PHYSICS AND SIMULATION OF QUASI-BALLISTIC TRANSPORT IN NANOSCALE TRANSISTORS

A Thesis

Submitted to the Faculty

of

Purdue University

by

Jung-Hoon Rhew

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2003

To my beloved Father and Mother for their sacrifice and support for my long education

ACKNOWLEDGMENTS

I am greatly indebted to my advisor, Prof. Mark Lundstrom, for a tremendous and organized endeavor he has made throughout his career to mature his students into inspired and responsible professionals; I believe he deserves the title "professor" more than anyone else. Also, I am grateful to his wife, Mrs. Mary Lundstrom, for her support and sacrifice, which are often not recognized but undoubtedly essential to making his endeavor possible. I am obliged to Prof. Supriyo Datta for his insightful advice and inspiring lectures, which were indispensable to my research. I'd like to thank Dr. R. K. Smith at Agere for valuable discussion and Prof. Thomas Downar for his service in my committee and his lectures on numerical analysis. Also, I am thankful to Prof. Daniel Elliott who willingly agreed to be my committee member in a difficult time. I would like to thank my former and current colleagues at Purdue, Dr. Kausar Banoo, Dr. Ramesh Venugopal, Dr. Zhibin Ren, Dr. Sebastien Goasguen, Anisur Rahman, Sayed Hasan, Jing Guo, and Jing Wang. Finally, I would like to express my gratitude to the following institutions for supporting my research and providing wonderful opportunities to me: Semiconductor Research Corporation (SRC), National Science Foundation (NSF), Agilent, and DURINT.

TABLE OF CONTENTS

Page

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

ABSTRACT

Rhew, Jung-Hoon. Ph.D., Purdue University, December, 2003. Physics and Simulation of Quasi-Ballistic Transport in Nanoscale Transistors. Major Professor: Mark S. Lundstrom.

The formidable progress in microelectronics in the last decade has pushed the channel length of MOSFETs into decanano scale and the speed of BJTs into hundreds of gigahertz. This progress imposes new challenges on device simulation as the essential physics of carrier transport departs that of conventional approaches because the effects of quasi-ballistic transport and quantum phenomena on device and circuit performance are becoming more important. Although a full quantum approach such as nonequilibrium Green's function method naturally accommodates the two issues and a semiclassical first principle such as Monte Carlo simulation resolves at least the issue of quasi-ballistic transport, their heavy computational prevents them from playing a major role in exploring a wide range of device design options in practice. Hence, it is of great interest to develop a new macroscopic approach for the simulation of nanoscale devices operating near the ballistic limit. This motivates the present study to explore the feasibility of such an attempt. Therefore, the purpose of this study is to understand essential physics of quasi-ballistic transport and its implications to nanoscale device simulation based on macroscopic transport models. The study is composed of three parts; one is to understand the essential physics of quasi-ballistic transport in a device context, another is to identify the limitations of commonly used transport models in assessing nanoscale devices, and the other is to explore new macroscopic transport models valid from the diffusive to the ballistic limit.

1. INTRODUCTION

1.1 Background

The formidable progress of semiconductor technology in the last decade has allowed transistors to operate in the regime where quasi-ballistic transport and quantum effects start playing a significant role in determining device and circuit performance. Scaling down of CMOS devices has already reached the decananometer node [4] and the speedup of bipolar transistors has been achieved over 300 GHz [5] and is now facing a 500 GHz barrier of maximum operation frequency. For the design of these ultra high performance devices, empirical exploration becomes more and more demanding in terms of time and cost to achieve acceptable device performance near the edge of the device technology (e.g., [6]). Therefore, predictive device simulation to guide the experiments is receiving a serious attention [7]. However, the simulation tools must be reliable in order to be predictive; to avoid unreasonable performance assessment in exploring a wide range of design options, it is required to understand the essential physics of carrier transport in nanoscale devices.

In principle, a full quantum approach such as nonequilibrium Green's function (NEGF) method [8] is the ultimate, naturally accommodating quasi-ballistic transport and quantum effects, and a first principle semiclassical approach such as Monte Carlo (MC) simulation [9] and a full Boltzmann solver [10] resolves at least the issue of quasi-ballistic transport. However, their computational burden is so tremendous that even some of the rigor inherent in them has to be sacrificed to speed up simulation. Even so, they are yet to be adequate to investigating a wide range of design options in practice. Thus, device engineers still mostly rely on semiclassical macroscopic transport models such as the drift-diffusion (DD) and the energy transport/hydrodynamic (ET/HD) models in the analysis and design of various transistors. But, some of

the assumptions in these semiclassical macroscopic transport models implemented in conventional technology computer aided design (TCAD) tools start to break down. These are 1) the assumption of collision-dominated transport, and 2) neglect of quantum effects and degenerate carrier statistics, especially in MOSFET simulation. For the past few years, device researchers have begun to notice that the DD model underestimates the on-current of MOSFETs [11] and that despite a success in modeling hot carrier effects, advanced transport models (e.g., ET/HD) may substantially overestimate the on-current [12]. Also, there have been studies on transport in high-speed BJTs with the base width below 100 nm questioning the validity of those macroscopic transport models [13] [14]. As for the quantum effects, the phenomenological models such as the density gradient method [15] and the effective potential approach [16] are implemented into semiclassical transport models. However, their limitations and applicability are not fully appreciated yet.

Since there exist no device simulation tools to confidently and efficiently examine and assess new design approaches for nanoscale transistors, extensive attempts have been made to build a new simulation tool. Those attempts can be divided into three categories: 1) one is to modify existing conventional macroscopic models [11] [17], 2) another to build a combined usage of existing techniques [18], and 3) the other to develop a new efficient transport model from the Boltzmann transport equation (BTE) [19]. There are many issues yet to be clarified as discussed in the review by Grasser, et al. [20]. Therefore, it seems worth critically examining the feasibility of such attempts based on a thorough understanding of the essential physics of carrier transport in nanoscale transistors. In this thesis, we focus on understanding the physics associated with quasi-ballistic transport in nanoscale transistors.

1.2 Quasi-Ballistic Transport in Nanoscale Transistors

In semiconductor devices, there are two types of carrier motion as shown in Fig. 1.1: 1) deterministic motion that follows the Newton's law (denoted as arrows) and 2) random motion due to scattering that follows Fermi's golden rule (the abrupt changes between arrows). In Fig. 1.2, L stands for the feature size of a device, e.g., the channel length of a MOSFET, and ℓ denotes the critical length that determines the device performance, e.g., the on-current of a MOSFET as discussed in [21] [22]. The critical length is the length between the thermionic injection point (the top of the source-to-channel barrier in a MOSFET or the beginning of the base in a BJT) and the beginning of a region that acts as a carrier sink (the beginning of the high-field channel region in a MOSFET or the beginning of the high-field collector in a BJT).

As we scale down a MOSFET or increase the speed of a BJT, ℓ decreases. As illustrated in Fig. 1.1, if the mean-free-path λ is much shorter than ℓ, the carrier motion can be approximated as drift-diffusion motion ignoring the effect of deterministic motion on the determination of important physical quantities (e.g., current) required for device design and circuit simulation. However, in high performance devices, λ becomes comparable to ℓ and the effect of deterministic motion comes into play (think of the circled region in Fig. 1.1 as a device).

