Moraru • M. Tabe (✉)

Research Institute of Electronics, Shizuoka University, Johoku 3-5-1, Naka-ku, Hamamatsu,

Shizuoka-ken 432-8011, Japan

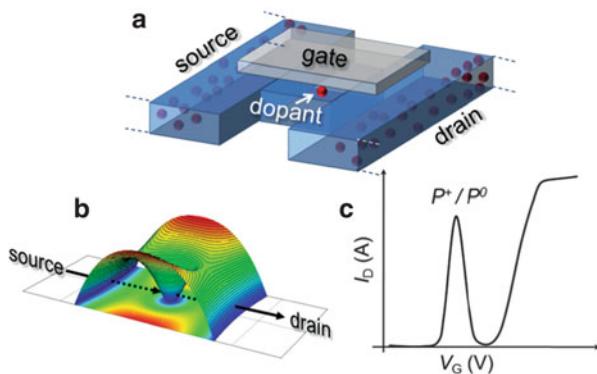
e-mail: [daniel@rie.shizuoka.ac.jp](mailto:daniel@rie.shizuoka.ac.jp); [romtabe@rie.shizuoka.ac.jp](mailto:romtabe@rie.shizuoka.ac.jp)

of the end of the downscaling trend at different stages of the transistor's evolution, the limitations were overcome by technological innovation and focused research. However, nowadays, commercial transistors reached channel dimensions on the order of only several tens of nanometers. It becomes obvious that fundamental and technological limits of the transistors will soon become an insurmountable barrier in front of further miniaturization.

One of the main problems that nanoscale transistors are facing is related to the discrete distribution of dopant (impurity) atoms in the channel. Together with the reduction of channel dimensions, the number of dopant atoms existing in the channel is also strongly reduced [2]. One important research direction is oriented toward understanding [3–5] and improving the impact of the discreteness of the dopant distribution [6, 7] by advancing the doping processes and fabrication technology. This research provides valuable information on the basic physics that could be dominant in nanoscale and could allow several more generations of downscaling. Nevertheless, it is obvious that a radically new transport mechanism must be considered for the later generations of electronics. Alternative technologies, such as FinFETs [8] or a variety of other new approaches [9], have been studied as viable alternatives for conventional silicon transistors, but these new technologies still require more study and other alternatives may be available.

As a significantly different approach, our group, along with several other groups around the world, focused on a different research direction. This approach continues on the basis of devices still built on the well-developed silicon platform and using knowledge accumulated through decades of experience for this material. The transport mechanism, however, is significantly different than for the conventional silicon devices, i.e., *single-electron tunneling mediated by individual dopant atom potential wells*, working as tiny quantum dots (QDs). Over the past few years, several papers have been published on the topic of *single-dopant transistors*, revealing various fundamental aspects of these basic devices [10–16].

The basic principle of single-dopant transistor operation is illustrated in Fig. 13.1 in an ideal design. We can imagine that, in a nanoscale silicon channel, between well-defined source and drain leads, one dopant atom can be introduced. For instance, let's consider the situation of a phosphorus (P) donor atom between *n*-type source and drain leads. When such a system is formed in a field-effect transistor (FET) structure, a gate can control the potential of the channel and, implicitly, the potential at the donor location. In particular, for negatively large gate voltages,  $V_G$ , the channel will be depleted of electrons, which means that the donor atom will also be positively ionized by removing its extra electron. Under these conditions, the ionized donor atom ( $P^+$ ) introduces a Coulomb potential well in the channel, which can be treated as a natural, really atomic QD. When  $V_G$  is swept gradually in the positive direction, the donor's ground state is aligned with the lead's Fermi level, and electrons can be transported from source to drain by tunneling. In particular at low temperatures, where thermally activated carriers are limited, this tunneling mechanism will allow the observation of electrical characteristics ( $I_D - V_G$  characteristics) containing a current peak. This mechanism is basically similar to the case of conventional single-electron transistors



**Fig. 13.1** (a) An ideal structure of a single-dopant transistor, with one dopant (e.g., a P donor) located between  $n$ -type source and drain leads. (b) Electron potential landscape induced by an ionized donor in the channel. The donor potential well works as a quantum dot. (c) Expected  $I_D - V_G$  characteristics, exhibiting one peak corresponding to the one-electron occupancy of a regular donor atom at usual temperatures

(SETs) with QDs artificially designed, but in *single-dopant transistors* the QD is induced by a single dopant.

Considering typical dopants for silicon, such as phosphorus (P) and arsenic (As) as donors or boron (B) as acceptor, it is known that such dopants can basically accommodate only one extra carrier (electron or hole) at usual temperatures [17, 18]. A second electron/hole may be introduced in a so-called shallow dopant, but its binding energy is so small (only a few meV) [19] that it is not expected to be observable in usual transport measurements, unless the temperature is maintained within the cryogenic range. Hence, in the  $I_D - V_G$  characteristics, a single current peak is expected to be seen, corresponding to tunneling transport via a single donor in the channel before the regular FET characteristics set in. This single peak can be considered as a basic feature to identify single-dopant transistor tunneling operation.

In realistic devices, however, it is still quite challenging to precisely position one dopant atom in the channel. Therefore, most devices investigated so far in terms of individual dopants contain dopants in the channel which either randomly diffuse from the highly-doped source and drain electrodes [10, 12, 14, 15] or are intentionally (randomly) doped with a relatively low doping concentration [11, 13]. In both situations, individual dopants can be addressed by electrical measurements and single-dopant transistors can be identified, albeit not necessarily in all devices studied. These multiple-dopant channels are the systems that will be mainly treated here.

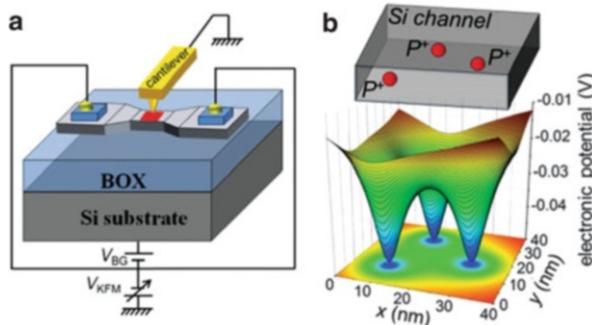
In this chapter, we focus on three main points essential to elucidate for a complete and smooth further progress of single-dopant devices. *First*, in Sect. 13.2, we address the issue of *dopant observation in devices under normal operation*. We employ a nondestructive method, low-temperature Kelvin probe force microscope (LT-KFM), which offers improved capabilities for dopant characterization in devices under normal operation conditions. *A second issue*,

described in Sect. 13.3, is the *identification of single-electron transport via individual dopant atoms* even in transistor nano-channels containing more than just one dopant. We demonstrated that, in spite of the presence of a large number of dopants, i.e., in dopant-rich environments, it is still possible to electrically address one or only a few dopant atoms, which allows some flexibility in designing single-dopant devices. Section 13.4 presents our approach to push the operation temperature of single-dopant transistors toward room temperature, for practical applications. We proposed a specific channel nano-design that allows an enhancement of the dopant's tunnel barrier and, implicitly, higher tunneling-operation temperatures. Finally, in an attempt to outline a new world of *atom-level applications starting from individual dopants*, we briefly present in Sect. 13.5 some recent results of our research on several functionalities of dopant-based devices, originating either from dopant–dopant or dopant–photon interactions.

## 13.2 Charging in Individual Dopants Observed by Low-Temperature Kelvin Probe Force Microscopy

As outlined so far, the individuality of dopants becomes more significant in the electrical characteristics of nanoscale transistors and it even gives rise to a new family of devices, *single-dopant devices*. In this background, it is crucial to observe directly the spatial distribution of ionized dopant potentials in the device channel in order to, eventually, correlate it with the electrical characteristics. Several reports on shallow-dopant profile or carrier concentration measurements by different techniques are available in literature, but they usually require special preparation of the samples and/or cannot perform accurate measurements on dopants located in the channel of devices under operation. Scanning tunneling microscopy (STM) was used to identify and characterize individual dopants in the top atomic layers from the surface [20–22]. Scanning capacitance microscopy (SCM) [23], for instance, is another interesting technique that provided information about capacitive coupling between a tip and the charges in a sample, but it does not yet have the required accuracy. Atom probe tomography [24–26], another promising technique for eventually resolving the three-dimensional distribution of dopant atoms in nano-channels, is a destructive technique and does not yet allow high-precision measurements of all dopants in the structures.

One attractive technique that could overcome most fundamental limitations is Kelvin probe force microscopy (KFM) [27], which allows, in principle, measurements with higher spatial resolution and higher sensitivity to charges located deeper in the device structure. We already demonstrated the possibility of observing individual dopants (both donors and acceptors) in the channel of devices under normal operation, by using our own designed low-temperature Kelvin probe

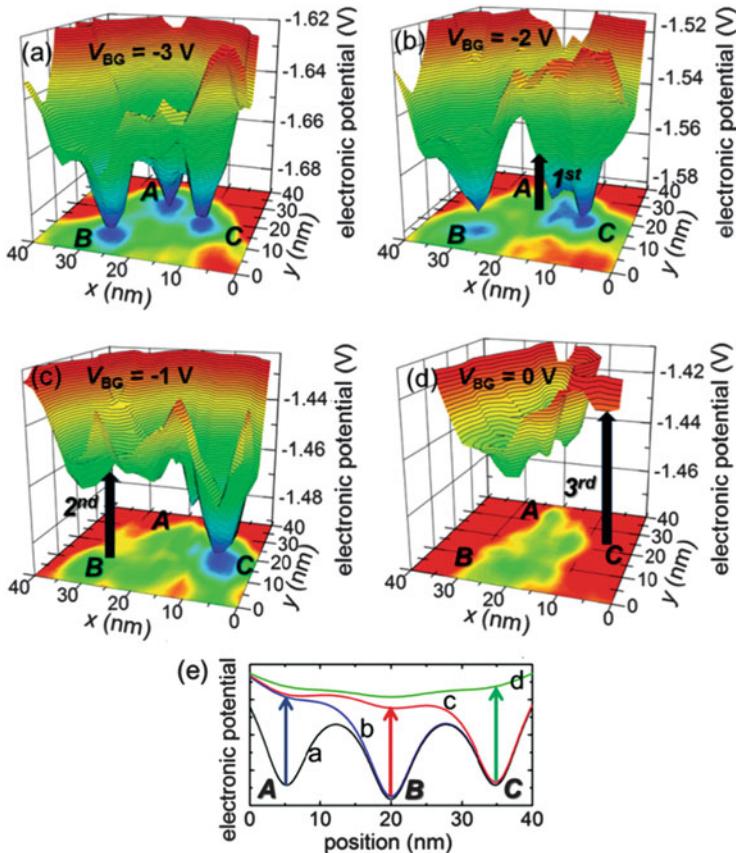


**Fig. 13.2** (a) Schematic representation of the KFM measurement setup. SOI-FET (without top gate) is biased from an external circuit as a regular transistor, while a cantilever scans over the channel surface and measures the electron potential. The device and measurement system are included in an ultrahigh vacuum chamber, and measurements can be performed at variable temperatures. (b) Calculated electronic potential landscape expected to be induced by an arrangement of a few ionized P donors in the device channel (after [31])

force microscopy technique (LT-KFM) [28–30]. Our LT-KFM technique allows KFM measurements in ultrahigh vacuum chamber, at temperatures from 13 to 300 K, with the possibility of biasing the devices with regular FET external biasing circuit [28]. More recently, we analyzed the effect of a gate voltage (substrate voltage for the KFM devices) on the charge occupancy of individual donors in a nanoscale channel [31]. By comparing results at 13 and 300 K, we could observe not only electronic potential wells induced by the phosphorus (P) donors but also distinct effects of electron injection in P donors, reflecting different electron transport mechanisms.

For the purpose of KFM measurements, we fabricated silicon-on-insulator (SOI) FETs without top gate, with the *p*-type Si substrate ( $N_A \cong 1 \times 10^{15} \text{ cm}^{-3}$ ) working as back gate, as schematically shown in Fig. 13.2a. Channel thickness is about 15 nm, with length and width of about 500 and 200 nm, respectively. Top Si layer was uniformly doped with phosphorus, in a structure similar to that of a junctionless transistor, at the concentration  $N_D \cong 1 \times 10^{18} \text{ cm}^{-3}$ , which corresponds to an average inter-donor distance of  $\sim 10$  nm. The sample was inserted in the KFM measurement chamber in ultrahigh vacuum ( $< 5 \times 10^{-7}$  Pa), and the electrodes were connected to external voltage sources. We measured the KFM surface potential images in the channel region of the device. Source and drain electrodes were grounded in this experiment, which allows us to study the static charge distribution in the channel in the absence of current flow. When several ionized P donors exist in the scan area, as shown in Fig. 13.2b, it is expected that the potential landscape will contain Coulomb wells induced by individual donors.

A KFM potential image for the back gate voltage  $V_{BG} = -3$  V, taken at 13 K, is shown in Fig. 13.3a. The applied  $V_{BG}$  works to ionize the donors and deplete electrons from the SOI channel [28–30]. Hence, the channel potential contrast is



**Fig. 13.3** (a–d) Electronic potential landscape measured by LT-KFM at low temperature ( $T = 13$  K), with back gate voltage,  $V_{BG}$ , increasing from  $-3$  to  $0$  V in  $1$  V steps. Successively, the potential wells of individual P donors vanish one by one, as also illustrated schematically in (e) (after [31])

primarily formed by the ionized P donors. In the shown area, three potential wells can be seen. Each well has a spatial extension of  $\sim 10$  nm and an electronic potential depth of  $10\text{--}40$  mV. These features suggest that each potential well is created by a different ionized P donor. In order to allow injection of electrons from grounded source and drain electrodes into the channel,  $V_{BG}$  is increased from  $-3$  to  $0$  V in  $1$  V steps. As a result, significant changes in the potential landscape can be seen. First, at  $V_{BG} = -2$  V (Fig. 13.3b), one of the potential wells, A, disappears, while, at  $V_{BG} = -1$  V, a second potential well, B, successively disappears (Fig. 13.3c). The last remaining potential well, C, disappears at  $V_{BG} = 0$  V (Fig. 13.3d). These potential changes are illustrated schematically in Fig. 13.3e as successive flattening of neighboring dopant-induced potential wells. These localized modifications are ascribed to successive single-electron filling in donors, since it is expected that each

P donor potential is almost neutralized and compensated by the capture of one electron.

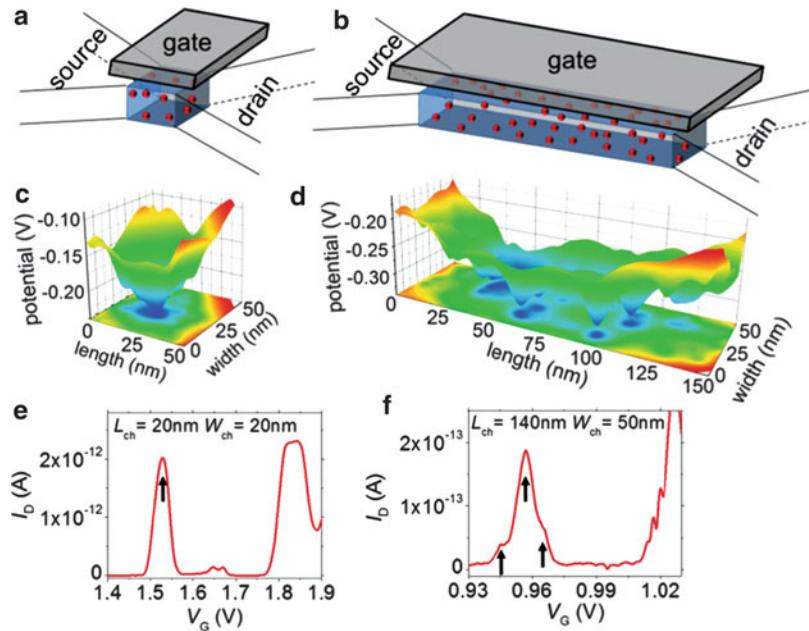
This situation is significantly different at higher temperatures [31]. For a temperature of 300 K, it was found that, regardless of the  $V_{BG}$  values, all donors are thermally activated, i.e., ionized. Therefore, spatially extended free electrons screen the dopant potentials and electron charging was observed only as delocalized potential increase by successively-injected electrons.

The results shown here illustrate the fact that LT-KFM is able to detect not only the potential of individual dopant atoms but also the electron injection, one by one, into dopants in a strongly localized manner. It also reveals the important effect of increasing temperature on the localization of the electrons at the level of individual dopants. These are the grounds on which the operation of single-dopant devices is established, since this operation consists of single-electron tunneling via individual dopant atoms, i.e., the current is formed by successive single-electron tunneling via a dopant atom. Based on this knowledge, we proceed to the electrical characterization of doped-channel nanoscale transistors with the purpose of revealing the features of tunneling transport mediated by individual dopant atoms.

### 13.3 Single-Electron Tunneling via an Individual Dopant Atom in Nanoscale Silicon Transistors

As revealed by the direct measurements using LT-KFM and presented in the previous section and our previous work [31], it is now possible to directly observe the fact that an ionized dopant atom works as a quantum dot (QD). When one dopant is coupled to electrodes in a transistor structure, electrons should be transported between source and drain via the donor-QD by successive injection into and extraction from the dopant well. Ideally, precise placement of one dopant in the channel is desirable and it has been recently demonstrated even experimentally by an STM atomic manipulation technique [16], but it remains quite complex for full CMOS compatibility. Nanostructures containing many dopant atoms can be, on the other hand, fabricated with conventional doping techniques. Under these conditions, it is essential to understand whether electron transport can be controlled by individual dopants even in such many-dopant environments.

In these structures, the overall potential landscape is not modulated by only one dopant atom but by the superposed potentials of many dopants [13, 28–30, 32]. For nano-channels doped with phosphorus (P) donors in a random process, the lowest electronic potential is most likely formed close to the channel center as a consequence of the entire set of dopants and the superposition of their long-range potentials [13]. When the channel minimum conduction band energy is shifted close to the source Fermi level by the gate voltage, transport occurs through the dopant-induced QDs. We demonstrated, by experiments and simple simulations, that even in such dopant-rich environments, individual P donors can be addressed



**Fig. 13.4** (a, b) Schematic representation of the P-doped channel of two devices, with short and, respectively, long channel length. (c, d) Channel electronic potential landscapes calculated by the superposition of randomly located P donors in the channels shown in (a) and (b), respectively. (e, f)  $I_D - V_G$  characteristics measured at low temperature ( $T = 17$  K) and for small source-drain bias ( $V_D = 10$  mV). Arrows indicate inflections (sub-peaks) seen on the first observable current peak (above the noise level) (after [16])

electrically by monitoring the electrical characteristics, i.e., the  $I_D - V_G$  characteristics, in particular at low temperatures and small source-drain biases.

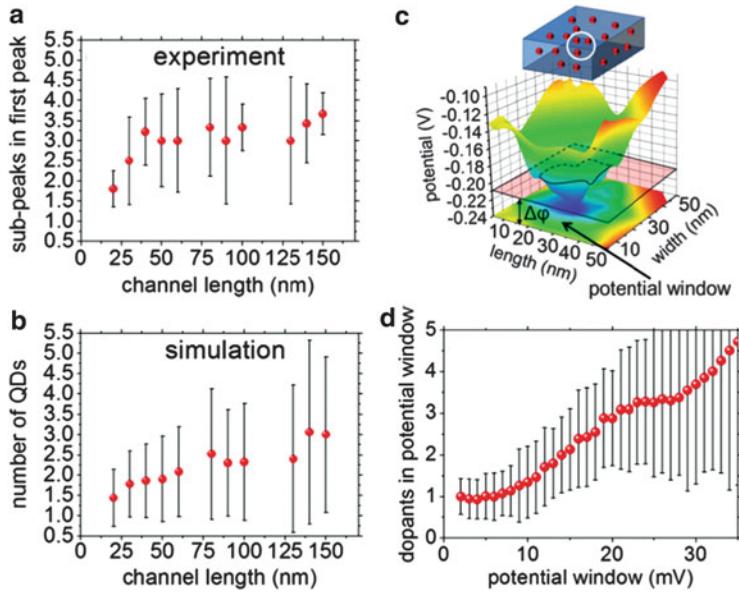
For understanding single-electron transport properties of P-doped nano-channel FETs, we fabricated devices using similar processes as for the ones used for KFM observation. However, channel constrictions of smaller lengths (20–150 nm) were defined between the two wider fan-shaped pads of Si for source and drain (as illustrated in Fig. 13.4a, b). After gate oxidation for a 10 nm-thick  $\text{SiO}_2$ , a wide Al front gate was formed on top of the nanoscale-doped channel. Doping concentration was estimated to be  $N_D \cong 1\text{--}5 \times 10^{18} \text{ cm}^{-3}$ , which means that, in the device channel, there are many dopants, not only one, that modulate the channel potential.

Figure 13.4c, d shows typical examples of dopant-induced potential landscapes simulated for random arrangements of dopant atoms with Coulomb potentials (all calculated [32] for doping concentration  $N_D = 1 \times 10^{18} \text{ cm}^{-3}$ ) in a short nanostructure (Fig. 13.4c) and a long nanostructure (Fig. 13.4d). In certain locations of the channel, potential minima will be formed and conduction starts when the channel minimum potential is shifted close to the source Fermi level by applying a positive  $V_G$  to the gate. It is expected that one dopant or an array of a few dopants

with comparable potentials will control the initial stages of transport. The  $I_D - V_G$  characteristics contain non-periodic current oscillations, which can be ascribed to transport by single-electron tunneling through dopant-induced QDs. The two lower panels in Fig. 13.4 (e, f) show the first peaks (first observable current peaks above 10 fA, when  $V_G$  is increased) of  $I_D - V_G$  characteristics measured at 17 K for FETs with constriction lengths of 20 nm and 140 nm, respectively. As a statistical tendency, while the short-channel FETs exhibit smooth single-peak (or double-split peak) current oscillations, longer-channel FETs typically exhibit multiple-split peak features (or inflections). The observed inflections are marked by arrows in the  $I_D - V_G$  characteristics. The splitting of a current peak may be treated as an indication of the formation of a multiple-QD array in the channel, while the number of sub-peaks can be considered an indication of the number of QDs [33]. These results suggest, thus, that in the shortest channels, containing a small number of donors, it is usual to observe transport mediated by single P donors. The situation becomes more complex in longer channels, in which it is likely that arrays of several P donors control the transport at the initial stages.

