Lining Zhang · Mansun Chan *Editors*

Tunneling Field Effect Transistor Technology

Springer

Tunneling Field Effect Transistor Technology

Lining Zhang · Mansun Chan
Editors

Tunneling Field Effect Transistor Technology

Springer

Editors
Lining Zhang
Hong Kong University of Science and Technology
Hong Kong
China

Mansun Chan
Hong Kong University of Science and Technology
Hong Kong
China

ISBN 978-3-319-31651-2 ISBN 978-3-319-31653-6 (eBook)
DOI 10.1007/978-3-319-31653-6

Library of Congress Control Number: 2016935208

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

Power consumptions have been a dominant constraint in nanoscale CMOS technologies. Different techniques to reduce computational power spanning from the architecture level to the fundamental semiconductor devices level are actively explored. One possible solution from the device perspective is to decrease the operation voltage without sacrifice of the switching properties. While its applicability was proved from the circuit theory, a lot of efforts in the electron device society have been gathered on devices with possible steep slopes that go beyond the traditional MOSFETs. Tunnel field-effect transistors are one representative of the steep slope devices. Their operations are based on the controlled switching of quantum tunneling, instead of the thermionic emissions. Historically, the study of similar concept may date back to the 1970s when the physicist used the gated tunnel junction to study the two-dimensional electron gas. Later, in the 1980s the interband quantum tunneling was observed in a DRAM trench transistor and people started to think about a device concept based on controlled interband tunneling. More device proposals followed in the 1990s. In 2004, a steep slope of 40 mV/dec was observed in carbon nanotube transistors and was attributed to the band-to-band quantum tunneling. At almost the same time people were intensively looking for solutions of the ever-increasing CMOS power problem. Since then the tunneling field-effect transistor (TFET) or devices with equivalent mechanisms but different names bloomed and attracted wide attention from the electronic device community as a promising low power device. Till date, TFET is an active research topic and is attracting attention from the industry for further development.

Knowledge sharing among different researchers, including people working on the device process, people working on device physics and modeling, people working on circuit designs, and people working on new materials and physics is an essential accelerator to incubate the technology and push it from research to applications. There are seven chapters in this book covering the TFET fabrications, TFET modeling, and also simulations of the TFET-based circuit design techniques. Chapter 1 covers a review of the steep slope devices including TFET. A holistic review on the research background and six kinds of steep slope devices are

provided. After brief introductions to each device's operations and the latest advances, a more detailed discussion of the TFET operation and several TFET performance boosters are summarized. Chapter 2 reviews the fabrication process and characterization methods of a variety of TFETs. Starting from the conventional lateral p-i-n TFET, the chapter discusses the tunnel junction formations including the doped junctions and the doping-less electron-hole bilayer. Going forward, the chapter summarizes the TFETs of homojunction and heterojunction, with material systems from Si/Ge, III–V compound semiconductors to the latest transition metal dichalcogenides. Characterization methods of the TFETs threshold voltages and subthreshold swings are provided. Chapter 3 discusses the compact models of TFETs. After providing a brief review of the TFET modeling in the literature, a complete SPICE model including the descriptions of current-voltage and charge-voltage characteristics are formulated based on detailed investigations of the TFET operations. Advanced effects in TFETs like the gate leakage and short channel effects are further discussed toward a full compact model. Challenges in the heterojunction TFET modeling are briefly discussed. Chapter 4 focuses on the challenges and designs of TFET-based digital circuits. Although promising for low voltage operations, TFETs have unique properties like unidirectional conduction, delayed saturation, enhanced Miller capacitance, imbalanced complementary logic, and larger variations. After describing these design challenges, the chapter proposes the all *n*-type pass-transistor logic to bypass the imbalanced complementary issue and the dual oxide device design to mitigate the issues due to enhanced Miller capacitances. Designs of the SRAM are investigated with a proposal of hybrid TFET–MOSFET cell. Chapters 5–7 cover more fundamental physics properties and the device designs of advanced TFETs. Chapter 5 reviews two atomistic simulation methodologies, namely the density functional theory (DFT) and tight binding (TB) within the Keldysh nonequilibrium Green's function (NEGF) framework. A new nonequilibrium vertex correction method is integrated with the NEGF-DFT to study disorder scattering in graphene TFETs. The NEGF-TB method is demonstrated by simulating the electric characteristics of a monolayer transition metal dichalcogenide TFET. Chapter 6 introduces another atomistic simulation method, the reduced-order $\boldsymbol{k} \cdot \boldsymbol{p}$ method, to accelerate the three-dimensional quantum transport study of TFETs. Basic theoretical background of the eight-band $\boldsymbol{k} \cdot \boldsymbol{p}$ Hamiltonian and the reduced-order NEGF equation, together with the spurious band elimination are described. The method is used to study the InAs-based homojunction TFET and the GaSb/InAs heterojunction TFET. Chapter 7 covers the device designs and optimizations of the carbon nanotube TFETs with the NEGF-TB method. After introducing the basic carbon nanotube properties, the chapter goes on to discuss device operation mechanisms. Doping engineering and gate dielectric engineering are developed to enhance the TFET performances. A barrier-controlled TFET is also proposed theoretically based on the atomistic simulations.

We are deeply grateful to all the chapter authors for their great efforts and outstanding chapters. When initiating this book on the tunneling field-effect transistor technology, all authors agreed that it was the right time to review the research

efforts on TFETs of the past decade and to gather together the latest research results. Bearing this in mind, every author spent their valuable time as a promise to make a comprehensive, authoritative, insightful, and up-to-date book for the purpose of knowledge sharing and dissemination. We sincerely hope that this edited book can serve as a platform for readers to have access to the current full frame of the tunneling field-effect transistor technology and to stimulate further interests into the next stage.

Clear Water Bay, Hong Kong

Lining Zhang
Mansun Chan

Contents

1 Steep Slope Devices and TFETs 1
Lining Zhang, Jun Huang and Mansun Chan

2 Tunneling FET Fabrication and Characterization 33
Tao Yu, Judy L. Hoyt and Dimitri A. Antoniadis

3 Compact Models of TFETs. 61
Lining Zhang and Mansun Chan

4 Challenges and Designs of TFET for Digital Applications 89
Ming-Long Fan, Yin-Nien Chen, Pin Su and Ching-Te Chuang

5 Atomistic Simulations of Tunneling FETs 111
Fei Liu, Qing Shi, Jian Wang and Hong Guo

6 Quantum Transport Simulation of III-V TFETs with Reduced-Order $k \cdot p$ Method 151
Jun Z. Huang, Lining Zhang, Pengyu Long, Michael Povolotskyi and Gerhard Klimeck

7 Carbon Nanotube TFETs: Structure Optimization with Numerical Simulation 181
Hao Wang

Index 211

Chapter 1
Steep Slope Devices and TFETs

Lining Zhang, Jun Huang and Mansun Chan

Abstract Reducing energy dissipations per function with the integrated circuit (IC) chips is always an appealing research topic. Techniques in the fundamental electronic device levels are being pursued besides of those in the architecture level. In this chapter, we introduce several device candidates with a common feature of steep slope as possible solutions for lower power computations. The ever increasing power densities with the complementary metal-oxide-semiconductor (CMOS) technologies and the behind reasons are reviewed first. Implications are reached that a device with steep slopes beyond the Boltzmann limitations helps. Then, several devices realizing steep slopes beyond that of the MOS field-effect-transistor (FET) technology are introduced, including the impact ionization FETs, the electro-mechanical FETs, the piezoelectric transistor, the ferroelectric FETs, the feedback FETs, and the tunneling FETs (TFETs). Afterward, we analyze the key features of the basic TFET operations and characteristics in details. Finally, several widely studied performance boosters for the TFET technology are also reviewed from device structures to doping and material engineering.

1.1 Reducing the CMOS Power with Steep Slope Devices

1.1.1 The CMOS Power Problem

Following Moore's law, scaling of semiconductor devices has gone with a relentless cadence in the past half century. Thanks to the effort of dimension minimization, the transistor density or roughly function density in integrated circuits

L. Zhang (✉) · M. Chan
Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
e-mail: lnzhang@ieee.org

J. Huang
School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA

L. Zhang and M. Chan (eds.), *Tunneling Field Effect Transistor Technology*,
DOI 10.1007/978-3-319-31653-6_1

(IC) has increased four orders, while price per transistor or roughly cost paid to one function has decreased six orders. At the same time, the transistor speed has increased four orders. These revolutionary changes in the semiconductor technology have pushed us into the information age, and now into a fantastic mobile information age.

These benefits from scaling are accompanied by tremendous increases in the IC power densities. Historical data indicate that the power densities of central processing units (CPU) by Intel had been increasing almost exponentially from nearly 2 W/cm^2 of the i386 with the 1.5 μm process, to nearly 100 W/cm^2 of the Pentium IV with the 0.13-μm process [1], as shown in Fig. 1.1. If following the same trend, we can predict that the CPU power density may reach that of a nuclear reactor, a rocket nozzle quickly, which imposes a power bottleneck on the complementary metal-oxide-semiconductor (CMOS) technology. Considering wide applications of semiconductor devices in our modern life, an article in Forbes magazine in 1999 [2] reported that electronic communication and information processing account for 10 % of US electrical usage. Later in 2011, another Forbes magazine article [3] estimated that cloud computation/storage facilities' share of US electrical usage is more than 10 %. This huge energy consumption by the IC chips is also named as the CMOS power crisis.

Bearing this huge CMOS power consumption in mind, one natural question to ask is, what is the physical limit on the energy dissipation of information processing? The differences between the physics law and the reality will create opportunities for us to overcome the power crisis. Actually, the issues of physical limitations on the silicon CMOS technology have been studied widely [4–7]. Meindl et al. [5] derived that the limit on the energy consumption in a binary switching of a metal-oxide-semiconductor field-effect transistor (MOSFET) is

$$E_s = kT \ln 2 \tag{1.1}$$

by assuming a single electron device, where k is the Boltzmann constant and T is the temperature. Later, Wang et al. [7] considered the energy relaxation time (t_{re}) and revised Eq. (1.1) as

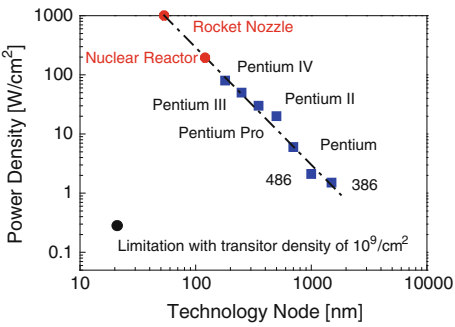

Fig. 1.1 Power densities of Intel's CPUs in history increase significantly with the CMOS scaling

$$E_s = \frac{1}{2}\alpha t_{re} + kT\ln 2 \tag{1.2}$$

where α is the energy switching speed. For example, with a 0.7-V operation and 1-GHz switching, the switching speed is $\alpha = 7 \times 10^8$ eV/s. While the energy relaxation time is in the order of picosecond, the second term in Eq. (1.2) is usually dominant. With the frequency increasing to around 52 GHz, the first term in Eq. (1.2) will be comparable to the second term, and the minimum energy dissipation is around 2kTln2. Assuming a transistor density of $10^9/cm^2$ and all the transistors are switching simultaneously at 52 GHz, the minimum power density is about 0.3 W/cm^2. In Fig. 1.1, this limitation is also plotted as a reference. Despite the above worst-case analysis, the derived power density is much smaller than those in real technologies. A huge room is there for reductions of the CMOS power density.

Till this end, it is necessary to know what caused the dramatic increase in the CMOS chip power density along with the technology scaling. In fact, the CMOS power crisis is a Gordian knot of continuous shrinking of MOSFET dimensions. Taking a CMOS inverter in Fig. 1.2 as an example, we can derive the power consumption as functions of transistors' size, operation voltage, and frequency.

At the fall edge of the input, the load capacitance C is charged by a current from the power source V_{dd}. The energy lost on the PMOS is $E_{pmos} = (1/2)CV_{dd}^2$, and the energy stored in the load capacitance is the same $E_{load} = (1/2)CV_{dd}^2$. At the following rise edge of the input, the energy stored in the capacitance E_{load} is lost through the NMOS. So the total energy consumption per switching period is $E_{total} = CV_{dd}^2$. The frequency (f)-dependent power consumption in reverting the output state is then $P_{dyn} = f \cdot CV_{vdd}^2$, which is defined as the dynamic power. In the steady input state, one leakage current I_{leak} from the power source to the ground leads to a second power consumption $P_{leak} = I_{leak}V_{dd}$, which is defined as the leakage power. The third contribution to the total power consumption is the short circuit current I_{sc}-induced power $P_{sc} = I_{sc}(t)V_{dd}$ which is present during the input/output flip. The total power consumption of an inverter is then given by

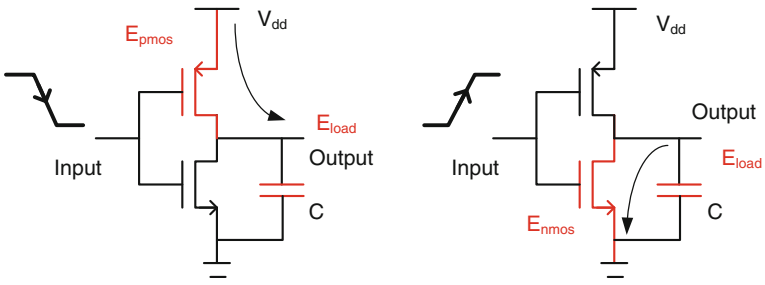

Fig. 1.2 Operations of the CMOS inverter. Energy is consumed during the binary switching

$$P = f \cdot CV_{dd}^2 + I_{leak}V_{dd} + I_{sc}(t)V_{dd} \tag{1.3}$$

With Dennard's scaling in the constant electric field scheme [8], we can calculate the power density of a CMOS chip with the technology scaling, i.e., the smaller transistor size, higher operational frequency, scaled gate oxide capacitance, and scaled operational voltages:

$$P = \alpha^2 \left[\alpha f \cdot \frac{1}{\alpha} C \frac{1}{\alpha^2} V_{dd}^2 + I_{leak} \frac{1}{\alpha} V_{dd} + I_{sc} \frac{1}{\alpha} V_{dd} \right] \tag{1.4}$$

We see that the dynamic power term is expected to be constant. However, in the practices the constant field scaling is not exactly adopted and the operation voltage scales slower than the device dimension. Figure 1.3 plots the operation voltage and threshold voltage scaling in multiple CMOS technology generations. There are several consequences. First, the dynamic power density along with the geometry scaling is increased due to the unmatched V_{dd} scaling. Second, the leakage power is increased due to the increase in the leakage current. Since the MOSFET sub-threshold current changes exponentially with gate voltages, the leakage current also increases exponentially with the threshold voltage V_{th} scaling.

We see that due to the deviations from Dennard's scaling, both the dynamic power and leakage power are increased. Figure 1.4 plots some statistics [10] on the dynamic and leakage power along with scaling, which actually explains the power density trend in Fig. 1.1.

1.1.2 Power Reduction with Steep Slope Devices

We notice in Fig. 1.3 that the supply voltage V_{dd} scaling cannot continue after it approaches 1.0 V from the 0.13-μm technology node. The unsustained voltage scaling further deteriorates the power consumption. However, the supply voltage trend is due to a physical limitation given below.

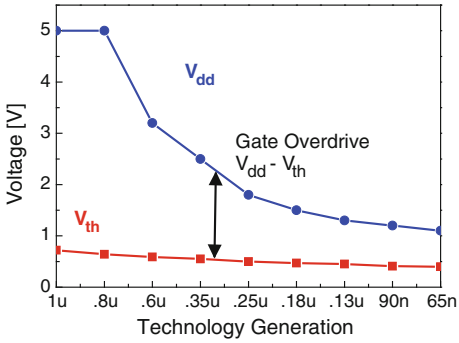

Fig. 1.3 The supply and threshold voltage scaling through the CMOS technology generations [9]

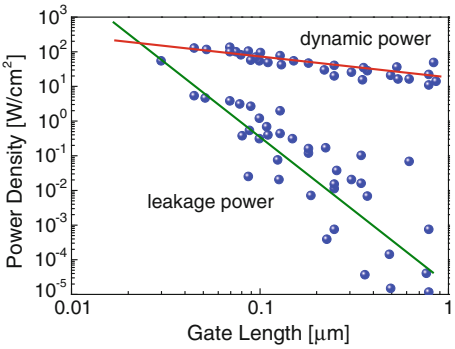

Fig. 1.4 The dynamic and leakage power density with the CMOS scaling

Figure 1.5 plots the general MOSFET switching properties. Separated by the threshold voltage V_{th}, there are the subthreshold and super-threshold regions. Instead of an ideal turn-off in the subthreshold region, the transistor's current is decreased gradually with decreasing the gate voltage. A term of subthreshold swing (SS) is defined as the gate voltage needed to change the transistor current by one order of magnitude. Supposing the threshold current is I_{th}, the off-state current is

$$I_{off} = I_{th} \exp\left(-\frac{2.3V_{th}}{SS}\right) \tag{1.5}$$

which is just the leakage current I_{leak} in Eq. (1.3). In order to control the leakage power component in Eq. (1.3), the threshold voltage V_{th} cannot be reduced too aggressively when SS is fixed. In the practical scaling [11–20], SS actually increases for few technology generations and then is maintained around 100 mV/dec for some time as shown in Fig. 1.6. This sets a limitation for the V_{th} scaling if the leakage current should be suppressed at some target level. It is also obvious from Eq. (1.5) that with reduced SS smaller V_{th} can be used to keep the same I_{leak}. In the state-of-the-art FinFET technology at 22-nm generation node, SS is reduced to around 65 mV/dec at room temperature [19]. This significant decrease in the SS allows further reductions of the threshold voltages.

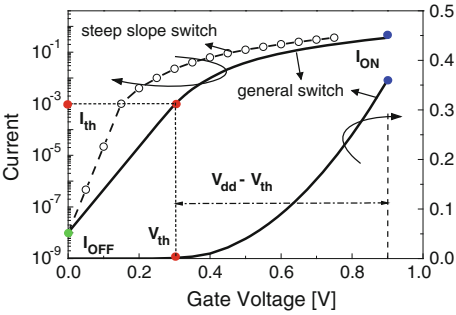

Fig. 1.5 Current–voltage characteristics of the general switch and a steep slope switch

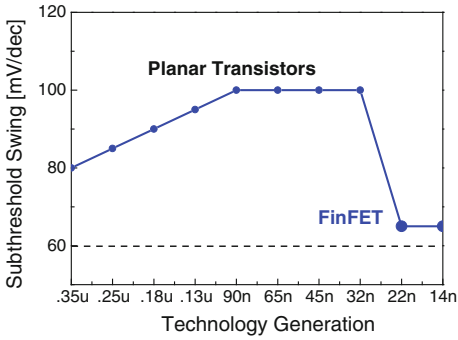

Fig. 1.6 Evolutions of the subthreshold swing in the CMOS technologies

The turn-on properties in the super-threshold region of Fig. 1.5 are approximated by

$$I_{on} = I_0(V_{dd} - V_{th})^2 \tag{1.6}$$

where I_0 is determined by the transistor materials and structures. The on-state current affects the charging/discharging duration in Fig. 1.2, hence the circuit speed. A simple model of the circuit frequency is written as

$$f \approx \frac{I_{on}}{CV_{dd}} = \frac{I_0}{C}\frac{(V_{dd} - V_{th})^2}{V_{dd}} \tag{1.7}$$

In order to increase the circuit operation frequency, either the first term or second term in right-hand side of Eq. (1.7) should be enhanced. It is obvious that an increase in V_{dd} with a fixed V_{th} leads to the monotonic increase of the second term. However, the requirements on V_{dd} posed by the power and speed are contractive to each other. Increasing the first term in right-hand side of Eq. (1.7) is one way to solve the power/speed conflicts. At the same time, V_{dd} scaling should be controlled in order to have an overall optimization of the device performances. Historically, V_{th} is roughly 1/3 of V_{dd} in optimized Si MOSFETs to induce enough on-state current I_{on} as well as reasonable off-state current I_{off} [21]. Concurrent considerations of the leakage power, dynamic power, and device/circuit speed lead to the voltage scaling results in Fig. 1.3.

Actually the voltage swing in the subthreshold region of the general switch in Fig. 1.5 is wasted since it does set a lower bound for the V_{dd} as shown in Eq. (1.7), also a lower bound for the energy consumption per circuit state switch. Keeping it in mind, one natural proposal is that SS should be scaled as aggressively as possible so that only a small V_{th} is necessary to maintain reasonably small leakage current in Eq. (1.5). One example of such a switch with smaller SS but the same I_0 is plotted in Fig. 1.5. This kind of device will avoid the conflicts between the requirements on V_{dd} from power and speed. A V_{dd} of 0.75 V can be used with both lower power and higher speed. In the extreme case where V_{th} is close to zero without any sacrifice in

the leakage current, V_{dd} of 0.4 V can be used to achieve the same speed as that of the general switch with V_{dd} of 0.9 V in Fig. 1.5. As a result, the power consumption will only be around 20 % of the general MOSFET switch. Although the physical power limit in Fig. 1.1 cannot be reached by just reducing SS, the proposal does help bringing down the CMOS power to a large extent. Devices that can achieve small SS are labeled as steep slope devices.

1.2 Steep Slope Devices

Figure 1.6 shows that a slope of 65 mV/dec is realized in the 22-nm and 14-nm technology generation by using the 3-dimensional fin-shaped MOSFET structure instead of the planar MOSFET structure. One nature question is, can even lower SS be achieved with the MOSFET technology?

The MOSFET is based on the drift–diffusion transport mechanism. An n-type MOSFET schematic is shown in Fig. 1.7, together with its operation principles. For planar transistors, the channel is doped with a different polarity from their source/drain. In the off state, there is a high barrier for electrons in the source to climb over and form the current conduction. By increasing the gate voltage, the barrier height is reduced and a hole depletion region at the channel surface is created. At the transistor drain side, a larger depletion actually exists due to the

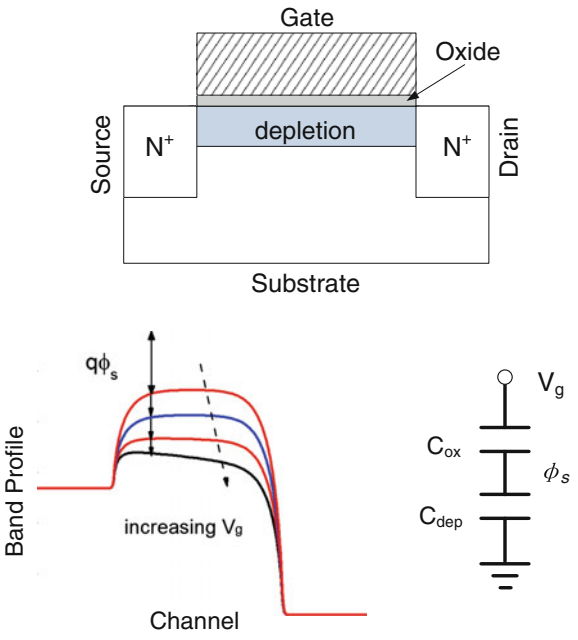

Fig. 1.7 The MOSFET schematic and its working principles

reversely biased drain–channel junction. According to the Boltzmann statistics, the electron density near the source is increased, more than that near the drain. Due to the diffusion of electrons from the source to drain, the current increases with gate voltage. Further increasing in the gate voltage leads to the super-threshold operations where the electrons drift will dominate. Let us check the subthreshold here. The physics quantity that corresponds to the barrier height is the surface potential. In the subthreshold operations, the surface potential is found by the simple capacitance model with gate oxide capacitance C_{ox} and depletion capacitance C_{dep} components:

$$\phi_s = \frac{C_{ox}}{C_{ox} + C_{dep}} V_g \tag{1.8}$$

Charge density at the source side is given by the Boltzmann statistics:

$$Q_s \propto \exp\left(\frac{q\phi_s}{kT}\right) \tag{1.9}$$

The subthreshold current is derived from the basic diffusion neglecting the drain side charge:

$$I_{sub} \propto \frac{W}{L} Q_s \tag{1.10}$$

Accordingly, the subthreshold swing SS is given by

$$SS = \left(1 + \frac{C_{dep}}{C_{ox}}\right) \times 2.3 \frac{kT}{q} / dec = \left(1 + \frac{C_{dep}}{C_{ox}}\right) \times 60 \text{ mV/dec} \tag{1.11}$$

In MOSFET technologies, both C_{ox} and C_{dep} are positive so the minimum SS is 60 mV/dec when C_{dep} is nearly zero. The scaling of SS in Fig. 1.6 is explained with Eq. (1.11) as follows. With the planar structures, we gradually increase the transistor channel doping through technology generations to suppress short-channel effects. The depletion width decreases and its capacitance increases. As a result, there is a gradual increase of the subthreshold swing. In the four generations from 90 to 32 nm, cooptimizations of the transistor channel doping and gate oxide thickness lead to similar C_{dep}/C_{ox}, so SS is maintained around a constant value. Going to the FinFET structure, requirements on the channel doping are eased greatly and the associated C_{dep} is reset to almost zero, leading to a sharp decrease of the SS to near 60 mV/dec. Another essential message we get from Eq. (1.11) is that SS cannot go below 60 mV/dec with the traditional technology due to the physics limitation from (a) the charge diffusion mechanism in Eq. (1.10) and (b) Boltzmann statistics in Eq. (1.9). New mechanisms shall be explored in order to go beyond this limitation and realize even steeper slope devices.

1.2.1 Impact Ionization FETs

Let us imagine a general physics process triggered by a control variable. When this variable is below its critical value, the process is not started. Once the critical value is reached, the process is suddenly initiated. In this sense, this critical variable value is a boundary for two distinctive operation regions. If the differences are significant enough, the transition from one region to another can be potentially used to represent two states of a switching device, e.g., the off and on states. By associating the control variable to a gate voltage, a steep slope can potentially be achieved.

Impact ionization (avalanche multiplication) in semiconductors is a first kind of such physical process. One carrier, either electron or hole, incident on a junction is accelerated by the electric field and can induce band-to-band excitations and generate electron–hole pairs if it gains enough energy. The ionization rate describes the number of electron–hole pairs per unit distance and is given by [22]:

$$a\left(\xi_{//}\right) = \frac{q\xi_{//}}{E_I}\exp\left[-\frac{\xi_I\xi_P}{\xi_{//}^2}\right] \tag{1.12}$$

where $\xi_{//}$ is the electric field that accelerates the carrier, E_I is the effective ionization energy and ξ_I is the ionization scattering threshold field, ξ_P is the phonon scattering threshold field. With a larger acceleration field, the possibility to excite ionizations increases significantly. Note that the impact ionization formulation Eq. (1.12) is a localized process which only depends on the local field. Electrons and holes may have different ionization rate, but both follow the above formulation.

The multiplication factor M_p describing the increase in carrier numbers due to impact ionizations is derived by simply assuming the same electron and hole ionization rates:

$$1 - \frac{1}{M_p} = \int_0^d a\left(\xi_{//}\right)dx \tag{1.13}$$

where x is the field direction and d is the boundary where the field vanishes. The current after the multiplications on the incident current I_0 is written as

$$I_p = M_p \cdot I_0 \tag{1.14}$$

If the electric field in the switching device is controlled by one device terminal so that the following condition is satisfied at certain terminal voltage:

$$1 = \int_0^d a\left(\xi_{//}\right)dx \tag{1.15}$$

It means the multiplication factor approaches infinity and the amplification on the current is extremely large. Essentials behind the above derivations are that the

generated electron–hole pairs can be accelerated again to induce new electron–hole pairs. The avalanche multiplication is a positive feedback process. The maximum electric field in the space (0, d) is usually defined as the critical field. In fact this avalanche mechanism has been used in the IMPATT devices and avalanche photodetectors. Before Eq. (1.15) is reached, the dependence of the current on electric field is given by Eq. (1.12)–(1.14).

The impact ionization field-effect transistors [23–25] make advantages of the above impact ionization process to realize sharp switching. The device schematic (n-type) is shown in Fig. 1.8, together with its working principle: the gate modulation effect on the potentials along the channel. Impact ionizations are initiated in the gate underlap regions. When the gate voltage is small, there is only a small amount of electrons transferred from the n-doped drain to the channel, and the electric field along the channel direction is low. The multiplication factor is almost unit and the current flowing from source to the drain is just the leakage current. The device is defined to be in its off state. By increasing the gate voltage, more electrons are transferred from drain to the channel and the electrostatic potentials in the gate covered channel are increased. As a result, the electric field across the gate underlap region is increased and ionization rate also becomes larger. No significant change in the current will be observed until the multiplication factor reaches a significantly large value. When Eq. (1.15) is approached at a certain gate voltage, several orders of magnitude changes in the current will be triggered. The device is switched to its on state. It is noted that the dependences of electrostatic potentials in the channel on the gate voltage are gradually weakened due to the screening effects from the

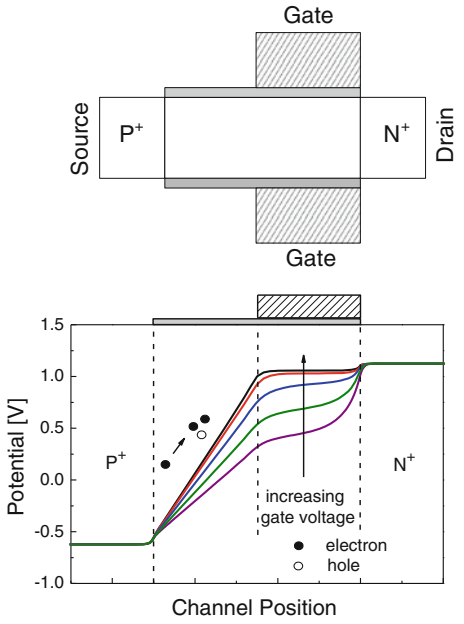

Fig. 1.8 Schematic of the impact ionization FET and its working principles

channel electrons. The electric field across the channel cannot be changed significantly after the avalanche multiplication. Compared to the MOSFET, electrostatics in the impact ionization FETs are similar to what Eqs. (1.8) and (1.9) describe. It is the amplification property of the avalanche breakdown Eq. (1.14) instead of the carrier diffusion Eq. (1.10) that leads to the steep slope.

Experimental realizations of impact ionization FETs had been reported in literatures [23–25]. A steep slope as small as 6 mV/dec [24] was achieved.

1.2.2 Electro-Mechanical FETs

In impact ionization FETs, the steep slopes are achieved by the current amplifications. If the surface potentials in Eq. (1.8) have sharp changes instead, steep slopes can potentially be obtained. From this perspective, electro-mechanical (EM) or nano-electro-mechanical (NEM) FETs were developed [26–30]. Although there are different configurations, the basic principle is the same: a bi-stable system with mechanical and electrostatic force can be transferred sharply from one state to another, triggered by a critical terminal voltage. Around this critical voltage, there is a sharp change in the surface potential and also a sharp change in the current, leading to significantly large current gain.

One example of the EM FETs is shown in Fig. 1.9, together with the physical explanations of the critical-state transition condition. The gate electrode can be

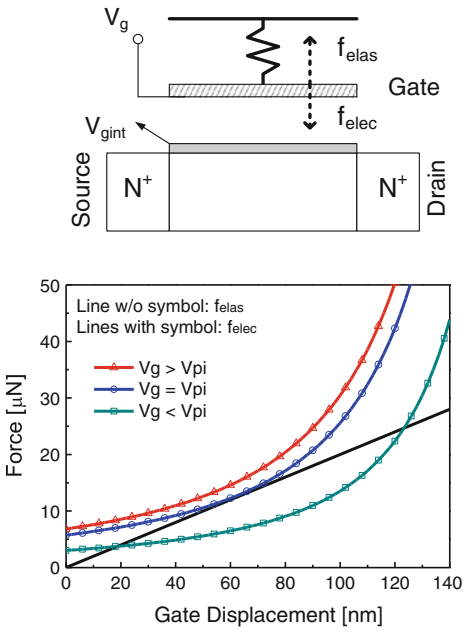

Fig. 1.9 Schematic of one electro-mechanical FET and its working principles

suspended with widely available techniques of the micro-electro-mechanical (MEM) process. When a gate voltage V_g is applied, there will be a voltage divider composed of the air gap capacitance and the gate oxide capacitance. The intrinsic voltage directly on the gate oxide V_{gint} is

$$V_{gint} = \frac{V_g}{1 + C_{ox}/C_{gap}} \tag{1.16}$$

As a result, there is an electric field-induced force (f_{elec}) on the suspended gate that tends to reduce the air gap. On the other hand, an elastic force (f_{elas}) is oppositely directional. The balance of these two forces determines the gate displacement d. Assuming the initial air gap distance is t_{gap0}, the spring elastic constant is k, the gate area is A, these two forces are formulated as

$$f_{elas} = k \cdot d \tag{1.17}$$

$$f_{elec} = \frac{\varepsilon_{gap}\left(V_g - V_{gint}\right)^2}{2} \frac{A}{\left(t_{gap0} - d\right)^2} \tag{1.18}$$

Figure 1.9 plots the two forces on the gate electrode for three given gate voltages applied on the EM device with the air gap thickness 190 nm. When increasing the gate voltage V_g, the system will be stabilized as indicated by the force curve crosspoint (the one with smaller displacement) when V_g is small. At certain gate voltage, the two crosspoints are reduced to one. Beyond this gate voltage, the electrostatic force is always larger than the elastic force, breaking the system balance and bringing down the gate electrode to directly sit on top of the gate dielectric. This critical gate voltage is defined as the pull-in voltage V_{pi}. Around V_{pi}, there is a jump of V_{gint} from that given by Eq. (1.16) to V_{pi} itself, inducing an amplification of the surface potential, hence the conduction current.

There is usually a hysteresis in the current–gate voltage characteristics of the EM FETs. A simplified view is given below. As shown in Fig. 1.9, the EM device is bi-stable under small gate bias. When decreasing the gate voltage from above V_{pi}, the EM device will firstly be stabilized at the crosspoint with larger displacement. The gate electrode moves to the oxide with further reduction of the gate voltage. At this stage, the current is still large. Until one of the crosspoint is larger than t_{gap0} in mathematics which is physically impossible, the EM device will switch to another stable state and be turned off.

Different types of the EM devices were developed in recent literatures. A suspended MOSFET with 2 mV/dec subthreshold swing was experimentally realized [28]. By turning to the accumulation mode FETs, the NEM FETs were proposed [27]. To facilitate the low-voltage and low-power applications, the micro-/nano-electro-mechanical relay switches and their logic were developed [29, 30] which eliminate the field-effect structures. They share the similar mechanism for the sharp switching between off and on sates.

1.2.3 Piezoelectric Transistor

Besides of the above EM FETs, another kind of steep slope device is potentially realized by using the mechanical stress-induced metal-semiconductor phase transition. The significant change in the conductance upon pressures can be used to represent the off and on states. If a transition with several orders of differences in the conductance can be achieved within a small terminal voltage range, then a steep slope will be obtained. A piezoelectric transistor (PET) has been proposed very recently [31] with the first demonstration [32]. It makes use of the internal transduction by converting the voltage into stress, and then to the conductance. Figure 1.10 shows the PET schematic.

A piezoelectric (PE) dielectric layer is placed between the gate and source electrodes. So the electric field from the voltage difference is turned into the expansion of the piezoelectric material. At the same time, the whole device structure is confined by a hard frame which is assumed to be strong enough so there is no strain at all. Then, the piezoresistance (PR) material is compressed generating a pressure. As a result, the conductance hence the current flowing from drain to source is increased due to the stress-induced resistance reduction. The mathematic description of the subthreshold swing is given below.

Assuming the piezoelectric material thickness is L_e and the piezoresistance material thickness is L_r, and there is no strain in electrodes, then their changes are

$$\Delta L_e = -\Delta L_r \tag{1.19}$$

With the PE material parameters, D_{33} the piezoelectric coefficient, S_{33_e} the compliance constant, and the induced stress T_{33_e}, together with the electric field E_{33}, the induced strain in PE is written as

$$S_{PE} = S_{33\_e} \cdot T_{33\_e} + D_{33} \cdot E_{33} = S_{33\_e} \cdot T_{33\_e} + D_{33} \frac{V_g}{L_e} \tag{1.20}$$

On the other hand, with the PR material parameter of S_{33_r} the compliance, T_{33_r} the stress, the strain is written as

$$S_{PR} = S_{33\_r} \cdot T_{33\_r} \tag{1.21}$$

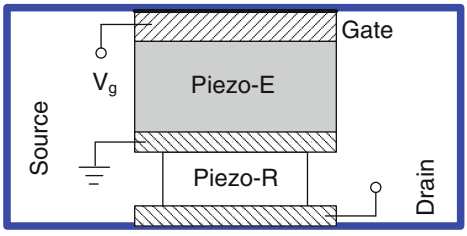

Fig. 1.10 Schematic of the piezoelectric transistor

With the cross section of PE material A, and PR material a, the equal force condition is

$$A \cdot T_{33\_e} = a \cdot T_{33\_r} \tag{1.22}$$

and the strain relationship is

$$L_e \cdot S_{PE} = -L_r \cdot S_{PR} \tag{1.23}$$

the transduction from voltage to PR stress is derived

$$T_{33\_r} = \frac{-V_g \cdot D_{33\_e}}{L_r \cdot S_{33\_r} + (a/A) \cdot L_e \cdot S_{33\_e}} \tag{1.24}$$

Assuming the resistivity of the PR material follows the function:

$$\log_{10}(R) = f(T_{33\_r}) \tag{1.25}$$

The subthreshold swing of the PET is obtained by combining Eqs. (1.24) and (1.25):

$$SS = \frac{L_r \cdot S_{33\_r} + (a/A) \cdot L_e \cdot S_{33\_e}}{D_{33\_e}} \cdot \frac{1}{f'(T_{33\_r})} \tag{1.26}$$

Materials with larger sensitivity of resistance on the stress help reducing the second term of right-hand side of Eq. (1.26), hence lead to steeper slope. For example, around 2-GPa stress can induce four orders of magnitude changes in the resistance of a typical PR material SmSe [32]. Meanwhile, by engineering the device geometry and choosing the proper PE/PR material combination, steep slope smaller than 60 mV/dec can potentially be achieved. For example, increasing the (A/a) ratio, reducing the PE/PR material thickness, and increasing the piezoelectric coefficients are effective ways for SS reduction. Device optimizations should consider the robustness, for example, (A/a) may have an upper limit.

1.2.4 Ferroelectric FETs

For the electro-mechanical FETs, it is the sudden change in the intrinsic gate voltage [hence the surface potential in Eq. (1.9)] due to the gap capacitance that induces the steep slope. As shown in Eq. (1.8), the surface potential is always smaller than the gate voltage in traditional MOSFETs since their oxide capacitance and depletion capacitance are always positive. If there are some physics processes that can break this limitation and allow the surface potential larger than the gate voltage, it can be clearly seen from the above discussions that the subthreshold

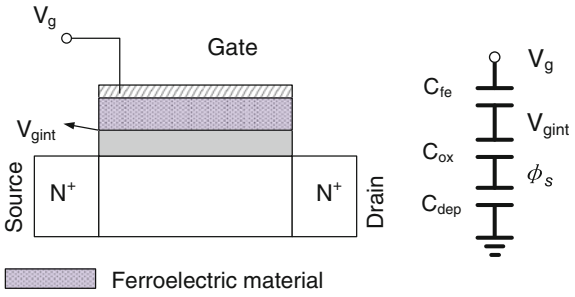

Fig. 1.11 Schematic of the ferroelectric FET and the voltage divider characteristics

swing may be brought down below 60 mV/dec. Such a kind of steep slope device was realized in the ferroelectric (FE) FETs [33]. Figure 1.11 shows the schematic of a FE FET.

In the FE FETs, a layer of ferroelectric material is added on top of the oxide dielectric of a normal MOSFET. The capacitance of the FE material is C_{fe}. Same as the voltage divider in the EM FETs above, an intrinsic gate voltage V_{gint} is induced at the interface of FE material and oxide, which controls the MOS channel surface potentials according to the classical theory Eq. (1.8). The simple voltage divider leads to the intrinsic gate voltage:

$$V_{gint} = \frac{V_g}{1 + C_{mos}/C_{fe}}, \; C_{mos} = \frac{C_{ox} \cdot C_{dep}}{C_{ox} + C_{dep}} \tag{1.27}$$

One unique properties of the ferroelectric material is that its capacitance can be negative within a bias window. As shown in Eq. (1.27), the intrinsic voltage is larger than the applied gate voltage V_g and still keeps positive when the absolute value of the ferroelectric capacitance is limited within the MOS capacitance. The negative capacitance is explained with the Landau–Khalatnikov equation [34] which describes the relationship between the areal polarization charge P, the external field E_{ext} and the Gibb's free energy U:

$$\nabla_p U = 0, \; U = \alpha P^2 + \beta P^4 + \gamma P^6 - E_{ext} \cdot P \tag{1.28}$$

where only the static condition is considered. α, β, γ are the ferroelectric material parameters, and $\alpha < 0$. One example of the ferroelectric material $BaTiO_3$ has the parameter of $\alpha = -10^7$ m/F. A straightforward derivation gives the external field as a function of the polarization charge:

$$E_{ext} = 2\alpha P + 4\beta P^3 + 6\gamma P^5 \tag{1.29}$$

By setting the MOSFET channel charge density Q = P, the two charge control equations across the MOS capacitance and the ferroelectric capacitances are given by

$$Q = C_{mos} \cdot V_{gint} \tag{1.30}$$

$$\frac{V_g - V_{gint}}{t_{fe}} = 2\alpha Q + 4\beta Q^3 + 6\gamma Q^5 \tag{1.31}$$

The voltage control equation is obtained when higher orders of the charge terms are negligible:

$$V_{gint} = \frac{1}{1 + C_{mos} \cdot 2\alpha t_{fe}} V_g \tag{1.32}$$

By combining Eq. (1.27) and (1.32), the negative capacitance property of the ferroelectric material is seen. Accordingly, the subthreshold swing SS is given by

$$SS = \left(1 + \frac{C_{dep}}{C_{ox}} + C_{dep} \cdot 2\alpha t_{fe}\right) \times 60 \text{ mV/dec} \tag{1.33}$$

Due to the large absolute value of the 'dielectric constants' of the ferroelectric material compared to oxide, a thicker layer is required to make the third term in Eq. (1.33) significant and to induce sub-60 mV/dec slope of the FE FETs. On the other hand, there is an upper limit of the ferroelectric layer thickness to avoid the instability. Experimental demonstrations of a FE FET with 13 mV/dec subthreshold swing were reported recently [35].

1.2.5 Feedback FETs

The positive feedback mechanism has also been proposed to realize steep slope devices. It is widely known that the negative feedback helps stabilizing a system while a positive feedback makes a system unstable. Similar to the electro-mechanical FETs discussed above, a strong positive feedback under a certain gate voltage can potentially bring significant changes in the device states and induce a steep slope. There are different types of positive feedback mechanisms that can be used. Figure 1.12 plots the schematics of two types of steep slope devices and their working principles.

The first type of feedback FET [36] is based on the gate-tuned, positively biased p-i-n junction, with negative charges near the n-doped source and positive charges near the p-doped drain shown in Fig. 1.12a, b. These extra charges are located in the gate underlap region and form two potential barriers for electrons and holes. When electrons are injected into the channel, some of them accumulating in the hole barrier region helps reducing the barrier height. The same is happening for holes. With the charge accumulation, the conduction current is increased and further reduces the barrier height. This positive feedback process induces sharp change in

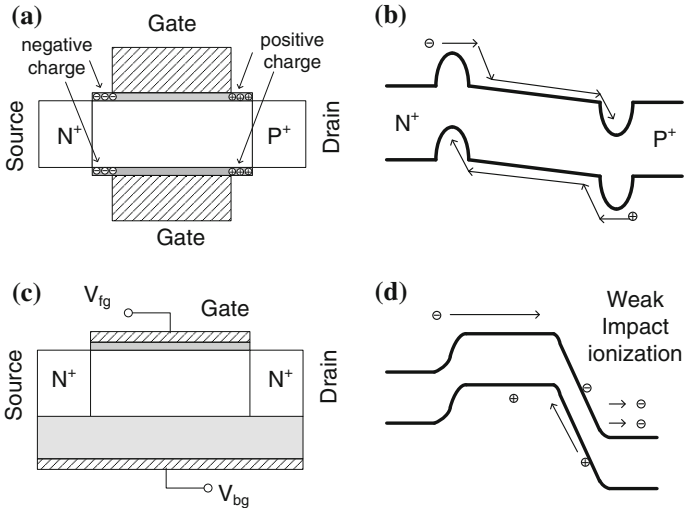

Fig. 1.12 Schematics of two kinds of feedback FETs and their working principles

the current with certain gate voltage. Those extra charges can come from the device process [37] or introduced by programming [38].

The second type of feedback FET [39, 40] is based on the floating body effect of the silicon-on-insulator (SOI) transistors and the weak impact ionizations, as shown in Fig. 1.12c, d. The diffusion current from the source to drain induces one electron–hole pair near the drain side due to the large electric field. The induced electron reaches the drain side together with the original electron, while the induced hole stays in the floating body of the SOI transistor. The positive charge reduces the source side barrier, so more electrons can take part in the diffusion and more holes accumulate in the body region. This positive feedback process also induces significant changes in the current with a small change in the gate bias, hence helps achieving the steep slopes.

Extremely small subthreshold slopes of 3.4 mV/dec [40], 2 mV/dec, or even 58 μV/dec [39], have been reported in the literatures with feedback FETs.

1.2.6 Tunnel FETs

It is well known that two mechanisms contribute to the breakdown of reversely biased p–n junction: One is the impact ionizations, and another is the band-to-band tunneling. Both mechanisms happen with high electric field and induce a large current change within a small voltage window. Similar to the impact ionization, the interband tunneling is also explored to realize steep slopes and the related transistors are called tunnel field-effect transistors (TFETs). Figure 1.13 shows the

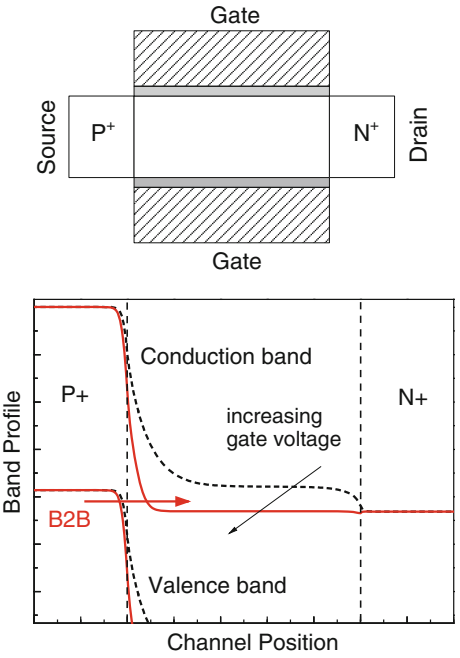

Fig. 1.13 Schematic of the homojunction tunnel FET and its working principles

schematic of one basic homojunction TFET and the operation principles. Its variation structures will be covered in later sections.

The basic TFET structure is very similar to that of the aforementioned impact ionization FET, however, with the gate electrode covering the whole channel region. With this structure, a high electric field is achieved but confined nearby the source/channel junction. Due to the absence of the underlap region with an almost constant field (Fig. 1.8), the impact ionization condition Eq. (1.15) is not satisfied. When the gate voltage is small, the interband tunneling is forbidden as indicated by the dotted line in Fig. 1.13. There is a SRH leakage current at this stage which represents the TFET off state. By increasing the gate voltage, the conduction band in the channel is brought below the valence band in the source, hence the band-to-band (B2B) tunneling starts. The carrier generation rate G_{tun} due to the interband tunneling is given by the Kane's model [41] with the local approximation:

$$G_{tun} = A\frac{\xi_{//}^2}{\sqrt{E_g}}\exp\left(-B\frac{E_g^{3/2}}{\xi_{//}}\right) \tag{1.34}$$

where E_g is the semiconductor band gap, A and B are two variables that are dependent on the material properties. It is shown in Fig. 1.13 that across the junction electric field is non-uniform so there is a spatial distribution of the carrier generations. Correspondingly, the tunneling current is obtained by integrating the generation rate along the channel direction:

$$I_{ds} = \int_{channel} G_{tun}(x)dx \tag{1.35}$$

With increasing the gate voltage, the potential inside the channel is also increased so is the electric field across the junction and the generation rate. Similar to the impact ionization FETs, the electrostatics inside the TFET channel follows the same gate modulation as Eq. (1.8) in MOSFETs. When the electric field is not very large, the tunneling current is limited and the subthreshold swing is written as

$$SS \approx 2.3 \times \frac{\xi_{//}^2}{B \cdot E_g^{3/2}} \times \frac{\partial V_g}{\partial \xi_{//}} \tag{1.36}$$

It is seen that when transforming the electrostatics to the tunneling by Eq. (1.34) instead of the thermal emission by Eq. (1.9), the limitation imposed by the 'kT' term in Eq. (1.11) is removed. The second term in Eq. (1.36) is material dependent and the third term is also device structure dependent. With proper options of the semiconductor material and device geometry, the slope given by Eq. (1.36) can be brought below 60 mV/dec at the room temperature. Similar to the impact ionization FETs, it is the amplification of the carrier transport mechanism different from the diffusion by Eq. (1.10) that contributes to the possible steep slope. Experimental demonstrations of TFETs with sub-60 mV/dec subthreshold swings have been reported widely [42–45].

1.2.7 Comparisons Between Steep Slope Devices

By choosing one representative of aforementioned steep slope devices reported in the literatures, we compare their properties in Fig. 1.14 by super-imposing their transfer

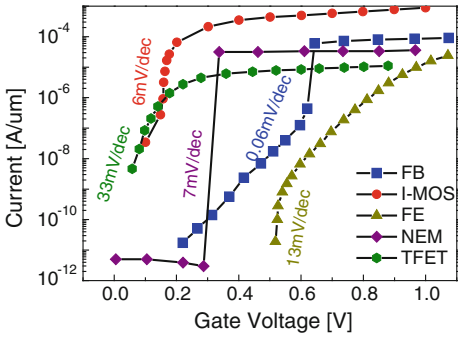

Fig. 1.14 Comparisons between the representatives of the steep slope devices, impact ionization FETs (I-MOS) [24], nano-electro-mechanical devices (NEM) [30], ferroelectric FETs (FE) (with current magnified by 10^5 times) [35], feedback FETs (FB) [39], and tunnel FETs (TFET) [44]

characteristics together. There is not yet any reported sub-60 mV/dec swing in PETs. In the figure, around 6 mV/dec subthreshold swing is achieved in an impact ionization FET [24]. However, it is generally believed that V_{dd} scaling with impact ionization FETs is challenging. On the other hand, it is shown that these transistors suffer from severe hot carrier effects which lead to significant threshold voltage shifts due to injected carriers into the gate dielectric. Very steep slopes (e.g., 7 mV/dec) have been achieved in the electro-mechanical device [30]. Due to the pull-in effect and the surface charge adhesion [29], its switching is usually hysteretic. A small slope of 13 mV/dec of the ferroelectric FET is also attractive [35]. Ferroelectric material growth and process compatibility are issues to be solved. An extremely small slope is realized by the feedback FET [39]. However, large voltage operation and complicated programming [38] make its application quite challenging. Subthreshold swing in the Si-based TFET is around 30 mV/dec [44], larger than those achieved by other steep slope devices. Although small driving current in TFETs is one common issue, there are a lot of techniques, e.g., those from materials or from device geometries that can potentially be used to improve their drivability. Variations of TFETs from the basic homojunction one are covered in the next sections.

1.3 TFETs Characteristics

In the above sections, we show the possible steep slope property of TFETs and qualitatively compare it with other kinds of steep slope devices. In this part, we look into the details of TFETs characteristics and find out some common issues in TFETs before making them more applicable in low-power circuits/systems. Numerical simulations by solving the Poisson's equation and the interband tunneling equation are used to reveal the basic TFET properties and their origins. Whenever possible, these device characteristics are confirmed with experimental TFETs.

Figure 1.15 plots the current–voltage characteristics of one TFET in double-gate configuration of Fig. 1.13. The channel is 10-nm-thick silicon, the gate oxide made

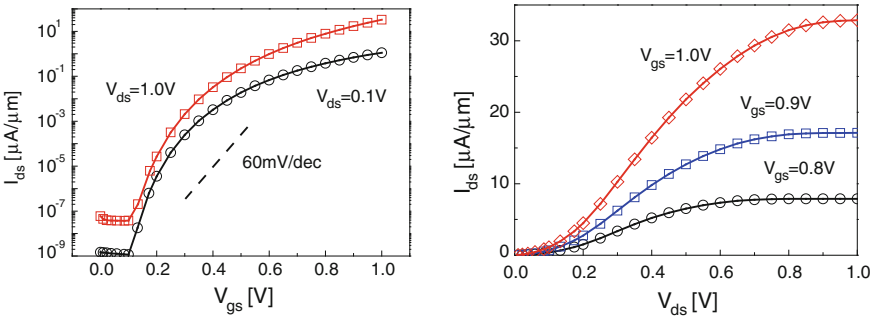

Fig. 1.15 Current–voltage characteristics of one simulated homojunction-based TFET

of SiO_2 is 1 nm thick, the channel width is 1 μm, and length is 50 nm. The source region is doped with p-type and a concentration of $10^{20}cm^{-3}$, the drain region is doped with n-type and the same concentration. A physics model to account for the interband tunneling from the WKB approximation [22] is included in the simulation. It induces a carrier generation term which is used in the drift–diffusion (DD) equation. The Poisson's equation coupled with the DD equation are solved together to obtain the TFET properties. Although more regular quantum transport simulations are also possible [46], the traditional DD-based simulations capture the essential physics and provide a quick view into the device internal.

The simulated TFET in Fig. 1.15 shows sub-60 mV/dec swing within a voltage window of 0.2 V (0.1–0.3 V), with the turn-on of the interband tunneling at 0.1 V. Correspondingly, the steep slope property is confined below a certain current around nA/μm. Meanwhile, the on-state current with the operation voltage of 1 V is around tens of μA/μm, much smaller than that in MOSFETs. On the other hand, the TFET current shows super-linear dependence on its drain voltage as shown in Fig. 1.15. The sub-60 mV/dec swing being confined within small current levels, the small on-state current, and the super-linear output characteristics are observed in experimental Si p-i-n TFETs. Table 1.1 summaries the reported devices in the literatures.

In the following, we look into the internal of the TFET in Fig. 1.15 and find out the reason for the above three characteristics. As shown in Fig. 1.13 and Eq. (1.36), it is the gate adjustment of the band profile in the channel (or electrostatic potentials) and the electric field across the tunnel junction that determines the subthreshold slope. If we look at the channel and N^+-doped drain in Fig. 1.13, we expect that the free carrier concentrations in the channel increase when the gate voltage pulls down the conduction band. The TFET channel is similar to the

Table 1.1 Recent reports of the experimental Si p-i-n TFETs properties

Work	Minimum SS (mV/dec)	Current window with steep slope	On-state current@Vdd	Super-linear output
Mayer et al. [42]	42	<0.1 nA/μm	0.02 uA/um@1.0 V	Yes
Jeon et al. [43]	46	<0.1 nA/μm	1.2 uA/um@1.0 V	Yes
Choi, et al. [44]	~33	500 nA/μm	12.1 uA/um@1.0 V	NA
Gandhi, et al. [45]	30	<10 pA/μm	NA	NA
Gandhi, et al. [45]	50	<0.1 nA/μm	0.02 uA/um@1.0 V	Yes
Richter et al. [47]	76	NA	1.3 uA/um@3.5 V	Yes

MOSFET channel in strong inversion. It is known that the screening effects from these free carriers reduce significantly the dependence of channel potentials on the gate voltage, compared to the subthreshold region where the channel potentials follow the gate voltage as Eq. (1.8). Since the channel potential relative to the source region determines the electric field across the tunnel junction, the gate control over the electric field is reduced. As a result, the subthreshold slope given by Eq. (1.36) is increasing with larger gate voltage. Figure 1.16 plots the changes of channel potentials and the maximum electric field with the gate voltages which confirm the above analysis. It is the charge screening effect in the p-i-n TFET channel that causes the increase of the subthreshold slope. At the same time, the tunneling current when the screening effect starts depends on the absolute electric field and the material properties like the band gap and carrier tunnel mass as given in Eqs. (1.34) and (1.35). Si has a relatively large band gap and carrier mass for tunneling, leading to the small current for the possible 60 mV/dec swing in both the simulated and experimental TFETs.

With increasing the gate voltage, free carrier concentrations in the channel increases. The p-i-n TFET is equivalent to a tunnel diode with gate-tuned 'doping' concentrations of the channel region. Eventually the electric field in the tunnel diode is also doping dependent and the maximum field will be reached. In this scenario, the tunneling current will be determined by the material properties. The

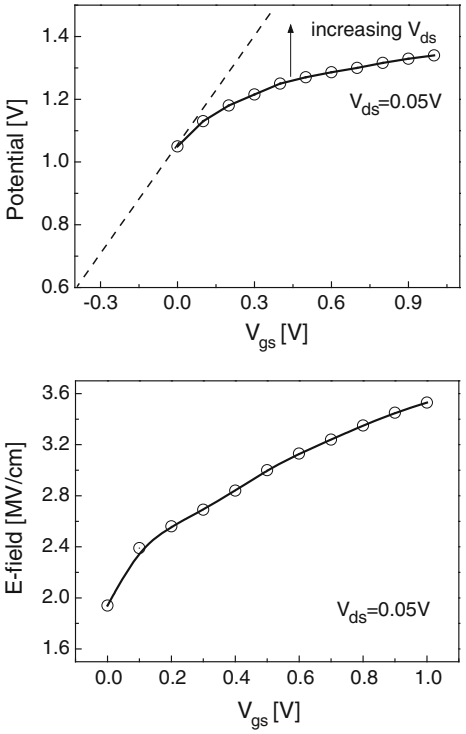

Fig. 1.16 Dependences of the channel potential and tunnel junction field on the gate voltage in TFETs

large band gap and tunnel mass correspond to small generation rate in Eq. (1.34), limiting the on-state current of the Si p-i-n TFETs.

The super-linear output is also due to the TFET channel charge effect. The dashed line in Fig. 1.16 shows the channel potentials without considering the channel charge. When the TFET drain voltage is increased, the free carriers inside the channel are reduced and the channel potentials are increased with the same gate voltage. As a result, the electric field across the tunnel junction is also increased as it is determined by the channel potentials. This means when the TFET drain voltage is small it codetermines the tunnel junction status with the gate voltage. Due to the exponential dependence of the tunneling current on the electric field, the TFET output characteristics are also exponential within a certain drain voltage window.

From the discussions in the first part, the switches for operation voltage and power consumption reductions are expected to have comparable on-state current as MOSFETs, steep slopes for current changes in several orders of magnitude. Another figure-of-merit of the steep slope devices is that the on-state channel resistance should be small in order to improve the circuit states switching speed simply shown in Fig. 1.2. It means that the super-linear output characteristics are not beneficial to the circuit applications. In the next part, techniques to enhance the TFET performances are covered.

1.4 Techniques for TFETs Performance Enhancements

1.4.1 Geometry Engineering

From the discussions in Sect. 1.2.6 and the above Sect. 1.3, it is the gate control over the tunnel junction electric field that determines the TFET subthreshold slope and on-state current. Techniques to improve the sensitivities of junction fields on the gate voltage will be effective performance enhancements. The nature length used to describe the short-channel effects in MOSFETs [48] is an indicator of the device electrostatic integrity. A smaller nature length represents a more abrupt potential profile and better electrostatic integrity. The nature length is also applicable to describe the tunnel junction potential profiles. A smaller nature length means a more effective transform of the gate field to the tunnel junction field, hence favoring the steeper slope and larger on-state current. Reasons for the geometry engineering can be obtained from the available nature length theory: (1) the nature length of gate-all-around (or nano-wire) is smaller than the double-gate, and smaller than the single-gate geometry; (2) thin equivalent oxide thickness (EOT) induces small nature length; (3) thin body thickness or radius leads to small nature length. As a result, double-gate or nano-wire TFETs, with small channel thickness or radius, together with a small EOT, can be used to enhance TFET performances. It has been confirmed by numerical simulation [49].

1.4.2 Doping Engineering

Without changes of the material, the second technique to enhance the Si TFET on current is adding a pocket doping (different polarity from the source) region as illustrated in Fig. 1.17. The pocket region is depleted, leaving the ionized positive charge. As shown in Fig. 1.17, the depletion charge contributes another electric field component to the intrinsic field in Fig. 1.16. While this pocket region does not change the potentials in the i-region for given gate and drain voltages, the larger field across the tunnel junction helps increasing the generation rate, hence the tunneling current. At the same time, this additional field also means that the current is increased right after the interband tunneling is turned on. As a result, the minimum subthreshold slope is further reduced. In the discussions of basic TFETs, the super-linear output is attributed to the drain voltage control over the junction field. In the pocket doped TFET, the drain voltage only changes the intrinsic field without affecting the pocket depletion. Overall the drain effect on the tunnel junction is weakened. The super-linear output characteristics are expected to be reduced compared to the basic p-i-n TFETs.

The pocket doping engineering based on the Si TFET was proposed in Ref. [50]. Optimizations of the doping concentrations, pocket width, and the dopant activation schemes were explored, and recently, a p-type pocket (P^+ doping with boron) TFET was demonstrated [51] with the subthreshold swing of 46 mV/dec, on-state current of 1.4 μA/μm and without obviously super-linear output. The current window with sub-60 mV/dec swing is limited to <10 pA/μm.

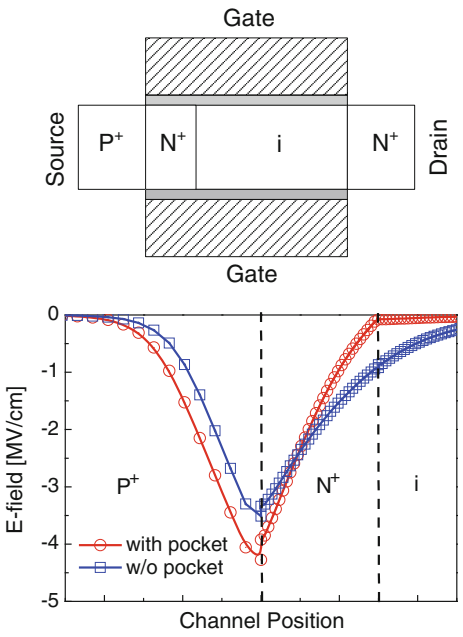

Fig. 1.17 The pocket doping technique for TFET performance enhancements

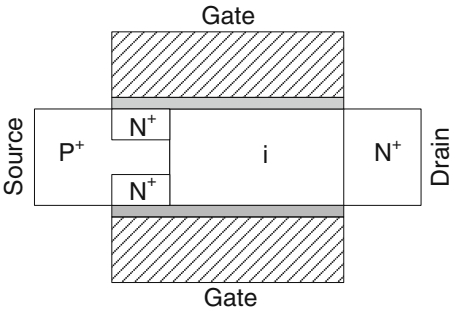

Fig. 1.18 A gate field-aligned tunneling structure is proposed to enhance TFET performances

Another similar doping engineering TFET was proposed and optimized [52], with its schematic shown in Fig. 1.18. By aligning the gate electric field with the tunneling direction, it is expected from numerical simulations that larger volume for tunneling can be achieved; hence, the TFET on-state current can be increased.

1.4.3 Material Engineering

From the basic tunneling physics in Eq. (1.34), semiconductor materials with smaller band gap favor larger tunneling generation rate and current with a given electric field. Correspondingly, the third technique to enhance the TFET performance is using materials of smaller band gap as the channel. Meanwhile, considerations of the semiconductor/dielectric interface, the doping and the overall process integration should be included at the same time.

Germanium (Ge) has a band gap of 0.66 eV, making it an option to enhance the TFET performance. By using Ge in the whole TFET channel [42], the on-state current increases by around 2700 times compared to the Si-based one. SiGe with different mole fraction is also applicable. Another option is to use Ge/Si heterojunction as the tunnel junction in TFETs. With the electron affinity of 4.0 eV (compared to the 4.05 eV of Si) and a small band gap, Ge and Si form the type-II (staggered) heterojunction. Figure 1.19 plots the junction band profile when the Ge/Si n-type TFET is in its off and on states. With the two-band model [53] of the interband tunneling process, the electron wave penetrating into the energy barrier initially shows the Ge property, then reaches the Si lattice, and decays into the Si conduction band. The effective energy barrier for this tunneling is 0.61 eV $[E_g(Ge) - \Delta E_c]$, slightly smaller than the one in pure Ge. In addition, the process integration of Ge into the Si technology is not a problem as in modern CMOS Ge is used to induce compressive strain to increase hole mobility in p-type MOSFETs [15]. With the Ge source and Si channel in TFETs, a minimum subthreshold swing of 40 mV/dec together with 0.4 μA/μm on-state current under 0.5 V operation is achieved [54]. It is also possible to combine the material engineering and doping engineering to further improve the TFET performance.

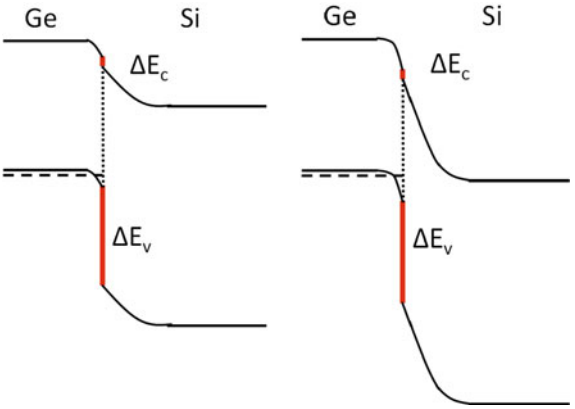

Fig. 1.19 Band alignments in the Ge/Si hetero-junction-based n-type TFET

The formulation for the tunneling barrier (or in another term, the effective band gap) of type-II heterojunction between material a and b is generalized as follows:

$$E_{g,eff} = E_{g,a} - \Delta E_c = E_{g,b} - \Delta E_v \tag{1.37}$$

It can be used to search other staggered heterojunctions for TFETs applications.

Some III–V materials and their alloys, such as InAs and InGaAs, have small band gaps (e.g., 0.36 eV of InAs, 0.58 eV of $In_{0.7}Ga_{0.3}As$, and 0.74 eV of $In_{0.53}Ga_{0.47}As$). Theoretical simulations [55] and experimental demonstrations [56, 57] of these materials-based TFETs show enhanced performances including the on-state current and linear output. Staggered heterojunctions can be formed by III–V materials or their alloys. III–V and IV material-based heterojunctions like the InAs/Si junction are also proposed for TFET applications. Table 1.2 summarizes the widely explored heterojunctions for TFETs.

Another kind of heterojunction with the broken-gap (type-III) alignment is also proposed for TFET applications [66, 67]. The junction is usually composed of GaSb (E_g = 0.75 eV) and InAs (E_g = 0.36 eV) with the valence band of GaSb higher than the InAs conduction band by ΔE_{vc} = 0.09 eV. Figure 1.20 shows the heterojunction band profiles by assuming that GaSb is doped with p-type and its Fermi level (the dashed line) aligns with the InAs conduction band. Initially, the

Table 1.2 Recent reports of the staggered heterojunction-based TFETs

Heterojunctions	TFET mode	Effective E_g	Work
InAs/$Al_{0.45}Ga_{0.55}Sb$	n-type	40 meV	[58, 59]
$GaAs_{0.35}$ $Sb_{0.65}$/$In_{0.7}Ga_{0.3}As$	n-type	0.25 eV	[60]
$In_{0.53}Ga_{0.47}As$/$In_{0.7}Ga_{0.3}As$	n-type	0.59 eV	[61] with <60 mV/dec slope
InAs/Si	p-type	0.2 eV	[62–64]
Ge/Si	n-type	0.6 eV	[65]

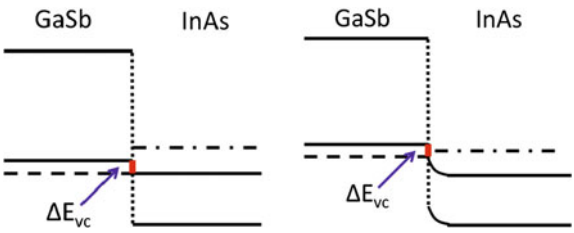

Fig. 1.20 Band alignments in the GaSb/InAs broken-gap heterojunction-based n-type TFET

first sub-band in InAs (the dash-dot line) due to certain quantum confinements are higher than the GaSb valence band, there is not tunneling window available and the TFET is in its off state. With electric field from the gate electrode, the first InAs sub-band is brought below the GaSb valence band and interband tunneling is made possible. Compared to the junction with doping modulation in Fig. 1.13 or Fig. 1.19, the electric field across the tunnel junction can be assumed to be infinitely large. As a result, a significant improvement in the on-state current is expected. The switch from the off to the on state accompanied with huge changes of tunneling current means a quite steep slope. At the same time, the drain voltage modulation on the tunneling current is weakened significantly leading to the linear output characteristics. Numerical simulations confirm that a constant steep slope (as small as several mV/dec) and CMOS comparable on-state current can be obtained with the broken-gap junction-based TFETs [68]. Experimental demonstrations of the GaSb/InAs-based TFETs were reported with record high on current of 180 μA/μm, but without the sub-60 mV/dec slope yet [69].

Fabrications of the III–V materials-based homojunction and heterojunction TFETs will be covered in Chap. 2 of this book. Material-engineered TFETs with the two-dimensional and one-dimensional semiconductors will be covered in later chapters.

1.5 Summary

In this chapter, we reviewed several steep slope devices as possible building blocks in low-power applications. From the analysis of CMOS power consumption, we revealed that reducing the device subthreshold swings promote reductions in circuit operation voltages and power. Devices using different physics mechanisms for steep slopes are introduced and compared with their operation principles. The TFET as one of the promising candidates among steep slope devices is given special emphasis. We investigated the TFET device physics and summarized the issues of the Si TFETs. Finally, we provided brief analysis on several techniques to improve TFETs performances, including the geometry engineering, the doping, and materials engineering.

Acknowledgement This work was supported by the Hong Kong's University Grant Committee via the Area of Excellence project AoE-P04-08. We would like to thank Prof. Cary Yang of Santa Clara University for his inputs on the steep slope devices.