Both types of carrier motion are present in first principles such as MC simulation or the BTE, but the effect of deterministic motion is missing in conventional macroscopic models due to the assumption of collision-dominated transport, which is valid when $\ell \gg \lambda$. In this drift-diffusion transport regime: 1) a device can be approximated as a bulk-like channel, and consequently the boundary conditions of macroscopic transport models are derived as if a periodic boundary condition were imposed in the corresponding microscopic description, 2) current is described as drift and diffusion motion, and 3) consequently only scattering (λ) determines the current. In the quasi-ballistic transport regime (i.e., $\ell \sim \lambda$), however, we need to consider the effect of deterministic motion. This means: 1) that a device consists of not only a channel but

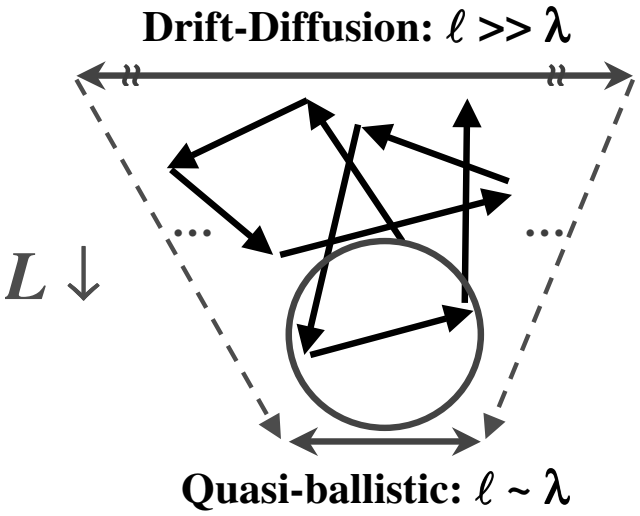

Fig. 1.1. Illustration of carrier motion in semiconductor devices. The arrows represent deterministic motion and the abrupt changes in between denote random motion due to scattering. If $\ell \gg \lambda$, carrier motion can be approximated as drift and diffusion while transport becomes quasi-ballistic if $\ell \sim \lambda$.

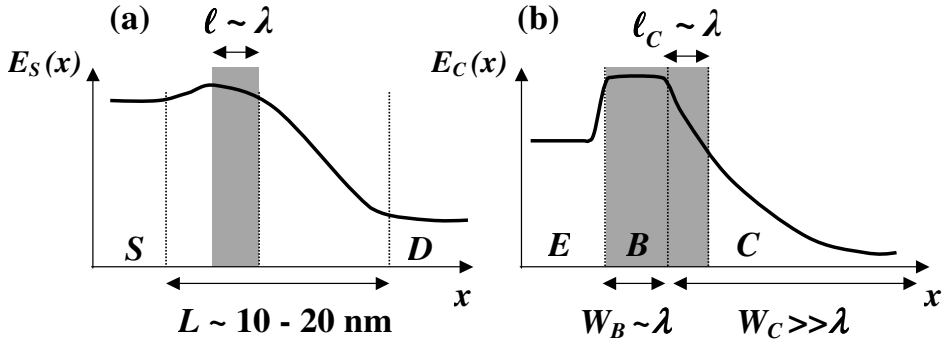

Fig. 1.2. (a) An energy band profile of a nanoscale MOSFET with channel length L is shown. Carriers are injected from the source, cross the critical region of length ℓ, and finally exit through the drain. (b) A conduction band profile of an *npn* BJT is shown. The critical regions are the thin base with its width $W_B \sim \lambda$ and the beginning of the collector region with the critical length $\ell_C \sim \lambda$. The collector length W_C is usually much longer than λ.

also contacts because carrier momentum imposed by the contacts can be maintained long enough to affect the device performance, 2) that current is transmission between the contacts through the channel [23], and consequently 3) that both scattering (λ) and deterministic motion (embedded in ℓ) determine the current. In this quasi-ballistic regime, we can no longer use periodic boundary conditions but should adopt so-called *inflow* boundary conditions even at a macroscopic level, with which we assign constraints only on carriers that come *into* the device from contacts [24]. Note that the inflow boundary conditions are inherent in the first principles such as the NEGF, MC simulation, and the BTE.

The mobility issue in short channel high electron mobility transistors (HEMTs) discussed in [25] confirms the above statement. Shur observed that the field effect mobility μ_{eff} extracted from measurements for a short channel FET is considerably smaller than that of a long channel mobility μ_0, and that this reduction is related to a finite electron acceleration time in the channel. Notice that $\mu_{eff} = \mu_0$ if we ignore the effect of the finite electron acceleration time. To account for the finite acceleration time, he introduced the ballistic mobility μ_B, which is defined as

$$\mu_B \equiv \frac{qL}{m^* \pi \upsilon_T}, \tag{1.1}$$

where q is electron charge, L is the channel length, m^* is effective mass, and υ_T ($= \sqrt{2k_BT/\pi m^*}$) is the thermionic emission velocity. Then, he calculated the short channel mobility μ_{eff} using Matthiessen's rule as

$$\mu_{eff}^{-1} = \mu_0^{-1} + \mu_B^{-1}, \tag{1.2}$$

and matched the experimental values. Note that the finite acceleration time L/υ_T in Eq. 1.1 originates from the effect of deterministic carrier motion as it describes the time required for a carrier to travel over a length L with the finite velocity υ_T given at the beginning of the channel ($\ell \sim L$ under a low-field condition [21]). Even in the 150 nm gate AlGaAs/GaAs HEMT he measured, the effect of deterministic motion affects the current since the long channel mobility is much higher than in Si. Hence,

we can expect the same thing to happen in Si MOSFETs if the channel length reduces further down to decanano scale and the long channel mobility μ_0 improves.

Now we further clarify this issue as we explain it in relation to the scattering theory for Si MOSFETs [21] as discussed by Wang, et al. [26]. The I–V equation Shur used to extract μ_{eff} is derived from the drift-diffusion equation, which is given as

$$I_{DS} = \mu_{eff} C_{ox} \frac{W}{L} (V_{GS} - V_T) V_{DS} \tag{1.3}$$

assuming

$$\mu_{eff} = \mu_0, \tag{1.4}$$

where C_{ox} is the effective oxide capacitance, W is the width and L is the channel length, and V_T is the threshold voltage. However, according to the scattering theory derived from McKelvey's flux model [27] [28] that embraces the three implications of quasi-ballistic transport mentioned above, the I–V equation is expressed as [26]

$$I_{DS} = \mu_0 C_{ox} \frac{W}{L (1 + \mu_0/\mu_B)} (V_{GS} - V_T) V_{DS}, \tag{1.5}$$

where we utilize that [21]

$$\mu_0 = \frac{q\lambda}{m^* \pi \upsilon_T}, \tag{1.6}$$

and that the transmission coefficient, which is derived from mesoscopic transport [8] and adopted in the scattering theory, can also be expressed using Eqs. 1.1 and 1.6 as

$$\mathbf{T} = \frac{\lambda}{L + \lambda} = \frac{\mu_0}{\mu_B + \mu_0}. \tag{1.7}$$

Thus, comparing Eq. 1.3 and Eq. 1.5, we can see that $\mu_{eff} \neq \mu_0$ but that μ_{eff} should be given by Eq. 1.2 in order for the conventional I–V model to explain the experimental I–V data. In other words, the conventional I–V model should be modified to the scattering theory model in Eq. 1.5 so that μ_{eff} should satisfy Eq. 1.2.

Figure 1.3 shows a generic relation between measured I_{DS} vs. the long channel mobility μ_0, and summarizes this issue. Note that the measured current cannot exceed the ballistic current I_{DSball}. In the drift-diffusion transport regime, i.e., $L \gg \lambda$,

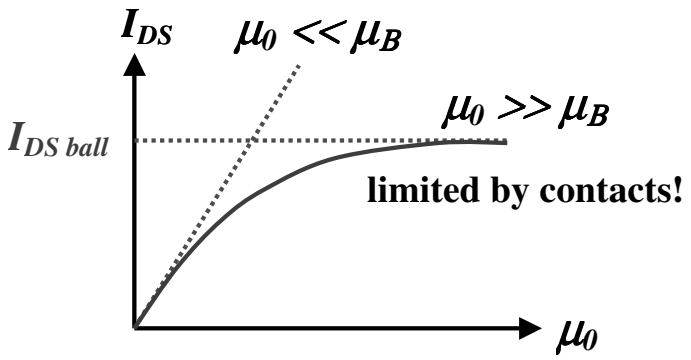

Fig. 1.3. A generic relation between measured I_{DS} and the long channel mobility μ_0. In the drift-diffusion regime, i.e., $\mu_0 \ll \mu_B$ (or equivalently $\lambda \ll L$), the current is proportional to μ_0. However, in the quasi-ballistic regime, the ballistic mobility μ_B becomes comparable to μ_0. In the near-ballistic regime, i.e., $\mu_0 \gg \mu_B$, the current is limited at the finite ballistic value I_{DSball}.

the current is proportional to the long channel mobility μ_0 (or λ) since μ_B is very large. However, as μ_0 increases or equivalently L decreases so that $L \sim \lambda$, transport becomes quasi-ballistic and the current does not exceed but approaches the ballistic limit I_{DSball}, and the mobility extracted for short channel MOSFETs is no longer μ_0 but $\mu_0 \cdot \mu_B/(\mu_0 + \mu_B) < \mu_0$; deterministic carrier motion comes into play through μ_B. Also, note that the long channel mobility μ_0 is still important in nanoscale FETs [29] as it can be related to the mean-free-path λ, but that its effect on current is limited by deterministic motion.