To provide a statistical image of our results, we counted the number of sub-peaks in the first measured current peak for 5–10 devices for each channel length ( $L_{ch}$ ) value ( $L_{ch}$  was changed in 11 steps from 20 to 150 nm). Figure 13.5a contains the statistical results of the average number of sub-peaks (dot symbols) and the standard deviation from the average (error bars), extracted from the experiments. Although the dispersion of the data is considerable, it is obvious that there is an increasing trend of the number of sub-peaks (i.e., the number of dopant-induced QDs) with  $L_{ch}$ . In order to further confirm this point, we performed simulations of dopant-induced potential landscapes for a statistical number of cases. For simplicity, we studied the behavior of isolated channels in the absence of the influence from source/drain doping. We considered nanostructures of different lengths (20–150 nm) but same width (50 nm) and thickness (10 nm). In these structures, we randomly introduced dopants with Coulomb potentials to a concentration  $N_D = 1 \times 10^{18} \text{ cm}^{-3}$ . The potential at each point is the result of the superposition of all dopant potentials. For monitoring the QD array structure, we choose a potential window of 30 mV from the bottom of the channel potential. This value is close to our approximate estimation of the charging energy from the experimental characteristics. Figure 13.5b plots the average number of QDs (dot symbols) obtained from 50 different dopant arrangements for each  $L_{ch}$  value, together with the error bars. The results indicate that the number of QDs exhibits, indeed, an increasing trend as a function of  $L_{ch}$ . The trend is in good agreement with the experimental results, which confirms that the number of dopant QDs is statistically controllable by the channel length. More detailed analysis is required for providing more definite conclusions on the quantitative trends.

Based on the simulations, it is also possible to evaluate how many P donors create the estimated QDs. For this purpose, we defined a potential window with the lower level fixed at the bottom of the potential in the channel. We gradually increased the higher level of the potential window and monitored the number of dopants and the number of QD inside the window, as illustrated in Fig. 13.5c.



**Fig. 13.5** (a) Number of sub-peaks (inflections) observed within the envelope of the first current peak (observed at lowest  $V_G$ 's above the noise level) as a function of channel length, extracted from 5 to 10 devices for each value of the channel length. (b) Number of QDs observed within a potential window of 30 mV, starting from the minimum channel potential, as a function of channel length. The data was extracted from a statistical number of simulated dopant-induced potentials for a  $50 \times 50 \text{ nm}^2$  channel area. Error bars are also indicated on both graphs. (c) An illustration of the basic procedure for evaluating the number of donors responsible for transport. (d) Number of dopants within a variable potential window, defined with the lowest level corresponding to the bottom of the channel potential

Figure 13.5d shows the number of dopants per QD statistically estimated from simulations for  $50 \times 50 \text{ nm}^2$  nanostructures, as shown above, for a doping concentration,  $N_D = 1 \times 10^{18} \text{ cm}^{-3}$ . It can be observed that, within a window of  $\sim 10 \text{ mV}$ , each QD contains on average one dopant, i.e., the observed QDs are formed by individual P donors. Above  $\sim 30 \text{ mV}$ , the average number of dopants in each QD is  $\sim 3$ . These findings indicate that, for the first observable current peaks, when the source Fermi level most likely “scans” the bottom of the channel potential, the QDs responsible for tunneling transport are mostly induced by individual donors. Hence, even in such donor-rich environments, single-electron tunneling via single P donors can be identified as the main transport mechanism, allowing us to develop and study *single-donor transistors* with the transport donors coupled (surrounded) by other donor atoms.

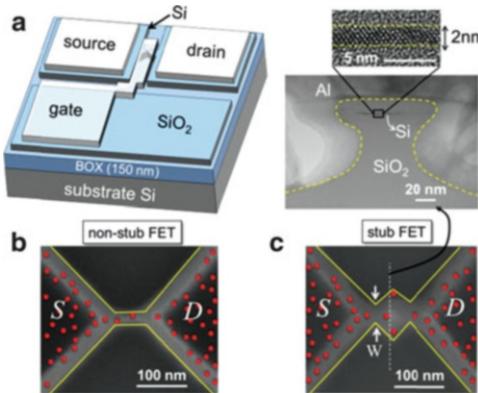
At this stage, we demonstrated that single-electron transport through single-dopant dots can be achieved even in randomly doped dopant-rich environments. This is due to the effect of the large number of dopant atoms that create a potential landscape favorable for accessing even a single dopant atom. From these preliminary

results it was also suggested that individual dopants can be identified easier in channels which are significantly isolated from source/drain electrodes [13]. This point was examined in more detail in our next work, in which we also took fully into consideration the quantum effects that should be observable in nanostructured channels.

### 13.4 Specially Patterned Single-Dopant Transistors with Tunneling Operation at Elevated Temperatures

Recent progress in silicon nanotechnology allowed electrical measurements of electron or hole tunneling through individual dopants located in the channel of silicon transistors. Results reported so far have been mostly obtained for transistors having channels without any special patterns. In these studies, dopants maintain their shallow ground states, and tunneling transport is reported only at low temperatures ( $T < 15$  K). In addition, for most reports on single-dopant devices, the final target application is quantum computing [34–36], for which low temperatures are suitable because of longer coherence time. Until recently, there were no reports focused on tunneling operation via dopants at elevated temperatures, despite the fact that, for applications toward CMOS-based electronics, higher tunneling-operation temperature is crucial. Several conditions must be met in order to be able to observe clear signatures of single-electron tunneling via dopants at high temperatures. One important requirement is for the charging energy to be significantly larger than  $k_B T$ , which is a first limiting factor for shallow dopants (charging energy is known to be on the order of several tens of meV only). Furthermore, at high temperatures, thermally activated transport, i.e., transport over the tunnel barrier of electrons found in the high-energy tail of the Fermi–Dirac distribution, may become quickly dominant. For practical applications, considering this last limiting condition, we focus on developing a channel design that would allow not only identification of single-electron tunneling transport via individual P donors but also operation at elevated temperatures [37].

In order to enhance the tunnel barrier height and, implicitly, increase the temperature up to which purely tunneling operation can be maintained, we attempted to take advantage in device design of phenomena that are specific for dopants in nanostructures. For single dopants embedded in silicon nanowires, it was recently found by simulation studies [38, 39] that dielectric confinement (and, in extremely small nanostructures, even quantum confinement) leads to a significant modification of the electronic properties of dopants as compared with the bulk Si case. In particular, it was reported that the ionization energy of dopants in nanowires is drastically increased, leading to an effective deactivation of the dopants and to a loss in conductivity, as observed also experimentally [40]. The significant impact of the dielectric confinement is due, basically, to the fact that dopants embedded in silicon nanostructures are mostly surrounded by dielectric material, and not by a large silicon matrix, as is the case for bulk. Due to this

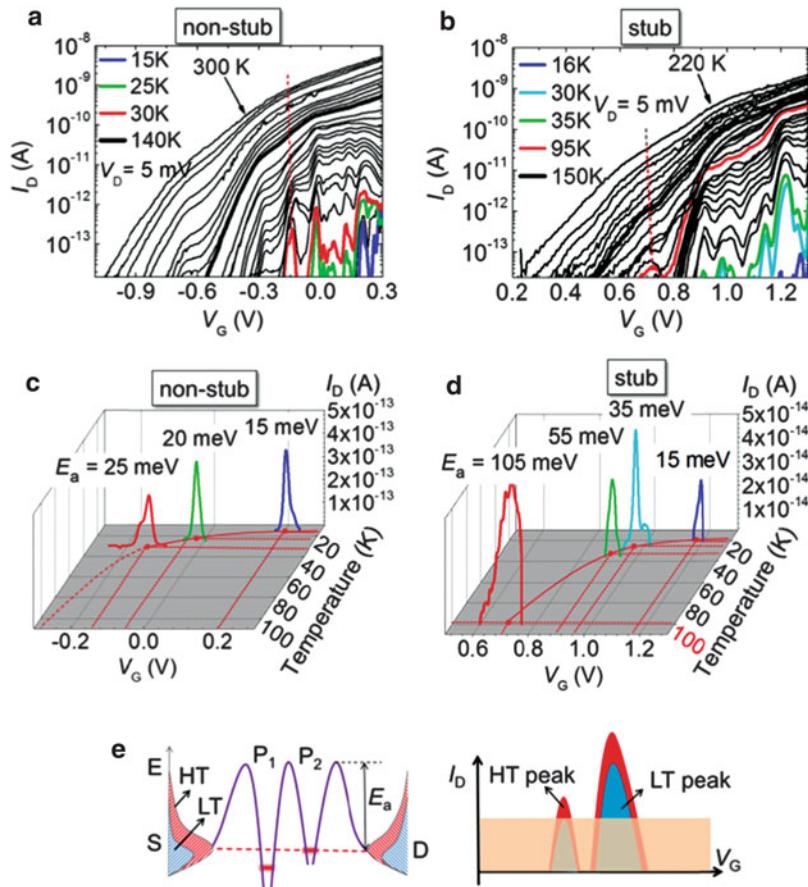


**Fig. 13.6** (a) Bird's eye view of SOI-FETs studied in this work. As shown on the right by the cross-section TEM image, top Si layer is extremely thin (only  $\sim 2$  nm). (b, c) Two types of channel patterns (non-stub-channel and stub-channel) were studied, as shown by the SEM top-view images. The boundaries of the channel region are delineated, and P donors are illustrated as red spheres according to some random distribution. In the case of stub-channel FETs, it is possible to find P donors located in the edge of the stub region, where dielectric confinement should be strongest (after [37])

configuration, screening of the dopant's extra charge by the silicon matrix is strongly reduced, leading to a modification of the dopant's ground state, i.e., a further deepening of the ground state below the conduction band edge for the case of a donor. Effectively, such enhanced ionization energy means that the donor's tunnel barrier is also enhanced, which is our main objective for the device design.

Based on the above reports and on our own *ab initio* simulations [13], we focused on specific channel design which could provide optimal conditions for the observation of dielectric confinement effect in P donors randomly introduced in the channels of nanoscale SOI-FETs (as shown in Fig. 13.6a). We found that a promising design consists of a stub-channel, as illustrated in Fig. 13.6c using a scanning electron microscope (SEM) image of one of the smallest devices; a random P donor arrangement is also illustrated for suggesting the possible favorable location of individual donors. For reference, we also fabricated devices without any special pattern of the channel, as shown in Fig. 13.6b. It should be mentioned that, in the vertical direction, donors are embedded in an ultrathin ( $\sim 2$ -nm-thick) Si layer (as seen from the cross-sectional transmission electron microscope (TEM) image as inset). For both structures, donors located in the channel should experience significant dielectric confinement effect in the vertical direction. However, only when a donor is located within the edge of the stub region, dielectric confinement effect becomes strong in the lateral direction as well, because it is mostly surrounded by SiO<sub>2</sub>, which is quite different than for the case of non-stub FETs.

For non-stub- and stub-channel FETs,  $I_D - V_G$  characteristics were measured at a small source-drain voltage,  $V_D = 5$  mV. Temperature was changed as a parameter from  $\sim 15$  to  $\sim 300$  K. Figure 13.7a, b shows representative sets of  $I_D - V_G$



**Fig. 13.7** (a, b) Temperature-dependent  $I_D - V_G$  characteristics measured for smallest devices with different channel patterns: non-stub channel [(a)] and stub channel [(b)]. (c, d) The last observable current peak (for different temperatures) as a function of both  $V_G$  and  $T$ . For stub-channel FETs, the final peak emerges at elevated temperatures ( $T = 100$  K), the highest temperature reported so far for single-dopant transistors. (e) Schematic model of the reason why higher temperature allows the observation of current peaks (after [37])

characteristics for two smallest devices with different channel patterns. At lowest temperatures ( $\sim 15$  K), the  $I_D - V_G$  characteristics exhibit a number of isolated current peaks, as seen both in Fig. 13.7a, b. As argued up to this point, these peaks are due to electron tunneling transport through donor-induced QDs formed in the channel. By raising the temperature in the range of 20–100 K, several new current peaks successively emerge at smaller  $V_G$ 's for both types of devices.

In Fig. 13.7c, d, for clarity, only the temperature-associated lowest- $V_G$  current peaks are extracted from the full  $I_D - V_G$  characteristics and are plotted in the  $V_G$ -temperature plane. These peaks successively appear with increasing temperature and are ascribed to tunneling via P donors with deeper ground-state energies.

We cannot observe the current peaks of these deep donors at low temperatures ( $\sim 15$  K), since the tunneling rate is too small due to the high potential barriers. (The detectable current level in the present system is around  $\sim 1 \times 10^{-14}$  A.) With increasing temperature, however, due to broadening of the Fermi–Dirac electron distribution in the reservoir and other thermal effects [41], as illustrated in Fig. 13.7e, the tunneling rate is enhanced and current peaks successively emerge, exceeding the detectable current level. It is found that, for the stub-channel FET, the last emerging current peak appears at  $T \cong 100$  K (Fig. 13.7d), which is the highest temperature reported so far for single-dopant transistors operating in single-electron tunneling (SET) mode. Moreover, the SET feature survives as a prominent hump up to  $\sim 150$  K, as shown by the thick  $I_D - V_G$  curve in Fig. 13.7b. At high temperatures, above  $\sim 100$  K, a number of SET peaks are significantly broadened and overlap each other, indicating a change from the Coulomb blockade mechanism to the resonant tunneling mechanism, primarily due to the reduction of the tunnel resistance.

From the temperature dependence of the electrical characteristics, such as data shown in Fig. 13.7a, b, it is possible to extract the barrier height at different  $V_G$ 's based on the analysis of Arrhenius plots. In particular, these values are most reliable as extracted from the data taken at highest temperatures, where it is certain that the thermally activated transport is the dominant transport mechanism [10]. Details of this extraction procedure and the results as a function of device dimensions can be found in our work [37]. Here, for simplicity, the values extracted for the barrier height for  $V_G$ 's corresponding to the last observable current peaks are indicated in Fig. 13.7c, d. It can be seen that, for the peak emerging at the highest temperature of  $\sim 100$  K, the barrier height is large ( $\sim 100$  meV), much larger than the value known for the ionization energy for P donors in bulk Si ( $\sim 44$  meV). These findings strongly suggest that, in stub-channel FETs, P donors, most likely located in the edge of the stub region, are exposed to a strong dielectric confinement which, as predicted by theory, leads to the detectable enhancement of their ionization energy (barrier height).

The findings described above and insights provided by *ab initio* simulations [37] suggest that the nano-channel's specific design is the critical factor in enhancing the tunneling operation temperature of single-dopant devices toward room temperature. If such operation can be realized by further optimization of the channel design or another innovative technique, single-dopant devices can be considered in the future for more practical applications. In this sense, it is necessary to investigate and clarify what kind of applications are suitable or achievable using dopant atoms in silicon nanodevices. The preliminary research that we carried out on this wide topic [42] and more recent progress will be briefly outlined in the following section.

### 13.5 Extended Applications of Dopant-Based Devices

In the previous sections of this chapter, we focused mainly on the characterization of individual dopants, either by KFM observations or by electrical characteristics. However, it is natural to expect that more complex functionality can be obtained

when dealing with systems of more than one dopant, because of the interactions between dopants, as well as interactions with external factors, such as absorbed photons and phonons. In order to propose single-dopant or multiple-dopant devices as viable candidates for future generations of electronics, it is essential to study possible functions involving a few dopants in silicon nanostructures.

In recent years, we addressed several topics dealing with functionalities arising from multiple-dopant systems and dopant–photon interactions [42]. However, it is beyond the scope of this chapter to enter deeply into the details of each topic, and more information can be found in our relevant publications. Here, we would like to offer only an overview of the wide range of possibilities that this field can provide, not only in terms of innovative applications, built from the fundamental level of atoms, but also regarding new physics possible to be revealed by utilizing these systems.

A single dopant, either a donor or an acceptor, can be indeed the building block of an atomic-level transistor. As we presented earlier, it is important to make efforts in research toward improving their fabrication and their operation so that they can be realized in a reproducible, reliable, and efficient manner. But, if in the vicinity of such a transport donor, for instance, another donor can be found, then we have the core unit for a dopant-atom-based memory device. The satellite donor atom can store an electron as a single-electron memory node and its charge state can be detected using the single-electron tunneling current flowing through the transport donor. We demonstrated such operation by carefully examining the charging effects detectable in nanoscale transistors containing more than one donor [43]. We found that the interaction of donors with nearby Si/SiO<sub>2</sub> interface plays also a key role in the transfer of one electron between donors. This kind of operation basically takes us one step further, now into atomic scale, following previous demonstrations of single-electron memories using semiconductor QDs [44, 45].

A similar observation of single-electron trapping and detrapping in a dopant atom was done under light illumination, in devices with basically the same structure, but only lacking the metallic top gate. We found that, by the absorption of individual visible-light photons into an ultrathin SOI channel and subsequent generation of an electron-hole pair, the elementary charges, separated in the electric field, can be captured by the ionized donors (in an *n*-doped channel). The electron is released after some time, leading to the observation of random telegraph signals (RTS) that depend strongly on light intensity [46]. This is a fundamental demonstration of single-photon detection using an array of QDs induced by individual donor atoms [47], similar in principle to previous works on photon detectors using semiconductor QDs [48, 49]. The impact of photogenerated carriers can also be detected in nanoscale *pn* junctions, where both donors and acceptors interact in creating appropriate conditions for the capture of elementary charges in individual dopant atoms [50]. In such nanoscale *pn* junctions, we also found that the dopant properties drastically change, most likely as a result of significant dielectric confinement and interdiffusion of donors and acceptors. This was understood from direct measurements of LT-KFM on nanoscale *pn* junctions, which revealed special properties of dopants in the depletion layer [51].

By considering systems with three or more dopants, e.g., P donors, we demonstrated also single-electron turnstile operation, both by experiments [52, 53] and by simulations [54, 55]. Such single-electron turnstile function was originally proposed using metallic QDs [56], then demonstrated also with semiconductor QDs [57]. These devices were considered attractive because of their ability to control electron transfer not only in space, through a specific array of QDs, but also in time, thus overcoming, to some extent, the stochastic nature of the single-electron tunneling phenomenon. Once realized with donor-atom arrays, it becomes possible to design and implement dopant-atom switches which could work as fundamental building blocks for the integrated circuits of future atomic-level electronics.

It is, thus, possible to design and couple a wide variety of applications starting from the fundamental structure of a dopant atom working as a QD, then involving interaction with neighboring dopants and with individual photons. All the functionalities that could be extracted from the analysis of single-dopant or multiple-dopant transistors will definitely provide the solid foundation necessary for building up the field of *dopant-atom-based electronics*. Exciting physics phenomena at atomic scale in these silicon nanodevices have been already revealed, while others are still expecting to be revealed, making this field a rapidly growing one, with chances to become a viable candidate for future generations of electronics.

## 13.6 Conclusions

In this chapter, we introduced our basic approach for moving beyond the present limitations that the device miniaturization trend is facing. We suggest that atomic-level electronics can be taken into consideration as an important candidate because it still builds upon the well-developed platform created for silicon nanodevices but focuses on a significantly different transport mechanism as compared to conventional devices: *single-electron tunneling via individual dopant atoms*. Based on this concept, a new family of electronic devices, *single-dopant devices*, can be developed and used to study fundamental physics related to the interaction between elementary charge carriers, single dopant atoms, and single photons.

After briefly outlining the essential points of this conceptual change, we present the direct observation of discrete dopants in operating devices using low-temperature Kelvin probe force microscopy. Then, our research on single-dopant transistors was introduced, starting from our demonstration of tunneling transport via an individual donor even in dopant-rich environments. Channel design can be considered for obtaining improved properties of dopants in nanostructured channels, properties completely different than in bulk silicon and which are promising for allowing eventually room-temperature operation of single-dopant devices. Finally, a range of applications can be designed, and we took a few basic steps by preliminary demonstrations of the feasibility of several dopant-based devices. The findings

included in this chapter, together with a rich and fast-paced development of the field in recent literature, guarantee an exciting future for atom-level electronics, with a smooth transition from present-day silicon-based electronics entering deeper and deeper into nanoscale.

**Acknowledgments** The authors are grateful for the collaboration and useful discussions with Prof. H. Mizuta, Prof. Y. Ono, Prof. T. Shinada, Prof. R. Jablonski, and Prof. D. Hartanto. Important contributions on various topics discussed in this chapter have been brought by many coworkers over the past few years, and we deeply appreciate them: Mr. T. Mizuno, Dr. M. Ligowski, Dr. E. Hamid, Dr. A. Udharto, Dr. M. Anwar, Dr. R. Nowak, Dr. S. Purwiyanti, and all the past and present members of our research group. This work was partially supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan (23226009, 22310085, and 25630144).