References

1. S. Borkar, Low power design challenges for the decade. in *Proceedings of ASP-DAC*, pp. 293–296, 2001
2. P. Huber, Dig more coal—the PCs are coming. Forbes, pp. 70–72, May 1999
3. M. Mills, Opportunity in the Internet's voracious energy appetite: the cloud begin with coal. Forbes, 31 May 2011
4. R.W. Keyes, R. Landauer, Minimal energy dissipation in logic. IBM J. Res. Dev. **14**, 152–157 (1970)
5. J.D. Meindl, J.A. Davis, The fundamental limit on binary switching energy for terascale integration. IEEE J. Solid-State Cir. **35**(10), 1515–1516 (2000)
6. R.W. Keyes, Fundamental limits of silicon technology. Proc. IEEE **89**, 227–239 (2001)
7. K.L. Wang, K. Galatsis, R. Ostroumov, A. Khitun, Z. Zhao, S. Han, Nanoarchitectonics for hetero-geneous integrated nanosystems. Proc. IEEE **96**(2), 212–229 (2008)
8. R. Dennard, F.H. Gaensslen, H. Yu, V. Rideout, E. Bassous, A.R. Leblanc, Design of ion-implanted MOSFET's with very small physical dimensions. IEEE J. Solid-State Circuits sc-**9**, 256–268 (1974)
9. P. Packan, Device and circuit interactions. IEEE IEDM Short Course (2007)
10. E.J. Nowak, Maintaining the benefits of CMOS scaling when scaling bogs down. IBM J. Res. Dev. **46**, 169–180 (2002)
11. M. Bohr, S. Ahmed, L. Brigham, R. Chau, R. Gasser, R. Green, W. Hargrove, E. Lee, R. Natter, S. Thompson, K. Weldon, S. Yang, A high performance 0.35 μm logic technology for 3.3 V and 2.5 V operation. IEDM Tech. Dig. 273–276 (1994)
12. M. Bohr, S.S. Ahmed, S.U. Ahmed, M. Bost, T. Ghani, J. Greason, R. Hainsey, C. Jan, P. Packan, S. Sivakumar, S. Thompson, J. Tsai, S. Yang, A high performance 0.25 μm logic technology optimized for 1.8 V operation. IEDM Tech. Dig. 847–850 (1996)
13. S. Yang, S. Ahmed, B. Arcot, R. Arghavani, P. Bai, S. Chambers, P. Charvat, R. Cotner, R. Gasser, T. Ghani, M. Hussein, C. Jan, C. Kardas, J. Maiz, P. McGregor, B. McIntyre, P. Nguyen, P. Packan, I. Post, S. Sivakumar, J. Steigerwald, M. Taylor, B. Tufts, S. Tyagi, M. Bohr, A high performance 180 nm generation logic technology. IEDM Tech. Dig. 197–200 (1998)
14. S. Tyagi, M. Alavi, R. Bigwood, T. Bramblett, J. Brandenburg, W. Chen, B. Crew, M. Hussein, P. Jacob, C. Kenyon, C. Lo, B. Mcintyre, Z. Ma, P. Moon, P. Nguyen, L. Rumaner, R. Schweinfurth, S. Sivakumar, M. Stettler, S. Thompson, B. Tufts, J. Xu, S. Yang, M. Bohr, A 130 nm generation logic technology featuring 70 nm transistors, dual Vt transistors and 6 layers of Cu interconnects. IEDM Tech. Dig. 567–570 (2000)
15. T. Ghani, M. Armstrong, C. Auth, M. Bost, P. Charvat, G. Glass, T. Hoffmann, K. Johnson, C. Kenyon, J. Klaus, B. McIntyre, K. Mistry, A. Murthy, J. Sandford, M. Silberstein, S. Sivakumar, P. Smith, K. Zawadzki, S. Thompson, M. Bohr, A 90 nm high volume manufacturing logic technology featuring novel 45 nm gate length strained silicon CMOS transistors. IEDM Tech. Dig. 978–980 (2003)
16. P. Bai, C. Auth, S. Lalakrishnan, M. Bost, R. Brain, V. Chikarmane, R. Heussner, M. Hussein, J. Hwang, D. Ingerly, R. James, J. Jeong, C. Kenyon, E. Lee, S. Lee, N. Lindert, M. Liu, Z. Ma, T. Marieb, A. Murthy, R. Nagisetty, S. Natarajan, J. Neirynck, A. Ott, C. Parker, J. Sebastina, R. Shaheed, S. Sivakumar, J. Steigerwald, S. Tyagi, C. Weber, B. Woolery, A. Yeoh, K. Zhang, M. Bohr, A 65 nm logic technology featuring 35 nm gate lengths,

enhanced channel strain, 8 Cu interconnect layers, low-k ILD and 0.57 μm^2 SRAM cell. IEDM Tech. Dig. 657–660 (2004)
17. K. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom, M. Bost, M. Brazier, M. Buehler, A. Cappellani, R. Chau, C. Choi, G. Ding, K. Fischer, T. Ghani, R. Grover, W. Han, D. Hanken, M. Hattendorf, J. He, J. Hicks, R. Huessner, D. Ingerly, P. Jain, R. james, L. Jong, S. Joshi, C. Kenyon, K. Kuhn, K. Lee, H. Liu, J. Maiz, B. Mcintyre, P. Moon, J. Neirynck, S. Pae, C. Parker, D. Parsons, C. Prasad, L. Pipes, M. Prince, P. Ranade, T. Reynolds, J. Sandford, L. Shifren, J. Sebastian, J. Seiple, D. Simon, S. Sivakumar, P. Smith, C. Thomas, T. Troeger, P. Vandervoorn, S. Williams, K. Zawadzki, A 45 nm logic technology with high-k + metal gate transistors, strained silicon, 9 Cu interconnect layers, 193 nm dry patterning, and 100 % Pb-free packaging. IEDM Tech. Dig. 247–250 (2007)
18. P. Packan, S. Akbar, M. Armstrong, D. Bergstrom, M. Brazier, H. Deshpande, K. Dev, G. Ding, T. Ghani, O. Golonzka, W. Han, J. He, R. Heussner, R. James, J. Jopling, C. Kenyon, S-H. Lee, M. Liu, S. Lodha, B. Mattis, A. Murthy, L. Neiberg, J. Neirynck, S. Pae, C. Parker, L. Pipes, J. Sebastian, J. Seiple, B. Sell, A. Sharma, S. Sivakumar, B. Song, A. St. Amour, K. Tone, T. Troeger, C. Weber, K. Zhang, Y. Luo, S. Natarajan, High performance 32 nm logic technology featuring 2nd generation high-k + metal gate transistors. IEDM Tech. Dig. 659–662 (2009)
19. C.-H. Jan, U. Bhattacharya, R. Brain, S. Choi, G. Curello, G. Gupta, W. Hafez, M. Jang, M. Kang, K. Komeyli, T. Leo, N. Nidhi, L. Pan, J. Park, K. Phoa, A. Rahman, C. Staus, H. Tashiro, C. Tsai, P. Vandervoorn, L. Yang, J.-Y. Yeh, P. Bai, A 22 nm Soc platform technology featuring 3-D tri-gate and high-k/metal gate, optimized for ultralow power, high performance and high density SoC applications. IEDM Tech. Dig. 44–47 (2012)
20. S. Natarajan, M. Agostinelli, S. Akbar, M. Bost, A. Bowonder, V. Chikarmane, S. Chouksey, A. Dasgupta, K. Fischer, Q. Fu, T. Ghani, M. Giles, S. Govindaraju, R. Grover, W. Han, D. Hanken, E. Haralson, M. Haran, M. Heckscher, R. Heussner, P. Jain, R. James, R. Jhaveri, I. Jin, H. Kam, E. Karl, C. Kenyon, M. Liu, Y. Luo, R. Mehandru, S. Morarka, L. Neiberg, P. Packan, A. Paliwal, C. Parker, P. Patel, R. Patel, C. Pelto, L. Pipes, P. Plekhanov, M. Prince, S. Rajamani, J. Sandford, B. Sell, S. Sivakumar, P. Smith, B. Song, K. Tone, T. Troeger, J. Wiedemer, M. Yang, K. Zhang, A 14 nm logic technology featuring 2nd generation FinFET transistors, air-gapped interconnects, self-aligned double patterning and a 0.0588 μm^2 SRAM cell size. IEDM Tech. Dig. 71–74 (2014)
21. R. Chau, S. Datta, M. Doczy, B. Doyle, B. Jin, J. Kavalieros, A. Majumdar, M. Metz, M. Radosavljevic, Benchmarking nanotechnology for high-performance and low-power logic transistor applications. IEEE Tran. Nanotechnology **4**(2), 153–158 (2005)
22. S.M. Sze, *Physics of Semiconductor Devices*, 2nd edition (John Wiley & Sons, 1981)
23. K. Gopalakrishnan, P.B. Griffin, J.D. Plummer, I-MOS: a novel semiconductor device with a sub-threshold slope lower than kT/q. IEDM Tech. Dig. 289–292 (2002)
24. D. Sarkar, N. Singh, K. Banerjee, A novel enhanced electric-field impact-ionization MOS transistor. IEEE Electron Dev. Lett. **31**(11), 1175–1177 (2010)
25. E. Toh, G. Wang, L. Chan, G. Lo, G. Samudra, Y. Yeo, Strain and materials engineering for the I-MOS transistor with an elevated impact-ionization region. IEEE Trans. Electron Dev. **54** (10), 2778–2785 (2007)
26. A.M. Ionescu, V. Pott, R. Fritschi, K. Banerjee, M.J. Declercq, P. Renaud, C. Hibert, P. Fluckiger, G. Racine, Modeling and design of a low-voltage SOI suspended-gate MOSFET with a metal-over-gate architecture. Proc. ISQED 496–501 (2002)
27. H. Kam, D. Lee, R. Howe, T.J. King, A new nano-electro-mechanical field-effect transistor design for low power electronics. IEDM Tech. Dig. 464–466 (2005)
28. N. Abele, R. Fritschi, K. Boucart, F. Casset, P. Ancey, A. M. Ionescu, Suspended-gate MOSFET: bringing new MEMS functionality into solid-state MOS transistor. IEDM Tech. Dig. 479–481 (2005)
29. T.J. King Liu, J. Jeon, R. Nathanael, H. Kam, V. Pott, E. Alon, Prospects for MEM logic switch technology. IEDM Tech. Dig. 424–427 (2014)

30. N. Xu, J. Sun, I. Chen, L. Hutin, Y. Chen, J. Fujiki, C. Qian, T.J. King Liu, Hybrid CMOS/BEOL-NEMS technology for ultra-low-power IC applications. IEDM Tech. Dig. 677–680 (2014)
31. D. Newns, G. Martyna, B. Elmegreen, X. Liu, T. Theis, Extended abstract, in *Proceedings of the 15th US-Japan Seminar on Dielectric and Piezoelectric Ceramics*, Castle Park Hotel, Kagoshima, Japan, 6–9 Nov 2011
32. P. Solomon, B. Bryce, M. Kuroda, R. Keech, S. Shetty, T. Shaw, M. Copel, L. Hung, A. Schrott, C. Armstrong, M. Gordon, K. Reuter, T. Theis, W. haensch, S. Rossnagel, H. Miyazoe, B. Elmegreen, X. Liu, S. Trolier-Mckinstry, G. Martyna, D. Newns, Pathway to the piezoelectric transduction logic device. Nano Lett. **15**, 2391–2395 (2015)
33. S. Salahuddin, S. Datta, Can the subthreshold swing in a classical FET be lowered below 60 mV/dec? IEDM Tech. Dig. 1–4 (2008)
34. S. Salahuddin, S. Datta, Use of negative capacitance to provide voltage amplification for low power nanoscale devices. Nano Lett. **8**, 405–410 (2008)
35. G. Salvatore, D. Bouvet, A.M. Ionescu, Demonstration of subthreshold swing smaller than 60 mV/ dec in Fe-FET with P(VDF-TrFE)/SiO_2 gate stack. IEDM Tech. Dig. 1–4 (2011)
36. A. Padilla, C. Yeung, C. Shin, C. Hu, T.J. King Liu, Feedback FET: a novel transistor exhibiting steep switching behavior at low bias voltages. IEDM Tech. Dig. 171–174 (2008)
37. Q. Zhou, S. Huang, H. Chen, C. Zhou, Z. Feng, S. Cai, K.J. Chen, Schottky source/drain Al_2O_3/InAlN/ GaN MIS-HEMT with steep sub-threshold swing and high on/off current ratio. IEDM Tech. Dig. 777–780 (2011)
38. C. Yeung, A. Padilla, T.J. King Liu, C. Hu, Programming characteristics of the steep turn-on/off feedback FET. VLSI Tech. Dig. 176–177 (2009)
39. Z. Lu, N. Collaert, M. Aoulaiche, B. De Wachter, A. De Keersgieter, J.G. Fossum, L. Altimime, M. Jurczak, Realizing super-steep subthreshold slope with conventional FDSOI CMOS at low-bias voltages. IEDM Tech. Dig. 407–410 (2010)
40. J. Zhang, M. De Marchi, P. Gaillardon, G. De Micheli, A schottky-barrier silicon FinFET with 6.0 mV/ dec subthreshold slope over 5 decades of current. IEDM Tech. Dig. 339–342 (2014)
41. E.O. Kane, Theory of tunneling. J. Appl. Phys. **32**(1), 83–91 (1961)
42. F. Mayer, C. Le Royer, J.-F. Damlencourt, K. Romanjek, F. Andrieu, C. Tabone, B. Previtali, S. Deleonibus, Impact of SOI, $Si_{1-x}Ge_xOI$ and GeOI substrates on CMOS compatible tunnel FET performance. IEDM Tech. Dig. 163–166 (2008)
43. K. Jeon, W.-Y. Loh, P. Patel, C. Y. Kang, J. Oh, A. Bowonder, C. Park, C.S. Park, C. Smith, P. Majhi, H.-H. Tseng, R. Jammy, T.-J. K. Liu, C. Hu, Si tunnel transistors with a novel silicided source and 46 mV/dec swing. VLSI Tech. Dig. 121–122 (2010)
44. W.Y. Choi, B.-G. Park, J.D. Lee, T.-J.K. Liu, Tunneling field effect transistors with subthreshold swing (SS) less than 60 mV/dec. IEEE Electron Device Lett. **28**(8), 743–745 (2007)
45. R. Gandhi, Z. Chen, N. Singh, K. Banerjee, S. Lee, Vertical Si-nanowire n-type tunneling FETs with low subthreshold swing (≤ 50 mV/decade) at room temperature. IEEE Electron Device Lett. **32**(4), 437–439 (2011)
46. Q. Shi, L. Zhang, Y. Zhu, L. Liu, M. Chan, H. Guo, Atomic disorder scattering in emerging transistors by parameter free first principle modeling. IEDM Tech. Dig. 733–736 (2014)
47. S. Richter, C. Sandow, A. Nichau, S. Trellenkamp, M. Schmidt, R. Luptak, K.K. Bourdelle, Q.T. Zhao, S. Mantl, Ω-gated silicon and strain silicon nanowire array tunneling FETs. IEEE Electron Device Lett. **33**(11), 1535–1537 (2012)
48. Y. Taur, T.H. Ning, *Fundamentals of Modern VLSI Devices* (Cambridge University Press, Cambridge, 1998)
49. L. Zhang, J. He, M. Chan, A Compact Model for Double-Gate Tunneling Field-Effect-Transistors and Its Implications on Circuit Behaviors. IEDM Tech. Dig. 143–146 (2012)
50. V. Nagavarapu, R. Jhaveri, J.C.S. Woo, The tunnel source (pnpn) n-MOSFET: a novel high perform-ance transistor. IEEE Trans. Electron Dev. **55**(4), 1013–1019 (2008)
51. A. Tura, Z. Zhang, P. Liu, Y. Xie, J.C.S. Woo, Vertical silicon p-n-p-n tunnel nMOSFET with MBE-grown tunneling junction. IEEE Trans. Electron Dev. **58**(7), 1907–1913 (2011)

52. C. Hu, P. Patel, A. Bowonder, K. Jeon, S. Kim, W. Loh, C. Kang, J. Oh, P. Majhi, A. Javey, T.S. King Liu, R. Jammy, Prospect of tunneling green transistor for 0.1 V CMOS. IEDM Tech. Dig. 387–390 (2010)
53. H. Flietner, The E(k) relation for a two-band scheme of semiconductors and the application to the metal-semiconductor contact. Phys. Stat. Sol. **54**, 201–208 (1972)
54. S. Kim, H. Kam, C. Hu, T.J. King Liu, Germanium-source tunnel field-effect transistors with record high I_{on}/I_{off}. VLSI Tech. Dig. 178–179 (2009)
55. M. Luisier, G. Klimeck, Atomistic full-band design study of InAs band-to-band tunneling field-effect transistors. IEEE Electron Device Lett. **30**(6), 602–604 (2009)
56. H. Zhao, Y. Chen, Y. Wang, F. Zhou, F. Xue, J. Lee, $In_{0.7}Ga_{0.3}As$ tunneling field-effect transistors with an Ion of 50 μA/μm and a subthreshold Sswing of 86 mV/dec using HfO_2 gate oxide. IEEE Electron Device Lett. **31**, 1392–1394 (2010)
57. S. Mookerjea, D. Mohata, R. Krishnan, J. Singh, A. Vallett, A. Ali, T. Mayer, V. Narayanan, D. Schlom, A. Liu, S. Datta, Experimental demonstration of 100 nm channel length $In_{0.53}Ga_{0.47}As$-based vertical inter-band tunnel field effect transistors (TFETs) for ultralow-power logic and SRAM applications. IEDM Tech. Dig. 949–952 (2009)
58. Y. Lu, G. Zhou, R. Li, Q. Liu, Q. Zhang, T. Vasen, S. Chae, T. Kosel, M. Wistey, H. Xing, A. Seabaugh, P. Fay, Performance of AlGaSb/InAs TFETs with gate electric field and tunneling direction aligned. IEEE Electron Device Lett. **33**(5), 655–657 (2012)
59. R. Li, Y. Lu, G. Zhou, Q. Liu, S. Chae, T. Vasen, W. Hwang, Q. Zhang, P. Fay, T. Kosel, M. Wistey, H. Xing, A. Seabaugh, AlGaSb/InAs tunnel field-effect transistor with on-current of 78 μA/μm at 0.5 V. IEEE Electron Device Lett. **33**(3), 363–365 (2012)
60. D. Mohata, R. Bijesh, S. Mujumdar, C. Eaton, R. Engel-Herbert, T. Mayer, V. Narayanan, J. Fastenau, D. Loubychev, A. Liu, S. Datta, Demonstration of MOSFET-like on-current performance in arsenide/antimonide tunnel FETs with staggered heterojunctions for 300 mV logic applications. IEDM Tech. Dig. 781–784 (2011)
61. G. Dewey, B.C. Kung, J. Boardman, J.M. Fastenau, J. Kavalieros, R. Kotlyar, W.K. Liu, D. Lubyshev, M. Metz, N. Mukherjee, P. Oakey, R. Pillarisetty, M. Radosavljevic, H. Then, R. Chau, Fabrication, characterization, and physics of III-V heterojunction tunneling field effect transistors for steep sub-threshold swing. IEDM Tech. Dig. 785–788 (2011)
62. K. Tomioka, M. Yoshimura, T. Fukui, Steep-slope tunnel field-effect transistors using III-V nanowire/ Si heterojunction. VLSI Tech. Dig. 47–48 (2012)
63. K. Moselund, H. Schmid, C. Bessire, M. Bjork, H. Ghoneim, H. Riel, InAs-Si nanowire heterojunction tunnel FETs. IEEE Electron Device Lett. **33**(10), 1453–1455 (2012)
64. H. Riel, K. Moselund, C. Bessire, M.T. Bjork, A. Schenk, H. Ghoneim, H. Schmid, InAs/Si heterojunc-tion nanowire tunnel diodes and tunnel FETs. IEDM Tech. Dig. 391–394 (2012)
65. M. Kim, Y. Wakabayashi, R. Nakane, M. Yokoyama, M. Takenaka, S. Takagi, High I_{on}/I_{off} Ge-source ultrathin body strained-SOI tunnel FETs. IEDM Tech. Dig. 331–334 (2014)
66. J. Smith, S. Das, J. Appenzeller, Broken-gap tunnel MOSFET: a constant-slope sub-60 mV/dec transistor. IEEE Electron Device Lett. **32**(10), 1367–1369 (2011)
67. L. Register, M. Hasan, S. Banerjee, Stepped broken-gap heterobarrier tunneling field-effect transistor for ultralow power and high speed. IEEE Electron Device Lett. **32**(6), 743–745 (2011)
68. M. Luisier, G. Klimeck, Performance comparisons of tunneling field-effect transistors made of InSb, carbon, and GaSb-InAs broken gap heterostructures. IEDM Tech. Dig. 913–916 (2009)
69. G. Zhou, R. Li, T. Vasen, M. Qi, S. Chae, Y. Lu, Q. Zhang, H. Zhu, J. M. Kuo, T. Kosel, M. Wistey, P. Fay, A. Seabaugh, H. Xing, Novel gate-recessed vertical InAs/GaSb TFETs with record high I_{on} of 180 μA/μm at V_{ds} = 0.5 V. IEDM Tech. Dig. 777–780 (2009)

Chapter 2
Tunneling FET Fabrication and Characterization

Tao Yu, Judy L. Hoyt and Dimitri A. Antoniadis

Abstract Since the early demonstration of the conventional p + −i − n + Tunneling FETs (TFETs), various tunneling junction designs as well as the introduction of new material systems enabled the performance of TFETs to improve by orders of magnitude. Different properties and considerations of the material systems require well designed processes and novel processes rarely seen in the CMOS technology also emerged. The technology of TFET fabrication has been evolving dramatically ever since. This chapter introduces a number of techniques in the previous studies on the fabrication technology for the TFETs. In addition, some characterization methods on the fabricated devices are also discussed for more efficient diagnosis and optimization on the TFETs.

2.1 Introduction

Since the first demonstration of the TFET by Appenzeller et al. in 2004 [1], extensive studies have been conducted to explore the potential of TFETs as an alternative technology for the future ultralow power CMOS [2–4]. Conventional lateral TFETs consist of heavily doped source and drain with opposite doping type and lightly doped or intrinsic channel as depicted in Fig. 2.1. The typical process flow for these TFETs is not particularly different from a MOSFET, which is summarized in Fig. 2.2. However, this generic design has two issues: (1) ambipolar behavior [3] and (2) poor drive current compared to the CMOS technology [2, 4]. Silicon, which is the most abundant semiconductor material, is the ideal choice for the TFETs considering the CMOS compatibility. Nevertheless, since silicon has indirect bandgap with E_g = 1.12 eV, the band-to-band tunneling process requires phonon assistance and the tunneling probability is poor such that the achievable ON current is in the order of 10^{-7} A/μm. Even worse, due to the presence of possible

T. Yu (✉) · J.L. Hoyt · D.A. Antoniadis
Microsystems Technology Laboratories, Massachusetts Institute of Technology,
77 Massachusetts Ave., Cambridge, MA 02139, USA
e-mail: yut@mit.edu

L. Zhang and M. Chan (eds.), *Tunneling Field Effect Transistor Technology*,
DOI 10.1007/978-3-319-31653-6_2

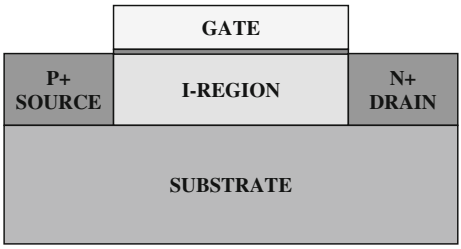

Fig. 2.1 Schematic view of a conventional lateral TFET consisting of the p+ source, intrinsic channel, and n+ drain

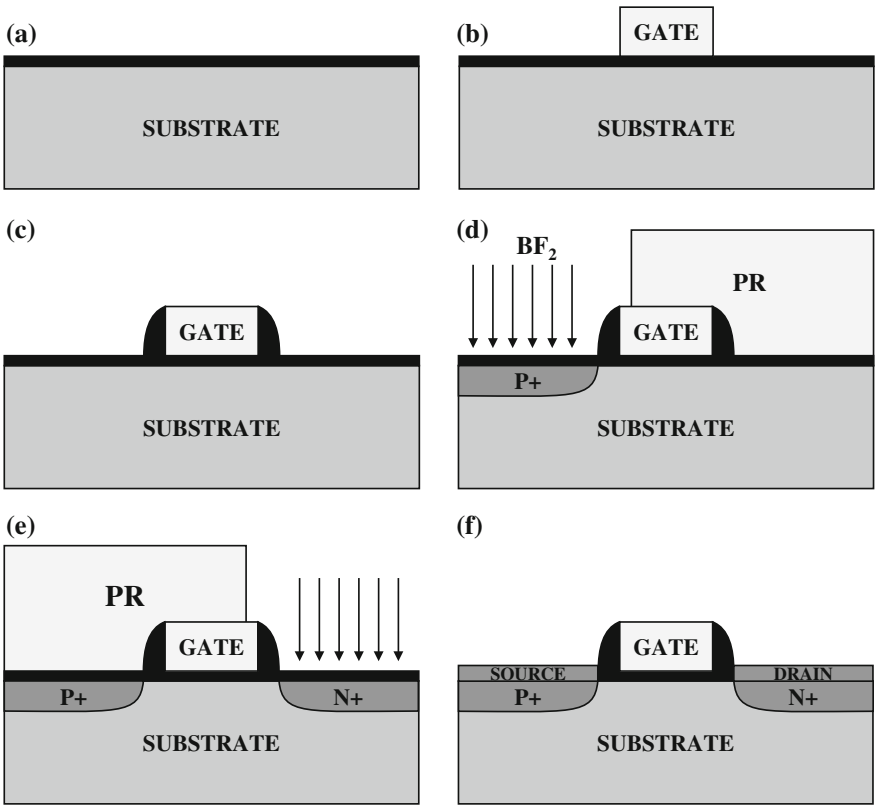

Fig. 2.2 Typical process flow of a conventional n-type lateral TFET with p+ source and n+ drain. **a** Gate oxide deposition. **b** Gate formation. **c** Sidewall spacer. **d** Source implantation. **e** Drain implantation. **f** Post-implantation annealing and metalization

parasitic thermionic currents, sub-60 mV/dec is rarely seen in the published experimental literature. These setbacks raised doubts on the potential of the TFETs as the future CMOS technology.

During years of research, efforts are made to improve the performance of TFETs in subthreshold swing and/or drive current making them competitive to the state-of-the-art CMOS technology [5–16]. In the design of the TFETs, the most essential component is the tunneling junction. Doping concentration, doping gradient/steepness, junction geometry, and dopant activation techniques were widely investigated to form the optimal configuration of the available technologies. On the other hand, tunneling properties of heterojunctions including group IV, III–V, and 2D material (graphene, transition metal dichalcogenide (TMD), black phosphorous, etc.) are extensively studied. The material properties (bandgap, density of states, etc.) and different combinations of material systems provide numerous choices for the TFET design. And achievements have been made demonstrating the TFETs with subthreshold slope as steep as 21 mV/dec [7], while some others showed drive current comparable to MOSFETs [10, 11].

To achieve the aforementioned performance, the technology for TFETs, however, is no longer straightforward. The device geometry and the tunneling junction become much more complicated, and the choice of material systems makes the fabrication process even more challenging. Additionally, overcomplicated fabrication processes failed to exhibit reliable and reproducible results. Analyzing the fabricated TFETs for the next iteration of design also becomes difficult. Traditional technique used to analyze the MOSFETs can no longer provide sufficient information on the TFETs, and new characterization methods are required to diagnose the possible issues with the design. In this chapter, commonly used TFET technologies for various TFET designs are discussed; several new characterization techniques are also described in detail for simplified design diagnosis procedure. The following sections are arranged as follows: Sect. 2.2 will describe several techniques for achieving superior tunneling junction formation; Sect. 2.3 will focus on the fabrication of the TFETs with different material systems and device geometries; and Sect. 2.4 will introduce several characterization methods for facilitating diagnosis on the device design and understand the potential performance limit for the design.

2.2 Tunneling Junction

In this section, various technologies are described for the improved tunneling junction designs. Although not all doping technologies are available for different material systems, similar ideas are still applicable for either group IV or III–V semiconductors. On the other hand, recent study showed that the semiconductors can also be “doped” electrostatically instead of using dopants. Section 2.2.2 will be devoted for the discussion of dopingless junctions.

2.2.1 Source/Drain Formation

The design of the source/drain doping profile is critical for the performance of TFETs in many aspects, including the subthreshold slope, tunneling current, and ambipolar behavior. For the homojunction TFETs, the tunneling conductance G_{tun} can be modeled with Kane's model [17]:

$$G_{tun} = A\frac{E^2}{\sqrt{E_g}}\exp\left(-B\frac{E_g^{3/2}}{|E|}\right)$$

where A and B are the parameters for the model, E_g is the bandgap of the material, E is the electric field across the tunneling junction:

$$E = \frac{E_g}{q \cdot W_{t,\min}}$$

and $W_{t,min}$ is the minimum tunneling distance. The 1D band diagram of the tunneling junction is shown in Fig. 2.3, where the minimum tunneling distance is highlighted. It can be seen that the minimum tunneling distance is basically the depletion length of the tunneling junction at the onset of the band-to-band tunneling. Therefore, for a given material, the requirement for achieving steep subthreshold swing and large ON current is a sharp doping profile and efficient gate modulation on $W_{t,min}$ (i.e., $-\frac{\partial W_{t,\min}}{\partial V_G}$).

To achieve sharp doping profile in the tunneling junction, explorative studies were carried out investigating the impacts of doping and dopant activation techniques on the resultant doping profile for both group IV and III–V materials. For Si-, Ge-, or SiGe-based TFETs, ion implantation is the most common approach for junction doping due to well-developed CMOS technology. On the other hand, the resulted doping profile is Gaussian distribution with the standard deviation ranging from a few nm to 100 nm depending on the dopant species and the energy [18]. The formed doping area is shallow and yet sometimes not sharp enough, which is not ideal for TFETs. Therefore, many previous studies are based on SOI substrates [9, 14, 19–21], while the TFETs on bulk Si can still demonstrate outstanding performance after optimization [5, 8, 22, 23]. Huang et al. [8] reported the results

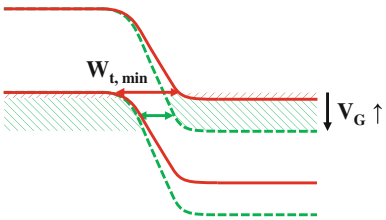

Fig. 2.3 1D band diagram of the source (tunneling) junction at different V_G bias. The shadowed region is overlapped state where tunneling can take place. And the minimum tunneling distance $W_{t,min}$ is highlighted

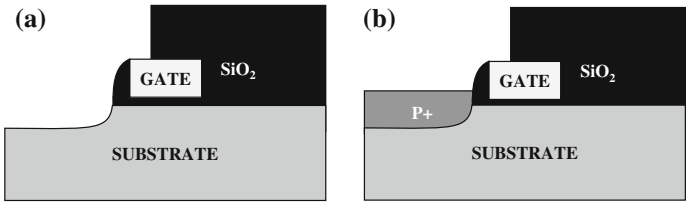

Fig. 2.4 Typical process flow of source recess and regrowth for steep source junction. **a** Source recess. **b** Source regrowth

of TFETs with striped gate structure on bulk Si using BF_2 and As ion implantation for source/drain formation, achieving 50 mV/dec minimum subthreshold swing. The ion implantation energy for BF_2 and As is 40 and 50 keV, respectively, followed by 5 s rapid thermal annealing (RTA) at 1050 °C. Demonstration of SOI, SiGeOI, and GeOI TFETs are reported by Mayer et al. [19] and Krishnamohan et al. [21]. Ion-implanted source/drain in both literature showed excellent subthreshold swing at low V_D, which is the indication that the doping profile achieved with ion implantation can still be sufficient for steep subthreshold switching.

To further optimize the tunneling junction technology, especially for non-Si material system with lower thermal budget, alternative doping techniques and annealing methods are also explored. One of the widely used techniques is the in situ doping with epitaxial growth to avoid dopant diffusion or lattice damage during dopant activation. For conventional lateral TFETs, source and drain can be formed with recess and selective-area regrowth. Figure 2.4 illustrates this step (n-type TFET as example) in contrast to step (d) in Fig. 2.2. Kim et al. [20] reported a Ge source TFET with recess/regrowth technique demonstrating subthreshold swing below 60 mV/dec and moderate ON current of 0.42 μA/μm. In addition, selective-area regrowth can also introduce strain to the channel, which can be beneficial to the tunneling process. Villalon et al. [9] reported p-type TFET on extremely thin SOI (ETSOI) substrate with strained SiGe channel and raised source/drain, demonstrating at least two decades higher ON current than previous results. This was also achieved with relatively lower RTA temperature at 950 °C, resulting less diffusion of As into the SiGe channel. On the other hand, TFETs with vertical tunneling direction are also fabricated on epitaxial substrates to exploit the steep junction achieved by high-quality epitaxial process. This is especially common in the III–V TFETs where the heterostructure can also be used for tunneling junction engineering [6, 11, 13]. Besides in situ doping with epitaxial growth, diffusion, which was long replaced by ion implantation, was again experimented to form the steep doping profile in the junction. Noguchi et al. demonstrated an InGaAs TFET with Zn-diffused source junction, showing high ON/OFF ratio and steep subthreshold swing [12]. Box-like Zn diffusion profile was achievable because the diffusion coefficient of Zn in InGaAs is proportional to the square of Zn concentration. Although only demonstrated in III–V material system, it is worth mentioning because such diffusion property might also exist for various dopant/substrate combinations.

Besides RTA, alternative annealing methods are also investigated, such as laser annealing and microwave annealing (MWA) [24]. It is worth to notice that the temperature used for microwave annealing can be as low as 490 °C as opposed to 900–1050 °C in RTA or laser annealing, resulting steeper doping profiles measured with SIMS. Yet, TFETs with steep subthreshold swing have not been demonstrated, although a comparison between the TFETs using RTA and MWA showed much improved subthreshold swing and ON current.

Source/drain technologies have been the major development in TFETs for the recent decade. More sophisticated junction designs and processes were also proposed, such as pocket doping [5, 23] and dopant segregation with alloyed contacts [23]. Heterojunction is also another direction in the pursuit for advance tunneling junction design. It would be expected that more complicated tunneling junction and fabrication process will be proposed to further enhance the tunneling property in the TFETs, and the tunneling junction design will no longer be the bottleneck in the TFET optimization.

2.2.2 Dopingless Junction

Although source/drain doping technologies have been significantly progressed during the past decade, it was still concerned that the doping fluctuation and the dopant states might limit the achievable subthreshold swing in the TFETs. Particularly, when the dimension of the devices becomes as small as a few tens of nanometers, single dopant-induced doping fluctuation can be serious as already observed in nanoscale MOSFETs. On the other hand, heavy doping results in dopant states in the bandgap of the material and the degenerate dopant states will extend the band edge into the bandgap and form a much sloped band edge. When the TFETs are operating in the subthreshold regime, the weak band-to-band tunneling through the band edge contributes to the subthreshold current and degrades the subthreshold swing of the devices. Therefore, doping the semiconductor without dopant states or actual dopants can further improve the subthreshold swing of the TFETs.

In fact, it has been observed in the threshold voltage adjustment in MOSFETs by engineering the gate metal with different workfunctions. By applying the capping materials, the bands of the semiconductor can be bent and introduce excess electrons/holes into the conduction/valence band. This process is depicted in Fig. 2.5, where a high/low workfunction metal gate bends the bands upward/downward and accumulates holes/electrons at the surface. Therefore, a novel type of TFET structure named “electron–hole bilayer TFETs” (E-H bilayer TFETs) was proposed utilizing both electrostatically induced electrons and holes [25]. The schematic view of the E-H bilayer TFET is shown in Fig. 2.6, where a double-gate structure is used. The two gates have different workfunctions and preferably induce electrons and holes at zero bias. Therefore, the electrons in the valence band of the bottom surface can tunnel into the conduction band of the top surface with sufficient gate bias, as depicted in Fig. 2.7.

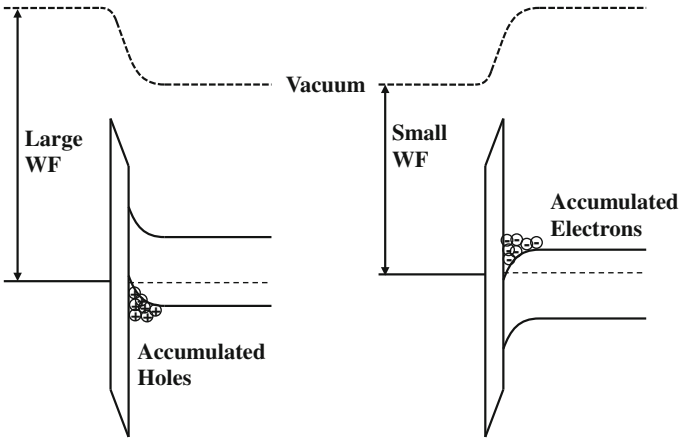

Fig. 2.5 Band diagrams of MOS structure with large and small workfunction metal gate. The holes and electrons are accumulated at the surface of the semiconductor at equilibrium

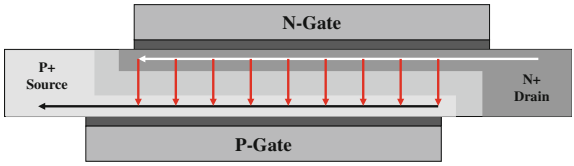

Fig. 2.6 Schematic view of the electron–hole bilayer TFET. The device is designed with slightly misaligned double-gate structure, and the electron and hole wells are formed electrostatically on the top and bottom surface

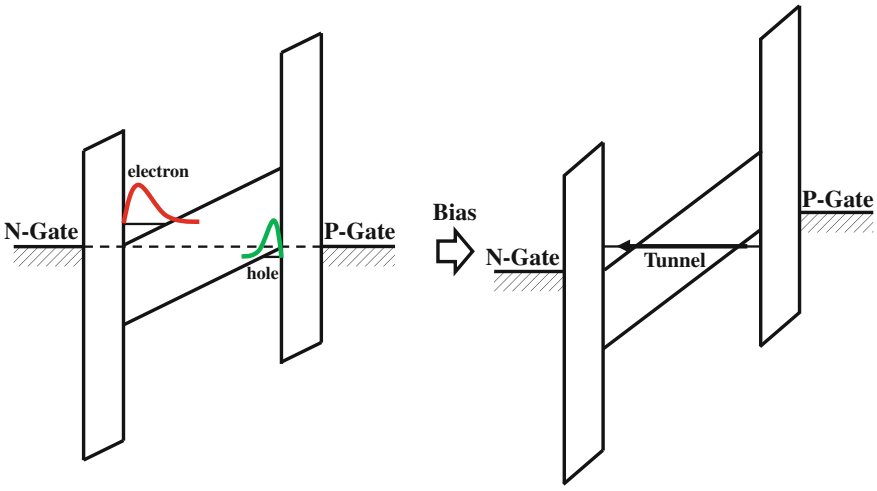

Fig. 2.7 Band diagrams of the E-H bilayer TFET in OFF (*left*) and ON (*right*) state. The eigenvalues of the electron and hole well are aligned as the voltage is applied to both the n- and p-gate with 180° phase difference

On the other hand, the parameter design for the TFETs is critical due to the trade-off between the device performance and the fabrication difficulty. Since the tunneling takes place from the bottom to top of the semiconductor slab, the thickness of the body should be thin enough to create sufficient field and shorten the tunneling distance; the bandgap of the material should also be small enough to allow tunneling in a reasonable bias condition. Teherani et al. first studied the impact of quantization energy and leakage current on the E-H bilayer TFETs in various material systems, proposing that a 15-nm InAs slab provides the optimal TFET performance with reasonable metal gate workfunctions and bias conditions [26]. Agarwal et al. [27] also investigated systematically on the engineering of the E-H bilayer TFETs in terms of gate efficiency and ON-state conductance, which corroborated the preference on the InAs E-H bilayer TFETs.

However, fabricating an InAs E-H bilayer TFET is extraordinarily challenging. Firstly, the InAs layer should be epitaxially grown and the thickness is controlled by the growth. This is due to the extreme sensitivity of the quantization energy in the electron and hole wells at the surface of the device on the thickness of the channel. Unlike the FinFETs, the thickness variation and the surface roughness from an etched InAs fin are not acceptable for the E-H bilayer TFET. Therefore, forming the bottom metal gate is one of the most challenging steps in fabricating an InAs E-H bilayer TFET. In fact, bonding and etchback process can be used to fabricate this structure, whose process flow is shown in Fig. 2.8. This process can be regarded as a revised version of creating the ultrathin body III–V-on-insulator (UTB III–V-O-I) substrate [28], except that the bottom gate is buried under the oxide before bonding.

Additionally, high-quality high-k gate dielectric on the InAs has been investigated extensively in the III–V CMOS community. Yet optimization on the high-k/InAs interface is not a trivial process, and the optimization on the InAs surface passivation and high-k deposition processes is necessary. Lin et al. demonstrated a high-quality HfO_2/InAs interface after the digital etch process [29]. The alternating oxidation-acid process removes the InGaAs/InAs layer by layer, leaving a smooth passivated surface for the high-k dielectric deposition. In this process, most of the InGaAs capping layer was etched with dry etch, which inevitably induced roughness and defects. Nevertheless, with 7–8 cycles of digital etch, the roughness and defects were removed and a fresh InAs surface was exposed.

Last but not least, contacting 15 nm (if not less) InAs is non-trivial at all. In case of MOSFETs, heavily doped capping layers were usually used for the contact. However, both n-type and p-type contact layers are needed in this case, making the process further complicated to either (1) pre-grow p+ and n+ capping layer on both sides of the InAs channel, respectively, then pattern, and recess; (2) expose both source and drain after the front gate process and then selectively regrow n+ and p+ contact layers. With either process, the fabrication complication is much worse than III–V MOSFETs or any other TFET processes. The yield and reproducibility of fabricating the E-H bilayer TFETs will become a major issue.

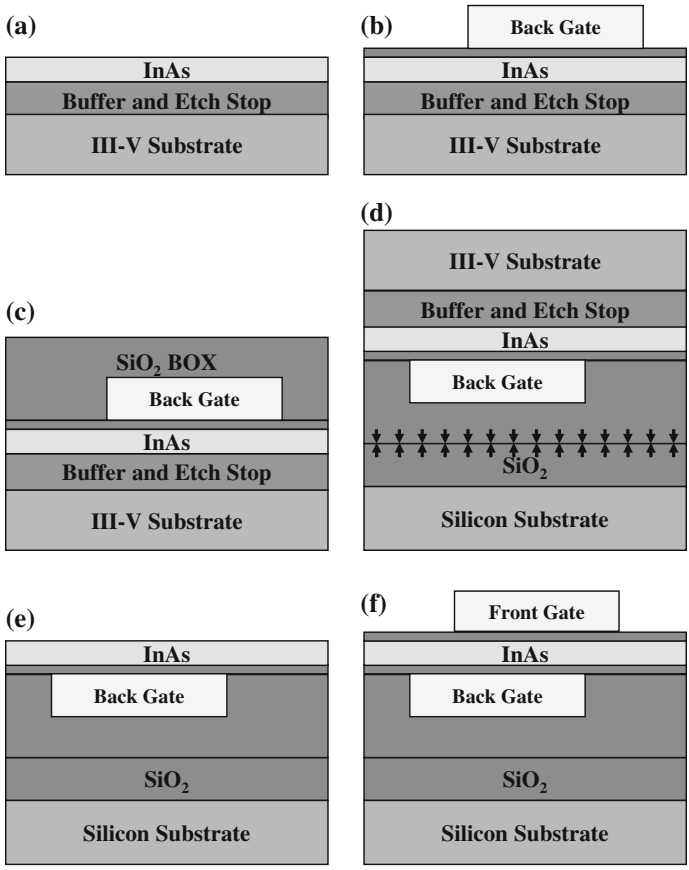

Fig. 2.8 Proposed double-gate process flow for the E-H bilayer TFET with ultrathin InAs epitaxial layer using bonding and etchback. **a** InAs epitaxial growth. **b** Back gate formation with high-k dielectrics. **c** SiO_2 box deposition and planarization. **d** Wafer bonding to a Si handle wafer with thermal oxide. **e** III–V donor wafer etch back. **f** Front gate formation

2.3 Alternative Material System and Device Geometry

This section focuses on TFET technologies on non-Si platform, including Si/Ge hetereostructure, III–V material systems, III–V/Si heterostructures, and single-layer transition metal dichalcogenide (TMD). The major achievement in recent TFET structural optimization has been in the development of various staggered gap (type-II) heterostructures. As shown in the band diagram in Fig. 2.9a, due to the direct/SRH recombination in the vicinity of tunneling interface, the bandgap of the material is a trade-off since larger bandgap material not only reduce recombination but also tunneling efficiency. On the other hand, as in Fig. 2.9b, the type-II heterostructure provides a small effective bandgap ($E_{g\text{-}eff}$) for tunneling, which is

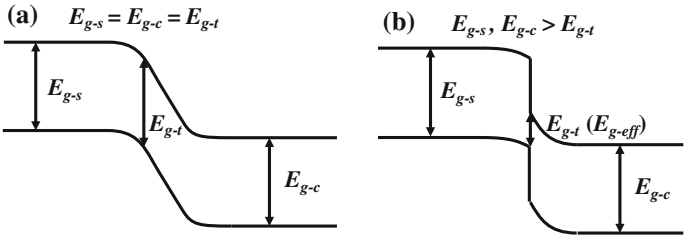

Fig. 2.9 Band diagrams of **a** homojunction and **b** heterojunction. The bandgap at source, tunneling interface, and channel are denoted by $E_{g\text{-}s}$, $E_{g\text{-}t}$, and $E_{g\text{-}c}$. Type-II heterojunction design can effectively suppress recombination in the source and channel, while the effective bandgap seen from tunneled carriers remains small

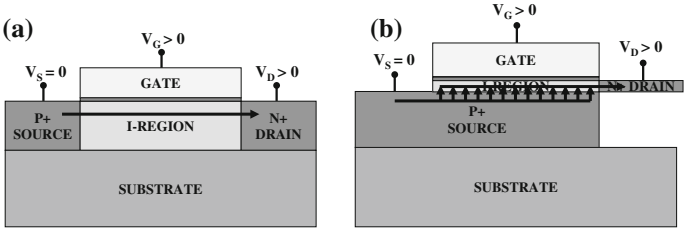

Fig. 2.10 Schematics of TFETs with tunneling direction **a** normal to the gate electric field; **b** aligned to the gate electric field

only seen at the tunneling interface, while the bandgap on both sides is large enough to suppress the recombination. Nevertheless, the bandgap at the interface is not only small to tunneling, but also to interface recombination and trap-assisted tunneling, which has been identified as the major source of parasitic thermal currents in heterostructure TFETs. Trade-off between tunneling efficiency and parasitic thermal currents with different $E_{g\text{-}eff}$ still exists in the optimization of the device design.

Besides heterostructure, tunneling direction with respect to the gate electric field and the dimensionality of the source/channel material (e.g. bulk, quantum well, nanowire, etc.) are two other perspectives in the geometrical design of the TFETs. In the conventional TFETs, the tunneling direction is perpendicular to the gate electric field, as illustrated in Fig. 2.10a. The electric field gradient along the junction line can degrade the overall gate modulation on the tunneling junction, and thus, only a very small region near the surface carries the tunneling current. To improve the gate modulation over the junction area, alternative device design embedding the source under the channel allows the tunneling align with the gate electric field (*a.k.a.* line-TFETs), as depicted in Fig. 2.10b. It has been shown that the line-TFETs can achieve both steeper subthreshold swing and higher ON current by an order of magnitude [3]. On the other hand, TFETs with different quantum confinement configurations have also been discussed, including quantum wells and nanowires with various thicknesses/diameters [30]. Exploring all different possible

combinations of material systems and geometrical design facilitates the advancements of TFETs as one of the candidates for the next-generation ultralow power CMOS technology.

2.3.1 Si/Ge/SiGe Heterojunction

Si, Ge, and SiGe alloys are considered the most potential material system for the TFETs due to their natural abundance and well-established fabrication technology. Benefit from decades of efforts to promote Si-based CMOS technology, available advanced processes developed for mass production, such as high-k dielectric, SiGe integration, and selective-area epitaxial growth, can also be used for TFET technology. Therefore, investigating the relative band alignments between the Si/Ge and Si/SiGe heterojunctions further promotes the performance of Si-based TFETs. Teherani et al. [31] reported the studies on the band alignment of the Si/Ge heterojunction, especially with strain, revealed that the Si/Ge or Si/SiGe heterojunction forms a type-II heterostructure, which is beneficial for achieving improved TFET performance. Particularly, for strained-Si (2 %)/strained-Ge (−2.7 %) grown on 42 % (Ge content) relaxed SiGe buffer, the effective bandgap of the material system is $E_{g\text{-}eff}$ = 122 meV, and the valence band offset is ΔE_v = 755 meV, as shown in Fig. 2.11.

Early demonstration of Si/Ge heterojunction TFET was made by Kim et al. [20], although only poly-Ge source with selective-area deposition was used. The key process is the source recess and boron-doped poly-Ge deposition step, which is the same as demonstrated in Fig. 2.4. Demonstrated device performance is shown to achieve improved subthreshold swing (<60 mV/dec) and ON/OFF ratio (>10^6) compared against the previous Si-based TFET results, yet the absolute drive current density per width still suffers from low tunneling efficiency mostly due to the poly-Ge source. A recent study on Ge source TFET with epitaxial boron-doped Ge source was presented on IEDM 2014 by Kim et al. [14]. Single-crystalline

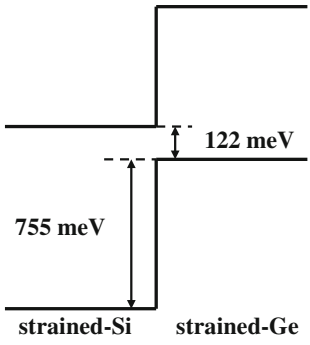

Fig. 2.11 Band alignment of strained-Si and strained-Ge on relaxed SiGe buffer with 42 % Ge. The resulting effective bandgap $E_{g\text{-}eff}$ is 122 meV, and the valence band offset ΔE_v is 755 meV

boron-doped Ge was epitaxially grown on SOI and strained-SOI substrates as the source. Higher strain level of the strained-SOI substrate showed reduced leakage current and almost 3X higher tunneling current, which may be due to the change in Si band structure when strain is present, so that the high-k/Si barrier is larger and the tunneling efficiency from Ge to Si is enhanced. On the other hand, improvements in SS are also observed with higher post-metallization annealing (PMA) temperature, which is due to the suppressed high-k/Si interface traps, and result in more efficient gate modulation.

Vertical nanowire TFETs are also attractive to TFET community due to its inherently small footprint and potential for high-density integration. Rooyackers et al. [32] proposed a complementary integration scheme for heterostructure vertical nanowire TFETs on IEDM 2013. Both n-type and p-type TFETs were fabricated on the same substrate with anisotropically etched i-Si nanowires. Different dopant species were ion implanted, and Ge source was deposited to finish the device structures. In fact, the source can be replaced by various small bandgap/electron affinity materials, so that the formed tunneling junction is the type-II heterostructure. Although the fabrication process is more complicated, the demonstrated scheme for complementary integration allows integrating low-bandgap material on Si platform with small footprint, providing high drive current and low leakage current.

2.3.2 III–V Homojunction

Further step toward band engineering for the tunneling junction of TFETs is III–V material system, which provides wide variety of bandgaps and band alignments to choose. Moreover, the small electron mass and direct bandgap in narrow-gap III–V materials also made them preferable for tunneling. Bandgap engineering with III–V homojunction enables the optimization on the trade-off between the leakage current and tunneling efficiency. Specifically, the main source of leakage current floor in TFETs comes from the gate leakage and the carrier recombination process, including direct recombination and Shockley–Reed–Hall recombination [33, 34]. With larger bandgap, the carrier recombination process can be suppressed, while the tunneling efficiency is also reduced due to larger tunneling barrier. In contrast, small bandgap material can provide larger tunneling current, while the leakage current floor is not well controlled.

$In_xGa_{1-x}As$ is the most popular material for III–V homojunction TFETs, especially $In_{0.53}Ga_{0.47}As$ which is lattice matched to InP and can be epitaxially grown with extremely low-defect densities [35]. $In_xGa_{1-x}As$ has a wide range of tunable bandgap from $\sim$0.35 eV (InAs) to $\sim$1.42 eV (GaAs) and electron effective mass as low as 0.023–0.046 m_0 [36]. These parameters made InGaAs preferable material for TFETs except that a steep junction is difficult to fabricate. As mentioned in

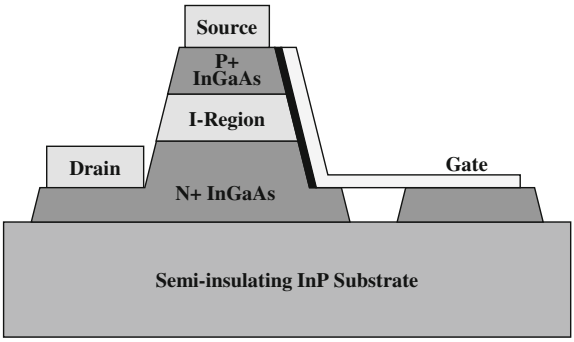

Fig. 2.12 Schematic view of the vertical InGaAs TFET with sidewall gate

Sect. 2.2.1, doping InGaAs with Zn using diffusion has shown to be successful in forming box-like doping profile and sharp tunneling junction. With this method, the fabrication process is completely compatible to the well-established MOSFET process, and complementary integration is achievable with donor dopant with similar diffusion property as Zn in InGaAs.

Alternatively, Dewey et al. [6] demonstrated a vertical TFET structure with a sidewall gate as shown in Fig. 2.12. With the in situ doping during epitaxial growth, steep doping profile can be achieved without ion implantation and annealing. However, there are also three issues come with this approach. First, the surface roughness of the sidewall is usually worse than the surface due to the plasma etching damage, making the high-k/III–V interface susceptible to interface traps and degrades gate modulation; second, the channel length is determined by the thickness of the i-layer, which means that it is impossible to integrate TFETs with different channel length; and lastly, integrating both n-type and p-type TFET on the same substrate becomes challenging due to different requirements for the junction. Fortunately, these issues can be solved with more advanced processes, or potentially with novel circuit design paradigms in the future. Aspect ratio trapping is one of the most active fields of study for heterointegration [37, 38]. Selectively growing III–V material on Si with high aspect ratio trenches enables high-quality material grown without lattice matching. The schematic steps of this process are shown in Fig. 2.13. Performing the selective-area growth with different epistructures makes integration of both n-type and p-type TFETs and TFETs with different gate length possible. Then, the sidewall of the III–V fins can be exposed by removing the oxide, leaving the sidewall surface without plasma damage. Moreover, additional digital etch process [29] can also remove the sidewall surface in a well-controlled manner. Alternating self-limiting oxidation and acid cleaning, the surface of the InGaAs can be uniformly etched 0.8–1.5 nm/cycle (depending on the material and oxidation condition), which can remove the surface damage within several cycles.

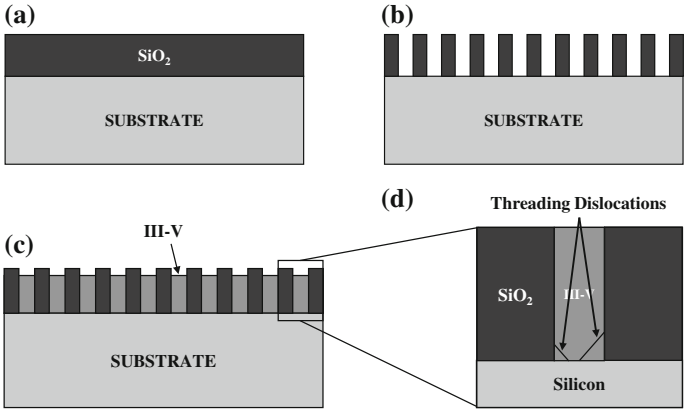

Fig. 2.13 Schematic view of the selective-area growth of III–V material on Si with high aspect ratio threading dislocation trapping trenches. **a** Si substrate with thick SiO_2. **b** High aspect ratio trenches are etched in SiO_2. **c** III–V material grown in the trenches. **d** Threading dislocations are terminated by the high aspect ratio trenches

2.3.3 III–V Heterojunction

Heterostructure tunneling junction is another important reason that III–V material systems are intriguing to the TFET community. There are many aspects in heterojunction engineering for TFETs, including bandgap engineering, band alignment engineering, and strain engineering. As mentioned in Sect. 2.3.1, type-II heterostructure is the most intriguing heterostructure for TFETs, and the most commonly used type-II heterostructure includes $In_xGa_{1-x}As/In_yGa_{1-y}As$, $In_xGa_{1-x}As/GaAs_ySb_{1-y}$, and $InAs/Al_{1-x}Ga_xSb$. Unlike Si/Ge system, source recess and regrowth are rarely seen due to the overcomplicated process and demanding growth conditions [39]. The starting substrate for the III–V heterojunction TFETs are, therefore, mostly epitaxially pre-grown and the tunneling interface is always aligned to the surface of the substrate. Dewey et al. also presented the first $In_{0.53}Ga_{0.47}As/In_{0.7}Ga_{0.3}As$ heterostructure TFET with sub-60 mV/dec subthreshold swing [6]. By inserting a 6-nm $In_{0.7}Ga_{0.3}As$ layer at the tunneling interface, the tunneling barrier is reduced from 0.74 eV to 0.59 eV, resulting improved ON current and hence steeper switching at low current level. Further investigation was reported by Zhao et al. [15], where nanowire InGaAs/InAs heterojunction TFETs are demonstrated. Taking advantage of the digital etch mentioned above, nanowires as thin as 15 nm can be fabricated using top-down approach, as shown in Fig. 2.14. Then, the surround gate and source/drain contacts are formed with planarization and recess steps, such that the alignment of the gate with respect to the tunneling junction can be controlled within 10-nm. The minimum room temperature subthreshold swing reported for a single nanowire TFET is 75 mV/dec. Temperature-dependent measurements showed that the subthreshold swing has

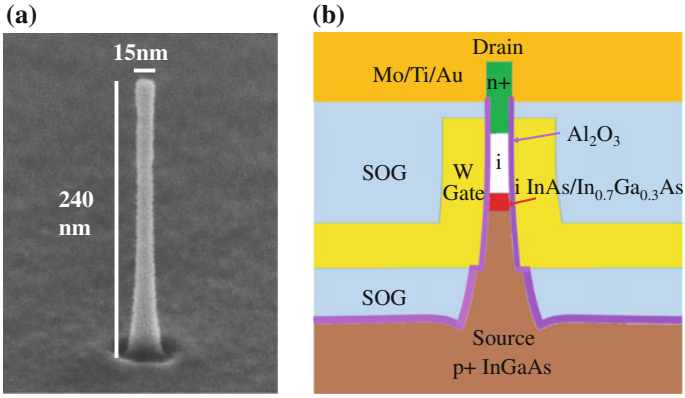

Fig. 2.14 **a** $In_{0.53}Ga_{0.47}As$ nanowire with 15 nm diameter and 240 nm length fabricated with top-down technology. **b** Schematic view of the InGaAs/InAs heterostructure nanowire TFET. © 2014 IEEE. Reprinted, with permission, from [15]

strong temperature dependence, indicating the contribution of thermally activated current to the room temperature subthreshold swing, most likely from trap-assisted tunneling or tunneling-assisted generation.

$In_xGa_{1-x}As/GaAs_ySb_{1-y}$ is another potential material system becoming more popular as the effective bandgap, $E_{g\text{-}eff}$, between $In_xGa_{1-x}As$ and $GaAs_ySb_{1-y}$ can vary from 270 meV (lattice matched to InP) all the way to −100 meV (lattice matched to GaSb). This unique property of the $In_xGa_{1-x}As/GaAs_ySb_{1-y}$ heterostructure gives plenty of room for the band alignment engineering to achieve the optimized device performance. Mohata et al. studied the impact of different the growth interfaces between InGaAs and GaAsSb on the device performance [40]. The demonstrated devices have the same device structure as shown in Fig. 2.12 but with different process flow. It was concluded that the In-As-terminated interface yields 3–4 orders of magnitude lower OFF current, which gives ON/OFF ratio greater than 10^4 and minimum subthreshold swing 169 mV/dec. This is due to the defects formed when the strain in the Ga-As-terminated interface relaxed, while the In-As-terminated interface is pseudomorphic without significant defect formation.

An extreme of the $In_xGa_{1-x}As/GaAs_ySb_{1-y}$ material system is InAs/GaSb which forms approximately −100 meV broken gap heterostructure (as in bulk). Zhou et al. demonstrated a vertical InAs/GaSb TFET with ON current as high as 180 μA/μm [11]. This TFET adopted the line-TFET design with 6 nm InAs, which is schematically shown in Fig. 2.15. With the ultrathin InAs channel design, the gate modulation on the tunneling interface is strong and uniform and the quantization energy in the InAs results in almost zero-gap heterojunction. In order to achieve the line-TFET structure without parasitic tunneling path, undercutting the GaSb is necessary. With isotropic NH_4OH wet etch, GaSb can be selectively etched and result in the GaSb pillar. With the drain capping layer/contact metal and the gate stack, the InAs layer atop is remained and provides a conduction channel from the tunneling junction to the drain.

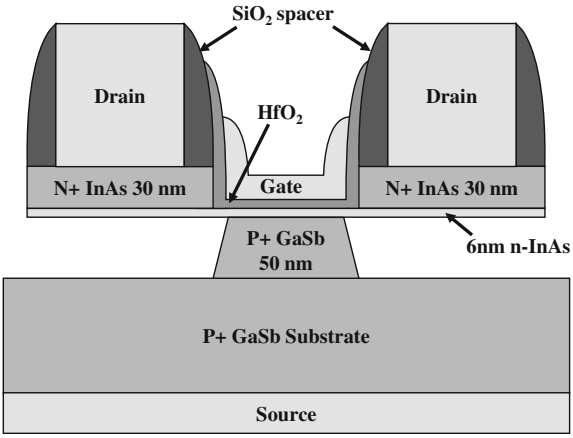

Fig. 2.15 Schematic view of the InAs/GaSb vertical TFET with tunneling direction aligned to the gate electric field. GaSb pillar structure was fabricated by undercutting the GaSb from both sides to avoid direct tunneling under the drain contact

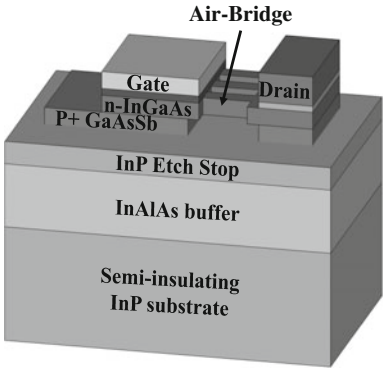

Fig. 2.16 InGaAs/GaAsSb quantum-well TFET with ultrathin 15 nm n-InGaAs on 15 nm p+ -GaAsSb. Air-bridge structure was fabricated to isolate the drain and the tunneling area

Alternative line-TFET structure with $In_{0.53}Ga_{0.47}As/GaAs_{0.5}Sb_{0.5}$ heterostructure was the $In_{0.53}Ga_{0.47}As/GaAs_{0.5}Sb_{0.5}$ quantum-well TFETs reported by Yu et al. [13] Different from the previous GaSb pillar supported structure, the TFET shown in Fig. 2.16 forms a 1-μm InGaAs air bridge, connecting the $In_{0.53}Ga_{0.47}As$ channel and the drain. With the air-bridge structure, the undercutting distance of the $GaAs_{0.5}Sb_{0.5}$ is much reduced from the entire contact size (up to microns) down to the width of the air bridge (150 nm). Moreover, this device structure also enables quantum-well to quantum-well tunneling with both ultrathin $In_{0.53}Ga_{0.47}As$ and $GaAs_{0.5}Sb_{0.5}$, which can potentially achieve dimensionality switching and gives steep subthreshold swing over a wide range of drive current [30]. The challenge in the device fabrication of the $In_{0.53}Ga_{0.47}As/GaAs_{0.5}Sb_{0.5}$ quantum-well TFETs is the undercutting of the GaAsSb. Since the GaAsSb layer is only 15 nm thick, the solution may be blocked by the surface tension and result in non-uniform undercut along the air bridge. Also, different from GaSb, NH_4OH solution does not etch GaAsSb. Therefore, 100:1 H_2O_2:NH_4OH with ultrasonication is used in the process

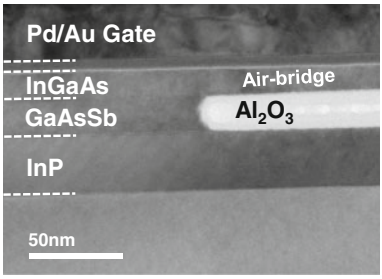

Fig. 2.17 TEM image of the InGaAs/GaAsSb quantum-well TFET at the edge of the air bridge. Successful undercut was achieved with highly selective GaAsSb wet etch over InGaAs

to undercut the GaAsSb and form the air-bridge structure. The cross-sectional TEM image of the finished air-bridge structure shown in Fig. 2.17 confirmed the success of the process.

2.3.4 III–V on Silicon

Recent progress in heterointegration of III–V materials on Si substrate has enabled the possibility of using III–V/Si heterojunction in TFETs. Since the extensive studies have demonstrated superior tunneling properties in III–V material systems, especially $In_xGa_{1-x}As$, integrating a III–V channel directly on Si substrate can potentially achieve improved TFET performance with lower leakage current and the possibility of similar manufacturing cost as the state-of-the-art CMOS technology. However, due to the 11.6 % lattice mismatch between Si and InAs, it is almost impossible to grow wafer-scale defect-free InAs directly on Si without a buffer. Nevertheless, it is possible to grow nanoscale InAs features on Si using selective-area epitaxial growth such that the InAs relaxes and the misfit dislocations with local strain only appear at the interface. Tomioka et al. [41] reported the selective-area metalorganic vapor phase epitaxy (MOVPE) of InAs nanowires directly on Si substrate. The selective-area property only allows the InAs to grow in the opening of the hardmask and perpendicular to the substrate surface. Figure 2.18 shows the transmission electron microscopy (TEM) images of the InAs nanowire where the expansion of the nanowire due to the lattice relaxation and the misfit dislocations at the heterointerface is observed. Furthermore, core–shell or even core–multishell nanowire structure can also be achieved with similar technology [42]. As shown in Fig. 2.19, a δ-doping layer, capping layers, etc., were grown on the exterior of the $In_{0.7}Ga_{0.3}As$ nanowire. Although TFET results were not reported in [42], it can be expected that the modulation doping technology or advanced gate stacks used for the planar TFETs are again available for nanowire TFET, which can further improved the extraordinary results reported in [7].

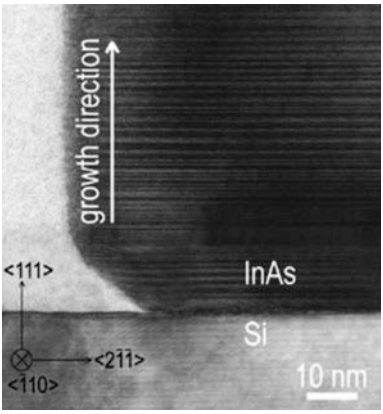

Fig. 2.18 TEM image of the InAs nanowire directly grown on Si. The 11.6 % strain was relaxed and resulted in the dilated InAs nanowire. Reprinted with permission from [41]. Copyright (2008) American Chemical Society

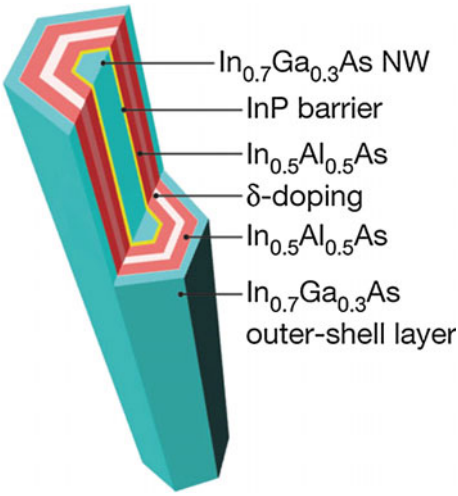

Fig. 2.19 Cross-sectional view of the core–multishell nanowire structure. Modulation doping was achieved with this design. Reprinted by permission from Macmillan Publishers Ltd: Nature [42], copyright (2012)

2.3.5 2D Material System

Rapid growth in the research on the 2D material system inspired the attempts on fabricating TFETs using 2D material. Graphene, MoS_2, WSe_2, black phosphorous, etc., are popular 2D materials with different band properties. Especially, transition metal dichalcogenides (TMDs) are one of the most promising material systems exhibiting a wide selection of bandgaps and band alignments [43, 44]. The TFET structures using 2D materials resemble those with bulk material systems, consisting of heavily doped source/drain and intrinsic channel, where doping these 2D materials is also essential to create high-performance TFETs. However, doping 2D materials have been challenging tasks, especially for complementary doping in the same piece of materials. For the past years, doping TMDs have been a popular topic

in the research of TMD-based transistors [45–51]. For example, WSe_2 and MoS_2 are one of the most popular TMD materials for fabricating MOSFETs due to their relatively high electron mobility and finite bandgap. Hence, effectively doping WSe_2 and MoS_2 in a well-controlled manner has been extensively studied in the community of 2D materials.

Previous studies have demonstrated that potassium and NO_2 are the doping species for n-type and p-type, respectively [45, 48]. Fang et al. [48] reported an experimental result of degenerate doping of few-layer MoS_2 and WSe_2 by surface charge transfer with potassium, where the sheet charge density of 10^{12}–10^{13} cm^{-2} has been achieved. The doping procedure was carried out by exposing the sample substrate to the K vapor in a sealed chamber with a K dispenser. Longer exposure time results in higher measured sheet charge transferred into the MoS_2/WSe_2. XPS results also confirmed that the K doping process on MoS_2/WSe_2 is similar to those reported on graphite and molecular films. Similar process of p-type doping WSe_2 with NO_2 was also reported, which contributed to a fabricated WSe_2 p-type MOSFET with the subthreshold swing of $\sim$60 mV/dec and the I_{ON}/I_{OFF} ratio greater than 10^6 [45]. These aforementioned results confirmed that doping MoS_2 and WSe_2 with both n-type and p-type dopants is feasible for the purpose of TFET fabrication. Unfortunately, due to the possible surface reaction with air/water, achieving air-stable doping for MoS_2 is rather challenging [50].

Alternative doping techniques are proposed for non-degenerate/degenerate doping of MoS_2 with various dopant sources, including spin-on chemicals, polymers, and plasma [46, 47, 49, 50]. Du et al. [47] investigated molecular doping of MoS2 with polyethyleneimine (PEI), which resulted in 2.6 times reduction in the sheet resistance of MoS_2. The doping procedure is rather straightforward: soak in PEI solution (0.02 wt% in methanol) for 24 h and then rinse with methanol and blow dry with N_2. The finished device exhibited 5.6 nm root-mean-square (RMS) surface roughness, which is due to the incorporation of PEI molecule in the monolayer MoS_2. In addition, similar approach using benzyl viologen (BV) was also studied by Kiriya et al. [50]. The doping of MoS_2 with BV can be performed by either drop-casting BV solution onto the substrate or immersion of the substrate into the BV solution for 12 h, then blow dry with N_2. The achieved sheet charge in the few-layer MoS_2 is around 1.2×10^{13} cm^{-2}, and it is worth mentioning that the doping with BV is air-stable, where the conductance characteristics of the doped MoS_2 flake remained almost the same after 1 day of exposure to air (Fig. 2.20).

Recently, Park et al. [51] demonstrated a wide-range controllable n-type doping technique for MoS_2 with phosphorous silicate glass. The doping procedure is relatively more complicated than the previously mentioned processes. The process flow includes a two-step annealing process at 700–900 °C for diffusion and 500 °C for activation, respectively, followed by an optical activation with power greater than 5 μW for wavelength λ = 655 nm, or 10 μW for λ = 520 and 785 nm. By controlling the annealing temperature of the first annealing step, the resulting sheet charge in the MoS_2 can range from 3.6×10^{10} cm^{-2} to 8.3×10^{12} cm^{-2}, which is so far the broadest among the published results.

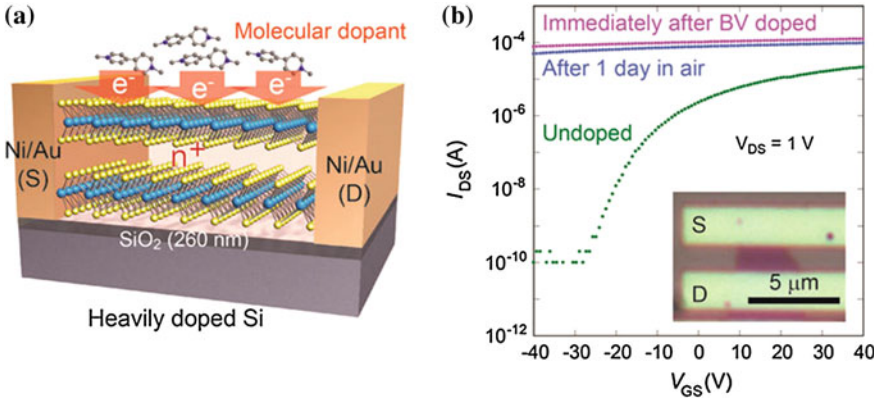

Fig. 2.20 **a** Schematic view of doping multilayer MoS_2 with molecular dopant BV. **b** The I-V characteristics of a MoS_2 sheet with and without BV doping. Reprinted with permission from [50]. Copyright (2014) American Chemical Society

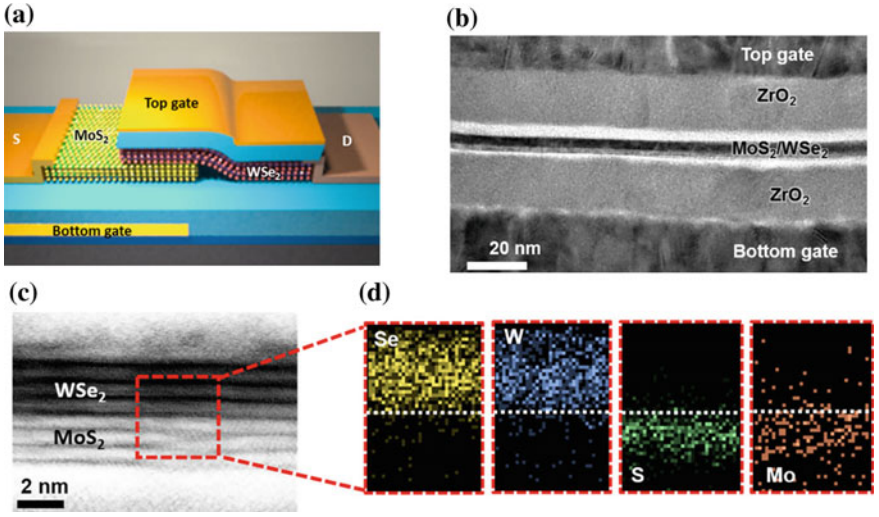

Fig. 2.21 **a** Three-dimensional schematic of the device (not to scale). **b** Cross-sectional TEM image of a representative device, showing the symmetric dual-gate structure. **c** High-resolution STEM image of the same heterostructure, consisting of 4 layers of MoS_2 and WSe_2. **d** EDS mapping of the heterostructure. Reprinted in part with permission from [53]. Copyright (2015) American Chemical Society

Similar to the III–V material systems, heterostructure is also available to the TMD materials due to the difference in the bandgaps and band alignments between different TMD materials. In particular, MoS_2 and WSe_2 are intrinsically n- and p-type and they form type-II band alignment [44], which is preferred in the heterostructure design in the TFETs. Interestingly, although epitaxial growth of

MoS_2/WSe_2 heterostructure is challenging [52], single-/multilayer MoS_2 and WSe_2 can be easily stacked together and coupled by van der Waals force. Roy et al. [53] demonstrated the band-to-band tunneling device with stacked WSe_2/MoS_2 heterostructure. The schematic view of the device is shown in Fig. 2.21a, where the overlapping between WSe_2 and MoS_2 is the region where band-to-band tunneling takes place. By positioning the WSe_2 on top of MoS_2 with transferring techniques commonly used for 2D materials, the WSe_2 is held down by the van der Waals force between the two layers. It can be seen from the high-resolution STEM image of the heterostructure in Fig. 2.21b, c that the two layers are perfectly on top of each other forming the heterojunction. And the EDS mapping of the heterostructure also confirmed the materials. The measurement results of the fabricated device showed excellent rectifying characteristics, which can be further referred from the original literature [53].

2.4 TFET Characterization

This section focuses on some characterization aspects of the TFETs, including threshold voltage and a new technique of extracting achievable subthreshold swing of a given tunneling junction from the backward/Esaki diode characteristics. Due to the differences in the operation mechanism and bias conditions between MOSFETs and TFETs, some of the conventional characterization methodologies for the MOSFETs are no longer applicable to the TFETs. One most obvious difference is the asymmetry in the output characteristics (i.e., fixed gate–source voltage, sweep drain–source voltage), as schematically shown in Fig. 2.22, due to the asymmetry of the source and drain. The bias conditions of the TFETs are also different from the MOSFETs. For example, the threshold voltage (V_T) of a MOSFET is determined by the onset of strong inversion, while V_T of a TFET should be determined by the

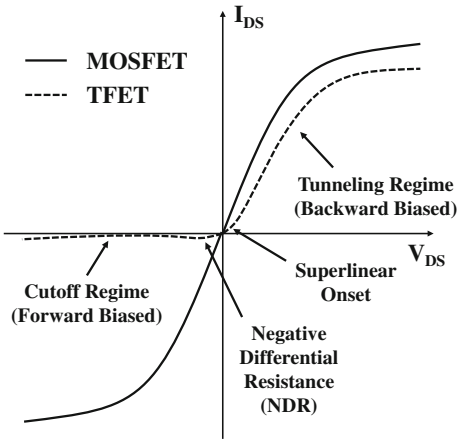

Fig. 2.22 Comparison between the output characteristics of a MOSFET and a TFET. The most obvious discrepancy in the output characteristics between MOSFETs and TFETs is in the negative V_{DS} domain, where the MOSFETs exhibit almost symmetrical behavior, while the TFETs has negative differential resistance and cutoff regime in the negative V_{DS} domain

onset of strong band-to-band tunneling, which is already in the "strong inversion regime" of the MOSFET counterpart. Therefore, some characterization methodologies of the TFETs should be revised and corroborated by the physics and be invariant to the geometric dimensions.

2.4.1 Threshold Voltage (V_T)

As just mentioned above, the V_T of the MOSFETs and TFETs should be defined at different bias conditions based on the switching mechanism. On the other hand, the most commonly used definition is by constant current (i.e., the gate voltage at where the drain current reaches 10^{-7} A/μm). Yet, it is unclear whether this V_T definition still gives physical insights on the TFETs. Boucart and Ionescu investigated the robustness of different definitions of the threshold voltage for the TFETs with the assistance of TCAD simulation results [54]. It was shown that the constant current extraction yields relatively lower V_T than the saturation of the tunneling distance ($W_{t,min}$), and the empirical current level (10^{-7} A/μm) used for the extraction is considerably arbitrary for the TFETs with various tunneling junctions. Boucart and Ionescu also proposed an alternative V_T extraction scheme which defines the V_{TG} as the gate voltage where the derivative of the transconductance (dg_m/dV_G) reaches the maximum. The proposed V_T extraction scheme physically defined the transition between the quasi-exponential and linear dependence of the drive current on the gate bias. With the transconductance-based extraction method, the extracted V_{TG} does not decrease with L_g, whereas the constant current method gives lower V_T due to the increase of subthreshold current induced by the direct tunneling leakage in the TFETs with smaller L_g (as shown in Fig. 2.23).