1.3 Overview of the Thesis

As illustrated in the previous section, the effect of deterministic carrier motion, which is no longer negligible in the quasi-ballistic regime, is missing in macroscopic description of carrier transport. Hence, the purpose of this thesis is to examine whether and how we can bring the missing piece back to the macroscopic level. In order to achieve the purpose, this thesis is directed toward understanding essential physics of quasi-ballistic transport and its implications to nanoscale transistor simulation based

on macroscopic transport models. Note that by nanoscale transistors we mean transistors near the ultimate performance limits, which, we assume, are more advanced than current-day technology. Therefore, our intention is to identify the fundamental limitations of conventional transport models that may become serious problems in the simulation of *the* nanoscale transistor rather than to criticize the use of them in current technology. The study is composed of three parts: 1) to understand the essential physics of quasi-ballistic transport in a device context, 2) to identify the limitations of commonly used transport models in assessing nanoscale devices, and 3) to explore new macroscopic transport models valid from the diffusive to the ballistic limit.

Prior to understanding quasi-ballistic transport, the essential physics of carrier transport in the ballistic limit need to be understood to see the effect of pure deterministic motion, and its implications to macroscopic transport models should be identified. Hence, the study of ballistic transport in nanoscale MOSFETs is performed in Chapter 2; it establishes new insights on ballistic transport, which provides a foundation to the rest of the thesis.

In Chapter 3, conventional transport models such as DD (drift-diffusion), ET (energy transport), and Bude's DD [11] are benchmarked against the ballistic results in a simple but reasonable nanotransistor structure. The results demonstrate that existing models do not provide reliable predictions of near-ballistic transport due to the assumptions of collision-dominated transport.

In Chapter 4, based on the understanding developed in the previous chapters, we formulate a drift-diffusion type macroscopic model that can accurately describe the ballistic transport in a nanoscale MOSFET. This is done by developing the directed-moment approach, which is an extension of McKelvey's flux method that satisfies the three implications of quasi-ballistic transport discussed above.

In Chapter 5, we implement a simple scattering model into the ballistic drift-diffusion equation developed in Chapter 4 and formulate a quasi-ballistic drift-diffusion model for a nanoscale MOSFET where velocity saturation is not likely to occur.

In Chapter 6, we turn our attention to quasi-ballistic transport in high-speed silicon based BJTs, identify fundamental limitations of the DD, the one-flux, and the ET models, and resolve the issue on whether macroscopic transport models are applicable to the simulation of high-speed BJTs. We explain in relation to the theory of impact ionization why no macroscopic model based on the moments of the BTE can overcome these limitations, but that despite the fundamental limitations, conventional macroscopic models can be used to assess transit time and cutoff frequency of current-day practical SiGe BJTs.

Finally, in Chapter 7, we summarize what we have learned through this research and propose future work.

2. A NUMERICAL STUDY OF BALLISTIC TRANSPORT IN A NANOSCALE MOSFET

In this chapter, we examine the physics of ballistic transport in a nanoscale MOSFET as reflected in the shape of the distribution function. We calculate the electron distribution function in the ballistic limit by solving the 1D steady-state Boltzmann transport equation self-consistently with the 2D Poisson equation in an n-channel ultra-thin-body nanoscale double-gate SOI MOSFET. In equilibrium, symmetry of the distribution function is achieved through balanced carrier injections from the source and drain contacts. Under bias, the distribution function displays distinctive features of ballistic transport—a discontinuous asymmetric shape and the development of a discontinuous ballistic peak. We discuss the implications of ballistic transport to modeling of nanoscale MOSFETs using moment-based macroscopic transport models.

2.1 Introduction

As semiconductor devices are scaled down to their limits, carrier transport may approach the ballistic limit [30]. In GaAs n^+-n-n^+ structures, hot electron, quasi-ballistic transport has been widely examined theoretically [31] [32]. Hesto, et al. first evaluated the distribution function in the quasi-ballistic transport regime in a submicron GaAs n^+-n-n^+ structure using Monte Carlo simulation in 3D momentum space [31]. Subsequently, several papers have examined quasi-ballistic transport. Among those, the work of Baranger, et al., who solved the Boltzmann transport equation (BTE) in one spatial and one momentum-space dimensions with the relaxation time approximation (RTA), is especially thorough [32]. In this work, the electron distribution function shows the general features of ballistic transport—asymmetry at

the top of the source-to-channel barrier, the development of a ballistic peak along the channel.

Previous studies using 1D devices are, however, not adequate for understanding the physics of ballistic transport in a nanoscale MOSFET because they fail to describe the following factors that interplay through the 2D self-consistent potential. First, the 1D n^+-n-n^+ structures do not capture the physical mechanism associated with current saturation in a MOSFET [33]. Second, they do not consider degenerate carrier statistics and vertical quantum confinement, which are indispensable features of ultra small MOSFETs. Finally, the presence of scattering in those works clouds the features directly originating from pure ballistic transport, which sets the ultimate performance limits on the device [33]. Thus, it is worth investigating the electron distribution function in the ballistic limit in an ultra small MOSFET, not only to assess the ultimate performance limits of the device but also to clarify the physics of ballistic transport as reflected in the shape of the distribution function and to identify its implication to nanoscale device modeling.

As a vehicle for this study, we take a model 10 nm double-gate SOI MOSFET depicted in Fig. 2.1 [34]. To describe ballistic transport along the channel, we solve the 1D ballistic BTE self-consistently with the 2D Poisson equation. A semiclassical approach is adequate because it has been recently demonstrated that the MOSFETs operate classically down to channel length of about 10 nm [34] [35]. In Appendix A, we provide a theoretical explanation for the validity of the semiclassical approach down to this length scale. Also, we focus on the steady-state on-current characteristics, which are relevant to the high-speed operation of digital circuits [36]. In a related paper [22], we examined the essential physics of nanoscale MOSFETs. In this chapter, we discuss the semiclassical solution approach, explain why a semiclassical treatment retains validity at 10 nm length scales, and explain the ballistic distributions throughout the device—not just at the top of the source-to-channel barrier.

We begin the chapter by explaining how to obtain the analytic solution of the 1D, steady-state ballistic BTE for given boundary conditions in Section 2.2. In Section 2.3,

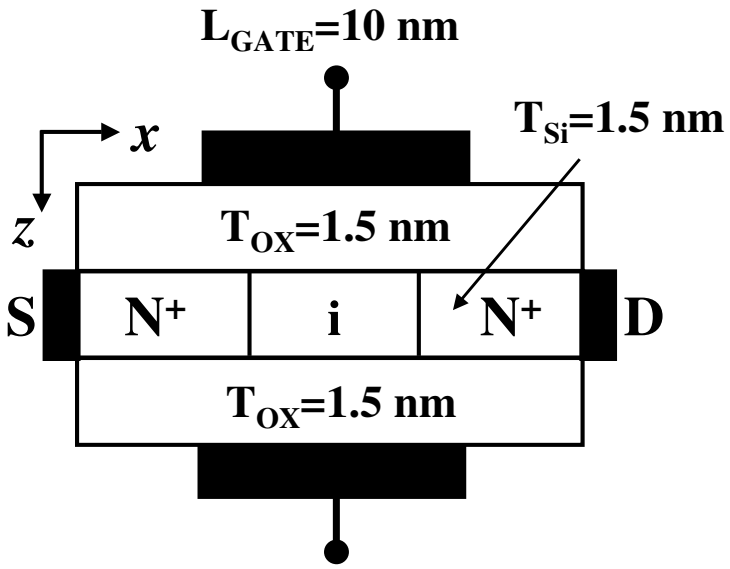

Fig. 2.1. Schematic of the model device

we present the model MOSFET, our simulation method, and the results. Then, we discuss modeling of nanoscale MOSFETs using moment-based macroscopic transport models in Section 2.4. Finally, we summarize in Section 2.5.