## References

1. Moore, G.: Cramming more components onto integrated circuits. *Electronics* **38**, 114–117 (1965)
2. Asenov, A.: Random dopant induced threshold voltage lowering and fluctuations in sub-0.1  $\mu\text{m}$  MOSFET's: A 3-D “atomistic” simulation study. *IEEE Trans. Electron Dev.* **45**, 2505–2513 (1998)
3. Mizuno, T., Okamura, J., Toriumi, A.: Experimental study of threshold voltage fluctuation due to statistical variation of channel dopant number in MOSFET's. *IEEE Trans. Electron Dev.* **41**, 2216–2221 (1994)
4. Ohtou, T., Sugii, N., Hiramoto, T.: Impact of parameter variations and random dopant fluctuations on short-channel fully depleted SOI MOSFETs with extremely thin BOX. *IEEE Electron Device Lett.* **28**, 740–742 (2007)
5. Pierre, M., Wacquez, R., Jehl, X., Sanquer, M., Vinet, M., Cueto, O.: Single-donor ionization energies in a nanoscale CMOS channel. *Nat. Nanotechnol.* **5**, 133–137 (2010)
6. Shinada, T., Okamoto, S., Kobayashi, T., Ohdomari, I.: Enhancing semiconductor device performance using ordered dopant arrays. *Nature* **437**, 1128–1131 (2005)
7. Li, Y., Yu, S.-M.: Comparison of random-dopant-induced threshold fluctuation in nanoscale single-, double-, and surrounding-gate field-effect transistors. *Jpn. J. Appl. Phys.* **45**, 6860–6865 (2006)
8. Hisamoto, D., Lee, W.-C., Kedzierski, J., Takeuchi, H., Asano, K., Kuo, C., Anderson, E., King, T.-J., Bokor, J., Hu, C.: FinFET—a self-aligned double-gate MOSFET scalable to 20 nm. *IEEE Trans. Electron Dev.* **47**, 2320–2325 (2000)
9. Wong, H.-S.P.: Beyond the conventional transistor. *IBM J. Res. Dev.* **46**, 133–168 (2002)
10. Sellier, H., Lansbergen, G.P., Caro, J., Collaert, N., Ferain, I., Jurczak, M., Biesemans, S., Rogge, S.: Transport spectroscopy of a single dopant in a gated silicon nanowire. *Phys. Rev. Lett.* **97**, 206805 (2006)
11. Ono, Y., Nishiguchi, K., Fujiwara, A., Yamaguchi, H., Inokawa, H., Takahashi, Y.: Conductance modulation by individual acceptors in Si nanoscale field-effect transistors. *Appl. Phys. Lett.* **90**, 102106 (2007)
12. Lansbergen, G.P., Rahman, R., Wellard, C.J., Woo, I., Caro, J., Collaert, N., Biesemans, S., Klimeck, G., Hollenberg, L.C.L., Rogge, S.: Gate-induced quantum-confinement transition of a single dopant atom in a silicon FinFET. *Nat. Phys.* **4**, 656–661 (2008)
13. Tabe, M., Moraru, D., Ligowski, M., Anwar, M., Jablonski, R., Ono, Y., Mizuno, T.: Single-electron transport through single dopants in a dopant-rich environment. *Phys. Rev. Lett.* **105**, 016803 (2010)

14. Tan, K.Y., Chan, K.W., Möttönen, M., Morello, A., Yang, C., van Donkelaar, J., Alves, A., Pirkkalainen, J.-M., Jamieson, D.N., Clark, R.G., Dzurak, A.S.: Transport spectroscopy of single phosphorus donors in a silicon nanoscale transistor. *Nano Lett.* **10**, 11–15 (2010)
15. Prati, E., Belli, M., Cocco, S., Petretto, G., Fanciulli, M.: Adiabatic charge control in a single donor atom transistor. *Appl. Phys. Lett.* **98**, 053109 (2011)
16. Fuechsle, M., Miwa, J.A., Mahapatra, S., Ryu, H., Lee, S., Warschkow, O., Hollenberg, L.C. L., Klimeck, G., Simmons, M.Y.: A single-atom transistor. *Nat. Nanotechnol.* **7**, 242–246 (2012)
17. Kohn, W., Luttinger, J.M.: Theory of donor states in silicon. *Phys. Rev.* **98**, 915–922 (1955)
18. Pantelidis, S., Sah, C.T.: Theory of impurity states in semiconductors. *Solid State Commun.* **11**, 1713–1718 (1972)
19. Taniguchi, M., Narita, S.:  $D^-$  state in silicon. *Solid State Commun.* **20**, 131–133 (1976)
20. Jäger, N.D., Urban, K., Weber, E.R., Ebert, P.: Nanoscale dopant-induced dots and potential fluctuations in GaAs. *Appl. Phys. Lett.* **82**, 2700–2702 (2003)
21. Schofield, S.R., Curson, N.J., Simmons, M.Y., Ruess, F.J., Hallam, T., Oberbeck, L., Clark, R. G.: Atomically precise placement of single dopants in Si. *Phys. Rev. Lett.* **91**, 136104 (2003)
22. Nishizawa, M., Bolotov, L., Kanayama, T.: Simultaneous measurement of potential and dopant atom distributions on wet-prepared Si(111):H surfaces by scanning tunneling microscopy. *Appl. Phys. Lett.* **90**, 122118 (2007)
23. Goragot, W., Takai, M.: Measurement of shallow dopant profile using scanning capacitance microscopy. *Jpn. J. Appl. Phys.* **43**, 3990–3994 (2004)
24. Thomson, K., Booske, J.H., Larson, D.J., Kelly, T.F.: Three-dimensional atom mapping of dopants in Si nanostructures. *Appl. Phys. Lett.* **87**, 052108-1-3 (2005)
25. Inoue, K., Yano, F., Nishida, A., Takamizawa, H., Tsunomura, T., Nagai, Y., Hasegawa, M.: Dopant distributions in n-MOSFET structure observed by atom probe tomography. *Ultramicroscopy* **109**, 1479–1484 (2009)
26. Perea, D.E., Hemesath, E.R., Schwalbach, E.J., Lensch-Falk, J.L., Voorhees, P.W., Lauhon, L. J.: Direct measurement of dopant distribution in an individual vapour-liquid-solid nanowire. *Nat. Nanotechnol.* **4**, 315–319 (2009)
27. Nonnenmacher, M., O’Boyle, M.P., Wickramasinghe, H.K.: Kelvin probe force microscopy. *Appl. Phys. Lett.* **58**, 2921–2923 (1991)
28. Ligowski, M., Moraru, D., Anwar, M., Mizuno, T., Jablonski, R., Tabe, M.: Observation of individual dopants in a thin silicon layer by low temperature Kelvin probe force microscope. *Appl. Phys. Lett.* **93**, 142101-1-3 (2008)
29. Tabe, M., Moraru, D., Ligowski, M., Anwar, M., Yokoi, K., Jablonski, R., Mizuno, T.: Observation of discrete dopant potential and its application to Si single-electron devices. *Thin Solid Films* **518**, S38–S43 (2010)
30. Anwar, M., Kawai, Y., Moraru, D., Nowak, R., Jablonski, R., Mizuno, T., Tabe, M.: Single-electron charging in phosphorus donors in Si observed by low-temperature Kelvin probe force microscope. *Jpn. J. Appl. Phys.* **50**, 08LB10-1-5 (2011)
31. Anwar, M., Nowak, R., Moraru, D., Udharto, A., Mizuno, T., Jablonski, R., Tabe, M.: Effect of electron injection into phosphorus donors in silicon-on-insulator channel observed by Kelvin probe force microscopy. *Appl. Phys. Lett.* **99**, 213101-1-3 (2011)
32. Evans, G.J., Mizuta, H., Ahmed, H.: Modelling of structural and threshold voltage characteristics of randomly doped silicon nanowires in the Coulomb-blockade regime. *Jpn. J. Appl. Phys.* **40**, 5837–5840 (2001)
33. Waugh, F.R., Berry, M.J., Mar, D.J., Westervelt, R.M., Campman, K.L., Gossard, A.C.: Single-electron charging in double and triple quantum dots with tunable coupling. *Phys. Rev. Lett.* **75**, 705–708 (1995)
34. Kane, B.: A silicon-based nuclear spin quantum computer. *Nature* **393**, 133–137 (1998)
35. Loss, D., DiVincenzo, D.P.: Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998)

36. Hollenberg, L.C.L., Dzurak, A.S., Wellard, C., Hamilton, A.R., Reilly, D.J., Milburn, G.J., Clark, R.G.: Charge-based quantum computing using single donors in semiconductors. *Phys. Rev. B* **69**, 113301-1-4 (2004)
37. Hamid, E., Moraru, D., Kuzuya, Y., Mizuno, T., Anh, L.T., Mizuta, H., Tabe, M.: Electron-tunneling operation of single-donor-atom transistors at elevated temperatures. *Phys. Rev. B* **87**, 085420-1-5 (2013)
38. Diarra, M., Niquet, Y.-M., Delerue, C., Allan, G.: Ionization energy of donor and acceptor impurities in semiconductor nanowires: importance of dielectric confinement. *Phys. Rev. B* **75**, 045301 (2007)
39. Li, B., Slachmuylders, A.F., Partoens, B., Magnus, W., Peeters, F.M.: Dielectric mismatch effect on shallow impurity states in a semiconductor nanowire. *Phys. Rev. B* **77**, 115335-1-10 (2008)
40. Björk, M.T., Schmid, H., Knoch, J., Riel, H., Riess, W.: Donor deactivation in silicon nanostructures. *Nat. Nanotechnol.* **4**, 103–107 (2009)
41. Morgan, N.Y., Abusch-Magder, D., Kastner, M.A., Takahashi, Y., Tamura, H., Murase, K.: Evidence for activated conduction in a single electron transistor. *J. Appl. Phys.* **89**, 410–419 (2001)
42. Moraru, D., Udhiaarto, A., Anwar, M., Nowak, R., Jablonski, R., Hamid, E., Tarido, J.C., Mizuno, T., Tabe, M.: Atom devices based on single dopants in silicon nanostructures. *Nanoscale Res. Lett.* **6**, 479-1-6 (2011)
43. Hamid, E., Moraru, D., Tarido, J.C., Miki, S., Mizuno, T., Tabe, M.: Single-electron transfer between two donors in nanoscale thin silicon-on-insulator field-effect transistors. *Appl. Phys. Lett.* **97**, 262101 (2010)
44. Yano, K., Ishii, T., Hashimoto, T., Kobayashi, T., Murai, F., Seki, K.: Room-temperature single-electron memory. *IEEE Trans. Electron Dev.* **41**, 1628–1638 (1994)
45. Fujiwara, A., Takahashi, Y., Murase, K., Tabe, M.: Time-resolved measurement of single-electron tunneling in a Si single-electron transistor with satellite Si islands. *Appl. Phys. Lett.* **67**, 2957–2959 (1995)
46. Udhiaarto, A., Moraru, D., Mizuno, T., Tabe, M.: Trapping of a photoexcited electron by a donor in nanometer-scale phosphorus-doped silicon-on-insulator field-effect transistors. *Appl. Phys. Lett.* **99**, 113108-1-3 (2011)
47. Tabe, M., Udhiaarto, A., Moraru, D., Mizuno, T.: Single-photon detection by Si single-electron FETs. *Phys. Status Solidi A* **208**, 646–651 (2011)
48. Shields, A.J., O'Sullivan, M.P., Farrer, I., Ritchie, D.A., Hogg, R.A., Leadbeater, M.L., Norman, C.E., Pepper, M.: Detection of single photons using a field-effect transistor gated by a layer of quantum dots. *Appl. Phys. Lett.* **76**, 3673–3675 (2000)
49. Nuryadi, R., Ishikawa, Y., Tabe, M.: Single-photon-induced random telegraph signal in a two-dimensional multiple-tunnel-junction array. *Phys. Rev. B* **73**, 045310-1-5 (2006)
50. Udhiaarto, A., Moraru, D., Purwiyanti, S., Kuzuya, Y., Mizuno, T., Mizuta, H., Tabe, M.: Photon-induced random telegraph signal due to potential fluctuation of a single donor-acceptor pair in nanoscale Si p-n junctions. *Appl. Phys. Express* **5**, 112201-1-3 (2012)
51. Nowak, R., Moraru, D., Mizuno, T., Jablonski, R., Tabe, M.: Effects of deep-level dopants on the electronic potential of thin Si pn junctions observed by Kelvin probe force microscope. *Appl. Phys. Lett.* **102**, 083109-1-4 (2013)
52. Moraru, D., Ono, Y., Inokawa, H., Tabe, M.: Quantized electron transfer through random multiple tunnel junctions in phosphorus-doped silicon nanowires. *Phys. Rev. B* **76**, 075332 (2007)
53. Moraru, D., Ligowski, M., Yokoi, K., Mizuno, T., Tabe, M.: Single-electron transfer by inter-dopant coupling tuning in doped nanowire silicon-on-insulator field-effect transistors. *Appl. Phys. Express* **2**, 071201 (2009)
54. Yokoi, K., Moraru, D., Ligowski, M., Tabe, M.: Single-gated single-electron transfer in nonuniform arrays of quantum dots. *Jpn. J. Appl. Phys.* **48**, 024503 (2009)

55. Yokoi, K., Moraru, D., Mizuno, T., Tabe, M.: Electrical control of capacitance dispersion for single-electron turnstile operation in common-gated junction arrays. *J. Appl. Phys.* **108**, 053710 (2010)
56. Geerligs, L.J., Anderegg, V.F., Holweg, P.A.M., Mooij, J.E., Pothier, H., Esteve, D., Urbina, C., Devoret, M.H.: Frequency-locked turnstile device for single electrons. *Phys. Rev. Lett.* **64**, 2691–2694 (1990)
57. Ono, Y., Zimmerman, N.M., Yamazaki, K., Takahashi, Y.: Turnstile operation using a silicon dual-gate single-electron transistor. *Jpn. J. Appl. Phys.* **42**, L1109–L1111 (2003)

# Chapter 14

## Single-Electron Transistor and Quantum Dots on Graphene

Lin-Jun Wang, Tao Tu, Li Wang, Cheng Zhou, and Guo-Ping Guo

**Abstract** Graphene has been proclaimed to be a new revolutionary material for electronics. In particular, graphene-based transistors have developed rapidly and are now considered an option for post-silicon electronics. Quantum FinFET and quantum computation are keeping on attracting scientists' research interest since it came out. It reveals a cornucopia of new physics and potential applications and has always maintained a very active research environment; many striking research efforts and advanced technologies emerge. New ideas and issues are continuously proposed.

Spin-based semiconductor quantum dot for the solid-state quantum information process has been considered as a very promising direction. A lot of excellent efforts have been made in this area, but there are also many difficulties to deal with, such as how to extend the spin coherence time. Scientists have tried many methods to solve this problem: one is to use new material, such as graphene to substitute currently used traditional gallium arsenide semiconductor material. Research on nanoscale transistors switching with only a single electron exemplifies that there are a number of unresolved problems that material scientists should tackle in the future for making the graphene dreams come true.

In this chapter, we will talk about all kind of graphene-based quantum dot devices, including single dot, single dot with integrated single-electron transistor

---

L.-J. Wang (✉)

Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China  
e-mail: [wanglj@ustc.edu.cn](mailto:wanglj@ustc.edu.cn)

T. Tu • L. Wang • C. Zhou • G.-P. Guo

Key Lab of Quantum Information, Chinese Academy of Science, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China  
e-mail: [tutao@ustc.edu.cn](mailto:tutao@ustc.edu.cn); [waly sclw@mail.ustc.edu.cn](mailto:waly sclw@mail.ustc.edu.cn); [kfrank@mail.ustc.edu.cn](mailto:kfrank@mail.ustc.edu.cn); [gpguo@ustc.edu.cn](mailto:gpguo@ustc.edu.cn)

(SET) charge detector, double dot in series, and double dot in parallel. We investigate the properties of devices by doing the low-temperature quantum transport measurements. Detailed descriptions on the fabrication methods of graphene quantum dot devices are described. And also the information of the ground states and excited states and the relevant energy scales and capacitances of the graphene quantum dot are investigated, as denoted by the presence of characteristic Coulomb blockade diamond diagrams. A twin-dot structure in which the larger dot serves as a single-electron transistor (SET) to read out the charge state of the nearby gate-controlled small dot has been fabricated. A high SET sensitivity of  $10^{-3}e/\sqrt{\text{Hz}}$  allowed us to probe Coulomb charging as well as excited state spectra of the QD, even in the regime where the current through the QD is too small to be measured by conventional transport means. Graphene double quantum dot devices with multiple electrostatic gates are investigated by low-temperature transport measurements; the honeycomb charge stability diagrams reveal the interdot coupling strength changed from weak to strong regime by tuning both the in-plane plunger gates and back gate. A large interdot tunnel coupling strength for this system allows for the observation of tunnel-coupled molecular states extending over the whole double dot.

## 14.1 Introduction

Graphene, a single layer of carbon atoms densely packed in a honeycomb crystal lattice, forming a perfectly stable and clean two-dimensional crystal with very few defects, is a rapidly rising star in the field of materials science and condensed matter physics. It exhibits exceptionally high crystal and electronic quality and has already revealed a cornucopia of new physics and potential applications. Owing to its unusual linear electronic spectrum  $E = v \cdot k$  at low energies  $E$ , its charge carriers mimic relativistic particles and are described with the Dirac equation rather than the Schrödinger equation; it has led to the emergence of a new paradigm of “relativistic” condensed matter physics, where quantum relativistic phenomena, some of which are unobservable in high-energy physics, can now be mimicked and tested in tabletop experiments. More generally, graphene is 2D material only one atom thick; on this basis, it offers new routes into low-dimensional physics that has never ceased to surprise and continues to provide an abundant ground for applications [1–4].

As time goes by, graphene has changed from being the exclusive domain of condensed matter physicists to being explored by those in the electron-device community. In particular, graphene-based transistors have developed rapidly and are now considered an option for post-silicon electronics. Mainly resting on the unique band structure properties, graphene has been proclaimed to be a new revolutionary material for electronics. The potential advantage of graphene may be the possibility of making devices with channels that are extremely thin that will allow graphene field-effect transistors to be scaled to shorter channel lengths and higher speeds without encountering the adverse short channel effects that restrict

the performance of existing devices. Meanwhile electron mobility in this material does not suffer extensively from surface contaminations and is surprisingly high even at room temperature. But the excellent mobility of graphene may not, as is often assumed, be its most compelling feature from a device perspective, rather the extremely thin channel that will allow graphene field-effect transistors to be more practically useful [5, 6].

In comparison to extremely high-quality semiconducting materials, such as silicon and GaAs, the understanding of electronic transport in graphene is still in its infancy. However, many details about the potential performance of graphene transistors in real applications remain unclear. Research on nanoscale transistors switching with only a single electron exemplifies that there are a number of unresolved problems that material scientists should tackle in the future for making the graphene dreams come true. There are a lot of outstanding challenges for graphene transistors including opening a sizeable and well-defined bandgap in graphene and fabricating graphene nanoribbons with well-defined widths and clean edges [5].

During the past few years significant progresses have been achieved in the implementation of electron spin qubits in semiconductor quantum dots [7–9]. For future nanoscale electronics like the realm of quantum information processing, graphene could be the basis material for the realization of spin qubits with very long coherence times. In order for quantum computation to become a reality, the effects of the interactions of qubits with their environment must be minimized [10]. Because of the weak spin-orbit coupling and largely eliminated hyperfine interaction in graphene, it is highly desirable to coherently control the spin degree of freedom in graphene nanostructures [11]. Demonstration of graphene quantum dot behavior is an important and first step toward realizing such promise. Graphene quantum dot sometimes is also called single-electron transistor because of the one-by-one electron transport behavior through the dot.

## 14.2 Graphene Single-Electron Transistor

Graphene-based quantum dot systems are considered to be a promised candidate for future quantum computer and quantum information science. The low-energy quasi-particles in single-layer graphene behave as Dirac fermions [12], and the Klein tunneling effect leads to the fact that it is important to form a quantum dot (electron/hole-bound states) in graphene. There are several theoretical proposals for trapping the massless Dirac fermions in graphene, such as using suitable transverse states in graphene ribbons [11, 13], using the presence of a bandgap in bilayer graphene by applying an electric field [14, 15], or using inhomogeneous magnetic fields [16]. However, at present, only one way, etching a central island of about nanometers in size to form a graphene quantum dot, has been experimentally realized [17, 18].

Here we first talk about graphene single quantum dot, also called graphene single-electron transistor; at large sizes ( $>100$  nm), they behave as conventional single-electron transistors, exhibiting periodic Coulomb blockade peaks. In order to investigate graphene quantum dot, we need to investigate both ground state and excited states. Nevertheless, the obvious identification of the excited states in a graphene quantum dot, which has been well developed to deduce the spin filling in conventional GaAs quantum dots, has so far remained on the list aiming at quantum information processing.

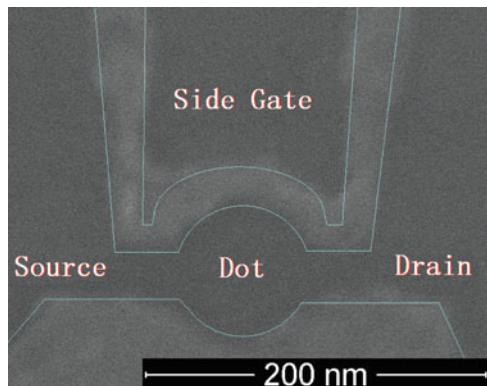
*Sample fabrication.* The fabrication of graphene nanostructures has so far been mainly based on the scotch tape technique followed by plasma etching, which has been mostly developed by Geim and Novoselov in Manchester [1, 19] and by Kim and coworkers at Columbia University [20]. Other techniques include unzipping of carbon nanotubes [21, 22] or local anodic oxidation with a scanning force microscope [23]. Here we discuss the basic process steps of the standard process.

The graphene flakes were produced by mechanical cleaving of nature graphite with adhesive tape, and the tape is pressed onto a preprocessed oxidized Si wafer; the thickness of the  $\text{SiO}_2$  layer is 100 nm. Thin flakes were found by optical microscopy and single-layer graphene flakes were selected by Raman spectroscopy measurement. The quality and cleanliness can be further inspected by atomic force microscopy (SFM). With conventional electron-beam lithography using PMMA mask and oxygen plasma etching, we carved single-layer graphene to a configured pattern. The flakes are usually contacted by deposition of 2–5 nm chromium (or titanium) and 50 nm gold on the electron-beam structured mask. The evaporation is followed by a subsequent lift-off of the metal-covered PMMA mask in hot acetone.