Furthermore, Lee and Choi [55] discussed the impact of the inversion layer on the turn-on of the TFETs. Similar to the MOSFETs, an inversion layer will form at

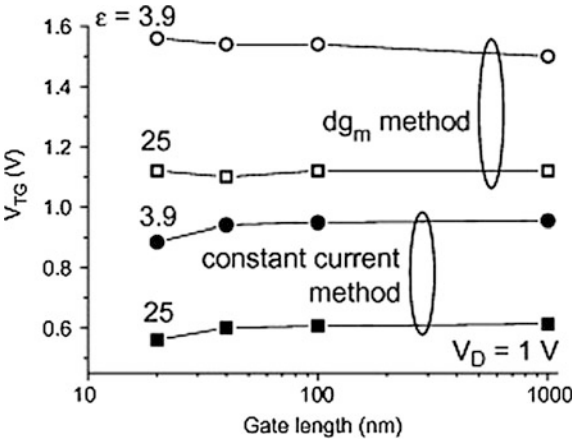

Fig. 2.23 Extracted threshold voltage using constant current method and maximum transconductance slope (dg_m) method at V_D = 1 V. Constant current method shows V_T roll-off due to the increase in subthreshold leakage current, while the dg_m method is still robust for short-channel devices. Reprinted from solid-state electronics [54], Copyright (2008), with permission from Elsevier

the surface of the channel with $V_{GS} = V_{INV}$. The electric field from the gate is then screened by the inversion layer and cannot modulate the bands further. On the other hand, the onset of the tunneling process is related to another voltage V_{ON}. If the inversion layer was formed before the tunneling process saturated, the increase of the drive current would transit from semi-exponential to linear due to the screening from the inversion charge. The corresponding threshold voltage V_T is the same as the MOSFETs ($V_T = V_{INV}$). On the contrary, if the tunneling process saturated (i.e., $W_{t,min}$ saturates) with V_{GS} before the inversion layer is formed, the corresponding V_T should be related to V_{ON} instead of V_{INV}. It has been shown in [55] that both V_{INV} and V_{ON} are V_{DS} dependent: A larger V_{DS} gives smaller V_{ON} until saturation and larger V_{INV} almost linearly. This behavior is expected since the tunneling process involves the source junction, while the inversion layer comes from the drain. Therefore, it is more reasonable to extract V_T of a TFET with the aforementioned method with high V_{DS}, while the constant current method still works for the low V_{DS} case.

2.4.2 Extracting Subthreshold Swing from Diode

The fabrication processes for the TFETs become much more complicated over the years of progress. It is challenging and may take several iterations to develop a complete process flow for a TFET, so that the performance optimization for the devices is getting more and more difficult. Particularly, optimizing the tunneling junction for the steepest subthreshold swing is one of the most essential goals. However, with the non-idealities in other parts of the device, such as high-k/semiconductor interface traps and series resistance, optimizing the tunneling junction alone is infeasible and may be obscured. Therefore, it is desired to be able to extrapolate TFET performance from the tunneling junction alone, i.e., the tunneling diode. It is not only much simpler in analysis but also more fabrication feasible to work on a tunneling diode than a complete TFET.

Agarwal and Yablonovitch [56] recently proposed that plotting the absolute conductance against the applied voltage of the diode can reflect, to certain extent, the achievable subthreshold swing of a TFET An example is shown in Fig. 2.24, where the current–voltage characteristics of two diodes are plotted. The black curve represents the characteristics of a backward diode, while the red curve plots an Esaki diode with NDR present. Although the I-V characteristics are totally different and it is difficult to judge which one may give better tunneling steepness as a tunneling junction, it can be seen that they actually give similar tunneling steepness at 54 and 60 mV/dec, respectively. This means that the potentially achievable TFET subthreshold swing with these tunneling junctions can actually achieve sub-60 mV/dec, disregard of any other non-idealities induced from the gate.

The physical interpretation of the relation between the conductance steepness and the TFET subthreshold swing can be derived from the basic equation of the tunneling current density

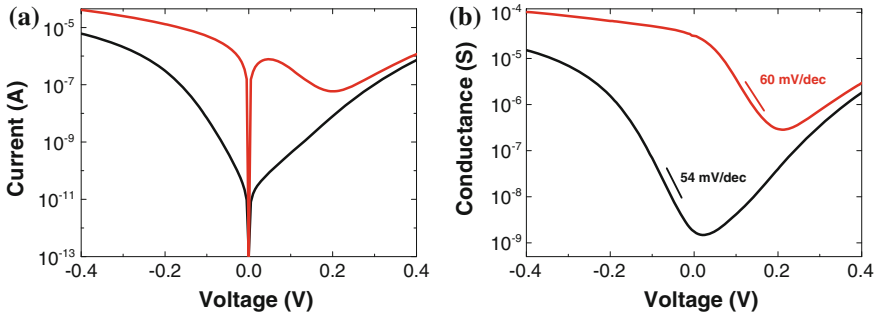

Fig. 2.24 **a** I-V characteristics of two diodes: One is a backward diode and the other is an Esaki diode. **b** Corresponding absolute conductance versus voltage plot of the two diodes shows similar steepness around 60 mV/dec

$$J \propto \int (f_C - f_V) \times T \times D_J(E) \times dE$$

where f_C and f_V are the Fermi occupancy in the conduction band and valence band of the n- and p-region of the diode, respectively; T is the tunneling probability; $D_J(E) = N_C(E)N_V(E)$ is the joint density of states between the valence band in the p-region and the conduction band in the n-region.

To the first order, with small enough applied voltage ($V_a < 8$ kT or 200 mV), $(f_C - f_V) \propto V_a$, so that

$$G = \frac{J}{V_a} \propto \frac{\int (f_C - f_V) \times T \times D_J(E) \times dE}{\int (f_C - f_V) dE}$$

and

$$\text{conductance slope} = \frac{dV_a}{d \log G} = \frac{dV_a}{d\langle T \times D_J(E)\rangle} \approx \frac{dV_a}{dT} \approx SS$$

where the pointed bracket represents the weighted average by $(f_C - f_V)$.

Although the extrapolation gives certain information on the achievable subthreshold swing of the TFETs, the absolute value of the conductance steepness may not be directly translated to the subthreshold swing of the TFETs due to the weighted average by $(f_C - f_V)$ [57]. However, it is useful for comparing the potential performance of different tunneling junctions. Those with steeper conductance steepness close to the origin usually exhibit steeper subthreshold swing if they were made into TFETs. Even the non-idealities in the tunneling junction, such as carrier recombination and trap-assisted tunneling, can also degrade the conductance steepness just as they do on the subthreshold swing of the TFETs. Additionally, the conductance steepness in the regime of negative differential resistance (NDR) is meaningless. This is due to the unreliable I–V measurement in the NDR region as

neither a regulated voltage source nor a current source is stable in this region. Extra measurement techniques, such as adding a parallel resistor to the device, are necessary to establish reliable measurements.

2.5 Conclusion

The fabrication technologies of the TFETs are introduced in the chapter, from the simplest generic CMOS-like process to the complicated process for E-H bilayer TFETs. The evolvement of the TFET design requires novel technology such as selective-area epitaxy, novel annealing technology, and bonding/etchback. The active research in the field of TFETs enabled these technologies, and significant progresses were made to improve the device performance by orders of magnitude. On the other hand, two novel characterization methods for TFETs are also discussed to give some insights on the analysis on the devices. Particularly, extracting subthreshold swing from the diode characteristics can be useful for the optimization of the tunneling junction when fabricating the complete TFET is challenging.

Acknowledgments Part of the work presented in this chapter was supported by the Center for Energy Efficient Electronics Science (NSF Award 0939514).

References

1. J. Appenzeller, Y.-M. Lin, J. Knoch, P. Avouris, Band-to-Band Tunneling in Carbon Nanotube Field-Effect Transistors. Phys. Rev. Lett. (2004). doi:10.1103/PhysRevLett.93.196805
2. A.C. Seabaugh, Q. Zhang, Low-voltage tunnel transistors for beyond CMOS logic. Proc. IEEE **98**, 2095–2110 (2010). doi:10.1109/JPROC.2010.2070470
3. A.M. Ionescu, H. Riel, Tunnel field-effect transistors as energy-efficient electronic switches. Nature **479**, 329–337 (2011). doi:10.1038/nature10679
4. H. Lu, A. Seabaugh, Tunnel Field-effect transistors: state-of-the-art. Electron Devices Soc. IEEE J. **2**, 44–49 (2014). doi:10.1109/JEDS.2014.2326622
5. C. Hu, P. Patel, A. Bowonder, et al, Prospect of tunneling green transistor for 0.1 V CMOS, in: *2010 IEEE International Electron, Devices Meeting (IEDM)*, pp. 16.1.1–16.1.4 (2010)
6. G. Dewey, B. Chu-Kung, J. Boardman, et al, Fabrication, characterization, and physics of III-V heterojunction tunneling field effect transistors (H-TFET) for steep sub-threshold swing, in *2011 IEEE International, Electron Devices Meeting (IEDM)*, pp 33.6.1–33.6.4 (2011)
7. K. Tomioka, M. Yoshimura, T. Fukui, Steep-slope tunnel field-effect transistors using III-V nanowire/Si heterojunction, in *2012 Symposium on VLSI Technology (VLSIT)*, pp. 47–48 (2012)
8. Q. Huang, R. Huang, Z. Zhan, et al, A novel Si tunnel FET with 36 mV/dec subthreshold slope based on junction depleted-modulation through striped gate configuration, in *Electron Devices Meeting (IEDM), 2012 IEEE International*, pp. 8.5.1–8.5.4 (2012)
9. A. Villalon, C. Le Royer, M. Casse, et al, Strained tunnel FETs with record I_{ON}: first demonstration of ETSOI TFETs with SiGe channel and RSD, in *2012 Symposium on VLSI Technology (VLSIT)*, pp 49–50 (2012)

10. RBijesh R, Liu H, Madan H, et al, Demonstration of $In_{0.9}Ga_{0.1}As/GaAs_{0.18}Sb_{0.82}$ near broken-gap tunnel FET with I_{ON} = 740 μA/μm, G_M = 70μS/μm and Gigahertz switching performance at V_{DS} = 0.5 V, in *2013 IEEE International*, *Electron Devices Meeting (IEDM)*, pp 28.2.1–28.2.4 (2013)
11. G. Zhou, R. Li, T. Vasen, et al, Novel gate-recessed vertical InAs/GaSb TFETs with record high I_{ON} of 180 μA/μm at V_{DS} = 0.5 V, in *2012 IEEE International*, *Electron Devices Meeting (IEDM)*, pp. 32.6.1–32.6.4 (2012)
12. M. Noguchi, S. Kim, M. Yokoyama, et al, High I_{on}/I_{off} and low subthreshold slope planar-type InGaAs tunnel FETs with Zn-diffused source junctions, in *2013 IEEE International*, *Electron Devices Meeting (IEDM)*, pp. 28.1.1–28.1.4 (2013)
13. T. Yu, J.T. Teherani, D.A. Antoniadis, J.L. Hoyt, Quantum-well Tunnel-FETs with tunable backward diode characteristics. IEEE Electron Device Lett. **34**, 1503–1505 (2013). doi:10.1109/LED.2013.2287237
14. M. Kim, Y. Wakabayashi, R. Nakane, et al, High I_{on}/I_{off} Ge-source ultrathin body strained-SOI tunnel FETs. In: *2014 IEEE International*, *Electron Devices Meeting (IEDM)*, pp. 13.2.1–13.2.4 (2014)
15. X. Zhao, A. Vardi, J.A. del Alamo, InGaAs/InAs heterojunction vertical nanowire tunnel fets fabricated by a top-down approach, in *2014 IEEE International*, *Electron Devices Meeting (IEDM)*, pp. 25.5.1–25.5.4 (2014)
16. B. Rajamohanan, R. Pandey, V. Chobpattana et al., 0.5 V supply voltage operation of $In_{0.65}Ga_{0.35}As/GaAs_{0.4}Sb_{0.6}$ tunnel FET. IEEE Electron Device Lett. **36**, 20–22 (2015). doi:10.1109/LED.2014.2368147
17. L. Zhang, M. Chan, SPICE modeling of double-gate tunnel-FETs including channel transports. IEEE Trans. Electron Devices **61**, 300–307 (2014). doi:10.1109/TED.2013.2295237
18. J.D. Plummer, M. Deal, P.D. Griffin, *Silicon VLSI technology: fundamentals, practice, and modeling*, 1st edn. (Prentice Hall, Upper Saddle River, NJ, 2000)
19. F. Mayer, C. Le Royer, J.F. Damlencourt, et al, Impact of SOI, $Si_{1-x}Ge_xOI$ and GeOI substrates on CMOS compatible Tunnel FET performance, in *Electron Devices Meeting, 2008*, *IEDM 2008*, *IEEE International*, pp. 1–5 (2008)
20. S.H. Kim, H. Kam, C. Hu, T.-J.K. Liu, Germanium-source tunnel field effect transistors with record high I_{ON}/I_{OFF}, in *2009 Symposium on VLSI Technology*, pp. 178–179 (2009)
21. T. Krishnamohan, D. Kim, S. Raghunathan, K. Saraswat, Double-gate strained-ge heterostructure tunneling FET (TFET) with record high drive currents and ≪60 mV/dec subthreshold slope, in *Electron Devices Meeting*, *IEDM 2008*, *IEEE International*, pp. 1–3 (2008)
22. Y. Yang, S. Su, P. Guo, et al, Towards direct band-to-band tunneling in P-channel tunneling field effect transistor (TFET): technology enablement by Germanium-tin (GeSn), in *Electron Devices Meeting (IEDM)*, *2012 IEEE International*, pp. 16.3.1–16.3.4 (2012)
23. Q. Huang, R. Huang, C. Wu, et al, Comprehensive performance re-assessment of TFETs with a novel design by gate and source engineering from device/circuit perspective, in *Electron Devices Meeting (IEDM)*, *2014 IEEE International*, pp. 13.3.1–13.3.4 (2014)
24. Y.-R. Jhan, Y.-C. Wu, Y.-L. Wang et al., Low-temperature microwave annealing for tunnel field-effect transistor. IEEE Electron Device Lett. **36**, 105–107 (2015). doi:10.1109/LED.2014.2386213
25. L. Lattanzio, L. De Michielis, A.M. Ionescu, Complementary Germanium electron-hole bilayer tunnel FET for sub-0.5-V operation. IEEE Electron Device Lett. **33**, 167–169 (2012). doi:10.1109/LED.2011.2175898
26. J.T. Teherani, S. Agarwal, E. Yablonovitch et al., Impact of quantization energy and gate leakage in bilayer tunneling transistors. IEEE Electron Device Lett. **34**, 298–300 (2013). doi:10.1109/LED.2012.2229458
27. S. Agarwal, J.T. Teherani, J.L. Hoyt et al., Engineering the electron-hole bilayer tunneling field-effect transistor. IEEE Trans. Electron Devices **61**, 1599–1606 (2014). doi:10.1109/TED.2014.2312939

28. L. Czornomaz, N. Daix, D. Caimi, et al, An integration path for gate-first UTB III-V-on-insulator MOSFETs with silicon, using direct wafer bonding and donor wafer recycling, in *Electron Devices Meeting (IEDM), 2012 IEEE International*, pp. 23.4.1–23.4.4 (2012)
29. J. Lin, X. Zhao, D.A. Antoniadis, J.A. del Alamo, A novel digital etch technique for deeply scaled III-V MOSFETs. IEEE Electron Device Lett. **35**, 440–442 (2014). doi:10.1109/LED.2014.2305668
30. S. Agarwal, E. Yablonovitch, Using dimensionality to achieve a sharp tunneling FET (TFET) turn-on, in *Device Research Conference (DRC), 2011 69th Annual, IEEE*, pp. 199–200 (2011)
31. J.T. Teherani, W. Chern, D.A. Antoniadis et al., Extraction of large valence-band energy offsets and comparison to theoretical values for strained-Si/strained-Ge type-II heterostructures on relaxed SiGe substrates. Phys. Rev. B **85**, 205308 (2012). doi:10.1103/PhysRevB.85.205308
32. R. Rooyackers, A. Vandooren, A.S. Verhulst, et al, A new complementary hetero-junction vertical tunnel-FET integration scheme, in *Electron Devices Meeting (IEDM), 2013 IEEE International*, pp 4.2.1–4.2.4 (2013)
33. S. Mookerjea, D. Mohata, T. Mayer et al., Temperature-dependent I-V characteristics of a vertical tunnel FET. IEEE Electron Device Lett. **31**, 564–566 (2010). doi:10.1109/LED.2010.2045631
34. T. Yu, J.T. Teherani, D.A. Antoniadis, J.L. Hoyt, Effects of substrate leakage and drain-side thermal barriers in $In_{0.53}Ga_{0.47}As/GaAs_{0.5}Sb_{0.5}$ quantum-well tunneling field-effect transistors. Appl. Phys. Express **7**, 094201 (2014). doi:10.7567/APEX.7.094201
35. W.K. Liu, D. Lubyshev, J.M. Fastenau et al., Monolithic integration of InP-based transistors on Si substrates using MBE. J. Cryst. Growth **311**, 1979–1983 (2009). doi:10.1016/j.jcrysgro.2008.10.061
36. I. Vurgaftman, J.R. Meyer, L.R. Ram-Mohan, Band parameters for III–V compound semiconductors and their alloys. J. Appl. Phys. **89**, 5815–5875 (2001). doi:10.1063/1.1368156
37. J.Z. Li, J. Bai, J.-S. Park et al., Defect reduction of GaAs epitaxy on Si (001) using selective aspect ratio trapping. Appl. Phys. Lett. **91**, 021114 (2007). doi:10.1063/1.2756165
38. J.G. Fiorenza, J.-S. Park, J. Hydrick et al., (Invited) Aspect ratio trapping: a unique technology for integrating ge and III-Vs with silicon CMOS. ECS Trans. **33**, 963–976 (2010). doi:10.1149/1.3487628
39. H.C. Tseng, Y.Z. Ye, High-performance, graded-base AlGaAs/InGaAs collector-up heterojunction bipolar transistors using a novel selective area regrowth process. IEEE Electron Device Lett. **20**, 271–273 (1999). doi:10.1109/55.767095
40. D.K. Mohata, R. Bijesh, Y. Zhu, et al, Demonstration of improved heteroepitaxy, scaled gate stack and reduced interface states enabling heterojunction tunnel FETs with high drive current and high on-off ratio, in *2012 Symposium on VLSI Technology (VLSIT)*. pp. 53–54 (2012)
41. K. Tomioka, J. Motohisa, S. Hara, T. Fukui, Control of InAs nanowire growth directions on Si. Nano Lett. **8**, 3475–3480 (2008). doi:10.1021/nl802398j
42. K. Tomioka, M. Yoshimura, T. Fukui, A III-V nanowire channel on silicon for high-performance vertical transistors. Nature **488**, 189–192 (2012). doi:10.1038/nature11293
43. M. Chhowalla, H.S. Shin, G. Eda et al., The chemistry of two-dimensional layered transition metal dichalcogenide nanosheets. Nat Chem **5**, 263–275 (2013). doi:10.1038/nchem.1589
44. C. Gong, H. Zhang, W. Wang et al., Band alignment of two-dimensional transition metal dichalcogenides: application in tunnel field effect transistors. Appl. Phys. Lett. **103**, 053513 (2013). doi:10.1063/1.4817409
45. H. Fang, S. Chuang, T.C. Chang et al., High-performance single layered WSe2 p-FETs with chemically doped contacts. Nano Lett. **12**, 3788–3792 (2012). doi:10.1021/nl301702r
46. M. Chen, H. Nam, S. Wi et al., Stable few-layer MoS2 rectifying diodes formed by plasma-assisted doping. Appl. Phys. Lett. **103**, 142110 (2013). doi:10.1063/1.4824205

47. Y. Du, H. Liu, A.T. Neal et al., Molecular doping of multilayer field-effect transistors: reduction in sheet and contact resistances. IEEE Electron Device Lett. **34**, 1328–1330 (2013). doi:10.1109/LED.2013.2277311
48. H. Fang, M. Tosun, G. Seol et al., Degenerate n-Doping of few-layer transition metal dichalcogenides by potassium. Nano Lett. **13**, 1991–1995 (2013). doi:10.1021/nl400044m
49. S. Chuang, C. Battaglia, A. Azcatl et al., MoS2 P-type transistors and diodes enabled by high work function MoOx contacts. Nano Lett. **14**, 1337–1342 (2014). doi:10.1021/nl4043505
50. D. Kiriya, M. Tosun, P. Zhao et al., Air-stable surface charge transfer doping of MoS2 by benzyl viologen. J. Am. Chem. Soc. **136**, 7853–7856 (2014). doi:10.1021/ja5033327
51. H.-Y. Park, M.-H. Lim, J. Jeon et al., Wide-range controllable n-Doping of molybdenum disulfide (MoS2) through thermal and optical activation. ACS Nano **9**, 2368–2376 (2015). doi:10.1021/acsnano.5b00153
52. J.H. Yu, H.R. Lee, S.S. Hong et al., Vertical heterostructure of two-dimensional MoS2 and WSe2 with vertically aligned layers. Nano Lett. **15**, 1031–1035 (2015). doi:10.1021/nl503897h
53. T. Roy, M. Tosun, X. Cao et al., Dual-gated MoS2/WSe2 van der waals tunnel diodes and transistors. ACS Nano **9**, 2071–2079 (2015). doi:10.1021/nn507278b
54. K. Boucart, A.M. Ionescu, A new definition of threshold voltage in Tunnel FETs. Solid-state electronics **52**, 1318–1323 (2008). doi:10.1016/j.sse.2008.04.003
55. W. Lee, W. Choi, Influence of inversion layer on tunneling field-effect transistors. IEEE Electron Device Lett. **32**, 1191–1193 (2011). doi:10.1109/LED.2011.2159257
56. S. Agarwal, E. Yablonovitch, Band-edge steepness obtained from esaki/backward diode current-voltage characteristics. IEEE Trans. Electron Devices **61**, 1488–1493 (2014). doi:10.1109/TED.2014.2312731
57. R.M. Iutzi, E.A. Fitzgerald, Microstructure and conductance-slope of InAs/GaSb tunnel diodes. J. Appl. Phys. **115**, 234503 (2014). doi:10.1063/1.4883756

Chapter 3
Compact Models of TFETs

Lining Zhang and Mansun Chan

Abstract Rapid developments in the TFETs' process and rising interests in evaluating their potential in low-power circuits/systems require a TFET compact model for SPICE simulations. In this chapter, we discuss the essential device physics of TFETs, propose necessary simplifications of their complex operations, and develop a core model for homojunction TFETs. At first, we analyze the roles of TFET channel charge in affecting their subthreshold swing and superlinear output. Bearing this in mind, we divide the TFET structure into three distinctive regions for the purposes of considering the channel charge and at the same time getting a closed-form solution of the device electrostatics. After that, we find a simplification to the integration formulation of the interband tunneling physics to derive the current model. With a straightforward derivation, we obtain the terminal charge model and therefore finish the core model development. Around this core, we are adding advanced effect modules and specifically introduce the gate leakage module and short-channel effect module here. Finally, we analyze the basic operations of heterojunction TFETs and possible challenges in their model developments.

3.1 A Review of TFET Modeling

Thanks to the developments of device processes such as those covered in Chap. 2 and in the references therein, TFETs have gained a lot of performance improvements in terms of the subthreshold swings, on state current and linear output resistances. Other performance boosters such as those introduced in Chap. 1 are also being explored by different research groups. Industries (such as Intel, IBM, and Toshiba) became interested in TFETs and reported either their fabricated devices [1–3] or their projections of TFETs' properties [4, 5]. Noticeably, there are quite a few research initiatives across the world dedicated on the TFETs (e.g., the

L. Zhang (✉) · M. Chan
Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
e-mail: lnzhang@ieee.org

L. Zhang and M. Chan (eds.), *Tunneling Field Effect Transistor Technology*,
DOI 10.1007/978-3-319-31653-6_3

STEEPER project [6] and the E^{2}SWITCH project [7] in Europe, the STEEP project funded by DARPA [8]). In order to evaluate the possible energy efficient applications of TFETs, circuit designers began to investigate the building blocks of TFET-based digital circuits and also to explore their analog applications, which will be covered in Chap. 4 later on.

Looking back to the history of the integrated circuit (IC) industry, a complete infrastructure has been identified as an important reason for its success in the past few decades. This infrastructure is described briefly as follows, which shows that a standard compact model bridges the circuit designers, electronic design automation (EDA) vendors, and the foundries. The foundries characterize their device technologies with specific compact device models and include the model information like the model identity and parameters in the process design kits (PDK). The circuit designers obtain the PDK corresponding to the chosen device technology in their designs from foundries, use circuit simulators provided by the EDA vendors which support the same compact models, and provide their designs back to the foundries. The accurate device models guarantee that the final taped-out chips function almost the same as the simulated designs. It is the effective communications via compact device models between different parties that contribute to the success of IC industry. When the TFET technology is going to be brought from research and development to real practices, a compact device model playing similar roles will be eventually needed.

A 'good' compact device model should generally satisfy several requirements:

- A compact device model reproduces the device terminal current–voltage characteristics accurately and quickly;
- A compact device model includes descriptions of equally important capacitance–voltage characteristics accurately, since the model will be used to simulate both the static circuit and the dynamic circuit characteristics;
- A compact device model is implementable into a SPICE engine to perform the circuit simulations and thus needs to fulfill the requirements of the SPICE algorithm.

Basically, there are three kinds of compact model: a table lookup model, a physics-based analytic model, and a behavioral model. The table lookup model stores the measured, discrete data of an electric device in a table and obtains the output by searching and/or interpolating. For example, during iterations in SPICE simulations of nonlinear circuits, certain interpolations are necessary to get the output for the intermediate input that is not included in the table. To have a good table lookup model, we have to do lots of measurements, not only the DC but also the capacitance characteristics, according to the 'good' model standard given above. For devices with different design parameters, we have to prepare separate tables. The physics-based model uses analytical equations to describe the devices' current/capacitance–voltage characteristics from solutions of physical laws which govern the device operation. It can be continuous and usually valid for devices with different parameters. A third kind of model lies between the above two models, which can be named a behavioral model. Without referring to the devices' physics,

it uses functions to fit the discrete data in a table lookup mode and therefore also eliminates the interpolation algorithms in SPICE simulations. However, similar to the table lookup model, different parameter sets are needed to fit different devices. Overall, a physics-based model, if available, is still preferred over the other two kinds of models [9].

At the beginning of TFET-based circuit research, a table lookup model [10] or a behavioral model was utilized by the circuit designers since a physical model was missing. Yibin et al. proposed a SPICE behavioral model in [11] and used it in dc simulations of a TFET-based inverter. In total, 20 parameters are used. Due to the unphysical nature of the pure fitting parameters, it is non-trivial to extract them. A table lookup model is built to evaluate the impacts of TFET technology on low-power SRAM designs in [10]. Similarly, circuit designers made use of different table lookup models to investigate more applications of TFETs in low-power digital circuits, such as the SRAMs, and even the CPU cores [12].

Physics-based current models of TFETs were initially investigated by a group of researchers at IMEC [13–15]. With the assumptions of no source depletion, no channel charge, and infinite channel length/thickness, authors proposed to separate the point and line tunneling and derived the tunnel path and current model [13]. Later, they included the drain voltage effects based on their point and line tunneling frame [15]. Later in 2010, Bardon et al. proposed a pseudo-two-dimensional current model for double-gate (DG) TFETs [16]. Authors noticed the source/drain depletions in TFETs due to the large lateral electric field across the tunneling junction and included the effects into their model. As stated by the authors, they focused on the transitions from the off-state to the onset operation regions and also neglected the influences of the channel mobile charge, similarly to the assumptions in IMEC's models. Following it, TFET current models were reported in [17] for SOI TFET, but source depletion effect was removed. Meanwhile, the tunneling current equation is given by an implicit integral, which is not suitable for SPICE implementations. In 2011, Wan et al. reported another analytical current model for SOI TFETs [18]. They interpreted TFETs as a serial combination of a tunneling diode and a conventional MOSFET. Actually, this is in accordance with the original idea of TFET, which uses a third terminal to tune the carrier transport in tunneling diodes. With this interpretation, authors built an analytical current model with a threshold voltage parameter and the drain voltage effects in TFETs were included through the MOSFET part.

Compact models of charge-based devices should include both the terminal current and the terminal charge descriptions. For TFETs, until 2011, Yang et al. proposed using modified BSIM3 capacitance model equations to account for the terminal capacitance properties [19]. It is good for understandings of TFET properties, but not adequate for SPICE simulation: (1) There is not a corresponding tunneling current model since BSIM3 cannot be used to describe the current transport properties of TFETs. (2) There is no proof that the charge conservation is held for the capacitance model proposed. A terminal charge model is preferred to derive the device dynamic response instead of a capacitance model. In later 2011, P.M. Solomon et al. proposed a relatively complete model, including descriptions

of both the current and capacitance characteristics [20]. Authors used the proposed model to benchmark the TFETs with FinFETs. In their model, the tunneling current is calculated by numerical integrations and no implicit expressions of terminal charges are given.

In a physics-based device model development, one general scheme is usually followed: building a core model first and then integrating more advanced effect modules into it. Semiconductor devices are simple in the sense that there is always one master equation to govern their principle operations. At the same time, semiconductor devices are complex from the perspective that all the physical effects such as different carrier scatterings and series resistance are all alive inside them. Correspondingly, the core model is targeted to capture the essential operations, while the advanced modules are added to make the model applicable to fit the measured devices. In the following sections, this scheme is also used in describing the TFET models.

3.2 A Core Model of Homojunction TFETs

3.2.1 TFET Operations

The very basic TFET properties have been introduced in Chap. 1. More details about the TFET operations are elaborated here before we proceed to the core model. Figure 3.1 compares transfer characteristics of one MOSFET and one TFET both in DG and in n-type configurations, with the same channel thickness t_{ch}, channel length, gate oxide thickness, gate work function, and drain voltages (V_{ds} = 0.05 V and 0.5 V). Silicon with higher density of state (DOS) is used as the channel material for comparisons in Fig. 3.1, and in later sections, InAs is also used as an example of materials with lower DOS. The device data are obtained from commercial TCAD device simulator, which solves the Poisson's equations with Fermi statistics and the transport equations. It is assumed that the ambipolar conductions in TFETs have been eliminated [21].

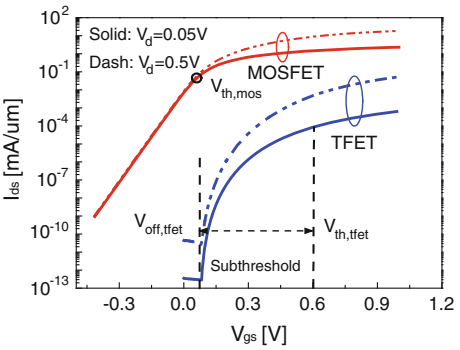

Fig. 3.1 Transfer characteristics of a DG p–i–n TFET and a DG n–i–n MOSFET with the only difference of source doping polarity

A standard second derivative (SD) method [22] is used to extract the MOSFET (Fig. 3.1, $V_d = 0.05$ V) threshold voltage $V_{th,mos}$. Between it and the flatband voltage lies the MOSFET subthreshold region. For MOSFETs, the threshold voltage is basically the gate voltage which separates the weak and strong channel inversions. However, only with gate voltages around or larger than $V_{th,mos}$, the interband tunneling current in the TFET starts to dominate over the leakage component. The TFET threshold voltage supposed to identify the transition between distinctive operations in the tunneling-dominant regions is clearly not defined as same as MOSFETs. Instead, an offset gate voltage, $V_{off,tfet}$, is defined as the gate voltage where the leakage to tunneling transition happens. If the same SD method is applied to the TFET (Fig. 3.1, $V_d = 0.05$ V), its threshold voltage, $V_{th,tfet,}$ is obtained to be around 0.6 V. Between the offset voltage and the threshold voltage is the TFET subthreshold region. In the TFET, there is a strong drain voltage modulation on the current. Some term like the drain-induced barrier thinning is used to describe the drain modulations on the tunneling current [23]. Another feature in the TFET subthreshold region is that the subthreshold swing changes significantly with the gate voltage and the current, unlike that in the MOSFET which almost manifests a constant. These subthreshold characteristics have been confirmed previously in many experimental homojunction TFETs [24–26].

A simplified formulation of Eq. (1.34) is used to qualitatively explain the TFET properties, with the tunnel distance W_{tun} and tunnel decay length λ_{tun}:

$$I_{ds} \propto \exp\left(-\frac{W_{tun}}{\lambda_{tun}}\right) \tag{3.1}$$

The subthreshold swing is then derived:

$$\frac{1}{SS} = \frac{\partial \log(I_{ds})}{\partial V_{gs}} \propto -\frac{1}{\lambda_{tun}} \times \frac{\partial W_{tun}}{\partial V_{gs}} \tag{3.2}$$

While the first term in the right-hand side of Eq. (3.2) depends on the material properties, the second term depends on the TFET electrostatics. A minimum tunnel distance $W_{t,min}$ is used [27, 28] to represent the integral of multiple tunnel paths when the gate voltage goes beyond the offset voltage, and Fig. 3.2a plots its dependences on the gate voltage. Generally, the tunnel distance decreases quickly when the gate voltage just exceeds the offset voltage, and then, its dependence on the gate voltage is gradually weakened. This inverse function-like dependence is responsible for the varying SS in TFETs. The electrostatic potentials in the TFET channel far from the tunnel junction are shown in Fig. 3.2b. The surface potential under the offset voltage is used as the zero potential reference. Further increasing in the gate voltage leads to larger surface and center potentials in the DG MOS structures. With the conduction band in the channel approaching the quasi-Fermi level, carriers in the channel start to screen the gate field and the electrostatic potentials gradually saturate, especially the center potentials. This is in accordance with Fig. 3.1 that the MOS part of the TFET channel enters into the strong inversion

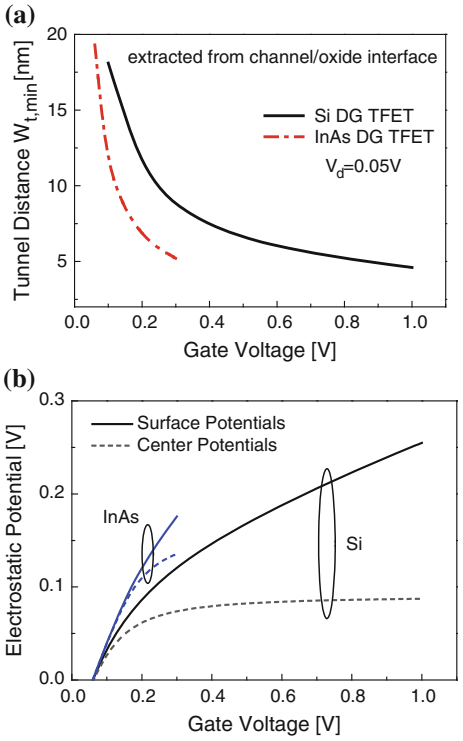

Fig. 3.2 **a** The minimum tunnel distances in TFETs show 1/sqrt(x)-like dependence on the gate voltage. **b** The electrostatics in TFETs is responsible for the tunnel behaviors. TFETs with lower DOS materials such as InAs have less screening effects, hence steeper slopes

region. A gradually weaker dependence of the potentials on gate voltages determines the tunnel distance behaviors. Another expression for the subthreshold swing is as follows:

$$\frac{1}{SS} = -\frac{1}{\lambda_{tun}} \times \frac{\partial W_{tun}}{\partial \varphi_{mos}} \times \frac{\partial \varphi_{mos}}{\partial V_{gs}} \tag{3.3}$$

Although the third term follows the same gate control as that in MOSFETs, the electrostatics in the second term makes the difference and contributes to the steep slope in TFETs.

If the channel charge screening effects are weakened, potentials in the channel follow more tightly on the gate voltage. As a result, the third term in Eq. (3.3) is enhanced and an even steeper slope is expected. Therefore, semiconductor materials of lower DOS having less charge screening are potentially steep slope boosters. Figure 3.2 compares the TFETs with silicon (higher DOS) and InAs (lower DOS) as channel materials. The InAs TFET has less screening effect, and its channel potential increases beyond that in the Si-based TFET as shown in Fig. 3.2b. The dependence of the tunnel distances on gate voltages is more sensitive leading to steeper slopes. The screening effect finally comes into the picture when the Fermi level in the channel goes far into the conduction band with further increase of the

gate voltage. However, a larger voltage window with steep slope is expected before that.

As the electrostatics in TFETs is codetermined by their gate and drain voltages, there is also a drain modulation effect on the current. In many experimental TFETs, the drain modulation is strong that the tunneling current shows an exponential dependence on the drain voltage. This leads to the superlinear output introduced in Chap. 1. As in the above analysis, the electrostatics in the TFET channel far from the tunnel junction is almost one-dimensional. At the same time, the drain–source voltage drop can be assumed to happen across the tunnel junction to the first-order approximation [28]. Under this scenario, the quasi-Fermi level E_{imref} (for majority carriers of the drain side) is almost a constant in the MOS channel part of TFETs. Therefore, the electrostatics in main part of the TFET channel is the same as that at the drain side of a long-channel MOSFET.

Mathematic descriptions of the drain modulation effect are given below. With the source-side degeneracy ($\Delta E = E_v - E_f$) and a small drain voltage V_d, the channel energy bands are as follows:

$$E_c - E_{imref} = \Delta + qV_d - q\varphi_{mos} \tag{3.4}$$

and the surface carrier concentration is given with the Fermi integral $F_{1/2}$:

$$n = N_c \frac{2}{\sqrt{\pi}} F_{1/2}\left(\frac{q\varphi_{mos} - \Delta E - qV_d}{kT}\right) \tag{3.5}$$

Electrostatics can be obtained from the Poisson's equation:

$$\frac{d^2\varphi_{mos}}{dx^2} = \frac{2q}{\sqrt{\pi}\varepsilon_{ch}} N_c F_{1/2}\left(\frac{q\varphi_{mos} - \Delta E - qV_d}{kT}\right) \tag{3.6}$$

If the effective DOS of the conduction band N_c is large and source degeneracy is small, there is a significant amount of carrier in the channel. As a result, the electrostatics, for example the surface potential, will be changed by the drain voltage through the right-hand side of Eq. (3.6). On the other hand, if N_c is small and the source degeneracy is large, so the right-hand side of Eq. (3.6) is ignorable, and there will be no obvious control of the drain voltage on the electrostatics. Figure 3.3a and b plots the conduction band and quasi-Fermi-level profiles in Si-based and InAs-based TFET with the gate voltages a bit larger than the offset voltages. The conduction band DOS of InAs ($N_c = 8.7 \times 10^{16}$ cm^{-3}) is much smaller than that of Si ($N_c = 3.2 \times 10^{19}$ cm^{-3}), so the drain voltage modulation on the surface potential in the InAs TFET is almost ignorable with a similar source degeneracy.

Figure 3.3c plots the output characteristics of InAs and Si TFETs with normalized current and normalized drain voltage. With the gate biased in the subthreshold region, an exponential dependence of tunneling current on the drain voltage in the Si TFET is obvious, while in the InAs TFET a nearly linear output is

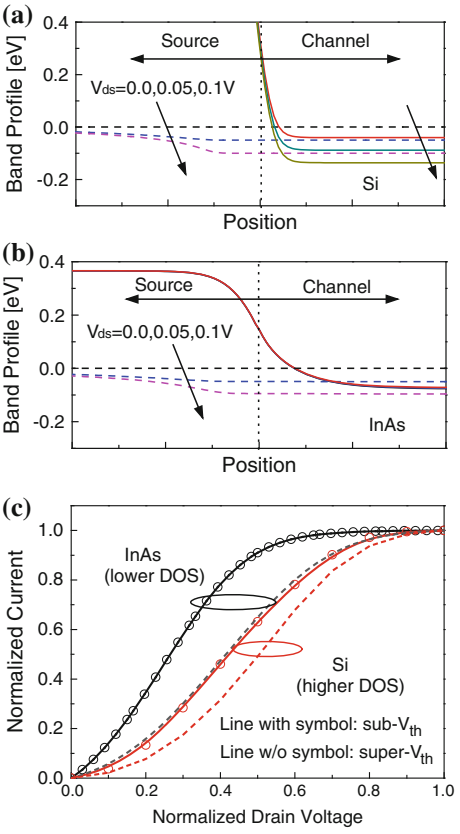

Fig. 3.3 **a** In TFETs with higher DOS materials like Si, the drain voltage strongly modulates the electrostatic potentials due to the channel charge. **b** The drain modulations on the surface potentials are less significant in TFETs with lower DOS material-like InAs. **c** In the subthreshold regions of TFETs with lower DOS, output characteristics are more linear compared to the above-threshold region

observed. The differences are attributed to the above drain modulation effects on the surface potential, hence the tunnel junction. Equation (3.6) shows that an increase of the potential leads to an increase of the charge concentrations, so at certain gate voltages the drain modulation on the electrostatic potential hence the tunneling currents is also partially recovered in TFETs with lower DOS material. Figure 3.3c shows that with the gate biased in the above-threshold operations, the output characteristics of the InAs TFET are also becoming superlinear. The drain modulations on the channel electrostatics can also explain the reduced Miller capacitance [29] in TFETs with lower DOS material. In the subthreshold region, the TFET terminal charge or capacitances are obtained straightforwardly from the above electrostatic analysis. The channel charge in TFETs with lower DOS material is less even in the orders of magnitude and further reduced by the drain voltage. As a result, the capacitance C_{gd} (Miller capacitance) as the gradient of channel charge–drain voltage characteristics is smaller.

The essential point from the above analysis is that the channel charge plays an important role in determining the varying subthreshold swing and the superlinear outputs. As a result, the carrier charge has to be included in the TFET core model.

3.2.2 The Electrostatic Potential Model

The CMOS industry is moving to the DG or FinFET technologies, based on which TFETs can be developed. As an example, a model for DG TFET is introduced here. Its schematic with an intrinsic body is shown in Fig. 3.4. As shown in the figure, t_{ch} is the body thickness, t_{ox} is the gate oxide thickness, L_g is the gate length, W_g is the gate width, and N_s and N_d are the source and drain doping concentrations, respectively. An n-type TFET is considered; thus, the source region is p+ doped.

Band-to-band tunneling (BTBT) happens across the source/body junction, and generated electrons are transported to drain through drift diffusion (DD). A TFET is actually equivalent to a gated tunneling diode and a DG MOSFET in series coupled by the sharing node. Region I is the depleted source region which is confirmed by TCAD simulations and also regarded as an important behavior of TFETs. Region II is part of the intrinsic TFET channel. The diode spans over regions I and II, while region III is the DG MOSFET channel. The contour shows the electron generation rate due to BTBT in the tunnel diode. Physical parameters and a coordinate system for modeling are also defined in Fig. 3.4. The potential profile and tunneling current in the diode depend on the potential and electric field boundary conditions at the sharing node, which also determine the DD current in the MOSFET. While the current continuity requires that the tunneling current is equal to the DD current, a perturbation approach is used to solve this coupled problem. By assuming the DG MOSFET in equilibrium (no voltage drops), the zero-order potential solutions in the diode are obtained. Then, with a BTBT model, one can calculate the tunneling current through the gated tunneling diode which brings the DG MOSFET into a non-equilibrium state. For practical TFETs with on-state currents up to the order of a few μA/μm (much smaller than the DD current of a DG MOSFET with the same geometry parameters and at the same biasing voltages), TCAD simulations show that the higher-order perturbation corrections are not significant. The zero-order solutions of electrostatic potential and tunneling current are reasonably accurate and serve well as the starting point of a core model. It also represents the

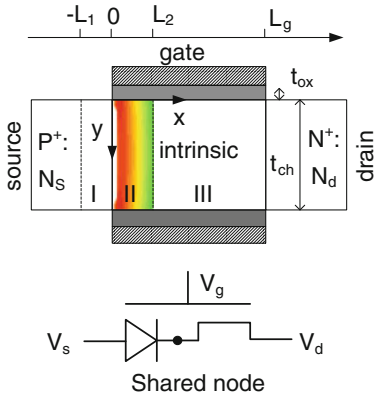

Fig. 3.4 A cross-sectional schematic of a DG TFET and its partition into a serial combination of a gated tunnel diode and a DG MOSFET. The contour plots the electron generation rate

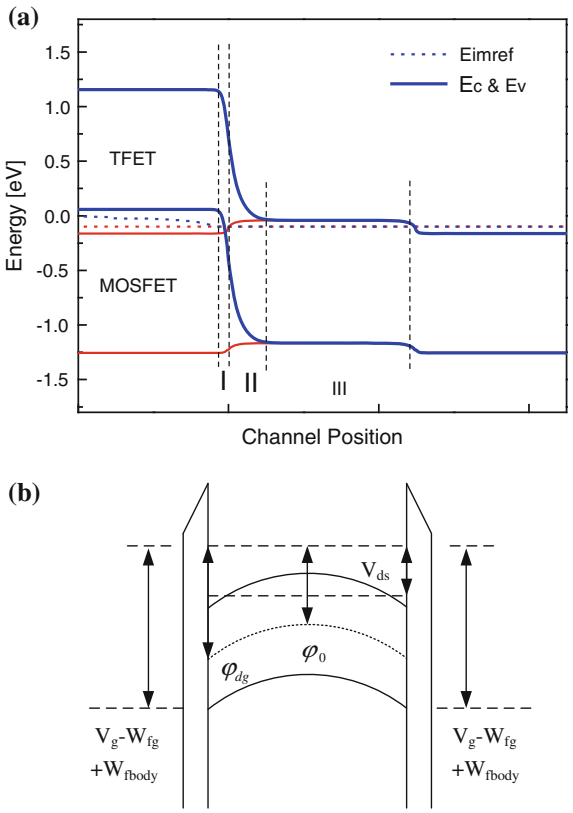

Fig. 3.5 Band diagrams **a** in a DG TFET along x and in a DG MOSFET along channel direction **b** in a DG TFET along y in region III

gate length independence of the current in TFETs shown in both TCAD simulations and the experiments.

Figure 3.5a compares the band profiles at y = 0 along the channel of a TFET and a MOSFET with the same DG structure. In the on state of TFETs, inversion electrons induced in region III make its resistance much smaller than the resistance of the tunneling junction. This exactly agrees with our analysis in Sect. 3.2.1. As a result, a large part of the drain voltage is dropped at the reverse-biased tunneling junction. This further makes the potential gradient in the inverted channel (region III) negligible due to the low tunneling current, especially at small V_d. As shown in Fig. 3.5a, the band profiles in region III of the DG TFET are nearly identical to that in the DG MOSFET. With increasing V_d before saturation, the channel electric field is increased but still small. Therefore, the inverted TFET channel (region III) is assumed to be equivalent to the channel of a DG MOSFET with the same source/drain voltage. It means that in region III, the electrostatic potentials can be obtained by solving the one-dimensional Poisson's equation along y with the charge in considerations. The right boundary condition for region II is mixed: (a) zero electric field and (b) the potentials determined by region III [30]. A band diagram along y in region III is shown in Fig. 3.5b. W_{fg} is the gate work function,

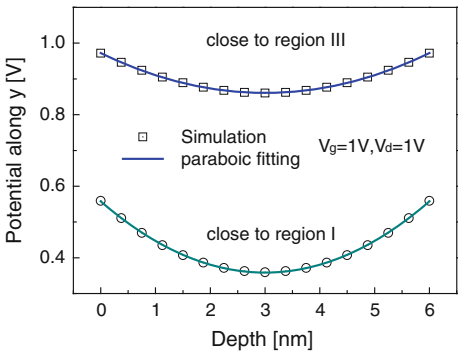

Fig. 3.6 The potential profile along y in region II of the DG TFETs

W_{fbody} is the work function of intrinsic silicon, φ_{dg} is the surface potential, and φ_0 is the center potential. The source-side work function W_{fs}, will be referred later. Potentials defined in Fig. 3.5 and the following derivations are referred to the zero potential in the undepleted source side. Unlike a regular DG MOSFET, source depletion in TFETs is formed due to the body–source electric field along x and the gate–source fringing field.

In region II, the electrostatic potential profiles are two-dimensional, with electric field from the gate electrode and from the drain electrode. While the potential profile along the channel direction is to be determined, the vertical potential follows a parabolic profile, as shown in Fig. 3.6. So a parabolic potential is assumed in region II, which is similar to that observed in short-channel MOSFETs [31]. In the depleted source region I, the electrostatic potential also follows a two-dimensional profile due to its similarity to a MOSFET with heavily doped channel.

It is already known that there are inversion electrons in region III. A gradual change in electron concentrations is expected in region II from the right boundary to the left boundary corresponding to the potential profiles in Fig. 3.5a. This is confirmed in Fig. 3.7. The charge non-uniformity makes the Poisson's equation in region II unsolvable and has to be simplified. Without losing the generality, it is assumed that there is only the intrinsic charge in region II, similar to the full depletion approximation in the p–n junction.

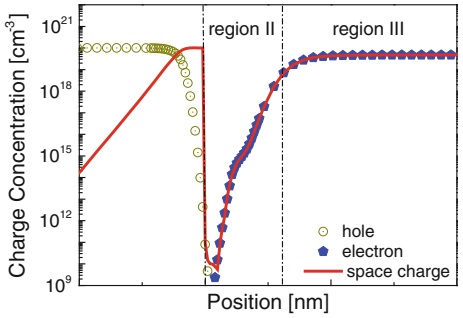

Fig. 3.7 The charge distribution along x from the source to the channel of TFETs

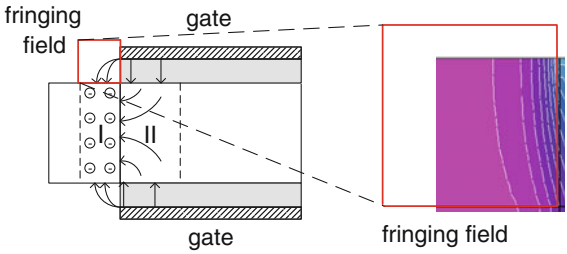

Fig. 3.8 The gate fringing field in the source depletion

The electrostatic problem in region I is also complicated. First, there is a gate fringing field as shown in Fig. 3.8 that will affect the electrostatic potentials. In principle, this fringing field can be handled with the conformal mapping technique which is used in modeling the very similar gate underlap structure in DG MOSFETs and gate-all-around MOSFETs [32]. The conformal mapping suggests that the fringing field in Fig. 3.8 could be approximated as the normal gate electric field in a MOS structure with an effective gate oxide thickness $t_{ox,eff} = \frac{m\pi}{2} t_{ox}$ [33]. This effective gate oxide method will be used in the following analysis. Secondly, the lateral depletion width across the source film is not uniform as verified by the TCAD simulations. The depletion region width at the interface is larger than that at the film center, but is more uniform for TFETs with smaller channel thickness. Again, this makes the Poisson's equation in region I unsolvable even the fully depletion is assumed. A uniform width is assumed in the following analysis.

After identifying the TFET operations and its essential device physics, deriving the potential and charge model is straightforward. To include the channel inversion charge effect in region III, the Poisson's equation with Fermi statistic in DG MOS structure is given by [34]:

$$\frac{d^2\varphi}{dy^2} = \left(\frac{1}{2}\frac{2q^2 n_i}{\varepsilon_{si} kT}\right) \frac{F_{1/2}\left[\left(q\varphi - qV_{ds} - E_g/2\right)/kT\right]}{F_{1/2}\left(-E_g/2kT\right)} \tag{3.7}$$

where φ is the potential, V_{ds} is the drain voltage, and $F_{1/2}$ is the Fermi–Dirac integral of order 1/2 [35]. The gate voltage control over the surface potential is given from Gaussian law:

$$V_{gs} - V_{fb} - \varphi_{dg} = \frac{\varepsilon_{si}}{C_{ox}} \frac{d\varphi_{dg}}{dy}\bigg|_{y=0} \tag{3.8}$$

From the above Eqs. (3.7) and (3.8), the surface potential φ_{dg} in region III can be calculated for the given gate and drain voltages. Details about the analytical solutions of the surface potentials are available in [28] and will not be repeated here.

As stated earlier, similar to the full depletion approximation used in the diode modeling, it is assumed that the only charge contribution in region II is the intrinsic carrier and any inversion charge especially near $x = L_2$ is ignored. Solving the

two-dimensional (2D) Poisson's equation in region II with the parabolic approximation leads to the surface potential profile [30]:

$$\varphi_{s2}(x) = B\,\exp(-\frac{x-L_2}{\lambda_{II}}) + C\,\exp(\frac{x-L_2}{\lambda_{II}}) + \left(V_{gs} - V_{fbs}\right) \tag{3.9}$$

in which λ_{II} is the natural length of the DG MOS structure [36]. B and C are two unknown coefficients, and L_2 is also undetermined yet. Corresponding to the zero-order approximation discussed previously, the mixed boundary conditions at the internal node between regions II and III are as follows:

$$\varphi_{s2}(L_2) = \left(V_{bi,s} + \varphi_{dg}\right), \quad \frac{d\varphi_{s2}(L_2)}{dx} \approx 0 \tag{3.10}$$

and $V_{bi,s}$ is the built-in potential of the source/channel junction. B and C are derived with Eq. (3.10), and a new form of the surface potential profile in region II is obtained with V_{fbs} = W_{fg}-W_{fs}, the flatband voltage of gate fringing DG MOS structure:

$$\varphi_{s2}(x) = \left(V_{gs} - V_{fbs}\right) - \left[V_{gs} - V_{fbs} - V_{bi,s} - \varphi_{dg}\right]\cosh\left(\frac{x-L_2}{\lambda_{II}}\right) \tag{3.11}$$

Considering the fringing field in Fig. 3.8 on the region I with the method of conformal mapping, together with the zero potential and field boundary conditions at $x = -L_1$, the potential profile is written as follows:

$$\varphi_{s1}(x) = \frac{qN_{seff}}{2\varepsilon_{si}}(x+L_1)^2 \tag{3.12}$$

where $N_{seff} = 2\frac{\varepsilon_{si}}{q}\left(\frac{qN_s}{2\varepsilon_{si}} - \frac{\varepsilon_{ox}}{\varepsilon_{si}}\frac{V_{gs}-V_{fbs}}{t_{ox}t_{ch}\pi/2}\right)$. In Eq. (3.12), L_1 is undetermined yet.

By matching the potential and electric fields at the boundary between the regions I and II based on the above Eqs. (3.11) and (3.12), final decisions about the length of these two regions and the overall potential profiles across these two regions are obtained. Without going into the detailed derivations [28], the two lengths L_1 and L_2 are listed here:

$$\begin{aligned}
L_1 &= \sqrt{\frac{2\varepsilon_{si}\varphi_s(0)}{qN_{seff}}} \\
L_2 &= \lambda_{II}\cosh^{-1}\left[-\frac{\varphi_s(0) - \left(V_{gs} - V_{fbs}\right)}{\left[\left(V_{gs} - V_{fbs}\right) - \left(V_{bi,s} + \varphi_{dg}\right)\right]}\right] \\
\varphi_s(0) &= -\sqrt{\left[V_{gs} - V_{fbs} - \left(V_{bi,s} + \varphi_{dg}\right)\right]^2 + 2\left(V_{gs} - V_{fbs}\right)\Phi + \Phi^2} \\
&\quad + \left(V_{gs} - V_{fbs} + \Phi\right), \Phi = \frac{qN_{seff}\lambda_{II}^2}{\varepsilon_{si}}
\end{aligned} \tag{3.13}$$

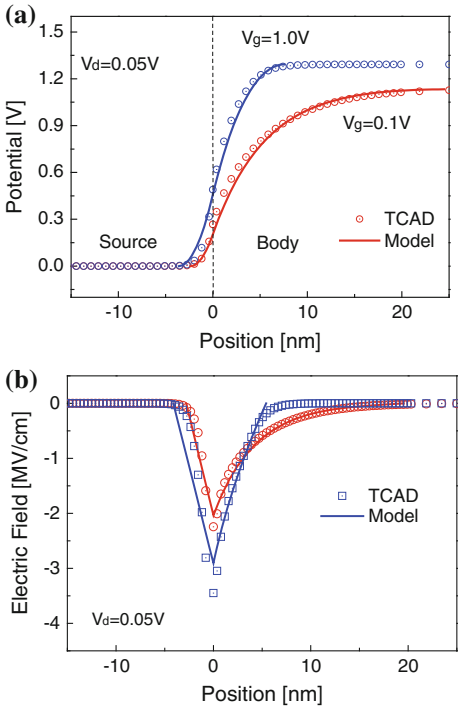

Fig. 3.9 The potential and electric field profile in the DG TFET

Figure 3.9 plots the surface potential profile across the tunnel junction and also the electric field. Overall good agreements between the modeling results and those from TCAD simulations are obtained. The lengths of region II become smaller as we increase the gate voltage. At the same time, larger gate voltage induces larger source depletion region, which shows that the effect of gate fringing field is captured. With the above potential model, we capture the essential device physics that we introduced in Sect. 3.2.1.

3.2.3 The Current Model

With an accurate electrostatic potential model in the above section, the interband tunneling current model is derived in this section. Physically, the interband tunneling current is obtained by summarizing all the paths in the energy window as the tunneling probability changes along different paths. If this method is used [16, 17], a numerical integral cannot be avoided. Tunneling through different paths in the energy space is mapped to a distribution of the electron generation rate in the real space, as illustrated in Fig. 3.10. To obtain the total current, an integral of the electron generation rate over the entire volume region II is necessary:

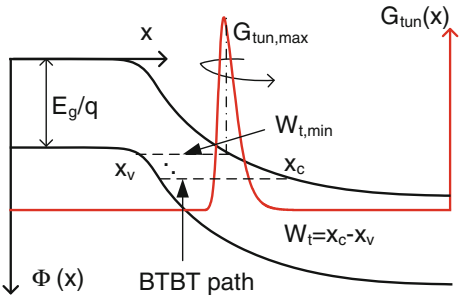

Fig. 3.10 The electron generation rate due to BTBT reaches its maximum value along the tunneling path with the smallest distance $W_{t,min}$. It decays exponentially with increasing tunnel distances

$$I_{ds} = q \cdot \int G_{tun} d\Omega_{tun} \tag{3.14}$$

For SPICE simulation, an analytical expression of the electron generation rate is necessary to derive a closed-form solution of the drain current. As shown in Fig. 3.10, there are different BTBT paths across the source/body junction with their classical turning points x_c and x_v and the tunnel distance $W_t = x_c - x_v$. The electron generation rate due to BTBT has a peak value $G_{tun,max}$ along the tunnel path which has the minimum tunnel distance $W_{t,\min}$ [27] because of the largest tunneling probability. Moving rightward or leftward, the generation rate drops almost exponentially since the tunneling distance is increased. Fortunately, the minimum tunneling distance can be found from the potential profile model in the above section:

$$W_{t,\min} = L_2 - \lambda_{II} \cosh^{-1}\left(\frac{V_{gs} - V_{fbs} - \varphi_I}{V_{gs} - V_{fbb} - \varphi_{dg}}\right) - \sqrt{\frac{2\varepsilon_{si}}{qN_{seff}}\left(\varphi_I - \frac{E_g}{q}\right)} + L_1$$

$$\varphi_I = \left(V_{gs} - V_{fbs}\right) + \frac{qN_{seff}\lambda_{II}^2}{\varepsilon_{si}} - \sqrt{\left(\frac{qN_{seff}\lambda_{II}^2}{\varepsilon_{si}}\right)^2 + \left(V_{gs} - V_{fbb} - \varphi_{dg}\right)^2 + 2\frac{qN_{seff}\lambda_{II}^2}{\varepsilon_{si}}\left(V_{gs} - V_{fbs} - \frac{E_g}{q}\right)} \tag{3.15}$$

Following the Kane's model, the peak electron generation rate is calculated by:

$$G_{tun,\max} = A \cdot \frac{E_g^{3/2}}{q^2} \cdot \frac{1}{W_{t,\min}^2} \exp\left[-\frac{W_{t,\min}}{\lambda_{tun}}\right] \tag{3.16}$$

With the assumption of an exponential profile of the carrier generation rate, the integral in Eq. (3.14) is transformed to an equivalent form:

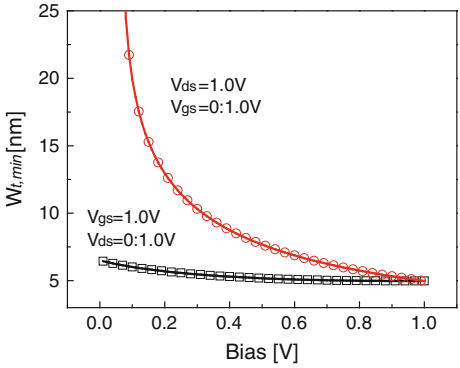

Fig. 3.11 The dependences of minimum tunneling distances $W_{t,min}$ on the external bias

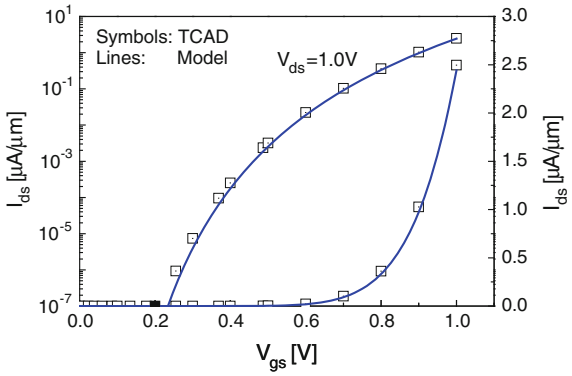

Fig. 3.12 The transfer characteristics of TFETs are reproduced by the current model

$$\int_{-\infty}^{\infty} G_{tun}(x)dx = 2\lambda_{tun}G_{tun,\max} \tag{3.17}$$

which further leads to the final current model formulation:

$$I_{ds} = q \cdot G_{tun,\max} \cdot (2\lambda_{tun} \cdot W \cdot t_{ch}) \tag{3.18}$$

Figure 3.11 plots the minimum tunneling distance $W_{t,min}$ as a function of the gate and drain voltage in one sample DG TFET. $W_{t,min}$ decreases quickly with the gate voltage which reproduces the effects as in Fig. 3.2 and leads to the steep slope properties. Depending on the decay length, the tunneling is not significant where $W_{t,min}$ is much larger than about 10 nm. At the same time, $W_{t,min}$ depends on the drain bias and asymptotically comes to its saturation value with larger V_{ds}. Figures 3.12 and 3.13 show that the current model discussed above reproduces the general TFET characteristics.

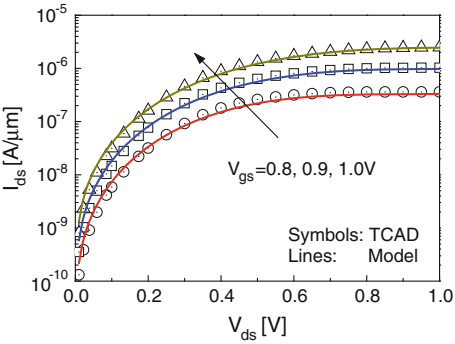

Fig. 3.13 The output characteristics of TFETs are reproduced by the current model including descriptions over the superlinear region

3.2.4 The Charge Model

With the electrostatic potential model in Sect. 3.2.2, the terminal charge and capacitances are derived in this section. Figure 3.14 shows schematics of the electric field in a DG TFET and the corresponding terminal capacitances before and after drain saturation. The region definition is exactly the same as in Sect. 3.2.2. Before saturation, major parts of the gate electric field are terminated by the inversion electrons in region III which dominate the total channel charge. Another part of the gate field is terminated by the source-side charge. The lateral electric field E_x in region III is much smaller, consistent with the zero-field approximation in developing the current model. In saturation regions, inversion charges in region III close to the drain side are depleted. The electric field from the drain goes

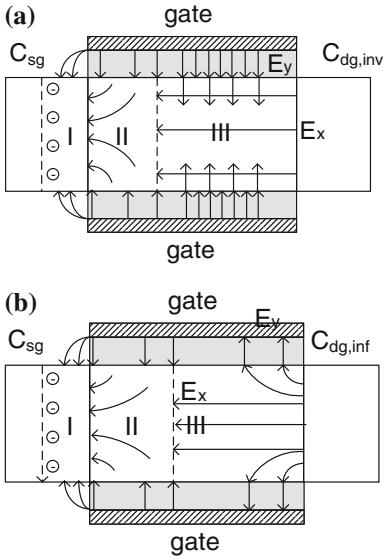

Fig. 3.14 Cartoons of the electric field in a DG TFET operating in **a** non-saturation and **b** saturation regions together with the corresponding terminal capacitances

partially through the channel region and is terminated by the gate. A peak electric field (E_m) is observed at the drain side from TCAD simulations similar to that in MOSFETs [37]. The charge corresponding to the lateral field gradient is defined as the inner fringing charge, which contributes an inner fringing capacitance [38] or the lateral field-induced capacitance in the saturation regions. In total, the inversion charge and inner fringing charge are two composite components that contribute to the total channel charge. In the drain saturation regions, the lateral potential gradient is larger than that before the saturation region. In principle, the channel charge should be obtained with an integral. For simplicity, a zero field is still assumed to find the inversion charge without losing generality since the inner fringing charge is dominant, especially for TFETs with small channel length.

Since the inversion charge in region II close to x = L_2 is ignored as discussed in the previous sections, the channel inversion charge is obtained by integrating the charge density in region III along x from L_2 to L_g. Due to the small potential gradient in region III, inversion charge profile is approximately uniform before saturation. As a result, the channel inversion charge is given:

$$Q_{channel,inv} = -2W_g(L_g - L_2)C_{ox}(V_{gs} - V_{fbs} - V_{bi} - \varphi_{dg}) \tag{3.19}$$

In saturation regions, the drain-side inner fringing charge $Q_{channel,inf}$ is a major component in the total channel charge. The maximum electric field at the drain side is approximated by [37]:

$$E_m = \left(V_d - V_{bi,d} - \varphi_{dg}\right)/\lambda_{II} \tag{3.20}$$

where $V_{bi,d}$ is the built-in potential of the drain/body junction. The inner fringing charge is calculated according to Gaussian law:

$$Q_{channel,inf} = W_g t_{si} \varepsilon_{si} E_m \tag{3.21}$$

The summation of the channel inversion charge from Eq. (3.19) and the fringing charge from Eq. (3.21) is a reasonable approximation for the total channel charge. It is given by:

$$Q_{channel} = Q_{channel,inv} + Q_{channel,\mathrm{inf}} \tag{3.22}$$

At the source side, depletion charge is the only composition of the total charge in region I both before and after saturation. A full depletion is assumed in region I, and the depletion width is still assumed to be uniform across the junction. This leads to the source depletion charge expression:

$$Q_{s,dep} = -qN_{seff} L_1 t_{ch} \tag{3.23}$$

Figures 3.15 and 3.16 show the modeled source depletion charge and channel charge of a DG TFET together with terminal charges obtained from TCAD. In the

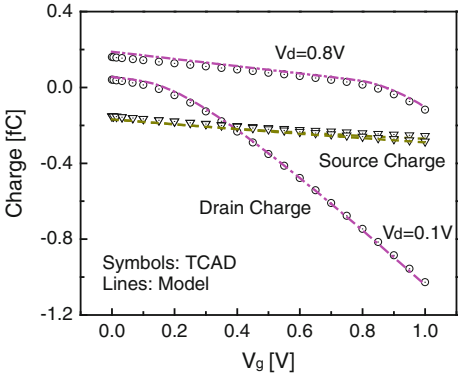

Fig. 3.15 Dependence of the source and drain terminal charges on the gate voltage from the charge model and TCAD simulations

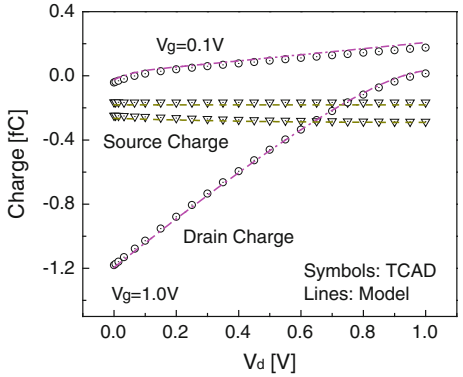

Fig. 3.16 Dependence of the source and drain terminal charges on the drain voltage from the charge model and TCAD simulations

simulations, the non-local BTBT model is switched on and the sinusoidal steady-state analysis [39] is used to get the impedance matrix of TFET, including the terminal capacitances. As shown in the figure, the modeled channel charge and source depletion charge are identical to the simulated drain terminal charge Q_d and source terminal charge Q_s, respectively.

$$Q_d = Q_{channel} \tag{3.24}$$

$$Q_s = Q_{s,dep} \tag{3.25}$$

Therefore, the proposed charge model indicates that the source terminal charge is entirely composed of depletion charge and all of the channel inversion charge is attributed to drain terminal. A 100/0 (the ratio of Q_d to Q_s) partition of channel inversion charge is confirmed for TFETs. This is actually expected from the device operations. When the device is in its off state, the inversion charge concentrations in the TFET channel can be ignored. After increasing the gate voltage, the conduction band edge in the channel is brought down by the injected electrons, and these electrons are provided by the electron reservoir drain side. When the TFET is turned on, the tunneling current also provides to the channel with tunneled electrons.

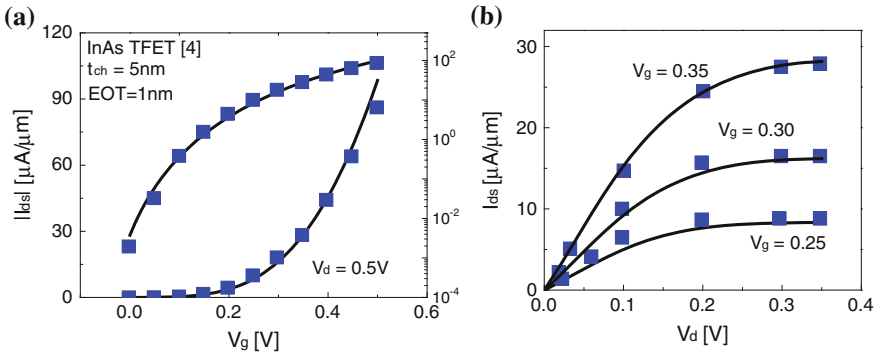

Fig. 3.17 The core model reproduces the InAs TFET properties

However, the same amount of electrons will be collected by the drain side due to the current continuity law. So under the quasi-static conditions, the electron charges in the channel are only provided by the drain side.

3.2.5 Model Applications

By combining the above current and charge model, we complete a compact model core for TFETs. It can be used to reproduce the 'ideal' TFET properties (without advanced effects inside, for example, the TFETs from atomistic simulations). Figure 3.17 shows that after parameter extractions, the core model captures the InAs-based homojunction TFETs. We make the core model available through the *i*-MOS platform [40].

3.3 Modeling Advanced Effects of Homojunction TFETs

In the above section, a core model of homojunction TFET is introduced in great details. To apply it in practices, advanced effect modules shall be integrated. Figure 3.18 lists several advanced effects that should be considered. In this section,

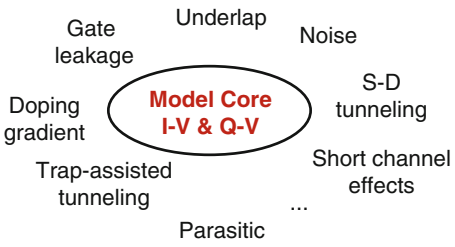

Fig. 3.18 Advanced effect modules with the model core for the TFET technology

we will take the gate leakage and short-channel effects in TFETs as examples and introduce their modeling.

3.3.1 Modeling the Gate Leakage

A first consideration is the gate leakage effect. We note that a large oxide electric field is needed to induce the channel carrier and turn on the TFETs. Under this strong oxide field, the gate leakage is more significant than its counterpart in MOSFETs if the same effective oxide thickness (EOT) is used. At the same time, thinner gate oxide (hence stronger junction field) is a performance booster for TFETs. There will be optimizations of the EOT for drivability and leakage. Meanwhile, the gate leakage will be mainly due to the direct tunneling considering the expected low-voltage operation.

In Sect. 3.2.4, we show the 100/0 charge partition in the TFET channel. Intuitively, this is because the tunnel junction plays a role of barrier for channel carrier to flow backward to the source. It will also affect the gate leakage partition. Figure 3.19 shows the schematic of the gate leakage in different TFET operations. For TFETs biased in the non-saturation region, the constant potential in region III determines that the gate leakage across the channel is almost uniform. Since the leakage path to the source side is blocked, all the current will flow into the drain. After TFETs enter into the saturation, the local oxide field near the drain is pointed from the channel to the gate electrode, so is the gate leakage. At the same time, the source is still not involved in the leaking. The gate leakage components are in accordance with the charge components in Fig. 3.14.

With the potential model we introduce in Sect. 3.2.2 and the well-known directly tunneling formulations, it is straightforward to obtain the gate leakage module. Figure 3.20a plots gate leakage and drain current in one sample TFET confirming the 100/0 partition of the gate leakage. Figure 3.20b demonstrates the flexibility of the leakage module in reproducing the device TCAD data for different device geometries.

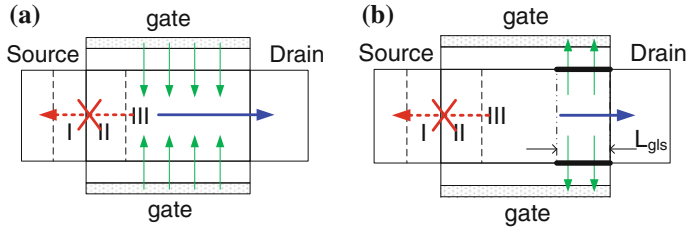

Fig. 3.19 Schematics of the gate tunneling in TFETs **a** before saturation and **b** after saturation

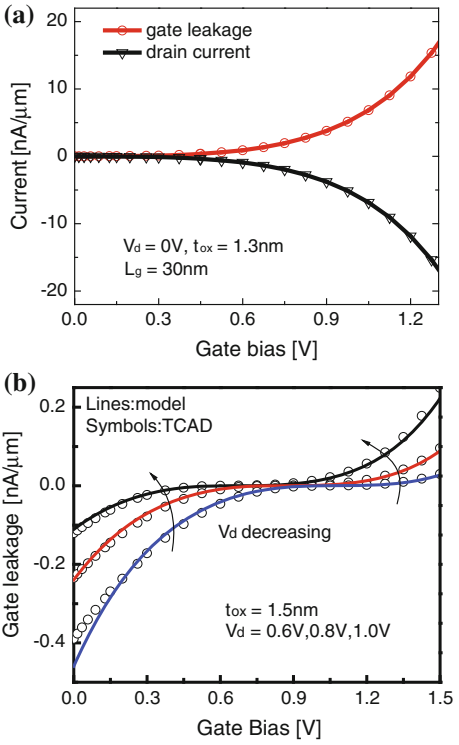

Fig. 3.20 **a** Partition of the gate leakage is 100 % to the drain as confirmed with TCAD simulations. **b** The leakage module is verified in (**b**)

3.3.2 *Modeling the Short-Channel Effects*

A second consideration is the short-channel effect in TFETs. We note in Sect. 3.2.1 that when the TFET is turned on, its channel is already filled with free carriers injected from the drain. On the other hand, what we learn about the short-channel effects in MOSFET is that they are significant in the subthreshold regions and become less important in the strong inversions. The fundamental reason for short-channel effects is the competitive control over the conduction by the gate and drain voltages. When the MOSFET enters the strong inversion, the electric field from the drain terminal is screened by the free carriers, reducing the effect from the drain side. Coming back to TFETs, the drain effect on the tunneling junction is also screened by the free carriers in the channel if the drain voltage is small, which means the short-channel effects purely from the gate length scaling are not significant. In another perspective, under the gate voltage around the offset voltage $V_{off,tfet}$, the length of region II is only around ten nanometers. It means before the channel length is scaled to be around the length of region II, short-channel effects due to geometry are not important.