2.2 Theory of the Ballistic BTE

2.2.1 1D ballistic BTE and the boundary conditions

Electrons in the channel of our model, ultra-thin-body MOSFET are subject to strong quantum confinement in z-direction and move ballistically in the x-y plane. Due to the strong vertical quantum confinement, all conduction electrons are accommodated in the lowest subband and respond in the x-y plane with an effective mass, m_t^*, which is the transverse effective mass of the ellipsoidal conduction valleys of Si [37]. We also assume a parabolic bandstructure because the use of a non-parabolic

band makes little difference for the bias condition to be used (no greater than 0.6 V). In 1D transport along the x-direction, the 1D BTE in steady-state is

$$\upsilon_x \frac{\partial f_B\left(x,\vec{p}\right)}{\partial x} - q\mathscr{E}_x \frac{\partial f_B\left(x,\vec{p}\right)}{\partial p_x} = 0, \tag{2.1}$$

where the subscript B denotes the ballistic distribution. The 2D momentum vector is $\vec{p} = p_x\hat{x} + p_y\hat{y}$, and the electric field is

$$\mathscr{E}_x = \frac{1}{q}\frac{dE_S\left(x\right)}{dx}, \tag{2.2}$$

where $E_S(x)$ stands for the lowest subband energy profile, and the total energy, $E\left(x,\vec{p}\right)$, is

$$E\left(x,\vec{p}\right) = \frac{p_x^2}{2m_t^*} + \frac{p_y^2}{2m_t^*} + E_S\left(x\right). \tag{2.3}$$

To complete the definition of the problem, boundary conditions for Eq. 2.1 must be specified. We focus on the ideal intrinsic device performance ignoring the source/drain series resistance, thus assume that scattering in the source and drain regions are cast into the perfect absorbing contacts that maintain thermal equilibrium [8]. At the source contact, the distribution for the incoming flux (the positive half) is an equilibrium Fermi-Dirac distribution,

$$f_B\left(x=0,\vec{p}\right) = f_o\left(E\left(0,\vec{p}\right);\mu_1\right) = \frac{1}{1+\exp\left(\frac{E(0,\vec{p})-\mu_1}{k_BT}\right)}, \tag{2.4}$$

if $p_x > 0$, and similarly at the drain contact for the negative half,

$$f_B\left(x=L,\vec{p}\right) = f_o\left(E\left(L,\vec{p}\right);\mu_2\right) = \frac{1}{1+\exp\left(\frac{E(L,\vec{p})-\mu_2}{k_BT}\right)}, \tag{2.5}$$

if $p_x < 0$, where $f_o\left(E;\mu\right)$ is the Fermi-Dirac distribution, and μ_1 and μ_2 are the Fermi levels in the source and drain regions respectively. The total electron energies, $E\left(0,\vec{p}\right)$ and $E\left(L,\vec{p}\right)$, are obtained from Eq. 2.3.

2.2.2 The solution of the 1D ballistic BTE

Using direct substitution, one can readily show that $g\left(E(x,\vec{p})-\mu\right)$, any function of total energy with the ambiguity of a Fermi-level μ, satisfies the ballistic BTE

in Eq. 2.1. (The same result occurs in collision-dominated equilibrium [38].) To completely specify the solution, we need to identify the functional form $g(\bullet)$ and the Fermi-level μ at the boundaries and inside the device. At the boundaries, Eqs. 2.4 and 2.5 determine $g(\bullet)$ and μ.

To specify $g(\bullet)$ and μ inside the device, we need to understand how the states are occupied as illustrated in Fig. 2.2, a plot of the lowest subband energy profile and a parabolic dispersion relation between the longitudinal kinetic energy E_x and longitudinal momentum p_x in the channel of our model transistor under bias. At $E_{\max}$, the subband energy at the top of the barrier, the semiclassical transmission changes abruptly from 0 to 1. Consequently, ballistic electrons injected from the source with their longitudinal kinetic energy greater than the barrier height transmit freely from source to drain while those below the barrier are reflected and return to the source. In the E_x-p_x parabola, $+p_x$ states above the energy barrier (solid line) will be occupied by carriers injected from the source. Likewise, $-p_x$ states (dashed line) are filled with drain-injected carriers. Below the barrier, the states can be seen only from the drain, thus they (dashed line) are occupied by the drain-injected carriers. In a similar way, the occupation of states by ballistic electrons in other regions can be understood.

According to the above description, the functional form $g(\bullet)$ is given as an equilibrium Fermi function, f_o,

$$f_B(x,\vec{p}) = g(E(x,\vec{p})) = f_o(E(x,\vec{p}) - \mu), \tag{2.6}$$

where μ is either the source Fermi-level μ_1 or the drain Fermi-level μ_2. Consequently, the solution of Eq. 2.1 is given for $|p_x| > p_{x\max}$ (above the barrier) as

$$f_B(x,\vec{p}) = \begin{cases} f_o(E(x,\vec{p});\mu_1) = \frac{1}{1+\exp\left(\frac{E(x,\vec{p})-\mu_1}{k_BT}\right)} & p_x > p_{x\max} \\ f_o(E(x,\vec{p});\mu_2) = \frac{1}{1+\exp\left(\frac{E(x,\vec{p})-\mu_2}{k_BT}\right)} & p_x < -p_{x\max} \end{cases}, \tag{2.7}$$

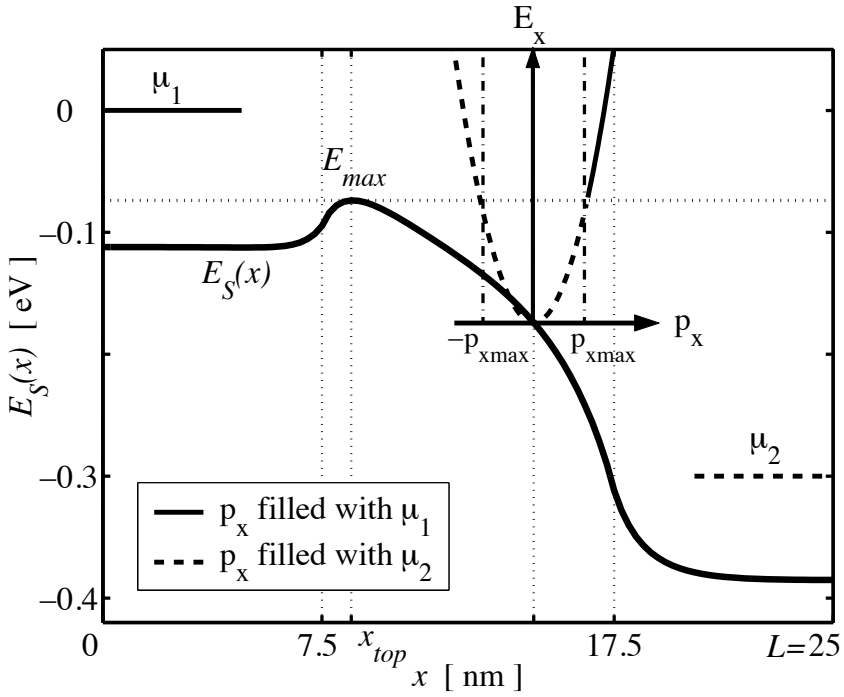

Fig. 2.2. The lowest subband profile of the model transistor in Fig. 2.1 under the bias at $V_{GS1} = V_{GS2} = 0.6$ V and $V_{DS} = -(\mu_2 - \mu_1)/q = 0.3$ V. The parabolic $E_x(p_x)$ relation illustrates how the states are occupied.