Figure 14.1 shows one of the defined quantum dots of 90 nm in diameter, connected by 30-nm-wide constrictions to source and drain contacts. The magnitude of the current through the dot depends on the tunneling rate between the dot and the reservoir on the source  $\Gamma_s$  and on the drain  $\Gamma_d$ . Here, Si wafer was used as the back gate and there is also a side gate nearby. There are two relevant physical mechanisms to form electron/hole-bound states in the single-layer graphene quantum dot. First, the designed quantum dot is connected to the source and drain with two short and narrow constrictions (see Fig. 14.1), which are used as a quantum point contact and provide quantum barriers to confine electrons/holes in the quantum dot. Second, the applied gate voltage can give rise to an additional gap in the sample; therefore electron/hole-bound states can more easily exist in such nanostructures in the presence of an electrostatic confinement potential.

*Transport properties of the device in low temperature.* The device usually was first immersed into a liquid helium storage dewar at 4.2 K to test the functionality of the gates. Then it can be placed to a top-loading dilution refrigerator equipped with filtered wiring and low-noise electronics at the base temperature of 10 mK. Standard ac lock-in technique with an excitation voltage of 20  $\mu\text{V}$  at 11.3 Hz is employed in the measurement.

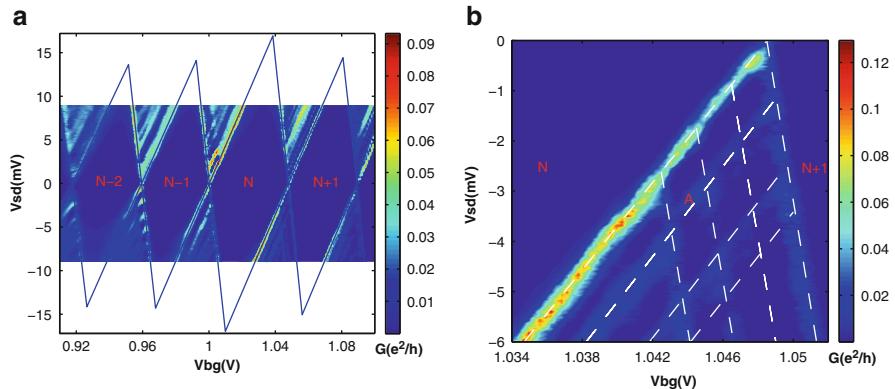
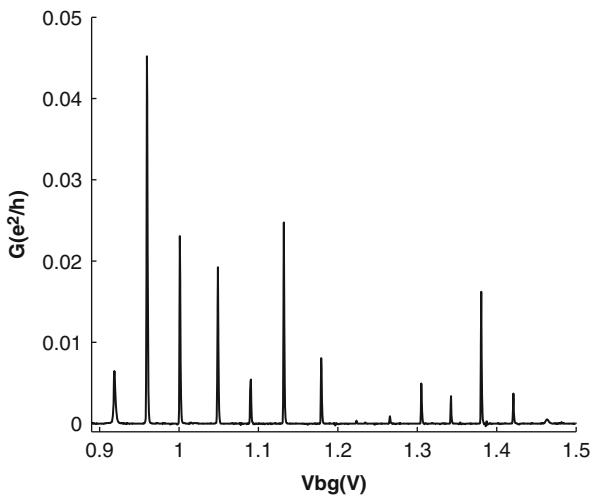
**Fig. 14.1** Scanning electron micrograph (SEM) of an etched gate tunable quantum dot on single-layer graphene. The dot has a diameter of 90 nm and is connected by 30-nm-wide constrictions to the source and drain contacts



Characterization of the quantum dot in single-layer graphene yields the low-temperature transport properties, which are dominated by the standard Coulomb Blockade (CB) oscillations over a wide range of back gate voltages, as denoted by the presence of characteristic quasi-periodic peaks in the linear differential conductance  $G = dI/dV$  (shown in Fig. 14.2). The response of the linear differential conductance change to the side gate has a similar effect to back gate  $V_g$  [24]. For a quantum dot system in equilibrium, single-electron tunneling is only possible when there is a energy level corresponding to the ground states of the dot, which is tuned by the gate voltage, within the source drain bias window set by the source and drain electrochemical potentials, and the charges in the quantum dot can change between  $N$  and  $N + 1$ . However, if no level in the dot is in the bias window, the electron number is fixed at  $N$ . By sweeping the back gate voltage and measuring the current flow through the dot, successive tunneling processes transform into the CB oscillation traces obtained in Fig. 14.2. The distance between the CB peaks provides insight into the energy spectrum of the dot [25]. On the one hand, for the average peak-to-peak separation in Fig. 14.2, we identify the CB oscillation periodicity,  $\Delta V_g \approx 41$  mV, yielding the capacitance between the back gate and quantum dot,  $C_g = e/\Delta V_g \approx 3.9$  aF. On the other hand, we use a simplified model  $C_g = 2\epsilon_0(\epsilon + 1)D$  [17], where  $\epsilon_0$  is the vacuum permittivity,  $\epsilon \approx 4$  is an effective dielectric constant accounting for the presence of a  $\text{SiO}_2$  layer under the single-layer graphene, and  $D$  is the quantum dot diameter. This value of  $C_g$  is estimated to be 7.97 aF, two times larger than the value extracted from the experimental observations. This interesting discrepancy can be ascribed to the electrostatic screening effect of the source and drain electrodes in such a short-channel device [17, 26].

More quantitative information on the graphene quantum dot can be obtained from a measurement of the differential conductance  $dI/dV$  as a function of the back gate voltage  $V_{bg}$  and source-drain bias voltage  $V_{sd}$ . The resulting stability diagram, termed as the well-known Coulomb diamond, is shown in Fig. 14.3a. The height of the Coulomb diamonds yields the charging energy  $E_c \approx 14.1$  meV, corresponding to a total capacitance of the dot  $C_\Sigma = e^2/E_c = C_s + C_d + C_g \approx 11.3$  aF,

**Fig. 14.2** Coulomb blockade oscillations. Conductance of the graphene single quantum dot is given over a wide range of back gate voltage  $V_{bg}$  for source-drain bias voltage  $V_{sd} = 0$



**Fig. 14.3** (a) Coulomb diamonds. The conductance of a graphene single quantum dot is given over a wide range of back gate voltages  $V_{bg}$  and source-drain bias voltages  $V_{sd}$ . Solid lines are guides indicating the region of Coulomb diamonds. (b) Zooming into the nonzero bias voltage region obviously reveals the excited states. Dashed lines are guides indicating the evolution of the excited states with  $V_{sd}$  and  $V_{bg}$

where  $C_s$  and  $C_d$  are the capacitances of the source and drain tunnel barriers, respectively. Given a ratio  $C_d/C_s \approx 2.4$ , obtained directly from the slopes of the diamond edges, we estimate  $C_s \approx 2.2$  aF and  $C_d \approx 5.3$  aF. The discrete energy spectrum of the graphene quantum dot is further revealed by the presence of additional lines parallel to the diamond edges. These lines indicate that the quantum dot is in the high bias regime where the source–drain bias is so high that multiple dot levels involving an excited state can participate in electron tunneling. An enlarged plot of the differential conductance in the high bias regime ( $\Lambda$ -shaped region) is shown in Fig. 14.3b, which is very useful for finding the energies of

the excited states. In the  $\Lambda$ -shaped region, if an excited state level falls within the bias window, a new channel is available for electrons tunneling through the dot and an additional line appears in the stability diagram corresponding to the current change. The line will terminate at the CB region and the bias window exactly equals the energy level spacing [7]. For example, the special lines denoted by A in Fig. 14.3b correspond to single-electron tunneling via the first excited state of  $N + 1$  electrons. Then the level spacing can be read off directly on the  $V_{sd}$  axis as about 1.1 meV. Other lines in the diagram are due to the higher excited states of the dot.

The transport experiment on a gate tunable quantum dot in single-layer graphene to probe the electron ground and excited states is the first step toward the realization of graphene-based solid-state qubit. The device tunability bodes well for future quantum transport studies. Along this line, an interesting problem how to obtain information on the spin state of electrons on a graphene quantum dot through excited-state spectroscopy under a magnetic field could be addressed in the near future.

### 14.3 A Graphene Quantum Dot with a Single-Electron Transistor as an Integrated Charge Sensor

Graphene single quantum dot is the most basic building block for the study of graphene-based qubit. Here as we go further, we use charge sensor to investigate the charge state of the transistor. As we know for spin-based quantum computation it is highly desirable to coherently control the spin degree of freedom recently there was a striking advance on experimental production of graphene single or double quantum dots [3, 5, 11, 17, 18, 27–32], which is an important first step toward such promise. The measurement of individual electrons or its spins in GaAs quantum dots (QDs) has been realized by so-called charge detection via a nearby quantum point contact (QPC) or single-electron transistor (SET) [33, 34]. In particular, the combination of high-speed and high charge sensitivity has made SET useful in studying a wide range of physical phenomena such as discrete electron transport [33, 35, 36], qubit readout [37–39], and nanomechanical oscillators [40, 41]. So far, most SETs have been using Al/AlOx/Al tunnel junctions. Here we realize a simple and reliable technique to fabricate graphene SET, making it an attractive substitute for use in various charge detector applications.

We design an all-graphene nanocircuit integration with a SET as charge readout for a QD. In conventional semiconductor systems, the gate-defined structure limits the distance between the QD and the detector. However, the QD and the SET presented here in the same material are defined in a single etching step, and the distance between the graphene nanostructures is determined by the etched area, which enables optimized coupling and sensing ability.

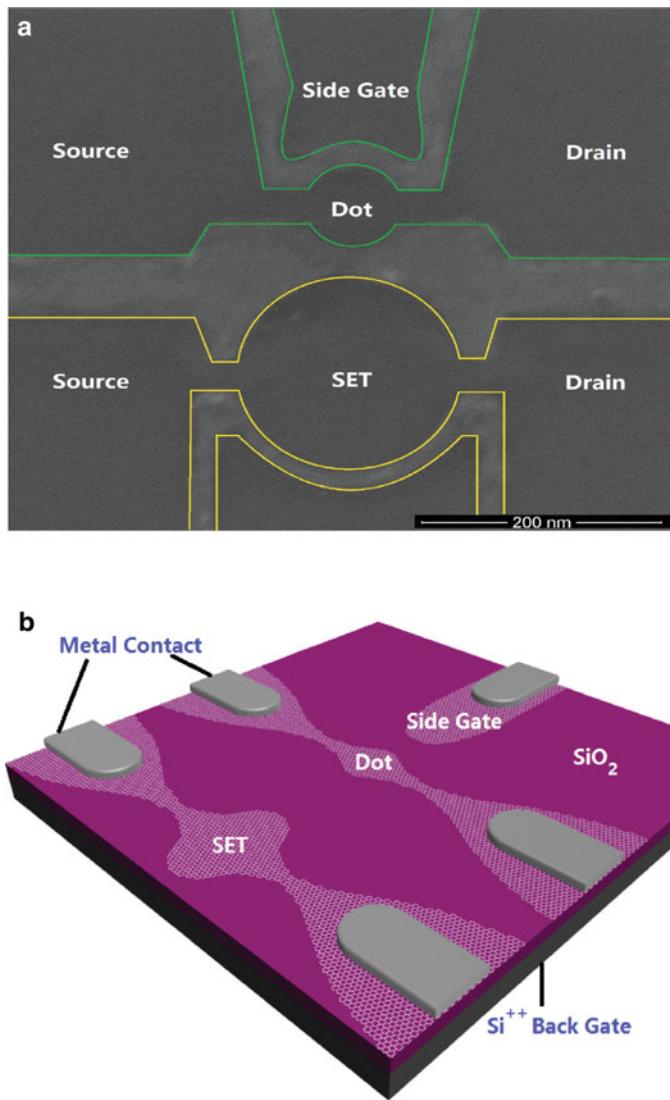
The SET is placed in close proximity to the QD giving rise to a strong capacitive coupling between the two systems. Once an additional electron occupies the QD,

the potential in the neighboring SET is modified by capacitive interaction that gives rise to a measurable conductance change. Even if charge transport through the QD is too small to be measured by conventional transport means, the SET charge sensor also allows measurements. These devices demonstrated here provide robust building blocks in a practical quantum information processor.

The devices have been fabricated using the same method as the graphene single quantum dot device which we have mentioned in the last section. One of our defined sample structures with a quantum dot and proximity SET is shown in Fig. 14.4. The quantum dot is an isolated central island of diameter 90 nm, connected by 30 nm wide tunneling barriers to source and drain contacts.

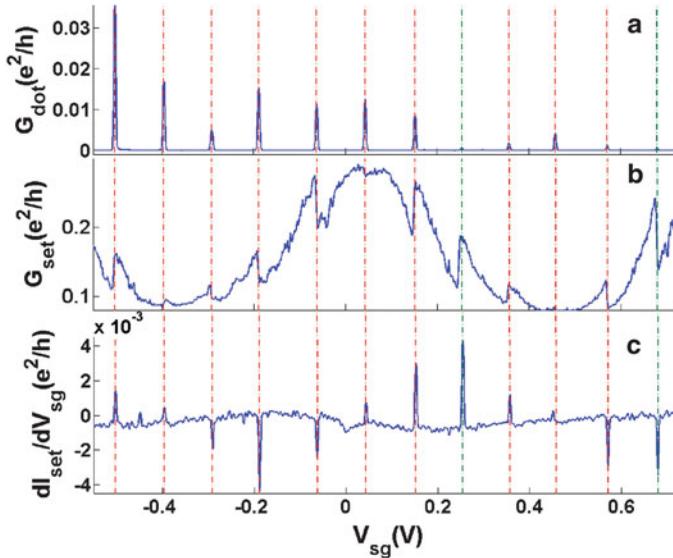
Here, the Si wafer was used as the back gate and there is also a graphene side gate near the small dot. The SET has a similar pattern while the conducting island has a much larger diameter (180 nm). Electronic transport through both the devices exhibits Coulomb blockade (CB) characteristics with back/side gate voltage. The distance between the CB peaks is determined by the sum of charging and quantum confinement energies, and the former contribution becomes dominant for our devices with diameter  $> 100$  nm [42]. Accordingly, we refer to it as a SET rather than a QD. The device was first immersed into a liquid helium storage dewar at 4.2 K to test the functionality of the gates. The experiment was carried out in a top-loading dilution refrigerator equipped with filtered wiring and low-noise electronics at the base temperature of 10 mK.

In the measurement, Standard ac lock-in technique is employed. Figure 14.5a shows the conductance through the dot GQD for applied side gate voltage  $V_{sg}$ . Clear CB peaks are observed related to charging of the tunable dot on the graphene. The dashed green lines in the range 0.2–0.7 V for side gate voltages show that the current through the dot becomes too small to be seen clearly. Figure 14.5b shows the conductance through the SET versus side gate voltage  $V_{sg}$ . The SET is as close as possible to the QD and in this way charging signals of the dot were detected by tracking the change in the SET current. The addition of one electron to the QD leads to a pronounced change in the conductance of the charge detector by typically 30 %. The slope of the SET conductance is the steepest at both sides of its CB resonances giving the best charge readout signal. To offset the large current background, we used a lock-in detection method developed earlier for GaAs dot [43]. A square-shaped pulse was superimposed on the dc bias on side gate voltage  $V_{sg}$ . A lock-in detector in sync with the pulse frequency measured the change in SET current due to the pulse modulation. Figure 14.5c shows a typical trace of the lock-in signal of the transconductance through the SET  $dI_{SET}/dV_{sg}$ . These sharp spikes or dips originate from the change in the charge on the dot by one electron. It shows essentially the same features as Fig. 14.5a, but is much richer, especially in the regime where the direct dot current is too small to be seen clearly. The vertical dashed lines in Fig. 14.5 illustrate that the SET sensor signals correspond to the QD transport measurements perfectly and indicate that the SET is reliable. We also note that the individual charge event measurement has also been demonstrated in a graphene QD with a QPC detector based on graphene nanoribbon [44].



**Fig. 14.4** (a) Scanning electron microscope image of the etched sample structure. The bar has a length of 200 nm. The upper small quantum dot as the main device has a diameter of 90 nm while the bottom single-electron transistor as charge sensor has a diameter of 180 nm. The bright lines define barriers and the graphene side gate. (b) Schematic of a representative device

More quantitative information on the system can be obtained from the measurement of the height response of the peak at 0.152 V in Fig. 14.5c as a function of the modulating pulse frequency on the side gate. The resulting diagram for the SET  $dI_{\text{SET}}/dV_{\text{sg}}$  gain magnitude is shown in Fig. 14.6. The dashed green line indicates the gain of 0.707 ( $-3 \text{ dB}$ ), corresponding approximately to the bandwidth of



**Fig. 14.5** (a) Conductance through the quantum dot vs. the side gate voltage. (b) The example of conductance through the single-electron transistor for the same parameter ranges as in panel (a). The steps in conductance have about 30 % change of the total signal and are well aligned with the CB in panel (a). (c) Transconductance of the single-electron transistor for the same parameters as in panel (a). The *spikes* and *dips* indicate the transitions in the charge states by addition of single electron in quantum dot. In particular, the *dashed green lines* show that the charge detection can allow measurement in the regime where the current through the dot is too small to be seen clearly by direct means. The *vertical dashed red lines* are a guide for the eyes to relate features in these graphs

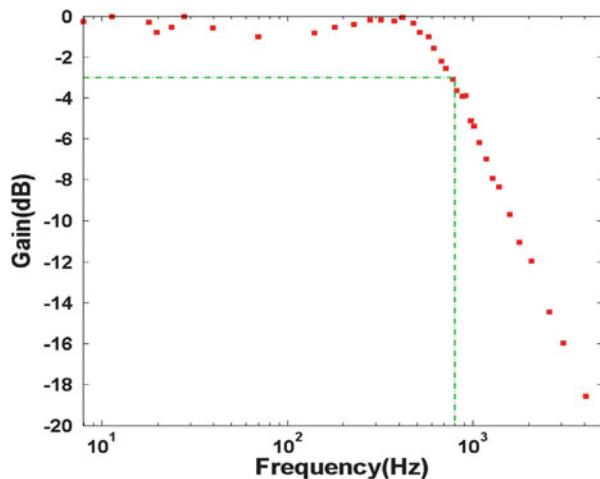
800 Hz of the SET device. By applying a signal of  $5 \times 10^{-2}$  electrons on the back gate of the SET and measuring the signal with a signal-to-noise ratio of 1, we achieved a charge sensitivity of  $10^{-3} e/\sqrt{\text{Hz}}$ , which is similar to that obtained previously in a GaAs QD and superconducting Al SET detector system [45].

The system can be simply considered as a resistor–capacitor circuit (RC circuit), and the bandwidth is limited by the resistor and capacitance of the cable connecting the SET and the room temperature equipment. As a result, we would expect that the bandwidth can be greatly improved by adding a cold amplifier [46]. It is also expected that adding a side gate near the SET to independently set the SET operating point to about 25 k $\Omega$  can obviously enhance the bandwidth.

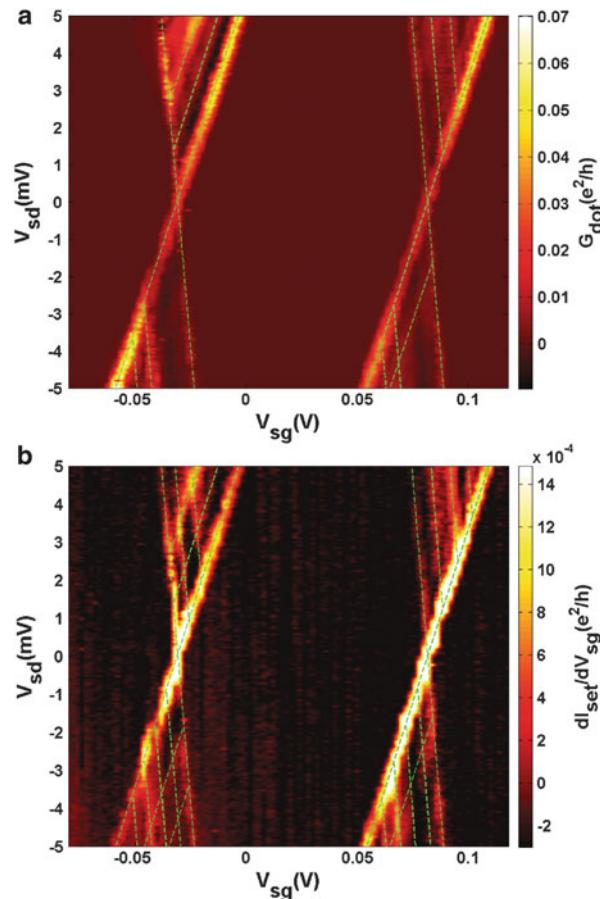
The information contained in the signal goes beyond simple charge counting. For instance, the stability diagram measurement can reveal excited states, which is crucial to get information of the spin state of electrons on a quantum dot [47]. Figure 14.7a shows Coulomb diamonds for the conductance through the dot GQD versus bias voltage  $V_{sd}$  and side gate voltage  $V_{sg}$ . For comparison, Fig. 14.7b shows the transconductance of the SET  $dI_{SET}/dV_{sg}$  as a function of the same parameters.