In the long-channel TFETs, the tunneling current saturates when the surface potential in region III saturates due to the increase of the drain voltage. On the other

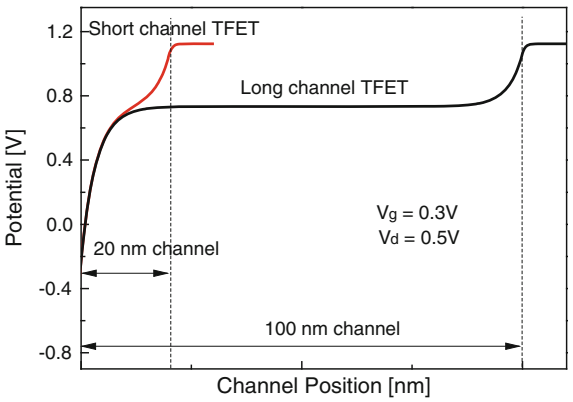

Fig. 3.21 In short-channel TFETs, the drain electric field penetrates into the tunnel junction

hand, the surface potential saturation is due to the depletion of the free carriers from the channel. Near or around the saturation drain voltage, the electric field from the drain terminal starts penetrating into the channel. Similar to the channel length modulation effect in the MOSFETs, the effective channel length is also reduced in TFETs. Although even in short-channel TFETs the tunneling current only depends on the tunnel junction, the drain-modulated channel region may eventually reach region II and affected the potential profiles across the junction. On the other hand, we note that across the tunnel junction, a potential difference corresponding to the material band gap is a necessary condition to induce the interband tunneling, and how much the additional field from the drain side will affect the junction field depends on the device biasing voltages. Overall the short-channel effects in TFETs are less than those in MOSFETs.

Figure 3.21 compares the surface potential profiles along the channel of two TFETs when they both worked in the 'saturation' region (with $V_d = V_{dd} = 0.5$ V, $V_g = 0.6V_{dd} = 0.3$ V,). When the channel length is 100 nm, the tunnel junction and the channel length modulation region are totally separated. Electric field at the edge of region II in the channel side is still almost zero. However, when the channel is scaled to 20 nm, the tunnel junction and the channel length modulation region meet. As a result, the electric field from the drain side helps increasing the field across the tunnel junction and also the current. With the mathematic language, the one-dimensional assumption in Eqs. (3.7) and (3.8) becomes invalid in the saturation region. Modeling the drain voltage effects on the tunneling junction requires solutions of the two-dimensional Poisson's equation in the channel length modulation region, which is a challenging task similar to the corresponding module in MOSFETs.

A model for short-channel effects in TFETs was reported recently [41]. By directly solving the quasi-two-dimensional Poisson's equation in the TFET channel, the potential model was derived. The current modeling follows the similar minimum tunnel distance method as discussed in Sect. 3.2.3, and a closed-form current model was obtained.

3.4 Modeling Heterojunction TFETs

Among several TFET performance boosters we introduced in Chap. 1, the heterojunction TFETs (HTFET) have been intensively explored. There are raising interests in developing a compact model for HTFETs [42]. Similar to the development of the homojunction TFET model in previous sections, the model core for HTFET will be built first including descriptions of its current and terminal charge properties.

One main feature in HTFETs is that the on-state current can be largely enhanced, even close to the thermal emission current of MOSFETs. This has two implications on the core model development. The first one is about the zero-order approximation of the coupled quantum tunneling and DD transport problem in TFETs. With limited current density and large carrier concentrations in the channel, the electric field in region III of the homojunction TFETs is assumed to be zero. While the current density is increased a lot in HTFETs, the zero-field approximation in region III probably should be eliminated. The detailed value of the channel electric field depends on the carrier mobility in the DD. In this scenario, the current continuity along the TFET channel between the tunneling current and the DD current should be solved to derive the final current. On the other hand, solutions of the electrostatic potentials are needed to derive the tunneling current, which requires the boundary conditions between the tunnel diode and the MOS channel. Figure 3.22 shows this difficulty in HTFET model development. The second implication is about the terminal charge modeling. With the zero-field approximation, the channel charge is easily found by Eq. (3.19) without integral along the channel. Eventually, in HTFETs, we may need to revisit the charge formulations. By inserting an internal node between the tunnel junction and MOS channel, we reported one solution for the complex coupled problem [28].

Another feature in HTFETs is inside the tunneling heterojunction. Figure 1.19 in Chap. 1 plots the band diagram of the Ge/Si heterojunction. For convenience, it is repeated here in Fig. 3.23 but expanded to generally represent the staggered heterostructure. Along the tunnel path (dashed line), the electron wave first has the properties of material A in the HTFET source and then has the properties of material B after entering the HTFET channel. With the complex band picture, the wave decaying appears in material A with its imaginary branch near the top of the valence band (bold arrow) and continues in material B with its imaginary branch

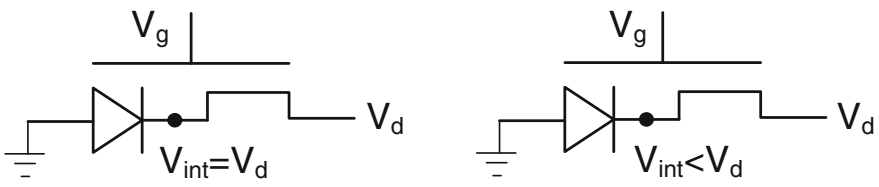

Fig. 3.22 In HTFETs, the carrier transport in the tunnel junction and the MOS channel may be coupled

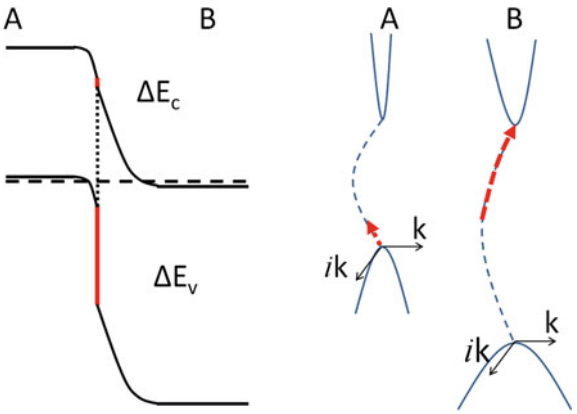

Fig. 3.23 A general description of the interband tunneling in the heterojunction, showing the challenge in HTFET modeling

near the bottom of the conduction band (bold arrow). In principle, the tunneling probability in the heterojunction depends on both materials. At the same time, different paths in the tunneling energy window are not exactly the same. In other words, portions of the tunnel path are dynamically changing with the bias. It is important to find one reasonable approximation to derive a closed-form solution for the tunneling current.

There are simplified cases, for example, the InAs/Si HTFET. Due to the low doping in the InAs source, it is shown [43] that most part of the tunneling happens inside the InAs source, instead of across both InAs and Si. The above general descriptions serve as a reference, and we will need to analyze the specific cases of the HTFETs.

3.5 Summary

In this chapter, we discuss the TFET model developments to support the evaluations of their low-power circuit applications. After reviewing the TFET modeling status, we proceed to analyze the essential device physics of TFETs and propose a framework for self-consistently modeling the current and terminal charges. While the equation derivations we introduce in the chapter are relatively straightforward, it is the reasonable approximations of the complex device characteristics to set up these equations that make the compact model development interesting. We finally describe the possible challenges to develop the heterojunction TFET model.

Acknowledgement This work was supported by the General Research Fund from the Research Grant Council of Hong Kong under project number 611012. We would also like to thank Prof. Zhiping Yu and Mr. Peng Wu of Tsinghua University for the discussions on short-channel effects in TFETs.

References

1. G. Dewey, B.C. Kung, J. Boardman, J.M. Fastenau, J. Kavalieros, R. Kotlyar, W.K. Liu, D. Lubyshev, M. Metz, N. Mukherjee, P. Oakey, R. Pillarisetty, M. Radosavljevic, H. Then, R. Chau, Fabrication, characterization, and physics of III-V heterojunction tunneling field effect transistors for steep sub-threshold swing. IEDM Tech. Dig. 785–788 (2011)
2. H. Riel, K. Moselund, C. Bessire, M.T. Bjork, A. Schenk, H. Ghoneim, H. Schmid, InAs/Si heterojunc-tion nanowire tunnel diodes and tunnel FETs. IEDM Tech. Dig. 391–394 (2012)
3. M. Goto, Y. Kondo, Y. Morita, S. Migita, A. Hokazono, H. Ota, M. Masahara, S. Kawanaka, The guideline of Si/SiGe heterojunction design in parallel plate style TFETs for Si CMOS platform implementation. SSDM Tech. Dig. 852–853 (2014)
4. U.E. Avci, S. Hasan, D.E. Nikonov, R. Rios, K. Kuhn, I.A. Young, Understanding the feasibility of scaled III-V TFET for logic by bringing atomistic simulations and experimental results. VLSI Tech. Dig. 183–184 (2012)
5. U.E. Avci, I.A. Young, Heterojunction TFET scaling and resonant-TFET for steep subthreshold slope at sub-9 nm gate length. IEDM Tech. Dig. 96–99 (2013)
6. http://steeper-project.org/index.php. Accessed June 2015
7. http://www.e2switch.org/index.php. Accessed June 2015
8. https://www.fbo.gov/index?s=opportunity&mode=form&id=bf6fe92c53588c57892b5a787cf66146&tab=core&tabmode=list&=. Accessed June 2015
9. Y. Tsividis, Operation and Modeling of the MOS transistor. (Oxford University Press, New York)
10. D. Kim, Y. Lee, J. Cai, I. Lauer, L. Chang, S.J. Koster, D. Sylvester, D. Blaauw, Low power circuit design based on heterojunction tunneling transistors (HETTs). Proc. ISLPED 219–224 (2009)
11. Y. Hong, Y. Yang, L. Yang, G. Samudra, C.H. Heng, Y.C. Yeo, SPICE behavioral model of the tunneling field-effect transistor for circuit simulation. IEEE Trans. Circuits Syst. II Express Briefs **56**(12), 946–950 (2009)
12. V. Saripalli, A.K. Mishra, S. Datta, V. Narayanan, An energy efficient heterogeneous CMP based on hybrid TFET-CMOS cores. Proc DAC 729–734 (2011)
13. W.G. Vandenberghe, A.S. Verhulst, G. Groeseneken, B. Soree, W. Magnus, Analytical model for point and line tunneling in a tunnel field-effect-transistor. Proc. SISPAD 137–140 (2008)
14. A.S. Verhulst, B. Soree, D. Leonelli, W.G. Vandenberghe, G. Groeseneken, Modeling the single-gate, double-gate, and gate-all-around tunnel field-effect-transistor. J. Appl. Phys. **107** (2), 024518 (2010)
15. A.S. Verhulst, D. Leonelli, R. Rooyachers, G. Groeseneken, Drain voltage dependent analytical model of tunnel field-effect transistors. J. Appl. Phys. **110**(2), 024510 (2011)
16. M.G. Bardon, H.P. Neves, R. Puers, C.V. Hoof, Pseudo-two-dimensional model for double-gate tunnel FETs considering the junctions depletion regions. IEEE Trans. Electron. Devices **57**(4), 827–834 (2010)
17. M.J. Lee, W.Y. Choi, Analytical model of single-gate silicon-on-insulator tunneling field-effect transistors. Solid State Elec. **30**(1), 110–114 (2011)
18. J. Wan, C.L. Royer, A. Zaslavsky, S. Cristoloveanu, A tunneling field effect transistor model combining interband tunneling and channel transport. J. Appl. Phys. **110**(10), 104503 (2011)
19. Y. Yang, X. Tong, L. Yang, P. Guo, L. Fan, Y. Yeo, Tunneling Field-effect transistor: capacitance components and modeling. IEEE Electron. Devices Lett. **31**(7), 752–754 (2010)
20. P.M. Solomon, D.J. Frank, S.O. Koswatta, Compact model and performance estimation for tunneling nanowire FET. Proc. DRC 197–198 (2011)
21. A.S. Verhulst, W.G. Vandenberghe, K. Maex, G. Groeseneken, Tunnel field-effect transistor without gate-drain overlap. Appl. Phys. Lett. **91**, 053102 (2007)
22. A. Ortiz-Conde, F. Sánchez, J. Muci, A. Barrios, J.J. Liou, C. Ho, Revisiting MOSFET threshold voltage extraction methods. Micro. Relia. **53**(1), 90–104 (2013)

23. D. Mohata, B. Rajamohanan, T. Mayer, M. Hudait, J. Fastenau, D. Lubyshey, A. Liu, S. Datta, Barrier-engineered arsenide-antimonide heterojunction tunnel FETs with enhanced drive current. IEEE Electron. Device Lett. **33**(11), 1568–1570 (2012)
24. K. Moselund, M.T. Björk, H. Schmid, H. Ghoneim, S. Karg, E. Lörtscher, W. Riess, H. Riel, Silicon nanowire tunnel FETs: low temperature operation and influence of high-k gate dielectric. IEEE Trans. Electron. Devices **58**(9), 2911–2916 (2011)
25. J. Nah, E. Liu, K. Varahramyan, D. Dillen, S. McCoy, J. Chan, E. Tutuc, Enhanced-performance germanium nanowire tunneling field-effect transistors using flash-assisted rapid thermal process. IEEE Electron. Device Lett. **31**(12), 1359–1361 (2010)
26. S. Richter, C. Sandow, A. Nichau, S. Trellenkamp, M. Schmidt, R. Luptak, K. Bourdelle, Q. Zhao, S. Mantl, Omega-gated silicon and strained silicon nanowire array tunneling FETs. IEEE Electron. Device Lett. **33**(11), 1535–1537 (2012)
27. P.M. Solomon, J. Jopling, D.J. Frank, C. D'Emic, O. Dokumaci, P. Ronsheim, W. Haensch, Universal tunneling behavior in technologically relevant P/N junction diodes. J. Appl. Phys. **95**(10), 5800–5812 (2004)
28. L. Zhang, M. Chan, SPICE modeling of double-gate tunnel FETs including channel transports. IEEE Trans. Electron. Devices **61**(2), 300–307 (2014)
29. S. Mookerjea, R. Krishnan, S. Datta, V. Narayanan, On enhanced Miller capacitance effect in interband tunnel transistor. IEEE Electron. Devices Lett. **30**(10), 1102–1104 (2009)
30. L. Zhang, X. Lin, J. He, M. Chan, An analytical charge model for double-gate tunneling FETs. IEEE Trans. Electron. Devices **59**(12), 3217–3223 (2012)
31. K.K. Young, Short-channel effect in fully depleted SOI MOSFET's. IEEE Trans. Electron. Devices **36**(2), 399–402 (1989)
32. L. Zhang, S. Wang, C. Ma, J. He, C. Xu, Y. Ma, Y. Ye, H. Liang, Q. Chen, M. Chan, Gate underlap design for short channel effects control in cylindrical gate-all around MOSFETs based on an analytical model. IETE Tech. Rev. **29**(1), 29–35 (2012)
33. S. Lin, J. Kuo, Modeling the fringing electric field effect on the threshold voltage of FD SOI nMOS devices with the LDD/sidewall oxide spacer structure. IEEE Trans. Electron. Devices **50**(12), 2559–2564 (2003)
34. E. Arnold, Charge-sheet model for silicon carbide inversion layers. IEEE Trans. Electron. Devices **46**(3), 497–503 (1999)
35. X. Aymerich-Humet, F. Serra-Mestres, J. Millan, A generalized approximation of the Fermi-Dirac integrals. J. Appl. Phys. **54**(5), 2850–2851 (1983)
36. R. Yan, A. Ourmazd, K.F. Lee, Scaling the Si MOSFET: from bulk to SOI to bulk. IEEE Trans. Electron. Devices **39**(7), 1704–1710 (1992)
37. M. Chan, T. Y. Man, J. he, X. Xi, C. Lin, X. Lin, P. K. Ko, A. M. Niknejad, C. Hu, Quasi-2D compact modeling for double-gate MOSFET. WCM Tech. Dig. 108–113 (2004)
38. S. Agrawal, J.G. Fossum, A physical model for fringe capacitance in double-gate MOSFETs with non-abrupt source/drain junctions and gate underlap. IEEE Trans. Electron. Devices **57**(5), 1069–1075 (2010)
39. S.E. Laux, Techniques for small signal analysis of semiconductor devices. IEEE Trans. Electron. Device **32**(10), 2028–2037 (1985)
40. http://i-mos.org. Accessed June 2015
41. P. Wu, J. Zhang, L. Zhang, Z. Yu, Channel potential based compact model of double-gate tunneling FETs considering channel length scaling. SISPAD Tech. Dig. (2015)
42. Y. Taur, J. Wu, J. Min, An analytic model for heterojunction tunnel FETs with exponential barrier. IEEE Trans. Electron. Devices **62**(5), 1399–1404 (2015)
43. A. Schenk, Rhyner, R, M. Luisier, C. Bessire, Analysis of Si, InAs and Si-InAs tunnel diodes and tunnel FETs using different transport models. SISPAD Tech. Dig. 263–266 (2011)

Chapter 4
Challenges and Designs of TFET for Digital Applications

Ming-Long Fan, Yin-Nien Chen, Pin Su and Ching-Te Chuang

Abstract This chapter reviews the challenges and designs of digital TFET circuits. Several fundamental features of TFET such as unidirectional conduction, delayed saturation, and enhanced Miller capacitance are described with emphasis on their impacts on the functionality and robustness of logic and SRAM circuits. For TFET logic circuits, structural innovations and device design are demonstrated to facilitate compact circuit design and performance improvement. For SRAM, the advantages of hybrid TFET-MOSFET 8T SRAM cell in stability and efficiency of WRITE-assisted circuit to enhance performance are addressed. Moreover, the variability and backgate bias technique for TFET digital circuit design are highlighted.

4.1 Introduction

Tunnel FET (TFET) that utilizes band-to-band tunneling to conduct current enables sub-kT/q subthreshold transition for greener computation under extremely low supply voltage (V_{DD}). For state-of-the-art SoC, digital circuit occupies significant portion of the chip area and dominates the power consumption. Thus, the benefits, constraints, and trade-offs of TFET for digital applications merit extensive examination.

M.-L. Fan · Y.-N. Chen · P. Su (✉) · C.-T. Chuang
Department of Electronic Engineering and Institute of Electronics,
National Chiao Tung University, 1001 Ta-Hsueh Rd., Hsinchu 300, Taiwan
e-mail: pinsu@mail.nctu.edu.tw

M.-L. Fan
e-mail: mlfan.ee95@gmail.com; mlfana@tsmc.com

Y.-N. Chen
e-mail: snoopyfairy@gmail.com

C.-T. Chuang
e-mail: chingte.chuang@gmail.com

L. Zhang and M. Chan (eds.), *Tunneling Field Effect Transistor Technology*,
DOI 10.1007/978-3-319-31653-6_4

Recently, large amount of research efforts have focused on the device-level optimizations to enhance the ON-state current (I_{ON}) of TFET while maintaining satisfactory subthreshold characteristics (e.g., leakage and subthreshold slope) through device designs and various performance boosters. However, there are several fundamental challenges/opportunities resulting from the unique characteristics of TFET that need to be carefully understood and addressed in order to exploit the full potential of TFET for digital circuit applications. This chapter begins with the review of some fundamental and unique features of TFET at device level and points out their implications for digital circuit applications. Various logic and SRAM circuit designs are then proposed to overcome the limitations and exploit the benefits of TFET with the intention to inspire more innovations for the emerging TFET technology and circuits.

4.2 Characteristics/Challenges of TFET

4.2.1 Unidirectional Conduction

Unlike traditional MOSFET, TFET uses a reversely biased p–i–n gated diode to enable steep subthreshold slope by suddenly generating significant band-to-band tunneling current at the source/channel junction. Due to its asymmetrical structure, the output characteristics (I_D–V_{DS}) of TFET under forward and reverse drain biases (V_{DS}) are quite different (Fig. 4.1a). For an n-type TFET (the source, channel, and drain are doped with p-type, intrinsic, and n-type impurities, respectively), drastically lower current is observed for the p–i–n diode operating under moderately negative V_{DS} (before the forward conduction of diode), while TFET behaves like a conventional MOSFET under positive drain voltage (i.e., reverse p-i-n diode) [1–3]. Figure 4.1b shows the schematic of an asymmetrical n-type TFET with bold line indicating the location of source/channel tunneling junction. With the unidirectional

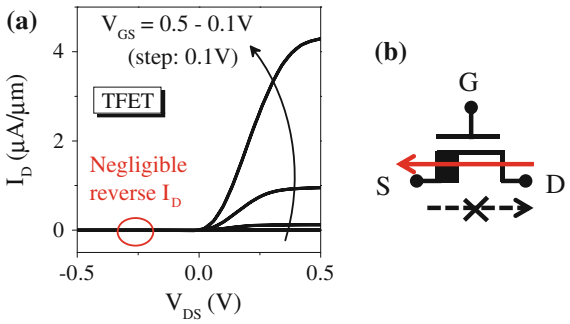

Fig. 4.1 **a** The output characteristics of TFET operating in positive and negative V_{DS} showing the unidirectional conduction and **b** the schematic/current conduction of an n-type TFET where the bold line indicates the tunneling junction between source/channel

conduction and asymmetrical structure, there are several concerns for TFET-based circuits: (1) special care for the design of source/drain regions with the correct orientation [2], (2) extra area overhead to separate/expand unexchangeable source/drain regions [2, 4], (3) potential circuit failure or reliability issues due to the disabled path to charge/discharge circuit internal nodes [1], and (4) loss of bidirectional pass-gate applications. On the other hand, the extremely low conductivity from the opposite direction of TFET offers opportunities for innovative circuit designs (e.g., compact multiplexer gate [1], half-select disturb-free SRAM cell [4], and RF rectifier [2]). Several circuit designs specific to TFET are addressed in Sects. 4.3 and 4.4.

4.2.2 *Delayed Saturation*

TFET exhibits superior current drive and subthreshold slope at low-V_{DD} operation. However, as V_{DD} increases to moderate level (≧0.2 V, depending on the device design), its output characteristics exhibit a broad, soft transition before the current reaches saturation. Figure 4.2a shows the output characteristics of TFET and MOSFET at V_{GS} = 0.5 V where the "delayed saturation" characteristics and lower current drive of TFET can be clearly seen [4–6]. In MOSFET, the gate voltage induces a (low resistivity) conducting channel, and the current saturation is governed by the "pinch-off" of the channel near the drain. In TFET, due to the lightly doped "intrinsic" channel region in the p–i–n structure, the resistivity of the channel is high. Physically, the TFET can be regarded as a source–channel tunneling junction in series with a resistor (i.e., channel resistance). At low V_{DS}, most of the applied drain voltage drops across the tunneling junction, resulting in direct thinning of tunneling barrier and thus causing the current to increase rapidly as V_{DS} increases. In this region, the critical tunneling length is reduced, and the available states for tunneling and the occupancy probability determine the transmission rate

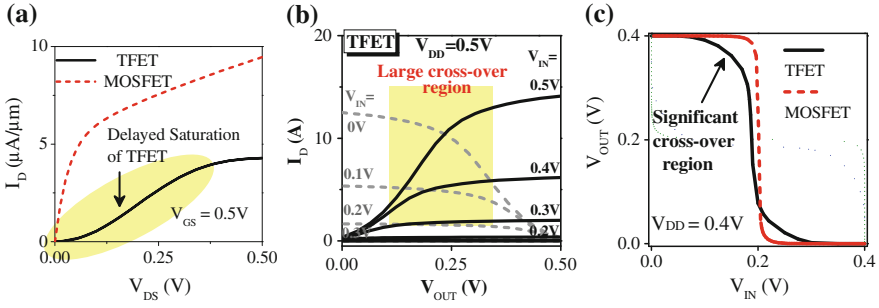

Fig. 4.2 **a** Comparison of output characteristics for TFET and MOSFET and **b** the switching characteristics of complementary TFETs in an inverter with the resulting voltage transfer curve shown in **c**

and the delayed output behavior [7]. It is desirable to have more available states/carriers for tunneling; hence, a degenerate source design is preferable for TFET [7, 8]. At medium and higher V_{DS}, most of the applied voltage drops across the channel resistance, so the current increases slowly, causing the saturation characteristics.

The "delayed saturation" of TFET leads to several impacts on TFET digital circuits. Figure 4.2b shows the switching characteristics of complementary TFETs in an inverter with the resulting voltage transfer curve (VTC) shown in Fig. 4.2c. As shown, the broad, soft transition characteristics of TFET result in large crossover region/current between the n-type and p-type TFETs, thus directly degrading the sharpness of inverter VTC [5, 9] and the stability of TFET SRAM cell [4, 5, 10], which will be discussed later.

4.2.3 *Enhanced Miller Capacitance*

For TFET, the capacitance performance differs from MOSFET. In MOSFET, once the inversion channel is formed, the source and drain regions are connected to the inversion channel with identical carrier polarity, and the gate capacitance (C_{GG}) is comparably partitioned into C_{GS} (gate–source capacitance) and C_{GD} (gate–drain capacitance). However, due to the source–channel barrier of TFET, the gate capacitance (C_{GG}) is essentially connected by the channel carriers to the drain; thus, C_{GD} is shown to dominate the total gate capacitance (Fig. 4.3a) [11, 12]. The large portion of C_{GD} in TFET increases Miller capacitance due to the enhanced capacitive coupling between gate and drain terminals (Fig. 4.3c). This large Miller capacitance induces significant overshoot/undershoot during circuit switching, causing large active power consumption and undermining the performance advantages of TFET (Fig. 4.3b) [8, 11–13]. The enhanced Miller coupling effect can be mitigated under certain circumstances: (1) circuit with heavy output loading capacitance (C_L) to

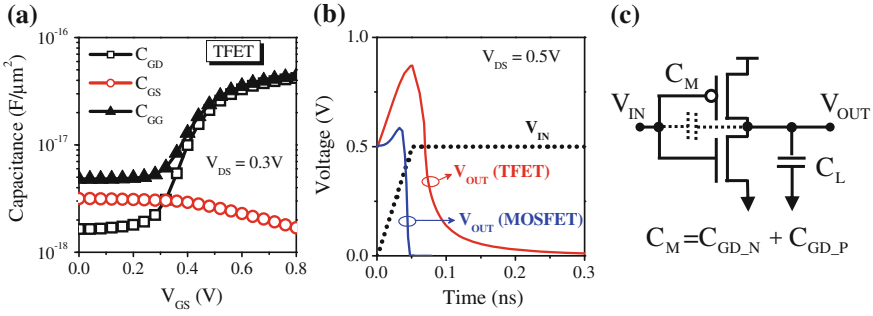

Fig. 4.3 **a** Capacitance–voltage characteristics of TFET and **b** the transient waveforms of TFET and MOSFET inverters. **c** Indication of Miller capacitance to cause significant overshoot/undershoot in the transient operation of TFET circuits

reduce the coupling factor (Fig. 4.3c) [11], (2) materials with lower density of states (DOSs) and bandgap to decrease gate capacitance and improve I_{ON} [11], and (3) device engineering in gate insulator, metal gate, or drain-side underlap [12]. The improvements in the performance and power of various TFET logic circuits are elaborated in Sect. 4.3.2.

4.2.4 Imbalanced Complementary TFETs

For low-power applications, comparable current drives between n-type pull-down and p-type pull-up networks are important to ensure satisfactory switching performance and sufficient/symmetric noise margin. Due to the change of electric field and allowable states around the tunneling junction, the source concentration influences the I_{ON} and transmission probability [14, 15]. For III–V materials that are promising for increasing the I_{ON} of TFET [e.g., heterojunction TFET (HTFET)], the source of a p-type TFET tends to become degenerate because of its lower conduction-band density of states, thus increasing the difficulty by simultaneously optimizing I_{ON} and subthreshold behaviors [16].

Figure 4.4(a) shows the design of HTFETs with different source doping levels (N_{SOURCE}). Note that for p-type HTFET (pHTFET) with N_{SOURCE} = 4E19 cm^{-3}, the lower conduction-band DOS of III–V materials makes the source region degenerately doped and moves the probability of available state ($1 - f_F(E)$) upward accordingly (Fig. 4.4b). In such case, the thermionic Fermi tail dominates the tunneling current and results in the degraded (MOSFET-like) subthreshold characteristics and lower current compared with that of n-type HTFET (nHTFET) under low V_{GS}. On the other hand, the decrease in N_{SOURCE} (e.g., N_{SOURCE} = 1E18 cm^{-3}) improves the subthreshold characteristics at the expense of lower I_{ON} arising from the reduced electric field near the tunneling junction. This deteriorates the CMOS matching, and thus, impractical sizing up of p-type TFET is required [17].

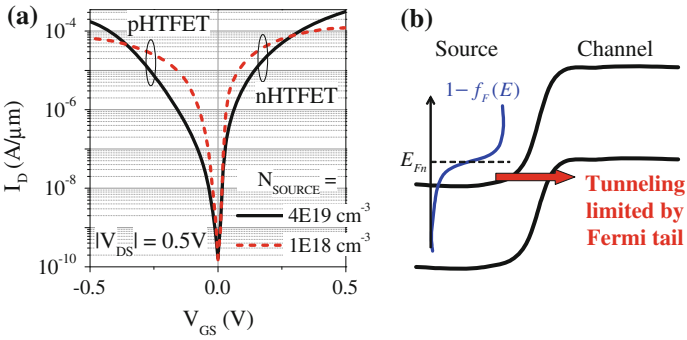

Fig. 4.4 **a** I_D–V_{GS} of complementary HTFETs with different source doping concentrations and **b** the band diagram of source–channel junction with source available Fermi states showing the cause of inferior subthreshold characteristics of p-type TFET with degenerate source [16]

Depending on the operating voltage or power–performance target, the optimization of suitable source doping concentration is important for TFET circuit design [14, 16, 18].

4.2.5 Variability

With the scaling of device dimensions, random variations have emerged as crucial concerns to degrade the viability and advantages of TFET. Among various variation sources, random telegraph noise (RTN) coming from the trapping/detrapping of the carriers at the interface trap and work-function variation (WFV) associated with the grain granularity characteristic of metal gate exhibit significant impacts on TFET [19, 20, 21].

Figure 4.5a shows the dependence of RTN amplitude ($\Delta I_D/I_D$) on the position of a single acceptor-type trap (carry a negative charge in trapped state) placed across the gate insulator/silicon sidewall interface. Because of the exponential dependence of the tunneling current on critical tunneling path, significant impact of RTN is observed for trap located near the tunneling junction (Region C′) and the influence decreases toward the drain side. Figure 4.5b shows the corresponding energy band diagrams of TFET along the channel length direction with a trapped acceptor-type trap at various locations. As shown, in the presence of a negatively charged trap near the tunneling junction, the band peaks up in the vicinity of the trap location, thus reducing the critical tunneling length and resulting in large RTN impact.

Figure 4.6a shows the influence of a single donor-type trap (carry a positive charge in detrapped state) placed at various locations. As shown, drastic degradations in OFF-state current (I_{OFF}) and significantly large RTN amplitude (inset of Fig. 4.6a) are observed. The broader region with severe susceptibility can be shown in Fig. 4.6b that illustrates the altered/shortened critical tunneling length even for the cases with trap located away from the tunneling junction. Compared with the conventional MOSFET (e.g., FinFET), TFET exhibits higher sensitivity to RTN.

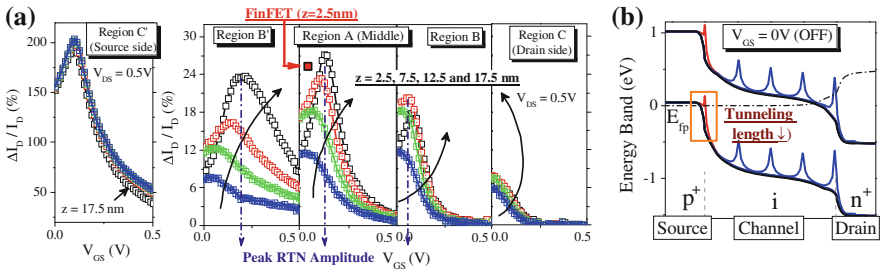

Fig. 4.5 **a** Dependence of RTN amplitude ($\Delta I_D/I_D$) on the location of a single acceptor-type trap at sidewall gate insulator/silicon interface and **b** the energy band diagrams along the channel length direction with a single acceptor trap placed at various locations. Regions C′, B′, A, B, and C separate the channel regions from source to drain in sequence. (z axis: fin height [19])

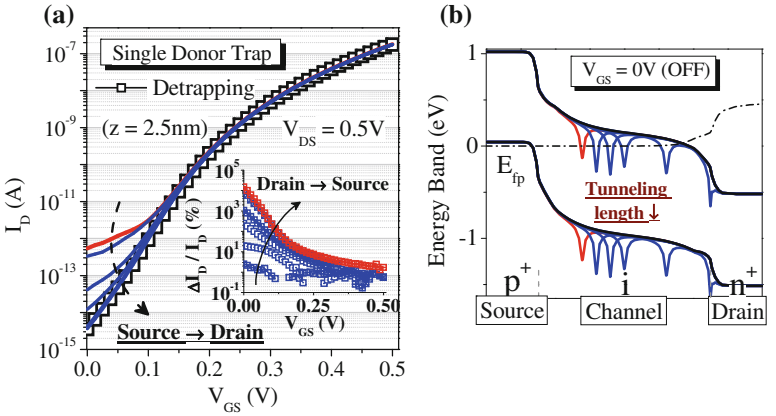

Fig. 4.6 **a** Impact of a single donor-type trap placed across the TFET sidewall interface along the channel length direction and **b** the associated energy band diagrams under OFF state [19]

The impact of WFV on the I_{ON} and I_{OFF} for TFET and FinFET under identical I_{OFF} and metal-gate grain patterns is compared in Fig. 4.7a. For TFET, the WFV influences I_{ON} and I_{OFF} through the grain patterns near the source and drain side, respectively. It is shown in Fig. 4.7b that higher I_{ON} and I_{OFF} occur for the TFET with small-work-function grain pattern. Because of its varying subthreshold swing and different dependence of I_{ON} and I_{OFF} on location of grain pattern, TFET shows weaker correlation between I_{ON} and I_{OFF} ($\rho = 0.49$) as opposed to the closer linkage ($\rho = 0.95$) found in the FinFET (attribute to the common dependence on threshold voltage). This implies that if one optimizes WFV for I_{ON}, it does not necessarily result in an improvement in the I_{OFF} variation, which may possibly emerge as a potential drawback of TFET. On the other hand, in terms of the optimization for I_{ON} and I_{OFF}, the lower correlation in TFET decouples the linkage between ON and

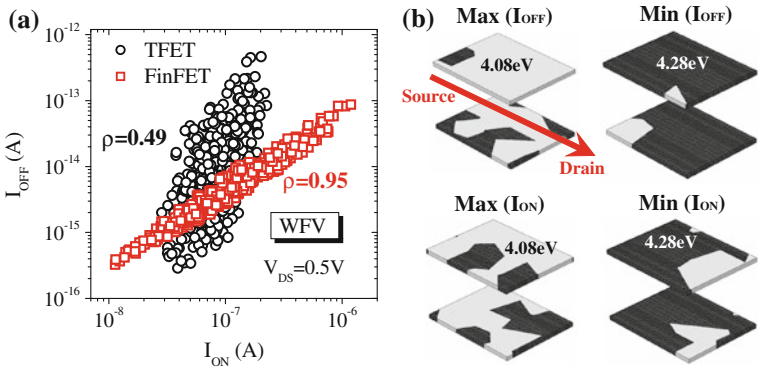

Fig. 4.7 **a** Comparison of I_{ON} and I_{OFF} for TFET and FinFET in the presence of WFV and **b** the metal-gate grain patterns associated with the maximum/minimum I_{ON} and I_{OFF} (*gray* and *black regions* stand for the grain with smaller and higher work-function value, respectively) [19]

OFF state, thus enabling us to independently and simultaneously improve I_{ON} and I_{OFF}.

The impacts of WFV on III-V HTFET, homojunction TFET, and FinFET devices have been investigated and compared in [21]. Due to the broken-gap nature, HTFET shows significantly steeper subthreshold slope and higher susceptibility to WFV near OFF state. For the I_{ON} variation, both the HTFET and homojunction TFET possess better immunity to WFV than the III–V FinFET. It has also been shown [21] that, for HTFET, the source-to-gate underlap design can suppress the impact of WFV on I_{OFF} variation with compromise of I_{ON} and subthreshold swing.

4.2.6 *Backgate Biasing Design*

In addition to the use of III–V materials to enhance the drive current of TFET, backgate biasing (V_{BS}) provides another design knob to optimize the leakage/delay of circuits [16, 22, 23]. Figure 4.8a compares the impact of V_{BS} on the I_D–V_{GS} characteristics of complementary HTFETs and MOSFETs under comparable I_{OFF}. As shown, HTFET exhibits significantly higher V_{BS} efficiency in modulating I_D (defined as I_D ($|V_{BS}|$ = 0.5 V)/I_D (V_{BS} = 0 V)) in the vicinity of OFF state. The significantly higher I_{OFF} modulation efficiency comes from the drastic reduction in the critical tunneling length under forward V_{BS} (i.e., V_{BS} = 0.5 V and −0.5 V for nHTFET and pHTFET, respectively). On the other hand, the impact of V_{BS} on HTFET rapidly decreases with the increasing V_{GS} and becomes considerably lower than that in the MOSFET. Figure 4.8b shows the I_{ON}/I_{OFF} comparisons of complementary HTFETs at various V_{DD}. For nHTFET, the forward and reverse V_{BS} are 0.5 V and −0.5 V, respectively, while the conditions are −0.5 and 0.5 V for pHTFET, respectively. As shown, reverse V_{BS} improves the I_{ON}/I_{OFF} ratio of HTFET with the aid of noticeable I_{OFF} suppression. At higher V_{DD}, the applicable

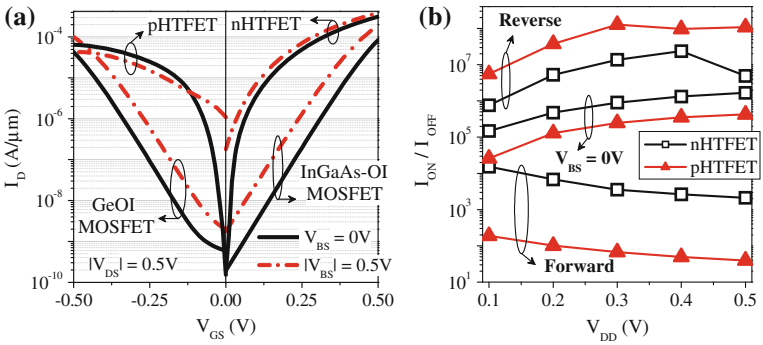

Fig. 4.8 **a** I_D–V_{GS} characteristics of complementary HTFETs and MOSFETs (composed of InGaAs-OI NFET and GeOI PFET) with different V_{BS} and **b** the impact of V_{BS} on the I_{ON}/I_{OFF} of HTFET at various V_{DD} [16]

reverse V_{BS} is larger and I_{ON}/I_{OFF} increases accordingly until significant ambipolar current (coming from the undesired tunneling current happening near the drain–channel junction) becomes dominant to decrease the I_{ON}/I_{OFF} ratio. Under forward V_{BS}, I_{OFF} increases considerably to degrade the I_{ON}/I_{OFF} ratio, especially at high V_{DD}.

4.3 Logic

4.3.1 ALL N-type TFET Design

As described in Sect. 4.2.4, the lack of matching n-type and p-type TFETs increases the difficulties in designing complementary TFET circuits. To mitigate the inferior subthreshold characteristics of p-type TFET, all n-type TFET pass-transistor logic (PTL) circuit has been developed to maintain the advantages of TFET under low-V_{DD} operation. The PTL circuit utilizes transistors as switches to pass logic signals between nodes of a circuit.

Figure 4.9 shows the schematic of PTL-based two-input AND using n-type TFETs [24]. Compared with the MOSFET counterpart, additional n-type TFET (enclosed in dashed/red box) is required to ensure correct operation under A = "low" and B = "high." Without this modification, the prohibition of TFET (controlled by signal B) to conduct current from source (indicated with bold line) to drain hinders the discharging of output node. Moreover, due to the inevitably poor "high" of using n-type transistor only, a weak p-type pull-up MOSFET is employed to restore a solid "high" in the output node.

For pass-gate application, using two n-type TFETs oriented in the opposite direction and operating in parallel is a straightforward approach to solve the unidirectional limitation. Instead of using two n-type TFETs to achieve bidirectional conduction, an alternative design with pass-gate TFETs placed in the same direction to merely discharge the output node is shown in Fig. 4.10 [15]. In such design, extra dynamic precharging network (one TFET connected to the output node (Y) and the other one controlled by another clock signal, CLK2) is added to provide an initial high-state output voltage before evaluation. Besides, the inclusion of TFET controlled

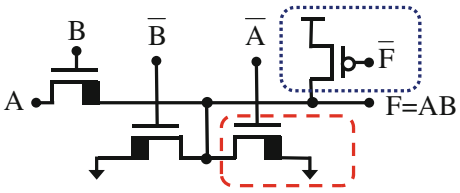

Fig. 4.9 Schematic of a PTL two-way AND using entire n-type TFET. The dashed/dotted boxes show extra n-type TFET and p-type MOSFET to ensure correct logic function and restore good "1", respectively [24]

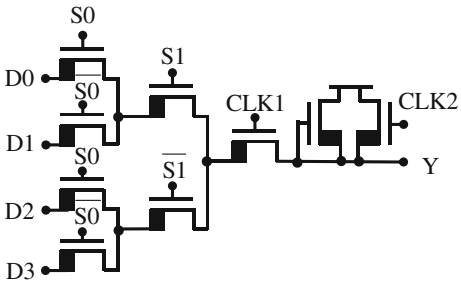

Fig. 4.10 Schematic of a 4-to-1 PTL multiplexer using precharged n-type TFET design [15]

by CLK1 is beneficial to prevent the possible charge sharing and isolate the inputs from the output node. Under suitable V_{DD} operating range, this design utilizes the unidirectional conduction feature to facilitate compact structure and lower internal capacitance.

4.3.2 Complementary TFET Logic Design

In addition to the constraint of unidirectional current conduction, the large Miller capacitance of TFET described in Sect. 4.2.3 is of particular importance and should be reduced for better performance and smaller dynamic power consumption. There are several device designs proposed to decrease the contribution of C_{GD} [12]: (1) dual-oxide (DOX), (2) drain-side underlap (DUND), and (3) dual-metal-gate work function (DWF). By placing low-κ and high-κ gate dielectrics close to the drain and source sides, respectively, Fig. 4.11 shows the impact of DOX technique on the I_{ON}, I_{OFF}, and C_{GD} of TFET under various lengths (portion) of high-κ gate dielectric (L_{HK}). It is shown that I_{ON} is less sensitive to the change in L_{HK} while

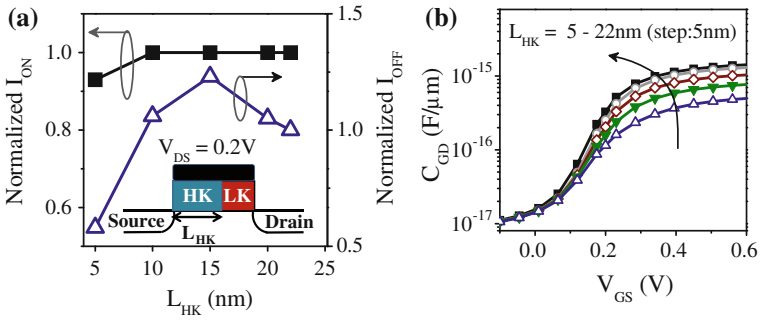

Fig. 4.11 **a** Normalized I_{ON} and I_{OFF} and **b** C_{GD}–V_{GS} characteristics of TFET with different lengths of high-k gate dielectric (L_{HK}) using dual-oxide design [12]

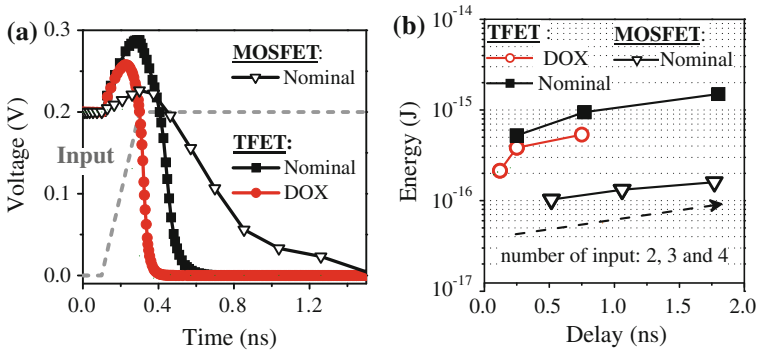

Fig. 4.12 **a** Comparison of switching transient waveforms of two-way NAND and the **b** dynamic energy versus delay with various inputs (two, three, and four) for TFET (nominal, DOX) and MOSFET (nominal) NANDs [13]

I_{OFF} non-monotonically varies with L_{HK}. In Fig. 4.11b, C_{GD} is found to decrease with the increasing component of low-κ material near the drain side (smaller L_{HK}), thus exhibiting potential for better switching efficiency. Because of its effectiveness to mitigate Miller capacitance, several important logic circuits designed with DOX TFET are evaluated and compared with the nominal MOSFET and TFET counterparts [13].

Figure 4.12a shows the transient waveforms of a two-way NAND during bottom switching transition. With DOX technique, the amount of overshoot is effectively reduced, thus enabling better switching performance compared with the nominal TFET. Furthermore, TFET with superior current drivability enhances discharging action than MOSFET under low-V_{DD} operation. The dynamic energy versus delay among various device designs under different input numbers is summarized in Fig. 4.12b. As expected, DOX TFET improves the performance (delay) and the enhancement increases with more inputs. However, the large Miller capacitive coupling renders the dynamic energy of TFET still inferior to that of MOSFET. Thus, more explorations to further reduce the enhanced Miller capacitance are necessary.

The comparisons of three-stage unloaded inverter chain and bus driver (with different bus loadings, C_L) using DOX TFET, nominal TFET, and MOSFET are shown in Fig. 4.13. For inverter chain (Fig. 4.13a), the DOX TFET offers the best delay, but still suffers from higher dynamic energy due to larger overshoot/undershoot than that of MOSFET. From the aspect of energy–delay product (EDP) that correlates with the energy efficiency of logic gates, DOX TFET exhibits comparable and superior switching energy efficiency to the nominal MOSFET and TFET counterpart, respectively. In Fig. 4.13b, comparisons of dynamic energy versus delay under different loadings are illustrated for various bus drivers. Due to

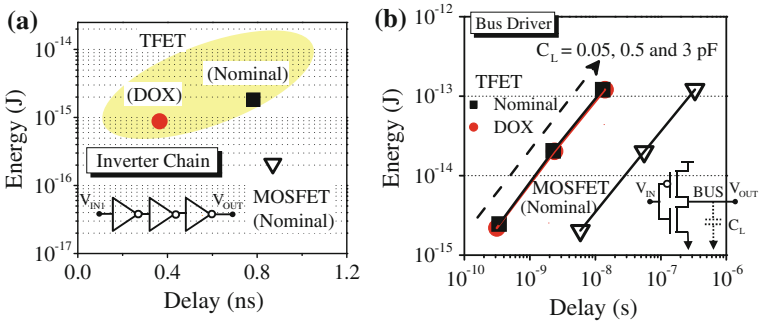

Fig. 4.13 **a** The dynamic energy versus delay for three-stage unloaded inverter chain and **b** bus driver under different bus-line loadings using different device designs [13]

the suppression of Miller capacitance in DOX TFET and with significant C_L, the contribution of superior current drivability in TFET increases such that TFET enables better dynamic energy and delay (thus, EDP) than the design with nominal MOSFET. Moreover, compared with the nominal TFET design, DOX TFET exhibits enhancements for the case with light loading where Miller capacitance is noticeable ($C_L = 0.05$ pF), whereas the advantages decrease with increasing loading (see Sect. 4.2.3).

Due to its better ability to eliminate the risk of data write-back through transmission gate under low-V_{DD} operation, the clocked CMOS latch exhibits potential for TFET application [25]. The schematic of a clocked CMOS latch is shown in Fig. 4.14a, while the dynamic energy versus CLK-to-Q delay is summarized in Fig. 4.14b. Compared with the nominal TFET latch, DOX design with reduced Miller capacitive coupling drastically enhances the delay and energy consumption to facilitate better energy efficiency.

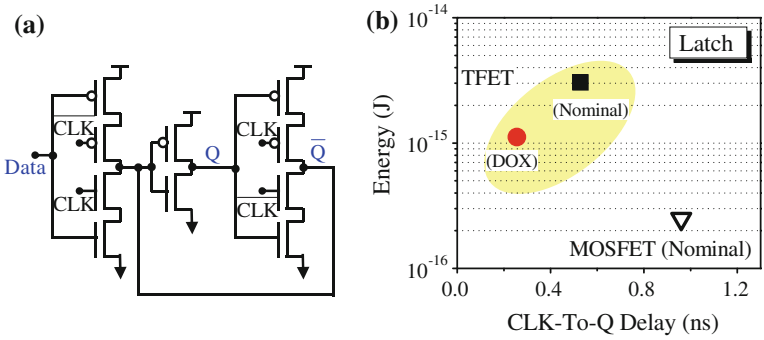

Fig. 4.14 **a** Schematic of clocked CMOS latch [25] and **b** the comparison of dynamic energy versus CLK-to-Q delay of the latch using nominal TFET, MOSFET, and DOX TFET [13]

4.4 SRAM

4.4.1 Problem of Conventional 6T SRAM Cell

Figure 4.15a shows the conventional 6T SRAM cell composing of two cross-coupled inverters to store the data and two pass-gate transistors [controlled by word line (WL)] to access the cell. During READ operation, the precharged bit line (BL/BLB) is released and the low-going bit line starts to discharge through the cell pass-gate and pull-down transistors from the side of storing "low" after the activation of WL. For WRITE operation, one of the bit line is pulled down by the WRITE driver to pull down the corresponding cell storage node. In such case, bidirectional current conduction is necessary for the pass-gate transistors, and the functionality of the conventional 6T TFET SRAM cell is impeded due to unidirectional conduction characteristic of TFET (Sect. 4.2.1).

The 8T SRAM cell (Fig. 4.15b) [26] with decoupled READ and WRITE paths appears to be a promising candidate for TFET applications. For 8T SRAM cell, the READ stability is improved with the elimination of READ disturb through the dedicated READ stack (solid/blue line in Fig. 4.15b) which separates the cell storage nodes from READ current path. Furthermore, the unidirectional WRITE access TFETs eliminate WRITE half-select disturb (i.e., half-select cells on the selected row perform "dummy" 6T SRAM like READ), thus facilitating bit-interleaving architecture to enhance soft error immunity. However, there are two drawbacks for the 8T TFET cell: (1) large crossover region of TFET (Sect. 4.2.2) that degrades the HOLD/READ static noise margin (HSNM/RSNM) and WRITE SNM (WSNM), and (2) lack of push-pull action during WRITE due to the unidirectional conduction of pass-gate TFETs (Sect. 4.2.1) that degrades the WRITE ability.

4.4.2 Hybrid TFET-MOSFET SRAM Cell Design

Based on the pros and cons of standard 8T TFET SRAM cell, Fig. 4.16 shows a hybrid TFET-MOSFET 8T SRAM cell utilizing MOSFET cross-coupled inverters

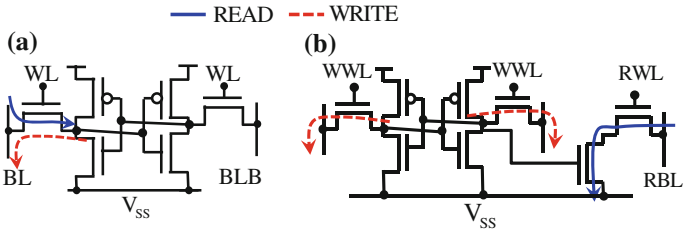

Fig. 4.15 Schematics of **a** the conventional 6T SRAM cell and **b** standard 8T SRAM cell associated with the corresponding READ (*solid/blue line*) and WRITE (*dashed/red line*) paths

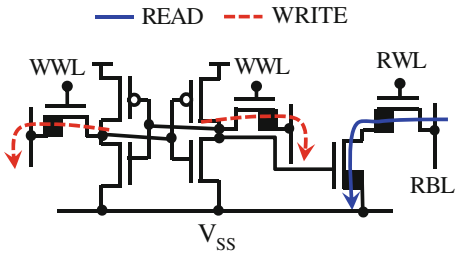

Fig. 4.16 Structure and the corresponding READ/WRITE paths of hybrid TFET-MOSFET 8T SRAM cell with MOSFET as internal cross-coupled inverter pair [4]

for robust HSNM and RSNM, and dedicated TFET READ stack for enhanced READ stability and performance [4]. Moreover, the use of pass-gate TFET facilitates bit-interleaving design and offers superior WRITE ability due to the current disparity between pass-gate TFET and pull-up p-type MOSFET at low-voltage operation.

The comparisons of HSNM, RSNM, WSNM, and half-select SNM (HSSNM, stability of the unselected cells on the selected row to perform dummy 6T SRAM like READ) among the hybrid TFET-MOSFET, TFET, and MOSFET 8T SRAM cells across various V_{DD} are shown in Fig. 4.17. For 8T cell, the RSNM equals to

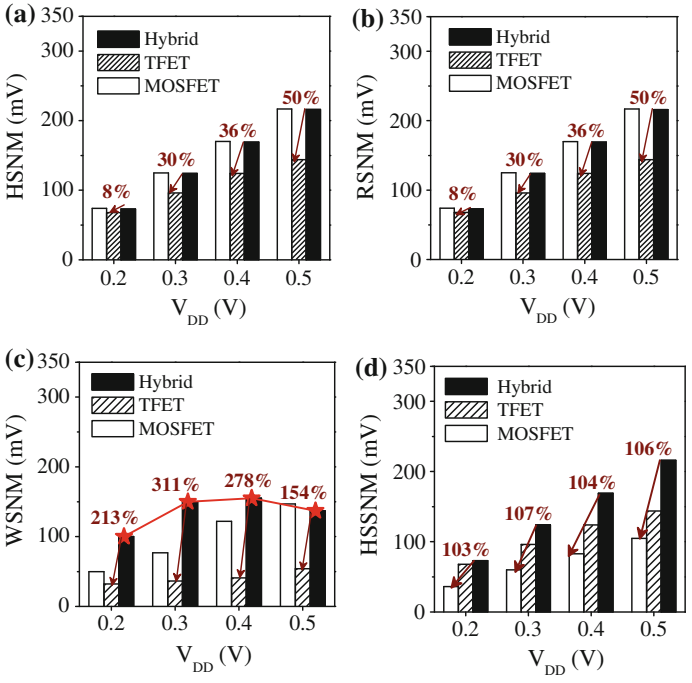

Fig. 4.17 Stability comparison of hybrid TFET-MOSFET, TFET, and MOSFET 8T SRAM cells across various V_{DD}: **a** HSNM, **b** RSNM, **c** WSNM, and **d** HSSNM [4]

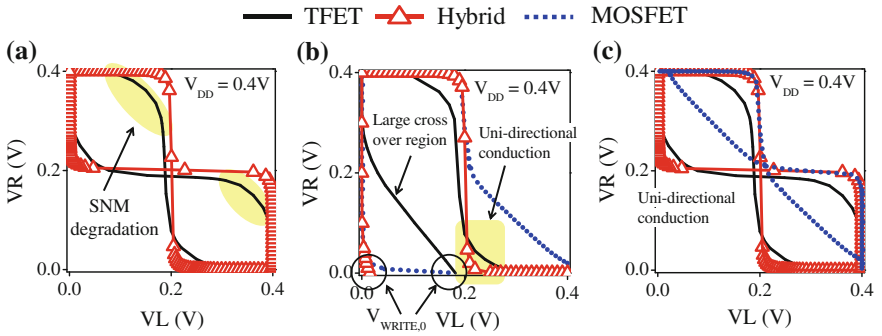

Fig. 4.18 Comparison of butterfly curves for hybrid TFET-MOSFET, TFET, and MOSFET 8T SRAM cells during **a** HOLD/READ, **b** WRITE, and **c** half-select operations at V_{DD} = 0.4 V [4]

HSNM. As shown in Fig. 4.17a, b, the hybrid TFET-MOSFET and MOSFET 8T cells exhibit comparable RSNM and HSNM, whereas the TFET 8T cell with significant crossover region to degrade the sharpness of inverter transition (Fig. 4.18a) possesses inferior RSNM/HSNM, particularly for $V_{DD} \geqq 0.3$ V. The actual amount of degradation depends on the output conductance of the manufactured TFET and MOSFET in weak inversions.

Figure 4.17c shows the WSNM of the three cells across various V_{DD}. Due to the unidirectional limitation of pass-gate TFET to deprive the push-pull action, the TFET 8T cell exhibits significant degradation in WSNM (Fig. 4.18b). Meanwhile, as shown in Fig. 4.18b, the large crossover region of TFET is observed to cause large $V_{WRITE,0}$ (determined by the current balance between pass-gate and pull-up transistors), thus further squeezing WSNM. From the comparisons, the significant reduction of $V_{WRITE,0}$ resulting from the large disparity between pass-gate TFET and pull-up MOSFET enables superior WSNM for hybrid TFET-MOSFET 8T SRAM cell.

Figure 4.17d shows the comparisons of HSSNM among three 8T SRAM cells. As shown, TFET and hybrid TFET-MOSFET 8T SRAM cells with unidirectional pass-gate TFET exhibit better HSSNM than that of MOSFET cell. The observed improvements attribute to the fact that with unidirectional pass-gate TFETs, the "dummy" READ current of the half-selected cells contributes negligible current to the storage nodes, therefore significantly mitigating half-selected disturb. This can be found in Fig. 4.18c with the corresponding butterfly curves. Furthermore, the hybrid TFET-MOSFET 8T cell with superior transition characteristic of the MOSFET-based cross-coupled inverters further improves the margin as compared with the TFET cell.

The performance comparisons of various low-V_{DD} 8T SRAM cells are illustrated in Fig. 4.19. During READ operation, the performance is estimated as the time from when selected READ WL (RWL) reaches half V_{DD} to when the READ BL (RBL) discharges to half V_{DD}. It is shown in Fig. 4.19a that with TFET

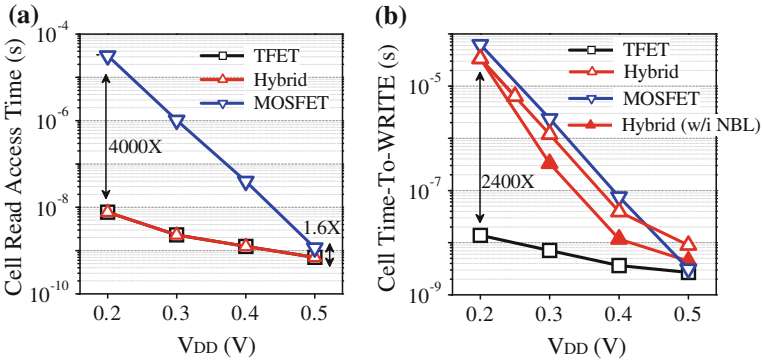

Fig. 4.19 Performance comparison of hybrid TFET-MOSFET, TFET, and MOSFET 8T SRAM cells: **a** cell READ access time and **b** cell Time-to-WRITE [4]

in the stacked READ path, TFET and hybrid TFET-MOSFET 8T SRAM cells significantly outperform the MOSFET counterpart at low-V_{DD} regime.

Figure 4.19b shows the Time-to-WRITE (the time from the 50 % activation of WRITE WL (WWL) to the time when the node voltage (pulling to "high") reaches 90 % V_{DD}). It is observed that both hybrid TFET-MOSFET and MOSFET 8T SRAM cells exhibit substantial longer Time-to-WRITE. Note that the WRITE process consists of two phases: (1) the competition between pass-gate TFET (or MOSFET) and holding pull-up MOSFET to pull down the high-state node voltage, and (2) the pull-up MOSFET from the opposite cell inverter to charge the opposite cell "low" storage node and trigger the latching mechanism to complete the whole operation. Thus, due to the use of low-current p-type MOSFET at low V_{DD}, the pull-up of the opposite cell "low" node voltage (second phase of the WRITE operation) in the hybrid TFET-MOSFET and MOSFET 8T SRAM cells is slow to degrade WRITE performance.

4.4.3 SRAM Peripheral Circuits

The inferior WRITE performance of the hybrid TFET-MOSFET 8T SRAM cell (Fig. 4.19b) can be improved with several WRITE-assisted circuits: (1) collapsing cell V_{DD} [27], (2) raising cell V_{SS} [28], (3) boosted WWL [29], and (4) negative WBL (NBL) voltage [30]. Among these techniques, collapsing cell V_{DD} and raising cell V_{SS} may degrade the stability of unselected cells on the selected column, while the boosting WWL introduces aggravated half-selected disturb for other unselected cells on the selected row. As such, the NBL technique with a low-going boosting control signal to capacitively coupling (through C_{BOOST}) a transient negative voltage to WBL is used to increase the V_{GS} and V_{DS} of the pass-gate transistor, thus improving the WRITE ability and WRITE performance of hybrid TFET-MOSFET

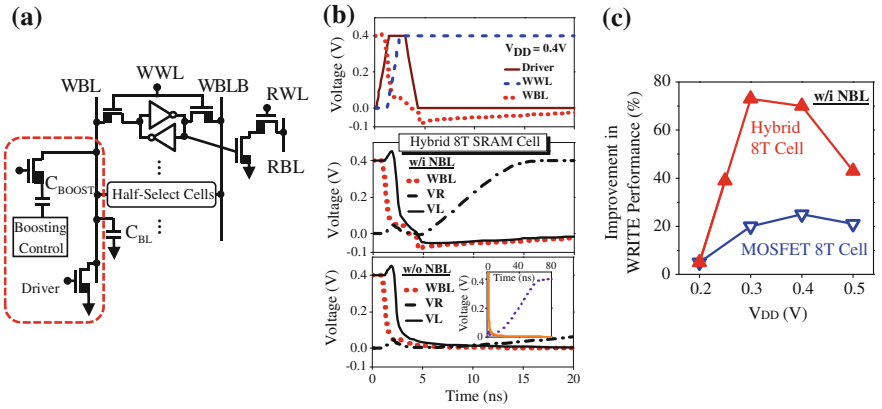

Fig. 4.20 **a** Schematic of negative bit-line (NBL) WRITE-assisted circuit and **b** waveforms of the hybrid 8T TFET-MOSFET cell with and without NBL. Inset of the figure shows the expanded time for the cell to WRITE without NBL. **c** Comparison of improvement in Time-to-WRITE with NBL [4]. VL/VR are voltages for left/right storage nodes, respectively

and MOSFET 8T cells (Fig. 4.20a). Figure 4.20b shows the transient waveforms for the hybrid TFET-MOSFET 8T cell with and without NBL during WRITE operation. It is shown that NBL enhances the WRITE action and performance. Moreover, compared with the MOSFET 8T cell, the hybrid cell is found to exhibit larger improvement using NBL WRITE-assisted circuit for V_{DD} below 0.5 V (Fig. 4.20c). The superior enhancement can be attributed to the better subthreshold swing of pass-gate TFET that offers more current increase with NBL.

Besides the assisted circuits to improve cell stability/performance, the robustness of several TFET sense amplifiers is addressed in this section. Two commonly used differential small-signal sense amplifiers are discussed: (1) current latch sense amplifier [31] (CLSA, in Fig. 4.21) and (2) voltage latch sense amplifier [32] (VLSA, in Fig. 4.22). For the activation of CLSA and VLSA, the sense enable (SE) signal goes to high state to sense the bit-line (BL/BLB) differential voltage through current and voltage modulation, respectively. The voltage difference in bit line is amplified through the current/voltage mismatch of two branches inside CLSA and VLSA. With variations, the offset voltage (V_{OS}) is used as the indicator to quantify the robustness of differential sense amplifiers [32]. In the presence of RTN that is recognized as an important variation source for TFET (Sect. 4.2.5), the originally symmetrical/balanced strength of two branches connecting to the BL and BLB is altered and the required voltage to compensate the imbalance is measured as V_{OS}. For correct sensing operation, the minimum bit-line differential voltage should be larger than the maximum value of V_{OS} to compensate variations.

To assess the impacts of RTN on TFET CLSA and VLSA, the possible trapping/detrapping combinations are binary-coded for the most critical TFETs of the sense amplifiers as labeled in Figs. 4.21a and 4.22a [33]. Among the 32 V_{OS}

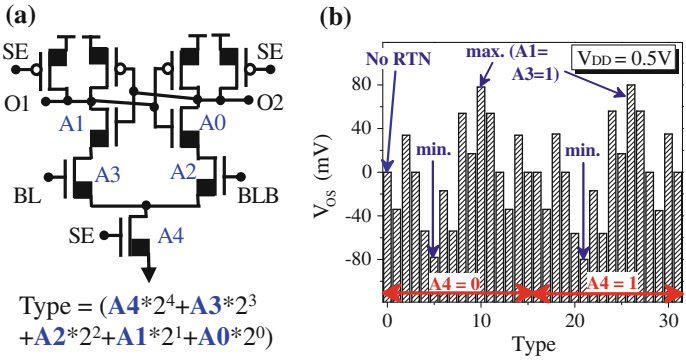

Fig. 4.21 **a** Definition of trapping/detrapping combinations for the RTN analysis of TFET current-latched sense amplifier and **b** the 32 values of V_{OS} with single trap at the worst position (tunneling junction) [33]

combinations coming from the five selected TFETs that are vulnerable to RTN (Fig. 4.21b), the trapping/detrapping in the devices connected to BL/BLB (A2 and A3) exhibits the highest impact on the robustness of CLSA. The maximum value of V_{OS} occurs in the RTN configuration with A1 and A3 devices both in trapped state. For VLSA, the RTN in pull-up (A0/A1) and A4 transistors is shown to have negligible impact on V_{OS}. It is shown in Fig. 4.22b that the existence of trapping/detrapping in pull-down devices (A2 or A3) yields the maximum V_{OS} fluctuation. Due to the significant changes in drain current, TFET-based CLSA requires larger V_{OS} to mitigate the impact of RTN and becomes inferior to VLSA. The single-ended large-signal inverter sense amplifier is also evaluated, and the results are shown together with that for CLSA and VLSA in Fig. 4.23. Compared with VLSA and inverter sense amplifier, CLSA is more susceptible to RTN.

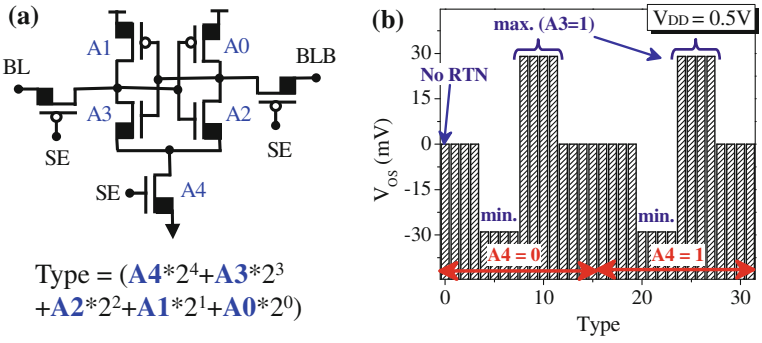

Fig. 4.22 **a** Definition of possible RTN combinations for TFET voltage-latched sense amplifier with **b** the 32 combinations of V_{OS} for single trap around the tunneling junction [33]

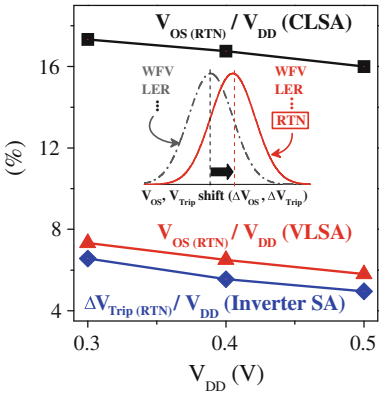

Fig. 4.23 Impact of RTN on differential (CLSA and VLSA) and single-ended inverter sense amplifier at various V_{DD} [33]

4.5 Summary

This chapter illustrates the challenges of TFET and the possible solutions for logic and SRAM applications. Under the constraint of unidirectional conduction, adequate design for the orientation of source/drain of pass-gate-based logic circuits and separation of READ/WRITE paths are demonstrated to improve functionality. The undesired delayed saturation can be mitigated with the hybrid combination of TFET and MOSFET. Furthermore, device engineering such as dual oxide, drain-side underlap, and dual-metal-gate work function is shown to alleviate the performance loss of TFET in the presence of enhanced Miller capacitance. Other design concerns (e.g., unbalanced complementary TFET design and variability) unique to TFET need to be thoroughly understood and addressed. In summary, early-stage analysis on the device–circuit interactions for TFET digital applications is important and should be conducted in parallel with device optimizations to enable TFET as a viable technology for future IC industry.

Acknowledgement This work was supported in part by the Ministry of Science and Technology in Taiwan under Contracts MOST 104-2911-I-009-301 (I-RiCE), MOST 102-2221-E-009-136-MY2, and MOST 103-2221-E-009-196-MY2, and by the Ministry of Education in Taiwan under the ATU Program. The authors are grateful to the National Center for High-Performance Computing in Taiwan for the software and facilities.

References

1. D.H. Morris, U.E. Avci, R. Rios, I.A. Young, Design of Low Voltage Tunneling-FET logic circuits considering asymmetric conduction characteristics. IEEE J. Emerg. Sel. Top. Circ. Syst. **4**(4), 380–388 (2014)
2. H. Liu, S. Datta, V. Narayanan, Steep switching tunnel FET: a promise to extend the energy efficient roadmap for Post-CMOS Digital and Analog/RF applications," in *Proceedings of ISLPED* (2013), pp. 145–150

3. D. Kim, Y. Lee, J. Cai, I. Lauer, L. Chang, S.J. Koeter, D. Sylvester, D. Blaauw, Low Power Circuit Design Based on Heterojunction Tunneling Transistors (HETTs), in *Proceedings of ISLPED* (2009), pp. 219–224
4. Y.-N. Chen, M.-L. Fan, V.P.-H. Hu, P. Su, C.-T. Chuang, Evaluation of stability, performance of ultra-low voltage MOSFET, TFET, and mixed TFET-MOSFET SRAM cell with write-assist circuits. IEEE J. Emerg. Sel. Top. Circ. Syst. **4**(4), 389–399 (2014)
5. V. Saripalli, S. Datta, V. Narayanan, J.P. Kulkarni, Variation-tolerant ultra low-power heterojunction tunnel FET SRAM design, in *Proceedings of IEEE Nano. Arch.* (2011), pp. 45–52
6. A. Mallik, A. Chattopadhyay, Drain-dependence of tunnel field-effect transistor characteristics: the role of the channel. IEEE Trans. Electron. Device **58**(12), 4250–4257 (2011)
7. L.D. Michielis, L. Lattanzio, A.M. Ionescu, Understanding the superlinear onset of Tunnel-FET output characteristics. IEEE Electron. Device Lett. **33**(11), 1523–1525 (2012)
8. L. Knoll, Q.-T. Zhao, A. Nichau, S. Trellenkamp, S. Richter, A. Schafer, D. Esseni, L. Selmi, K.K. Boudelle, S. Mantl, Inverters with strained Si nanowire complementary tunnel field-effect transistors. IEEE Electron. Device Lett. **34**(6), 813–815 (2013)
9. D. Esseni, M. Guglielmini, B. Kapidani, T. Rollo, M. Alioto, Tunnel FETs for ultralow voltage digital VLSI circuits: Part I—device-circuit interaction and evaluation at device level. IEEE Trans. VLSI Syst. **22**(12), 2488–2498 (2014)
10. Y.-N. Chen, M.-L. Fan, V.P.-H. Hu, P. Su, C.-T. Chuang, Design and analysis of robust tunneling FET SRAM. IEEE Trans. Electron. Devices **60**(3), 1092–1098 (2013)
11. S. Mookerjea, R. Krishnan, S. Datta, V. Narayanan, Effective capacitance and drive current for tunnel FET (TFET) CV/I estimation. IEEE Trans Electron. Devices **56**(9), 2092–2098 (2009)
12. Y.-N. Chen, M.-L. Fan, V.P.-H. Hu, P. Su, C.-T. Chuang, Investigation of tunneling FET device designs for improving circuit switching performance and energy, Extended Abstracts of Solid State Devices and Materials (SSDM), (2013), pp. 84–85
13. Y.-N. Chen, M.-L. Fan, V.P.-H. Hu, P. Su, C.-T. Chuang, Evaluation of Sub-0.2 V high-speed low-power circuits using hetero-channel MOSFET and tunneling FET Devices, IEEE Trans Circ. Syst–I **61**(12), 3339–3347
14. U.E. Avci, R. Rios, K.J. Kuln, I.A. Young, Comparison of power and performance for the TFET and MOSFET and considerations for P-TFET, in *Proceedings of IEEE International Conference on Nanotechnology* (2011), pp. 869–872
15. R. Mukundrajan, M. Cotter, V. Saripalli, M.J. Irwin, S. Datta, V. Narayanan, Ultra Low Power Circuit Design Using Tunnel FETs," in *Proceedings of ISVLSI* (2012), pp. 153–158
16. M.-L. Fan, V.P.-H. Hu, Y.-N. Chen, C.-W. Hsu, P. Su, C.-T. Chuang, Investigation of backgate-biasing effect for ultra-thin-body III-V heterojunction tunnel FET. IEEE Trans. Electron. Devices **62**(1), 107–113 (2015)
17. E. Baravelli, E. Gnani, A. Gnudi, S. Reggiani, G. Baccarani, TFET inverters with n-/p-devices on the Same technology platform for low-voltage/low-power applications. IEEE Trans. Electron. Devices **61**(2), 473–478 (2014)
18. W.G. Vandenberghe, A.S. Verhulst, K.-H. Kao, K.D. Meyer, B. Soree, W. Magnus, G. Groeseneken, A model determining optimal doping concentration and material's band gap of tunnel field-effect transistors. Appl. Phys. Lett. **100**(19), 193–196 (2012)
19. M.-L. Fan, V.P.-H. Hu, Y.-N. Chen, P. Su, C.-T. Chuang, Analysis of single-trap-induced random telegraph noise and its interaction with work function variation for tunnel FET. IEEE Trans. Electron. Devices **60**(6), 2038–2044 (2013)
20. U.E. Avci, R. Rios, K. Kuhn, I.A. Young, Comparison of performance, switching energy and process variations for the TFET and MOSFET in Logic, in *Proceedings of VLSI Symposium Tech. Dig.* (2011), pp. 124–125
21. C.-W. Hsu, M.-L. Fan, V.P.-H. Hu, P. Su, Investigation and simulation of work-function variation for III–V broken-gap heterojunction tunnel FET. IEEE J. Electron. Device Soc. **3**(3), 194–199 (2015)
22. A. Guo, P. Matheu, T.-J.K. Liu, SOI TFET I_{ON}/I_{OFF} enhancement via back biasing. IEEE Trans. Electron. Device **58**(10), 3283–3285 (2011)

23. P. Matheu, B. Ho, Z.A. Jacobson, T.-J.K. Liu, Planar GeOI TFET performance improvement with back biasing. IEEE Trans. Electron. Device **59**(6), 1629–1635 (2012)
24. S.H. Kim, Z.A. Jacobson, P. Patel, C. Hu, T.-J.K. Liu, Tunnel FET-based pass-transistor logic for ultra-low-power applications, in *Proceedings of Device Research Conference (DRC)* (2011), pp. 133–134
25. S. Jain, S. Khare, S. Yada, V. Ambili, P. Salihundam, S. Ramani, S. Muthukumar, M. Srinivasan, A. Kumar, S.K. Gb, R. Ramanarayanan, V. Erraguntla, J. Howard, S. Vangal, S. Dighe, G. Ruhl, P. Aseron, H. Wilson, N. Borkar, V. De, S. Borkar, A 280mV-To-1.2V wide-operating-range IA-32 processor in 32 nm CMOS, in *Proceedings of ISSCC Dig. Tech.* (2012), pp. 66–67
26. L. Chang, D.M. Fried, J. Hergenrother, J.W. Sleight, R.H. Dennard, R.K. Montoye, L. Sekaric, S.J. McNab, A.W. Topol, C.D. Adams, K.W. Guarini, W. Haensch, Stable SRAM cell design for the 32 nm node and beyond, in *Proceedings of VLSI Symposium Tech. Dig.* (2005), pp. 128–129
27. N. Verma, A.P. Chandrakasan, A 256 Kb 65 nm 8T subthreshold SRAM employing sense-amplifier redundancy. IEEE J. Solid-State Circ. **43**(1), 141–149 (2008)
28. M.-H. Tu, J.-Y. Lin, M.-C. Tsai, S.-J. Chou, C.-T. Chuang, Single-ended subthreshold SRAM with asymmetrical write/read assist. IEEE Trans. Circ. Syst I, Regul. Pap. **57**(12), 3039–3047 (2010)
29. I.J. Chang, J.J. Kim, S.P. Park, K. Roy, A 32 Kb 10T Subthreshold SRAM array with bit-interleaving and differential read scheme in 90 nm CMOS. IEEE J. Solid-State Circ. **44**(2), 650–658 (2009)
30. M. Yabuuchi, K. Nii, Y. Tsukamoto, S. Ohbayashi, Y. Nakase, H. Shinohara, A 45nm 0.6V cross-point 8T SRAM with negative biased read/write assist, *in Proceedings of Symposium VLSI circuits dig.* (2009), pp. 158–159
31. T. Kobayashi, K. Nogami, T. Shirotori, Y. Fujimoto, A current-controlled latch sense amplifier and a static power-saving input buffer for low-power architecture. IEEE J. Solid-State Circuits **28**(4), 523–527 (1993)
32. B. Wicht, T. Nirschl, D. Schmitt-Landsiedel, Yield and speed optimization of a latch-type voltage sense amplifier. IEEE J. Solid-State Circ. **39**(7), 1148–1158 (2004)
33. M.-L. Fan, V.P.-H. Hu, Y.-N. Chen, P. Su, C.-T. Chuang, Investigation of single-trap-induced random telegraph noise for tunneling FET devices, 8T SRAM cell, and sense amplifiers, in *Proceedings of IEEE International Reliability Physics Symposium (IRPS)* (2013), pp. CR.1.1–CR.1.6

Chapter 5
Atomistic Simulations of Tunneling FETs

Fei Liu, Qing Shi, Jian Wang and Hong Guo

Abstract With continuous scaling of semiconductor devices, the number of atoms in transistors becomes countable. Various effects related to the device atomic structure, such as random dopants, edge roughness, and channel-oxide interface, have great impact on device performance. Therefore, it is valuable to study material electronic properties and device transport characteristics at the atomic level. In this chapter, we review the atomistic modeling methods of density functional theory (DFT) and tight-binding (TB) model within the Keldysh non-equilibrium Green's function (NEGF) framework. To investigate impurity scattering in devices, the framework of non-equilibrium vertex correction (NVC) with NEGF–DFT is reviewed. The NEGF–DFT–NVC approach can give the statistic transport information of nanodevices with atomic disorder and is applied to study disorder effects in graphene TFETs. Due to the diffusive impurity scattering, the band-to-band tunneling current is substantially reduced in graphene TFETs with atomic disorder. At last, atomistic simulations of monolayer transition metal dichalcogenide (TMDC) TFETs are carried out by using the NEGF–TB method. It is revealed that the orientation-dependent transport is determined by conduction sub-bands and the atomic structure along the transport direction.