and for $|p_x| < p_{x\max}$ (below the barrier)

$$f_B(x,\vec{p}) = \begin{cases} f_o\left(E(x,\vec{p});\mu_1\right) = \frac{1}{1+\exp\left(\frac{E(x,\vec{p})-\mu_1}{k_BT}\right)} & 0 < x < x_{top} \\ f_o\left(E(x,\vec{p});\mu_2\right) = \frac{1}{1+\exp\left(\frac{E(x,\vec{p})-\mu_2}{k_BT}\right)} & x_{top} < x < L \end{cases} \tag{2.8}$$

where

$$p_{x\max}(x) = \sqrt{2m_t^*\left[E_{\max} - E_S(x)\right]}. \tag{2.9}$$

In this sense, a special kind of equilibrium can be said to exist, even though the device itself may be very far from equilibrium [8]. Each state is occupied according to a Fermi function. The appropriate Fermi-level to use is that of the contact from which the state was occupied. Because the appropriate Fermi levels change abruptly with energy, highly nonequilibrium overall distribution functions can result.

2.3 Results

2.3.1 The model device and the description of simulation

Our model device shown in Fig. 2.1 is one of the devices examined in [34]. The channel length is 10 nm; the oxide thickness is 1.5 nm for both top and bottom gates, and the silicon body thickness 1.5 nm. The source and drain regions are doped at $1.0 \cdot 10^{20}$ cm^{-3}, and the channel is undoped. The gate structure is symmetric with a mid-gap gate material (top/bottom). The channel junctions are abrupt and no gate-to-source/drain overlap is assumed. Transport occurs along x-direction with transverse y-direction along the device width (assumed to be wide). Quantum confinement is along z-direction and is strong enough to justify the one subband model.

We adopt the semiclassical quasi-2D simulation approach in [34] where the 2D transport solution is obtained by solving two 1D problems separately—1D semiclassical transport along x-direction and 1D quantum confinement in z-direction. In this method, sheet carrier density, $n_s(x)$, in x-y plane is calculated from the approach of Section 2.2. Next, $n_s(x)$ is weighted by the amplitude squared of the wave function

$\phi(z)$ obtained from a 1D Schrödinger solver to yield volume carrier density $n(x,z)$, that is

$$n(x,z) = n_s(x)\ |\phi(z)|^2 . \tag{2.10}$$

Then, we insert $n(x,z)$ into the 2D Poisson equation to obtain the 2D potential profile. From the modified $E_S(x)$, a new $n_s(x)$ is obtained. The procedure is repeated until it converges. This approach is valid when the vertical quantum confinement is so strong as to be insensitive to the longitudinal change of the potential [34].

2.3.2 Equilibrium distributions

The computed distribution functions in equilibrium are shown in Fig. 2.3. The 1D cross-sections and 2D plots of $f_B(\upsilon_x, \upsilon_y)$ in equilibrium ($V_{DS} = 0.0$ V) are calculated at $x = 5$ nm (source region) and $x = x_{top} = 12.5$ nm (top of the barrier) under $V_{GS1} = V_{GS2} = 0.6$ V. In the 1D plots, the solid (dashed) lines denote the states populated by the source (drain) contact with Fermi-level μ_1 (μ_2). In equilibrium, the symmetry of the distribution is achieved through the balanced injection from each contact instead of detailed balance due to collision as in the diffusive regime. The volume under the surface at $x = 5$ nm (which represents the carrier density) is greater than that at $x = x_{top}$ because more charge is available in the source region than in the channel. The shape of the distribution at $x = 5$ nm is Fermi-Dirac whereas that at $x = x_{top} = 12.5$ nm is closer to Maxwellian because it is less degenerate in the channel than in the source/drain regions.

2.3.3 Distributions at the top of the source barrier and current control mechanism

Figures 2.4 and 2.5 show 1D cross-sections and 2D plots of $f_B(\upsilon_x, \upsilon_y)$ at the top of the source barrier under $V_{GS1} = V_{GS2} = 0.6$ V with different drain biases. At the top of the barrier, the source injection fills the whole positive υ_x states of the distribution

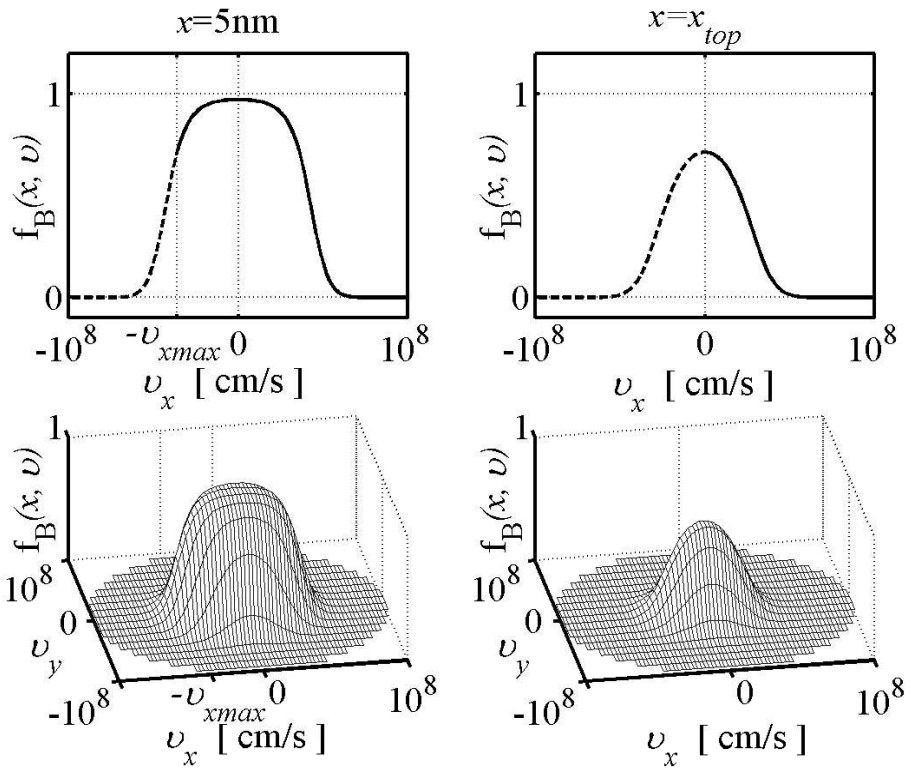

Fig. 2.3. The 1D cross-sections and 2D plots of $f_B\,(\upsilon_x, \upsilon_y)$ in equilibrium ($V_{DS} = 0.0$ V) are shown at $x = 5$ nm (source region) and $x = x_{top} = 12.5$ nm (top of the barrier) under $V_{GS1} = V_{GS2} = 0.6$ V. The solid (dashed) lines denote the states populated by the source (drain) contact Fermi-level μ_1 (μ_2)

.