A perfect match between the QD transport measurements and the detector signal is observed. Moreover, the discrete energy spectra of the graphene quantum dot

**Fig. 14.6** The magnitude of the SET signal  $dI_{\text{SET}}/dV_{\text{sg}}$  as a function of the modulating pulse frequency. The dashed green line illustrates that the bandwidth of the SET device is about 800 Hz corresponding to a gain of 0.707 ( $-3 \text{ dB}$ ). Due to the stray capacitances, the response decreases rapidly after 800 Hz



**Fig. 14.7** (a) Plot of the differential conductance of the quantum dot as a function of the bias voltage and the side gate voltage applied on the dot. From the lines parallel to the edges of Coulomb diamonds, we can identify the excited states. (b) Transconductance of the single-electron transistor with the same parameters as in panel (a). Perfect matching with panel (a) and resolving more excited states spectra indicate that the single-electron transistor can be used as a highly sensitive charge detector. Data in panels (a) and (b) were recorded simultaneously during a single sweep. Dashed green lines are the guide for identifying the excited states



are revealed by the presence of additional lines parallel to the diamond edges. These lines indicate that the quantum dot is in the high bias regime where the source-drain bias is high enough that the excited states can participate in electron tunneling [7]. The excited states become much more visible in the SET charge detector signal than the direct measurement. All of these features have been seen in the GaAs QD with QPC [7], but here we achieve the goal with an all-graphene nanocircuit of QD with SET. In the previous reports, the QD and QPC detector are separated by typically 100 nm in width. In the present case, the SET detector is 50 nm from the edge of the QD. Therefore it is expected that the capacitance coupling between the QD and SET is enhanced compared to the conventional case realized in semiconductor QD and QPC. This enhanced coupling leads to a larger signal-to-noise ratio of the SET detector signal that can be exploited for time-resolved charge measurement or charge/spin qubit readout on the QD.

In conclusion, we realize a simple fabrication process that produces a quantum dot and a highly sensitive single-electron transistor charge detector with the same material, graphene. Typically the addition of a single electron in QD would result in a change in the SET conductance of about 30 %. The charging events measured by both the charge detector and direct transport through the dot perfectly match, and more excited state information beyond the conventional transport means is also obtained. The devices demonstrated here represent a fascinating avenue toward realizing a more complex and highly controllable electronic nanostructure formed from molecular conductors such as graphene.

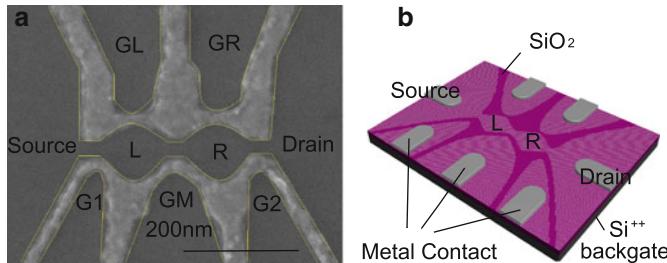
#### 14.4 Controllable Tunnel Coupling and Molecular States in A Graphene Double Quantum Dot

To realize quantum computation, the effects of interactions between qubits and their environment must be minimized [10]. Because of the weak spin-orbit coupling and largely eliminated hyperfine interaction in graphene, it is highly desirable to coherently control the spin degree of freedom in graphene nanostructures for quantum computation [11]. However, the low-energy quasi-particles in single-layer graphene behave as massless Dirac fermions [3, 48], and the relativistic Klein tunneling effect leads to the fact that it is hard to confine electrons within a small region to form quantum dot in graphene using traditional electrostatical gates [5, 11]. It is now possible to etch a graphene flake into nano-constrictions in size, which can obtain electron-bound states and thus act as quantum dots. As a result, the rough edges also lift the valley degeneracy, which could suppress the exchange coupling between spins in the graphene quantum dots [11, 17]. Recently, there was a striking advance on experimental production of graphene single [17, 18, 24, 49–51] or double quantum dots [30, 31, 52–54], which is an important first step toward such promise.

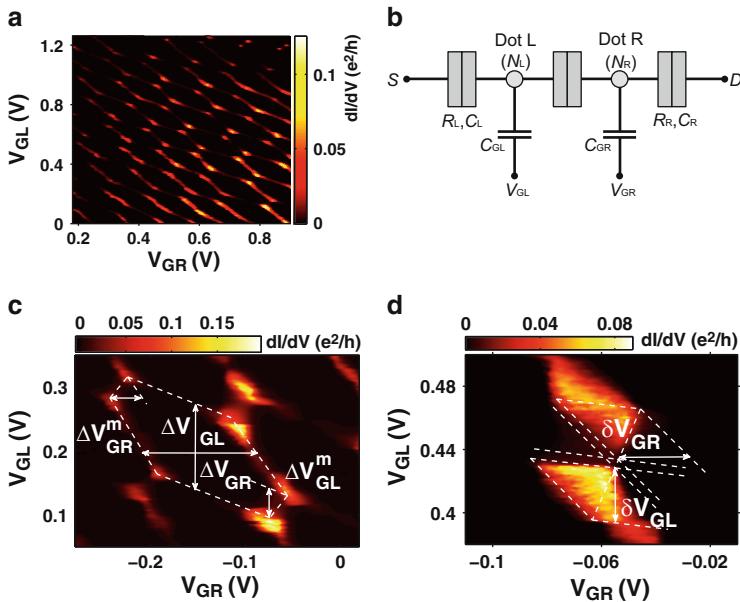
Previously, the charge stability diagram in coupled quantum dot systems has been studied by the classical capacitance model [55]. However, the quantum effect should also manifest itself [56]. In particular, the tunnel coupling  $t$  between the two dots in a double dot is an important quantity, because it can affect the geometry of the overall charge stability diagram. Furthermore, several different spin qubit operations can be performed by controlling this tunnel rate as a function of time. For approaches based on single-electron spin qubit, utilizing  $t$  enables the  $\sqrt{\text{SWAP}}$  gate operations between two qubits [57]. In an architecture in which each qubit is composed of two-electron single-triplet states, control of  $t$  in the presence of a nonuniform magnetic field enables universal single qubit rotations [58].

Here, we talk about an experimental demonstration and electrical transport measurement in a tunable graphene double quantum dot device. Depending on the strength of the interdot coupling, the device can form atomic-like states on the individual dots (weak tunnel coupling) or molecular-like states of the two dots (strong tunnel coupling). We also extract the interdot tunnel coupling  $t$  by identifying and characterizing the molecule states with wave functions extending over the whole graphene double dot. The result implies that this artificial graphene device may be useful for implementing two-electron spin manipulation.

A scanning electron microscope image of the defined sample structure with double quantum dot is shown in Fig. 14.8a, b. The double quantum dot has two isolated central islands of diameter 100 nm in series, connected by  $20 \times 20$  nm narrow constriction to source and drain contacts (S and D electrodes) and  $30 \times 20$  nm narrow constriction with each other. These constrictions are expected to act as tunnel barriers due to the quantum size effect. In addition, the highly P-doped Si substrate is used as a back gate and five lateral side gates, labeled the left gate G1, right gate G2, center gate GM, and GL(R), which are expected for local control. All of the side gates are effective: gates GL, GR, and G2 have very good effect on two dots and middle barrier, while gates G1 and GM have weak effect on those. The device was first immersed into a liquid helium storage dewar at 4.2 K to test the functionality of the gates. The experiment was carried out in a He3 cryostat equipped with filtered wiring and low-noise electronics at the base temperature of 300 mK. In the measurement, we employed the standard AC lock-in technique with an excitation voltage  $20 \mu\text{V}$  at 11.3 Hz. Figure 14.9a displays the differential conductance through the graphene double quantum dot circuit as a function of gate voltages  $V_{\text{GL}}$  and  $V_{\text{GR}}$ . Here, the measurement was recorded at  $V_{\text{sd}} = 20 \mu\text{V}$ ,  $V_{\text{G}1} = 0 \mu\text{V}$ ,  $V_{\text{GM}} = 0 \mu\text{V}$ ,  $V_{\text{G}2} = 0 \mu\text{V}$ , and  $V_{\text{bg}} = 2.5 \text{ V}$ . The honeycomb pattern is clearly visible and uniform over many times. Each cell of the honeycomb corresponds to a well-defined charge configuration (NL, NR) in the nearly independent dots, where NL and NR denote the number of electrons on the left and right dot, respectively. The conductance is large at the vertices, where the electrochemical potentials in both dots are aligned with each other, and the Fermi energy in the leads and resonant sequential tunneling are available. These vertices are connected by faint lines of much smaller conductance along the edges of the honeycomb cells.



**Fig. 14.8** (a) Scanning electron microscope image of the structure of the designed multiple gated sample studied in this work. The double quantum dot has two isolated central islands of diameter 100 nm in series, connected by  $20 \times 20$  nm tunneling barriers to source and drain contacts (S and D) and  $30 \times 20$  nm tunneling barrier with each other. These gates are labeled as G1, GL, GM, GR, and G2 and gates GM, G1, and G2 are used to control the coupling barriers between the dots as well as the leads. Gates GL and GR are used to control and adjust the energy level of each dot. (b) Schematic of a representative device

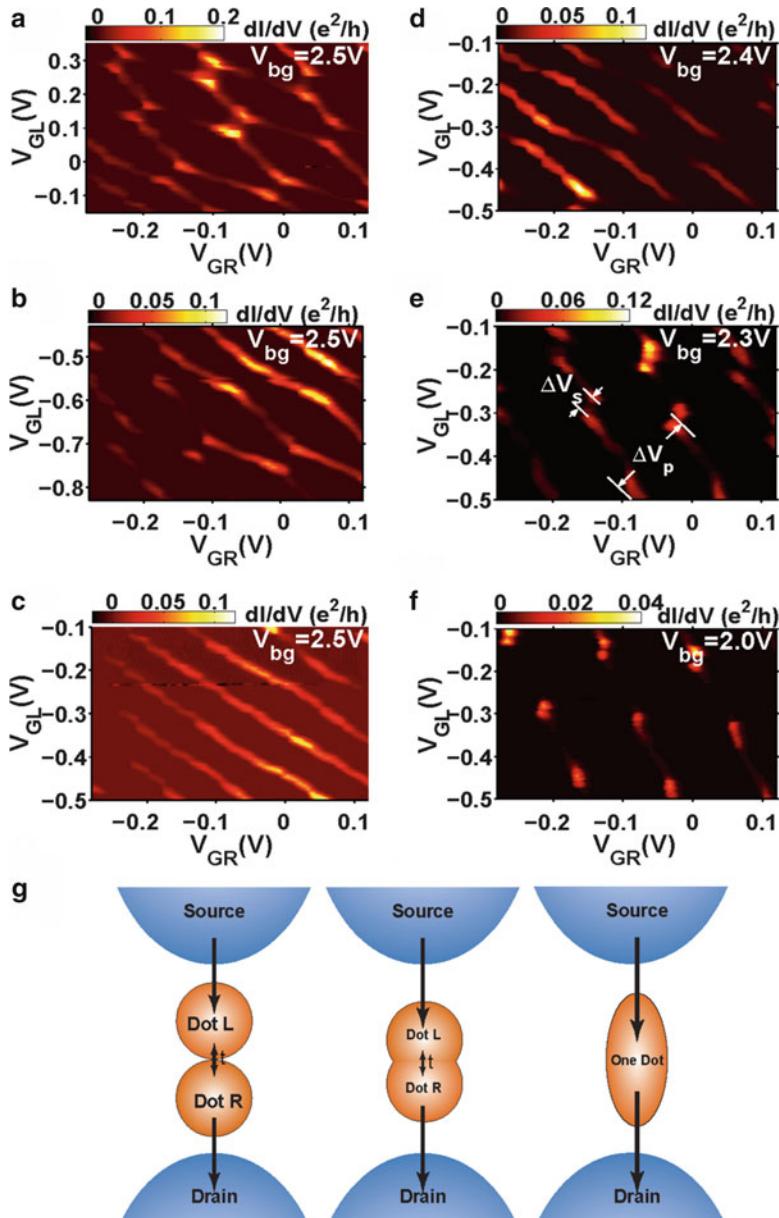


**Fig. 14.9** (a) Color-scale plot of the differential conductance versus voltage applied on gate GL ( $V_{GL}$ ) and gate GR ( $V_{GR}$ ) at  $V_{sd} = 20 \mu\text{V}$ ,  $V_{G1} = 0 \mu\text{V}$ ,  $V_{GM} = 0 \mu\text{V}$ , and  $V_{G2} = 0 \mu\text{V}$ . The honeycomb pattern we got stands for the typical charge stability diagram of coupled double quantum dots. (b) Pure capacitance model of a graphene double-dot system. Zoom-in of a honeycomb structure (c) and a vertex pair (d) at  $V_{sd} = 900 \mu\text{V}$

At these lines, the energy level in one dot is aligned with the electrochemical potential in the corresponding lead and inelastic co-tunneling processes occur. The observed honeycomb pattern resembles the charge stability diagram found for weakly coupled GaAs double quantum dot [55]. Such similarities indicate that

graphene quantum dot devices will continue to share features with well-studied semiconductor quantum dot systems. The energy-level statistics of single graphene quantum dot was probed and shown to agree well with the theory of chaotic Dirac billiards [17]. It is interesting and important to know whether these Dirac fermions' behaviors can be realized and observed in graphene double quantum dot. Nevertheless, it will be studied in the future work. More quantitative information such as double-dot capacitances can be extracted using an electrostatic model as shown in Fig. 14.9b [55]. First, the capacitance of the dot to the side gate can be determined from measuring the size of the honeycomb in Fig. 14.9c as  $C_{\text{GL}} = e/\Delta V_{\text{GL}} \approx 1.27 \text{ aF}$  and  $C_{\text{GR}} = e/\Delta V_{\text{GR}} \approx 1.49 \text{ aF}$ . Next, the capacitance ratios can be determined from measuring the size of the vertices in Fig. 14.9d at finite bias  $V_{\text{sd}} = 900 \mu\text{V}$  as  $\alpha_L = |V_{\text{sd}}|/\delta V_{\text{GL}} = 0.029$  and  $\alpha_R = |V_{\text{sd}}|/\delta V_{\text{GR}} \approx 0.035$ . Using the relation  $C_{\text{GL}}/C_L = \alpha_L$  and  $C_{\text{GR}}/C_R = \alpha_R$ , we can obtain the typical values of dot capacitances as  $C_L \approx 44.8 \text{ aF}$  and  $C_R \approx 44.1 \text{ aF}$ , respectively. The amount of interdot coupling can be achieved by measuring the vertices splitting in Fig. 14.9c. Assuming that the capacitive coupling is dominant in the weakly coupled dot regime [55, 59], the mutual capacitance between dots is calculated as  $C_m = \frac{\Delta V_{\text{GL}}^m C_{\text{GL}} C_R}{e} = \frac{\Delta V_{\text{GR}}^m C_{\text{GR}} C_L}{e} \approx 9.2 \text{ aF}$ .

It has been expected that opening the interdot constriction by gate voltage will cause the tunnel coupling to increase exponentially faster than the capacitive coupling [60]. Figure 14.10a–c represents a selection of such measurements by holding the same  $V_{\text{GR}}$  and  $V_{\text{bg}}$  and scanning different ranges of  $V_{\text{GL}}$  between  $-0.5$  and  $0.35 \text{ V}$ . An evolution of conductance pattern indicates that the stability diagram changes from weak to strong tunneling regimes [55, 59]. The conductance near the vertices depends on the relative contributions of the capacitive coupling and tunnel coupling. For the former, the vertices become a sharpened point, while for the latter, the vertices become blurred along the edges of the honeycomb cell [61]. In Fig. 14.10b, the vertices are not obvious as those in Fig. 14.10a, which indicates a stronger tunnel coupling. The results suggest that two graphene dots are interacting with each other through the large quantum mechanical tunnel coupling, which is analogous to covalent bonding. We will analyze it in details below. An increase in interdot coupling also leads to much larger separation of vertices in Fig. 14.10b [59] and finally to a smearing of honeycomb features in Fig. 14.10c. In this case, the double dots behave like a single dot, as illustrated in Fig. 14.10g. We note that a similar evolution is observed for four different values of  $V_{\text{bg}}$  from  $2.5$  to  $2.0 \text{ V}$  at the same  $V_{\text{GL}}$  and  $V_{\text{GR}}$  regimes as shown in Fig. 14.10d–f. Thus, the interdot tunnel coupling could also be changed by  $V_{\text{GL}}$  or  $V_{\text{bg}}$ . This can be explained by the fact that the side gates and back gate may influence the central barrier through the existing capacitances between the gates and the central barrier. Similar to the definitions in [62], we define  $f = 2\Delta V_S/\Delta V_P$  with  $\Delta V_S$  representing the splitting between vertices in the diagonal direction and  $\Delta V_P$  the vertex pairs distance (Fig. 14.10e). Thus, the case  $f = 1$  stands for strong coupling limit where the double dots behave like a single dot, while the case  $f = 0$  represents weak coupling limit where the double dots behave like two isolated dots. This way,  $f$  should have a certain relationship with tunnel couplings which offers us a method



**Fig. 14.10** (a)–(c) Color-scale plot of the differential conductance versus voltage applied on gate GL ( $V_{GL}$ ) and gate GR ( $V_{GR}$ ) at  $V_{bg} = 2.5$  V for different  $V_{GL}$  regimes. (c)–(f) Color-scale plot of the differential conductance versus voltage applied on gate L ( $V_{GL}$ ) and gate R ( $V_{GR}$ ) for different back gate voltage  $V_{bg}$ . The trend of interdot tunnel coupling changing from weak to strong can be seen clearly

to measure the contribution of the interdot tunneling to the splitting of the vertex. In our double-dot sample, a clear evolution of  $f$  is obtained through scanning different regimes of  $V_{\text{GL}}$  with fixed  $V_{\text{GR}}$  (Fig. 14.10a–c). Through extracting  $\Delta V_{\text{S}}$  and  $\Delta V_{\text{P}}$ , we get  $f \approx 0.5$  for (a),  $f \approx 0.65$  for (b), and  $f \approx 1$  for (c), respectively. These values indicate that control of tunnel coupling as a function of such a gate voltage is conceivable.

Having understood the qualitative behavior of the graphene device in the strong coupling regime, we extract the quantitative properties based on a quantum model of graphene artificial molecule states [56, 61, 63]. Here, we only take into account the topmost occupied state in each dot and treat the other electrons as an inert core [55, 64]. In the case of neglected tunnel coupling, the nonzero conductance can only occur right at the vertices which are energy degenerate points as  $E(N_{\text{L}} + 1, N_{\text{R}}) = E(N_{\text{L}}, N_{\text{R}} + 1)$ . When an electron can tunnel coherently between the two dots, the eigenstates of the double-dot system become the superposed states of two well-separated dot states with the form:

$$\begin{aligned} |\Psi_{\text{B}}\rangle &= -\sin \frac{\theta}{2} e^{\frac{-ip}{2}} |N_{\text{L}} + 1, N_{\text{R}}\rangle + \cos \frac{\theta}{2} e^{\frac{ip}{2}} |N_{\text{L}}, N_{\text{R}} + 1\rangle, \\ |\Psi_{\text{A}}\rangle &= \cos \frac{\theta}{2} e^{\frac{-ip}{2}} |N_{\text{L}} + 1, N_{\text{R}}\rangle + \sin \frac{\theta}{2} e^{\frac{ip}{2}} |N_{\text{L}}, N_{\text{R}} + 1\rangle, \end{aligned}$$

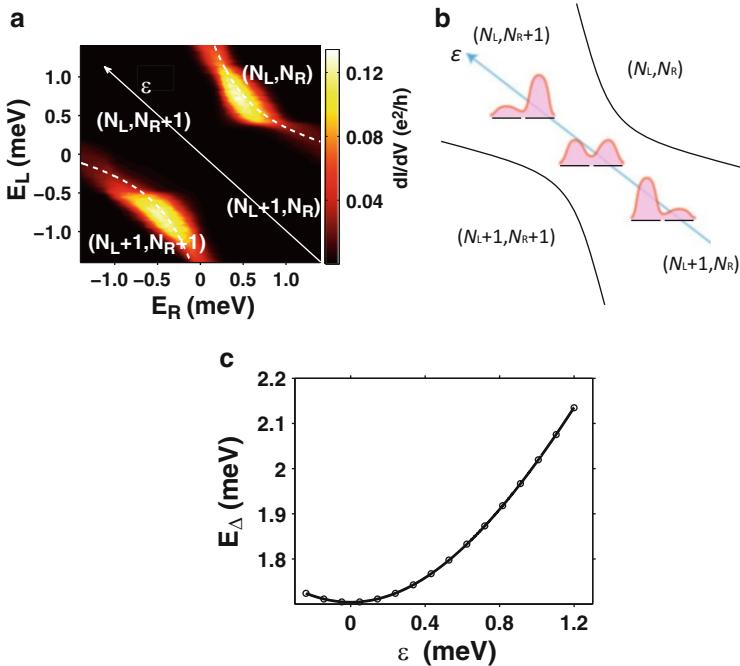
where  $\theta = \arctan(\frac{2t}{\epsilon})$ ,  $\epsilon = E_{\text{L}} - E_{\text{R}}$ , and  $E_{\text{L}}$  and  $E_{\text{R}}$  are the energies of state  $|N_{\text{L}} + 1, N_{\text{R}}\rangle$  and  $|N_{\text{L}}, N_{\text{R}} + 1\rangle$ , respectively. Thus  $|\Psi_{\text{B}}\rangle$  and  $|\Psi_{\text{A}}\rangle$  are the bonding and antibonding state in terms of the uncoupled dot, and the energy difference between these two states can be expressed by

$$E_{\Delta} = U' + \sqrt{\epsilon^2 + (2t)^2}. \quad (14.1)$$

Here  $U' = \frac{2e^2C_{\text{m}}}{C_{\text{L}}C_{\text{R}}-C_{\text{m}}^2}$  is the contribution from electrostatic coupling between dots [65].

Provided that the graphene double-dot molecule eigenstate  $|\Psi\rangle$  participates in the transport process, sequential tunneling is also possible along the honeycomb edges. In Fig. 14.11a, b, a color-scale plot of the differential conductance is shown at  $V_{\text{sd}} = 20 \mu\text{V}$  in the vicinity of a vertex. As expected, the visible conductance is observed at both the position of the vertex and the honeycomb edges extending from the vertex. Figure 14.11c shows a fit of the energy difference  $E_{\Delta}$  from the measured mount of splitting of the positions of the differential conductance resonance peak in the  $\epsilon$ -direction.