5.1 Introduction

With the continuation of device scaling, the gate length gets smaller than typical magnitude of the de Broglie wavelength of electrons and carrier phase coherence must be taken into account in transport simulations. Traditional semi-classical

Fei Liu and Qing Shi contributed equally.

F. Liu (✉) · J. Wang
Department of Physics, The University of Hong Kong, Pok Fu Lam, Hong Kong, China
e-mail: fliu003@gmail.com

Q. Shi · H. Guo
Department of Physics, McGill University, Montreal, QC H3A 2T8, Canada

L. Zhang and M. Chan (eds.), *Tunneling Field Effect Transistor Technology*,
DOI 10.1007/978-3-319-31653-6_5

transport theories derived from the Boltzmann transport equation (BTE) treat electrons and holes as semiclassical particles [1], and therefore are no longer capable of describing quantum transport at nanoscale. At the same time, quantum effects such as quantum confinement, quantum tunneling, and quantum interference are not negligible for device performance. Furthermore, the number of channel's atoms becomes countable at sub-10-nm technology nodes. Statistical variability related to atomic structure has become one of the major sources of degradation of the circuit and system performance [2, 3] and increases the power dissipation in contemporary chips, such as random discrete doping (RDD) and line edge roughness (LER). So, atomistic and quantum simulations are necessary to study carrier transport at nanoscale.

The NEGF formalism is based on rigorous quantum mechanism and provides a general approach to study quantum transport of nanoelectronic devices [4]. In the early 1960s, the method is first developed by Martin and Schwinger [5, 6] to solve various non-equilibrium problems in statistics. Then, Keldysh [7], Kadanoff and Baym [8] further developed the NEGF method to deal with non-equilibrium transport. Caroli et al. [9] first got the general formalism of current in terms of Green's functions in the early 1970s, whose modern form is obtained by Meir and Wingreen [10]. Nowadays, the theory has been adapted to address quantum transport in nanodevices such as molecular devices [11, 12], Si FETs [13, 14], 1D nanotube [15, 16] and nanowire FETs [17, 18], and FETs using 2D materials [19–21].

Tunneling FET (TFET) has emerged as a promising candidate to reduce the power dissipation in MOSFETs by reaching subthreshold swing (SS) smaller than 60 mV/decade at room temperature [22, 23]. Tunneling process in TFETs is a quantum phenomenon and can naturally be described by NEGF. Quantum transport of various TFETs has been extensively studied by the NEGF method. At nanoscale, kinds of atomistic effects have an important influence on device characteristics of TFETs including atomic defect, impurity, and channel–substrate interface. Thus, it is significative to atomically simulate device performance of TFETs. Atomistic simulations have been carried out in investigation of gate-all-around (GAA) nanowire TFETs [24, 25], graphene nanoribbon TFETs [26, 27], and transition metal dichalcogenide (TMDC) TFETs [28–31].

In this chapter, we start with a brief introduction about calculations of electronic properties at the atomic level. DFT is the most widely applied computational approach to study material's ground state properties. We review the basic theorems of DFT and then present the linear muffin-tin orbital (LMTO) method for solving the Kohn–Sham equation. Besides DFT, we also review the tight-binding (TB) approach which greatly reduces computation cost in comparison with DFT and is suitable for large-scale atomistic simulations. In the second part, we briefly review the quantum transport theory of NEGF and then review the theory of NEGF–DFT which combines the NEGF quantum transport theory with the DFT theory. So, the non-equilibrium quantum transport in open devices can be calculated from atomic first principles. Then, we present the coherent potential approximation (CPA) and the non-equilibrium vertex correction (NVC) theory within the framework of NEGF–DFT for calculating atomic disorder scattering in

nanodevices. In the last part, we present two works about atomistic simulations of TFETs. NEGF–DFT–CPA method is applied to study effects of disorder scattering on the quantum transport properties of a boron–nitrogen (B–N) co-doped graphene TFETs. In the other work, device physics in TFETs using two-dimensional monolayer TMDCs is investigated by atomistic simulations using the TB model.

5.2 Electronic Structure Theory

5.2.1 Density Functional Theory

5.2.1.1 The Born–Oppenheimer Approximation

Solid-state materials are formed by lots of bonded atoms together. Crystal structure is determined by quasi-fixed atomic nuclei while material electronic and optical properties are determined by free electron outside nucleus. In quantum mechanism, we know that all the information of a quantum system can be obtained by the wave function. The wave function in a solid material is obtained by solving Schödinger equation in nonrelativistic quantum theory:

$$H\Psi(r,R) = E\Psi(r,R) \tag{5.1}$$

where r represents the position of electron, and R is the position of nucleus. In a material, there are lots of electrons and nuclei. Therefore, the system Hamiltonian of solid-state material includes the kinetic energy of electrons and nuclei, interaction energy among these particles and potential under external field:

$$H = H_e + H_N + H_{e-H} + V_{ext} \tag{5.2}$$

where

$$H_e(r) = T_e(r) + U_e(r) = -\sum_i \frac{\hbar^2}{2m_e}\nabla_{r_i}^2 + \frac{1}{2}\sum_{i,j}\frac{e^2}{|r_i - r_j|} \tag{5.3}$$

$$H_N(r) = T_N(R) + U_N(R) = -\sum_i \frac{\hbar^2}{2M}\nabla_{R_i}^2 + \frac{1}{2}\sum_{i,j}\frac{Z_i Z_j e^2}{|R_i - R_j|} \tag{5.4}$$

$$H_{e-N} = -\sum_{i,j} U_{e-N}(r_i - R_j) = -\sum_{i,j}\frac{Z_j e^2}{|r_i - R_j|} \tag{5.5}$$

The first terms in Eqs. (5.3) and (5.4) represent kinetic energies of electrons and nuclei, respectively; the second terms represent the interaction energies among electrons and nuclei. Equation (5.5) describes the interaction between electrons and

nuclei; Z_j and R_j are the charge and the position of the j-th nucleus, respectively. Because there are lots of electrons and nuclei, we have to solve Schödinger equation with many variables. Direct solution of the system with so many particles is not possible and necessary approximation should be made. We know that the nucleus mass is much heavier than electron mass, so it is plausible to assume nucleus to vibrate at the equilibrium position. The separation of electronic motion and nuclear motion is known as the Born–Oppenheimer approximation [32]. In the approximation, the motion of electron is separated from nucleus motion and the Schödinger equation of many electrons can be obtained:

$$(T_e + U_{e-e} + U_{e-N} + V_{ext})\psi(r) = E\psi(r) \tag{5.6}$$

where T_e is the kinetic energy, U_{e-e} is the interaction energy between electrons, and U_{e-N} is the interaction energy between nucleus and electron. With the Born–Oppenheimer approximation, the Schödinger equation is greatly simplified; however, Eq. (5.6) is still a many-body equation of lots of electrons and cannot be solved directly.

5.2.1.2 The Hohenberg–Kohn Theorems and Kohn–Sham Equations

The Hohenberg–Kohn theorems are the theoretical foundation of DFT, which are obtained from Tomas-Fermi model [33]. The first Hohenberg–Kohn theorem states that the ground state electron density determines the ground state properties of an interacting electron system. The ground state wave function is the functional of the ground state electron density and all properties of the system in ground state are functionals of the ground state electron density. The second Hohenberg–Kohn theorem states that with given external potential, the minimal value of the total energy functional is obtained at the ground state electron density, and the energy is the ground state energy of the system.

The Hohenberg–Kohn theorems are rigorously based on quantum mechanism but do not show the specific way to calculate the ground state electron density. Kohn and Sham proposed to use free electron rather than the many-body interacting system [34]. The Kohn–Sham method avoids the difficulty of direct solving the Schödinger equation with many-body interactions and applies an approximation of solving an effective equation of single electron. The interaction is put in the exchange correlation functionals in the effective equation of single electron.

The total energy of an interacting many-body system is composed of three parts:

$$E_{tot} = T(\rho) + V(\rho) + U(\rho) \tag{5.7}$$

where ρ is the system electron density, and T, V, and U are the kinetic energy, potential at external field, and the electron interacting energy, respectively. In the Kohn–Sham method, the non-interacting kinetic energy functional $T_s(\rho)$ substitutes

the interacting kinetic energy functional $T(\rho)$. The $T_s(\rho)$ can be achieved from the non-interacting wave function:

$$T_s(\rho) = -\frac{\hbar^2}{2m_e}\sum_i \int d^3 r \phi_i^*(r)\nabla^2\phi_i(r) \tag{5.8}$$

where the total kinetic energy is the sum of kinetic energy of single free electron. From Eq. (5.8), T_s is the functional of free energy wave function which is also the functional of the system electron density; therefore, the total kinetic energy of free electrons is the functional of the system electron density: $T_s[\phi(\rho)]$. We can describe the system total energy by the non-interacting kinetic energy:

$$E(\rho) = T(\rho) + V(\rho) + U(\rho) = T_s[\phi(\rho)] + U_H(\rho) + E_{xc}(\rho) + V(\rho) \tag{5.9}$$

where U_H is the Hatree potential, $E_{xc}(\rho)$ represents the effective exchange correlation which has two parts: the difference between the interacting kinetic energy and non-interacting kinetic energy $T - T_s$; the difference between the interacting energy and Hatree potential $U - U_H$. After variational computing of free electron wave function, one can obtain:

$$\frac{\delta E}{\delta\phi_i^*} = \frac{\delta T_s}{\delta\phi_i^*} + \left[\frac{\delta U_H}{\delta n(r)} + \frac{\delta V}{\delta n(r)} + \frac{\delta E_{xc}}{\delta n(r)}\right]\frac{\delta n(r)}{\delta\phi_i^*} = 0 \tag{5.10}$$

where the wave function is orthogonal:

$$\langle\phi_i|\phi_j\rangle = \delta_{i,j} \tag{5.11}$$

with Eq. (5.11), Lagrange multiplier is induced to find the conditional extreme value:

$$\frac{\delta\left\{E - \sum_i E_i[\langle\phi_i|\phi_i\rangle - 1]\right\}}{\delta\phi_i^*} = 0 \tag{5.12}$$

E_i is the Lagrange multiplier, with the following equation:

$$\frac{\delta T_s}{\delta\phi_i^*} = -\frac{\hbar^2}{2m_e}\nabla^2\phi_i(r); \frac{\delta n(r)}{\delta\phi(r)^*} = \phi_i(r) \tag{5.13}$$

The Schödinger equation of free electron is obtained finally as:

$$\left(-\frac{\hbar^2}{2m_e}\nabla^2 + V_{KS}(r)\right)\phi_i(r) = E_i\phi_i(r) \tag{5.14}$$

where

$$V_{KS} = \frac{\delta U_H}{\delta n(r)} + \frac{\delta V}{\delta n(r)} + \frac{\delta E_{xc}}{\delta n(r)} \tag{5.15}$$

with the wave function from Eq. (5.14), the electron density is obtained:

$$n(r) = \sum_i f_i |\phi_i(r)|^2 \tag{5.16}$$

where f_i is the Fermi–Dirac distribution function of electron. Equations (5.14), (5.15), and (5.16) are Kohn–Sham equations. Hatree energy and exchange correlation energy depend on electron density function, which should be obtained from electron wave function. However, electron wave function is determined by V_{KS}. Hence, Kohn–Sham equations are nonlinear equations and should be solved self-consistently. From the initial guessed electron density function ρ_0, V_{KS} and $\phi_i(r)$ are calculated. Then, using the wave function the new electron density $\rho(r)$ is constructed for the next iteration. The computational process of Kohn–Sham equations is repeated until numerical convergence is reached. As long as electron density ρ is obtained, one can get the system total energy:

$$E_0 = \sum_i \varepsilon_i - \frac{e^2}{2} \int d^3r \int d^3r' \frac{n_0(r) n_0(r')}{|r - r'|} - \int d^3r V_{xc}(r) n_0(r) + E_{xc}[n_0] \tag{5.17}$$

5.2.1.3 Basis Set and Exchange Correlation Functional

In DFT calculations, there are two important choices: the basis set and the exchange correlation functional. Various wave function basis can be chosen for calculating material electronic structure such as the plane-wave basis sets, linear combination of atomic orbital (LCAO), linear combinations of Gaussian orbitals (LCGO), the linear augmented plane wave (LAPW), muffin-tin orbital (MTO). Tight-binding linear Muffin-Tin orbital (TB-LMTO) method is applied in transport calculations in this chapter. Even though the exchange correlation energy E_{xc} is the functional of electron density, there is no specific formalism of E_{xc}. A simple approximation is to construct the exchange correlation energy functional by using the local electronic density, which is called local density approximation (LDA) [35]. Based on LDA, the exchange correlation energy functional can be further improved by including the gradient of the electronic density, which is known as generalized gradient approximations (GGA) [36]. These approximations have been proved very successful in studying electronic properties of matter. Even though band gap of semiconductor is usually underestimated by LDA or GGA, DFT calculations using these exchange correlation functionals are computationally cheap. In particular for a open-device system with lots of atoms, LDA and GGA are efficient without

sacrifice of accuracy. In the following NEGF–DFT calculations, LDA is applied to study disorder effects in graphene TFETs.

5.2.2 *Tight-Binding Linear Muffin-Tin Orbital Method*

5.2.2.1 Atomic Sphere Approximation in Linear Muffin-Tin Orbitals

The technique used for graphene TFET simulations is built on the MTO technique. MTO is one of the oldest and most commonly used methods in electronic band theory. The reason that we use this basis is that the Hamiltonian under such basis can be very parse which make it ultra-efficient for large system computation. Besides, this basis is compatible with the CPA theory, which will be introduced later. In this theory, the space is divided into a series of non-overlapping spherical muffin-tin spheres (Wigner-Seitz cells), in each of which the electronic potential is assumed to be spherically symmetric and the interstitial of which to be a constant. In this work, we employ a more trivial but more computationally efficient form of this theory, i.e., the atomic sphere approximation (ASA) in MTO. We basically increase the radius of each of the spheres in MTO so that they are slightly overlapped in space. In interstitial region, we introduce additional empty spheres which together with the original atomic spheres make the lattice structure close-packed. Figure 5.1 shows the ASA filling of the silicon bulk lattice. The spheres in yellow are the empty spheres we introduced to make the silicon FCC lattice close-packed. The spirit of MTO–ASA is: (i) the use of spherically symmetric potential inside slightly overlapping, space-filling atomic Wigner–Seitz spheres centered at each individual site, and (ii) a complete neglect of the electronic kinetic energy $(E - V_0)$ in the volumeless interstitial region. Thus, when performing the integral in space, we just simply sum over the quantities within each of the spheres and neglecting those in the interstitial region. The potential here, like that in DFT, consists of

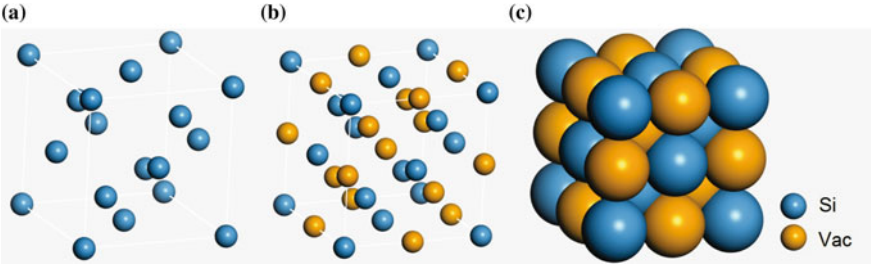

Fig. 5.1 Unit cell of silicon crystal. **a** Shows the basic structure of the unit cell of crystal silicon. **b** Shows the positions of vacuum spheres (in *yellow*) in the close-packed lattice. **c** Is the same with (**b**) but with exact sphere radius. These silicon and vacuum spheres together from a close-packed structure

Hartree potential, nuclei ion attractions, and exchange correlation potential produced by many-body effects. Then, the single electron Schrödinger equation has the form

$$\left[-\frac{\hbar^2}{2m_e}\nabla^2 + V(\mathbf{r}) - E\right]\psi(\mathbf{r}) = 0, \quad \mathbf{r} \in A \tag{5.18}$$

$$\nabla^2\psi(\mathbf{r}) = 0, \quad \mathbf{r} \in I \tag{5.19}$$

where A refers to the region inside muffin-tin spheres while I outside. Obviously, the interstitial part is exactly the Laplace equation and has two linearly independent solutions

$$\begin{aligned} J_L(\mathbf{r}) &= J_l(r)Y_L(\hat{\mathbf{r}}), \quad J_l(r) = \frac{1}{2(2l+1)}\left(\frac{r}{\omega}\right)^l, \\ K_L(\mathbf{r}) &= K_l(r)Y_L(\hat{\mathbf{r}}), \quad K_l(r) = \left(\frac{\omega}{r}\right)^{l+1}, \end{aligned} \tag{5.20}$$

where $L = (l, m)$ is the usual angular momentum index composed of orbital quantum number l and the magnetic quantum number is m; $\hat{\mathbf{r}}(= \mathbf{r}/r)$ is the unit vector parallel to $\mathbf{r}$; ω is the averaged Wigner–Seitz radius of the system; and $Y_L(\hat{\mathbf{r}})$ refers to spherical harmonics. For the part inside muffin-tin spheres of Eq. (5.18), as required by MTO, $V(\mathbf{r})$ is assumed to be spherically symmetric. The solution for a given energy E can be decomposed into a radial-dependent part and an angular-dependent part

$$\varphi_{\mathbf{R}L}(\mathbf{r}, E) = \varphi_{\mathbf{R}l}(r, E)Y_L(\hat{\mathbf{r}}), \tag{5.21}$$

where subscript $\mathbf{R}$ denotes the position of sphere, in which we are solving the Schödinger equation.

Now we have already obtained wave function $\varphi_{\mathbf{R}l}(r, E)$ for the region inside the spheres and $J_l(r)$, $K_l(r)$ for outside; one straightforward constraint condition is that they need to be matched at the boundaries (sphere surface). These boundary conditions lead to the traditional MTO's which are energy-dependent [37]. Once we have constructed the basis, we can expand the quantities needed for secular equation in variational theory under such basis and yield the eigenvalues of the system.

The variational principle procedure is used to build the secular equation. It can be briefly described as follows:

$$\begin{aligned} \delta \int \psi(\mathbf{r})\left[-\frac{\hbar^2}{2m_e}\nabla^2 + V(\mathbf{r})\right]\psi^*(\mathbf{r})\mathrm{d}^3\mathbf{r} = 0, \\ \int \psi(\mathbf{r})\psi^*(\mathbf{r})\mathrm{d}^3\mathbf{r} = 1, \end{aligned} \tag{5.22}$$

which is equivalent to solving the famous Schrödinger equation. Due to the second constraint in Eq. (5.22), the energy E enters the variational approach as a Lagrange multiplier. If we assume that $\psi(\mathbf{r})$ can be written as a linear combination of basis function $(\chi_i\mathbf{r})$, the variational principle leads to the secular equation

$$\det(EO_{ij} - H_{ij}) = 0, \tag{5.23}$$

where O_{ij} and H_{ij} are the overlap and Hamiltonian matrices with the expressions

$$\begin{aligned} H_{ij} &= \int \chi_i(\mathbf{r}) \left[-\frac{\hbar^2}{2m_e} \nabla^2 + V(\mathbf{r}) \right] \chi_j^*(\mathbf{r}) \mathrm{d}^3\mathbf{r}, \\ O_{ij} &= \int \chi_i(\mathbf{r}) \chi_j^*(\mathbf{r}) \mathrm{d}^3\mathbf{r}. \end{aligned} \tag{5.24}$$

One thing needed to specify is that the integral in Eq. (5.24) is performed over the whole space. In MTO technique, the orbital is chosen to be energy-dependent [37], and thus, the secular equation Eq. (5.23) becomes nonlinear with energy so that it cannot be reduced to a simple matrix eigenvalue problem.

Fortunately, this drawback can be removed by employing the linearized muffin-tin orbital (LMTO). In LMTO, the wave function inside spheres can be Taylor expanded around some energy $E_{v,\mathbf{R}l}$

$$\varphi_{\mathbf{R}l}(r, E) = \phi_{\mathbf{R}l}(r) + \dot{\phi}_{\mathbf{R}l}(r)(E - E_{v,\mathbf{R}l}) + \ldots, \tag{5.25}$$

where notation is introduced as

$$\phi_{\mathbf{R}l}(r) = \varphi_{\mathbf{R}l}(r, E_{v,\mathbf{R}l})$$

$$\dot{\phi}_{\mathbf{R}l}(r) = \dot{\varphi}_{\mathbf{R}l}(r, E_{v,\mathbf{R}l})$$

and the "dot" on top means derivative over energy. $E_{v,\mathbf{R}l}$ can be chosen arbitrarily and is usually taken in the center of the occupied part of the $\mathbf{R}l$-projected valence density of states [37]. In our simulation, only the first two terms in Eq. (5.25) were kept for the sake of simplicity. It will be showed that such truncation is accurate enough in our problems. It is also easy to show that $\phi_{\mathbf{R}l}(r)$ and $\dot{\phi}_{\mathbf{R}l}(r)$ are orthogonal to each other [37]. In other words, the wave function inside atomic spheres can be written as a linear combination of $\phi_{\mathbf{R}l}(r)$ and $\dot{\phi}_{\mathbf{R}l}(r)$. Recall that wave function outside can be written as linear combination of $K_l(r)$ and $J_l(r)$. Then, one last thing remaining for these orbitals is to smoothly match them up at the boundaries. A convenient way to smoothly match these functions at the sphere surfaces is using the Wronskian function [37]. If a function $f(r)$ is to be matched continuously, including the function value and the first derivative, at $r = r_0$ to a linear combination of two other functions $f_1(r)$ and $f_2(r)$, the matching condition has the form

$$f(r) \to \frac{\{f,f_2\}f_1(r) - \{f,f_1\}f_2(r)}{\{f_1,f_2\}} \tag{5.26}$$

where $\{f_1(r),f_2(r)\} = r^2[f_1(r)f_2'(r) - f_1'(r)f_2(r)]|_{r=r_0}$. Then if matching $K_l(r)$ and $J_l(r)$ with the rest two separately, we have:

$$\begin{aligned} K_l(r) &\to -\{K,\dot{\phi}\}_{\mathbf{R}l}\phi_{\mathbf{R}l}(r) + \{K,\phi\}_{\mathbf{R}l}\dot{\phi}_{\mathbf{R}l}(r), \\ J_l(r) &\to -\{J,\dot{\phi}\}_{\mathbf{R}l}\phi_{\mathbf{R}l}(r) + \{J,\phi\}_{\mathbf{R}l}\dot{\phi}_{\mathbf{R}l}(r), \end{aligned} \tag{5.27}$$

where $\{\ldots\}$ means estimating the Wronskian at the surface of the $\mathbf{R}l$th muffin-tin sphere. This provides us a way to construct the LMTO's (labeled as $\chi_{\mathbf{R}L}(\mathbf{r})$) for our problem. For interstitial region, we simply use the $K_l(r)$ since it decays at large r. For region of $r_{\mathbf{R}} < s_{\mathbf{R}}$, $\phi_{\mathbf{R}l}(r)$ and $\dot{\phi}_{\mathbf{R}l}(r)$ need to continuously match $K_l(r)$ at $r = s_{\mathbf{R}}$, then by using the matching condition of Eq. (5.27), $\chi_{\mathbf{R}L}(\mathbf{r})$ should be

$$\chi_{\mathbf{R}L}(\mathbf{r}) = -\{K,\dot{\phi}\}_{\mathbf{R}l}\phi_{\mathbf{R}l}(\mathbf{r}_{\mathbf{R}}) + \{K,\phi\}_{\mathbf{R}l}\dot{\phi}_{\mathbf{R}l}(\mathbf{r}_{\mathbf{R}}).$$

For region $r_{\mathbf{R}'} < s_{\mathbf{R}'}(\mathbf{R}' \neq \mathbf{R})$, we need to move the starting point of $r_{\mathbf{R}}$ from $\mathbf{R}$ to $\mathbf{R}'$. The relation

$$K_L(\mathbf{r}_{\mathbf{R}}) = -\sum_{L'} S_{\mathbf{R}L,\mathbf{R}'L'}J_{L'}(\mathbf{r}_{\mathbf{R}'})$$

helps us to achieve this. $S_{\mathbf{R}L,\mathbf{R}'L'}$ is the canonical structure constant and is defined as

$$\begin{aligned} S_{\mathbf{R}'L',\mathbf{R}''L''} = \sum_{L}(-1)^{l''+1}\frac{8\pi(2l-1)!!C_{LL'L''}}{(2l'-1)!!(2l''-1)!!} \\ K_L(\mathbf{R}''-\mathbf{R}'), \end{aligned} \tag{5.28}$$

where $C_{LL'L''}$ are the Gaunt coefficients [37]. Then by using the matching condition of Eq. (5.27), together with previous expressions for $\chi_{\mathbf{R}L}(\mathbf{r})$, the final LMTO has the form

$$\begin{aligned} \chi_{\mathbf{R}L}(r) &= -\{K,\dot{\phi}\}_{\mathbf{R}l}\phi_{\mathbf{R}l}(\mathbf{r}_{\mathbf{R}}) \\ &\quad + \{K,\phi\}_{\mathbf{R}l}\dot{\phi}_{\mathbf{R}l}(\mathbf{r}_{\mathbf{R}}), \quad r_{\mathbf{R}} \le s_{\mathbf{R}}, \\ &= \sum_{L'} S_{\mathbf{R}L,\mathbf{R}'L'}[\{J,\dot{\phi}\}_{\mathbf{R}'l'}\phi_{\mathbf{R}'L'}(\mathbf{r}_{\mathbf{R}'}) \\ &\quad - \{J,\phi\}_{\mathbf{R}'l'}\dot{\phi}_{\mathbf{R}'L'}(\mathbf{r}_{\mathbf{R}'})], \quad \mathbf{r}_{\mathbf{R}'} \le s_{\mathbf{R}'}, \\ &= K_L(\mathbf{r}_{\mathbf{R}}), \quad \mathbf{r} \in I. \end{aligned} \tag{5.29}$$

5.2.2.2 Hamiltonian and Overlap Matrix in LMTO-ASA

Once we defined the basis, we need to find the expression for some useful physical quantities such as Hamiltonian, overlap matrix, and Green's functions next. We insert these LMTO's into Eq. (5.24) and neglect unimportant high-order terms, and we obtained the expressions for matrix elements $H_{\mathbf{R}L,\mathbf{R}'L'}$ and $O_{\mathbf{R}L,\mathbf{R}'L'}$, which are all energy independent. Thus, Eq. (5.23) becomes a simple matrix eigenvalue problem and yields the eigenvalues as well as the wave functions of the system, which in turn gives a new potential. Such iterative process is performed until the potential reaches required convergence. In matrix form, the Hamiltonian and overlap matrix take the form

$$\begin{aligned} H &= \Big(\{K,\dot{\phi}\} - S\{J,\dot{\phi}\}\Big)E_v\Big(\{K,\dot{\phi}\} - \{J,\dot{\phi}\}S\Big) \\ &\quad - \Big(\{k,\dot{\phi}\} - S\{J,\dot{\phi}\}\Big)(\{K,\phi\} - \{J,\phi\}S), \\ O &= \Big(\{K,\dot{\phi}\} - S\{J,\dot{\phi}\}\Big)\Big(\{K,\dot{\phi}\} - \{J,\dot{\phi}\}S\Big). \end{aligned} \tag{5.30}$$

To simplify these expressions for convenience of calculation, we introduce a new function called "potential function"

$$P_{\mathbf{R}l} = \frac{\{K,\varphi(E)\}_{\mathbf{R}l}}{\{J,\varphi(E)\}_{\mathbf{R}l}}. \tag{5.31}$$

When inserting Taylor's expansion of $\varphi(E)$ Eq. (5.25) and keeping up to first order of $E - E_v$, we get

$$P_{\mathbf{R}l}(E) = \frac{E - C_{\mathbf{R}l}}{\Delta_{\mathbf{R}l} + \gamma_{\mathbf{R}l}(E - C_{\mathbf{R}l})}, \tag{5.32}$$

where quantities $C_{\mathbf{R}l}$, $\Delta_{\mathbf{R}l}$, and $\gamma_{\mathbf{R}l}$ are the so-called potential parameters and defined as

$$\begin{aligned} C_{\mathbf{R}l} &= E_{v,\mathbf{R}l} - \frac{\{K,\phi\}_{\mathbf{R}l}}{\{K,\dot{\phi}\}_{\mathbf{R}l}}, \\ \Delta_{\mathbf{R}l} &= \frac{\omega}{2}\frac{1}{\{K,\dot{\phi}\}^2_{\mathbf{R}l}}, \\ \gamma_{\mathbf{R}l} &= \frac{\{J,\dot{\phi}\}_{\mathbf{R}l}}{\{K,\dot{\phi}\}_{\mathbf{R}l}}. \end{aligned} \tag{5.33}$$

Insert these parameters into the Hamiltonian matrix and introduce a transformation U that satisfies $U^T U = O$, we eventually get the orthogonal Hamiltonian

$$\begin{aligned} H^{orth} =& U^T H U \\ =& C + \Delta^{1/2} S (1 - \gamma S)^{-1} \Delta^{1/2}. \end{aligned} \tag{5.34}$$

Roughly speaking, such formula is correct up to second order in $(E - E_v)$ for Hamiltonian and third order in $(E - E_v)$ for eigenvalues [38].

Next, we want the expression for Green's function, which is in demand for transport calculation. In matrix form, the Green's function matrix is defined as $G(z) = (z - H^{orth})^{-1}$. Insert Eq. (5.34) and after some algebra, we get

$$\begin{aligned} G(z) &= \lambda(z) + \mu(z) g(z) \mu(z), \\ g(z) &= [P(z) - S]^{-1} \end{aligned} \tag{5.35}$$

where

$$\begin{aligned} P(z) &= \frac{z - C}{\Delta + \gamma(z - C)}, \\ \mu(z) &= \frac{\sqrt{\Delta}}{\Delta + \gamma(z - C)}, \\ \lambda(z) &= \frac{\gamma}{\Delta + \gamma(z - C)}. \end{aligned} \tag{5.36}$$

It is worth noting that, except for S matrix, all these quantities in $G(z)$ are *diagonal* matrices. Then, the bottleneck for computation is the sparsity of S matrix. Unfortunately, S decays as $1/|\mathbf{R} - \mathbf{R}'|^{l + l' + 1}$, which is relatively slow for speed up. Then Anderson [39, 40] in 1980s, introduced the so-called tight-binding (TB) LMTO method by transforming the basis of the conventional LMTO basis into a new basis of the TB-LMTO's, under which all diagonal quantities stay still diagonal while S gets much sparser. In TB-LMTO used in our code, $S_{\mathbf{R}L,\mathbf{R}'L'}$ vanish when $|\mathbf{R} - \mathbf{R}'|$ goes beyond second-nearest neighbors. For details of the transformation, see Ref. [38] and we will only list the final equations to give a basic impression

$$\begin{aligned} G(z) &= \lambda^{\alpha}(z) + \mu^{\alpha}(z) g^{\alpha}(z) \mu^{\alpha}(z), \\ g^{\alpha}(z) &= [P^{\alpha}(z) - S^{\alpha}]^{-1}, \\ S^{\alpha} &= S[1 - \alpha S]^{-1}, \\ P^{\alpha}(z) &= [1 - P(z)\alpha]^{-1} P(z), \\ \mu^{\alpha}(z) &= \frac{\sqrt{\Delta}}{\Delta + (\gamma - \alpha)(z - C)}, \\ \lambda^{\alpha}(z) &= \frac{\gamma - \alpha}{\Delta + (\gamma - \alpha)(z - C)}. \end{aligned} \tag{5.37}$$

Again, we would like to mention that all quantities except for $G(z)$ and S^{α}, are diagonal matrices and we have omitted matrices subscripts $\{\ldots\}_{\mathbf{R}L}$ and $\{\ldots\}_{\mathbf{R}L,\mathbf{R}'L'}$ for all matrices for the sake of simplicity. Here, the $\alpha_{\mathbf{R}l}$'s are called screening constants and are equivalent for all close-packed system. Their values can be found in Ref. [38]. Two things are worthy mentioning here: (i) the Green's function $G(z)$ is independent of screening transform $\alpha_{\mathbf{R}l}$ so such transform will not affect the form of all physical quantities; (ii) these $\alpha_{\mathbf{R}l}$'s are applicable for almost all structures and materials.

5.2.3 Tight-Binding Formalism

The TB method was initially conceived by Slater and Koster in 1954 [41]. The method simplifies and parameterizes the Hamiltonian and also captures quantum and atomic effects without losing the accuracy. In TB formalism, the wave functions of electrons and holes are expanded as linear combinations of atomic orbitals. Electrons are assumed to be bonded at each atom and have interactions with neighbor atoms according to the real space lattice structure. The model gives good qualitative results and can couple with other methods to study transport, optical properties, and various kinds of many-body problems. Compared with DFT calculations, TB approximation can be more efficient and significantly reduce the computational cost, especially in transport simulations of a large device.

We first discuss atomic orbitals and orbital hybridization based on hydrogen atom. It is well known that there is only one electron outside the nucleus and the movement of electron can be solved rigorously. The nucleus of the hydrogen atom is composed of one proton and has positive charge. Electron moves at a coulomb field, and the Hamiltonian of electron can be described as follows:

$$H = -\frac{\hbar^2}{2m_e}\nabla + V_e(r) \tag{5.38}$$

where m_e is the electron mass, and V_e is the coulomb attraction energy:

$$V_e(r) = -\frac{e^2}{4\pi\varepsilon_0 r} \tag{5.39}$$

By solving the Schrödinger equation, electron wave function can be obtained and is written under the spherical coordinate as:

$$\psi_{nlm}(r) = R_{nl}(r)Y_{lm}(\theta,\phi) \tag{5.40}$$

where n is the principal quantum number ranging from 1, 2, 3, …, which describes the atom electron shell; l is the azimuthal quantum number and determines the orbital angular momentum; and m is the magnetic number. These three quantum

numbers determine the shape and energy level of electron wave function. $R_{nl}(\mathrm{r})$ are radial functions and describe the wave function change with radius. The lowest several radial functions are as follows:

$$\begin{aligned} n = 1 : R_{10} &= \frac{2}{a^{3/2}} e^{-r/a} \\ n = 2 : R_{20} &= \frac{1}{\sqrt{2}a^{3/2}} \left(1 - \frac{r}{2a}\right) e^{-r/2a}, \\ R_{21} &= \frac{1}{2\sqrt{6}a^{3/2}} \frac{r}{a} e^{-r/2a} \end{aligned} \tag{5.41}$$

Y_{lm} are spherical harmonics and represent the angular part of wave function. s, p, d orbitals are corresponding to l = 0, 1, 2, respectively. s orbital is the spherical symmetry:

$$Y_{\theta,\phi} = \frac{1}{\sqrt{4\pi}} \tag{5.42}$$

Three p orbitals have the same symmetry and are orthogonal:

$$Y_{1,-1}(\theta,\phi) = \sqrt{\frac{3}{4\pi}} x/r,\ Y_{1,0}(\theta,\phi) = \sqrt{\frac{3}{4\pi}} y/r,\ Y_{1,1}(\theta,\phi) = \sqrt{\frac{3}{4\pi}} z/r \tag{5.43}$$

Five orthogonal d orbitals are obtained as:

$$\begin{aligned} Y_{2,2}(\theta,\phi) &= \sqrt{\frac{15}{4\pi}} \frac{yz}{r^2},\ Y_{2,1}(\theta,\phi) = \sqrt{\frac{15}{4\pi}} \frac{xz}{r^2},\ Y_{2,0}(\theta,\phi) = \sqrt{\frac{15}{4\pi}} \frac{xy}{r^2} \\ Y_{2,-1}(\theta,\phi) &= \sqrt{\frac{15}{4\pi}} \frac{x^2 - y^2}{2r^2},\ Y_{2,-2}(\theta,\phi) = \sqrt{\frac{15}{4\pi}} \frac{3z^2 - r^2}{r^2} \end{aligned}$$

These orbital can be described in more concise form using Dirac's notation. For n = 1, there is only a s orbital: $|1s\rangle$; for n = 2, there are four orbitals: $|2s\rangle$, $|2p_x\rangle$, $|2p_y\rangle$, $|2p_z\rangle$; for n = 3, there are 9 orbitals: $|3s\rangle$, $|3p_x\rangle$, $|3p_y\rangle$, $|3p_z\rangle$, $|3d_{xy}\rangle$, $|3d_{yz}\rangle$, $|3d_{zx}\rangle$, $|3d_{x^2-y^2}\rangle$, $|3d_{3z^2-r^2}\rangle$. According to the orbital contribution to transport tight-modeling Hamiltonian of semiconductor can be constructed by these subshell orbitals (s, p, d, …). For example, in graphene, only one p orbital is enough to describe the lowest band near Fermi level [42]; while, for traditional semiconductor such as Si and Ge, s, p, and d orbitals should be applied to precisely obtain the band structure [43].

In solid materials, uncountable atoms are bonded together and there are a huge number of electrons. According to the Bloch theorem, wave function in solid crystals can be expanded by atomic orbitals on various atoms. The isolated atom orbital can also be applied to construct electron wave function in solid-state materials. In the most simplified TB model, the atomic orbitals are assumed to be

orthogonal and orbital hybridization is restricted within a certain distance between the first or second-nearest neighbors. When the distance between two atoms is far enough, the interaction can be neglected. Three-center integrals are also neglected and the electronic interaction only exists between two atoms [41].

Crystalline solids are periodically constructed by primitive cells. In a TB scheme, the electron wave functions can be expended as a Bloch sum of primitive cells:

$$\Phi_{\alpha lk}(r) = \frac{1}{\sqrt{N}} \sum_j e^{i(R_j + r_l)\cdot k} \phi_\alpha (r - R_j - r_l) \tag{5.44}$$

where N is the number of primitive cell in the system; α is the atomic orbital index; l denotes the atom position in a primitive cell; k is the wave vector; R_j is the position of the jth primitive cell; and r_l is the lth atom relative position within the primitive cell. Therefore, the eigenfunction in the bulk can be expressed as a linear combination of $\Phi_{\alpha lk}$:

$$\Psi_k = \sum_{\alpha,l} C_{\alpha l} \Phi_{\alpha lk} \tag{5.45}$$

In the representation, the Schödinger equation is given by:

$$\hat{H}\Psi_k = E\Psi_k \tag{5.46}$$

using Eq. (5.45):

$$\sum_{\alpha,l} C_{\alpha,l}(k) \hat{H} \Phi_{\alpha,l,k}(r) = E \sum_{\alpha,l} (k) \Phi_{\alpha,l,k}(r) \tag{5.47}$$

By multiplying $\Psi_k^\dagger$, we obtain:

$$\sum_{\alpha l} [H_{\alpha' l' \alpha l}(k) - E(k)\delta_{\alpha'\alpha}\delta_{l'l}] C_{\alpha,l}(k) = 0 \tag{5.48}$$

where

$$H_{\alpha' l' \alpha l}(k) = \langle \Phi_{\alpha' l' k} | \hat{H} | \Phi_{\alpha l k} \rangle, \langle \Phi_{\alpha' l' k} | \Phi_{\alpha l k} \rangle = \delta_{\alpha'\alpha}\delta_{l'l} \tag{5.49}$$

Equation (5.48) is secular equation whose eigenvalues are the electronics bands. Using Eq. 5.44, one can obtain the following:

$$\begin{aligned} H_{\alpha' l' \alpha l}(k) &= \frac{1}{N} \sum_{j'j} e^{i(R_j + r_l - R_{j'} - r_{l'})\cdot k} \langle \phi_{\alpha'}(r - R_{j'} - r_{l'}) | \hat{H} | \phi_\alpha (r - R_j - r_l) \rangle \\ &= \sum_j e^{i(R_j + r_l - R_{j'} - r_{l'})\cdot k} \langle \phi_{\alpha'}(r - R_{j'} - r_{l'}) | \hat{H} | \phi_\alpha (r - R_j - r_l) \rangle \end{aligned}$$

where $\langle \phi_{j'\alpha'} | \hat{H} | \phi_{j\alpha} \rangle$ is the electronic interaction between α' orbital at site j' and α orbital at site j. Due to the symmetry of periodical structure, the factor of 1/N is eliminated by summing over the lattice vectors. When interaction appears at on-site, the on-site integrals are the orbital energies:

$$\langle \phi_{js} | \hat{H} | \phi_{js} \rangle = E_s, \ \langle \phi_{jp} | \hat{H} | \phi_{jp} \rangle = E_p, \ \langle \phi_{jd} | \hat{H} | \phi_{jd} \rangle = E_d \tag{5.50}$$

When interaction exists between two nearest atoms, the inter-atom integrals are the orbital hopping energies and can be parameterized as Slater–Koster forms which can be found in Ref. [41].

5.3 Quantum Transport Theory

5.3.1 Landauer–Büttiker Formalism

In 1957, Landauer proposed a formula to study the conductance of metal with disorder and determined the electron conductance to be a function of tunneling possibility [44]:

$$G = \frac{e^2}{h} \frac{T}{R} \tag{5.51}$$

where T is the transmission possibility, and R is the reflection possibility. The measured device is assumed to be connected with two infinite leads. The leads are ideal conductors and the device is viewed as a potential barrier. So, the device can be modeled as electron wave function tunneling a barrier with finite height and the conductance is determined by the tunneling possibility. The Landauer formula presents the wave aspect of electron and is a quantum-mechanical result. From the equation, the conductance is infinite and there is no resistance when tunneling possibility is 100 %. Büttiker proposed another kind description [45]:

$$G = \frac{e^2}{h} T \tag{5.52}$$

The equation is apparently different from the Landauer formalism: Eq. (5.52) includes the resistance between the device and two leads, while Eq. (5.51) is just the device conductance. When device is large, the conductance can be described by Ohm law:

$$G = \frac{\sigma A}{L} \tag{5.53}$$

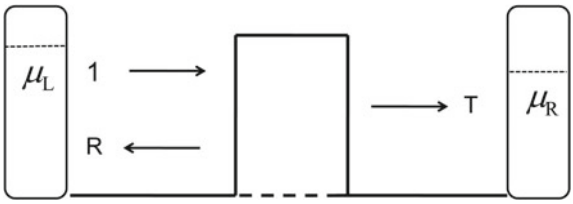

Fig. 5.2 Single-channel model with two reservoirs, μ_L and μ_R denote Fermi level of *left* and *right* reservoir, respectively

where σ is the conductivity which is the material intrinsic property, L is the length and A is the cross-sectional area. If the device length is scaled to be very short, the conductance should be infinite. While experimental results show that no matter how short the device is, the conductance is always finite and is consistent with Eq. (5.52).

How to understand the finite conductance? When device reaches nanoscale, quantum property presents and the Ohm's law does not work. Due to the quantum confinement, electron states are no more continuous and become discrete. In experiment, electrode is connected with measured device. There are infinite states from electrode injected to the device, while states in the device are limited due to quantum confinement. Only those states from electrodes matching with device states can get into the device, and the rest states are reflected which results in contact resistance.

Figure 5.2 shows a single-channel model. There is a barrier connected with two ideal leads, and a single quantum state is assumed. Electron is injected from left electrode to right electrode with +k. At the barrier interface, electron has a possibility to tunnel the barrier or be reflected to the left electrode. The two electrodes are the same material, so the tunneling possibility from left to the right is the same as that from the right to the left:

$$T_{L,R} = T_{R,L} = T \tag{5.54}$$

μ_L and μ_R are the Fermi energies of left and right electrode, respectively. μ_0 is chosen as a reference Fermi energy and smaller than μ_L and μ_R. The current from left to right is obtained to be:

$$I_{L,R} = -\frac{e}{h}T(\mu_L - \mu_0) \tag{5.55}$$

and the current from left to right is:

$$I_{R,L} = -\frac{e}{h}T(\mu_R - \mu_0) \tag{5.56}$$

Hence, the total current is:

$$I = I_{L,R} - I_{R,L} = -\frac{e}{h}T(\mu_L - \mu_R) \tag{5.57}$$

and then the conductance is obtained:

$$G = \frac{I}{V_L - V_R} = \frac{e^2}{h} T \tag{5.58}$$

The equation is the conductance of single-channel device and the Landauer–Büttiker equation [45]. The formalism can be extended and applied in a system with multi-channel and multi-probe [45]. The result is essential to the numerical calculation of the transmission coefficients by Green's function.

5.3.2 *Non-equilibrium Green's Function Formalism*

Figure 5.3 shows a two-probe device structure consisting of three parts: left lead, central region, and right lead. The two leads are connected with external electron reservoirs with constant electrochemical potentials μ_L and μ_R. When bias voltage is applied to two leads $\mu_L \neq \mu_R$, electron current flows from one lead through the central region to the other lead. Then, the whole system is at a non-equilibrium state. To calculate the non-equilibrium transport, electronic wave function should be obtained by solving the Schrödinger equation:

$$H\psi = E\psi \tag{5.59}$$

The Hamiltonian of the device includes terms describing the two leads, the central region, the interactions between leads and the central region, and electrostatic potential under bias voltage:

$$H = H_{leads} + H_{center} + H_{inter} + V_{ext} \tag{5.60}$$

There are open boundaries provided by semi-infinite electrodes which keep different chemical potential at the external bias voltage, which is different from a periodic system. The device boundary condition plays a crucial role in simulating non-equilibrium quantum transport properties.

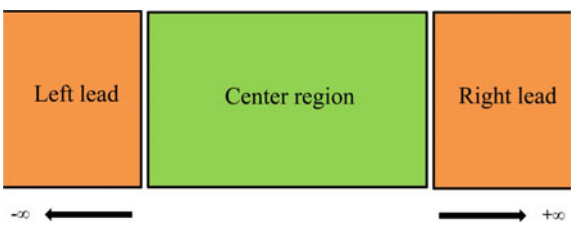

Fig. 5.3 Schematic illustration of a two-probe device, which has a center region connected with two semi-infinite leads

The Schrödinger equation can be solved by the NEGF method. The system Hamiltonian matrix can be divided as:

$$\mathbf{H} = \begin{pmatrix} H_{LL} & H_{LC} & 0 \\ H_{CL} & H_{CC} & H_{CR} \\ 0 & H_{RC} & H_{RR} \end{pmatrix} \tag{5.61}$$

where H_{LL}/H_{RR} is the left/right lead block Hamiltonian with infinite size, H_{CC} is the central region block Hamiltonian with finite size, H_{LC} and H_{CR} are the coupling matrixes between the lead and the central region. The system Green's function is:

$$\begin{aligned} \mathbf{G} &= \begin{pmatrix} G_{LL} & G_{LC} & G_{LR} \\ G_{CL} & G_{CC} & G_{CR} \\ G_{RL} & G_{RC} & G_{RR} \end{pmatrix} = (E^+ O - H)^{-1} \\ &= \begin{pmatrix} E^+ O - H_{LL} & -H_{LC} & 0 \\ -H_{CL} & E^+ O - H_{CC} & -H_{CR} \\ 0 & -H_{RC} & E^+ O - H_{RR} \end{pmatrix}^{-1} \end{aligned} \tag{5.62}$$

where $E^+ = E + i0^+$, E is the energy, 0^+ is a small number approaching zero, and O is the orbital hopping matrix. The Hamiltonian is an infinite matrix, so does the Green's function. However, the quantity of interest is the finite central region block of the Green's function G_{CC}, which can be solved as:

$$G_{CC} = [E^+ O - H_{CC} - H_{LC}^{\dagger}(E^+ O - H_{LL})^{-1} H_{LC} - H_{CR}(E^+ O - H_{RR})^{-1} H_{CR}^{\dagger}]^{-1} \tag{5.63}$$

where we define self-energies: $\Sigma_L = H_{LC}^{\dagger}(E^+ O - H_{LL})^{-1} H_{LC}$ and $\Sigma_R = H_{RC}^{\dagger}(E^+ O - H_{RR})^{-1} H_{RC}$, which describe the coupling of central region with infinite leads and have a finite size. The self-energy can be solved by an iterative technique [46]. By using the concept of self-energy, the infinite-sized problem is transformed to a finite-sized problem, which can be solved numerically.

The retarded Green's function is defined as

$$G^R = [E^+ O - H_{CC} - \Sigma_L - \Sigma_R]^{-1} \tag{5.64}$$

where H_{CC} is the device Hamiltonian with finite size. At equilibrium, the electron states in the devices are filled according to the Fermi level:

$$n(r) = \frac{s}{2\pi} diag\left[\int_{-\infty}^{+\infty} dEAOf(E-\mu)\right] \tag{5.65}$$

where s is the spin degeneracy, $A(E) = i(G^R - G^{R\dagger})$ is the spectral function and is the imaginary part of the device Green's function. At non-equilibrium, the Fermi level in the left lead is not the same as the right Fermi level. Electrons in the devices can be divided into two parts: from left lead with μ_L and from right lead with μ_R. Therefore, the electron density is calculated by integrating over all occupied states as follows:

$$n(r) = \frac{s}{2\pi} diag\left[\int_{-\infty}^{+\infty} dE(A_L f(E-\mu_L) + A_R f(E-\mu_R))O\right] \tag{5.66}$$

where

$$A^L = G\Gamma^L G^\dagger,\ A^R = G\Gamma^R G^\dagger \tag{5.67}$$

$$\Gamma_L = i(\Sigma^L - \Sigma^{L,\dagger}),\ \Gamma_R = i(\Sigma^R - \Sigma^{R,\dagger}) \tag{5.68}$$

From the spectral function, the device density of states (DOS) can be extracted:

$$DOS(E) = \frac{s}{2\pi} Tr[(A_L f(E-\mu_L) + A_R f(E-\mu_R))O] \tag{5.69}$$

Once electron density is obtained, the potential can be calculated by directly solving the Poisson equation including all external fields as electrostatic boundary conditions.

$$-\nabla(\varepsilon\nabla V) = q(N_D - N_A - n + p) \tag{5.70}$$

where N_D/N_A is the donor/acceptor concentration and n/p is the electron/hole density. The iteration between the transport equation and the Poisson equation is carried out until the self-consistency is achieved. After solving the potential and electron density from the iteration cycles, the transport properties can be calculated by the Landauer–Büttiker formalism [10]:

$$T(E) = Tr[\Gamma_L G^R \Gamma_R G^A] \tag{5.71}$$

and the current–voltage (I–V) characteristics under finite bias then is obtained:

$$I(V) = \frac{se}{h}\int_{-\infty}^{+\infty} dET(E)[f(E-\mu_L) - f(E-\mu_R)] \tag{5.72}$$

5.3.3 Random Disorder Scattering Within the LMTO Method

5.3.3.1 Coherent Potential Approximation

This section describes the CPA theory, a powerful method for handling configurational averaging of physical observables of substitutional random systems. It is developed based on the Green's function theory, in which all physical quantities have been expressed in terms of Green's functions, $G(E)$. So one main task of this section is to find computational methods and approximations for the configurational average of $G(E)$, i.e., $\overline{G(E)}$. Once $\overline{G(E)}$ is obtained, everything about the disorder system are known.

To describe the physical quantities of a random disorder system, we introduce an occupation index, $\eta_{\mathbf{R}}^{Q}$. It has the following meaning: $\eta_{\mathbf{R}}^{Q} = 1$ if an atom of the type Q occupies the site $\mathbf{R}$ and $\eta_{\mathbf{R}}^{Q} = 0$ otherwise. It satisfies the relation $\overline{\eta_{\mathbf{R}}^{Q}} = c_{\mathbf{R}}^{Q}$, where $c_{\mathbf{R}}^{Q}$ denotes the probability that atom of type Q occupies the site $\mathbf{R}$. Then, the Green's function in Eq. (5.37) has its random form as [47]

$$\begin{aligned} G_{\mathbf{R},\mathbf{R}'}(z) = &\sum_{Q} \lambda_{\mathbf{R}}^{\alpha,Q}(z)\eta_{\mathbf{R}}^{Q}\delta_{\mathbf{R},\mathbf{R}'} \\ &+ \sum_{Q,Q'} \mu_{\mathbf{R}}^{\alpha,Q}(z)\eta_{\mathbf{R}}^{Q} g_{\mathbf{R},\mathbf{R}'}^{\alpha}(z)\eta_{\mathbf{R}'}^{Q'}\mu_{\mathbf{R}'}^{\alpha,Q'}(z), \end{aligned} \tag{5.73}$$

where $\lambda_{\mathbf{R}}^{\alpha,Q}$ and $\mu_{\mathbf{R}}^{\alpha,Q}$ denote the value of $\lambda_{\mathbf{R}}^{\alpha}$ and $\mu_{\mathbf{R}}^{\alpha}$ when atom type Q occupies the atomic site $\mathbf{R}$. One thing should be mentioned is that here we have put the subscript of index L into the index $\mathbf{R}$ for the sake of simplicity, i.e., $\mathbf{R}L \rightarrow \mathbf{R}$. Then, we need to take its configurational average. We do it separately for diagonal parts and off-diagonal parts.

For diagonal parts, we have

$$\begin{aligned} \overline{G}_{\mathbf{R},\mathbf{R}}(z) &= \sum_{Q} c_{\mathbf{R}}^{Q}\overline{G}_{\mathbf{R},\mathbf{R}}^{Q}(z), \\ \overline{G}_{\mathbf{R},\mathbf{R}}^{Q}(z) &= \lambda_{\mathbf{R}}^{\alpha,Q}(z) + \mu_{\mathbf{R}}^{\alpha,Q}(z)\overline{g}_{\mathbf{R},\mathbf{R}}^{\alpha,Q}(z)\mu_{\mathbf{R}}^{\alpha,Q}(z), \\ \overline{g}_{\mathbf{R},\mathbf{R}}^{\alpha,Q}(z) &= \frac{1}{c_{\mathbf{R}}^{Q}}\overline{\eta_{\mathbf{R}}^{Q} g_{\mathbf{R},}^{\alpha}(z)}, \end{aligned} \tag{5.74}$$

where $\overline{G}_{\mathbf{R},\mathbf{R}}^{Q}(z)$ is called conditional average Green's function, which describes the decomposition of the averaged Green's function into contributions corresponding to the occupation of a particular lattice site $\mathbf{R}$ by an atom of type Q. Similarly, the off-diagonal parts take the form

$$\begin{aligned}\overline{G}_{\mathbf{R},\mathbf{R}'}(z) &= \sum_{Q,Q'} c_{\mathbf{R}}^{Q} \overline{G}_{\mathbf{R},\mathbf{R}'}^{Q,Q'}(z) c_{\mathbf{R}'}^{Q'}, \\ \overline{G}_{\mathbf{R},\mathbf{R}'}^{Q,Q'}(z) &= \mu_{\mathbf{R}}^{\alpha,Q}(z) \overline{g}_{\mathbf{R},\mathbf{R}'}^{\alpha,QQ'}(z) \mu_{\mathbf{R}'}^{\alpha,Q'}(z), \\ \overline{g}_{\mathbf{R},\mathbf{R}'}^{\alpha,QQ'}(z) &= \frac{1}{c_{\mathbf{R}}^{Q}} \overline{\eta_{\mathbf{R}}^{Q} g_{\mathbf{R},\mathbf{R}'}^{\alpha}(z) \eta_{\mathbf{R}'}^{Q'}} \frac{1}{c_{\mathbf{R}'}^{Q'}}.\end{aligned} \tag{5.75}$$

For derivation of $\overline{G(E)}$, we end here. One obvious thing is that the averaged Green's function, no matter diagonal parts or off-diagonal parts, can be expressed in terms of $\overline{g}^{\alpha}$, the configurationally averaged auxiliary Green's function. So how to calculate $\overline{g}^{\alpha}$ remains as a key problem of dealing with such random disorder systems. We apply the so-called CPA.

In CPA, we assume

$$\overline{g}^{\alpha} = (\mathscr{P}^{\alpha} - S^{\alpha})^{-1}, \tag{5.76}$$

where $\mathscr{P}^{\alpha}$ is a *diagonal* matrix. It describes the properties of effective non-random atoms which characterize the system after configurational averaging. It should be noted that the site-diagonal character of $\mathscr{P}^{\alpha}$ indicates that CPA neglects all local environmental effects of each atom. But CPA becomes exact in low concentration limit, weak scattering limit, and split-band limit [38]. Despite the lack of local environmental effects, CPA generally gives correct concentration or material-dependent trends for ground state physical quantities and this has been demonstrated by some of our published work [48, 49]. One more property of CPA is that it restores the translational invariant symmetry of the random disorder system, which enables us to apply the conventional Fourier transform method used for periodic structures.

The unknown $\mathscr{P}^{\alpha}$ needs to be determined self-consistently. Here, we apply the single-site approximation method, which is introduced by Velický [50] in 1968. We consider an atom of type Q occupying some given site $\mathbf{R}$ while all other sites are occupied by "effective" atoms corresponding to the translationally invariant effective medium $\mathscr{P}^{\alpha}$. Thus, we have only two cases, one with an atom A occupying site $\mathbf{R}$ with probability $c_{\mathbf{R}}^{A}$ and the other with an atom B occupying the same site $\mathbf{R}$ with probability $c_{\mathbf{R}}^{B}$. All other sites are described by the coherent potential function $\mathscr{P}^{\alpha}$. The scattering from such a single-impurity system is described by the single-site T-matrix

$$t_{\mathbf{R}}^{\alpha,Q}(z) = -\left[\left(P_{\mathbf{R}}^{\alpha,Q}(z) - \mathscr{P}_{\mathbf{R}}^{\alpha,Q}(z)\right)^{-1} + \overline{g}_{\mathbf{R},\mathbf{R}}^{\alpha}(z)\right]^{-1}. \tag{5.77}$$

Then the unknown coherent potential $\mathscr{P}^{\alpha}$ can be determined by the condition of vanishing average scattering due to the recovery of translational invariant symmetry of effective medium $\mathscr{P}^{\alpha}$, i.e.,

$$\sum_{Q} c_{\mathbf{R}}^{Q} t_{\mathbf{R}}^{\alpha,Q}(z) = 0. \tag{5.78}$$

This condition has to be satisfied by all sites $\mathbf{R}$ simultaneously and it is usually called the CPA condition. If inserting $t_{\mathbf{R}}^{\alpha,Q}$ into the above equation and using the relation $\sum_Q c_{\mathbf{R}}^{Q} = 1$, we get the expression as

$$\overline{g}_{\mathbf{R},\mathbf{R}}^{\alpha}(z) = \sum_{Q} c_{\mathbf{R}}^{Q} \overline{g}_{\mathbf{R},\mathbf{R}}^{\alpha,Q}(z), \tag{5.79}$$

which is equivalent to the CPA condition (5.78).

From now on, we do everything in a two-probe system [11], where random disorders only exist in the central region. Everything retains the same except for the Green's function. For $\overline{g}^{\alpha}$, we need to introduce a self-energy function (Σ) into the expression for center Green's function to describe the impact from the leads

$$\overline{g}^{\alpha} = (\mathscr{P}^{\alpha} - S^{\alpha} - \Sigma)^{-1}. \tag{5.80}$$

Equation (5.80) together with Eq. (5.79) gives the following closed self-consistent equation set

$$\begin{aligned}
\overline{g}_{\mathbf{R},\mathbf{R}}^{\alpha} &= \sum_{Q} c_{\mathbf{R}}^{Q} \overline{g}_{\mathbf{R},\mathbf{R}}^{\alpha,Q}, \\
\overline{g}^{\alpha} &= (\mathscr{P}^{\alpha} - S^{\alpha} - \Sigma)^{-1}, \\
\overline{g}_{\mathbf{R},\mathbf{R}}^{\alpha} &= [(\mathscr{P}^{\alpha} - S^{\alpha} - \Sigma)^{-1}]_{\mathbf{R},\mathbf{R}}, \\
\overline{g}_{\mathbf{R},\mathbf{R}}^{\alpha,Q} &= [(\tilde{P}_{\mathbf{R}}^{\alpha,Q} - S^{\alpha} - \Sigma)^{-1}]_{\mathbf{R},\mathbf{R}},
\end{aligned} \tag{5.81}$$

where $\tilde{P}_{\mathbf{R}}^{\alpha,Q}$ means to replace the $\mathbf{R}$th diagonal element of $\mathscr{P}^{\alpha}$ by the number $P_{\mathbf{R}}^{\alpha,Q}$, which can be called the conditional coherent potential.

We find that only diagonal parts of $\overline{g}^{\alpha}$ are needed to be determined $\mathscr{P}^{\alpha}$. So, we can introduce a coherent interactor $\Omega_{\mathbf{R}}$, which is a diagonal matrix. Then, the self-consistent equation set can be simplified as

$$\begin{aligned}
\overline{g}^{\alpha}_{\mathbf{R},\mathbf{R}} &= \sum_{Q} c^{Q}_{\mathbf{R}} \overline{g}^{\alpha,Q}_{\mathbf{R},\mathbf{R}}, \\
\overline{g}^{\alpha} &= \left(\mathscr{P}^{\alpha} - S^{\alpha} - \Sigma\right)^{-1}, \\
\overline{g}^{\alpha}_{\mathbf{R},\mathbf{R}} &= \left(\mathscr{P}^{\alpha}_{\mathbf{R}} - \Omega_{\mathbf{R}}\right)^{-1}, \\
\overline{g}^{\alpha,Q}_{\mathbf{R},\mathbf{R}} &= \left(P^{\alpha,Q}_{\mathbf{R}} - \Omega_{\mathbf{R}}\right)^{-1}.
\end{aligned} \tag{5.82}$$

Again, we want to mention that this equation set is closed and can be solved self-consistently. For detailed information of solving Eq. (5.82), we refer interested readers to Ref. [51]. Another thing worth mentioning is that the quantities, including Green's functions, coherent potentials, self-energy's, and coherent interactor can take either retarded or advanced form.

5.3.3.2 CPA at Non-equilibrium

This is the case for equilibrium system. For non-equilibrium system, we have another Green's function to deal with, i.e., $\overline{G}^{<}(E)$. Basically in LMTO–ASA framework, $\overline{G}^{<}_{\mathbf{R},\mathbf{R}} = \sum_{Q} c^{Q}_{\mathbf{R}} \overline{G}^{Q,<}_{\mathbf{R},\mathbf{R}}$ and $\overline{G}^{Q,<}_{\mathbf{R},\mathbf{R}} = \mu^{\alpha,Q}_{\mathbf{R}} \overline{g}^{\alpha,Q,<}_{\mathbf{R},\mathbf{R}} \mu^{\alpha,Q}_{\mathbf{R}}$. And these $\overline{g}^{\alpha,Q,<}_{\mathbf{R},\mathbf{R}}$ are needed to be solved self-consistently either so we need CPA equation set for $\overline{g}^{\alpha,<}$. Actually, equation set for $\overline{g}^{\alpha,<}$ can be obtained by applying the generalized Langreth theorem [51], which is an extension of the conventional Langreth theorem.

We know that Eq. (5.82) applies to retarded and advanced quantities. So generally, analytical continuation tells us that those equations also apply to complex contour quantities (Green's functions and self-energy), which actually satisfy exactly the same equation set as Eq. (5.82). Then we can apply the generalized Langreth theorem and obtain the following equation set

$$\begin{cases}
\overline{g}^{\alpha,r}_{\mathbf{R},\mathbf{R}} = \sum_{Q} c^{Q}_{\mathbf{R}} \overline{g}^{\alpha,Q,r}_{\mathbf{R},\mathbf{R}}, \\
\overline{g}^{\alpha,r} = \left(\mathscr{P}^{\alpha,r} - S^{\alpha} - \Sigma^{r}\right)^{-1}, \\
\overline{g}^{\alpha,r}_{\mathbf{R},\mathbf{R}} = \left(\mathscr{P}^{\alpha,r}_{\mathbf{R}} - \Omega^{r}_{\mathbf{R}}\right)^{-1}, \\
\overline{g}^{\alpha,Q,r}_{\mathbf{R},\mathbf{R}} = \left(P^{\alpha,Q,r}_{\mathbf{R}} - \Omega^{r}_{\mathbf{R}}\right)^{-1},
\end{cases} \tag{5.83}$$

$$\begin{cases}
\overline{g}^{\alpha,<}_{\mathbf{R},\mathbf{R}} = \sum_{Q} c^{Q}_{\mathbf{R}} \overline{g}^{\alpha,Q,<}_{\mathbf{R},\mathbf{R}}, \\
\overline{g}^{\alpha,<} = \overline{g}^{\alpha,r}\left(\Sigma^{<} - \mathscr{P}^{\alpha,<}\right)\overline{g}^{\alpha,a}, \\
\overline{g}^{\alpha,<}_{\mathbf{R},\mathbf{R}} = \overline{g}^{\alpha,r}_{\mathbf{R},\mathbf{R}}\left(\Omega^{<}_{\mathbf{R}} - \mathscr{P}^{\alpha,<}_{\mathbf{R}}\right)\overline{g}^{\alpha,a}_{\mathbf{R},\mathbf{R}}, \\
\overline{g}^{\alpha,Q,<}_{\mathbf{R},\mathbf{R}} = \overline{g}^{\alpha,Q,r}_{\mathbf{R},\mathbf{R}} \Omega^{<}_{\mathbf{R}} \overline{g}^{\alpha,Q,a}_{\mathbf{R},\mathbf{R}}.
\end{cases} \tag{5.84}$$

These 8 equations are the so-called non-equilibrium coherent potential approximation (NECPA) equation set [51]. It can be solved self-consistently. For details of how to solve them, we refer interested readers to read Ref. [51]. Once they are solved, in principle, everything about the random disorder system are solved.

5.3.3.3 Average and Fluctuation of Quantum Transport

This section discusses how to calculate transmission coefficient and its fluctuation. In LMTO–ASA framework, the transmission coefficient takes the following form:

$$\overline{T(E)} = \mathrm{Tr}\left[\overline{g^{\alpha,r}(E)\Gamma_L^{\alpha}(E)g^{\alpha,a}(E)\Gamma_R^{\alpha}(E)}\right], \tag{5.85}$$

where $\Gamma_{L,R}$ are the linewidth functions of the left/right lead and take the form $\Gamma_{L/R} = i(\Sigma_{L/R}^{r} - \Sigma_{L/R}^{a})$. Remember that self-energy's are non-random functions, so $\overline{T(E)}$ only requires calculating the average $\overline{g^{\alpha,r}(E)\Gamma_L^{\alpha}(E)g^{\alpha,a}(E)}$. In another aspect, $\overline{g}^{\alpha,<} = \overline{g^{\alpha,r}\Sigma^{\alpha,<}g^{\alpha,a}}$. We found that these two averages take the same mathematical form, which indicates they can be calculated with similar mathematical algebraic method. Actually, $\overline{T(E)}$ can be calculated by self-consistently solving the NECPA equation set, in which we set $f_L(E) = -i$ and $f_R(E) = 0$ in $\Sigma^{<}(E)$.

Next, we discuss how to calculate the variance of $T(E)$. The basic idea is to start from a NEGF-based quantum transport formulation of the variability $dT \equiv \sqrt{\overline{T^2} - \overline{T}^2}$ and carry out a further disorder ensemble average over T^2 *analytically* to deduce an *analytical formula* of δT. This analytical formula is then implemented in NEGF-based DFT [52] for further atomistic modeling of DDV. It is clear that the key problem of calculating dT is the trouble of calculating

$$\overline{T^2} = \mathrm{Tr}\left(\overline{g^r\Gamma_L g^a\Gamma_R g^r\Gamma_L g^a\Gamma_R}\right), \tag{5.86}$$

where we have omitted the configuration index α for the sake of simplicity. Actually, there is some method that can exactly calculate this quantity by applying the Γ-decomposition and its details can be found in Ref. [52]. But in real material, especially semiconductors, dopants or defects usually have very low concentration ($<1\ \%$). Therefore, we can make good use of this property and workout Eq. (5.86) in the limit of low concentration. This method is called the low-concentration approximation (LCA) and has been discussed in detail in Ref. [52]. It is developed based on the scattering theory and Feynman diagrams. Here, we will simply list the basic idea and some key results of LCA.

We start from the scattering form Green's function

$$\begin{aligned} g^r = \overline{g}^r + \sum_{\mathbf{R}} \overline{g}^r t^r_{\mathbf{R}} \overline{g}^r + \sum_{\mathbf{R}} \sum_{\mathbf{R}' \neq \mathbf{R}} \overline{g}^r t^r_{\mathbf{R}'} \overline{g}^r t^r_{\mathbf{R}} \overline{g}^r \\ + \sum_{\mathbf{R}} \sum_{\mathbf{R}' \neq \mathbf{R}} \sum_{\mathbf{R}'' \neq \mathbf{R}'} \overline{g}^r t^r_{\mathbf{R}''} \overline{g}^r t^r_{\mathbf{R}'} \overline{g}^r t^r_{\mathbf{R}} \overline{g}^r + \ldots \end{aligned} \tag{5.87}$$

Again, we mention that configuration index α is omitted here, and $t^r_{\mathbf{R}}$'s are random variables which can take the value $t^{r,Q}_{\mathbf{R}}$ with the probability of $c^Q_{\mathbf{R}}$. Then, we insert Eq. (5.87) and its advanced counterpart into Eq. (5.86). Basically, we will have infinite number of terms for $\overline{T^2}$ which makes it impossible to calculate. But here, since $c^Q_{\mathbf{R}}$ for dopants or defects are very small, we can rewrite those terms in order of $c^Q_{\mathbf{R}}$ and only keep the first-order terms [52]. With the help of Feynman diagram, we find that only 9 diagrams survive for dT and they can be written in a beautiful form as

$$\begin{aligned} dT^2 &= \sum_{\mathbf{R},Q>0} c^Q_{\mathbf{R}} \left(Y^{\alpha,Q}_{\mathbf{R}} + Y^{\beta,Q}_{\mathbf{R}} + Y^{\gamma,Q}_{\mathbf{R}} \right)^2, \\ Y^{\alpha,Q}_{\mathbf{R}} &= \mathrm{Tr}\left\{ t^{a,Q}_{\mathbf{R}} [\overline{g}^a \Gamma_R \overline{g}^r \Gamma_L \overline{g}^a]_{\mathbf{R},\mathbf{R}} \right\}, \\ Y^{\beta,Q}_{\mathbf{R}} &= \mathrm{Tr}\left\{ t^{r,Q}_{\mathbf{R}} [\overline{g}^r \Gamma_R \overline{g}^a \Gamma_L \overline{g}^r]_{\mathbf{R},\mathbf{R}} \right\}, \\ Y^{\gamma,Q}_{\mathbf{R}} &= \mathrm{Tr}\left\{ t^{r,Q}_{\mathbf{R}} [\overline{g}^r \Gamma_L \overline{g}^a]_{ii} t^{a,Q}_{\mathbf{R}} [\overline{g}^a \Gamma_R \overline{g}^r]_{\mathbf{R},\mathbf{R}} \right\}, \end{aligned} \tag{5.88}$$

where we have denoted that $Q = 0$ refers to the host atoms whose concentration is large compared to dopants or defects. One thing we want to mention at last is this method also applies to the case where there are more than one host atoms (e.g., GaAs with defects).

5.4 Applications

5.4.1 Atomic Disorder Scattering in Graphene TFETs

With device continuous miniaturization, electronic properties related to atomic structure are playing more and more important role in device performance. Atomic disorder is an important issue limiting the performance of nanoscale devices. Intensive studies have been carried out in investigation into various atomic disorder effects. Random dopant results in threshold voltage lowering and fluctuations and leads to the well-known device-to-device variability; due to the discrete atomic dopants, the semiconductor–metal contact resistance is greatly changed [54, 55]; the resistance of copper or graphene [56] interconnect wires is greatly increased by the

random grain boundary. Several theoretical methods have been applied to study atomic disorder effects in nanodevices, such as the density gradient correction in the drift-diffusion model [57] and the random-alloy approach in the TB model [58]. An atomic disorder not only modifies the microscopical atomic structure, but also changes the electronic properties. For simulating atomic disorder effects, DFT is the most realistic approach because it can study material electronic properties from atomic structure with parameter free. Transport calculations using DFT within the NEGF formalism have been realized in atomic device [11]. However, considering atomic disorder in device, brute force computation of different configurations has to be carried out to get statistical information, which is a bottleneck for the first principles calculations in dealing with disorder. Fortunately, the theory of the NECPA implemented in LMTO has been developed for calculating the average non-equilibrium transport properties with disorder scattering [51] and has previously been applied to simulate atomic disorders in magnetic tunnel junctions [59], copper interconnects [60], and Si nanochannels [61]. In this section, we introduce the work about studying atomic disorder effects in B–N co-doped graphene TFETs using first principles.