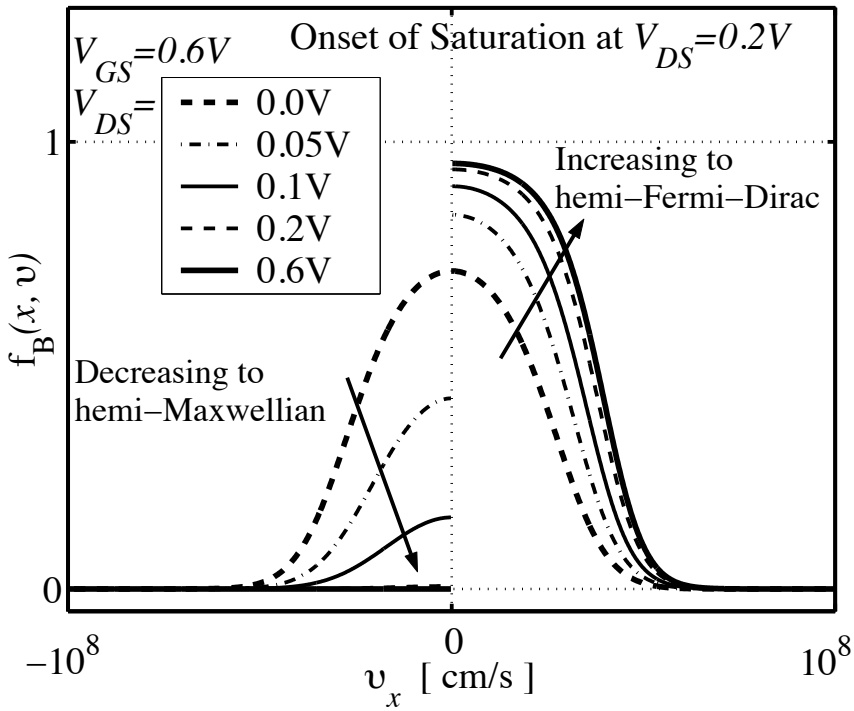

Fig. 2.4. 1D cross-sections of $f_B\,(\upsilon_x, \upsilon_y = 0)$ at the top of the barrier at $V_{GS1} = V_{GS2}$ = 0.6 V. Discontinuity occurs at $\upsilon_x = 0$. The distribution at $V_{DS} = 0.2$ V is almost identical to that at $V_{DS} = 0.6$ V because saturation begins at $V_{DS} = 0.2$ V.

while the negative υ_x states are populated only by the drain injection. At $V_{DS} = 0.0$ V (in equilibrium), the distribution is symmetric due to balanced injections.

As V_{DS} increases, the drain injection ($-$ half) becomes suppressed leaving the source injection (+ half) unbalanced. Hence, the current starts to flow (the current vs. V_{DS} plot in Fig. 2.6). Note, however, that the source injection increases to maintain the charge balance imposed by the gate (see the density vs. V_{DS} plot in Fig. 2.6). Thus, the area under the positive half of the distribution at $V_{DS} = 0.2$ V is approximately twice that of the positive half of the equilibrium distribution. For this to happen, the subband energy needs to be pulled down as V_{DS} increases to fill more $+\upsilon_x$ states until the onset of saturation as shown in $E_S(x)$ plot of Fig. 2.6. (See [34] for more details.) As a consequence, the reduced barrier seen from the source increases the degeneracy associated with the positive half, but the increased barrier seen from the drain makes

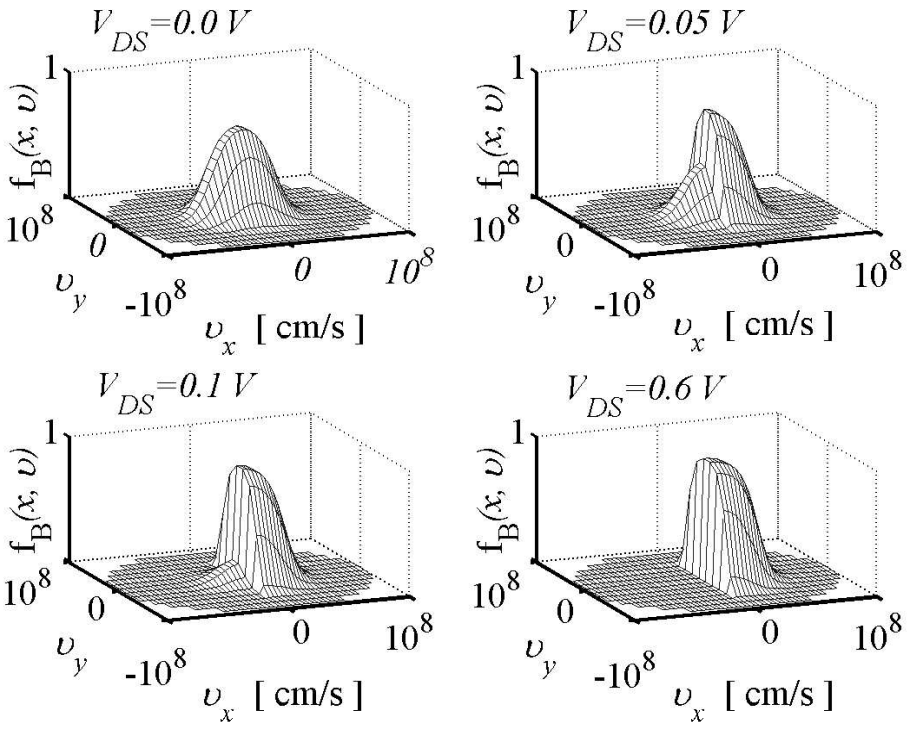

Fig. 2.5. Corresponding 2D plots of $f_B\,(\upsilon_x, \upsilon_y)$. Discontinuity occurs along υ_y axis but the distribution in υ_y direction remains equilibrium in shape.

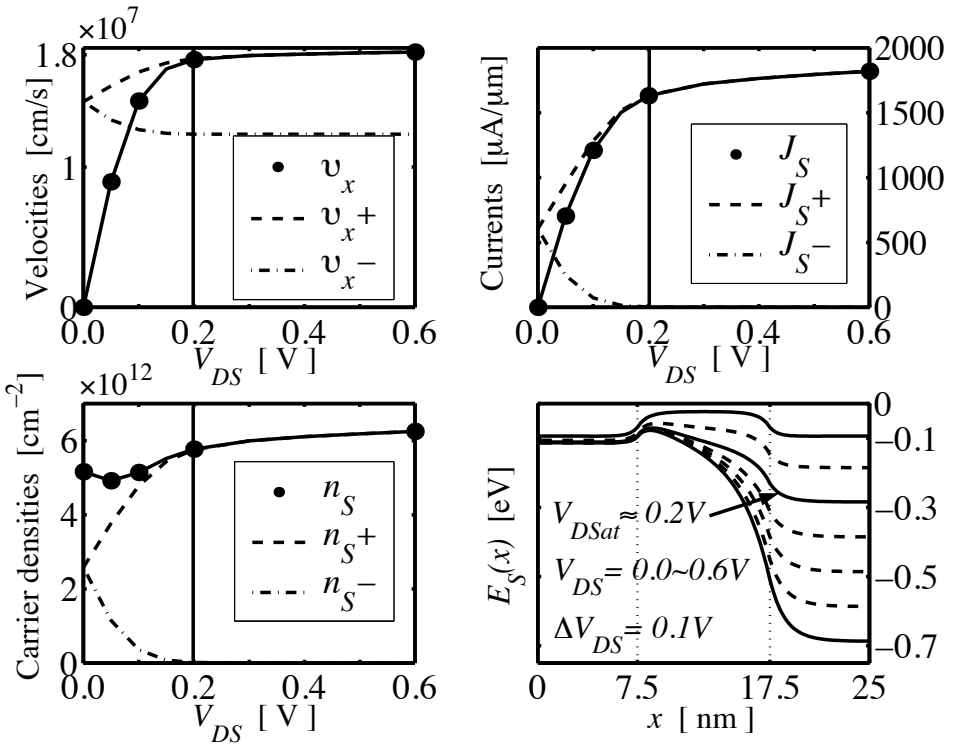

Fig. 2.6. Average velocity vs. V_{DS}, carrier density vs. V_{DS}, and current vs. V_{DS} plots calculated at the top of the barrier, and the lowest subband energy profile $E_S(x)$ under different V_{DS}. The dots correspond to the five bias points shown in Fig. 2.4.

it less degenerate for the negative half. Thus, the shape of the increasing positive half rapidly approaches hemi-Fermi-Dirac and its average velocity υ_x^+ reaches up to $1.8 \cdot 10^7$ cm/s while the diminishing negative half becomes more hemi-Maxwellian and its average velocity υ_x^- approaches the thermal velocity of a hemi-Maxwellian ($1.2 \cdot 10^7$ cm/s) as shown in the average velocity vs. V_{DS} plot in Fig. 2.6. At the total suppression of the drain injection (at $V_{DS} = 0.2$ V), the source injection saturates except the slight increase due to 2D electrostatics, which is explained in the next subsection. The device operates in linear region if $V_{DS} < 0.2$ V and it is in saturation region if $V_{DS} > 0.2$ V. Note that even in the ballistic limit, the current does saturate due to the gate control, which is absent in 1D device structures. The plots in Fig. 2.6 summarize this current control mechanism. In conclusion, the physics reflected in

the shape of the distribution function at the top of the barrier confirms the current control mechanism discussed in [21] [22].