Here, we use  $\epsilon = E_{\text{L}} - E_{\text{R}} = e\alpha_{\text{L}}V_{\text{GL}} - e\alpha_{\text{R}}V_{\text{GR}}$  to translate the gate voltage detuning  $V_{\text{GL}} - V_{\text{GR}}$  with the conversion factors  $\alpha_{\text{L}}$  and  $\alpha_{\text{R}}$  determined above. The fitting with (14.1) yields the values of tunnel coupling strength  $t \approx 727 \mu\text{eV}$  and  $U' = 209 \mu\text{eV}$ . Similar measurements have been performed in carbon nanotube double dots with  $t \approx 358 \mu\text{eV}$  and  $U' = 16 \mu\text{eV}$  [61] and semiconductor double dots with  $t \approx 80 \mu\text{eV}$  and  $U' = 175 \mu\text{eV}$  [63]. The fact that the tunnel coupling  $t$  is



**Fig. 14.11** (a) Color-scale plot of the differential conductance versus the energies of each dot  $E_L$  and  $E_R$  at  $V_{sd} = 20 \mu\text{V}$  near the selected two vertices with *dashed lines* as guides to the eye. (b) Schematic of a single anticrossing and the evolution from the state localized in each dot to a molecule state extending across both dots [17]. (c)  $E_\Delta$  dependence of the detuning  $\epsilon = E_L - E_R$ .  $E_\Delta$  (*circles*) is measured from the separation of the two high conductance wings in Fig. 14.11a. The *line* illustrates a fit of the data to (14.1).

dominant than capacitive coupling  $U'$  implies that the interdot tunnel barrier in the etched graphene double dot is much more transparent than those gated carbon nanotube or semiconductor double dot.

Finally, we discuss the relevance of graphene double-dot device for implementing a quantum gate and quantum entanglement of coupled electron spins. An operation has already been demonstrated in a semiconductor double-dot system using the fast control of exchange coupling  $J$  [57]. The operation time  $\tau$  is about 180 ps for  $J \approx 0.04 \text{ meV}$  corresponding to  $J \approx 0.16 \text{ meV}$ . In the present graphene device, we have obtained much larger  $t \approx 0.72 \text{ meV}$ , and the estimated  $\tau \approx 50 \text{ ps}$  is much shorter than the predicted decoherence time ( $\mu\text{s}$ ) [66]. The results indicate the ability to carry out two-electron spin operations in nanosecond timescales on a graphene device, four times faster than previously shown for semiconductor double dot.

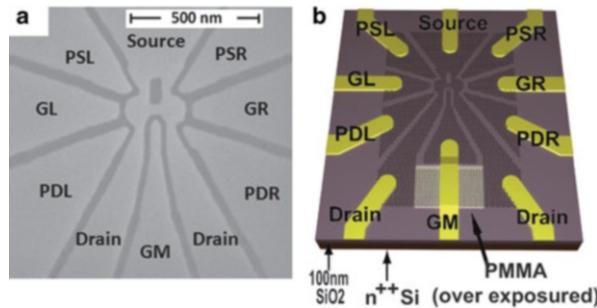
In conclusion, a graphene double quantum dot with multiple electrostatic gates has been measured and the transport pattern evolution in different gate configurations is observed. This way offers us a method to identify the molecular states as a quantum-mechanical superposition of double dot and measure the

contribution of the interdot tunneling to the splitting of the differential conductance vertex. The precisely extracted values of interdot tunnel coupling  $t$  for this system are much larger than those in previously reported semiconductor device. These short operation times due to large tunneling strength together with the predicted very long coherence times suggest that the requirements for implementing quantum information processing in graphene nanodevice are within reach.

## 14.5 Gate-Controlled Parallel-Coupled Double Quantum Dot on Both Single-Layer and Bilayer Graphene

Gate-controlled double quantum dot (DQD) system has been considered as a promising candidate of spin-based solid-state qubits for the quantum computation processing [67, 68]. Many efforts and progresses have been made in the double-dot devices research based on various materials, including GaA two-dimensional electron gas [7, 55], semiconductor nanowires [69, 70], and carbon nanotube [59, 71]. The natural two-dimensional material, graphene, has attracted extensive interest due to its distinguishing electronic quality and flexibility in device designs [3, 17, 24, 31, 32, 44, 52]. In addition, it is theoretically predicted that the spin decoherence time in graphene can be very long due to its weak spin-orbit coupling and largely eliminable hyperfine interaction, which has a significant meaning for spin-based quantum processors. In contrast to DQD in series, where the applied current passes through the double dot serially, the parallel-coupled double quantum dot (PDQD) requires two sets of entrances and exits, one for each dot. PDQD is an ideal artificial system for investigating the interaction and the interference. Rich physical phenomena, such as Aharonov-Bohm (AB) effect, Kondo regimes, and Fano effect, have been predicted to be observed in parallel PDQD [72-76]. Particular excitement is the prospect of accessing theoretically predicted quantum critical points in quantum phase transitions [77]. The graphene PDQD is an attractive system for investigating the quantum phase transitions due to its intrinsically large energy separation between on-dot quantum levels, thus offering a significant advantage over conventional systems as GaAs or silicon-based quantum dots. In this work, we present the design, fabrication, and quantum transport measurement of double-dot structure coupled in parallel, on both bilayer and single-layer graphene flakes, which may open a door to study the rich PDQD physical phenomena in this material the parallel graphene structure can be tuned from a strong coupling resulted artificial molecule state to a weak coupling resulted two-dot state by adjusting in-plane plunger gates. The tuning is found to be very reliable and reproducible, with good long-term stability on the order of days.

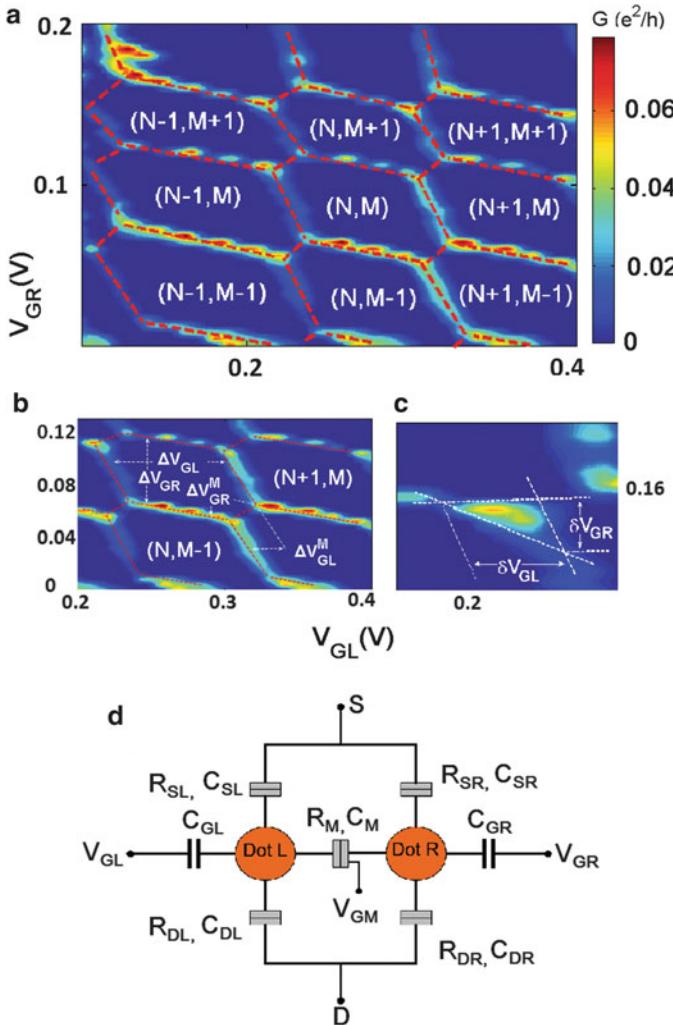
Graphene flakes are produced by mechanical cleaving of bulk graphite crystallites by Scotch tape [1]. For this kind of exfoliated graphene flakes on SiO<sub>2</sub> substrate, the mobility is normally about 15,000 cm<sup>2</sup>/(Vs) [3]. By using heavily doped Si substrate with 100 nm thick SiO<sub>2</sub> on top, we can identify monolayer,



**Fig. 14.12** (a) Scanning electron microscope image of the etched parallel-coupled graphene double-dot sample structure. The bar has a length of 500 nm. The diameters of the two dots are both 100 nm; constriction between the two dots is 35 nm in width and length. The four narrow parts connecting the dot to source and drain parts have a width of 30 nm. Seven in-plane plunger gates GL, GR, GM, PSL, PDL, PSR, and PDR are integrated around the dot for fine tuning. Gates GL, GR, and GM are respectively designed to adjust the energy level of left dot, right dot, and interdot coupling strength. (b) Schematic picture of the device. n-type heavily doped silicon substrate is used as a global back gate. A layer of overexposed PMMA is used as a bridge to make gate GM separated from the drain part of graphene

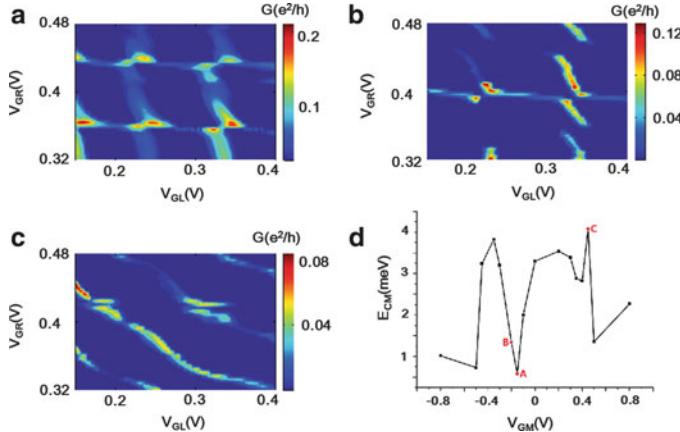
bilayer, and few layer graphenes through optical microscope. Monolayer and bilayer graphenes were further checked by Raman spectrum. Firstly, graphene flakes are transferred to the substrate with gold markers. Then, a layer of 50 nm thick polymethyl methacrylate (PMMA) is spun on the substrate for electron beam lithography (EBL) to form a designed pattern. After that, O<sub>2</sub>/Ar (50:50) plasma is used to remove unprotected parts of graphene. Next, an area of overexposed PMMA is used to separate a bridge plunger gate from the drain part of graphene [76, 78]. The final step is to make the metal contacts, which are defined by the standardized EBL process, followed by the E-beam evaporation of Ti/Au (2 nm/50 nm).

Figure 14.12a shows a scanning electron microscope (SEM) image of one sample with the same structure as the bilayer device we measured. Two central islands with diameter of 100 nm connect through 30 nm wide narrow constrictions to the source and the drain regions. Another narrow constriction (35 nm in both width and length) connects the two central islands. Seven in-plane plunger gates labeled as GL, GR, GM, PSL, PDL, PSR, and PDR are integrated in close proximity to the dots. GL, GR, and GM are, respectively, designed to adjust the energy level of left dot, right dot, and interdot coupling strength. And PSL and PDL (PSR, PDR) are used for the tuning of the coupling of the left (right) dot to source and drain. The n-type heavily doped silicon substrate is used as a global back gate. The bridge plunger gate GM is separated from the drain part of graphene by a layer of overexposed PMMA. All the devices were primarily tested to check the functionality of all the gates in a liquid helium storage dewar at 4.2 K. Then the samples were mounted on a dilution refrigerator equipped with filtering wirings and low-noise electronics at the base temperature of 10 mK. To maintain consistency, we will use the data from one sample only in the following.



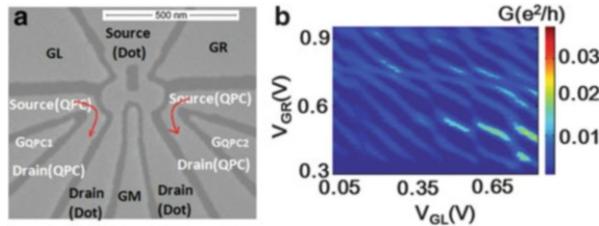
**Fig. 14.13** (a)–(c) PDQD conductance as a function of plunger gate voltage  $V_{GL}$  and  $V_{GR}$ . The red dash lines are guides to the eyes showing the honeycomb pattern.  $(N, M)$  represents the carriers in the left and right dot, respectively. (b) Zoom-in of the area  $(N, M)$  of the honeycomb pattern. (c) Zoom-in of a vertex pair with white dash lines. (d) Capacitance model for the analysis of the double-dot system. Graphene nano-constriction behavior as tunneling barriers, which are presented, for example, as  $R_{SL}, C_{SL}$  (a capacitance and a resistance coupled in parallel). Gates GL and GR are capacitively coupled to the dots;  $C_{GL}$  and  $C_{GR}$  represent the capacitance

Figure 14.13a shows color-scale plot of the measured differential conductance of the double dot as a function of  $V_{GL}$  and  $V_{GR}$  detected in standard ac lock-in technique with an excitation ac voltage  $20 \mu\text{V}$  at frequency of  $11.3 \text{ Hz}$ . A dc bias of  $0.3 \text{ mV}$  is applied, the back gate voltage  $V_{bg}$  is fixed at  $5 \text{ V}$ , and the middle plunger gate  $V_{GM}$  is  $-0.45 \text{ V}$ . The hexagon pattern characteristic for double dot



**Fig. 14.14** Interdot coupling vs. middle gate voltage  $V_{GM}$ . Conductance as a function of gate voltage  $V_{GL}$  and  $V_{GR}$  at  $V_{bg} = 3$  V,  $V_{bias} = -1$  mV, and the scan region of GL and GR is the same. (a)–(c) represent three different coupling regimes of the two dots. (a) Weak coupling regime,  $V_{GM} = -0.15$  V, (b) medium coupling regime,  $V_{GM} = -0.2$  V, and (c) strong coupling regime,  $V_{GM} = 0.45$  V. (d) shows coupling energy  $E_{CM}$  (V) as a non-monotonic function of the middle gate voltage  $V_{GM}$ . A, B, and C points here represent the corresponding coupling energy in (a), (b), and (c)

coupled in parallel is clearly visible. Figure 14.13b is a zoom-in of the area ( $N, M$ ) of the honeycomb pattern and Fig. 14.13c zoom-in of a vertex pair with white dashed lines. From the model of purely capacitively coupled dots as illuminated by Fig. 14.13d, the energy scales of the system can be extracted [31, 32, 52, 55]. The capacitance of the dot to the side gate can be determined from measuring the size of the honeycomb as shown in Fig. 14.13a, b;  $\Delta V_{GL} = 0.087$  V,  $\Delta V_{GR} = 0.053$  V,  $\Delta V_{GL}^M = 0.0261$  V, and  $\Delta V_{GR}^M = 0.0133$  V; therefore,  $C_{GL} = e/\Delta V_{GL} = 1.84$  aF and  $C_{GR} = e/\Delta V_{GR} = 3.0$  aF. With a large DC bias of 0.3 mV, we can get  $\Delta V_{GL} = 0.013$  V and  $\Delta V_{GR} = 0.01$  V as shown in Fig. 14.13c. The lever arm between the left (right) gate  $V_{GL}$  and the left (right) dot can be calculated as  $\alpha_{GL} = V_{bias}/\Delta V_{GL} = 0.023$  ( $\alpha_{GL} = V_{bias}/\delta V_{GR} = 0.03$ ). The total capacitances of the dots can then be calculated as  $C_L = C_{GL}/\alpha_L = 79.8$  aF and  $C_R = C_{GR}/\alpha_R = 100.4$  aF, the corresponding charging energy  $E_{CL} = \alpha_{GL} \cdot \Delta V_{GL} = 2.0$  meV and  $E_{CR} = \alpha_{GR} \cdot \Delta V_{GR} = 1.6$  meV, and the coupling energy between the two dots  $E_{CM} = \alpha_{GL} \cdot \Delta V_{GL}^m = 0.3$  meV. It is also noted that the lever arms between the left gate and the right dot and vice versa can be determined from the slope of the co-tunneling lines delimiting the hexagons. These crossing couplings only modify the results slightly and are neglected usually [31, 32, 52]. Here, by calculating dots area and carrier density (related to  $V_{bg}$ ), or from the Coulomb charging period, we estimate that each dot contains more than 20 electrons when  $V = 5$  V. By applying voltage to the middle plunger gate GM, the interdot coupling can be tuned efficiently. Figure 14.14a–c shows the charge stability diagrams of the PDQD in three different coupling regimes: (a) weak, (b) medium, and (c) strong. In these



**Fig. 14.15** (a) SEM image of single-layer graphene PDQD integrated with two QPCs. The *bar* has a length of 500 nm. (b) Characteristic honeycomb structure of the conductance through the PDQD as a function of two in-plane plunger gates voltage  $V_{GL}$  and  $V_{GR}$ , revealed by direct transport measurement of the PDQD at 4.2 K

measurements, back gate voltage  $V_{bg} = 3$  V, source–drain DC bias  $V_{bias}$  is set to  $-1$  mV, and the scan regions of GL and GR are the same. Only the voltage applied to the gate GM is adjusted as (a)  $V_{GM} = -0.15$  V, (b)  $V_{GM} = -0.2$  V, and (c)  $V_{GM} = 0.45$  V. By using the same model as in Fig. 14.13, we can calculate the corresponding coupling energy between the dots: (a)  $E_{CM} = 0.58$  meV, (b)  $E_{CM} = 1.34$  meV, and (c)  $E_{CM} = 4.07$  meV. The honeycomb diagrams of the parallel and serial DQD look similar except for the weak coupling regime, as shown in Fig. 14.14a. In this case, the lines delimiting the hexagons are more visible in comparison with serial DQD, because the leads have two parallel accesses to the dots in parallel DQD, which also enables correlated tunneling of two valence electrons simultaneously [79]. Figure 14.14d indicates the coupling energy changes with the gate voltage  $V_{GM}$ . As in the previous reports of graphene DQD in series [31, 52], the interdot coupling is non-monotonically dependent on the applied gate voltage. Although the detailed reasons for this non-monotony are undetermined, we assumed that one key factor will be the disorders in graphene introduced by either fabrication steps or substrate [80]. Many more efforts are still needed to address this issue for the realization of practical graphene-based nanodevices.

We have designed and fabricated an alternative structure of a PDQD integrated with two quantum point contact sensors (QPCs) in single-layer graphene, as shown in Fig. 14.15a. The integrated QPCs can be used as a noninvasive charge detector which may have various applications [7, 55, 81, 82]. As primary tests of the present structure, we can get similar charge stability diagram of the PDQD as in Fig. 14.15b by the direct quantum transport tests at 4.2 K. Although the noninvasive measurements by QPC are still under processing, no remarkable difference is founded between PDQD in bilayer and monolayer graphenes from direct transport measurement. Making tunable coupling double dot is the first step toward the quantum dot based quantum computation bits; the architectonics with integrated charge detector around double quantum dot demonstrated here offers the chance to achieve the charge or spin reading out, which is essential for the quantum computation device. Therefore, a lot of extended and follow-up works can be done on this basis in the future. Both bilayer and single-layer graphenes can be exploited in this application.

In conclusion, low-temperature quantum transport measurement of gate-controlled parallel-coupled double quantum dot on both bilayer and single-layer graphenes have been investigated. The interdot coupling strength can be largely tuned by graphene in-plane gates. With the quantum transport honeycomb charge stability diagrams, a common model of purely capacitively coupled double dot is used to extract all the relevant energy scales and parameters of graphene PDQD. Although many more effects are still needed to further upgrade and exploit the present designed graphene quantum dot system, the results have intensively demonstrated the promise of the realization of graphene nanodevice and desirable study of rich PDQD physical phenomena in graphene.