5.4.1.1 Device Model and Electronic Structure

Figure 5.4a shows the device structure of a graphene transistor with Al_2O_3 gate oxide and copper contacts. The channel under the oxide layer is co-doped with boron (B) and nitride (N) [62], where co-doping means the concentration of B and N is the same. With the CPA method, the average electronic and transport properties of such disordered system can be calculated. Figure 5.4b plots one of many

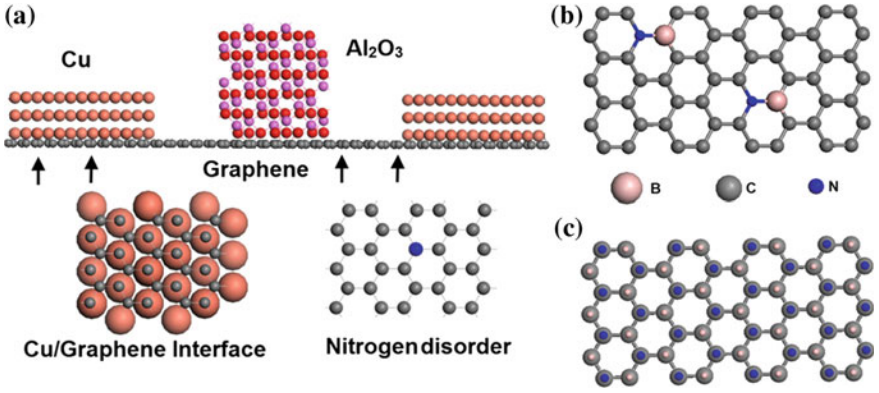

Fig. 5.4 **a** The device structure of the simulated graphene TFET. **b** A specific atomic configuration of B–N co-doped graphene. **c** The effective medium model of disordered graphene treated by the CPA method [53]

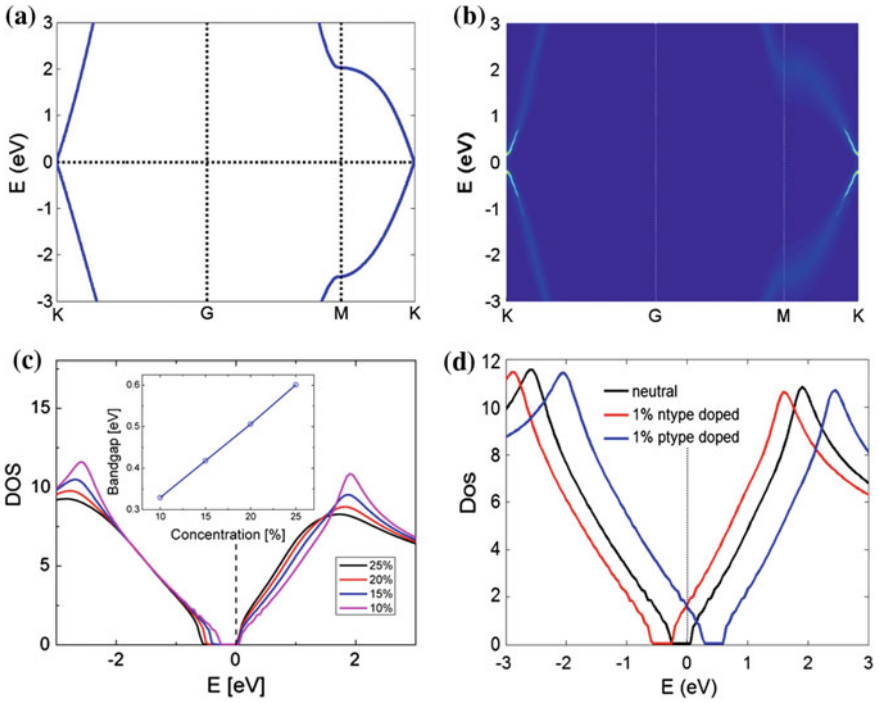

Fig. 5.5 **a** Band structure of pristine graphene; **b** CPA density of states (DOS) of doped graphene; **c** DOS of disordered graphene at different co-doping concentrations; **d** the averaged DOS versus energy of the B–N co-doped graphene with extra individual boron (p-type) or nitrogen (n-type) impurity atoms. Reprinted with permission from [53]. © IEEE 2014

atomic configurations for a given B–N concentration x%. Due to random doping, the atomic structure varies from one configuration to another; hence, the electronic structure is changed by the doping. Instead of repetitively computing different configurations, their average properties are calculated once by the effective medium model of CPA, which is represented in Fig. 5.4c. In the model, disorder doping is averaged to each atomic site of graphene with a probability of x%.

Figure 5.5a presents the calculated band structure of pristine graphene with zero band gap and Dirac cone. Once the graphene is co-doped by B and N dopants, the band structure is changed correspondingly. Figure 5.5b shows the disorder-averaged CPA band structure [63]. Due to B–N co-doping, disordered graphene has a band gap of 0.33 eV and no longer massless with 0.1 m_0 electron mass. These results have been benchmarked and agree with brute force DFT calculations of many disorder configurations using VASP. With the increasing of co-doping density, the band gap gets larger linearly as shown in Fig. 5.5c. The Fermi level is modified with the doping concentration change of B or N dopant and p-type or n-type graphene can be achieved as shown in Fig. 5.5d.

5.4.1.2 Modeling Transport with Disorder Scattering

Two different doping channels of simulated double-gate TFETs are shown in Fig. 5.6a, b. The channel length L_G is 10 nm and gate dielectric thickness EOT = 1.7 nm. In Fig. 5.6a, the doping of graphene TFETs is treated by the virtual crystal approximation (VCA), in which the atom electronic potential is shifted due to the average doping. Atomic disorder is considered in the whole channel and part of source/drain as shown in Fig. 5.6b and calculated by NECPA. The transverse direction perpendicular to transport direction is treated as periodic conditions. The whole simulated system has 2160 atoms calculated with the VCA theme and 4080 atoms with the NECPA theme.

Both the VCA and the NECPA themes give the same band profiles along the transport direction as shown in Fig. 5.6c. Figure 5.6d compares transmission spectrums of the VCA and NECPA themes. It is observed that the atomic disorder scattering has important influence on tunneling process. Compared to the VCA curve without disorder scattering, the transmission is suppressed in the NECPA theme due to atomic disorder scattering. In the disordered graphene TFETs, the inter-band tunneling at source–channel junction is diffusive as illustrated by the local density of states (LDOS) as in Fig. 5.7a; thus, the tunneling current is reduced. Figure 5.7b shows transfer characteristics of the graphene TFETs with or without atomic disorder. It can be found that drain current is decreased by atomic disorder scattering. It is noteworthy that disorder scattering has much larger influence on off-state current than on on-state current as shown in Fig. 5.7b. It is also found that subthreshold slope is increased by 20 % due to atomic disorder.

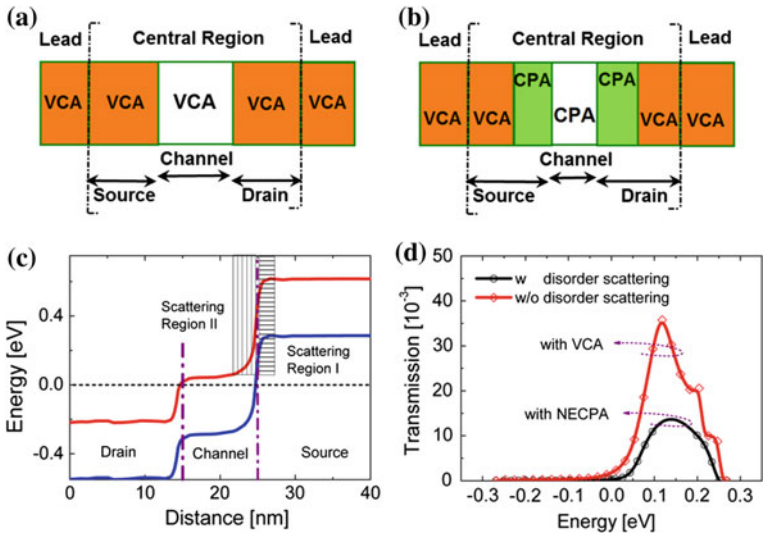

Fig. 5.6 **a** VCA- and **b** CPA-doped graphene channels. **c** Potential profiles of graphene TFETs and **d** corresponding transmissions. Reprinted with permission from [53]. © IEEE 2014

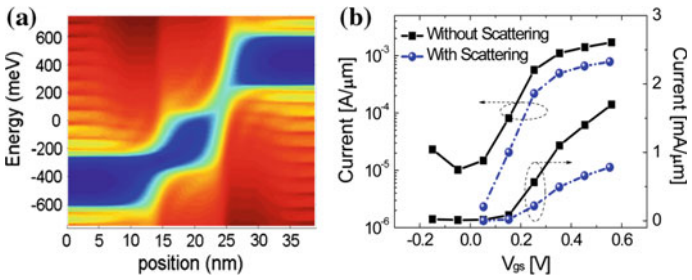

Fig. 5.7 **a** LDOS of the co-doped graphene TFET using the CPA method and **b** drain current as a function of gate voltage of the 10 nm graphene TFET with and without disorder scattering. Reprinted with permission from [53]. © IEEE 2014

5.4.2 Device Physics in TMDC TFETs

Recently, two-dimensional (2D)-layered semiconductor crystalline solids have been attracting increasing interest as promising channel materials for TFET applications, such as graphene nanoribbons (GNRs) [26, 27], TMDCs [65, 66], and phosphorene [67, 68]. Atomically, thin structures are beneficial for better gate control and leakage current can also be effectively suppressed than conventional bulk semiconductors. In planar 2D materials, high electric field at the tunnel junction is expected to enhance the tunneling current. At the same time, 2D materials have good mechanical properties and are attractive for flexible electronics [69]. Furthermore, these 2D materials have no surface dangling bonds, thus are immune to device variability due to interface disorder in TFETs using bulk semiconductors. Clean interface is also desired for reducing subthreshold slope (SS) [70]. More recently, layered TMDC TFETs have been extensively studied [30, 65, 66, 71, 72]. In this section, we discussed device physics in monolayer TMDC TFETs by atomic simulations within the NEGF formalism [73]. Electron-hole symmetry and sub-bands contributing to transport are well captured by three-band TB model. With the precise band structure, orientation-dependent transport and negative differential resistance in monolayer TMDC TFETs are predicted. For optimizing device performance, source/drain doping concentration and gate oxide thickness are tuned. The scaling behavior of TMDC TFETs is also discussed. At last, transport properties of six different monolayer TMDC materials are compared for TFET applications.

5.4.2.1 Electronic Structure

Figure 5.8a schematically illustrates the cross-sectional view of a double-gate TFET. Monolayer TMDCs are applied as the channel material. There are two natural transport directions with high symmetry as shown in Fig. 5.8b: the zigzag direction (ZD) and the armchair direction (AD). Generally, the band edge effective

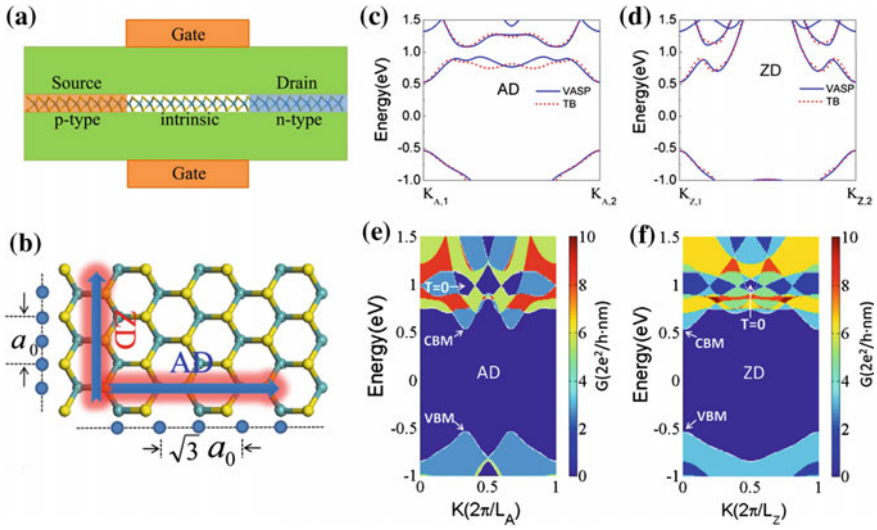

Fig. 5.8 **a** Schematic illustration of a double-gate monolayer TMDC TFET with 12-nm intrinsic channel and 10 nm p-type/n-type source/drain. 3 nm thick HfO_2 is applied as the gate oxide. **b** Atomistic structures of monolayer TMDC. Band structures and transmission (T) spectrums of monolayer $MoTe_2$ in (**c**, **e**) the armchair direction (AD) and (**d**, **f**) the zigzag direction (ZD). Reprinted with permission from [31, 64]. © IEEE 2015 2016

masses along the two directions are the same. However, the conduction sub-band properties along AD and ZD are different. The band structures calculated by the TB model and density functional theory (DFT, by VASP) along the AD and ZD are compared in Fig. 5.8c, d. Both TB and DFT results show a gap in the conduction band in AD (Fig. 5.8c) at around 1.0 eV. This gap corresponds to the conductance valley at $K = 2\pi/(3L_A)$ or $4\pi/(3L_A)$ in the transmission spectrum of AD. On the other hand, there is no such gap in ZD (see, Fig. 5.8d).

Figure 5.8e, f shows the ballistic transmission spectrums of $MoTe_2$ obtained by using the NEGF formalism, which indicate available transport channels in the wave vector (K) and energy (E) space. It can be observed that transmission in the AD is different from ZD. In AD, both the conduction band minimum (CBM) and the valance band maximum (VBM) are at $K = 2\pi/3L_A$ or $4\pi/3L_A$, and at the same K there are regions without transport channels in CB. Therefore, when carriers from source VB transport to drain CB, some of the carriers may tunnel to these regions without channels and then drain current is reduced [29]. However, there is no such physics in ZD TMDC TFETs. Channel-less region in CB does not exist at $K = 0$ or $K = 2\pi/L_Z$ where the CBM and VBM locate. Due to the orientation-dependent band structure, monolayer TMDC TFETs have different device characteristics in the two directions.

5.4.2.2 Current–Voltage Characteristics

Figure 5.9a shows the transfer characteristics of 12 nm double-gate monolayer $MoTe_2$ TFETs. The simulated device has top and bottom gate insulator layers of 3 nm HfO_2 ($\kappa = 25$). The source and drain are doped to p-type and n-type with the same doping concentration of $n_0 = 0.02$ dopant/atom, respectively. The device transport properties are simulated by self-consistently solving the Poisson equation and the open boundary Schrödinger equation within the NEGF formalism [29, 73]. The minimum current is achieved at $V_G = V_D/2$, where the gate-to-source voltage is the same as the drain-to-gate voltage and the tunneling barrier from source to drain is the largest. From Fig. 5.9a, an ambipolar behavior can be found in monolayer $MoTe_2$ TFETs, while n-type and p-type I_D–V_G curves in AD TFETs are asymmetric.

Figure 5.9a clearly shows that the transfer characteristics of $MoTe_2$ TFETs are orientation dependent. I_D in ZD is larger than that in AD in all studied gate voltage range. The drain currents at $V_G = 0.25$ V and $V_G = 0.75$ V of ZD TFETs are 212 and 305 larger than those of AD TFETs, respectively. The orientation transport is more significant than in TMDC FETs [74]. Generally, the transport properties mainly depend on electronic structure, doping density, and channel length. In the following, we discuss the orientation-dependent transport in TMDC TFETs.

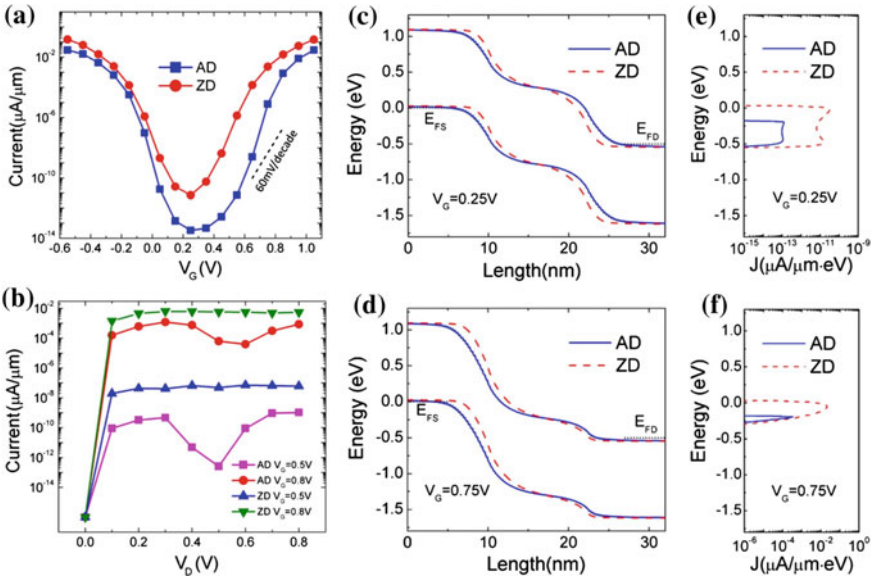

Fig. 5.9 **a** I_D versus V_G and **b** I_D versus V_D of 12-nm monolayer $MoTe_2$ TFETs in the armchair direction (AD) and the zigzag direction (ZD) at $V_D = 0.5$ V. **c**, **d** Potential profiles and **e**, **f** current densities (J) of monolayer $MoTe_2$ TFETs in AD and ZD at off-state ($V_G = 0.25$ V) and on-state ($V_G = 0.75$ V). Reprinted with permission from [31]. © IEEE 2016

First, band edge effective masses of monolayer $MoTe_2$ are isotropic and do not provide a reason for the orientation-dependent transfer characteristics. However, the conduction band is orientation dependent as presented in Fig. 5.8c, d. Figure 5.9e shows the current density in AD TFETs is smaller than ZD TFETs at -0.3 eV < E < 0 eV, corresponding to the region without transport channel in conduction band as shown in Fig. 5.8c, e. So, drain current is reduced in the energy region in AD TFETs. However, there is no such mechanism in ZD TFETs. It is also found that the orientation-dependent conduction band results in $I_D - V_G$ asymmetry of n-type and p-type AD TFETs as shown in Fig. 5.9a. For n-type TFETs, with the increasing of V_G the drain current mainly tunnels at the source–channel junction and encounters the energy region without channel in the drain CB. While, in p-type TFETs carrier tunnels at the channel–drain junction and thus the channel-less region in drain conduction band plays a less important role.

Secondly, the atomic structure also contributes the orientation-dependent transport in TMDCs TFETs. From Fig. 5.9c, d, different depletion region lengths in AD and ZD can be found, which is due to different atomic arrangements in AD and ZD. From Fig. 5.8b, the projected distances between two nearest Mo atoms are different: $a_0/2$ in ZD and $\sqrt{3}a_0/2$ in AD, where $a_0 = 3.557$ Å. As a result, projected 1D-doping densities along the two directions are different. For an abrupt doped p–n junction, the depletion region length is obtained by [75]:

$$W \propto \left[\left(\frac{1}{N_A} + \frac{1}{N_D}\right) V_{bi}\right]^{1/2} \tag{5.89}$$

where W is the depletion width, N_A/N_D is the 1D density of acceptors/donors in p-type/n-type region, and V_{bi} is the built-in potential. For the same number of dopant per atom in a $MoTe_2$ p–n junction, we can obtain:

$$N_{A,AD} < N_{A,ZD},\ N_{D,AD} < N_{D,ZD},\ V_{bi,AD} = V_{bi,ZD} \tag{5.90}$$

Therefore, the AD depletion width is longer than ZD, $W_{AD} > W_{ZD}$. Hence, a wider tunnel barrier in AD TFETs results in smaller current in Fig. 5.9a.

The channel-less region in conduction band also leads to giant negative differential resistance (NDR). The I_D–V_D of monolayer $MoTe_2$ TFETs in the two directions are compared in Fig. 5.9b. There is a giant NDR in AD TFETs and the peak-to-valley ratio (PVR) reaches 10^3. With the tuning of V_D, the contribution of channel-less region in AD is modified; hence, NDR is achieved. It is predicted that NDR is quiet general and exists in at least six TMDC materials (MoS_2, $MoSe_2$, $MoTe_2$, WS_2, WSe_2, and WTe_2) [29].

5.4.2.3 Doping Density and Geometry Dependence

Figure 5.10a, b compare I_D–V_G of 12-nm monolayer $MoTe_2$ TFETs with different doping densities, which is changed equally in source and drain. With the increasing of doping concentration, the drain current is increased at all gate voltages. Specifically, the I_D of ZD TFETs at $V_G = 0.75$ V is 1.3×10^{-2}, 2.4×10^{-3} and 4.8×10^{-5} μA/μm for $N_{S/D} = 1.5n_0, n_0$ and $0.5n_0$, respectively. The increasing of doping density not only enlarges the tunneling energy window from source VB edge to drain CB edge, but also decreases the depletion region length at source–channel and channel–drain junctions. Consequently, drain current gets larger with the increase of doping density for both n-type and p-type TFETs. On the other hand, SS is deteriorated at the same time: 48.0, 59.7, and 62.5 mV/decade for $N_{S/D} = 0.5n_0, n_0$ and $1.5n_0$, respectively.

Next, we discuss the effect of gate oxide thickness T_{ox} on I_D of monolayer $MoTe_2$ TFETs as shown in Fig. 5.10c. Device performance of $MoTe_2$ can be greatly improved by using thinner gate oxide layer. The off-state current at $V_G = 0.25$ V is decreased by 15 times while the on-state current at $V_G = 0.75$ V is increased by 37 times. As a result, SS is improved from 58.6 to 44.3 mV/decade. Hence, thinner insulator layer not only boosts the on-state current but also decreases the SS. The gate control of double-gate devices can be quantified by the natural decay length λ given by [76]:

$$\lambda = \sqrt{\frac{\varepsilon_{ch}}{2\varepsilon_{ox}}\left(1 + \frac{\varepsilon_{ox}}{4\varepsilon_{ch}}\frac{t_{ch}}{t_{ox}}\right)t_{ox}t_{ch}} \tag{5.91}$$

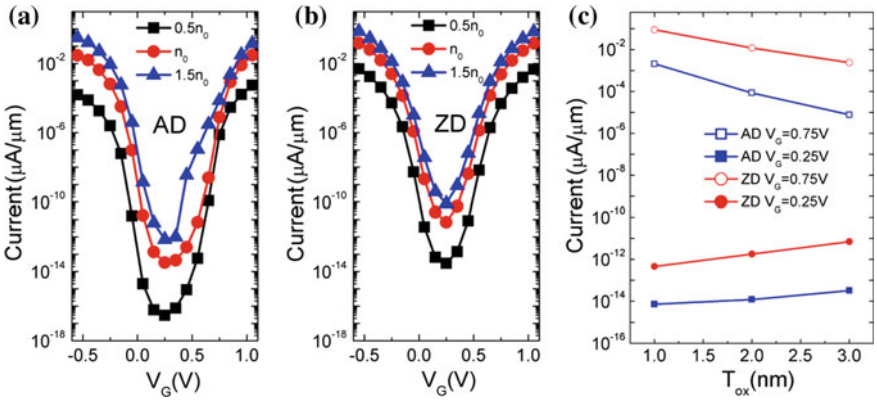

Fig. 5.10 I_D versus V_G of 12-nm monolayer MoTe_2 TFETs with different doping densities at V_D = 0.5 V: **a** the armchair direction (AD) **b** the zigzag direction (ZD) and **c** I_D as a function of HfO_2 oxide layer thickness (T_{ox}) in 12-nm monolayer $MoTe_2$ TFETs at $V_D = 0.5$ V. Reprinted with permission from [31]. © IEEE 2016

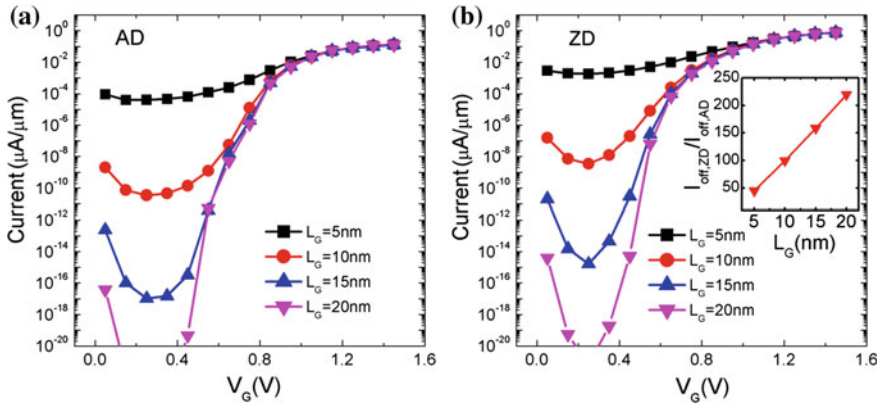

Fig. 5.11 I_D versus V_G of monolayer $MoTe_2$ TFETs with different gate lengths: **a** the armchair direction (AD) and **b** the zigzag direction (ZD). The inset figure shows the ratio of $I_{off,ZD}/I_{off,AD}$ as a function of the gate length at V_G = 0.25 V. Reprinted with permission from [31]. © IEEE 2016

where ε_{ch} and ε_{ox} are the channel and gate dielectric constant, respectively; t_{ch} and t_{ox} are the corresponding thicknesses. Using thinner gate oxide layer, the decay length λ is reduces and electric field in the tunneling junction is enlarged. As a result, on-state current is increased.

Figure 5.11 shows the transfer characteristics of monolayer $MoTe_2$ TFETs with different gate lengths at V_D = 0.5 V. Due to the increase of direct tunneling width, the off-state current decreases with the channel length as expected, while the saturation current is mainly determined by the source–channel tunneling junction for n-type TFETs, which does not change with L_G. As a result, the saturation current remains the same as shown in Fig. 5.11. From the inset figure in Fig. 5.11b, it can be found that the $I_{off,ZD}/I_{off,AD}$ ratio increases with the gate length at V_G = 0.25 V.

5.4.2.4 Comparison of Six Kinds of TMDC TFETs

In this section, device performances of six kinds of monolayer TMDC TFETs are compared. These devices have the same device structure with 3-nm HfO_2 oxide layer and 12-nm gate length. The channel materials are in the form of MX_2 where M = Mo, W and X = S, Se, Te. Band gaps and carrier effective masses of these materials are shown in Table 5.1, which determine device performances. I_{on} as a function of I_{on}/I_{off} ratio in ZD TFETs is shown in Fig. 5.12a. Monolayer MoS_2 TFETs have the lowest saturation current but can reach the largest I_{on}/I_{off} ratio. The band gap and carrier effective masses of monolayer WTe_2 are the smallest, so the largest on-state current can be achieved in monolayer WTe_2 TFETs when the I_{on}/I_{off} ratio is smaller than 10^7. At last, device performances of monolayer MoS_2 TFETs and monolayer WTe_2 TFETs in AD and ZD are presented in Fig. 5.12b, c.

Table 5.1 E_g is the band gap energy, $m_e(m_h)$ is the electron (hole) effective mass extracted from the tight-binding model, m_0 is the bare electron mass, and SS_{min} is the minimal subthreshold swing achieved in 12 nm ZD TFETs at $V_D = 0.5V$ [31]

	MoS_2	WS_2	$MoSe_2$	WSe_2	$MoTe_2$	WTe_2
E_g(eV)	1.66	1.81	1.43	1.54	1.07	1.07
$m_e(m_0)$	0.42	0.30	0.48	0.32	0.39	0.25
$m_h(m_0)$	0.54	0.39	0.59	0.41	0.54	0.34
SS_{min}	48.3	56.1	49.4	58.6	58.6	71.2

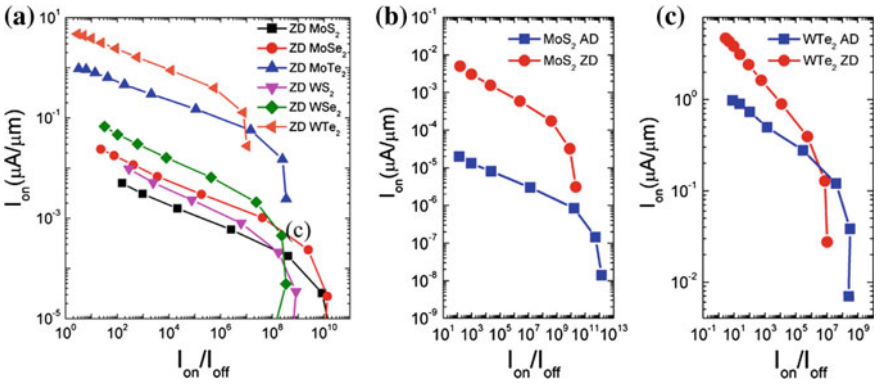

Fig. 5.12 I_{on} as a function of I_{on}/I_{off} ratio at a bias window of 0.5 V: **a** six kinds of monolayer TMDC TFETs in ZD, **b** monolayer MoS_2 TFETs, and **c** monolayer WTe_2 TFETs. Reprinted with permission from [31]. © IEEE 2016

Orientation-dependent transport also exists in the two kinds of TFETs. ZD TFETs have larger I_{on} but lower I_{on}/I_{off} ratio utmost limit.

5.5 Summary

In summary, we have reviewed the fundamental theory for electronic structure calculations. Using DFT, electronic structure of solid-state material can be calculated from atomic structure with parameter free. We specifically discussed the TB-LMTO method, which is a basis set for NEGF–DFT calculations. The TB model is also discussed and suitable for large-scale computing at the atomic level. With LMTO basis, CPA–NVC theory is developed to deal with impurity scattering in nanodevices. With the CPA–NVC method, the statistic transport information can be obtained without iteratively computing different atomic configurations. At last, two applications are reviewed. Diffusive transport of graphene TFETs with boron and nitrogen impurities is simulated with the parameter-free first principles

modeling methodology of the NEGF–DFT–NVC approach. It is revealed that the band-to-band tunneling current is substantially suppressed by the atomistic disorder. In another application, device physics in TMDC TFETs is investigated with NEGF using TB approach. By atomistic simulations, it is demonstrated that atomic arrangement and sub-bands have great impact on transport properties of TMDC TFETs at nanoscale. It is unquestioned that atomistic simulations of TFETs get more and more important when device dimension is continuously scaled down. It is valuable to further study various atomistic effects on device performance of TFETs.

Acknowledgements This work was supported by the University Grant Council (Contract No. AoE/P-04/08) of the Government of HKSAR, National Natural Science Foundation of China with No. 11374246 (J. Wang), and NSERC of Canada (H. Guo). We thank CLUMEQ, CalcuQuebec and Compute Canada for computation facilities.

References

1. M. Lundstrom, *Fundamentals of Carrier Transport* (Addison-Wesley Publishing Company, Reading, MA, 1990), vol. 10
2. A. Asenov, IEEE Trans. Electron. Devices **45**, 2505 (1998)
3. A. Asenov, A.R. Brown, J.H. Davies, S. Kaya, G. Slavcheva, IEEE Trans. Electron. Devices **50**, 1837 (2003)
4. S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, Cambridge, 1995)
5. P.C. Martin, J. Schwinger, Theory of many-particle systems. I. Phys. Rev. **115**, 1342 (1959)
6. J. Schwinger, J. Math. Phys. **2**, 407 (1961)
7. L.V. Keldysh, Sov. Phys. JETP **20**, 1018 (1965)
8. L.P. Kadanoff, G. Baym, *Quantum Statistical Mechanics* (Benjamin/Cummings, 1962)
9. C. Caroli, R. Combescot, P. Nozieres, D. Saint-James, J. Phys. C: Solid St. Phys. **4**, 916 (1971)
10. Y. Meir, N.S. Wingreen, Phys. Rev. Lett. **68**, 2512 (1992)
11. J. Taylor, H. Guo, J. Wang, Phys. Rev. B **63**, 245407 (2001)
12. J. Taylor, H. Guo, J. Wang, Phys. Rev. B **63**, 121104 (2001)
13. Z. Ren, R. Venugopal, S. Datta, M.S. Lundstrom, D. Jovanovic, J.G. Fossum, IEDM Tech. Dig. 715 (2000)
14. Z. Ren, R. Venugopal, S. Datta, M.S. Lundstrom, IEDM Tech. Dig. 5.4.1 (2001)
15. J. Guo, S. Datta, M. Lundstrom, IEEE Trans. Electron. Devices **51**, 172 (2004)
16. J. Guo, S. Datta, M.P. Anantram, M. Lundstrom, J. Comput. Electron. **3**, 373 (2004)
17. Y. Zheng, C. Rivas, R. Lake, K. Alam, T. Boykin, G. Klimeck, IEEE Trans. Electron. Devices **52**, 1097 (2005)
18. M. Luisier, A. Schenk, W. Fichtner, G. Klimeck, Phys. Rev. B **74**, 205323 (2006)
19. Y. Ouyang, Y. Yoon, J. Guo, IEEE Trans. Electron. Devices **54**, 2223 (2007)
20. Y. Yoon, K. Ganapathi, S. Salahuddin, Nano. Lett. **11**, 3768 (2011)
21. F. Liu, Y. Wang, X. Liu, J. Wang, H. Guo, IEEE Trans. Electron. Devices **61**, 3871 (2014)
22. A.C. Seabaugh, Q. Zhang, Proc. IEEE **98**, 2095 (2010)
23. A.M. Ionescu, H. Riel, Nature **479**, 329 (2011)
24. M. Luisier, G. Klimeck, IEEE Electr. Dev. Lett. **30**, 602 (2009)
25. M. Luisier, G. Klimeck, J. Appl. Phys. **107**, 084507 (2010)
26. P. Zhao, J. Chauhan, J. Guo, Nano. Lett. **9**, 684 (2009)
27. S.K. Chin, D. Seah, K.T. Lam, G.S. Samudra, G. Liang, IEEE Trans. Electron. Devices **57**, 3144 (2010)

28. A. Szabo, R. Rhyner, M. Luisier, Proc. IEEE IEDM 30.4.1 (2014)
29. F. Liu, J. Wang, H. Guo, Nanotechnology **26**, 175201 (2015)
30. H. Ilatikhameneh, Y. Tan, B. Novakovic, G. Klimeck, R. Rahman, J. Appenzeller, IEEE. J. Exploratory Solid-State Comput. Devices Circuits **1**, 12 (2015)
31. F. Liu, J. Wang, H. Guo, IEEE Trans. Elec. Dev. **63**, 311 (2016)
32. M. Born, R. Oppenheimer, Zur Quantentheorie der Molekeln. Ann. Phys. **389**, 457 (1927)
33. P. Hohenberg, W. Kohn, Phys. Rev. **136**, B864 (1964)
34. W. Kohn, L.J. Sham, Phys. Rev. **140**, A1133 (1965)
35. J.P. Perdew, A. Zunger, Phys. Rev. B **23**, 5048 (1981)
36. J.P. Perdew, K. Burke, M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996)
37. H.L. Skriver, *The LMTO Method* (Springer, Berlin, 1984)
38. I. Turek, *Electronic Structure of Disorder Alloys Surfaces and Interfaces* (Kluwer Academic Publishers, Boston, 1997)
39. O.K. Anderson, Phys. Rev. Lett. **53**, 2571 (1984)
40. O.K. Anderson, Phys. Rev. B **34**, 5253 (1986)
41. P.C. Slater, G.F. Koster, Phys. Rev. **94**, 1498 (1954)
42. P.R. Wallace, The band theory of graphite. Phys. Rev. **71**, 622–634 (1947)
43. T.B. Boykin, G. Klimeck, F. Oyafuso, Phys. Rev. B **69**, 115201 (2004)
44. R. Landauer, IBM J. Rev. Dev. **1**, 223 (1957)
45. M. Buttiker, Y. Imry, R. Landauer, S. Pinhas, Phys. Rev. B **31**, 6207 (1985)
46. M.P. Lopez Sancho, J.M. Lopez Sancho, J. Rubio, J. Phys. F. **15**, 851 (1985)
47. J. Kudrnovský, I. Drchal, Phys. Rev. B **41**, 7515 (1990)
48. J.S. Chawla, F. Zahid, H. Guo, D. Gall, Appl. Phys. Lett. **97**, 132106 (2010)
49. F. Zahid, Y. Ke, D. Gall, H. Guo, Phys. Rev. B **81**, 045406 (2010)
50. B. Velický, S. Kirkpatrick, H. Ehrenreich, Phys. Rev. **175**, 747 (1968)
51. Y. Zhu, L. Liu, H. Guo, Phys. Rev. B **88**, 205415 (2013)
52. Y. Zhu, L. Liu, H. Guo, Phys. Rev. B **88**, 085420 (2013)
53. Q. Shi, L. Zhang, Y. Zhu, L. Liu, M. Chan, H. Guo, Proc. IEEE IEDM 30.6.1 (2014)
54. H. Fang, M. Tosun, G. Seol, T.C. Chang, K. Takei, J. Guo, A. Javey, Nano. Lett. **13**, 1991 (2013)
55. N.H. Pour, Y. Anugrah, S. Wu, X. Xu, S.J. Koester, DRC Tech. Dig. 101 (2013)
56. Q. Yu, L.A. Jauregui, W. Wu, R. Colby, J. Tian, Z. Su, H. Cao, Z. Liu, D. Pandey, D. Wei, T. F. Chung, P. Peng, N.P. Guisinger, E.A. Stach, J. Bao, S. Pei, Y.P. Chen, Nat. Matter **10**, 443 (2011)
57. A. Asenov, F. Adamu-Lema, X. Wang, S.M. Amoroso, IEEE Trans. Elec. Dev. **61**, 2745 (2014)
58. T.B. Boykin, M. Luisier, A. Schenk, N. Kharche, G. Klimeck, IEEE Trans. Nano. **6**, 43 (2007)
59. Y. Ke, K. Xia, H. Guo, Phys. Rev. Lett. **105**, 236801 (2010)
60. J.S. Chawla, F. Zahid, H. Guo, D. Gall, Appl. Phys. Lett. **97**, 132106 (2010)
61. J. Maassen, H. Guo, Phys. Rev. Lett. **109**, 266803 (2012)
62. L. Ci, L. Song, C. Jin, D. Jariwala, D. Wu, Y. Li, A. Srivastava, Z.F. Wang, K. Storr, L. Balicas, F. Liu, P.M. Ajayan, Nat. Mater. **9**, 430 (2010)
63. NanoAcademic, Nanodsim Manual (2013)
64. F. Liu, Y. Wang, X. Liu, J. Wang, H. Guo, IEEE Electron. Device Lett. **36**, 1091 (2015)
65. K. Lam, X. Cao, J. Guo, IEEE Electron. Device Lett. **34**, 1331 (2013)
66. N. Ma, D. Jena, Appl. Phys. Lett. **102**, 132102 (2013)
67. J.W. Chang, C. Hobbs, Appl. Phys. Lett. **106**, 083509 (2015)
68. F. Liu, Q. Shi, J. Wang, H. Guo, Appl. Phys. Lett. **107**, 203501 (2015)
69. D. Akinwande, N. Petrone, J. Hone, Nat. Commun. **5**, 5678 (2014)
70. S. Agarwal, E. Yablonovitch, Proc. 69th Annu. DRC, 199 (2011)
71. X.W. Jiang, J. Gong, N. Xu, S.S. Li, L.W. Wang, Appl. Phys. Lett. **104**, 023512 (2014)
72. D. Sarkar, X. Xie, W. Liu, W. Cao, J. Kang, Y. Gong, S. Kraemer, M.P. Ajayan, K. Banerjee, Nature **526**, 91 (2015)

73. S. Datta, *Quantum Transport: Atom to Transistor* (Cambridge University Press, Cambridge, 2005)
74. F. Liu, Y. Wang, X. Liu, J. Wang, H. Guo, IEEE Electron. Device Lett. **36**, 1091 (2015)
75. R.F. Pierret, *Semiconductor Device Fundamentals* (Addison-Wesley, Boston, 1996)
76. K. Suzuki, T. Tanaka, Y. Tosaka, H. Horie, Y. Arimoto, IEEE Trans. Electron. Devices **40**, 2326 (1993)

Chapter 6
Quantum Transport Simulation of III-V TFETs with Reduced-Order $k \cdot p$ Method

Jun Z. Huang, Lining Zhang, Pengyu Long, Michael Povolotskyi and Gerhard Klimeck

Abstract III-V tunnel field-effect transistors (TFETs) offer great potentials in future low-power electronics application due to their steep subthreshold slope and large "on" current. Their 3D quantum transport study using non-equilibrium Green's function method is computationally very intensive, in particular when combined with multiband approaches such as the eight-band $\boldsymbol{k \cdot p}$ method. To reduce the numerical cost, an efficient reduced-order method is developed in this chapter and applied to study homojunction InAs and heterojunction GaSb–InAs nanowire TFETs. Device performances are obtained for various channel widths, channel lengths, crystal orientations, doping densities, source–pocket lengths, and strain conditions.

6.1 Introduction

Scaling the supply voltage enables reduction of power consumption of integrated circuits. In order to continue reducing the supply voltage without degrading the performance, steep subthreshold swing (SS) transistors are highly needed. Steep tunnel field-effect transistors (TFETs) can achieve sub-60 mV/dec SS at room temperature by using quantum mechanical band-to-band tunneling (BTBT) [1, 2]. However, TFETs generally suffer from low I_{ON} due to low tunneling probabilities. To enhance BTBT and increase I_{ON}, group III-V semiconductor-based TFETs are very attractive since III-V materials can provide low bandgap and small tunneling mass, and allow different band-edge alignments [1].

J.Z. Huang (✉) · P. Long · M. Povolotskyi · G. Klimeck
Network for Computational Nanotechnology, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA
e-mail: huang599@purdue.edu; junhuang1021@gmail.com

L. Zhang
Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Kowloon, Hong Kong

L. Zhang and M. Chan (eds.), *Tunneling Field Effect Transistor Technology*,
DOI 10.1007/978-3-319-31653-6_6

In order to achieve the best III-V TFET performances, it is required to systematically optimize various design parameters, such as the channel thickness, channel length, crystal orientations, and doping densities. In addition, many schemes have been proposed to further boost I_{ON}. The first scheme is to embed a pocket doping between the source and the channel [3, 4]. The pocket increases the electric field near the tunnel junction and thus improves the I_{ON} and SS. 2D quantum simulations have also been performed for this kind of device [5]. The second scheme is to replace the homojunction with a heterojunction, for instance, a GaSb/InAs broken-gap heterojunction [6, 7], or an InGaAs/InAs heterojunction [8]. Due to the band offset between the two materials, the tunneling barrier height and distance are greatly reduced. 2D and 3D quantum transport simulations have also been performed for these heterojunction TFETs [9–11]. Other schemes include strain engineering [12, 13], grading of the molar fraction in the source region [13, 14], adding a doped underlap layer between source and channel [15], and embedding a quantum well in the source [16].

To understand the device physics, predict the performance, and optimize the design parameters of these structures, an efficient quantum transport solver is highly needed. The BTBT process can be accurately accounted for by combining non-equilibrium Green's function (NEGF) approach [17] with tight-binding or eight-band $\boldsymbol{k} \cdot \boldsymbol{p}$ Hamiltonian. Unfortunately, these multiband NEGF studies require huge computational resources due to the large Hamiltonian matrix. To improve their efficiency, equivalent but greatly reduced tight-binding models can be constructed for silicon nanowires (SiNWs) [18], which greatly speed up the simulation of p-type SiNW MOSFETs even in the presence of phonon scattering. Recently, this method has been extended to simulate III-V nanowire MOSFETs and heterojunction TFETs [19]. Note that construction of the reduced tight-binding models requires sophisticated optimization process. A mode space $\boldsymbol{k} \cdot \boldsymbol{p}$ approach is also proposed for p-type SiNW MOSFETs and InAs TFETs [20], which has been employed to simulate strain-engineered and heterojunction nanowire TFETs [12, 13]. Though optimization process is not needed, this approach selects the modes only at the Γ point, i.e., at $k = 0$, which is inefficient to expand the modes that are far away from $k = 0$.

In this work, we propose to construct the reduced-order $\boldsymbol{k} \cdot \boldsymbol{p}$ models with multipoint expansion [21, 22]. We also extend this method to be able to simulate heterojunction devices. This efficient quantum transport solver is then applied to optimize device configurations such as crystal orientation, channel width, and channel length. Various performance boosters such as source pocket, heterojunction, and strain will be explored. Homojunction InAs and heterojunction GaSb/InAs nanowire TFETs will be the focus of this study.

The device structure is described in Sect. 6.2. The $\boldsymbol{k} \cdot \boldsymbol{p}$ method is developed in Sect. 6.3, where the eight-band $\boldsymbol{k} \cdot \boldsymbol{p}$ Hamiltonian and its matrix rotations are reviewed first. The rotated Hamiltonian is then discretized in a mixed real and spectral space. The accuracy of the $\boldsymbol{k} \cdot \boldsymbol{p}$ method is benchmarked by comparing the band structures with tight-binding method for several nanowire cross sections. The reduced-order NEGF method is developed in Sect. 6.4, where the reduced-order

NEGF equations are summarized first. Then, the problem of spurious bands, particular for the multipoint expansion, is identified. Afterward, a simple procedure to eliminate these spurious bands is proposed. The method is finally validated by checking the band structures as well as the *I–V* curves. In Sect. 6.5, extensive simulations are carried out to understand and optimize the TFETs under different application requirements. Conclusions are drawn in Sect. 6.6.

6.2 Device Structure

The n-type gate-all-around (GAA) nanowire TFET to be simulated is illustrated in Fig. 6.1. The nanowire is P++ doped in the source and N+ doped in the drain, while it is intrinsic in the channel. A thin layer of N++ doping is inserted between the source and the channel to form a source pocket. The nanowire is surrounded by the oxide layer, through which the gate controls the channel (and the pocket).

For homojunction TFETs in this study, all the source, pocket, channel, and drain are made of material InAs, because high "on" current is possible due to its small direct bandgap and light effective masses [23].

For heterojunction TFETs in this study, GaSb is used for the source, while InAs is used for the pocket, channel, and drain. These two materials form broken-gap heterojunction at the source–channel (or source–pocket) interface, though in reality staggered-gap heterojunction is formed due to lateral confinements [9, 10].

Different channel width, gate length, and pocket length will be studied. The channel crystal orientations will be varied from [100] (with (010) and (001) surfaces), [110] (with (−1,1,0) and (001) surfaces), to [111] (with (−1,1,0) and (−1, −1,2) surfaces). The uniaxial stress will be applied along the x direction, while the biaxial stress will be applied in the y and z directions.

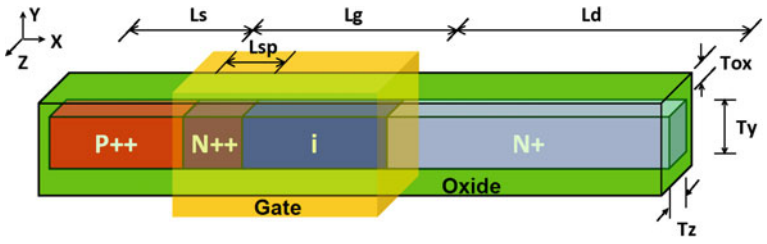

Fig. 6.1 GAA TFET with p-n-i-n doping profile. The transport direction is x, while the confinement directions are y and z. The source, source pocket, gate, and drain lengths are L_s, L_{sp}, L_g and L_d, respectively, with doping density in the source, pocket, and drain, N_s, N_{sp} and N_d, respectively. Nanowire width and thickness are T_z and T_y. Oxide layer thickness is denoted by T_{ox}; its dielectric constant is ε_{ox}

6.3 The $k \cdot p$ Method

6.3.1 The Eight-Band $k \cdot p$ Hamiltonian

To describe the band structure involving both the conduction and valence bands of III-V compound semiconductor materials, a widely used approach is the eight-band $\boldsymbol{k} \cdot \boldsymbol{p}$ model. When the eight basis functions are chosen to be spin-up and spin-down s and p atomic orbital-like states, the zincblende Hamiltonian can be written as [24–26],

$$\boldsymbol{H}^8 = \begin{pmatrix} \boldsymbol{H}^4 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H}^4 \end{pmatrix} + \begin{pmatrix} \boldsymbol{G}_{so} & \boldsymbol{\Gamma} \\ -\boldsymbol{\Gamma}^* & \boldsymbol{G}_{so}^* \end{pmatrix} + \begin{pmatrix} \boldsymbol{H}_{str}^4 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H}_{str}^4 \end{pmatrix}, \tag{6.1}$$

where the first part is spin-independent, the second part accounts for spin–orbit coupling, and the last part is deformation potential contribution due to strain.

With operator ordering (for heterostructures) taken into account, the four-band Hamiltonian $\boldsymbol{H}^4$ is

$$\boldsymbol{H}^4 = \begin{pmatrix} E_c \boldsymbol{I}_1 + \boldsymbol{H}_{cc} & \boldsymbol{H}_{cv} \\ \boldsymbol{H}_{vc} & E_v' \boldsymbol{I}_3 + \boldsymbol{H}_{vv} \end{pmatrix}, \tag{6.2}$$

where $\boldsymbol{I}_1$ and $\boldsymbol{I}_3$ are the 1×1 and 3×3 identity matrices, and

$$\boldsymbol{H}_{cc} = k_x A_c k_x + k_y A_c k_y + k_z A_c k_z, \tag{6.3}$$

$$\boldsymbol{H}_{cv} = \begin{pmatrix} iP^+ k_x + ik_x P^- & iP^+ k_y + ik_y P^- & iP^+ k_z + ik_z P^- \end{pmatrix}, \tag{6.4}$$

$$\boldsymbol{H}_{vc} = \begin{pmatrix} -ik_x P^+ - iP^- k_x & -ik_y P^+ - iP^- k_y & -ik_z P^+ - iP^- k_z \end{pmatrix}^T, \tag{6.5}$$

$$\boldsymbol{H}_{vv} = \begin{pmatrix} k_x L k_x + k_y M k_y + k_z M k_z & k_x N^+ k_y + k_y N^- k_x & k_x N^+ k_z + k_z N^- k_x \\ k_y N^+ k_x + k_x N^- k_y & k_y L k_y + k_z M k_z + k_x M k_x & k_y N^+ k_z + k_z N^- k_y \\ k_z N^+ k_x + k_x N^- k_z & k_z N^+ k_y + k_y N^- k_z & k_z L k_z + k_x M k_x + k_y M k_y \end{pmatrix}. \tag{6.6}$$

Here, the parameter $E_c = E_v + E_g$ is the conduction band edge with E_v being the valence band edge and E_g the bandgap. $E_v' = E_v - \Delta/3$ is the valence band edge in the absence of spin–orbit coupling, with Δ being the spin–orbit split-off energy. P is proportional to the momentum matrix element and can be evaluated by its equivalent energy $E_p = 2m_0 P^2/\hbar^2$. A_c is determined from the conduction band effective mass m_c^*,

$$A_c = \frac{\hbar^2}{2m_c^*} - \frac{2P^2}{3E_g} - \frac{P^2}{3(E_g + \Delta)}. \tag{6.7}$$

The parameters L, M, and N are related to the Luttinger parameters γ_1, γ_2, and γ_3,

$$L = -\frac{\hbar^2}{2m_0}(\gamma_1 + 4\gamma_2) + \frac{P^2}{E_g}, \tag{6.8}$$

$$M = -\frac{\hbar^2}{2m_0}(\gamma_1 - 2\gamma_2), \tag{6.9}$$

$$N = -\frac{\hbar^2}{2m_0}(6\gamma_3) + \frac{P^2}{E_g}. \tag{6.10}$$

While the widely used symmetrized operator ordering evenly divides the terms leading to $P^+ = P^- = P/2$ and $N^+ = N^- = N/2$, the correct Burt–Foreman ordering [27] divides the terms according to different bands' contribution, which leads to $P^+ = P, P^- = 0, N^- = M - \hbar^2/2m_0$ and $N^+ = N - N^-$.

The spin–orbit terms $\boldsymbol{G}_{so}$ and $\boldsymbol{\Gamma}$ are

$$\boldsymbol{G}_{so} = \frac{\Delta}{3}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{\Gamma} = \frac{\Delta}{3}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -i \\ 0 & -1 & i & 0 \end{pmatrix}. \tag{6.11}$$

The strain Hamiltonian is [28]

$$\boldsymbol{H}^4_{str} = \begin{pmatrix} \boldsymbol{H}^{cc}_{str} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H}^{vv}_{str} \end{pmatrix}, \tag{6.12}$$

where

$$\boldsymbol{H}^{cc}_{str} = a_c\left(\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}\right), \tag{6.13}$$

$$\boldsymbol{H}^{vv}_{str} = \begin{pmatrix} l\varepsilon_{xx} + m\left(\varepsilon_{yy} + \varepsilon_{zz}\right) & n\varepsilon_{xy} & n\varepsilon_{xz} \\ n\varepsilon_{xy} & l\varepsilon_{yy} + m(\varepsilon_{xx} + \varepsilon_{zz}) & n\varepsilon_{yz} \\ n\varepsilon_{xz} & n\varepsilon_{yz} & l\varepsilon_{zz} + m\left(\varepsilon_{xx} + \varepsilon_{yy}\right) \end{pmatrix}. \tag{6.14}$$

Here, $\left(\varepsilon_{xx}, \varepsilon_{yy}, \varepsilon_{zz}, 2\varepsilon_{yz}, 2\varepsilon_{xz}, 2\varepsilon_{xy}\right)$ is the strain vector in Voigt's notation, a_c is the deformation potential constant for the conduction band, and $m = a_v - b, l = a_v + 2b, n = \sqrt{3}d$ and a_v, b, d are the Pikus–Bir deformation potential constants for the valence bands. Note that we only keep the k independent terms.

The $\boldsymbol{k} \cdot \boldsymbol{p}$ parameters for III-V compounds and their alloys can be found in [29].

The $\boldsymbol{k} \cdot \boldsymbol{p}$ Hamiltonian matrix defined in the above is in terms of $\boldsymbol{k}$ in the crystal coordinate system (CCS). In practice, nanostructures can grow in different crystal directions, and thus, the quantization directions and periodic directions are aligned

with device coordinate system (DCS). Therefore, it is more convenient to work in the DCS, and it requires coordinate transformation of the Hamiltonian matrix.

We first define a 3 × 3 unitary rotation matrix from DCS to CCS $\boldsymbol{R}_{D\to C}$, so that the 3 × 1 $\boldsymbol{k}$ vectors in the CCS and DCS, i.e., $\boldsymbol{k}_C$ and $\boldsymbol{k}_D$, are related by

$$\boldsymbol{k}_C = \boldsymbol{R}_{D\to C} \cdot \boldsymbol{k}_D. \tag{6.15}$$

Note that the rows of $\boldsymbol{R}_{D\to C}$ are the coordinates of the CCS unit vectors in the DCS.

Then, we rotate the $\boldsymbol{k} \cdot \boldsymbol{p}$ matrix element by element. Each of the second order in $\boldsymbol{k}$ terms in the CCS is of the form $\boldsymbol{k}_C^T \boldsymbol{H}_C^{(2)} \boldsymbol{k}_C$. Substituting (6.15), we have,

$$\boldsymbol{k}_C^T \boldsymbol{H}_C^{(2)} \boldsymbol{k}_C = \left(\boldsymbol{k}_D^T \boldsymbol{R}_{D\to C}^T\right) \boldsymbol{H}_C^{(2)} \left(\boldsymbol{R}_{D\to C} \boldsymbol{k}_D\right) = \boldsymbol{k}_D^T \boldsymbol{H}_D^{(2)} \boldsymbol{k}_D. \tag{6.16}$$

From above, we can identify

$$\boldsymbol{H}_D^{(2)} = \boldsymbol{R}_{D\to C}^T \boldsymbol{H}_C^{(2)} \boldsymbol{R}_{D\to C}. \tag{6.17}$$

Each of the first order in $\boldsymbol{k}$ terms in the CCS is of the form $\boldsymbol{k}_C^T \boldsymbol{H}_{C,R}^{(1)} + \boldsymbol{H}_{C,L}^{(1)} \boldsymbol{k}_C$, where $\boldsymbol{H}_{C,R}^{(1)}$ is a 3 × 1 matrix and $\boldsymbol{H}_{C,L}^{(1)}$ is an 1 × 3 matrix. Substituting (6.15), we have,

$$\begin{aligned} \boldsymbol{k}_C^T \boldsymbol{H}_{C,R}^{(1)} + \boldsymbol{H}_{C,L}^{(1)} \boldsymbol{k}_C &= \boldsymbol{k}_D^T \boldsymbol{R}_{D\to C}^T \boldsymbol{H}_{C,R}^{(1)} + \boldsymbol{H}_{C,L}^{(1)} \boldsymbol{R}_{D\to C} \boldsymbol{k}_D \\ &= \boldsymbol{k}_D^T \boldsymbol{H}_{D,R}^{(1)} + \boldsymbol{H}_{D,L}^{(1)} \boldsymbol{k}_D. \end{aligned} \tag{6.18}$$

From above, we can identify

$$\boldsymbol{H}_{D,R}^{(1)} = \boldsymbol{R}_{D\to C}^T \boldsymbol{H}_{C,R}^{(1)}, \tag{6.19}$$

and

$$\boldsymbol{H}_{D,L}^{(1)} = \boldsymbol{H}_{C,L}^{(1)} \boldsymbol{R}_{D\to C}. \tag{6.20}$$

The $\boldsymbol{k}$ independent terms, such as the band edges and spin–orbit constants, do not need to rotate.

The strain and stress components are usually set in the DCS; however, the strain components in the CCS are those that enter the $\boldsymbol{k} \cdot \boldsymbol{p}$ Hamiltonian [30].

The rotation for the strain is given by

$$\boldsymbol{\epsilon}_{C,3\times 3} = \boldsymbol{R}_{D\to C} \cdot \boldsymbol{\epsilon}_{D,3\times 3} \cdot \boldsymbol{R}_{D\to C}^{-1}, \tag{6.21}$$

where $\boldsymbol{\epsilon}_{C,3\times 3}$ and $\boldsymbol{\epsilon}_{D,3\times 3}$ are the 3 × 3 strain matrices in the CCS and DCS, respectively. Similar rotation holds for $\boldsymbol{\sigma}_{C,3\times 3}$ and $\boldsymbol{\sigma}_{D,3\times 3}$, the stress matrices in the CCS and DCS.

It is convenient to transform directly the strain vector by

$$\boldsymbol{\epsilon}_{C,6} = \boldsymbol{R}_{6,D\to C} \cdot \boldsymbol{\epsilon}_{D,6}, \tag{6.22}$$

where $\boldsymbol{\epsilon}_{C,6}$ and $\boldsymbol{\epsilon}_{D,6}$ are the 6 × 1 strain vectors in the CCS and DCS, respectively. $\boldsymbol{R}_{6,D\to C}$ is the 6 × 6 transformation matrix, whose elements can be found by expanding Eq. (6.21). Similar rotation holds for $\boldsymbol{\sigma}_{C,6}$ and $\boldsymbol{\sigma}_{D,6}$, the stress vectors in the CCS and DCS.

Finally, $\boldsymbol{\sigma}_{C,6}$ can be converted to $\boldsymbol{\epsilon}_{C,6}$ via three elastic constants C_{11}, C_{12}, and C_{44},

$$\begin{pmatrix} \varepsilon_{C,xx} \\ \varepsilon_{C,yy} \\ \varepsilon_{C,zz} \\ \varepsilon_{C,yz} \\ \varepsilon_{C,xz} \\ \varepsilon_{C,xy} \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{12} & C_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & 2C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & 2C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & 2C_{44} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\sigma}_{C,xx} \\ \boldsymbol{\sigma}_{C,yy} \\ \boldsymbol{\sigma}_{C,zz} \\ \boldsymbol{\sigma}_{C,yz} \\ \boldsymbol{\sigma}_{C,xz} \\ \boldsymbol{\sigma}_{C,xy} \end{pmatrix}. \tag{6.23}$$

6.3.2 The Discretized Hamiltonian

For nanostructures, the periodicity is broken by the finite sizes and the external potentials. The eigen states can be found by solving the following coupled differential equation for envelop function $\boldsymbol{F}_m$ $(m = 1, 2, \ldots, 8)$,

$$\sum_{n=1}^{8} \left[\boldsymbol{H}_{mn}^{8}(-i\nabla) + V(\boldsymbol{r})\delta_{mn}\right]\boldsymbol{F}_n(\boldsymbol{r}) = E\boldsymbol{F}_m(\boldsymbol{r}) \tag{6.24}$$

where $V(\boldsymbol{r})$ is the slowly varying perturbed potential distribution, and operator $\boldsymbol{H}_{mn}^{8}(-i\nabla)$ is the element of $\boldsymbol{H}^{8}(\boldsymbol{k})$ with $\boldsymbol{k}$ replaced by the differential operator $-i\nabla$.

In order to solve (6.24) numerically, the operator needs to be discretized first. To have a discretized form that is compact and valid for arbitrary nanowire orientation, we rewrite the eight-band $\boldsymbol{k}\cdot\boldsymbol{p}$ operator (considering operator ordering) in (6.24) as,

$$\boldsymbol{H} = \sum_{\alpha,\beta=x,y,z} \partial_\alpha \boldsymbol{H}_{\alpha,\beta}\partial_\beta + \sum_{\alpha=x,y,z} \left(\boldsymbol{H}_{\alpha,L}\partial_\alpha + \partial_\alpha \boldsymbol{H}_{\alpha,R}\right) + \boldsymbol{H}_0 \tag{6.25}$$

where the matrices $\boldsymbol{H}_{\alpha,\beta}$, $\boldsymbol{H}_{\alpha,L}$, $\boldsymbol{H}_{\alpha,R}$, and $\boldsymbol{H}_0$ are the material- and orientation-dependent coefficients containing contributions from Löwdin's renormalization, spin–orbit interaction, and strain.

In Ref. [21], finite difference method (FDM) is adopted and it results in extremely sparse matrices. Therefore, the Bloch modes can be obtained efficiently with sparse matrix solvers. In fact, with shift-and-invert strategy implemented, the Krylov subspace-based eigenvalue solver converges very quickly, as the eigenvalues of interest (close to the valence band top) distribute in a very small area. However, it is found that the Krylov subspace method is less efficient in the eight-band case. The reason is that the eigenvalues of interest distribute over a larger area, as both conduction and valence bands are to be sought and between them there is a bandgap.

Therefore, the method used in Ref. [20] is employed, which is also generalized to arbitrary crystal orientations and to heterojunctions here. In this method, the transport direction is still discretized by FDM, while the transverse directions are discretized by spectral method. Spectral method has high spectral accuracy (i.e., the error decreases exponentially with the increase of discretization points N) if the potential distribution is smooth [31]. This is true for devices that do not have any explicit impurities or surface roughness. So, the Hamiltonian matrix size of a layer, i.e., N_t, can be kept very small (although it is less sparse or even dense), making direct solution of the eigenvalue problem possible.

To discretize the operator (6.25), the longitudinal component of the unknown envelope function is discretized with second-order central FDM,

$$\partial_x\left(\boldsymbol{H}_{x,R}\boldsymbol{\psi}\right)|_{x=x_i} \approx \frac{\boldsymbol{H}_{x,R}(x_{i+1})\boldsymbol{\psi}(x_{i+1}) - \boldsymbol{H}_{x,R}(x_{i-1})\boldsymbol{\psi}(x_{i-1})}{2\Delta x}, \tag{6.26}$$

$$\boldsymbol{H}_{x,L}\partial_x\boldsymbol{\psi}|_{x=x_i} \approx \boldsymbol{H}_{x,L}(x_i)\frac{\boldsymbol{\psi}(x_{i+1}) - \boldsymbol{\psi}(x_{i-1})}{2\Delta x}, \tag{6.27}$$

$$\begin{aligned}\partial_x\left(\boldsymbol{H}_{x,x}\partial_x\boldsymbol{\psi}\right)|_{x=x_i} \approx & \frac{\boldsymbol{H}_{x,x}(x_{i+1}) + \boldsymbol{H}_{x,x}(x_i)}{2(\Delta x)^2}\boldsymbol{\psi}(x_{i+1}) \\ & - \frac{\boldsymbol{H}_{x,x}(x_{i+1}) + 2\boldsymbol{H}_{x,x}(x_i) + \boldsymbol{H}_{x,x}(x_{i-1})}{2(\Delta x)^2}\boldsymbol{\psi}(x_i) \\ & + \frac{\boldsymbol{H}_{x,x}(x_{i-1}) + \boldsymbol{H}_{x,x}(x_i)}{2(\Delta x)^2}\boldsymbol{\psi}(x_{i-1}),\end{aligned} \tag{6.28}$$

where Δx is the grid spacing.

The transversal components are expanded using Fourier series [20], i.e.,

$$\phi_{p,q}(y_m, z_n) = \frac{2}{\sqrt{N_y N_z}}\sin\left(k_p y_m\right)\sin\left(k_q z_n\right), \tag{6.29}$$

where N_y and N_z are the number of real space grid points in the y and z directions, respectively, m and n $\left(1 \le m \le N_y, 1 \le n \le N_z\right)$ are the coordinates of the Rth grid

point in real space, and p and q $\left(1 \le p \le N_y, 1 \le q \le N_z\right)$ are the coordinates of the Sth grid point in the Fourier space,

$$k_p = \frac{p\pi}{T_y}, \quad k_q = \frac{q\pi}{T_z}, \tag{6.30}$$

where T_y (T_z) is the nanowire thickness in the y (z) direction. Note that hard wall boundary condition is enforced at the interfaces between the oxide layer and the semiconductor nanowire when the basis function (6.29) is used.

Operating (6.25) on (6.29), multiplying the result with (6.29), and performing integrations, we get the discretized form. It is block tridiagonal,

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{D}_1 & \boldsymbol{T}_{1,2} & \boldsymbol{0} & & \\ \boldsymbol{T}_{1,2}^{\dagger} & \boldsymbol{D}_2 & \boldsymbol{T}_{2,3} & & \\ & \ddots & \ddots & \ddots & \\ & & \boldsymbol{T}_{N_x-2,N_x-1}^{\dagger} & \mathbf{D}_{N_x-1} & \boldsymbol{T}_{N_x-1,N_x} \\ & & \boldsymbol{0} & \boldsymbol{T}_{N_x-1,N_x}^{\dagger} & \boldsymbol{D}_{N_x} \end{pmatrix}, \tag{6.31}$$

where $\boldsymbol{D}_i$ is the on-site Hamiltonian for layer i $(1 \le i \le N_x)$, $\boldsymbol{T}_{i,i+1}$ $(1 \le i \le N_x - 1)$ is the coupling Hamiltonian between adjacent layers, and N_x is the number of grids in the longitudinal direction x.

The (S, S') block of $\boldsymbol{D}_i$ can be written down using very simple prescription,

$$\begin{aligned} \boldsymbol{D}_i^{S,S'} = & \left[\left(\boldsymbol{H}_{y,L}^i + \boldsymbol{H}_{y,R}^i\right)\frac{4k_p'}{\pi}\frac{p}{p^2 - p'^2}\right]\boldsymbol{\delta}_{p+p',odd}\boldsymbol{\delta}_{q,q'} \\ & + \left[\left(\boldsymbol{H}_{z,L}^i + \boldsymbol{H}_{z,R}^i\right)\frac{4k_q'}{\pi}\frac{q}{q^2 - q'^2}\right]\boldsymbol{\delta}_{q+q',odd}\boldsymbol{\delta}_{p,p'} \\ & - \left[\left(\boldsymbol{H}_{y,z}^i + \boldsymbol{H}_{z,y}^i\right)\frac{4k_p'}{\pi}\frac{p}{p^2 - p'^2}\frac{4k_q'}{\pi}\frac{q}{q^2 - q'^2}\right]\boldsymbol{\delta}_{p+p',odd}\boldsymbol{\delta}_{q+q',odd} \\ & + \left[\boldsymbol{H}_0^i + \left(\boldsymbol{H}_{x,x}^{i+1} + 2\boldsymbol{H}_{x,x}^i + \boldsymbol{H}_{x,x}^{i-1}\right)\frac{1}{2(\Delta x)^2} + \boldsymbol{H}_{y,y}^i k_p^2 + \boldsymbol{H}_{z,z}^i k_q^2\right]\boldsymbol{\delta}_{p,p'}\boldsymbol{\delta}_{q,q'}, \end{aligned} \tag{6.32}$$

where (p, q) and (p', q') are the coordinates of the Sth and S'th grid points, respectively. $\boldsymbol{\delta}$ is Kronecker delta function, for instance, $\boldsymbol{\delta}_{q+q',odd}$ is equal to 1 (0) if $q + q'$ is an odd (even) number.

Similarly, the (S, S') block of $\boldsymbol{T}_{i,i+1}$ can be written as,

$$\begin{aligned}\boldsymbol{T}_{i,i+1}^{S,S'} = &\left[-\left(\boldsymbol{H}_{x,x}^{i+1}+\boldsymbol{H}_{x,x}^{i}\right)\frac{1}{2(\Delta x)^2}+\left(\boldsymbol{H}_{x,R}^{i+1}+\boldsymbol{H}_{x,L}^{i}\right)\frac{1}{2\Delta x}\right]\boldsymbol{\delta}_{p,p'}\boldsymbol{\delta}_{q,q'}\\ &-\left[\left(\boldsymbol{H}_{x,y}^{i+1}+\boldsymbol{H}_{y,x}^{i}\right)\frac{1}{2\Delta x}\frac{4k'_p}{\pi}\frac{p}{p^2-p'^2}\right]\boldsymbol{\delta}_{p+p',odd}\boldsymbol{\delta}_{q,q'}\\ &-\left[\left(\boldsymbol{H}_{x,z}^{i+1}+\boldsymbol{H}_{z,x}^{i}\right)\frac{1}{2\Delta x}\frac{4k'_q}{\pi}\frac{q}{q^2-q'^2}\right]\boldsymbol{\delta}_{q+q',odd}\boldsymbol{\delta}_{p,p'}.\end{aligned} \tag{6.33}$$

In this work, we use $\Delta x = 0.2\,\text{nm}$ and have limited S to be $1 \le S \le 183$ by employing the index scheme in [20]. This means 183 Fourier series are used to expand each wave function component, which is found to be sufficient. The dimension of $\boldsymbol{D}_i$ is thus $N_t = 183 \times 8 = 1464$.

6.3.3 Comparison with TB Results

Since $\boldsymbol{k}\cdot\boldsymbol{p}$ method is only valid in a small region around the Γ point, there is a concern whether the $\boldsymbol{k}\cdot\boldsymbol{p}$ method is accurate for small nanostructures. A comparison with full-band tight-binding (TB) results will help answer this question. In order to have a fair comparison for confined structures, the $\boldsymbol{k}\cdot\boldsymbol{p}$ parameters are fit to the bulk TB calculations. At first, the bulk band structure is computed using the sp^3s^* spin–orbit TB model, from which we have the bandgap, spit-off energy, electron effective mass, and heavy hole and light hole effective masses in both the [100] and [111] directions. These immediately determine $\boldsymbol{k}\cdot\boldsymbol{p}$ parameters $E_g, \Delta, m_c^*/m_0, \gamma_1, \gamma_2$, and γ_3 [24],

$$\frac{m_0}{m_{hh}^*(100)} = \gamma_1 - 2\gamma_2, \tag{6.34}$$

$$\frac{m_0}{m_{lh}^*(100)} = \gamma_1 + 2\gamma_2, \tag{6.35}$$

$$\frac{m_0}{m_{hh}^*(111)} = \gamma_1 - 2\gamma_3, \tag{6.36}$$

$$\frac{m_0}{m_{lh}^*(111)} = \gamma_1 + 2\gamma_3. \tag{6.37}$$

The remaining parameter E_p is then slightly reduced from experiment value so as to avoid spurious solution [32]. The fitted parameters for materials InAs and GaSb

Table 6.1 Material parameters for InAs and GaSb at $T = 300$ K

Parameters	Eg (eV)	Δ (eV)	m_c^*/m_0	γ_1	γ_2	γ_3	Ep (eV)	VBO (eV)
InAs	0.368	0.381	0.024	19.20	8.226	9.033	18.1	0
GaSb	0.751	0.748	0.042	13.27	4.97	5.978	21.2	0.56

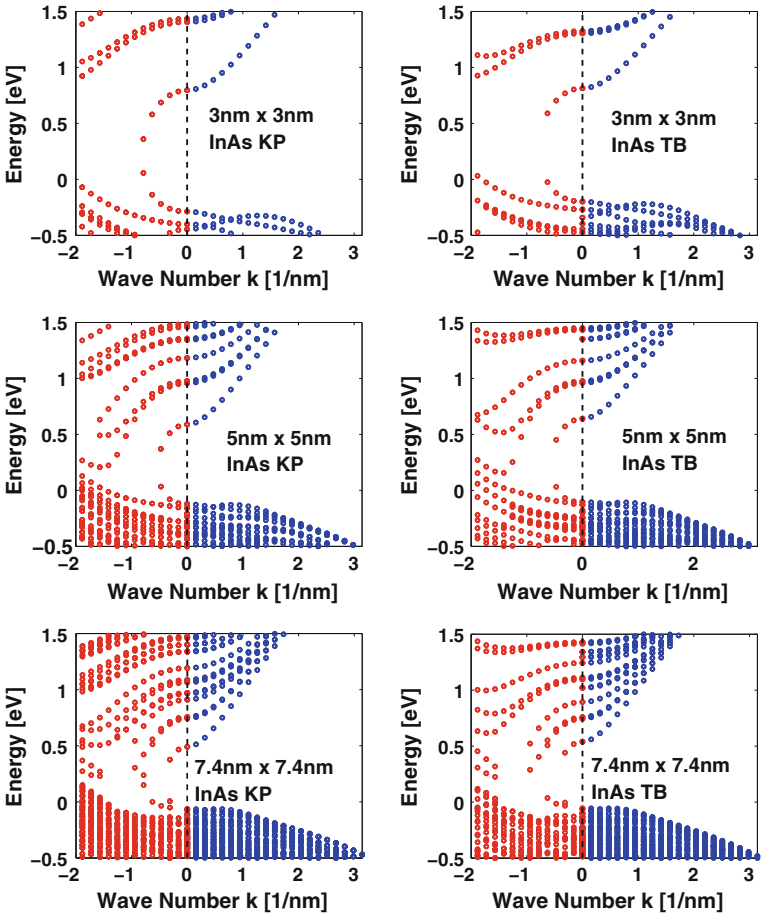

Fig. 6.2 Complex and real band structures of a 3 nm × 3 nm (*top*), a 5 nm × 5 nm (*middle*), and a 7.4 nm × 7.4 nm (*bottom*) cross-sectional InAs nanowire in the [100] orientation (with (010) and (001) surfaces). *Left* Eight-band $\boldsymbol{k} \cdot \boldsymbol{p}$ results; *right* sp^3s^* spin–orbit TB results

used in this work are list in Table 6.1, which slightly differ from those in [29]. The valence band offset (VBO) of GaSb relative to InAs is taken from [29].