2.3.4 Distribution along the channel under bias - development of ballistic peaks

In this subsection, we will discuss the shape of the distribution function at different locations along the channel under high gate and drain bias. Figure 2.7 shows 1D cross-sections of $f_B(v_x, v_y)$ at different locations along the channel under the bias $V_{GS1} = V_{GS2} = V_{DS} = 0.6$ V. The locations of interest along x-direction are selected non-uniformly to represent substantial changes. Note that the discontinuity occurs at $v_x = \pm v_{x\max}$ depending on whether it is at the source or the drain side where

$$v_{x\max}(x) = \frac{p_{x\max}(x)}{m_t^*}, \tag{2.11}$$

in which $p_{x\max}(x)$ is obtained from Eq. 2.9. The thick (thin) solid lines represent the states populated by the source (drain) contact. The plots in Fig. 2.8 exhibit the corresponding 2D distribution function $f_B(v_x, v_y)$ at four different locations: 1) the source region, 2) the top of the barrier, 3) the channel, and 4) the drain region.

In the source region, electrons injected from the source with $0 < v_x < v_{x\max}$ are reflected back by the barrier filling their negative counterpart while electrons with $v_x > 0$ have transmitted across the barrier, but there are almost no carriers injected from the drain contacts. Hence, the distribution is effectively zero for $v_x < -v_{x\max}$. Since the high-energy portion in the negative half is eliminated, evaporative cooling occurs in this region. As electrons go against the barrier from the source, the shape of the distribution function approaches hemi-Fermi-Dirac as in thermionic emission.

Down the potential barrier toward the drain, the source injection develops as a ballistic peak as shown in Fig. 2.7 while the injection from the drain is still negligible. As carriers are accelerated along the channel, the location of the peak (Eq. 2.11) moves toward higher velocity, i.e., velocity overshoot occurs. Note that the velocity overshoot makes the ballistic peak narrower as the carrier density decreases to

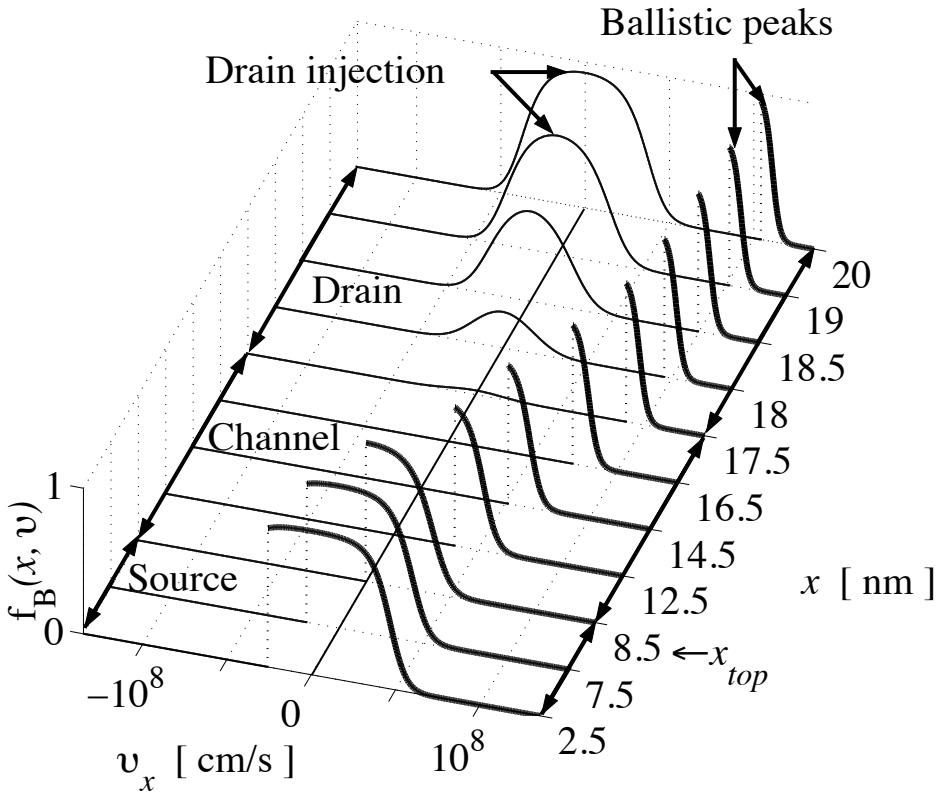

Fig. 2.7. 1D cross-sections of $f_B\,(v_x, v_y = 0)$ at different locations along the channel under the bias $V_{GS1} = V_{GS2} = V_{DS} = 0.6$ V. The thick (thin) lines represent the states populated by the source (drain) contact.

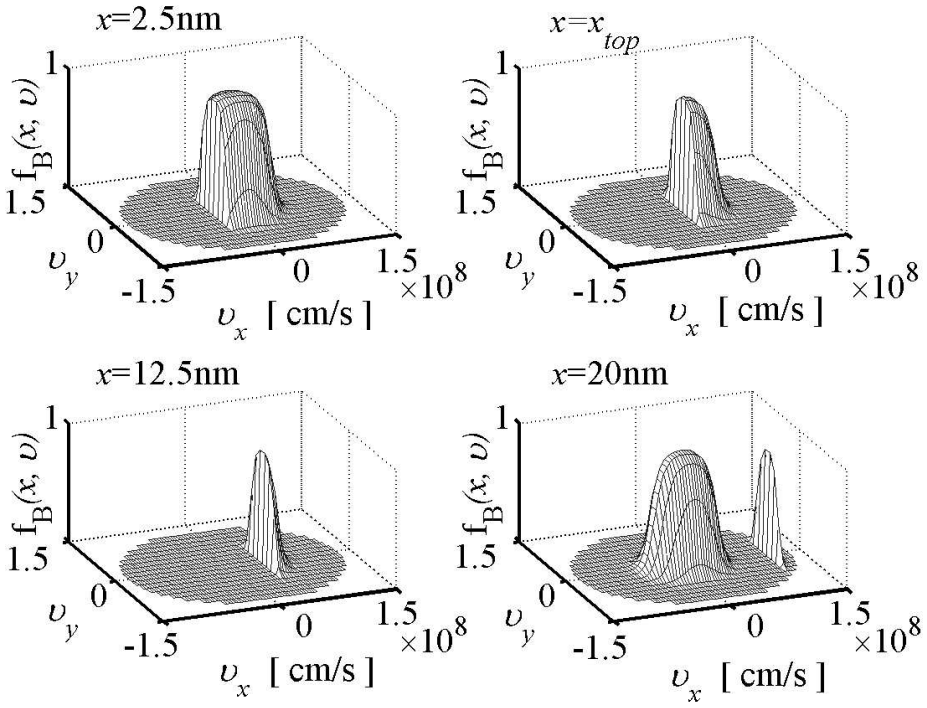

Fig. 2.8. Corresponding 2D plots of $f_B(\upsilon_x, \upsilon_y)$

preserve the current but the height of the peak remains the same because the total electron energy is conserved in the ballistic limit. Consequently, the velocity overshoot reduces the channel charge density helping the drain electric field penetrate more into the source barrier, which increases the current in the saturation region. However, the extreme velocity overshoot in the ballistic limit affects the current indirectly through self-consistent electrostatics. Although the on-current increase due to velocity overshoot is not negligible in bulk MOSFETs [11], it is found to be minor in our model double gate MOSFET due to strong gate control. Thus, it is confirmed that the on-current of a well-tempered MOSFET is mostly determined by the control mechanism at the top of the barrier.