## References

- Novoselov, K.S., Geim, A.K., Morozov, S.V., Jiang, D., Zhang, Y., Dubonos, S.V., Grigorieva, I.V., Firsov, A.A.: *Science* **306**, 666–669 (2004)
- Novoselov, K., Jiang, D., Schedin, F., Booth, T., Khotkevich, V., Morozov, S., Geim, A.: *Proc. Natl. Acad. Sci. USA* **102**, 10451 (2005)
- Geim, A.K., Novoselov, K.: *Nat. Mater.* **6**, 183–191 (2007)
- Neto, A., Guinea, F., Peres, N., Novoselov, K., Geim, A.: *Rev. Mod. Phys.* **81**, 109–162 (2009)
- Ihn, T., Güttinger, J., Molitor, F., Schnez, S., Schurtenberger, E., Jacobsen, A., Hellmüller, S., Frey, T., Droscher, S., Stampfer, C., Ensslin, K.: *Mater. Today* **13**, 44 (2010)
- Schwierz, F.: *Nat. Nanotechnol.* **5**, 487 (2010)
- Hanson, R., Kouwenhoven, L.P., Petta, J.R., Tarucha, S., Vandersypen, L.M.K.: *Rev. Mod. Phys.* **79**, 1217 (2007)
- Hanson, R., Awschalom, D.: *Nature* **453**, 1043 (2008)
- Taylor, J.M., Engel, H.A., Dur, W., Yacoby, A., Marcus, C.M., Zoller, P., Lukin, M.D.: *Nat. Phys.* **1**, 177 (2005)
- Fischer, J., Loss, D.: *Science* **324**, 1277 (2009)
- Trauzettel, B., Bulaev, D.V., Loss, D., Burkard, G.: *Nat. Phys.* **3**, 192 (2007)
- Castro Neto, A.H., Guinea, F., Peres, N.M.R., Novoselov, K.S., Geim, A.K.: *Rev. Mod. Phys.* **81**, 109 (2009)
- Silvestrov, P.G., Efetov, K.B.: *Phys. Rev. Lett.* **98**, 016802 (2007)
- Pereira Jr., J.M., Vasilopoulos, P., Peeters, F.M.: *Nano Lett.* **7**, 946 (2007)
- Recher, P., Nilsson, J., Burkard, G., Trauzettel, B.: *Phys. Rev. B* **79**, 085407 (2009)
- Martino, A.D., Dell'Anna, L., Egger, R.: *Phys. Rev. Lett.* **98**, 066802 (2007)
- Ponomarenko, L., Schedin, F., Katsnelson, M., Yang, R., Hill, E., Novoselov, K., Geim, A.: *Science* **320**, 356 (2008)
- Stampfer, C., Schurtenberger, E., Molitor, F., Guettinger, J., Ihn, T., Ensslin, K.: *Nano Lett.* **8**, 2378 (2008)
- Novoselov, K.S., Geim, A.K., Morozov, S.V., Jiang, D., Katsnelson, M.I., Grigorieva, I.V., Dubonos, S.V., Firsov, A.A.: *Nature* **438**, 197–200 (2005)
- Zhang, Y., Tan, Y.W., Stormer, H.L., Kim, P.: *Nature* **438**, 201–204 (2005)
- Jiao, L., Zhang, L., Wang, X., Diankov, G., Dai, H.: *Nature* **458**, 877–880 (2009)
- Dimiev, A., Price, B., Tour, J.: *Nature* **458**, 872–876 (2009)
- Neubeck, S., Ponomarenko, L., Freitag, F., Giesbers, A., Zeitler, U., Morozov, S., Blake, P., Geim, A., Novoselov, K.: *Small* **6**, 1469–1473 (2010)
- Wang, L.J., Cao, G., Tu, T., Li, H.O., Zhou, C., Hao, X.J., Su, Z., Guo, G.C., Jiang, H.W., Guo, G.P.: *Appl. Phys. Lett.* **97**, 262113 (2010)

25. Marcus, C.M., McEuen, P.L., Tarucha, S., Westervelt, R.M., Wingreen, N.: In: NATO ASI conference proceedings (1997)
26. Gelmont, B., Shur, M.S., Mattauch, R.J.: Solid State Electron. **38**, 731 (1995)
27. Avouris, P.Z., Chen, H., Perebeinos, V.: Nat. Nanotechnol. **2**, 605 (2007)
28. Miao, F., Wijeratne, S., Zhang, Y., Coskun, U.C., Bao, W., Lau, C.N.: Science **317**, 1530 (2007)
29. Girit, C., Bouchiat, V., Naaman, O., Zhang, Y., Crommie, M.F., Zettl, A., Siddiqi, I.: Nano Lett. **9**, 198 (2009)
30. Molitor, F., Knowles, H., Droscher, S., Gasser, U., Choi, T., Rouleau, P., Guttinger, J., Jacobsen, A., Stampfer, C., Ensslin, K., Ihn, T.: Europhys. Lett. **89**, 67005 (2010)
31. Liu, X.L., Hug, D., Vandersypen, L.M.K.: Nano Lett. **10**, 1623 (2010)
32. Moriyama, S., Tsuya, D., Watanabe, E., Uji, S., Shimizu, M., Mori, T., Yamaguchi, T., Ishibashi, K.: Nano Lett. **9**, 2891 (2009)
33. Lu, W., Ji, Z.Q., Pfeiffer, L., West, K.W., Rimberg, A.J.: Nature (London) **423**, 422 (2003)
34. Elzerman, J.M., Hanson, R., van Beveren, L.H.W., Witkamp, B., Vandersypen, L.M.K., Kouwenhoven, L.P.: Nature (London) **430**, 431 (2004)
35. Bylander, J., Duty, T., Delsing, P.: Nature (London) **434**, 361 (2005)
36. Gotz, G., Steele, G.A., Vos, W.J., Kouwenhoven, L.P.: Nano Lett. **8**, 4039 (2008)
37. Lehner, K.W., Bladh, K., Spietz, L.F., Gunnarsson, D., Schuster, D.I., Delsing, P., Schoelkopf, R.J.: Phys. Rev. Lett. **90**, 027002 (2003)
38. Duty, T., Gunnarsson, D., Bladh, K., Delsing, P.: Phys. Rev. B **69**, 140503(R) (2004)
39. Vijay, R., Devoret, M.H., Siddiqi, I.: Rev. Sci. Instrum. **80**, 111101 (2009)
40. Knobel, G., Cleland, A.N.: Nature (London) **424**, 291 (2003)
41. LaHaye, D., Buu, O., Camarota, B., Schwab, K.C.: Science **304**, 74 (2004)
42. Kouwenhoven, L.P., Marcus, C., McEuen, P.L., Tarucha, S., Westervelt, R.M., Wingreen, N. S.: In: Mesoscopic electron transport, Series E: Applied sciences, vol. 345, pp. 105–214. Kluwer Academic, Dordrecht (1997)
43. Elzerman, J.M., Hanson, R., van Beveren, L.H.W., Vandersypen, L.M.K., Kouwenhoven, L. P.: Appl. Phys. Lett. **84**, 4617 (2004)
44. Güttinger, J., Stampfer, C., Hellmüller, S., Molitor, F., Ihn, T., Ensslin, K.: Appl. Phys. Lett. **93**, 212102 (2008)
45. Berman, D., Zhitenev, N.B., Ashoori, R.C., Shayegan, M.: Phys. Rev. Lett. **82**, 161 (1999)
46. Vink, I.T., Nooitgedagt, T., Schouten, R.N., Vandersypen, L.M.K.: Appl. Phys. Lett. **91**, 123512 (2007)
47. Hanson, R., Witkamp, B., Vandersypen, L.M.K., van Beveren, L.H.W., Elzerman, J.M., Kouwenhoven, L.P.: Phys. Rev. Lett. **91**, 196802 (2003)
48. Geim, A.K.: Science **324**, 1530 (2009)
49. Stampfer, C., Güttinger, J., Molitor, F., Graf, D., Ihn, T., Ensslin, K.: Appl. Phys. Lett. **92**, 012102 (2008)
50. Schnez, S., Molitor, F., Stampfer, C., Güttinger, J., Shorubalko, I., Ihn, T., Ensslin, K.: Appl. Phys. Lett. **94**, 012107 (2009)
51. Güttinger, J., Seif, J., Stampfer, C., Capelli, A., Ensslin, K., Ihn, T.: Phys. Rev. B **83**, 165445 (2011)
52. Molitor, F., Droscher, S., Güttinger, J., Jacobson, A., Stampfer, C., Ihn, T., Ensslin, K.: Appl. Phys. Lett. **94**, 222107 (2009)
53. Wang, L.J., Guo, G.P., Wei, D., Cao, G., Tu, T., Xiao, M., Guo, G.C., Chang, A.M.: Appl. Phys. Lett. **99**, 112117 (2011)
54. Volk, C., Fringes, S., Terres, B., Dauber, J., Engels, S., Trelenkamp, S., Stampfer, C.: Nano Lett. **11**, 3581 (2011)
55. van der Wiel, W.G., de Franceschi, S., Elzermann, J.M., Fujisawa, T., Tarucha, S., Kouwenhoven, L.P.: Rev. Mod. Phys. **75**, 1 (2003)
56. Yang, S., Wang, X., Das Sarma, S.: Phys. Rev. B **83**, 161301(R) (2011)

57. Petta, J.R., Johnson, A.C., Taylor, J.M., Laird, E.A., Yacoby, A., Lukin, M.D., Marcus, C.M., Hanson, M.P., Gossard, A.C.: *Science* **309**, 2180 (2005)
58. Foletti, S., Bluhm, H., Mahalu, D., Umansky, V., Yacoby, A.: *Nat. Phys.* **5**, 903 (2009)
59. Mason, N., Biercuk, M.J., Marcus, C.M.: *Science* **303**, 655 (2004)
60. Kouwenhoven, L.P., Marcus, C.M., McEuen, P.L., Tarucha, S., Westervelt, R.M., Wingreen, N.: In: Sohn, L.L., Kouwenhoven, L.P., Schon, G. (eds.) *Mesoscopic electron transport*. Kluwer, Dordrecht (1997)
61. Gruber, M.R., Coish, W.A., Hoffmann, C., Weiss, M., Furer, J., Oberholzer, S., Loss, D., Schonenberger, C.: *Phys. Rev. B* **74**, 075427 (2007)
62. Livermore, C., Crouch, C.H., Westervelt, R.M., Campman, K.L., Gossard, A.C.: *Science* **274**, 1332 (1996)
63. Hatano, T., Stopa, M., Tarucha, S.: *Science* **309**, 268 (2005)
64. Golovach, V.N., Loss, D.: *Phys. Rev. B* **69**, 245327 (2004)
65. Ziegler, R., Bruder, C., Schoeller, H.: *Phys. Rev. B* **62**, 1961 (2000)
66. Fischer, J., Trauzettel, B., Loss, D.: *Phys. Rev. B* **80**, 155401 (2009)
67. Loss, D., DiVincenzo, D.P.: *Phys. Rev. A* **57**, 120 (1998)
68. Cerletti, V., Coish, W.A., Gywat, O., Loss, D.: *Nanotechnology* **16**, R27 (2005)
69. Fasth, C., Fuhrer, A., Bjork, M.T., Samuelson, L.: *Nano Lett.* **5**, 1487 (2005)
70. Pfund, A., Shorubalko, I., Leturcq, R., Ensslin, K.: *Appl. Phys. Lett.* **89**, 252106 (2006)
71. Sapmaz, S., Meyer, C., Beliczyński, P., Jarillo-Herrero, P., Kouwenhoven, L.P.: *Nano Lett.* **6**, 1350 (2006)
72. Holleitner, A.W., Decker, C.R., Qin, H., Eberl, K., Blick, R.H.: *Phys. Rev. Lett.* **87**, 256802 (2001)
73. López, R., Aguado, R., Platero, G.: *Phys. Rev. Lett.* **89**, 136802 (2002)
74. Ladrón de Guevara, M.L., Claro, F., Orellana, P.A.: *Phys. Rev. B* **67**, 195335 (2003)
75. Orellana, P.A., Ladron de Guevara, M.L., Claro, F.: *Phys. Rev. B* **70**, 233315 (2004)
76. Chen, J.C., Chang, A.M., Melloch, M.R.: *Phys. Rev. Lett.* **92**, 176801 (2004)
77. Dias de Silva, L.G.G.V., Ingersent, K., Sandler, N., Ulloa, S.E.: *Phys. Rev. B* **78**, 153304 (2008)
78. Huard, B., Sulpizio, J.A., Stander, N., Todd, K., Yang, B., Goldhaber-Gordon, D.: *Phys. Rev. Lett.* **98**, 236803 (2007)
79. Holleitner, A.W., Blick, R.H., Huttel, A.K., Eberl, K., Kotthaus, J.P.: *Science* **297**, 70 (2002)
80. Todd, K., Chou, H.T., Amasha, S., Goldhaber-Gordon, D.: *Nano Lett.* **9**, 416 (2009)
81. Guo, G.P., Li, C.F., Guo, G.C.: *Phys. Lett. A* **286**, 401 (2001)
82. Zhang, H., Guo, G.P., Tu, T., Guo, G.C.: *Phys. Rev. A* **76**, 012335 (2007)

# Chapter 15

## Terahertz Response in Schottky Warp-Gate Controlled Single Electron Transistors

Weihua Han and Seiya Kasai

**Abstract** Schottky wrap-gate controlled GaAs-based single electron transistor (SET) can operate as a high sensitivity terahertz (THz) detector. The reproducible THz photocurrent was observed at low temperature in the device at the normal incidence of CH<sub>3</sub>OH gas laser with the frequency 2.54 THz. The change of source-drain current induced by THz photon shows that a satellite peak was generated beside the resonance peak. THz photon energy can be characterized by the difference of gate voltage positions between the resonance peak and satellite peak, which indicates that the satellite peak exactly results from the THz photon-assisted tunneling. Both experimental results and theoretic analysis verify that the narrow spacing of double barriers is more effective for the enhancement of THz response.

### 15.1 Introduction

Terahertz (THz) wave frequency region has attracted considerable attention as the remaining frequency resource for applications such as high-capacity communications, biomedical image, chemical analysis of molecular spectroscopy, airport security, and remote sensing in space. The THz region of the electromagnetic spectrum filled the rather large wavelength range between 1 mm and 100 μm (3 GHz–3 THz), which corresponds to an approximate photon energy between 1.2 and 12.4 meV or to an equivalent blackbody temperature between 14 and 140 K, well below the ambient background on Earth. Sandwiched between traditional

---

W. Han (✉)

Engineering Research Center of Semiconductor Integrated Technology, Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, People's Republic of China  
e-mail: [weihua@semi.ac.cn](mailto:weihua@semi.ac.cn)

S. Kasai

RCIQE, Hokkaido University, Sapporo, Hokkaido 0608628, Japan  
e-mail: [kasai@rciqe.hokudai.ac.jp](mailto:kasai@rciqe.hokudai.ac.jp)

microwave and optical technologies, the THz frequency range remains one of the most elusive regions of the electromagnetic spectrum.

The low-dimensional semiconductor systems are one of most suitable candidates for working at the THz frequency regime. The conducting electrons in low-dimensional systems are confined within nanoscale spacing so that the quantized energies fall into the meV scale, corresponding to THz frequency regime. If THz wave couples into this kind of low-dimensional semiconductor quantum devices, THz photon-assisted electrons will modify the tunnel current. Tien–Gorden theory has described the photon-assisted tunneling (PAT) in superconducting tunnel junctions [1], through which the tunneling of electron with energy  $E$  can exchange photon energy  $\hbar\omega$  with microwave field, creating a new set of electron states  $E + n\hbar\omega$ .

Based on this idea, a series of these THz-photon detectors have been implemented in semiconductor nanostructures such as semiconductor superlattices, quantum point contact (QPC) devices, resonant tunneling diodes (RTD), and single electron transistors (SET). In 1993, Wyss et al. theoretically predicted that a pronounced far-infrared photon-induced current ministeps can be detected by QPC device defined in two-dimensional gas (2DEG) by split-gate electrodes [2]. The QPC device generally exhibits the current plateau behavior of one-dimensional (1D) electron system. With the help of incoming far-infrared photons, the transport electron gain the photon energy so as to transmit the barrier near the classical turning points. In 1995, Keay et al. observed the absolute negative conductance and new current steps and plateaus in superlattice driven by intense THz electric fields at temperature range of 8–15 K [3]. Superlattice consists of alternating layers with periodic potentials along the superlattice axis, which give rise to the formation of minibands and are separated by minigaps. At low dc bias the drift of electrons in superlattice is reduced because of the frequency modulation of the Bloch oscillation by intense THz frequency field. At higher dc bias the THz field opens new conduction channels in the neighboring well and the current increases resulting in the formation of current steps. In 1999, Asada research group observed the gradual current change of triple-barrier RTD at room temperature under THz irradiation from the classical square-law detection to photon-assisted tunneling with increasing THz photon energy [4]. The THz operation of RTD was based on an inter-subband resonance and depended on the applied bias voltage producing the correct subband alignment of the two wells. Electrons tunnel from emitter electrode to 2D resonant states in the wells; subsequently, electrons leave the well by tunneling through the collector barrier. In addition to the conventional resonant tunneling peak where the subband energy levels in the two quantum wells are aligned, two additional peaks appear due to THz photon-assisted tunneling with photon absorption and stimulated emission. These two peaks are located at the bias voltages where the photon energy equals the difference between the subband energy levels in the two quantum wells.

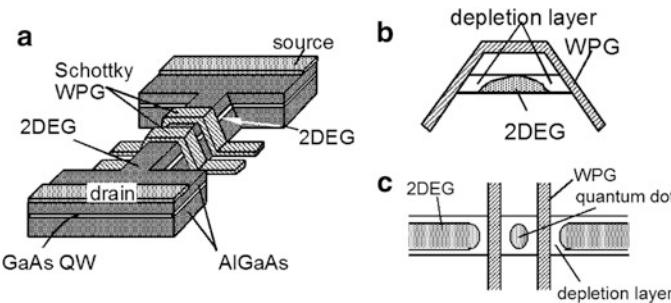
SET is known as a high sensitive quantum dot (QD) detector for THz Photons. It consists of a small conductive island weakly tunnel coupled to two reservoirs (source and drain). Coulomb blockade effect results in an oscillating behavior of

the conductance through the dot as a function of gate voltage. The energy to add an extra electron to a QD constitutes the charging energy  $e^2/C$  for a single electron and a finite energy difference  $\Delta\epsilon$  arising from the confinement. A tunneling current can flow through the QD when a discrete energy state is aligned to the Fermi energies of the leads. In 1994 Kouwenhoven demonstrated the effect of PAT through a GaAs/AlGaAs QD in the presence of a microwave signal [5]. When high frequency voltages drop across the two barriers, additional current peaks appear. Microwave radiation is treated as oscillating potentials across the source and drain tunnel junctions. Different ac amplitudes across the source and the drain barriers can be used to allow for an asymmetry in the coupling of microwaves to the dot. The ordinary Coulomb blockade oscillations are modified by the application of high frequency radiation. The photon energy acquired by the electrons allows them to tunnel through the states of higher energy above the Fermi energy, which are otherwise not accessible. The sidebands originate from matching the ground and excited states to the Fermi levels of the leads by a photon energy  $\hbar\omega$ . Such a QD detector offers the possibility to tune the discrete states of the dot easily by varying electro-static potentials, thus allowing a frequency-selective detection.

The QD detector of THz photon was developed by Komiyama using a cold (50 mK) SET in a high magnetic field in 2000 [6] and double QDs without magnetic field in 2002 [7]. Under THz illumination, an excited charge on a QD is sensed by a nearby SET. The mechanism is based on the polarization of charges in quantum dot created by far-infrared photon, which caused the conductance resonance peaks to shift in SET. Incident THz photons are coupled into the quantum dot via small dipole antennas. Within the quantum dot, an electron–hole pair created by the incident photon releases energy to the crystal lattices, which causes a polarization between two closely coupled electron reservoirs. Electron tunneling occurs, causing a shift in the gate voltage of the SET.

## 15.2 Schottky Wrap-Gate Single Electron Transistors

In recent years, a variety of fabrication methods for tunable semiconductor quantum dots has been reported. The most common scheme consists of patterning metallic top gate electrodes by electron beam lithography. For the realization of III-V compound semiconductor-based single electron circuit, it is important to develop a kind of effective SET with suitable structure for high-density planar integration by simple fabrication process. The conventional split gate control of 2DEG produces a rather weak and gradual confinement potential with soft-wall boundaries. In order to realize stronger confinements, a novel Schottky wrap-gate (WPG) SET was fabricated by the processes that double nano-sized Schottky gates wrapped around a GaAs-based heterostructure nanowire in a narrow spacing. Its advantages include not only stronger confinement potential for electrons but also high gate controllability and flexibility of device design.



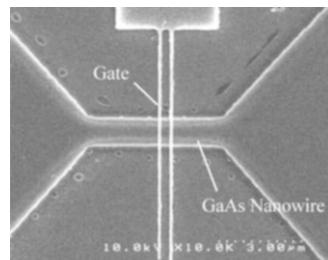
**Fig. 15.1** The schematic structure of (a) WPG SET, (b) the depletion of 2DEG in channel, and (c) a dot sandwiched by double barriers

The structure of WPG SET is a GaAs-based etched two-dimensional electron gas (2DEG) nanowire with two Schottky wrap gates around it, as Fig. 15.1a shows. The spacing distance of two wrap gates is about 100–300 nm. The gate width is about 100–200 nm. The width of nanowire is about 200–800 nm. With suitable negative voltages of double Schottky wrap gates, electrons under gates are depleted to form a quasi-one-dimensional electron gas in a short central channel (i.e., a quantum point contact structure) in Fig. 15.1b. Further negatively increasing the gate voltage, the electrons along the channel will be pinched off completely to form double barrier potentials. Then a quantum dot in Fig. 15.1c is sandwiched between both barriers. The size of quantum dot can be adjusted flexibly by double gates so that the energy level position and level spacing can be tuned effectively. The smaller the quantum dot become, the thicker double barriers need to be. Thus, only few current peaks resulting from resonant tunneling can be observed near the pinch-off gate threshold voltage.

### 15.3 Device Fabrications and Measurement

The WPG SETs were fabricated on a  $\delta$ -doped GaAs/AlGaAs heterostructure wafer. The GaAs/AlGaAs heterostructure wafer was grown by molecular beam epitaxy, consisting of a semi-insulating GaAs substrate, a 500-nm-thick buffer layer of undoped GaAs, a 100-nm-thick layer of undoped  $\text{Al}_x\text{Ga}_{1-x}\text{As}$ ,  $x = 0.33$ , a 20-nm-thick layer of undoped GaAs, a 10-nm-thick layer of undoped  $\text{Al}_x\text{Ga}_{1-x}\text{As}$ ,  $x = 0.33$ , a  $\delta$ -doped layer with carrier density  $3 \times 10^{12} \text{ cm}^{-2}$ , a 50-nm-thick layer of undoped  $\text{Al}_x\text{Ga}_{1-x}\text{As}$ ,  $x = 0.33$ , and a 10-nm-thick GaAs cap layer doped with Si at  $1 \times 10^{18} \text{ cm}^{-2}$ . Figure 15.2 shows a scanning electron microscope (SEM) image of the fabricated SET. The SET consists of a GaAs-based heterostructure nanowire surrounded by double nanoscale Schottky gates. The nanowires were patterned on top of the GaAs/AlGaAs heterostructure wafer chip using electron-beam

**Fig. 15.2** The SEM image of a fabricated Schottky-wrap gate SET



lithography and developing procedure. Isotropic wet etching in a cooling  $\text{H}_2\text{SO}_4:\text{H}_2\text{O}_2:\text{H}_2\text{O}$  solution was carried out to transfer the nanowire pattern onto GaAs/AlGaAs heterostructure. The two-dimensional electron gas (2DEG) which forms at the GaAs/AlGaAs heterointerface is located 70 nm below the sample surface and has a mobility  $\mu = 1.19 \times 10^5 \text{ cm}^2/\text{V s}$  and sheet electron density  $n_s = 6.43 \times 10^{11} \text{ cm}^{-2}$  at the low temperature 77 K. From these data we can extract a mean free path of  $l_e = 25.8 \mu\text{m}$  and Fermi wavelength in the 2DEG region of  $\lambda_F = 31 \text{ nm}$ . The two investigated devices are with the nanowire width of 680 nm and 580 nm, respectively. The ohmic contact of source and drain electrodes was fabricated by following procedures including photolithography, the evaporation of multiple metal layer Ge/Au/Ni/Au, lift off, and annealing at 430 °C. E-beam overlay exposure was used to pattern double gate stripes. The gate length is 100 nm and the gate spacing is 200 or 300 nm for different devices. Metal Cr/Au deposition and liftoff procedures transfer the nanoscale gate pattern onto the nanowire mesa. Schottky barriers of about 0.7 eV were formed between Cr and GaAs interface to stop leak current. Then  $\text{SiO}_2$  layer with the thickness of 30 nm covered the whole surface of sample for the passivation of surface states. At last, for electronic tests, electrode down-leads were impressed on the electrode windows of  $\text{SiO}_2$  layer.