Figure 6.2 compares the eight-band $\boldsymbol{k} \cdot \boldsymbol{p}$ and sp^3s^* spin–orbit TB band structures of three InAs nanowires with 3 nm × 3 nm, 5 nm × 5 nm, and 7.4 nm × 7.4 nm cross

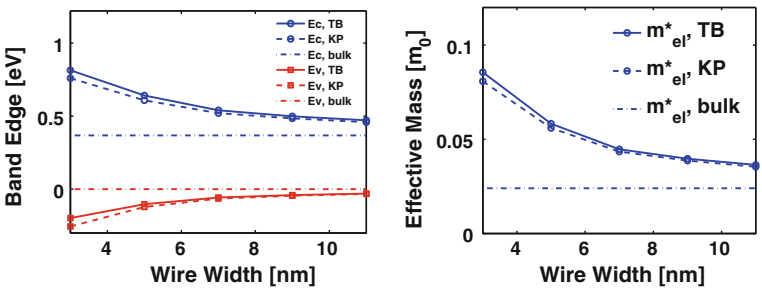

Fig. 6.3 Band edges (*left*) and electron effective mass (*right*) as functions of InAs wire cross section

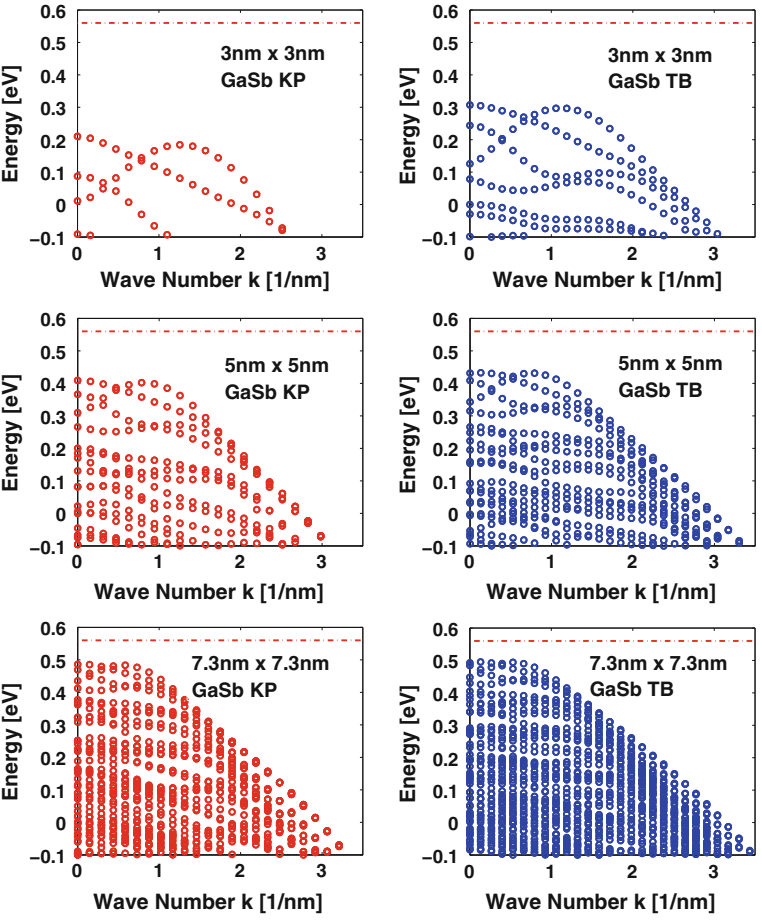

Fig. 6.4 Valence band structures of a 3 nm × 3 nm (*top*), a 5 nm × 5 nm (*middle*), and a 7.3 nm × 7.3 nm (*bottom*) cross-sectional GaSb nanowire in the [100] orientation (with (010) and (001) surfaces). *Left* Eight-band $\boldsymbol{k} \cdot \boldsymbol{p}$ results; *right* sp^3s^* spin–orbit TB results. The bulk valence band edge is shown in *dash-dot line*

sections, respectively. For $\boldsymbol{k}\cdot\boldsymbol{p}$, hard wall boundaries are imposed at the four surfaces, while for TB, the surface atoms are passivated with hydrogen atoms. Good matches are observed except for the 3 nm × 3 nm case, where the $\boldsymbol{k}\cdot\boldsymbol{p}$ model has larger separation of subbands though the bandgaps are close.

The band edges and effective masses as functions of nanowire cross-sectional size are plotted in Fig. 6.3. The $\boldsymbol{k}\cdot\boldsymbol{p}$ and TB results match quite well, except that $\boldsymbol{k}\cdot\boldsymbol{p}$ band edges are slightly shifted downward for small nanowires. These two models predict the same trends, i.e., as the nanowire size decreases, both the bandgap and the electron effective masses increase.

Figure 6.4 compares $\boldsymbol{k}\cdot\boldsymbol{p}$ and TB band structures of three GaSb nanowires with 3 nm × 3 nm, 5 nm × 5 nm, and 7.3 nm × 7.3 nm cross sections, respectively. Only valence band is shown since it is the most relevant for transport in heterojunction GaSb/InAs TFET application here. Similar to InAs case, quantitative matches are observed except for the 3 nm × 3 nm case where $\boldsymbol{k}\cdot\boldsymbol{p}$ model predicts lower valence band edge and larger subband energies.

6.4 Reduced-Order NEGF Method

6.4.1 Reduced-Order NEGF Equations

The NEGF equations for the retarded and lesser Green's function, $\boldsymbol{G}^R$ and $\boldsymbol{G}^<$, in the mixed real and Fourier space can be written as,

$$\left[E\boldsymbol{I}-\boldsymbol{H}-\boldsymbol{V}-\boldsymbol{\Sigma}^R(E)\right]\boldsymbol{G}^R(E)=\boldsymbol{I}, \tag{6.38}$$

$$\boldsymbol{G}^<(E)=\boldsymbol{G}^R(E)\boldsymbol{\Sigma}^<(E)\boldsymbol{G}^{R\dagger}(E), \tag{6.39}$$

where $\boldsymbol{H}$ is the block three-diagonal $\boldsymbol{k}\cdot\boldsymbol{p}$ Hamiltonian of the isolated device, $\boldsymbol{V}$ is potential term that is block diagonal, and $\boldsymbol{\Sigma}^R$ ($\boldsymbol{\Sigma}^<$) is the retarded (lesser) self-energy matrix due to the semi-infinite leads, which is nonzero only in the first and last blocks. Phonon scattering has a very modest effect on the *I*–*V* curve [12], and coherent transport is sufficient for III-V homojunction and heterojunction TFETs with direct bandgap [33, 34]; thus, it is excluded in this work.

As the matrices involved are very large, to solve $\boldsymbol{G}^R$ and $\boldsymbol{G}^<$ efficiently for many different energy E, the reduced-order matrix equations can be constructed,

$$\left[E\widetilde{\boldsymbol{I}}-\widetilde{\boldsymbol{H}}-\widetilde{\boldsymbol{V}}-\widetilde{\boldsymbol{\Sigma}}^R(E)\right]\widetilde{\boldsymbol{G}}^R(E)=\widetilde{\boldsymbol{I}}, \tag{6.40}$$

$$\widetilde{\boldsymbol{G}}^<(E)=\widetilde{\boldsymbol{G}}^R(E)\widetilde{\boldsymbol{\Sigma}}^<(E)\widetilde{\boldsymbol{G}}^{R\dagger}(E), \tag{6.41}$$

and the reduced-order Green's functions $\widetilde{\boldsymbol{G}}^R(E)$ and $\widetilde{\boldsymbol{G}}^<(E)$ are to be solved. Here, the reduced Hamiltonian, potential, self-energy, and Green's function are

$$\begin{aligned} \widetilde{\boldsymbol{H}} &= \boldsymbol{U}^\dagger \boldsymbol{H}\boldsymbol{U}, \quad \widetilde{\boldsymbol{V}} = \boldsymbol{U}^\dagger \boldsymbol{V}\boldsymbol{U}, \\ \widetilde{\boldsymbol{\Sigma}}^{R,<}(E) &= \boldsymbol{U}^\dagger \boldsymbol{\Sigma}^{R,<}(E)\boldsymbol{U}, \quad \widetilde{\boldsymbol{G}}^{R,<}(E) = \boldsymbol{U}^\dagger \boldsymbol{G}^{R,<}(E)\boldsymbol{U}, \end{aligned} \tag{6.42}$$

where $\boldsymbol{U}$ is a block diagonal transformation matrix containing the reduced basis $\boldsymbol{U}_i$ of each layer i (with dimension $N_t \times N_m$, where N_m is the number of reduced basis).

The $-i\widetilde{\boldsymbol{G}}^<(E)$ gives the electron density. The hole density $i\widetilde{\boldsymbol{G}}^>(E)$ is obtained by subtracting electron density from the spectral function

$$i\widetilde{\boldsymbol{G}}^>(E, x_i) = -2Im\{\widetilde{\boldsymbol{G}}^R(E, x_i)\} + i\widetilde{\boldsymbol{G}}^<(E, x_i). \tag{6.43}$$

In TFET, electrons can tunnel from valence band into conduction band and leave holes in the valence band. The charge density involving both electrons and holes is calculated by the method similar to Ref. [35].

$$\begin{aligned} \widetilde{\boldsymbol{Q}}(x_i) = (ie) \int dE \frac{1}{2}[sgn(E - E_N(x_i)) + 1] \cdot \widetilde{\boldsymbol{G}}^<(E, x_i) \\ + \frac{1}{2}[-sgn(E - E_N(x_i)) + 1] \cdot \widetilde{\boldsymbol{G}}^>(E, x_i), \end{aligned} \tag{6.44}$$

where $E_N(x_i)$ is the layer-dependent threshold (charge neutral level), which is taken as the mid-bandgap $E_N(x_i) = 0.5[E_v(x_i) + E_c(x_i)] + \bar{V}(x_i)$ where $\bar{V}(x_i)$ is the average potential of layer x_i. *sgn* is the sign function. This model basically says that if a carrier is above (below) the threshold, it is considered as an electron (hole). The required diagonal blocks of $\widetilde{\boldsymbol{G}}^{R,<}(E)$, i.e., $\widetilde{\boldsymbol{G}}^{R,<}(E, x_i)$, can be calculated with efficient recursive Green's function (RGF) algorithm [36], since the matrices are still block three diagonal after the transformation.

The integrated $\widetilde{\boldsymbol{Q}}(x_i)$ is then transformed back into real space,

$$Q(\boldsymbol{r}) = diag\left(\boldsymbol{U}'\boldsymbol{U}\widetilde{\boldsymbol{Q}}\boldsymbol{U}^\dagger \boldsymbol{U}'^\dagger\right), \tag{6.45}$$

where $\boldsymbol{U}'$ is the transformation matrix from Fourier space to real space. Note that only diagonal terms are needed, which can be utilized to relieve the computational cost of back transformation. The transmission coefficient (and then ballistic current) can be calculated directly in the reduced space.

The problem now is how to construct this transformation matrix $\boldsymbol{U}$ so that the reduced system is as small as possible, and yet it still accurately describes the original system. To construct the reduced basis $\boldsymbol{U}_i$ for layer i, the Hamiltonian of layer i is repeated to form an infinite periodic nanowire. The reduction comes from the fact that only the electrons near the conduction band bottom and valence band top are important in the transport process. To approximate the band structure over

that small region, $\boldsymbol{U}_i$ then consists of the sampled Bloch modes with energy lying in that region. Multiple point k space sampling and/or E space sampling can be employed as has been demonstrated for the three- and six-band cases [21]. Here, k space sampling is adopted since E space sampling is more costly and that the eight-band matrix is larger than the six- or three-band case.

6.4.2 Spurious Band Elimination

For three- and six-band $\boldsymbol{k} \cdot \boldsymbol{p}$ models, as is shown in [21], by sampling the Bloch modes at multiple points in the k space and/or E space, a significantly reduced Hamiltonian can be constructed that describes very well the valence band top, based on which p-type SiNW FETs are simulated with good accuracy and efficiency. However, direct extension of this method to eight-band $\boldsymbol{k} \cdot \boldsymbol{p}$ model fails. The problem is that the reduced model constructed by multipoint expansion generally leads to some spurious bands, a situation similar to constructing the equivalent tight-binding models [18], rendering the reduced model useless.

As an example, Fig. 6.5a plots the E-k dispersion for an ideal InAs nanowire orientated in the [100] direction. Figure 6.5b is the result using the reduced Hamiltonian $\widetilde{\boldsymbol{H}}$. The reduced basis $\boldsymbol{U}_i$ (i is arbitrary here) is constructed by sampling the Bloch modes evenly in the Brillouin zone (at $k = 0$, $\pm\boldsymbol{\pi}/4$, $\pm 2\boldsymbol{\pi}/4$, and $\pm 3\boldsymbol{\pi}/4$ [1/nm], as denoted in green lines in Fig. 6.5a), with the energy $E \in [E_v - 0.3\text{eV}, \text{E}_\text{c} + 0.8\text{eV}]$ (E_v and E_c are the confined valence and conduction band edges), which results in $N_m = 134$ modes. Note that the modes at negative k can be obtained by a transformation of those at positive k [21]. Clearly, the reduced

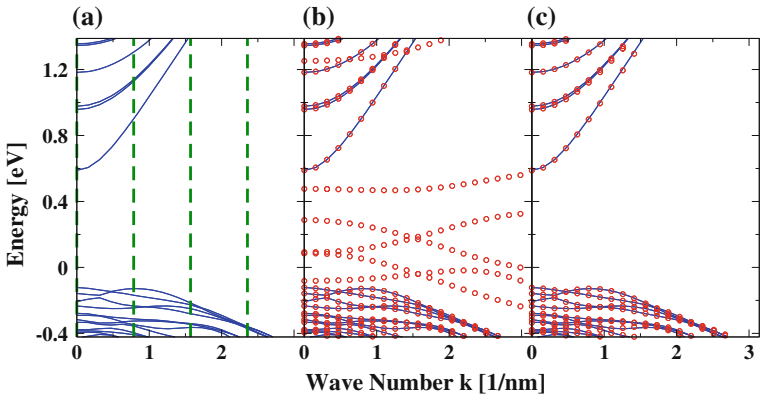

Fig. 6.5 E-k diagrams of a 5 nm × 5 nm InAs nanowire in the [100] direction. **a** Exact solution (*blue lines*) and the sampling lines (*dashed green lines*). **b** Comparison between exact solution (*blue lines*) and reduced-order model solution (*red circles*) with spurious bands showing up. **c** Comparison between exact solution (*blue lines*) and reduced-order model solution (*red circles*) with spurious bands removed

Hamiltonian reproduces quite well the exact dispersion in that energy window, demonstrating the number of sampling points is sufficient. However, there are also some spurious bands appearing in the conduction band and even in the bandgap, making the reduced model useless. Moreover, different sampling points or sampling windows would change the number and position of the spurious bands. This situation is not encountered in the three- or six-band model involving only the valence bands, or in the one-band effective mass model involving the conduction band only. It should be caused by the coupling between the conduction and valence bands which makes the eight-band model indefinite. The coupling is important for materials with narrow bandgaps.

The spurious bands must be suppressed. To this end, a singular value decomposition (SVD) is applied to the matrix $\boldsymbol{U}_i$. It is found that the singular values spread from a large value down to zero, indicating there are some linearly dependent modes. These linearly dependent modes give rise to null space of the reduced model and therefore must be removed. It is further found that the normal bands are mainly contributed by singular vectors having large singular values, in contrast to the spurious bands where singular vectors with small singular values also have large contribution. By removing the vectors with small singular values, i.e., vectors with $v \leq v_{th}$ where $v_{th} = 0.20$ is the threshold, a new reduced basis $\widetilde{\boldsymbol{U}}_i$ is generated with $\widetilde{N}_m = 76$. Using this new reduced basis, a new reduced Hamiltonian is constructed with its *E-k* diagram given in Fig. 6.5c. It is observed that all the spurious bands have been eliminated at a cost of slightly compromised accuracy. The reduction ratio is $\widetilde{N}_m/N_t = 76/1464 = 5.19\ \%$, which is quite significant.

The value of v_{th} is found to be crucial. A small v_{th} might be insufficient to remove all the spurious bands, while a large v_{th} may degrade the accuracy severely. Moreover, adjustment of v_{th} may be required when different sampling points or sampling energy windows are used. To determine v_{th} automatically, we propose a search process as follows:

1. Sample enough Bloch modes and store them in matrix $\boldsymbol{B}$. Suppose I points are sampled in the k space, and m_i modes with energy $E \in [E_1, E_2]$ are obtained at the ith point k_i $(1 \leq i \leq I)$, then the size of matrix $\boldsymbol{B}$ is $N_t \times N_m$, where $N_m = \sum_{i=1}^{I} m_i$.
2. Do SVD of $\boldsymbol{B}$, i.e., $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\dagger}$.
3. Set an initial value for v_{th}. Let us use $v_{th} = 0$ here.
4. Use v_{th} to construct a reduced basis $\widetilde{\boldsymbol{U}}$ by removing the singular vectors with $v < v_{th}$ in $\boldsymbol{U}$. The size of $\widetilde{\boldsymbol{U}}$ will be $N_t \times \widetilde{N}_m$.
5. Use $\widetilde{\boldsymbol{U}}$ to build a reduced Hamiltonian $\widetilde{\boldsymbol{H}}$. For each layer of $\widetilde{\boldsymbol{H}}$, the size will be $\widetilde{N}_m \times \widetilde{N}_m$.
6. Solve the *E-k* relation of $\widetilde{\boldsymbol{H}}$ for certain k_i, obtaining $\widetilde{m}_i$ modes with $E \in [E_1, E_2]$. It is found that $k_i = 0$ is a good choice.
7. If $\widetilde{m}_i > m_i$ (which means that there are still some spurious bands), increase v_{th} appropriately and go back to step 4. Otherwise, stop.

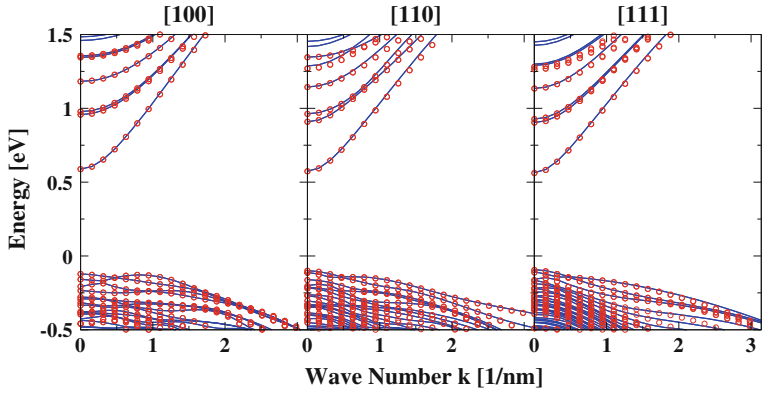

Fig. 6.6 *E-k* diagrams of 5 nm × 5 nm InAs nanowires in the [100], [110], and [111] orientations. The *blue lines* are the exact solutions, while the *red circles* are the solutions of the reduced-order models. The valid energy window of the reduced-order models is $[E_v - 0.3\text{eV}, E_c + 0.8\text{eV}]$. The orders of the reduced models are 76, 66, and 62, respectively

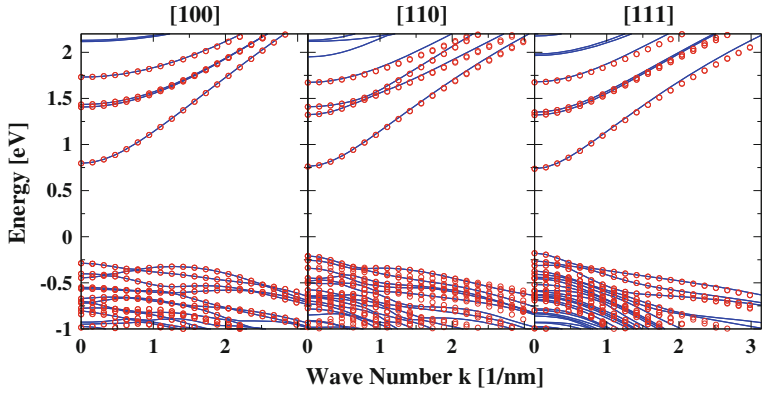

Fig. 6.7 *E-k* diagrams of 3 nm × 3 nm InAs nanowires in the [100], [110], and [111] orientations. The *blue lines* are the exact solutions, while the *red circles* are the solutions of the reduced-order models. The valid energy window of the reduced-order models is $[E_v - 0.55\,\text{eV}, E_c + 1.0\,\text{eV}]$. The orders of the reduced models are 66, 52, and 48, respectively

The above search process is fast, since step 5 and step 6 are much cheaper than step 1 although they have to be repeated many times. In fact, the complexity of step 1 is $I \times O\left(N_t^3\right)$, step 2 is $O\left(N_t N_m^2\right)$, step 5 is $O\left(\widetilde{N}_m N_t^2\right)$, and step 6 is $O\left(\widetilde{N}_m^3\right)$. Note that $\widetilde{N}_m < N_m < N_t$.

The $v_{th} = 0.20$ used earlier is the result of the above search process. The above process also gives good results for nanowires in the [110] and [111] directions, as shown in Fig. 6.6. Different energy windows $E \in [E_v - 0.2\text{eV}, \text{E}_\text{c} + 0.6\text{eV}]$ and $E \in [E_v - 0.4\text{eV}, \text{E}_\text{c} + 1.0\text{eV}]$ are tested, and again faithful results are obtained (not shown here). For other cross-sectional InAs nanowires such as the 3 nm × 3 nm

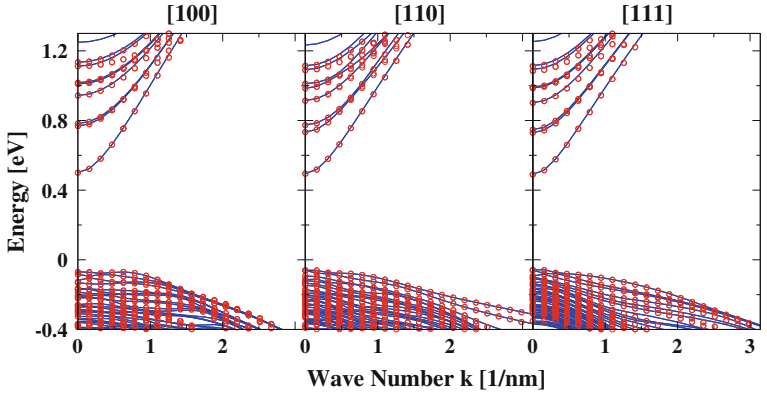

Fig. 6.8 *E-k* diagrams of 7 nm × 7 nm InAs nanowires in the [100], [110], and [111] orientations. The *blue lines* are the exact solutions, while the *red circles* are the solutions of the reduced-order models. The valid energy window of the reduced-order models is $[E_v - 0.25\,\mathrm{eV}, E_c + 0.7\,\mathrm{eV}]$. The orders of the reduced models are 104, 104, and 96, respectively

and 7 nm × 7 nm in the [100], [110], and [111] orientations, this method gives reliable results, as shown in Figs. 6.7and 6.8. It should be mentioned that this process results in a smaller basis set, which is different from the method for tight-binding models [18] where the basis is enlarged by putting in more modes to eliminate the spurious bands.

6.4.3 Error and Cost of the Reduced Models

Now, this reduced model can be applied to simulate the TFET as shown in Fig. 6.1. Reduced NEGF equations and Poisson's equation are solved self-consistently. To improve the efficiency, the reduced basis is constructed for an ideal nanowire with its potential term set to zero, so the reduced basis just needs to be solved only once for each material and it remains unchanged during the self-consistent iterations. The potential term in real devices then merely causes transitions between these scattering states. This assumption has been adopted in Ref. [18] with good accuracy demonstrated. As will be shown below, it is also a fairly good approximation for the GAA nanowire TFET here.

The I_{DS}-V_{GS} transfer characteristics of a 5 nm × 5 nm cross-sectional InAs homojunction TFET are plotted in Fig. 6.9a. Three curves are compared. In the first, second, and third *I*–*V* curve, the valid energy window is $[E_v - 0.2\mathrm{eV}, \mathrm{E_c} + 0.6\mathrm{eV}]$, $[E_v - 0.3\mathrm{eV}, \mathrm{E_c} + 0.8\mathrm{eV}]$, and $[E_v - 0.4\mathrm{eV}, \mathrm{E_c} + 1.0\mathrm{eV}]$, respectively. The sampling k points are all at $k = 0$, $\pm\pi/4$, $\pm 2\pi/4$, and $\pm 3\pi/4$ $[1/\mathrm{nm}]$. This leads to $\widetilde{N}_m = 48$, $\widetilde{N}_m = 76$, and $\widetilde{N}_m = 106$, with corresponding *I*–*V* curves denoted as I_{48}, I_{76}, and I_{106}. Here, I_{106} can be considered as the reference, since with larger energy window and more modes, the result is expected to have better accuracy. The

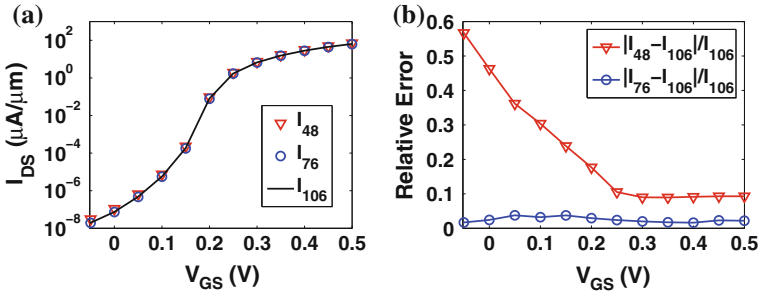

Fig. 6.9 **a** I_{DS}-V_{GS} transfer characteristics of a p-i-n InAs homojunction TFET as shown in Fig. 6.1. The nanowire is oriented in the [100] direction. $T_{ox} = 1$ nm, $\varepsilon_{ox} = 12.7$, $T_y = T_z = 5$ nm, $L_s = 15$ nm, $L_g = 20$ nm, $L_d = 30$ nm. The doping density is equal to 5×10^{19} cm^{-3} at the source and 5×10^{18} cm^{-3} at the drain. The drain bias is fixed to $V_{DS} = 0.3$V. **b** Relative errors of I_{48} and I_{76} with respect to I_{106}

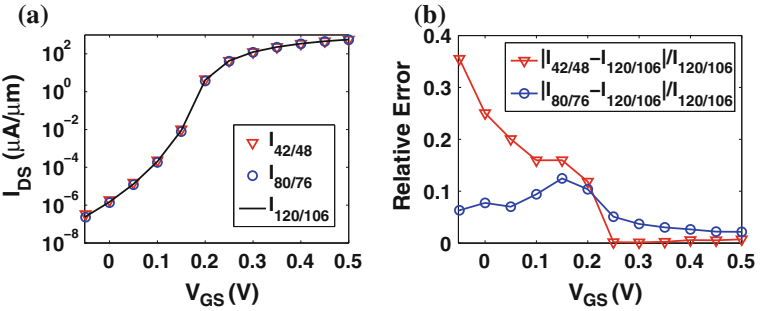

Fig. 6.10 **a** I_{DS}-V_{GS} transfer characteristics of a p-i-n GaSb/InAs heterojunction TFET as shown in Fig. 6.1. The device settings are the same as Fig. 6.9. **b** Relative errors of $I_{42/48}$ and $I_{80/76}$ with respect to $I_{120/106}$

relative errors of I_{48} and I_{76} relative to I_{106} are calculated and plotted in Fig. 6.9b. It is observed that I_{48} has large deviations with respect to I_{106}, especially when current is small. Instead I_{76} is very close to I_{106}, and the relative errors are less than 4 % for all bias points, indicating that the results have converged.

The I_{DS}-V_{GS} transfer characteristics of a 5 nm × 5 nm cross-sectional GaSb/InAs heterojunction TFET are plotted in Fig. 6.10a. Burt–Foreman operator ordering is used at the material interface though symmetrized ordering gives similar results in this case. Due to the small lattice mismatch between GaSb and InAs, strain is small and is neglected here. Again, three *I*–*V* curves are compared. In the first, second, and third *I*–*V* curve, the energy window is $[E_v - 0.3\text{eV}, \text{E}_\text{c} + 0.4\text{eV}]$, $[E_v - 0.4\text{eV}, \text{E}_\text{c} + 0.6\text{eV}]$, and $[E_v - 0.5\text{eV}, \text{E}_\text{c} + 0.8\text{eV}]$ for GaSb, $[E_v - 0.2\text{eV}, \text{E}_\text{c} + 0.6\text{eV}]$, $[E_v - 0.3\text{eV}, \text{E}_\text{c} + 0.8\text{eV}]$, and $[E_v - 0.4\text{eV}, \text{E}_\text{c} + 1.0\text{eV}]$ for InAs. The sampling k points are all at $k = 0$, $\pm\pi/4$, $\pm 2\pi/4$, and $\pm 3\pi/4$ [1/nm]. This leads to $\widetilde{N}_m = 42$, $\widetilde{N}_m = 80$, and $\widetilde{N}_m = 120$ for GaSb, $\widetilde{N}_m = 48$, $\widetilde{N}_m = 76$, and $\widetilde{N}_m = 106$ for InAs,

Table 6.2 List of run time for the TFET simulations

I–V curves	I_{48}	I_{76}	I_{106}	$I_{42/48}$	$I_{80/76}$	$I_{120/106}$
One iteration (minutes)	2.38	2.96	4.09	2.47	3.02	4.21
No. of iterations	41	47	43	87	82	98
Total (minutes)	97.6	139.1	175.9	214.9	247.7	412.6

with corresponding *I–V* curves denoted as $I_{42/48}$, $I_{80/76}$, and $I_{120/106}$. The relative errors of $I_{42/48}$ and $I_{80/76}$ relative to $I_{120/106}$ are calculated and plotted in Fig. 6.10b. It is observed that $I_{42/48}$ has very small deviations with respect to I_{106} when above threshold current, but large errors when below threshold current. In contrast, $I_{80/76}$ has much better accuracy below threshold but larger error above threshold. Overall, the error of $I_{80/76}$ is still acceptable for predictive device modeling.

Table 6.2 lists the run time details for generating the above *I–V* curves. Note that homogeneous energy mesh with grid size $\Delta E = 3$ meV is used which results in 359 energy points in total. Different energy points are calculated in parallel with 12 cores. All the simulations are performed on dual 8-core Intel Xeon-E5 CPUs. It is observed that the simulation time of one NEGF–Poisson iteration increases sub-linearly with $\widetilde{N}_m$; different $\widetilde{N}_m$ leads to small fluctuation of convergence (in terms of number of NEGF–Poisson iterations). In addition, the heterojunction TFET is harder to converge compared with homojunction case. Overall, the simulation time for one *I–V* curve took just a few hours, suitable for device design and optimization.

6.5 Simulation Results

The above benchmarked quantum transport solver is used to study various device configurations as described in Sect. 6.2. Homojunction TFETs are simulated first with both n-type and p-type devices considered. Then, various performance boosters are applied to the n-type devices, though the same ideas can be applied to p-type devices with qualitatively similar results expected. The device parameters are the same as those in Figs. 6.9 and 6.10 if not stated otherwise. In the following discussions, the current is normalized to the width of the nanowire to get unit of μA/μm. We consider high-performance (HP), low operating power (LOP), and low standby power (LSTP) applications, where the OFF currents are fixed to 10^{-1} μA/μm, 5×10^{-3} μA/μm, and 10^{-5} μA/μm, respectively.

6.5.1 Homojunction TFETs

Figure 6.11a compares the I_{DS}-V_{GS} curves with different gate lengths. It is found that the SS improves as gate length increases, while the turn-on characteristics

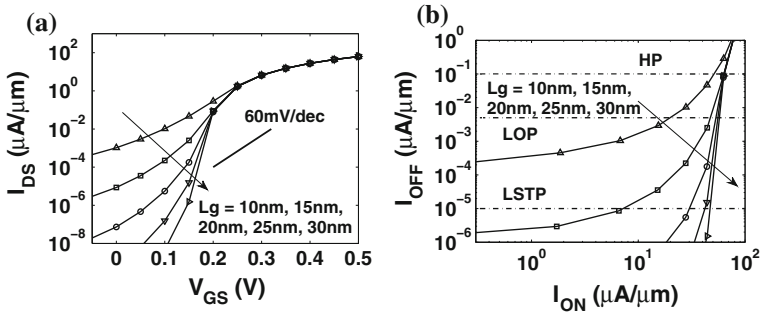

Fig. 6.11 **a** I_{DS}-V_{GS} curves (at $V_{DS} = 0.3$ V) and **b** I_{ON}-I_{OFF} (at $V_{DD} = 0.3$ V) of the 5 nm × 5 nm cross-sectional InAs nanowire homojunction n-type TFETs in the [100] orientation. Gate lengths of 10–30 nm are compared

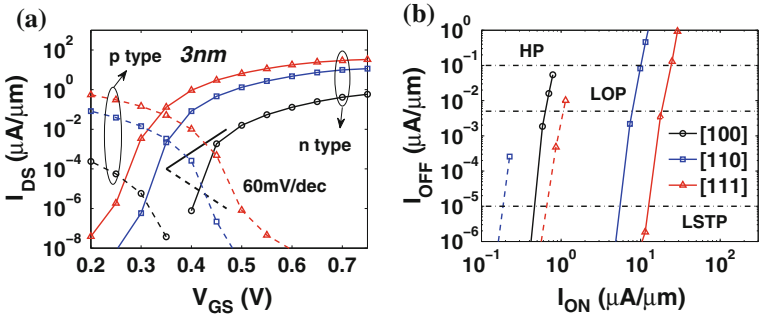

Fig. 6.12 **a** I_{DS}-V_{GS} curves (at $V_{DS} = 0.3$ V) and **b** I_{ON}-I_{OFF} (at $V_{DD} = 0.3$ V) of the 3 nm × 3 nm cross-sectional n-type and p-type InAs nanowire homojunction TFETs. Three transport orientations [100], [110], and [111] are considered

remain unchanged. This is understandable since longer gate length has less source-to-drain tunneling leakage. As a result, I_{ON} improves when I_{OFF} is fixed, as shown in Fig. 6.11b. The I_{ON} improvement is the largest for LSTP application and the smallest for the HP application. It is also found that I_{ON} will saturate when gate length becomes very long; the gate length at which I_{ON} saturates is shorter for HP application than for LSTP application. In the following simulations, we fix the gate length to be 20 nm.

Figures 6.12, 6.13, and 6.14 compare the I_{DS}-V_{GS} and I_{ON}-I_{OFF} characteristics of InAs nanowire TFETs for three cross-sectional sizes and for three transport directions. For p-type devices considered here, the doping density is set to 2×10^{19} cm^{-3} at the source and 5×10^{19} cm^{-3} at the drain, $L_s = 25$ nm, $L_d = 15$ nm. It is observed that, for small cross sections such as the 3 nm case, although the SS is very small, the I_{ON} are very limited, for all three orientations. This is due to their large electron effective masses and large bandgaps. While for large cross sections

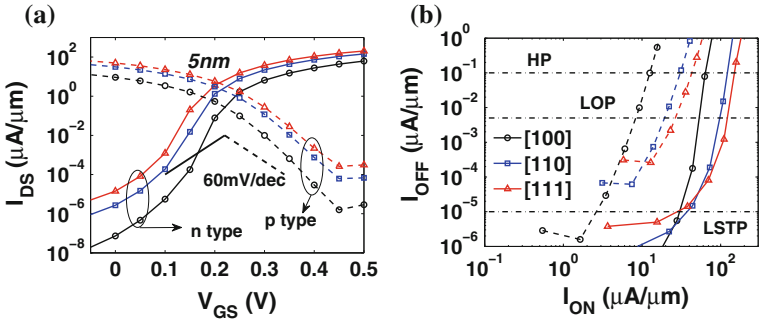

Fig. 6.13 The same as Fig. 6.12 but for 5 nm × 5 nm cross section

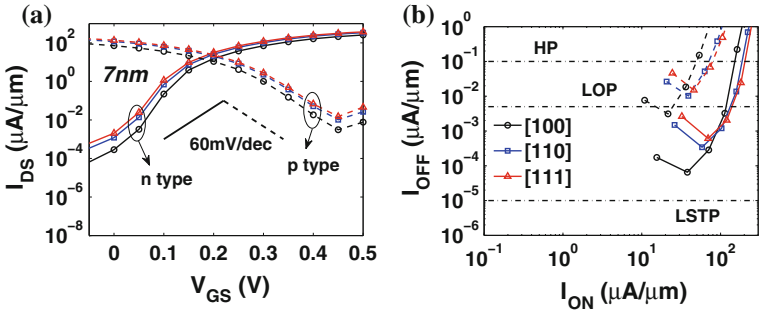

Fig. 6.14 The same as Fig. 6.12 but for 7 nm × 7 nm cross section

such as the 7 nm case, the SS degrades due to the smaller electron effective masses, smaller bandgaps, as well as the weaker electrostatic control; the I_{ON}, however, are very good for both HP and LOP applications. It should be noted that I_{OFF} are not sufficiently small which makes them unsuitable for LSTP application. The large I_{OFF} are due to the direct source-to-drain tunneling leakage and ambipolar tunneling leakage at the channel–drain junction, and these two components become more pronounced when bandgap becomes smaller. Therefore, for LSTP application, medium-sized cross section such as the 5 nm case should be a better choice; otherwise, the channel length needs to be increased and/or the drain doping density needs to be decreased to suppress the leakage.

For small wire cross section, the performances of the three orientations differ a lot. In particular, for the 3 nm case, [111] orientation gives the best I_{ON} for all three applications, while [100] is the worst. When the cross-sectional size increases, the three orientations tend to deliver similar performances. It also means that [111] orientation has the best cross-sectional scaling ability. In fact, when the confinement becomes stronger (as the nanowire size decreases), the band structure starts to differ from each other for the three orientations, as can be observed in Figs. 6.6, 6.7, and 6.8. In particular, the [100] orientation shows the fastest increase of bandgap, while

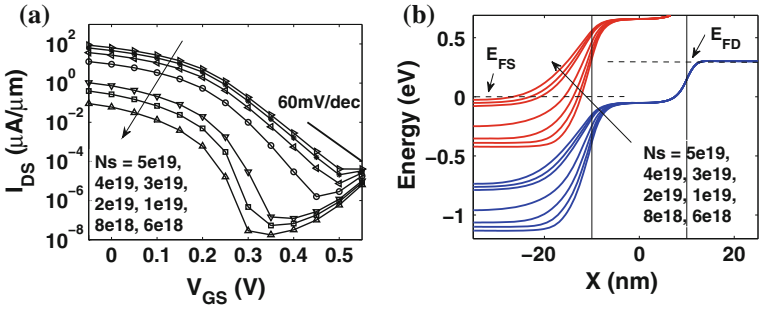

Fig. 6.15 **a** I_{DS}-V_{GS} curves (at $V_{DS} = 0.3$ V) and **b** potential profiles (at $V_{DS} = 0.3$ V and $V_{GS} = 0.3$ V) of the 5 nm × 5 nm cross-sectional InAs nanowire homojunction p-type TFETs in the [100] orientation. Source doping densities of 5×10^{19} cm^{-3}–6×10^{18} cm^{-3} are compared

the [111] case shows the slowest increase of bandgap, meanwhile the hole effective mass decreases (as a result of strongly anisotropic heavy hole band).

Comparing n-type and p-type devices, we found that p-type devices have worse SS and smaller I_{ON}. This is because the doping density has been set to 2×10^{19} cm^{-3} at the source side, lower than that of n-type ones (which is 5×10^{19} cm^{-3}). This doping density is a compromise of SS and I_{ON}. In fact, as shown in Fig. 6.15, lower doping leads to smaller I_{ON} as a result of less abrupt tunneling junction, while higher doping leads to worse SS (approaching 60 mV/dec) since larger Fermi degeneracy is created in the conduction band. The Fermi degeneracy creates thermal tail which counteracts the energy filtering functionality of TFETs. For 3 nm (7 nm) p-type TFETs here, smaller (larger) Fermi degeneracy in the source is observed because the electron mass and density of states increases (decreases) as cross section decreases (increases) (Fig. 6.3).

6.5.2 *Improvements of Homojunction TFETs*

As shown in Fig. 6.16a, b, the source–pocket TFETs can improve I_{ON} for all HP, LOP, and LSTP applications, by up to 50 μA/μm. With 2–5 nm pocket lengths, I_{ON} first increases and then saturates. Further increasing the pocket length will decrease I_{ON} (not shown here). This can be explained by plotting the band diagram and current spectra, as shown in Fig. 6.16c, d. With 2–5 nm pocket lengths, the source pocket first increases the electric field across the source–channel tunneling junction and then starts to form a quantum well. This quantum well creates a resonant state leading to a very sharp tunneling peak. However, this peak is too narrow in energy to help the total current.

It has been shown in [12] that uniaxial compressive stress and biaxial tensile stress reduce InAs nanowire bandgap and effective masses, which can be used to improve TFET performances. As shown in Fig. 6.17, the uniaxial compressive

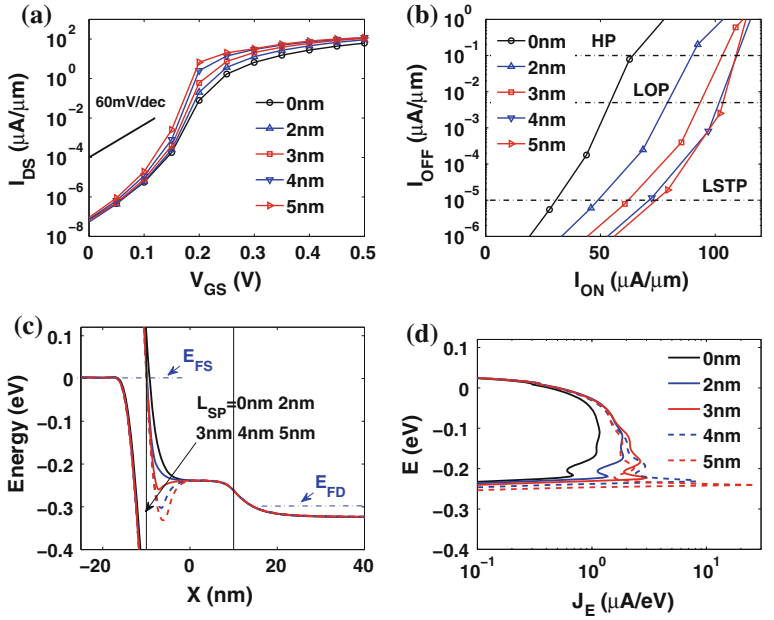

Fig. 6.16 **a** I_{DS}-V_{GS} curves (at $V_{DS} = 0.3$ V), **b** I_{ON}-I_{OFF} (at $V_{DD} = 0.3$ V), **c** potential profiles (at $V_{DS} = 0.3$ V and $V_{GS} = 0.45$ V), and **d** current spectra (at $V_{DS} = 0.3$ V and $V_{GS} = 0.45$ V), of the 5 nm × 5 nm cross-sectional source–pocket InAs homojunction TFETs in the [100] orientation, compared with no pocket case. The doping density of the pocket is $N_{sp} = 5 \times 10^{19}$ cm^{-3}. Pocket lengths of 2–5 nm are considered

stress degrades the SS, but still improved I_{ON} is observed for both HP and LOP applications, consistent with [12]. The degraded SS can be explained by the large Fermi degeneracy of the source (due to lighter hole effective mass) creating thermal tail. It is found here that the strain-induced I_{ON} improvement is more significant in the [100] orientation than in the [110] and [111] orientations, since the strain-induced bandgap and effective mass reductions are more pronounced in the [100] orientation. On the other hand, uniaxial tensile stress leads to increased bandgap and effective masses and thus degraded I_{ON} (no shown here). Biaxial strain in the cross-sectional plane is hard to realize in experiments and therefore is not considered here.

6.5.3 *Heterojunction TFETs*

As shown in Fig. 6.18a, b, the GaSb/InAs heterojunction TFETs significantly improve I_{ON} for all HP, LOP, and LSTP applications. [111] orientation gives the best I_{ON} for both HP and LOP applications, while [100] gives the best I_{ON} for LSTP

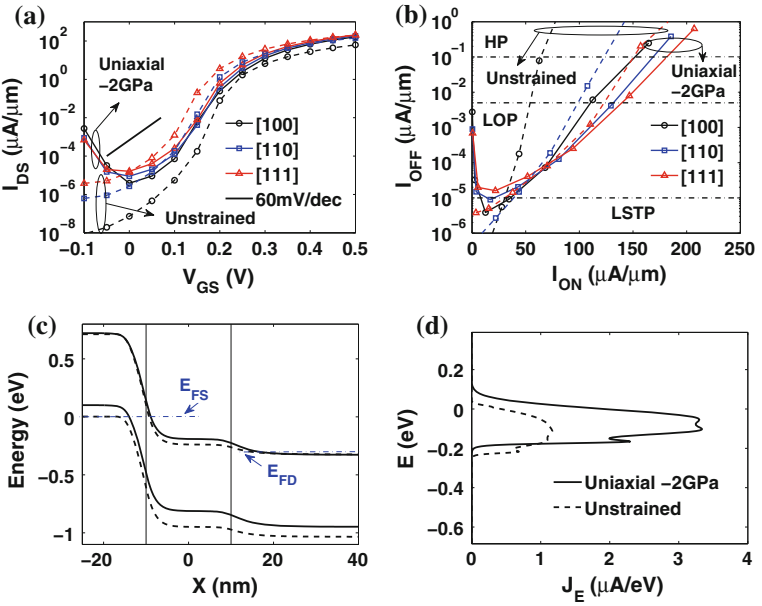

Fig. 6.17 **a** I_{DS}-V_{GS} curves (at $V_{DS} = 0.3$ V) and **b** I_{ON}-I_{OFF} (at $V_{DD} = 0.3$ V) of the 5 nm × 5 nm cross-sectional homojunction TFETs in the [100], [110], and [111] orientations, with uniaxial compressive stress along the transport direction, in comparison with unstrained cases. **c** Potential profiles (at $V_{DS} = 0.3$ V and $V_{GS} = 0.45$ V) and **d** current spectra (at $V_{DS} = 0.3$ V and $V_{GS} = 0.45$ V), of the 5 nm × 5 nm cross-sectional homojunction TFET in the [100] orientation, with uniaxial compressive stress along the transport direction, in comparison with unstrained case

application. In Fig. 6.18c, d, we compare the band diagrams and current spectra of the GaSb/InAs heterojunction TFET with the InAs homojunction TFET. It is clear that the smaller tunneling height and distance of the heterojunction TFET, in particular at the GaSb side, lead to around 10× larger tunneling current.

6.5.4 Improvements of Heterojunction TFETs

Employing the schemes for improving I_{ON} of homojunction TFETs, we get source–pocket heterojunction TFETs or strained heterojunction TFETs, which are expected to deliver even larger I_{ON}.

Indeed, as shown in Fig. 6.19a, b, the source pocket can improve I_{ON} for all HP, LOP, and LSTP applications, by up to 200 μA/μm. The optimal pocket length is found to be around 4 nm, beyond which I_{ON} will drop. The physics is similar to the source–pocket homojunction TFETs and will not be repeated here.

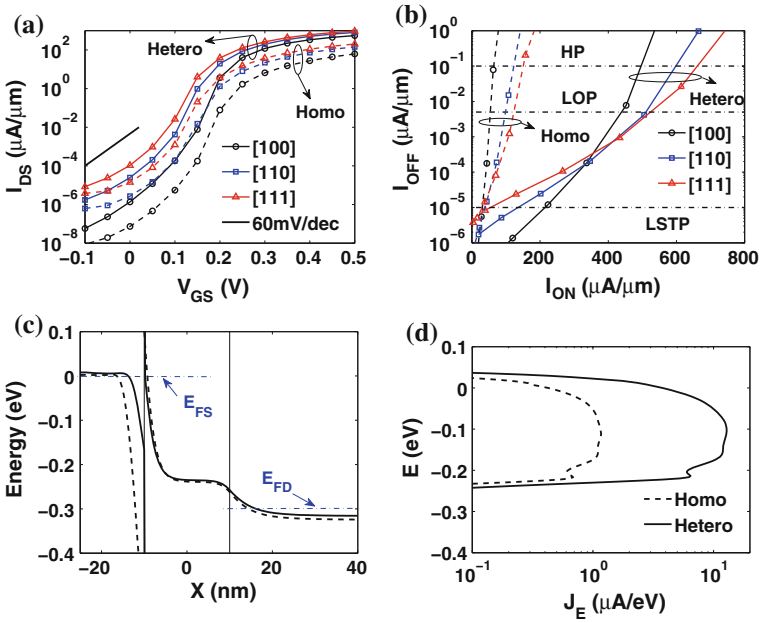

Fig. 6.18 **a** I_{DS}-V_{GS} curves (at $V_{DS} = 0.3$ V) and **b** I_{ON}-I_{OFF} (at $V_{DD} = 0.3$ V) of the 5 nm × 5 nm cross-sectional GaSb/InAs heterojunction TFETs in the [100], [110], and [111] orientations, compared with the homojunction cases. **c** Potential profiles (at $V_{DS} = 0.3$ V and $V_{GS} = 0.45$ V) and **d** current spectra (at $V_{DS} = 0.3$ V and $V_{GS} = 0.45$ V), of the 5 nm × 5 nm cross-sectional GaSb/InAs heterojunction TFETs in the [100] orientation, compared with the homojunction case

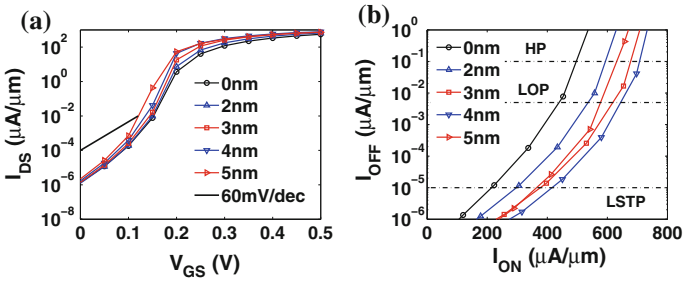

Fig. 6.19 **a** I_{DS}-V_{GS} curves (at $V_{DS} = 0.3$ V) and **b** I_{ON}-I_{OFF} (at $V_{DD} = 0.3$ V), of the 5 nm × 5 nm cross-sectional source–pocket GaSb/InAs heterojunction TFETs in the [100] orientation, compared with no pocket case. The doping density of the pocket is $N_{sp} = 5 \times 10^{19}$ cm^{-3}. Pocket lengths of 2–5 nm are considered

However, as shown in Fig. 6.20a, b, uniaxial compressive stress only slightly improves I_{ON} of [100] orientation for HP and LOP applications (and [110] orientation for HP application). In the [111] orientation, the stress even degrades I_{ON}. Again, uniaxial tensile stress leads to increased effective masses and thus degraded

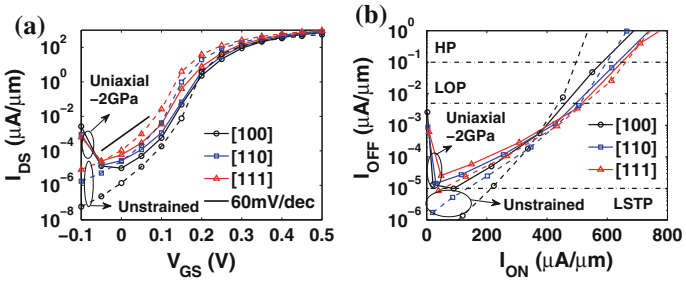

Fig. 6.20 **a** I_{DS}-V_{GS} curves (at $V_{DS} = 0.3$ V) and **b** I_{ON}-I_{OFF} (at $V_{DD} = 0.3$ V), of the 5 nm × 5 nm cross-sectional heterojunction TFETs in the [100], [110], and [111] orientations, with uniaxial compressive stress applied along the transport direction, in comparison with unstrained cases

I_{ON} (no shown here). The physics is similar to the strained homojunction TFETs and will not be repeated here.

6.6 Conclusions

To efficiently simulate III-V nanowire-based TFETs, a reduced-order $\boldsymbol{k} \cdot \boldsymbol{p}$ NEGF method is developed. Through comparison with TB method, the $\boldsymbol{k} \cdot \boldsymbol{p}$ method is shown to be able to describe quite well the band structures of very small nanowires. By introducing a spurious band elimination process, the reduced-order $\boldsymbol{k} \cdot \boldsymbol{p}$ models can be constructed for reproducing the original band structures in an energy window near the bandgap. The reduced models can also accurately capture the *I*–*V* characteristics of homojunction and heterojunction TFETs within a short simulation time.

InAs TFETs with different cross sections and channel orientations are compared, and it is found that [111] direction has the best cross-sectional scaling ability. Various performance boosters are studied. It is found that embedding source pockets can improve the "on" current due to the enhanced band bending at the source-to-channel junction, but this effect will saturate with increasing pocket length. Uniaxial compressive stress can also be used to boost the "on" current, which is found to be more effective in the [100] orientation than in the [110] and [111] orientations. Adopting GaSb/InAs heterojunction achieves a much larger "on" current due to the staggered-gap band alignment. Incorporating source pockets with proper pocket length into the heterojunction TFET is shown to further enhance the "on" current.

However, there is a large gap between theoretical projections and experiments [37]. In experiments, the device performances are usually degraded by nonidealities such as phonon/dopant-induced band tails, defect-assisted tunneling, interface roughness, and traps [37, 38]. To model these non-idealities, the $\boldsymbol{k} \cdot \boldsymbol{p}$ Hamiltonian needs to be modified properly to account for these defects and the transport solver

needs to be extended to incorporate various scattering events due to impurity, alloy, phonon, and surface roughness. These will be done in the future.

Acknowledgment The use of nanoHUB.org computational resources operated by the Network for Computational Nanotechnology funded by the US National Science Foundation under Grant Nos. EEC-0228390, EEC-1227110, EEC-0228390, EEC-0634750, OCI-0438246, OCI-0832623, and OCI-0721680 is gratefully acknowledged.

References

1. A.M. Ionescu, H. Riel, Tunnel field-effect transistors as energy-efficient electronic switches. Nature **479**, 329–337 (2011)
2. A.C. Seabaugh, Q. Zhang, Low-voltage tunnel transistors for beyond CMOS logic. Proc. IEEE **98**(12), 2095–2110 (2010)
3. V. Nagavarapu, R. Jhaveri, J.C. Woo, The tunnel source (pnpn) n-MOSFET: a novel high performance transistor. IEEE Trans. Electron. Devices **55**(4), 1013–1019 (2008)
4. R. Jhaveri, V. Nagavarapu, J.C. Woo, Effect of pocket doping and annealing schemes on the source-pocket tunnel field-effect transistor. IEEE Trans. Electron. Devices **58**(1), 80–86 (2011)
5. D. Verreck, A. Verhulst, K.-H. Kao, W. Vandenberghe, K. De Meyer, G. Groeseneken, Quantum mechanical performance predictions of p-n-i-n versus pocketed line tunnel field-effect transistors. IEEE Trans. Electron. Devices **60**(7), 2128–2134 (2013)
6. D.K. Mohata, R. Bijesh, S. Mujumdar, C. Eaton, R. Engel-Herbert, T. Mayer, V. Narayanan, J.M. Fastenau, D. Loubychev, A.K. Liu, S. Datta, Demonstration of MOSFET-like on-current performance in arsenide/antimonide tunnel FETs with staggered hetero-junctions for 300 mV logic applications, in *IEEE IEDM Technical Digest*, Dec 2011, pp. 33.5.1–33.5.4
7. A.W. Dey, B.M. Borg, B. Ganjipour, M. Ek, K.A. Dick, E. Lind, C. Thelander, L.-E. Wernersson, High-current GaSb/InAs(Sb) nanowire tunnel field-effect transistors. IEEE Electron. Devices Lett. **34**(2), 211–213 (2013)
8. X. Zhao, A. Vardi, J.A. del Alamo, InGaAs/InAs heterojunction vertical nanowire tunnel FETs fabricated by a top-down approach, in *IEEE IEDM Technical Digest*, Dec 2014, pp. 25.5.1–25.5.4
9. M. Luisier, G. Klimeck, Performance comparisons of tunneling field-effect transistors made of InSb, Carbon, and GaSb-InAs broken gap heterostructures, in *IEEE IEDM Technical Digest*, Dec 2009, pp. 1–4
10. Z. Jiang, Y. He, G. Zhou, T. Kubis, H.G. Xing, G. Klimeck, Atomistic simulation on gate-recessed InAs/GaSb TFETs and performance benchmark, in *DRC*, June 2013, pp. 145–146
11. U.E. Avci, I.A. Young, Heterojunction TFET scaling and resonant-TFET for steep subthreshold slope at sub-9 nm gate-length, in *IEEE IEDM Technical Digest,* Dec 2013, pp. 4.3.1–4.3.4
12. F. Conzatti, M.G. Pala, D. Esseni, E. Bano, L. Selmi, Strain-induced performance improvements in InAs nanowire tunnel FETs. IEEE Trans. Electron. Devices **59**(8), 2085–2092 (2012)
13. S. Brocard, M.G. Pala, D. Esseni, Design options for hetero-junction tunnel FETs with high on current and steep sub-threshold voltage slope, in *IEEE IEDM Technical Digest*, Dec 2013, pp. 5.4.1–5.4.4

14. S. Brocard, M.G. Pala, D. Esseni, Large on-current enhancement in hetero-junction tunnel-FETs via molar fraction grading. IEEE Electron. Devices Lett. **35**(2), 184–186 (2014)
15. A. Sharma, A.A. Goud, K. Roy, GaSb-InAs n-TFET with doped source underlap exhibiting low subthreshold swing at sub-10-nm gate-lengths. IEEE Electron. Devices Lett. **35**(12), 1221–1223 (2014)
16. M.G. Pala, S. Brocard, Exploiting hetero-junctions to improve the performance of III-V nanowire tunnel-FETs. IEEE J. Electron. Devices Soc. **3**(3), 115–121 (2015)
17. S. Datta, *Quantum Transport: Atom to Transistor* (Cambridge University Press, Cambridge, 2005)
18. G. Mil'nikov, N. Mori, Y. Kamakura, Equivalent transport models in atomistic quantum wires. Phys. Rev. B **85**(3), 035317-1–035317-11 (2012)
19. A. Afzalian, J. Huang, H. Ilatikhameneh, J. Charles, D. Lemus, J. Bermeo Lopez, S. Perez Rubiano, T. Kubis, M. Povolotskyi, G. Klimeck, M. Passlack, Y.-C. Yeo, Mode space tight binding model for ultra-fast simulations of III-V nanowire MOSFETs and heterojunction TFETs, in *Proceedings 18th International Workshop on Computer Electronics*, Sept 2015
20. M. Shin, Full-quantum simulation of hole transport and band-to-band tunneling in nanowires using the k · p method. J. Appl. Phys. **106**(5), 054505-1–054505-10 (2009)
21. J.Z. Huang, W.C. Chew, J. Peng, C.-Y. Yam, L.J. Jiang, G.-H. Chen, Model order reduction for multiband quantum transport simulations and its application to p-type junctionless transistors. IEEE Trans. Electron. Devices **60**(7), 2111–2119 (2013)
22. J.Z. Huang, L. Zhang, W.C. Chew, C.-Y. Yam, L.J. Jiang, G.-H. Chen, M. Chan, Model order reduction for quantum transport simulation of band-to-band tunneling devices. IEEE Trans. Electron. Devices **61**(2), 561–568 (2014)
23. M. Luisier, G. Klimeck, Atomistic full-band design study of InAs band-to-band tunneling field-effect transistors. IEEE Electron. Devices Lett. **30**(6), 602–604 (2009)
24. D. Gershoni, C. Henry, G. Baraff, Calculating the optical properties of multidimensional heterostructures: application to the modeling of quaternary quantum well lasers. IEEE J. Quantum Electron. **29**(9), 2433–2450 (1993)
25. P. Enders, M. Woerner, Exact block diagonalization of the eight-band Hamiltonian matrix for tetrahedral semiconductors and its application to strained quantum wells. Semicond. Sci. Technol. **11**(7), 983 (1996)
26. H.-B. Wu, S.J. Xu, J. Wang, Impact of the cap layer on the electronic structures and optical properties of self-assembled InAs/GaAs quantum dots. Phys. Rev. B **74**(20), 205329 (2006)
27. B.A. Foreman, Elimination of spurious solutions from eight-band $k \cdot p$ theory. Phys. Rev. B **56** (20), R12748 (1997)
28. T.B. Bahder, Eight-band $k \cdot p$ model of strained zinc-blende crystals. Phys. Rev. B **41**(17), 11992 (1990)
29. I. Vurgaftman, J.R. Meyer, L.R. Ram-Mohan, Band parameters for III-V compound semiconductors and their alloys. J. Appl. Phys. **89**(11), 5815–5875 (2001)
30. D. Esseni, P. Palestri, L. Selmi, *Nanoscale MOS transistors: semi-classical transport and applications* (Cambridge University Press, Cambridge, 2011)
31. A. Paussa, F. Conzatti, D. Breda, R. Vermiglio, D. Esseni, P. Palestri, Pseudospectral methods for the efficient simulation of quantization effects in nanoscale MOS transistors. IEEE Trans. Electron. Devices **57**(12), 3239–3249 (2010)
32. R.G. Veprek, S. Steiger, B. Witzigmann, Ellipticity and the spurious solution problem of $k \cdot p$ envelope equations. Phys. Rev. B **76**(16), 165320-1–165320-9 (2007)
33. M. Luisier, G. Klimeck, Simulation of nanowire tunneling transistors: From the Wentzel–Kramers–Brillouin approximation to full-band phonon-assisted tunneling. J. Appl. Phys. **107** (8), 084507 (2010)
34. S.O. Koswatta, S.J. Koester, W. Haensch, On the possibility of obtaining MOSFET-like performance and sub-60-mV/dec swing in 1-D broken-gap tunnel transistors. IEEE Trans. Electron Devices **57**(12), 3222–3230 (2010)

35. J. Guo, S. Datta, M. Lundstrom, M. Anantam, Toward multiscale modeling of carbon nanotube transistors. Int. J. Multiscale Comput. Eng. **2**(2), 257–276 (2004)
36. R. Lake, G. Klimeck, R.C. Bowen, D. Jovanovic, Single and multiband modeling of quantum electron transport through layered semiconductor devices. J. Appl. Phys. **81**(12), 7845–7869 (1997)
37. H. Lu, A. Seabaugh, Tunnel field-effect transistors: state-of-the-art. IEEE J. Electron. Devices Soc. **2**(4), 44–49 (2014)
38. U.E. Avci, D.H. Morris, I.A. Young, Tunnel field-effect transistors: prospects and challenges. IEEE J. Electron. Devices Soc. **3**(3), 88–95 (2015)

Chapter 7
Carbon Nanotube TFETs: Structure Optimization with Numerical Simulation

Hao Wang

Abstract The unique band structure makes carbon nanotube (CNT) an ideal vehicle for tunnel FET (TFET) studying. In this chapter, the structure of CNT-TFET is optimized with numerical simulation. The band structure of CNT is acquired with p_z orbital tight-binding model. Quantum mechanical simulation with the non-equilibrium Green's function is adopted describing the carrier transport. TFET is compared with conventional MOSFET with CNT as the channel material. A steeper than 60 mv/dec inverse subthreshold slope is obtained at the cost of a smaller on current and the ambipolar conduction behavior. The current modulation mechanism of TFET is discussed concerning both the occupancy probability and tunnel probability. The occupancy probability can be modulated with band alignment, and the tunnel probability can be modulated with the electric field or tunnel path. Several optimized TFET structures including doping engineering, dielectric engineering, and gate work function engineering are demonstrated for improved performances with increased on current and/or reduced ambipolar conduction.

7.1 Introduction

At the present era, the information technology spans to every part of human life in all kinds of human activities. People in our modern society benefit a lot from the easy access of all kinds of software running on the hardware of controllers and processors. Supporting at the bottom level, it is the semiconductor device founded with the success of silicon planar technology. With the rapid scaling pace of Moore's law, semiconductor devices shrink with a smaller and smaller size, reducing the power cost, enabling increased transistor density. However, it is well known that there is a hard wall when the size is approaching atomistic level. High thermal density, intrinsic parameter fluctuations, and other effects are obstructing and slowing down the scaling of semiconductor devices. What kind of device, with

H. Wang (✉)
School of Physics and Technology, Wuhan University, Wuhan 430072, Hubei, China
e-mail: wanghao@whu.edu.cn

L. Zhang and M. Chan (eds.), *Tunneling Field Effect Transistor Technology*,
DOI 10.1007/978-3-319-31653-6_7

what structure, of what material can be used to meet the greedy demand of semiconductor industry driven by the information technology in the beyond-CMOS age?

Carbon is an element of many different structural forms. It is also a key element forming life, including our human beings. Several distinct kinds of valence bonds can be formed due to the special mechanism of the hybridization of orbits. Carbon atom has six electrons including two strongly bound core electrons and four weakly bound valence electrons. The four electrons can take part in the forming of covalent bonds with three kinds of hybridizations: sp, sp^2, and sp^3. These hybridizations enable carbon isomers from zero dimension to three dimension including fullerene, carbyne, nanotubes, graphene, and diamond [1].

The early report of carbon nanotube (CNT) can be traced back to 1952 [2]. It has been extensively studied since the report by Iijima in 1991 [3]. Although it is multi-wall CNT that is observed in Iijima's work, single-wall CNT serves a better vehicle for the studies. In this chapter, only single-wall CNT is considered. In semiconductor device area, CNT is attractive due to its peculiar electronic properties [4, 5]. It can be considered as a one-dimensional quantum wire [6]. The strong C-C sp^2 bonds lead to the inherent immunity from electromigration which is very attractive for interconnect applications. The mobility of both electron and hole is very high in semiconductor CNT. Due to the absence of dangling bonds, it is chemically very robust on the surfaces, leading to the good compatibility with high-κ dielectrics. It is also thermally robust for carrying large current density. The large carrier mean-free path makes the transport near ballistic at room temperature. The symmetry between the conduction and valence bands is beneficial for complementary applications. The semiconducting and metallic CNTs can be used in devices and interconnects, respectively. Due to these advantages, CNT has been focused on by the semiconductor device community as a promising material.

In the traditional MOSFETs, the carrier transport from the source to the drain is modulated with an in-channel barrier. The carrier comes from thermal injection in source/drain at equilibrium. It yields the well-known physical limitation [7] of the inverse subthreshold slope (S) of $S(T) = \log(10)\cdot k_B T/q$, where k_B is the Boltzmann constant, T is the temperature, and q is the electron charge. This physical limitation is about 60 mV/dec at room temperature, which is an obstacle for further decrease of supply voltage and power consumption.

The carrier transport in tunnel FET (TFET) is based on the band-to-band tunneling (BTBT) mechanism which is not subject to the 60 mV/dec limitation [7]. The low bandpass filter [6] like behavior makes carriers governed by the Fermi distribution effectively cooled down. Thus, the carrier injection in TFET is referred to as "cold injection" [7, 8], and the TFET is supposed to be "green" [9] transistor. The TFET is regarded as the leading optional structure in electronic devices [10]. The first demonstrated TFET is by Appenzeller et al. [11] and his colleagues at IBM in 2004 with CNT as the channel material. An S of about 40 mV/dec is observed. TFET quickly draws a lot of attention from the scientific community. It has been developing rapidly both in theory modeling [12–18] and experimental fabrication [19, 20] of TFET.

This chapter describes the ongoing efforts on the structure optimization of CNT-based TFET with numerical simulation using the non-equilibrium Green's function (NEGF) [21, 22] quantum transport framework.

7.2 Carbon Nanotube Band Structure

The electronic band structure of CNT determines the unique electronic properties. A CNT can be supposed to be a hollow cylinder rolled up from a two-dimensional single-layer graphene sheet. The graphene band structure is the starting point of CNT band structure [1].

7.2.1 Graphene Band Structure

Graphene is a two-dimensional sheet of infinite size. The real-space and reciprocal-space lattice of graphene are shown in Fig. 7.1. The unit vectors are as follows:

$$a_1 = a_0\left(\frac{\sqrt{3}}{2}\hat{x} + \frac{1}{2}\hat{y}\right), \quad a_2 = a_0\left(\frac{\sqrt{3}}{2}\hat{x} - \frac{1}{2}\hat{y}\right), \quad a_0 = \sqrt{3}a_{cc} \tag{7.1}$$

$$b_1 = b_0\left(\frac{1}{2}\hat{x} + \frac{\sqrt{3}}{2}\hat{y}\right), \quad b_2 = b_0\left(\frac{1}{2}\hat{x} - \frac{\sqrt{3}}{2}\hat{y}\right), \quad b_0 = \frac{4\pi}{3a_{cc}}, \tag{7.2}$$

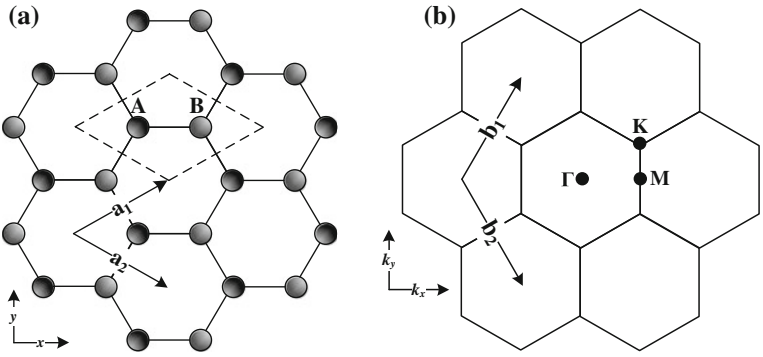

Fig. 7.1 **a** Real-space lattice of graphene. The unit cell is shown in the *dashed lines*, and the unit vectors are denoted with a_1 and a_2. **b** Reciprocal-space lattice of graphene. The unit vectors are denoted with b_1 and b_2. The high symmetry points are labeled with Γ, M, and K

where a_0 and b_0 are the lattice constants in real and reciprocal spaces, respectively, and a_{cc} is the nearest carbon atom distance.

There are two different sorts of carbon atoms, A and B, in graphene unit cell. In graphene (and CNT), the sp^2 hybridization is formed up with three valence electrons, in $2s$, $2p_x$, and $2p_y$ orbits. The other valence electron in $2p_z$ orbit is delocalized in π bonds and responsible for electrical transport properties.

The wave function of the delocalized single electron can be expressed with the linear combination of atomic orbitals (LCAO). In graphene, the wave function of π orbit electron $\Psi(r)$ can be written with the wave function of carbon atoms A and B.

$$\Psi(r) = c\sum_{r_A} e^{ikr_A}\xi(r - r_A) + d\sum_{r_B} e^{ikr_B}\xi(r - r_B), \tag{7.3}$$

where $\xi(r)$ is the electron wave function localized in $2p_z$ orbital in an isolated carbon atom. It can be substituted to the Schrodinger equation of

$$H\Psi(r) = E\Psi(r), \tag{7.4}$$

where H is the Hamiltonian and E is the energy. The Schrodinger equation becomes

$$\begin{aligned} c\sum_{r_A} e^{ikr_A}H\xi(r - r_A) + d\sum_{r_B} e^{ikr_B}H\xi(r - r_B) = Ec\sum_{r_A} e^{ikr_A}\xi(r - r_A) \\ + Ed\sum_{r_B} e^{ikr_B}\xi(r - r_B) \end{aligned} \tag{7.5}$$

With Eq. (7.5) left multiplied with $\xi^*(r - r_A)$, it leads to

$$\varepsilon_{2p}ce^{ikr_A} + td\sum_j e^{ikr_B} = Ece^{ikr_A} + sEd\sum_j e^{ikr_B}, \tag{7.6}$$

where

$$\begin{aligned} t &= \xi^*(r - r_A)H\xi(r - r_B) \\ s &= \xi^*(r - r_A)\xi(r - r_B) \\ \varepsilon_{2p} &= \xi^*(r - r_A)H\xi(r - r_A), \end{aligned} \tag{7.7}$$

where t is the coupling energy of $2p_z$ orbital between atoms A and B, s is the $2p_z$ wave functions overlap between atoms A and B, and ε_{2p} is the orbital energy of the $2p_z$ level, the summation over j runs over all neighboring unit cells including itself.

Equation (7.6) can be rewritten as follows:

$$c\left(E - \varepsilon_{2p}\right) + d(sE - t)\sum_j e^{ik(r_B - r_A)} = 0 \tag{7.8}$$

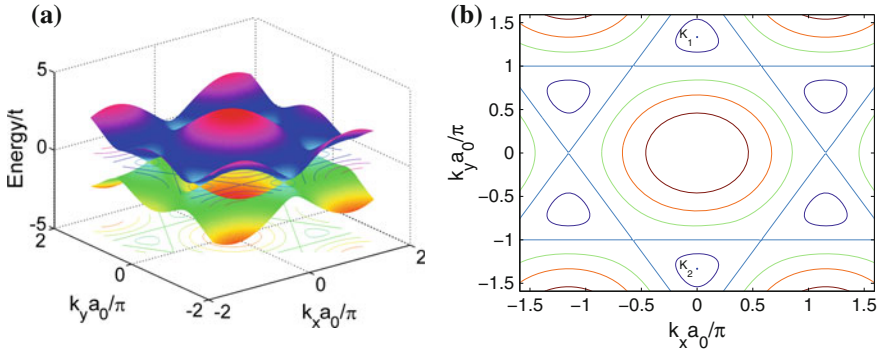

Fig. 7.2 **a** Surface plot of energy dispersion of graphene. The energy level is reduced with the hopping parameter t. **b** Contour plot of energy dispersion of graphene. Two non-equivalent K points of the six are denoted as K_1 and K_2

Similarly, by left multiplying Eq. (7.5) with $\xi^*(r - r_B)$, it leads to

$$d\left(E - \varepsilon_{2p}\right) + c(sE - t)\sum_j e^{ik(r_A - r_B)} = 0 \tag{7.9}$$

The wave function overlap s is very small and neglected, i.e., $s = 0$. Equations (7.8) and (7.9) can be solved for the energy dispersion. Considering only the nearest neighbor and running the summation over the three nearest atoms, the obtained energy dispersion can be calculated as follows:

$$E(k) = \varepsilon_{2p} \pm t\left[1 + 4\cos\left(k_y \frac{1}{2}a_0\right)\cos\left(k_x \frac{\sqrt{3}}{2}a_0\right) + 2\left(\cos\left(k_y a_0\right) + 1\right)\right]^{1/2}, \tag{7.10}$$

where k is the wave vector.

The energy dispersion is plotted in Fig. 7.2. It is shown that the conduction and valence bands meet at the six K points at the corners of the Brillouin zone with zero bandgap. With the nearest tight-binding approximation, the conduction and valence bands are fully symmetric. Note that there are two different types of non-equivalent K points, denoted as K_1 and K_2 as shown in Fig. 7.2.

7.2.2 CNT Band Structure

Carbon nanotube can be viewed as a rolled-up hollow sheet of graphene and uniquely specified with the chiral vector C_h. The chiral vector C_h is along the circumferential direction and forms up a periodic boundary condition. The translational vector T is along the nanotube axis and is normal to the chiral vector C_h.

The chiral vector can be expressed with the unit vector a_1 and a_2 of unrolled graphene lattice.

$$C_h = na_1 + ma_2, \quad (n, m \text{ are integers}, \ \ 0 \le |m| \le n) \tag{7.11}$$

For the case of $n = m$, i.e., $C_h = (n, n)$, the CNT is called armchair CNT. And for the case of $m = 0$, i.e., $C_h = (n, 0)$, which is called zigzag CNT. The diameter of the CNT d is as follows:

$$d = \frac{|C_h|}{\pi} = \frac{a_0\sqrt{n^2 + m^2 + nm}}{\pi} = \frac{a_{cc}\sqrt{3(n^2 + m^2 + nm)}}{\pi} \tag{7.12}$$

The translational vector T can be expressed as follows:

$$T = t_1 a_1 + t_2 a_2, \quad (t_1, t_2 \text{ are integers},) \tag{7.13}$$

As T is normal to C_h, i.e., $T \cdot C_h = 0$, the expression for t_1 and t_2 can be obtained as follows:

$$\begin{aligned} t_1 &= \frac{2m + n}{d_R}, \quad t_2 = -\frac{2n + m}{d_R} \\ d_R &= \begin{cases} d & \text{if} (n - m) \text{ is not a multiple of } 3 \\ 3d & \text{if } (n - m) \text{ is a multiple of } 3 \end{cases}, \end{aligned} \tag{7.14}$$

where d_R is defined with the d, which is the greatest common divisor of n and m. In each CNT unit cell, there are N graphene unit cells and $2N$ carbon atoms as follows:

$$N = \frac{2(n^2 + m^2 + nm)}{d_R}. \tag{7.15}$$

There are six K points in the Brillouin zone, and only two of them are independent. Due to the symmetric bands, they are also referred as the Fermi points. The energy dispersion near these points is most relevant for the transport properties. Let us choose the two points at

$$k_F = \left(0, \pm\frac{4\pi}{3a_0}\right) = \pm\frac{b_1}{3} \mp \frac{b_2}{3}. \tag{7.16}$$

With Taylor expansion at the points, the energy dispersion can be expressed as follows:

$$\begin{aligned} E(k) &= (k_x - 0)\left[\frac{\partial E}{\partial k_x}\right]_{k_F = \left(0, \pm\frac{4\pi}{3a_0}\right)} + \left(k_y \mp \frac{4\pi}{3a_0}\right)\left[\frac{\partial E}{\partial k_y}\right]_{k_F = \left(0, \pm\frac{4\pi}{3a_0}\right)} \\ &= \pm\frac{\sqrt{3}a_0 t}{2}|k - k_F|. \end{aligned} \tag{7.17}$$

Let us define

$$k' = k - k_F \tag{7.18}$$

and express k' along the chiral vector C_h direction (k'_c) and translational vector T direction (k'_t) as

$$k' = k'_c\hat{c} + k'_t\hat{t}, \tag{7.19}$$

where k'_t is continuous as the CNT is supposed to be infinitely long, and (k'_c) is quantized due to the periodical boundary condition determining the bandgap of CNT. Due to the wave vector quantization along the C_h direction,

$$k \cdot C_h = 2\pi q, \quad q \text{ is an integer.} \tag{7.20}$$

Along with Eqs. (7.16), (7.11), and (7.12),

$$k'_c = \frac{(k - k_F) \cdot C_h}{|C_h|} = \frac{2\pi(q \mp n \pm m)}{3|C_h|} = \frac{2(3q \mp n \pm m)}{3d}. \tag{7.21}$$

Substituting Eqs. (7.21) and (7.19) into (7.17), the energy dispersion of CNT can be obtained.

1. If $(n - m)$ mod 3 = 0, $\left|k'_c\right|_{\min} = 0$ and the bandgap of CNT is zero. The nanotube is metallic. For the lowest subband around the Fermi points, the energy dispersion and density of state $D(E)$ are

$$E(k) = \pm\frac{\sqrt{3}a_0 t}{2}\left|k'_t\right|, \quad D(E) = \frac{8}{\sqrt{3}\pi a_0 t}. \tag{7.22}$$

2. If $(n - m)$ mod 3 $\neq$ 0, the bandgap is not zero and the nanotube is semiconducting. The bandgap E_g is

$$E_g = 2E(k)_{\min} = \frac{\sqrt{3}a_0 t}{2}\left|k'_c\right|_{\min} = \frac{2a_{cc}t}{d} \approx \frac{0.8\,\text{eV}}{d(\text{in nm})}. \tag{7.23}$$

   And the bandgap of the ith subband is [23]

$$E_{gi} = \frac{2a_{cc}t}{d} \times \left[\frac{6i - 3 - (-1)^i}{4}\right], \quad i = 1, 2, 3\ldots \tag{7.24}$$

For each subband, the density of state is

$$D(E) = \frac{8}{3\pi a_{cc} t} \frac{|E|}{\sqrt{E^2 - \left(\frac{E_{gi}}{2}\right)^2} \Theta\left(|E| - \frac{E_{gi}}{2}\right)} \tag{7.25}$$

where $\Theta(x)$ is the step function as $\Theta(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}$.

Therefore, all (n, n) armchair CNTs are metallic. For a $(n, 0)$ zigzag CNT, it is metallic if n is a multiple of 3, and it is semiconducting if n is not a multiple of 3. In this chapter, only zigzag semiconducting CNTs are considered in the device simulations. Figures 7.3 and 7.4 show the band structure and DOS of an armchair (9, 9) and zigzag (13, 0) CNT. The (9, 9) CNT is metallic with zero bandgap, and the (13, 0) CNT is semiconducting with zero DOS at the bandgap region.

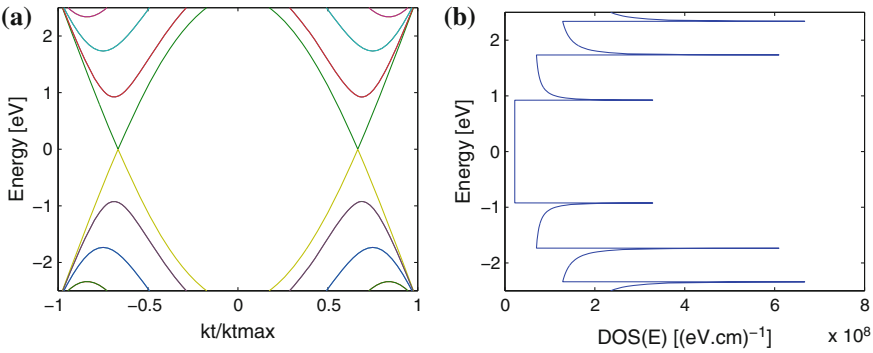

Fig. 7.3 **a** The band structure and **b** the DOS of (9, 9) armchair CNT. The DOS at the neutral point is not zero, and it is metallic with zero bandgap

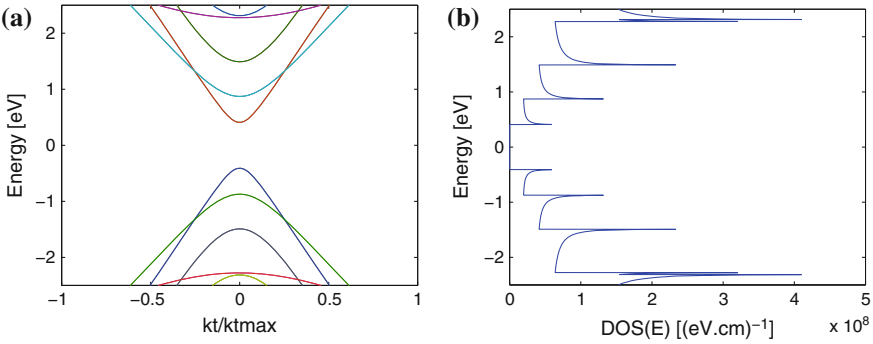

Fig. 7.4 **a** The band structure and **b** the DOS of (13, 0) zigzag CNT. The DOS at the neutral point is zero, and it is semiconducting with a nonzero bandgap

Obviously, the conduction and valence bands of the CNT are fully symmetric. In TFET, both the bands are totally involved. Thus, the CNT is not only the first demonstrated channel material [11] for TFET, but also an ideal vehicle in exploring TFET structure [24]. And it has been verified that the Hamiltonian with p_z orbit nearest neighboring tight binding can adequately capture the essential device physics [25].

7.3 NEGF Framework and the TCAD

In the optimization of the emerging semiconductor devices, the traditional trial-and-improve process is both time-consuming and resource-demanding. With the rapid development of information technology, computer becomes more and more powerful with reduced cost. Technology for computer-aided design (TCAD) becomes a powerful tool for researchers and engineers. Numerical simulation can be utilized in providing both guidelines for device improvement before fabrication, and deep physical insights into the phenomena observed in experiments.

7.3.1 The NEGF-Based Device Simulation Procedure

With the semiconductor device shrink to the nanoscale at molecular level, the treatment of electrons and holes as semiclassical particles becomes inappropriate. The carrier transport is imperative to be described as quantum mechanical entities at the atomistic scale. A direct solution to many-body Schrodinger equation would adequately provide accurate results, but the computation burden makes this approach unattractive.

In mathematics, Green's function is used to solve non-homogeneous boundary value problems for harmonic functions. The non-equilibrium Green's function theory was initiated in the condense matter physics area. With the mean field and single-particle approximation, the many-particle information is casted into self-energies. It is suitable for many-particle quantum system both in thermodynamic equilibrium and non-equilibrium. With a powerful conceptual and computational framework for carrier quantum transport, it has been widely adopted by the device physicists and engineers in molecular electronics and novel device modeling and simulation.

In the simulation of semiconductor device, there are two kernel governing physical equations [26]. One is the transport equation describing the carrier behavior driven by the electric field. The other is the Poisson equation describing the electric field induced by both outside bias and internal electron charge distribution. The two are strongly coupled to each other and need to be solved self-consistently. With the quantum transport theory, they are as follows:

$$\begin{cases} H\Psi = E\Psi \\ \nabla\varepsilon(\nabla\phi) = -q\left[-n(\phi) + p(\phi) + N_d^+ - N_a^-\right] \end{cases}, \tag{7.26}$$

where H is the Hamiltonian, E is the energy, ε is the dielectric constant, ϕ is the electrostatic potential, q is the electron charge, N_d^+ and N_a^- are ionized donor and acceptor concentrations, and n and p are electron and hole concentrations. In the Poisson equation, Dirichlet boundary conditions are used for all the gates with fixed bias, and Von Neumann boundary conditions are adopted for other area to let the potential float to any value with zero field [27]. The Schrodinger equation with open boundary conditions at the electronic terminals will be solved with NEGF technique.

In the NEGF formalism, the first step is to write down the Hamiltonian H of the device in an appropriate form. In the CNT device simulation, the nearest p_z orbitals nearest tight binding is adequate, as only the energy of the p_z orbitals is around the Fermi level and relevant in carrier transport. For a (n, 0) semiconducting zigzag CNT with N_c carbon rings, the matrix size of H is ($nN_c \times nN_c$). The open boundary conditions are handled in the self-energy matrices, Σ_s and Σ_d, where the indexes s and d denote for source and drain, respectively. With the nearest tight-binding model, the nonzero values in the Σ_s and Σ_d matrices are ($n \times n$). The self-energy represents the coupling between the device and the contacts due to the wave function interchange between the device and the outside electronic terminals. The electronic terminals, i.e., the source and drain, are supposed to be semi-infinite long reservoirs in thermal equilibrium characterized with the Fermi levels E_{Fs} and E_{Fd}. The self-energies can be obtained with a recursive relation [28] for the surface Green's function. Note that the self-energies are energy-dependent.

The retarded Green's function G is the key quantities in NEGF, and it can be expressed as follows:

$$G = [(E + i0^+)I - H - \Sigma_S - \Sigma_D]^{-1}. \tag{7.27}$$

As the matrix inversion is needed, the complexity of direct inverting is $O\left(n^3 \times N_c^3\right)$ [29]. Obviously, G is also energy-dependent and runs on all interested energies. Thus, the calculation of G is very demanding, especially when the matrix size is large. Many efficient techniques have been developed to reduce the computation burden. In the mode-space method [29, 30], the matrices H, Σ, and G are all transformed to a mode space. With only the lowest N_m ($N_m \ll n$) modes considered, the matrices size will be reduced from ($nN_c \times nN_c$) to ($N_mN_c \times N_mN_c$) [30]. In the recursive Green's function (RGF) algorithm, the block tridiagonal structure is exploited to reduce the complexity from $O\left(n^3 \times N_c^3\right)$ to $O(n^3 \times N_c)$ [31]. However, if the gate leakage current is accounted, the matrices would not be strictly block tridiagonal. Thus, the high-order elements will be neglected in RGF method underestimating the coupling effect through the gate terminal [32]. The contact block reduction method [33] is another efficient Green's function technique, which reduces the complexity to $\left(O\left(N_{eigen} \times N_A^2\right) + O\left(N_A^3\right)\right)$, where N_A is the size of the

boundary region coupling with outside contacts ($N_A = 2n$, here considering the source and drain) and N_{eigen} ($N_{eigen} \ll nN_c$) is the number of eigenstates used in spectral expansion for G.

Once G is obtained, all interested physical entities (e.g., the electron density, n, and the source-to-drain current, I) can be calculated. The electron density n can be calculated as

$$n = \frac{1}{\pi}\int dE[A_s f(E - E_{Fs}) + A_d f(E - E_{Fd})] \tag{7.28}$$

where A_s and A_d are the spectral functions for source and drain, respectively, and f is the Fermi distribution function. The spectral function can be expressed as

$$\begin{cases} A_s = G\Gamma_s G^+, \Gamma_s = i\left(\Sigma_s - \Sigma_s^+\right) \\ A_d = G\Gamma_d G^+, \Gamma_d = i\left(\Sigma_d - \Sigma_d^+\right) \end{cases}, \tag{7.29}$$

where Γ_s and Γ_d are the broadening matrices for the source and drain. The source-to-drain current can be calculated as

$$I = \frac{q}{\pi\hbar}\int dET(E)[f(E - E_{Fs}) - f(E - E_{Fd})] \tag{7.30}$$

where $\hbar$ is the reduced Planck's constant and T is the source-to-drain transmission coefficient which is computed as

$$T = \text{Trace}(\Gamma_s G \Gamma_d G^+). \tag{7.31}$$

Note that the matrices G, Σ, Γ, A, and T are all energy-dependent and needs to be calculated at all relevant energy level E. The spectral matrix A in (7.28) and the transmission coefficient T in (7.30) are small at very low energy, and the Fermi distribution function f is small for very high energy. Thus, only the intermediate energy levels are relevant in the calculations.

Once the charge concentration is obtained from NEGF in (7.28), the result needs to be substitute in the Poisson equation in (7.26) in iterative computation. Solving the two equations with simple relaxation leads to bad convergence due to the nonlinearity. In order to overcome this obstacle, a predictor–corrector method [34] is developed with an approximation expression of the electron density n as a function of the electrostatic potential ϕ. Thus, an explicit Jacobian can be obtained and the Poisson equation can be solved with Newton–Raphson iteration. The approximation relation between n and ϕ can be either deducted from perturbation theory [34] or with a simple exponential function [35]. A nonlinear iteration method named Anderson mixing is also found helpful for the convergence [36]. There are quite a few open-source semiconductor device simulators [37–39] with NEGF method by the device research community, which may help understand the principle and implementation of NEGF formalism.

7.3.2 The NEGF Simulation Toolkit of NanoTCAD ViDES

NanoTCAD ViDES [35, 39, 40] is a highly modularized extensible framework of two-dimensional and three-dimensional Schrodinger-Poisson solver powered with NEGF method developed by Giuseppe Iannanccone's group from University of Pisa. It is readily available to the scientific community through either the open-source release [39] or the online version in nanoHUB.org [41]. The calculation core is realized with C/FORTRAN of high computing performance. A wrapper in Python is used to integrate the kernel functions. Tunnel FET with two-dimensional (e.g., graphene, MoS_2) or three-dimensional material (e.g., CNT) can be simulated as well as MOSFET-like devices. With a Hamiltonian generation module, it is also feasible to simulate material not predefined. With the calculations implemented in the TCAD, the user can focus on the optimization of the device structure. The syntax of NanoTCAD ViDES is similar to other TCAD tools as listed in the Webpage [39]. All simulations in this work were performed with NanoTCAD ViDES.

7.4 Simple Gate-All-Around PIN TFET

7.4.1 Advantages and Problems

A tunnel FET is a gated p-i-n structure as shown in Fig. 7.5a. The source is p-doped, with the Fermi level in the valence band. The drain is n-doped, with the Fermi level in the conduction band. And the channel is intrinsic. Compared with the traditional MOSFET-like n-i-n structure as shown in Fig. 7.5b, the source dope type is different. Both the p-i-n and n-i-n structures are simulated with NEGF scheme using the p_z orbit atomistic description Hamiltonian with NanoTCAD ViDES.

A CNT with chirality of (13, 0) is considered. The bandgap is 0.75 eV with the carbon-to-carbon hopping energy of 2.7 eV. The diameter of CNT is 1 nm. The source and drain are doped with a molecular fraction of 5×10^{-3}, which is about 0.61/nm atoms compared to the carbon atom density of 122/nm. A 1-nm SiO_2 layer is used as the gate dielectric. The length of the CNT is 100 nm, with 20 nm source/drain and 60 nm channel. The gate is aligned to the channel with no overlap

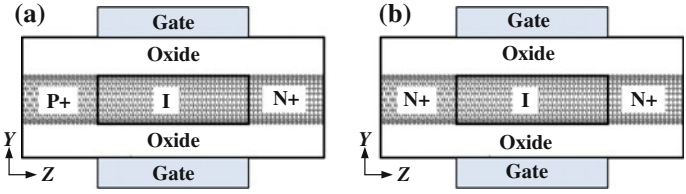

Fig. 7.5 Device structure of **a** p-i-n TFET and **b** n-i-n MOSFET

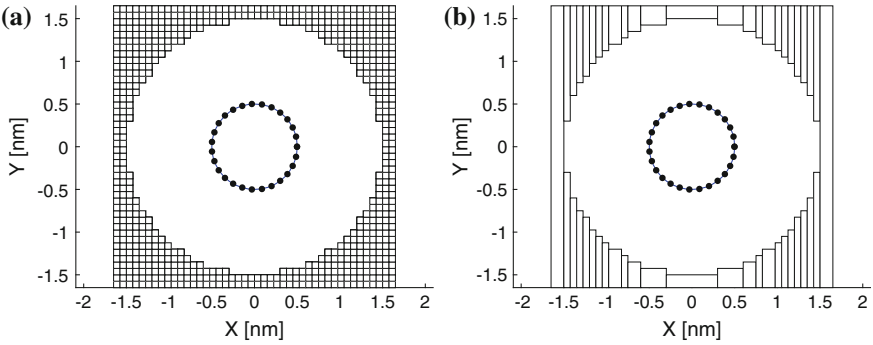

Fig. 7.6 Finite differential implementation of gate-all-around structure. **a** The grids in the gate area. **b** The defined hexahedron geometries in TCAD

or underlap. With a ballistic transport assumption, the device characteristic is not much sensitive to the channel length if only it is not too small (<10 nm). Gate-all-around structure is adopted for good electrostatic control. As both the Poisson equation and the NEGF equations are realized with finite differential method (FDM), the grids of the gate in the cross section as well as the carbon atoms are shown in Fig. 7.6a. These grids are defined with hexahedron geometries as shown in Fig. 7.6b.

A power supply of $V_{dd} = 0.4$ V is used. Both the drain and gate voltage are swept from 0 V to V_{dd} with the step of 0.025 V. Convergence is a general problem in computational electronics. It is related to a lot of factors including the grids, the device profiles, the physical model, and the initial guess. In general, smaller grids and good initial guesses are helpful for good convergence. The swept is started from flat band initial guess with zero drain/gate bias. For each bias point swept, only one voltage should be changed, either the drain or gate voltage. And the solution of one bias point can be used as the new initial guess for a new bias point.

The obtained output and transfer characteristics are shown in Fig. 7.7. From the transfer curves, it can be seen that the S of MOSFET submits to the 60 mV/dec physical limit and the TFET can break this limitation. In TFET, the S at small current is remarkably smaller than 60 mV/dec, implying a better switch device. This is the main advantage of TFET over MOSFET.

From the output curves, it can be seen that the maximum on current is notably smaller than the MOSFET with a similar doping profile. This is due to the special mechanism of TFET. The carrier must tunnel through the barrier to transport from the source to drain. It is well known that the tunnel probability through a barrier dependent exponentially on the barrier width. Thus, the tunnel current is very sensitive on the electrostatic profiles of the band diagrams. And the current is generally smaller than the barrier-controlled MOSFET device. In the simulated structure, the current of TFET is about 1/3 of the MOSFET. The current of real-fabricated TFET device would be smaller due to the non-ideal effects. Thus, the small on current is an important drawback of TFET.

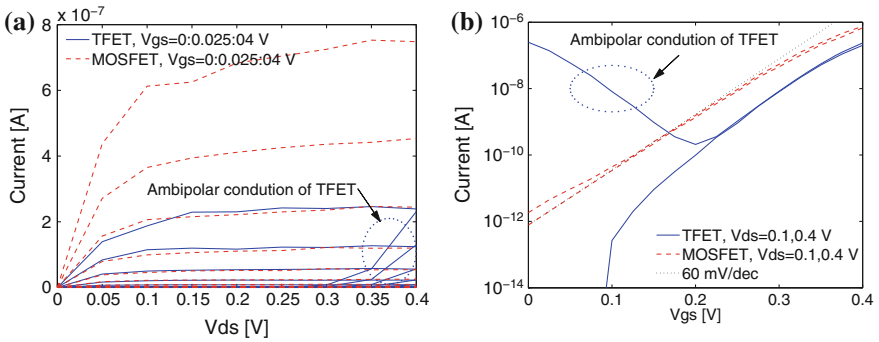

Fig. 7.7 The **a** output with Vgs swept from 0 to 0.4 V with a step of 0.025 V and **b** transfer characteristics of CNT-TFET and CNT-MOSFET. Severe ambipolar conduction can be observed in both the output and transfer curves. The current at (Vds = 0.4 V, Vgs = 0) is almost equal to the current at (Vds = 0.4 V, Vgs = 0.4 V) due to the full symmetric between the conduction band and valence band of CNT

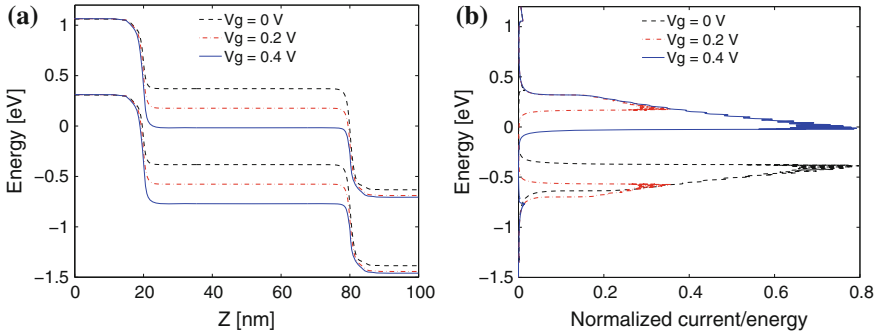

Fig. 7.8 TFET sturcture. **a** The band diagrams for different gate voltage with *Vds* = 0.4 V. **b** The current spectrum for different gate voltage with *Vds* = 0.4 V. Except for a small distribution higher than the minimum of conduction band at the source, the current spectrum mainly distributes at the energy window

It can be seen that the TFET exhibits severe ambipolar behavior. This phenomenon is similar to the Schottky-barrier FET [42] and MOSFET [43, 44] of CNT material. With the same doping concentration of the source and drain, the carrier tunnel at the source junction and the drain junction is equivalent with different types. The gate modulation on the current of TFET and MOSFET is shown in Figs. 7.8 and 7.9. With the increase of *Vgs*, the band diagrams are pushed down. For $Vgs < Vds/2$, the carrier mainly happen in the channel–drain junction, with carriers tunnel between the conduction band of channel and the valence band of the drain. And for $Vgs > Vds/2$, the tunnel current dominates at the source–drain junction with carriers tunnel between the conduction band of the source and valence band at the channel. The tunnel at the two junctions is equivalent for $Vgs = Vds/2$;

thus, the current spectrum is symmetric in the conduction band and valence band. And the minimum current is achieved at this point. Let us defined the off current at the minimum current point ($Vgs = Vds/2$ and $Vds = V_{dd}$) and the on current at the maximum current point ($Vgs = V_{dd}$, and $Vds = V_{dd}$). Obviously, the off current is rather large due to the ambipolar behavior. This is another main shortage of tunnel FET.

7.4.2 Tunnel Current Modulation Mechanisms

The gate controlling mechanism is different in TFET and MOSFET. As shown in Fig. 7.9, the current spectrum distributes in the conduction band higher than the barrier top in MOSFET. A tail current can be seen from the spectrum which is due to the carrier thermal injection. The current cannot be well turned off, and it determines the physical limitation of S. The gate pushes down the band diagrams in the channel to reduce the barrier height and increase the current. But the gate modulation in TFET is of two-folder effects. As discussed in [6, 45, 46], the tunnel current is determined by both the occupancy probability and the tunneling probability. On the one hand, with the band diagram pushed down in the channel region, the energy window overlaps between the valence/conduction band in the source/drain and the conduction/valence band in the channel changes. Band-to-band tunneling is only possible at the energy window region. On the other hand, the band-to-band tunneling barrier width is also modulated by the gate voltage. With the increase of gate voltage and band bending, the tunnel barrier width is also reduced; thus, the tunnel probability is increased.

The first effect is due to the modulation of the occupancy probability, and the second is due to the tunneling probability. The occupancy probability is determined by the availability of carrier in the origination of tunnel and the availability of

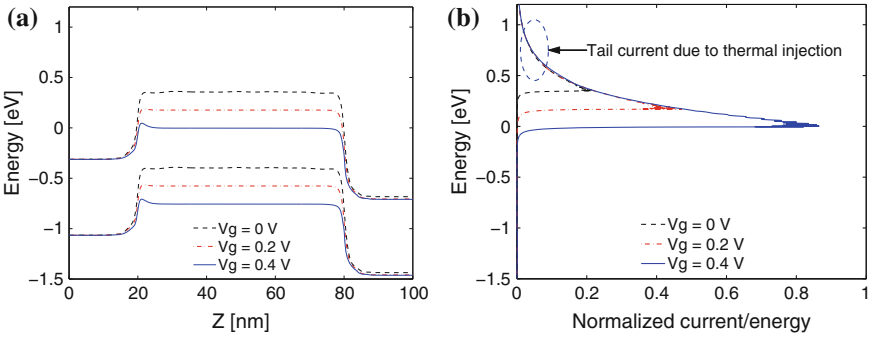

Fig. 7.9 MOSFET structure. **a** The band diagrams for different gate voltage with $Vds = 0.4$ V. **b** The current spectrum for different gate voltage with $Vds = 0.4$ V. Obviously, the current spectrum is mainly higher than the top of barrier in the channel

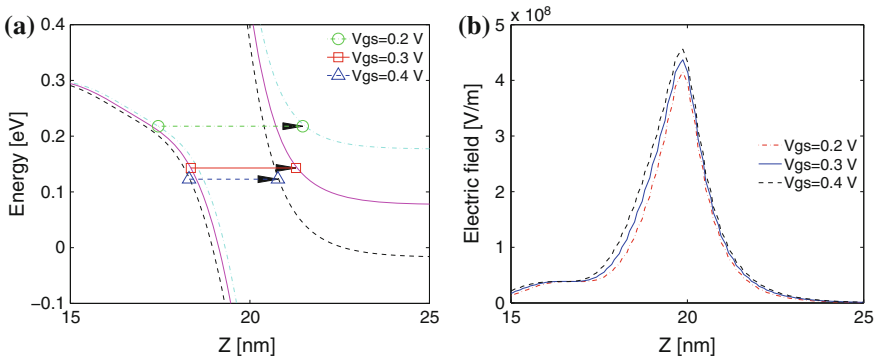

Fig. 7.10 **a** The band-to-band tunnel path at the source junction (Z = 20 nm) with different gate voltages. **b** The electric distribution at the source junction with different gate voltages

empty states of carriers in the destination [45]. It can be measured with the band alignment of the bands. This mechanism is also referred to as a band filtering effect [6], as shown in the current spectrum of TFET. The tunnel probability increases with a smaller imaginary wave vector of tunneling carrier [47], which is mainly determined by the material. For a given material, the tunnel probability can be modulated with the tunnel path [46].

As shown in Fig. 7.10a, the carrier band-to-band tunnel at the source–channel junction along the shortest distance is largest. And the shortest tunnel path decreases with the increased gate voltage due to the increased band diagram abruptness. Generally, a short tunnel path means a large electric field, and vice versa. Thus, in addition to the tunnel path, the tunnel probability can also be characterized with the electric filed. With the increased electric field, the band diagram changes more rapidly. Usually, the gate is aligned with the source–channel junction to let most band drop happens at the junction. From a qualitative point of view, there is a near-exponential dependence of tunnel probability T on the electric field E [47] as

$$T \propto \exp\left(-E^{-1}\right). \tag{7.32}$$

The electric field is not uniform in the tunnel area, and the maximum electric field is obtained at the junction. It is shown in Fig. 7.10b that the electric field increases with the increased gate voltage.

In summary, the TFET achieves a steeper S at the cost of reduced on current, and a large off current due to ambipolarity. The gate voltage modulates both the occupancy and the tunnel probability. The occupancy probability can be modulated with the band alignment in a bandpass filtering behavior. The tunnel probability can be increased with a smaller tunnel path or equivalently a larger electric field.

7.5 Optimization with Doping Engineering and Gate Dielectric Engineering

The main objects in TFET structure optimization are to mend up the shortages of TFET without hurting the benefits. Specifically, the on current needs to be boosted and the ambipolar leakage current needs to be reduced. And the smaller than 60 mV/dec *S* needs to be kept. Carbon nanotube can either be chemically doped or electrostatically doped with similar effects of change in the band diagrams. High-κ and low-κ material can be used in the gate dielectric. And the gate can be of different materials. Channel dope engineering, gate dielectric engineering, and gate work function engineering are the main measurements considered in the context.

7.5.1 Pocket Doping in the Source Side Junction

In order to improve the on current and reduce the ambipolar current, the band bending needs to be sharp in source tunnel junction and gentle in the drain tunnel junction. In the heterojunction TFET [48], the source and channel are composed of different materials. Thus, the source is of a smaller bandgap material, and the channel is of a lager bandgap material. The tunnel path at the heterointerface is reduced. In the n-pocket doping TFET [19], a thin n-layer is inserted in the source–channel junction increasing the band bending at the tunnel junction. Obviously, the heterojunction structure is not applicable in CNT-TFET. The pocket doping structure is an effective measurement in CNT-TFET. Its effects of increasing the on current will be explored.

The TFET with an n-pocket doping is shown in Fig. 7.11. The nanotube is of a p-n-i-n doping structure compared with the normal p-i-n structure. The thin n-layer is doped with 2×10^{-3} with a length of 10 nm. The drain is lightly p-doped with 5×10^{-4} in order to reduce the ambipolar behavior.

Figure 7.12 show the output and transfer curves of PIN TFET and PNIN TFET. The on current is obviously increased with steeper than 60 mV/dec *S*. The corresponding on state band diagrams and electric field are shown in Fig. 7.13. The band diagrams are similar except for the n-pocket region. The band bending is shaper in the PNIN TFET structure, and the tunnel path is reduced. Correspondingly, the electric field is notably increased at the tunnel junction. The n-pocket doping

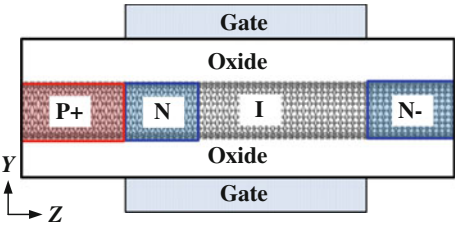

Fig. 7.11 The p-n-i-n TFET with an n-pocket doping at the source junction

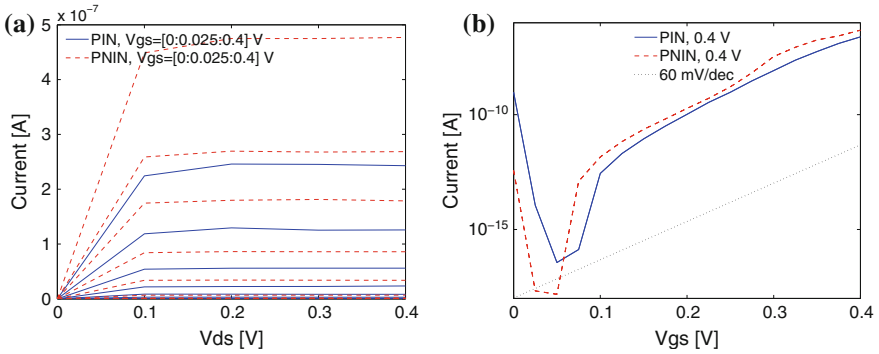

Fig. 7.12 The **a** output with Vgs swept from 0 to 0.4 V with a step of 0.025 V and **b** transfer characteristics of PIN TFET and PNIN TFET

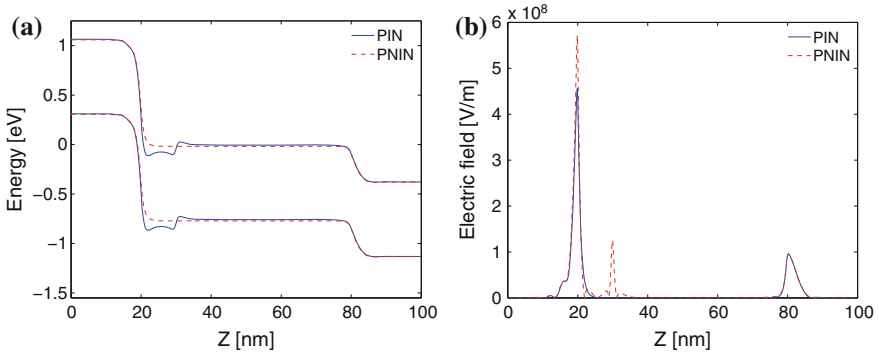

Fig. 7.13 The on state **a** band diagrams and **b** electric field distribution of PIN TEFT and PNIN TFET

technique [19] has been widely adopted and studied in TFET structure. The device performances can be further optimized by changing the size, position, and concentration of the doping region. And the PNIN structure is also found helpful for the device reliability [49].

7.5.2 *Doping Engineering in the Drain Side Junction*

The channel–drain junction tunnel causes the ambipolarity with large leakage current which is harmful. The doping engineering can be applied at the drain-side junction to reduce the unwanted tunnel for smaller leakage current.

In Section 7.5.1, the drain is doped at a lower level to reduce the ambipolar leakage current, but this will lead to poor ohmic contacts at the terminal.

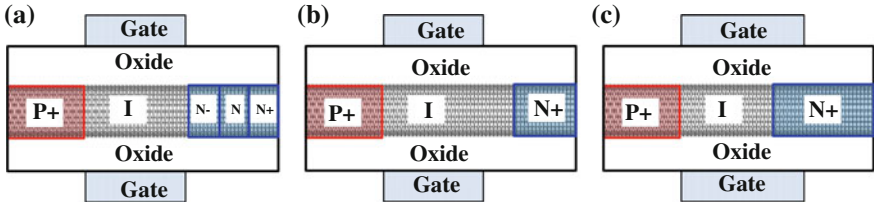

Fig. 7.14 TFET structure with **a** staircase doping, **b** drain underlap, **c** drain overlap

An equivalent doping density of the source and drain would be preferred. The ambipolar behavior has also been a problem in traditional CNT with both Schottky-barrier FETs and MOSFETs. Linear doping [43] and staircase doping [44] profiles have been used to relieve this problem. With a similar principle, three structures will be simulated and compared. They are shown in Fig. 7.14 as follows: (1) TFET with staircase doping at the drain side; (2) TFET with drain underlap [50]; (3) TFET with drain overlap [51]. Note that the latter two can be viewed as special cases of the first one.

The source side of the TFETs is doped with a molecular fraction of 5×10^{-3}. The drain side is designed with doping engineering. In the staircase doping structure, there are three segments with different doping levels. The N+ segments are doped with a molecular fraction of 5×10^{-3}. The N segment is lightly doped with a molecular fraction of 2×10^{-3}, and the N− segment is doped with a molecular fraction of 5×10^{-4}. The underlap/overlap between the channel and drain is 20 nm in the other two structures. Note that the parameters can be optimized for better performances.

The obtained band diagrams at off state ($Vds = 0$, $Vgs = V_{dd}$) as well as the transfer curves are shown in Fig. 7.15. It is shown that the drain-side leakage current has been remarkably reduced with the structures. In the drain underlap structure, the band diagram change gently in the drain side. The leakage current is reduced due to increased tunnel path. In the drain underlap structure, there is no abrupt change. In the drain overlap structure, there are two abrupt changes due to the drain doping and gate voltage, respectively. But the band change ΔE is smaller than the bandgap as shown in Fig. 7.15a; thus, the abrupt change does not lead to an available tunnel path.

7.5.3 *Low-κ Window in the Source*

The gate dielectric engineering is an important measurement in device optimization. In tradition MOSFET, high-κ material is well known to be beneficial. In TFET, several different structures have been proposed with combined high-κ and low-κ materials [52–54]. The key point is to improve the electric field to increase the carrier tunnel.

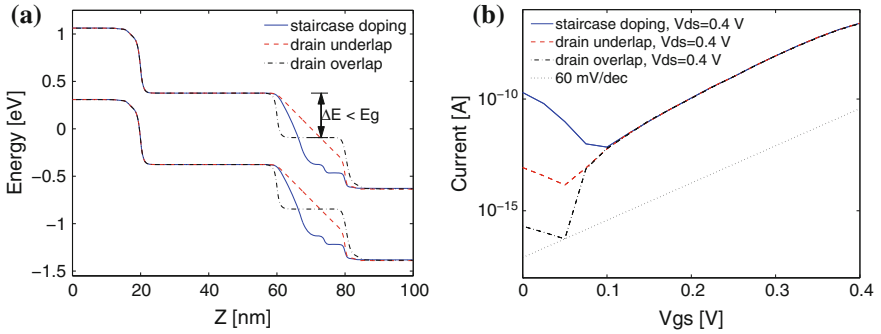

Fig. 7.15 The off-state band diagrams (**a**) and transfer curves (**b**) of different drain-side doping profiles

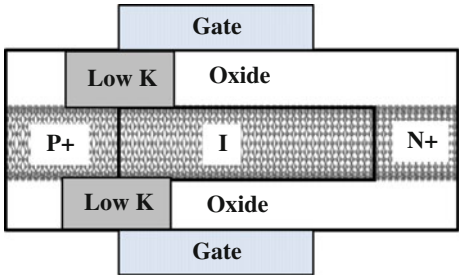

Fig. 7.16 The structure of TFET with a low-κ window

As shown in Fig. 7.16, low-κ material is used in the source–channel junction area. The structure is compared with high-κ only and low-κ only TFETs. The window size is 4 nm to the left side of the source–channel junction and is 2 nm to the right side of the source–channel junction. In the high-κ only structure, HfO_2 is used as the dielectric, and SiO_2 is used in the low-κ only structure. In the low-κ window structure, SiO_2 is used in the low-κ window and with HfO_2 in other areas. A gate–drain underlap is used to restrain the leakage current in the drain side.

The transfer curves and on state electric field distributions are shown in Fig. 7.17. The on current of the low-κ window structure and the low-κ only structure is notably larger than the high-κ only structure. The electric field of the low-κ window structure is almost the same with the low-κ only structure in the low-κ window area and is similar with the high-κ only structure. It can be confirmed that the on current is boosted with the low-κ material at the tunnel junction.

In summary, the TFET performances are optimized by modulation of the electric field and junction abruptness. The BTBT can be enlarged with increased junction abruptness and high electric field. On the other hand, the leakage current can be restrained with gentle band diagram and reduced electric field. It is worth noting that abrupt band diagram does not necessarily imply a large leakage current if the band change ΔE is smaller than the bandgap E_g.

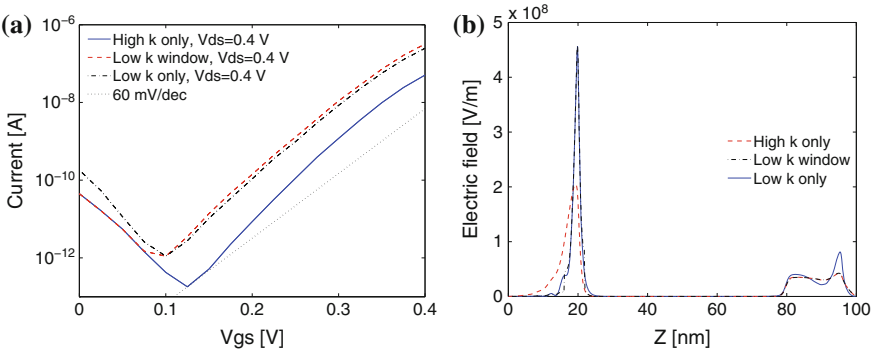

Fig. 7.17 The transfer curves (**a**) and the electric field distributions (**b**) of the low-κ window TFET structure compared with the high-κ only or low-κ only structure

7.6 Barrier-Controlled Tunnel FET

In order to mend up the shortage of TFET, a lot of work has been performed with various principles. It is well known that neither the small on current nor the large leakage current is a severe problem in traditional barrier-controlled MOSFET devices. The barrier-controlled tunnel FET (BC-TFET) is a structure that combines the merits of both TFET and traditional MOSFET [55]. The principle is shown in Fig. 7.18. It can be viewed as a normal TFET in series with a MOSFET. The carrier injection at the source side is determined by the internal TFET with cold carrier injection leading to the smaller than 60 mv/dec *S*. And the current modulation is mainly accomplished with the internal MOSFET. The gate voltage is used to modulate both the internal TFET and MOSFET. The barrier of the internal MOSFET can be formed up with different mechanism. The examples with gate work function engineering and doping engineering will be introduced in this section.

7.6.1 BC-TFET with Gate Work Function Engineering

Gate work function engineering has been adopted for device optimization in both TFET and MOSFET. Multi-segment gate has been reported in carbon nanotube

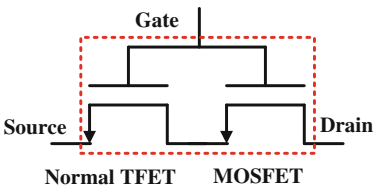

Fig. 7.18 The BC-TFET can be viewed as a normal TFET in series with a MOSFET

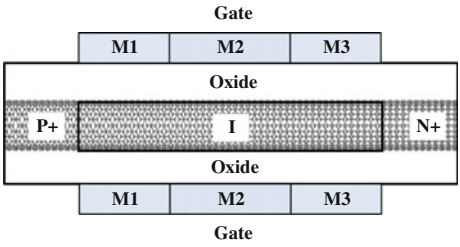

Fig. 7.19 The structure of BC-TFET with gate work function engineering

device fabrication [56]. The BC-TFET structure with gate work function engineering is shown in Fig. 7.19. The gate is composed of several segments with different materials with different work functions.

The three segments of gate are labeled with M1, M2, and M3, respectively. And their work functions are denoted with W_{M1}, W_{M2}, and W_{M3}, respectively. The source and drain are heavily doped with P+ and N+ type, respectively. And the channel is intrinsic. Referring to the principle in Fig. 7.18, the source and M1 forms up an internal TFET. With W_{M1} and W_{M3} smaller than W_{M2}, the internal MOSFET is formed up with M1, M2, and M3, respectively. With a given gate voltage, there will be a barrier in M2 region due to the work function difference. The barrier can be used to modulate the source-to-drain current. Thus, the structure is named as barrier-controlled TFET.

The device parameters of the BC-TFET are as follows. The CNT is of 100 nm long with the chirality of (13, 0). The bandgap of the nanotube is 0.75 eV, and the diameter is about 1 nm. Both the source and drain are 20 nm long. And their molecular doping density is of 5×10^{-3}. And the length of M1, M2, and M3 gate region is 10, 30, and 20 nm, respectively. The gate work function of M2 is set the same as CNT (W_{CNT}). And W_{M1} and W_{M3} are smaller than W_{M2} as $W_{M1} = W_{M3} = W_{M2} - 3/4E_g$.

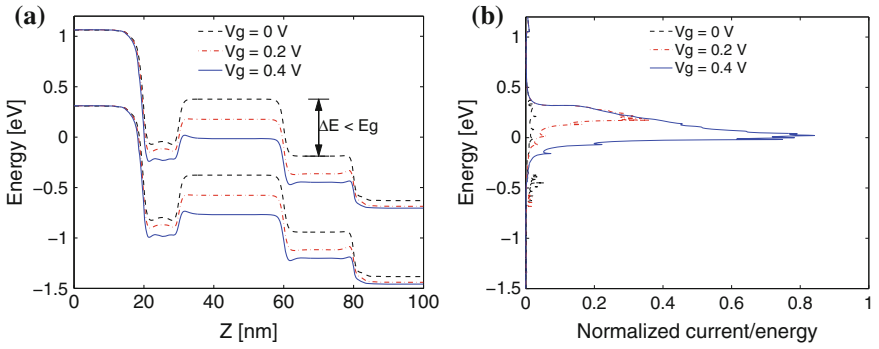

Fig. 7.20 BC-TFET structure. **a** The band diagrams for different gate voltage with Vds = 0.4 V. **b** The current spectrum for different gate voltage with Vds = 0.4 V

The gate modulation of current is shown in Fig. 7.20. The results for BC-TFET can be compared with the previous results for normal TFET and MOSFET. Compared with the normal TFET as shown in Fig. 7.8, the current spectrum of BC-TFET also resides in an energy window. The upper tail is cut off by the source valence band and channel conduction band. But it is different from the normal TFET that there is only one spectrum peak in BC-TFET as the drain current leakage due to ambipolarity is restrained. Compared to the MOSFET as shown in Fig. 7.9, the band diagrams of the M1, M2, and M3 regions are very similar to MOSFET. The current is controlled with the barrier at M2 region. Thus, the ambipolar behavior is restrained. On the other hand, in MOSFET, the tail current (see Fig. 7.9) is determined with the production of DOS and Fermi function, leading to the physical limitation of 60 mV/dec S. In BC-TFET, there is no such a limitation as the tail current is cut off with the energy band diagrams.

The BC-TFET structure can be compared with other measurements in TFET optimization. The effect of M1 segment gate is similar to the pocket doping in the PNIN structure. The junction abruptness is formed with both doping density difference and the source–channel junction and the work function difference of W_{M1} and W_{CNT}. The role of M3 segment gate is similar to the drain overlap structure. The abrupt band diagram is buffered with the M3 segment. Thus, the energy band change ΔE between different segments becomes smaller than the bandgap E_g to reduce the leakage current. As ΔE is smaller than E_g, M2 and M3 can be viewed as a whole bandpass filter. Thus, the BC-TFET structure can be compared with the heterojunction technique used in Si/Ge TFET. Obviously, a heterojunction cannot be formed up in the CNT. But if we view M2 and M3 as a whole, then the combination of M2 and M3 can be supposed to be an equivalent wideband material compared to the M1 segment of narrowband. The thinner bandgap (M1) material is used in the source–channel side, and the thicker bandgap (combination of M2 and M3) material is used in the channel–drain side, so that the BC-TFET can be viewed as an analogue to the heterojunction TFET structure.

To better understand the principle of BC-TFET, simulations will be performed in comparison with the normal TFET and dual-material gate TFET (DMG-TFET) [57]. The normal TFET can be viewed as $W_{M1} = W_{M2} = W_{M3} = W_{CNT}$. Two different DMG-TFETs will be simulated. They are $W_{M1} = W_{M2} = W_{CNT}$, $W_{M3} = W_{CNT} - 3/4E_g$, denoted with DMG-TFET 1, and $W_{M1} = W_{CNT} - 3/4E_g$, $W_{M2} = W_{M3} = W_{CNT}$, denoted with DMG-TFET 2. The transfer curves and band diagrams of the different structures are shown in Fig. 7.21. In normal TFET, severe ambipolar behavior is observed. The leakage current is reduced in other three structures due to the increased width of bandpass filter. The BC-TFET shows the smallest leakage current as the barrier due to the work function difference of the three gate segments. In DMG-TFET 2 and the BC-TFET, the on current is larger than the other two due to the increased band bending at M1 segment. In summary, BC-TFET shows the best performance both in on current and off current.

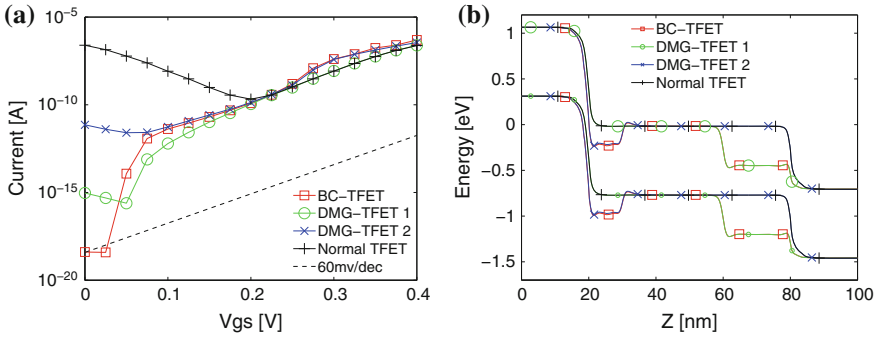

Fig. 7.21 **a** The transfer curves and **b** the on state band diagrams of different structures

7.6.2 An Alternative BC-TFET Structure Implementation

The BC-TFET described in Sect. 7.6.1 is realized with gate work function engineering. It is also feasible with other measurement. As an example, a BC-TFET implementation with the combination of gate work function and channel doping engineering [58] will be shown in this section. As depicted in Fig. 7.22, the BC-TFET is realized with a PNIN structure. Different from the PNIN BC-TFET, an N-pocket doping is utilized to increase the band bending at the source–channel junction. The gate is of two segments, M1 and M2, with the work functions of W_{M1} and W_{M2}. The parameters are as follows: $W_{M1} = W_{CNT}$, $W_{M2} = W_{M1} - 3/4E_g$. The lightly N doping density is 3×10^{-3}.

The obtained on/off band diagrams and electric field distributions are shown in Fig. 7.23. It is shown that similar band diagram to PNIN BC-TFET is obtained. The N-pocket doping in the channel near the tunnel junction plays the same role with the M1 gate segment in the PNIN BC-TFET. There are three submits of electric field due to the band bending at both on and off state. The device is turned on/off with the in-channel barrier. The current is cut off with the alignment of conduction band of the source and the valence band of the channel although the electric filed at the source–channel junction is still rather large at off state.

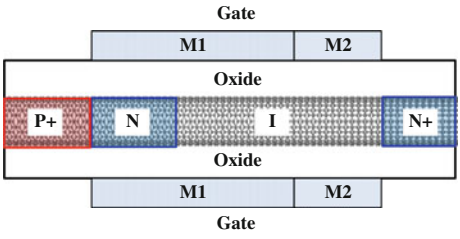

Fig. 7.22 A PNIN BC-TFET implemented with both gate work function and doping engineering

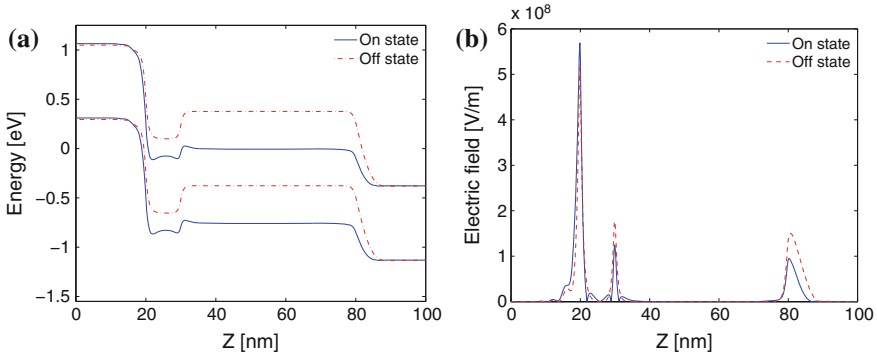

Fig. 7.23 The on/off **a** band diagrams and **b** electric field distributions of PNIN BC-TFET

7.6.3 The Concept Analysis of BC-TFET

In TFET, the tunnel current is modulated with two main factors. On the one hand, the bandgap of the material plays the row of bandpass filter. The alignment of different of band diagrams affects the leakage current. On the other hand, at the energy window where the tunnel is possible, the tunnel path determines the magnitude of the tunnel probability and current. In normal TFET, both the two factors are directly controlled with the gate voltage. But it is different in BC-TFET. As shown in Fig. 7.24a, the shortest tunnel path between the valence band of source and conduction band of the channel is at a lower energy than the top of barrier energy. Therefore, carrier tunnel at this shortest tunnel path is forbidden due to the bandpass filter of the in-channel barrier. Only the tunnel path at a higher level than the in-channel barrier is allowed, and it is referred to as the effective tunnel path. With the increase of gate voltage, the top of barrier is reduced monopoly as shown

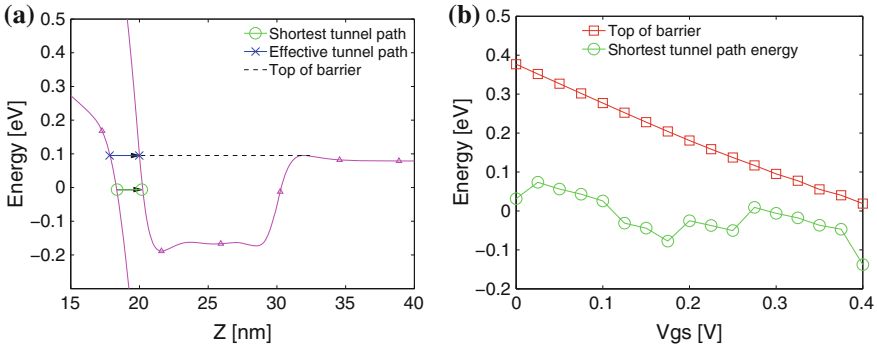

Fig. 7.24 **a** The depiction of shortest tunnel path and effective tunnel path in BC-TFET. **b** The shortest tunnel path energy and top of barrier energy as a function of gate voltage

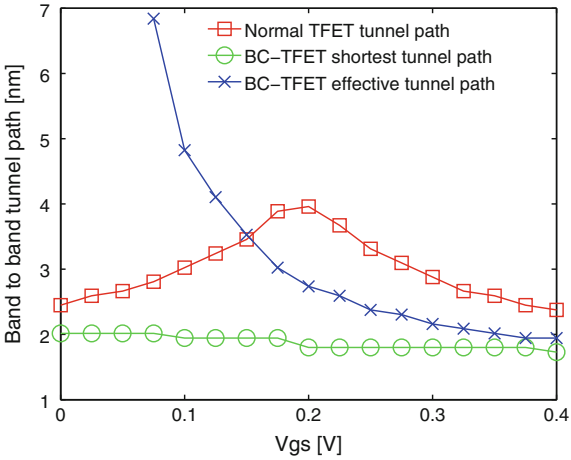

Fig. 7.25 Barrier modulation of the tunnel path in BC-TFET compared with the normal TFET

in Fig. 7.24b. The shortest tunnel path energy level is not monopoly with the gate voltage, and it is generally much lower than the top of barrier.

As a result, the top of barrier is modulated with the gate voltage, and the effective tunnel path is determined by the top of barrier energy. It is shown in Fig. 7.25 that the effective tunnel path deceases monopoly with the increase of gate voltage. But the shortest tunnel path is mainly determined by the difference gate work function and doping density between the source and the channel. And it changes much gently with the gate voltage. Note that there is no available effective tunnel path as the top of the in-channel barrier is higher than the maximum valence band of the source when $Vgs < 0.075$ V. In normal TFET, the tunnel path is directly modulated with the gate voltage. It increase with the gate voltage when $Vgs < V_{dd}/2$ and decrease when $Vgs > V_{dd}/2$. Thus, the tunnel current decreases and then increases with gate voltage as shown in Fig. 7.21a. But in BC-TFET, the effective tunnel path determined by the in-channel of barrier is modulated with the gate voltage.

In summary, BC-TFET is superior because it combines the merits of both TFET and traditional MOSFET. The barrier control mechanism improves the gate modulation mending up the weakness of large leakage current of TFET. The increased band bending at the tunnel junction boosts the on current. And the cold carrier injection keeps the advantage of steep S of TFET.

7.7 Summary

In addition to the material of the first fabricated TFET, the unique band structure of CNT makes it an ideal vehicle for studying the physics of TFET. With the nearest p_z orbit tight-binding method, the band structure is obtained from two-dimensional

graphene. The quantum transport simulation framework of NEGF is revisited with both the theory and numerical implementation. The TCAD powered with NEGF is introduced and adopted in TFET structure optimization. With comparison between MOSFET and TFET of CNT material, TFET is superior with a smaller than 60 mV/dec *S,* breaking the physical limitation of MOSFET restrained with carrier thermal injection. The small on current and ambipolar conductivity are the main shortages of TFET.

The on current in TFET is mainly determined with both the occupancy probability and the tunneling probability. The occupancy probability is determined by the availability of the initial states and final states of tunneling which can be understood with the bandpass filtering analogue. The tunneling probability is determined with the tunnel path or, equivalently, the electric field at the tunnel junction.

The TFET structure can be optimized with manipulation of both the occupancy probability and the tunnel probability. Examples are presented with doping engineering, gate dielectric engineering, and gate work function engineering. The electric field at the source–channel junction is enlarged to increase the tunneling probability in the pocket doping structure and low-κ window structure. And it is decreased at the channel–drain junction to reduce the leakage current in the staircase doping structure and gate–drain underlap structure. The occupancy probability is reduced with the misalignment of the band diagrams in the gate–drain overlap structure and the BC-TFET structure. The BC-TFET structure combines the merits of both TFET and tradition MOSFET and exhibits both a high on current and a small leakage current.

References

1. R. Saito, G. Dresselhaus, M.S. Dresselhaus, *Physical Properties of Carbon Nanotubes* (World Scientific Publishing Company, 1998)
2. L. Radushkevich, V. Lukyanovich, On the structure of carbon formed by the thermal decomposition of carbon monoxide (CO) to the contact with iron. Russ. J. Phys. Chem. **26**, 88 (1952)
3. S. Iijima, Helical microtubules of graphitic carbon. Nature **354**, 56–58 (1991)
4. A. Javey, J. Kong, *Carbon Nanotube Electronics* (Springer, Berlin, 2009)
5. H.S.P. Wong, D. Akinwande, *Carbon Nanotube and Graphene Device Physics* (Cambridge University Press, Cambridge, 2011)
6. J. Knoch, S. Mantl, J. Appenzeller, Impact of the dimensionality on the performance of tunneling FETs: bulk versus one-dimensional devices. Solid-State Electron. **51**, 572–578 (2007)
7. J. Appenzeller, L. Yu-Ming, J. Knoch, C. Zhihong, P. Avouris, Comparing carbon nanotube transistors—The ideal choice: a novel tunneling device design. IEEE Trans. Electron Devices **52**, 2568–2576 (2005)
8. J. Appenzeller, J. Knoch, M.T. Bjork, H. Riel, H. Schmid, W. Riess, Toward nanowire electronics. IEEE Trans. Electron Devices **55**, 2827–2845 (2008)
9. P.A. Patel, Steep turn on/off 'Green' tunnel transistors. Ph.D. thesis, EECS Department, University of California, Berkeley (2010)

10. D.E. Nikonov, I.A. Young, Overview of beyond-CMOS devices and a uniform methodology for their benchmarking. Proc. IEEE **101**, 2498–2533 (2013)
11. J. Appenzeller, Y.M. Lin, J. Knoch, P. Avouris, Band-to-band tunneling in carbon nanotube field-effect transistors. Phys. Rev. Lett. **93**, 196805 (2004)
12. L. Zhang, X. Lin, J. He, M. Chan, An analytical charge model for double-gate tunnel FETs. IEEE Trans. Electron Devices **59**, 3217–3223 (2012)
13. Y. Taur, J. Wu, J. Min, An analytic model for heterojunction tunnel FETs with exponential barrier. IEEE Trans. Electron Devices **62**, 1399–1404 (2015)
14. E.-H. Toh, G. H. Wang, L. Chan, G. Samudra, Y.-C. Yeo, Device physics and guiding principles for the design of double-gate tunneling field effect transistor with silicon-germanium source heterojunction. Appl. Phys. Lett. **91**, 243505-3 (2007)
15. O.M. Nayfeh, C.N. Chleirigh, J. Hennessy, L. Gomez, J.L. Hoyt, D.A. Antoniadis, Design of tunneling field-effect transistors using strained-silicon/strained-germanium type-II staggered heterojunctions. IEEE Electron Device Lett. **29**, 1074–1077 (2008)
16. K. Ganapathi, S. Salahuddin, Heterojunction vertical band-to-band tunneling transistors for steep subthreshold swing and high on current. IEEE Electron Device Lett. **32**, 689–691 (2011)
17. M. Liu, Y. Liu, H. Wang, Q. Zhang, C. Zhang, S. Hu et al., Design of GeSn-based heterojunction-enhanced N-Channel tunneling FET With improved subthreshold swing and ON-State current. IEEE Trans. Electron Devices **62**, 1262–1268 (2015)
18. D.B. Abdi, M.J. Kumar, In-built N+ pocket p-n-p-n tunnel field-effect transistor. IEEE Electron Device Lett. **35**, 1170–1172 (2014)
19. R. Jhaveri, V. Nagavarapu, J.C.S. Woo, Effect of pocket doping and annealing schemes on the source-pocket tunnel field-effect transistor. IEEE Trans. Electron Devices **58**, 80–86 (2011)
20. V. Nagavarapu, R. Jhaveri, J.C.S. Woo, The tunnel source (PNPN) n-MOSFET: a novel high performance transistor. IEEE Trans. Electron Devices **55**, 1013–1019 (2008)
21. S. Datta, Nanoscale device modeling: the Green's function method. Superlattices Microstruct. **28**, 253–278 (2000)
22. S. Datta, *Quantum Transport: Atom to Transistor* (Cambridge University Press, Cambridge, 2005)
23. S.O. Koswatta, N. Neophytou, D. Kienle, G. Fiori, M.S. Lundstrom, Dependence of DC characteristics of CNT MOSFETs on bandstructure models. IEEE Trans. Nanotechnol. **5**, 368–372 (2006)
24. S.O. Koswatta, M.S. Lundstrom, D.E. Nikonov, Performance comparison between p-i-n tunneling transistors and conventional MOSFETs. IEEE Trans. Electron Devices **56**, 456–465 (2009)
25. S. Bruzzone, G. Iannaccone, N. Marzari, G. Fiori, An open-source multiscale framework for the simulation of nanoscale devices. IEEE Trans. Electron Devices **61**, 48–53 (2014)
26. D. Vasileska, S.M. Goodnick, *Computational Electronics* (Morgan & Claypool Publishers, 2006)
27. Z. Ren, Nanoscale MOSFETs: physics, simulation and design. Ph.D. thesis, Purdue University (2006)
28. R. Lake, G. Klimeck, R.C. Bowen, D. Jovanovic, Single and multiband modeling of quantum electron transport through layered semiconductor devices. J. Appl. Phys. **81**, 7845–7869 (1997)
29. M.P. Anantram, M.S. Lundstrom, D.E. Nikonov, Modeling of nanoscale devices. Proc. IEEE **96**, 1511–1550 (2008)
30. R. Venugopal, Z. Ren, S. Datta, M.S. Lundstrom, D. Jovanovic, Simulating quantum transport in nanoscale transistors: real versus mode-space approaches. J. Appl. Phys. **92**, 3730–3739 (2002)
31. M. Luisier, A. Schenk, W. Fichtner, Quantum transport in two- and three-dimensional nanoscale transistors: Coupled mode effects in the nonequilibrium Green's function formalism. J. Appl. Phys. **100**, 12 (2006)

32. H. Wang, G. Wang, S. Chang, Q. Huang, High-order element effects of the Green's function in quantum transport simulation of nanoscale devices. IEEE Trans. Electron Devices **56**, 3106–3114 (2009)
33. D. Mamaluy, M. Sabathil, P. Vogl, Efficient method for the calculation of ballistic quantum transport. J. Appl. Phys. **93**, 4628–4633 (2003)
34. A. Trellakis, A.T. Galick, A. Pacelli, U. Ravaioli, Iteration scheme for the solution of the two-dimensional Schrodinger-Poisson equations in quantum structures. J. Appl. Phys. **81**, 7880–7884 (1997)
35. G. Fiori, G. Iannaccone, G. Klimeck, A three-dimensional simulation study of the performance of carbon nanotube field-effect transistors with doped reservoirs and realistic geometry. IEEE Trans. Electron Devices **53**, 1782–1788 (2006)
36. H. Wang, G. Wang, S. Chang, Q. Huang, Accelerated solution of Poisson-Schrodinger equations in nanoscale devices by Anderson mixing scheme. IET Micro Nano Letters **4**, 122–127 (2009)
37. Z. Ren, S. Goasguen, A. Matsudaira, S.S. Ahmed, K. Cantley, Y. Liu, Y. Gao, X. Wang, M. Lundstrom, NanoMOS (2014). http://nanohub.org/resources/nanomos. doi:10.4231/D3GQ6R32D
38. S. Koswatta, J. Guo, D. Nikonov, MOSCNT: code for carbon nanotube transistor simulation. http://nanohub.org/resources/1989
39. G. Fiori, G. Iannaccone, NanoTCAD ViDES. http://vides.nanotcad.com/vides/
40. G. Fiori, G. Iannaccone, Three-dimensional simulation of one-dimensional transport in silicon nanowire transistors. IEEE Trans. Nanotechnol. **6**, 524–529 (2007)
41. G. Fiori, G. Iannaccone, NanoTCAD ViDES. https://nanohub.org/resources/vides. doi 10.4231/D3P26Q456
42. J. Guo, S. Datta, M. Lundstrom, A numerical study of scaling issues for Schottky-barrier carbon nanotube transistors. IEEE Trans. Electron Devices **51**, 172–177 (2004)
43. I. Hassaninia, M.H. Sheikhi, Z. Kordrostami, Simulation of carbon nanotube FETs with linear doping profile near the source and drain contacts. Solid-State Electron. **52**, 980–985 (2008)
44. H. Zhou, M. Zhang, Y. Hao, Performance Optimization of MOS-Like carbon nanotube FETs with realistic contacts. IEEE Trans. Electron Devices **57**, 3153–3162 (2010)
45. L. De Michielis, L. Lattanzio, K.E. Moselund, H. Riel, A.M. Ionescu, Tunneling and occupancy probabilities: how do they affect tunnel-FET behavior? IEEE Electron Device Lett. **34**, 726–728 (2013)
46. L. De Michielis, L. Lattanzio, A.M. Ionescu, Understanding the superlinear onset of tunnel-FET output characteristic. IEEE Electron Device Lett. **33**, 1523–1525 (2012)
47. M.A. Khayer, R.K. Lake, Drive currents and leakage currents in InSb and InAs nanowire and carbon nanotube band-to-band tunneling FETs. IEEE Electron Device Lett. **30**, 1257–1259 (2009)
48. A.S. Verhulst, W.G. Vandenberghe, K. Maex, S. De Gendt, M.M. Heyns, G. Groeseneken, Complementary silicon-based heterostructure tunnel-FETs with high tunnel rates. IEEE Electron Device Lett. **29**, 1398–1401 (2008)
49. C. Wei, C.J. Yao, G.F. Jiao, H. Darning, H.Y. Yu, L. Ming-Fu, Improvement in reliability of tunneling field-effect transistor with p-n-i-n structure. IEEE Trans. Electron Devices **58**, 2122–2126 (2011)
50. A.S. Verhulst, W.G. Vandenberghe, K. Maex, G. Groeseneken, Tunnel field-effect transistor without gate-drain overlap. Appl. Phys. Lett. **91**, 053102–053103 (2007)
51. D.B. Abdi, M.J. Kumar, Controlling Ambipolar current in tunneling FETs using overlapping gate-on-drain. IEEE J. Electron Devices Soc. 1–1 (2014)
52. C. Anghel, P. Chilagani, A. Amara, A. Vladimirescu, Tunnel field effect transistor with increased ON current, low-k spacer and high-k dielectric. Appl. Phys. Lett. **96** (2010)
53. M. Schlosser, K.K. Bhuwalka, M. Sauter, T. Zilbauer, T. Sulima, I. Eisele, Fringing-induced drain current improvement in the tunnel field-effect transistor with high-k gate dielectrics. IEEE Trans. Electron Devices **56**, 100–108 (2009)

54. C. Woo Young, L. Woojun, Hetero-gate-dielectric tunneling field-effect transistors. IEEE Trans. Electron Devices **57**, 2317–2319 (2010)
55. H. Wang, S. Chang, Y. Hu, H. He, J. He, Q. Huang, F. He, G. Wang, A novel barrier controlled tunnel FET. IEEE Electron Device Lett. **35**, 798–800 (2014)
56. S.J. Wind, J. Appenzeller, P. Avouris, Lateral scaling in carbon-nanotube field-effect transistors. Phys. Rev. Lett. **91**, 058301 (2003)
57. S. Saurabh, M.J. Kumar, Novel attributes of a dual material gate nanoscale tunnel field-effect transistor. IEEE Trans. Electron Devices **58**, 404–410 (2011)
58. H. Wang, S. Chang, J. He, Q. Huang, A Novel PNIN Barrier controlled tunnel FET. Adv. Mater. Res. **1096**, 497–502 (2015)

Index

A
Armchair CNT, 186
Atomic disorder scattering in graphene TFETs, 136
Atomistic simulations of tunneling FETs, 111

B
Backgate biasing design of TFETs, 96
Band-to-band tunneling (BTBT), 69
Band-to-band tunneling mechanism, 182
Barrier controlled tunnel FET (BC-TFET), 201
  alternative structure implementation, 204
  concept analysis of, 205
  with gate work function engineering, 201
Behavioral model, 62
Benzyl viologen (BV), 51
Born-oppenheimer approximation, 113

C
Carbon nanotube band structure
  graphene band structure, 183
Challenges of TFETs, 90
Characteristics of TFETs, 90
  backgate biasing design, 96
  complementary TFET logic design, 98
  delayed saturation, 91
  enhanced miller capacitance, 92
  imbalanced complementary TFETS, 93
  n-type TFET design, 97
  unidirectional conduction, 90
  variability, 94
Characteristics, tunnel FETs, 20
Characterization of TFETs, 53
  extracting subthreshold swing from diode, 55
  threshold voltage, 54
Charge model, 77
Coherent potential approximation (CPA), 112
Coherent potential approximation (CPA) theory, 131
Compact models of TFETs, 84
  charge model, 77
  current model, 74
  electrostatic potential model, 69
  gate leakage, 81
  heterojunction TFETs, 84
  homojunction TFETs, 80
  model applications, 80
  short-channel effects, 82
Complementary Metal-Oxide-Semiconductor (CMOS) technology
  power problem, 1
  reducing the power of, with steep slope devices, 1
Complementary TFET logic design, 98
Current–voltage characteristics, 142

D
Delayed saturation, 91
Dennard's scaling, 4
Density functional theory (DFT), 113
Device physics in TMDC TFETs, 140
Digital applications of TFETs
  challenges, 89
  designs, 89
Discretized Hamiltonian, 157
Disorder Scattering, modeling transport with, 139
Doping engineering, 197
  in drain side junction, 198
  pocket doping, 197
  for TFETs performance enhancement, 24
Dopingless junction, 38
Drain-side underlap (DUND) technique, 98

L. Zhang and M. Chan (eds.), *Tunneling Field Effect Transistor Technology*,
DOI 10.1007/978-3-319-31653-6

Dual material gate TFET (DMG-TFET), 203
Dual metal-gate work function (DWF) technique, 98
Dual oxide (DOX) technique, 98

E
Eight-band $\boldsymbol{k} \cdot \boldsymbol{p}$ Hamiltonian, 154
Electro-mechanical FETs, 11
Electron-hole bilayer TFETs (E-H bilayer TFETs), 38
Electronic structure theory, 113
  applications, 140
  basic set and exchange correlation functional, 116
  density functional theory, 113
Enhanced miller capacitance, 92

F
Fabrication of TFET
  device geometry, 41
  dopingless junction, 38
  III–V on silicon, 49
  III-V heterojunction, 46
  III-V homojunction, 44
  material system, 41
  Si/Ge/SiGe heterojunction, 43
  source/drain formation, 36
  tunneling junction, 35
  2D material system, 50
Feedback FETs, 16
Fermi integral, 67
Ferroelectric FETs, 14
Field-Effect-Transistor (FET) technology
  electro-mechanical FETs, 11
  impact ionization FETs, 9

G
Gate dielectric engineering, 197
Gate leakage, 81
Gate work function engineering, 201
Generalized gradient approximations (GGA), 116
Geometry engineering for TFETs performance enhancements, 23
Graphene band structure, 183
Graphene TFETs, 136
Green's function, 129

H
Heterojunction TFETs, 84, 174
Heterojunctions based TFETs, 26
Hohenberg-Kohn theorem, 114
Homojunction-based TFET, 20
Homojunction TFETs, 18, 80, 170, 173
Hybrid TFET-MOSFET SRAM cell design, 101

I
Imbalanced complementary TFETs, 93
Impact ionization FETs, 9

K
Kohn–Sham equation, 114
$\boldsymbol{k} \cdot \boldsymbol{p}$ method, 151, 154

L
Lagrange multiplier, 115
Landau–Khalatnikov equation, 15
Landauer–Büttiker Formalism, 126
Laplace equation, 118
Linear muffin tin orbitals (LMTO)
  atomic sphere approximation in, 117
  Hamiltonian matrix in, 121
  overlap matrix in, 121

M
Material engineering
  for TFETs performance enhancement, 25
Metal-organic vapor phase epitaxy (MOVPE), 49
Metal-oxide-semiconductor field-effect transistor (MOSFET), 3
  working principle, 7
Microwave annealing (MWA), 38
Modeling of TFETs, 61. *See also* Compact models of TFETs
Muffin tin orbital (MTO) technique, 117

N
N-type TFET design, 97
NanoTCAD ViDES, 192
Non-equilibrium Green's function (NEGF), 111, 112, 189
  device simulation procedure, 189
  Simulation Toolkit of NanoTCAD ViDES, 192
Non-equilibrium Green's function formalism, 128
Non-equilibrium vertex correction (NVC), 111
Non-equilibrium vertex correction (NVC) theory, 112

P
Performance enhancements techniques for TFETs, 23
  doping engineering, 24
  geometry engineering, 23
  material engineering, 25

Physics-based model, 62
Piezoelectric Transistor (PET), 13
Pocket doping engineering, 24
Poisson equation, 67
Polyethyleneimine (PEI), 51

Q
Quantum transport simulation of III-V TFETs, 151
  device structure, 153
Quantum transport theory, 126
  coherent potential approximation, 131
  CPA at non-equilibrium, 134
  Landauer-Büttiker formalism, 126
  non-equilibrium green's function formalism, 128
  random disorder scattering within LMTO method, 131

R
Random disorder scattering within LMTO method, 131
Recursive Green's function (RGF) algorithm, 190
Reduced-Order NEGF Method, 163
Reduced models, error and cost of, 168
Retarded Green's function, 129

S
Schödinger equation, 115
Schrödinger equation, 118
Second derivative (SD) method, 65
Short-channel effects, 82
Si/Ge/SiGe heterojunction, 43
Silicon-on-insulator (SOI) transistors, 17
Simple Gate-All-Around PIN TFET, 192
  advantages, 192
  problems, 192
  tunnel current modulation mechanisms, 195
Simulation Results, 170
6T SRAM cell, 101
Spurious Band Elimination, 165
SRAM peripheral circuits, 104
Steep slope devices, 1
  comparison between, 19
  feedback FETs, 16
  ferroelectric FETs, 14
  impact ionization FETs, 9
  power reduction with, 4
  reducing CMOS power with, 1
  tunnel FETs, 17
Subthreshold swing (SS)
  in CMOS technologies, 5

T
Table lookup model, 62
Technology for computer-aided design (TCAD), 189
III–V heterojunction, 46
III–V homojunction, 44
III–V on silicon, 49
III–V TFETs, 151
Threshold voltage, 54
Tight binding (TB), 160
Tight-binding (TB) method, 123
Tight-binding linear muffin tin orbital (TB-LMTO) method, 116, 117
Transition metal dichalcogenides (TMDs), 41, 50
2D material system, 50

U
Unidirectional conduction, 90

Z
Zigzag CNT, 186