The devices have been shielded in a vacuum test box in order to screen the disturbance of environment noises. In order to avoid heating electrons by electric field, the source/drain voltage is kept 0.2 mV, which is below  $k_B T/e \approx 0.862 \text{ mV}$  at 10 K. The current–voltage characteristics between source–drain current and gate voltage have been measured at low temperatures using Agilent semiconductor parameter analyzer. The detection experiments were performed by measuring the change of source–drain current with varying gate voltages in SET at temperature 6–7 K under the normal incident 2.54 THz far infrared laser, which radiates from  $\text{CH}_3\text{OH}$  gas pumped by  $\text{CO}_2$  laser. The size of conductive mesa is of the order of several hundred microns, comparable to the THz wavelength of 118 μm. The alternate THz electric field is perpendicular to THz wave propagating. Therefore, the normal incident THz laser wave is very helpful to induce electrons confined in the nanowire and the quantum dot. By adjusting the gate voltage, the change of level spacing in quantum dot allows tunable THz photon detection more effective. THz photocurrent response was obtained from the difference between source–drain

current with and without THz laser irradiation. In order to exclude the influence of random noises, source-drain current was measured repeatedly using a fine scanning step (0.01 mV) of the gate voltage. Reproducible response results could confirm the photocurrent closely relates to the absorption of THz photons.

## 15.4 THz Detection by GaAs-Based Schottky Wrap-Gate SET

At finite temperature the current peak of single electron resonant tunneling could be given by a Landauer-type formula for symmetric barriers [8],

$$I_{DS}(V_G) = G_0 \cdot V_{DS} \cdot \frac{1}{4k_B T} \int_{-\infty}^{\infty} \cosh^{-2}\left(\frac{E}{2k_B T}\right) \frac{\Gamma^2}{[E - e\alpha(V_N - V_G)]^2 + \Gamma^2} dE, \quad (15.1)$$

with  $G_0 = Ae^2/\hbar$  and  $A$  the temperature-independent energy-integrated strength of resonance,  $T$  low temperatures,  $\Gamma$  the level broadening resulting from resonant tunneling, and  $V_N$  the gate voltage at the resonance. The scaling factor  $\alpha = \Delta E/e\Delta V_G$  and quantized energy  $\Delta E$  is determined by quantum level spacing  $\Delta e$  and charging energy  $U = e^2/2C$  ( $C$  the total capacitance). For sufficiently low temperatures ( $\Gamma < k_B T < \Delta e < U$ ), electron transport across the device with low source-drain bias is dominated by quantum tunneling through a single level. In this temperature range, the full width of resonance peak at half maximum, arising from the thermal broadening of the Fermi distribution of the electrons, is linear in temperature: FWHM =  $3.52k_B T/e\alpha$ .

The current peak line shape modified by THz photon-assisted tunneling is given by Tien–Gorden theory [1]. In this theory,  $n$  photon absorption or emission is viewed as creating a new set of electron eigenstates with energies  $E + n\hbar\omega$ , where  $E$  is the eigenenergy of the original electron state without irradiation, and  $n$  is negative integer for the absorption and positive integer for the emission. The time dependence of the wave function for every one-electron state will be modified according to

$$\begin{aligned} \psi(x, t) &= \psi(x) \exp\left[-\frac{i}{\hbar} \int^t dt (E + eV_{ac} \cos \omega t)\right] \\ &= \psi(x) \sum_{n=-\infty}^{\infty} J_n(eV_{ac}/\hbar\omega) \exp[-i(E + n\hbar\omega)t/\hbar] \end{aligned}, \quad (15.2)$$

where  $E$  is the unperturbed energy of the Bloch state,  $J_n$  is the  $n$ th order Bessel function of first kind,  $V_{ac}$  is ac-coupling voltage, and  $\hbar\omega$  is photon energy. The applied ac voltage is assumed to modulate adiabatically the potential energy for each quantum level, which are equivalent to dc voltages  $n\hbar\omega/e$  applied across the

barrier with the probability amplitude  $J_n(eV_{\text{ac}}/\hbar\omega)e^{-in\omega t}$  for transitions between the modulated state and any unmodulated eigenstates. Thus, source-drain current modulated by absorption of THz photons can be given by

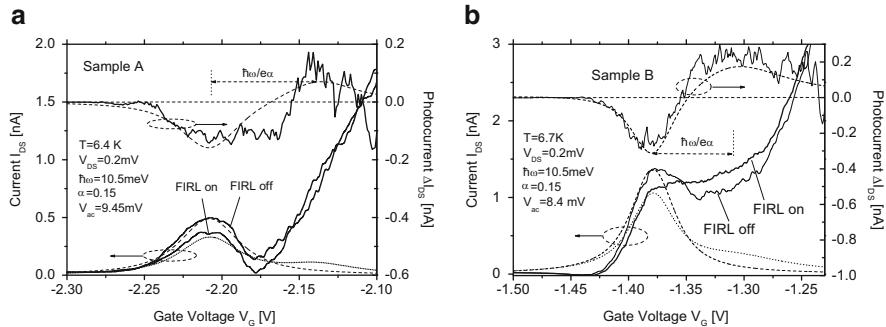
$$\begin{aligned}\tilde{I}_{\text{DS}}(V_G) = & G_0 \cdot V_{\text{DS}} \cdot \frac{1}{4k_B T} \int_{-\infty}^{\infty} \cosh^{-2} \left( \frac{E}{2k_B T} \right) \\ & \cdot \sum_{n=-\infty}^{\infty} \frac{J_n^2(eV_{\text{ac}}/\hbar\omega)\Gamma^2}{[E - e\alpha(V_N - V_G) + n\hbar\omega]^2 + \Gamma^2} dE.\end{aligned}\quad (15.3)$$

Therefore, THz photocurrent could be obtained from the difference of resonance peaks with and without THz irradiation, namely,

$$\Delta I_{\text{DS}}(V_G) = \tilde{I}_{\text{DS}}(V_G) - I_{\text{DS}}(V_G). \quad (15.4)$$

The response of THz photocurrent might provide a kind of spectrum in both frequency and intensity of THz wave. Since THz photocurrent is extracted from the resonance peaks of source-drain current, the separation of THz photocurrent is the most important for the spectrum. In order to obtain clear spectrum, the photon-assisted tunneling must be dominative in the electron quantum transport with the following challenges: (1) Charge energy must be larger than thermal energy ( $U > k_B T$ ) only because the operation of single electron tunneling through the SET depends on Coulomb blockade effect. (2) THz photon energy should exceed the level broadening (i.e.,  $\hbar\omega > \Gamma$ ) for the electrons on the quantum dot. When photon energy  $\hbar\omega > \Gamma$ , i.e.,  $\omega > 1/\tau$ , each electron experiences at least one cycle time of THz signal within the mean time  $\tau$  tunneling through double barriers [9]. If THz frequency  $\omega < 1/\tau$ , each electron on the dot can't go through a cycle time of THz field. In this adiabatic driving regime, electrons on the dot only see an essentially static potential and could not be induced by THz alternative electric field. (3) THz photocurrent might be observed clearly if THz photon energy is larger than thermal broadening of resonant levels  $\hbar\omega > 3.52k_B T$  [10]. However, THz photocurrents would be masked by thermal broadening of levels at higher temperatures. (4) The electrons absorbed THz photons should not release the THz photon energy easily by LO-phonon due to the strong quantum confinement in the nanostructure. The LO-phonon scattering is prohibited during electron tunneling process from electron reservoirs to quantum dot according to the Fermi-Golden rule and momentum selection rule.

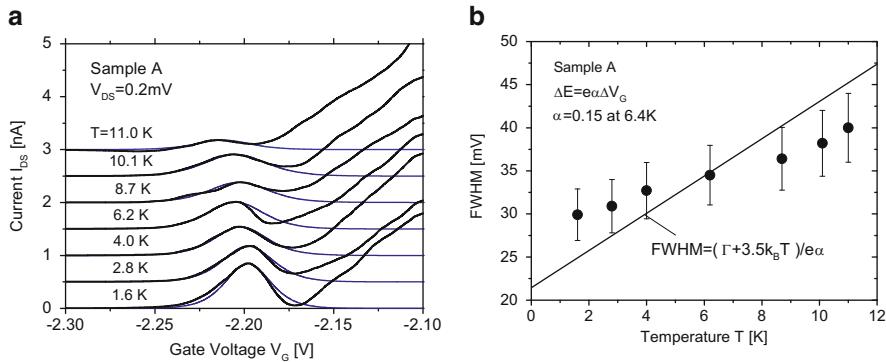
Reproducible THz photocurrents of two devices A and B with different quantum dot size have been obtained from the difference between source-drain current with and without THz laser irradiation, which shows in Figs. 15.3a and 15.3b respectively. Table 15.1 provides the confined space size of quantum dots for device A and B by comparing the scale parameters of nanowire and double gates. The resonant current peak and photocurrent peak in device B are more remarkable than that in device A. The improvement of local confinement in quantum dot can increase the level spacing to enhance the discreteness of resonance peaks. Figure 15.3 indicates that the increase in the local confinement of quantum dot is very useful for the



**Fig. 15.3** THz photocurrent response and source–drain currents with and without 2.525 THz normal incident far infrared laser (FIRL) irradiation for device A in (a) and device B in (b)

**Table 15.1** The structure parameters of device samples

Device no.	Nanowire		Gate	
	Width (nm)	Length ( $\mu\text{m}$ )	Length (nm)	Spacing (nm)
A	680	4	100	300
B	580	4	100	200



**Fig. 15.4** Temperature dependence of (a) current peaks and (b) their full width at half maximum (FWHM)

increase of THz response. If the photocurrent response results from the THz laser irradiation, the gate voltage difference ( $\Delta V_G = 70\text{ mV}$  for both devices) between resonance peak and satellite peak should be able to characterize THz photon energy (10.5 meV) by  $\hbar\omega = e\alpha\Delta V_G$ . In order to prove this assumption, the source–drain current with THz irradiation were simulated by Tien–Gorden theory for both devices. Simulated current curves agree well with characteristics of experimental source–drain current. Extracting from the curve fitting process, we obtained the

scaling factor  $\alpha = 0.15$  for both of devices. The value of the scaling factor is just what we hoped. Since the scaling factor is related to gate capacitance and total capacitance, the scaling factor  $\alpha$  indicates that both devices with the same gate length have the similar gate depletion layers. The scaling factor  $\alpha$  is very important for the scale of THz photon energy. The experiment on the temperature dependence of resonance peaks in sample A has been done for the verification of the scaling factor. The measured conductance characteristics as a function of temperatures are shown in Fig. 15.4a. With the increase of temperature, the resonant current peak width is basically broadened as  $(\Gamma + 3.52k_B T)/\alpha e$  in Fig. 15.4b. The level broadening of resonance peak  $\Gamma = 3$  meV, which extracts from the curve fitting process. The scaling parameter  $\alpha$  was estimated as 0.146 at 6.4 K. This result indicates that the satellite peak is exactly generated by THz photon.

## 15.5 Conclusions

Reproducible THz photocurrents have been detected by GaAs-based wrap-gate single electron transistor at low temperature. Analysis based on Tien–Gorden photon-assisted tunneling theory shows that the THz photon energy could be characterized by the gate voltage spacing between the resonance peak and satellite peak. Temperature dependence of resonance peak further verifies that the satellite peak is exactly generated by THz photon. Both experimental results and theoretic analysis show that the narrow spacing of double barriers is the most useful for the enhancement of THz response at higher temperature.

**Acknowledgments** The authors gratefully acknowledge the discussions with Prof. H. Hasegawa and Prof. F. H. Yang and the experimental help by Dr. M. Yumoto. This experimental work was supported by the Research Center of Integrated Quantum Electronics, Hokkaido University, Japan.

## References

1. Tien, P.K., Gordon, J.P.: Multiphoton process observed in the interaction of microwave fields with the tunneling between superconductor films. *Phys. Rev.* **129**, 647–651 (1963)
2. Wyss, R.A., Eugster, C.C., Alamo, J.A.D., Hu, Q.: Far-infrared photon-induced current in a quantum point contact. *Appl. Phys. Lett.* **63**, 1522–1524 (1993)
3. Keay, B.J., Zeuner, S., Allen, J.S.J., Maranowski, K.D., Gossard, A.C., Bhattacharya, U., Rodwell, M.J.W.: Dynamic localization, absolute negative conductance, and stimulated, multiphoton emission in sequential resonant tunneling semiconductor superlattices. *Phys. Rev. Lett.* **75**, 4102–4105 (1995)
4. Sashinaka, N., Oguma, Y., Asada, M.: Observation of terahertz photon-assisted tunneling in triple-barrier resonant tunneling diodes integrated with patch antenna. *Jpn. J. Appl. Phys.* **39**, 4899–4903 (2000)
5. Kouwenhoven, L.P., Jauhar, S., Orenstein, J., McEuen, P.L., Nagamune, Y., Motohisa, J., Sakaki, H.: Observation of photon-assisted tunneling through a quantum dot. *Phys. Rev. Lett.* **73**, 3443–3446 (1994)

6. Komiyama, S., Astaev, O., Antonov, V., Kutsuwa, T., Hirai, H.: A single-photon detector in the far-infrared range. *Nature* **403**, 405–407 (2000)
7. Komiyama, S., Astafiev, O., Antonov, V., Kutsuwa, T.: Single-photon detection of THz-waves using quantum dots. *Microelectron. Eng.* **63**, 173–178 (2002)
8. Foxman, E.B., McEuen, P.L., Meirav, U., Wingreen, N.S., Meir, Y., Belk, P.A., Belk, N.R., Kastner, M.A.: Effects of quantum levels on transport through a Coulomb island. *Phys. Rev. B* **47**, 10020–10023 (1993)
9. Büttiker, M., Landauer, R.: Traversal time for tunneling. *Phys. Rev. Lett.* **49**, 1739–1742 (1982)
10. Hu, Q.: Photon-assisted quantum transport in quantum point contacts. *Appl. Phys. Lett.* **62**, 837–839 (1993)

# Index

## A

Analog circuits, 152, 248  
Aspect ratio, 131–145  
Atomic functionality, 320

## C

Carrier mobility, 25–50, 170, 272, 274  
Characteristic fluctuation, 126, 128, 145–155  
Compact modeling, 264–271  
Correlation, 9, 26, 56, 99–121, 162, 164, 165, 173, 179, 182  
Coulomb blockade (CB), 216, 228, 285–301, 318, 326, 328–330, 332, 352, 353, 357

## D

Dopant distribution, 167, 237, 243, 306  
Dopant quantum dot, 305–307, 311  
Double gate MOSFET, 82  
2D potential profile, 94  
Drain extended, 247–261  
Dynamic characteristic fluctuation, 132, 150

## E

Electrical characteristics, 66, 125, 131, 140, 145, 156, 228, 232–243, 305, 306, 308, 312, 318  
Enhanced quantum effects, 285–301  
Ensemble Monte Carlo simulation, 108

## F

Fin doping, 56, 66–73, 76, 176, 190, 254  
Fin field-effect transistor (FinFET)  
    reliability, 66, 206, 252, 261, 264, 286  
    modeling, 3, 60, 81–95, 101–104, 263–282  
Fourier synthesis, 102, 166, 173, 178

## G

GaAs/AlGaAs, 353–355  
Gate stack, 56–58, 60, 64, 72–76, 100, 152, 216, 232  
Gate-all-around (GAA), 38, 50, 174, 186, 205–216, 220, 229, 276, 327, 328, 331, 332, 334, 336, 338, 343, 353–359  
Graphene  
    double quantum dot, 326, 336–343  
    quantum dot, 326–336, 339  
    single electron transistor, 327–331

## H

High voltage, 248–253, 261  
High-k/metal gate stack, 152  
Hot carrier effect (HCE), 264, 271–275, 281

## I

In situ doping, 233, 238, 239  
Independent double-gated (IDG), 227–243  
Inversion-mode (IM), 67, 69, 159, 160, 175–189, 197–199  
Inverter, 129, 130, 132, 133, 137–140, 145, 149–151, 156, 271

**J**

Junctionless (JL), 69, 159, 160, 175, 183, 189–197, 200, 309

**L**

Lattice relaxation, 214, 216, 217, 221

Line edge roughness (LER), 100–106, 112, 121, 146, 160–167

Low temperature, 5, 60, 71, 228, 230, 291–296, 300, 306–312, 315, 318, 320, 326, 328, 329, 348, 355, 356, 359

Low temperature characterization, 228

**M**

Mobility, 25–50, 55–76, 94, 101, 120, 128, 134, 135, 153, 170, 177, 191, 192, 197, 230, 234, 239, 255, 272, 274, 327, 343, 355

Mode-space representation, 3, 14–21

Multi-gate FET, 174, 205

Multi-phonon process, 205, 215–217, 221

Multi-switching functionality, 301

Multiple-gate architecture, 81

**N**

Nanoscale

device, 2, 94, 100, 116

silicon dot, 292

transistor simulation, 311, 313, 314

Nanowire, 25–50, 189, 205–221, 227–243, 265, 287, 289, 315, 343, 353–355, 357, 358

Non-equilibrium Green's function formalism (NEGF), 2, 4–10, 12–21

**O**

Oxide thickness fluctuation (OTF), 152, 153, 160, 172–173, 269

**P**

Photon-assisted tunneling (PAT), 352, 353, 356, 357, 359

Polycrystalline Silicon (Poly-Si), 60, 72, 228, 232, 234, 236, 238, 241, 243, 287–289

Process variation effect (PVE), 126, 129, 146, 152–155

**Q**

QM modeling, 92

Quantum

confinement, 3–4, 16, 17, 25–50, 61, 81, 90, 94, 101, 170, 205–221, 229, 230, 265, 275, 276, 278, 279, 281, 282, 286, 291, 292, 300, 315, 332, 357

dot, 285, 305–307, 311, 326–348, 352–355, 357

transport, 2–4, 293, 326, 331, 343, 348, 357

Quasi-planar, 82, 132–136, 138, 140–150, 153, 155, 156, 229

**R**

Random dopant fluctuation (RDF), 2, 66, 100, 126–131, 147–153, 155, 156, 160, 167–174, 176, 183–186, 188, 190, 195–200, 254

Random telegraph noise (RTN), 205–221

Real-space representation, 4, 6–9, 14, 21

Room-temperature charge stability, 290, 293–296, 301

**S**

Sensitivity analysis, 107–109, 111, 112, 121

Short channel effects (SCEs), 2, 26, 82–84, 100, 113, 131, 135, 140, 178, 186, 191, 194, 197, 227, 229, 230, 233, 235, 243, 264, 326

Si nanowire FET (SNWT), 206–218, 220, 221

Silicon-on-insulator (SOI), 26, 56, 57, 62, 64, 67, 68, 72, 76, 82, 83, 126, 131, 140, 174, 176, 189, 193, 207, 228, 254–256, 287, 289, 297, 309, 316, 319

Simulation, 1–22, 26, 60, 62, 82, 85, 89, 90, 95, 100–105, 108–110, 113, 118, 119, 121, 127, 128, 131, 132, 135, 137, 142, 145, 146, 153, 155, 160, 162, 166, 169–171, 173, 176–180, 184, 186, 188, 189, 192, 230, 234, 237, 253, 255, 259, 264, 265, 267–271, 277, 278, 280, 281, 311, 313–316, 318, 320

Single electron transistor, 243, 285, 286, 306, 325–348, 351–359

Single-dopant electronics, 305–321

Single-electron transport, 308, 312, 314

Single-electron tunneling, 286, 290, 296, 305–321

SNM model, 26, 27, 33, 36, 49, 50

SNWT. *See* Si nanowire FET (SNWT)  
Stacked devices  
Static random access memory, 128, 145,  
  156, 188  
Statistical analysis, 104–106  
Strain, 34, 37, 56, 61–66, 76, 264, 265,  
  275–282, 287, 298  
Subthreshold swing (SS), 131, 135, 140, 176,  
  190, 228  
Surface orientation, 56–62, 69–72, 74–76  
System on Chip (SoC), 247–261

**T**

Technology scaling, 160, 196, 197, 250  
Terahertz photon, 357  
Thin-Film Transistor (TFT), 228  
Threshold voltage, 2, 4, 66, 71, 82, 87, 92, 93,  
  95, 100, 101, 104, 105, 107–113, 115,  
  117, 118, 120, 121, 126, 132, 135, 141,  
  145, 148, 150, 156, 167, 173, 174,  
  176–178, 180–183, 186, 190, 193, 217,  
  230, 254, 255, 267–269, 272, 274, 275,  
  278, 282, 286, 290, 354

Threshold voltage fluctuation, 107, 111, 115,  
  126, 150  
Transfer characteristic fluctuation, 129, 132  
Tri-gate, 100, 132, 133, 135, 136, 138,  
  140–150, 152, 153, 155, 156, 174,  
  205, 206, 220, 229, 254, 255  
Tunneling, 3, 4, 214, 216, 229, 237–239, 243,  
  255, 286, 290, 293, 296, 305–321,  
  327–332, 336–339, 341, 345–347,  
  352–354, 356, 357, 359

**U**

Ultrascaled FinFET, 285–301  
Uniaxial strain, 264, 265, 275–277, 279, 281

**V**

Variability, 2, 99–121, 159–200

**W**

Work-function variation/variability, 99–121,  
  160, 173–174