Near the drain, the thermal equilibrium injection from the drain contact suddenly becomes important. Electrons injected from the drain are reflected by the barrier, and build up a symmetric, thermal distribution. While the thermal portion remains in equilibrium, the ballistic peak from the source causes heating in the drain region.

2.4 Discussion

As device dimensions continue to shrink, operation near the ballistic limit becomes a possibility and the question of whether commonly used macroscopic approaches lose validity becomes important. Nekovee, et al. [39] examined this issue and found that lower order models fail near the ballistic limit because of the ballistic peak that cannot be fit with low order polynomials. They also showed that a critical voltage occurs, above which there exists no unique solution. This was shown to occur when the average velocity within the device reaches the thermal velocity.

The results of our simulation of a ballistic silicon-MOSFET support the observations of Nekovee, et al. made for a GaAs n^+-n-n^+ diode. Specifically, Fig. 2.7 shows the strong ballistic peak that develops, and it seems clear that such distributions will be impossible to fit with low order polynomials. Also, our results show that the distribution at the top of the source barrier develops a highly asymmetric shape

at relatively low drain biases ($\sim$ 0.05 V, as shown in Fig. 2.4). As the distribution approaches a hemi-Fermi Dirac shape, the average velocity approaches the thermal velocity. Above this rather small bias, Nekovee, et al. showed that no unique solution of the macroscopic equation exists.

Although it seems clear that macroscopic transport equations of the conventional kind will fail in the ballistic limit, they may continue to be useful in nanoscale devices. Recent work shows that nanoscale MOSFETs, for example, operate well below the ballistic limit (e.g., [40]). It is important, however, to recognize this ballistic limitation and to be able to identify it in simulations. A simple way is to see if average velocities at a critical point (e.g., the top of the barrier) exceed the thermal velocity. It would, of course, be preferable to have a simple, macroscopic transport model that behaves gracefully in the ballistic limit. Our work supports that of Nekovee, et al., which suggested "that it is not possible to treat predominantly ballistic transport phenomena with any low-order systems of moment equations." [39] We believe, however, that it is worth exploring the application of flux equations, which can treat different carrier populations with different distributions [27] [28].

2.5 Summary

In this chapter, we examined ballistic transport in a model nanoscale MOSFET with particular emphasis on the shape of the distribution function at different positions under bias. In equilibrium, the symmetry of the distribution function is achieved through balanced carrier injections instead of detailed balance as in diffusive transport. Under a drain bias, we verified that the inversion layer density at the top of the source barrier is approximately constant, as controlled by the gate voltage, and that the distribution function approaches a hemi-Fermi Dirac distribution when the drain bias exceeds about $2k_BT/q$. These simulation results confirm previous assumptions [22]. We also examined the ballistic distributions throughout the transistor and confirmed cooling and heating in the source and drain regions, the development of a

ballistic peak, and velocity overshoot and consequent current increase. Finally, the results reported here also support the conclusions of Nekovee, et al. [39] that conventional, macroscopic moment-based transport models cannot describe transport in the quasi-ballistic regime.

3. BENCHMARKING MACROSCOPIC TRANSPORT MODELS FOR NANOTRANSISTOR TCAD

We benchmark macroscopic transport models against the ballistic limit in a realistic MOSFET structure to assess their validity in the near-ballistic regime and identify the challenges to a new macroscopic model valid from the diffusive to the ballistic limit. The macroscopic transport models tested in this study are drift-diffusion, Bude's drift-diffusion, and an energy transport model. We perform the benchmarking study with quasi-2D simulation on a model 10 nm ultra-thin body double-gate MOSFET. We confirm from our results that none of those macroscopic models successfully provides reliable predictions for nanotransistors operating in the near-ballistic regime and we explain the reason for the failure.

3.1 Introduction

Recently, aggressive scaling down of MOSFETs has made device engineers concerned about ballistic transport in the nanoscale regime where the validity of conventional macroscopic transport models is questionable [30]. For the past few years, device researchers have begun to note that drift-diffusion underestimates the on-current of nanoscale MOSFETs [11] and that despite the success in modeling hot carrier effects, advanced transport models (e.g., energy transport) may substantially overestimate the on-current in the near-ballistic regime [12]. There have been previous studies on ballistic transport in n-i-n structures [12] [32], but the essential physics of charge control and current limits in transistors [22] cannot be fully appreciated with those structures. Also, there have been attempts to extend the use of the conventional macroscopic models into the near-ballistic regime by adjusting parameters in those models [11] [41]. Hence, we benchmark macroscopic transport

models against the ballistic results obtained in [1] in a realistic, 10 nm ultra-thin body double-gate MOSFET (DGFET) depicted in Fig. 2.1 to assess their validity in the near-ballistic regime and to identify challenges to a macroscopic model valid from diffusive to ballistic regime.

The macroscopic transport models to be benchmarked are drift-diffusion (DD) in [11], Bude's drift-diffusion [11], and an energy transport (ET) model [42]. The assumptions in those models appear in transport parameters and hierarchy closures. Among them, the most important is the low-field mobility (μ_0) because it is involved in the lowest order polynomial approximation of the Boltzmann transport equation (BTE), and it is also proposed that μ_0 continues to play a critical role in the on-current performance even in the near-ballistic regime [21] [29]. Moreover, there are attempts to enhance μ_0 considerably to achieve suggested on-current performance targets [43]. Therefore, we benchmark the conventional models against the ballistic limit in our model device by changing μ_0 from 0 (diffusive limit) to 500 $\mathrm{cm}^2/\mathrm{V}\cdot\mathrm{s}$ (near-ballistic regime). Note that although we benchmark 1D transport models, our quasi-2D simulation [1] captures quantum confinement in the vertical direction and 2D electrostatics.

This chapter is organized as follows. In Section 3.2, we present the conventional models benchmarked in our study and show the benchmarking results in Section 3.3. Then, we discuss the results in Section 3.4 and finally summarize in Section 3.5.

3.2 Conventional Macroscopic Transport Models

Starting with the momentum balance equation of the BTE, we obtain [44]

$$\vec{J}_n = qn\mu_n\vec{\mathscr{E}} + 2\mu_n\nabla\cdot\overleftrightarrow{W}. \tag{3.1}$$

In the DD model, we ignore the drift component (asymmetry) in the tensor $\overleftrightarrow{W}$, and assume that the thermal energy in $\overleftrightarrow{W}$ that corresponds to carrier temperature T_C is distributed isotropically (symmetry), and further that $\nabla T_C = 0$ (local transport) to

end the hierarchy [44]. Then, Eq. 3.1 becomes the well-known drift-diffusion equation (1D),

$$J_{nx} = qn\mu_n \mathscr{E}_x + qD_n \frac{dn}{dx}, \tag{3.2}$$

where x is the direction of carrier transport, and $D_n = \mu_n k_B T_C/q$. Therefore, the DD model is valid only when the net carrier transport does not disturb the nearly symmetric carrier distribution and $\nabla T_C = 0$. Since there is no restriction to low-field, near-equilibrium conditions as far as ∇T_C is small [44], the DD model employs the Caughey-Thomas field-dependent mobility model [45] to describe high-field transport as well,

$$\mu_n = \mu_0 \left[1 + \left(\mu_0 \mathscr{E}_x / \upsilon_{sat}\right)^{\beta}\right]^{-1/\beta}, \tag{3.3}$$

where μ_0 is the low-field mobility, υ_{sat} is the saturation velocity, and β is a constant [45]. Measured values for electrons in Si are $\upsilon_{sat} \approx 1.0 \cdot 10^7$ cm/s and $\beta = 2$ [46]. Also, $T_C = T_L$, where T_L is lattice temperature, is usually assumed although it is not correct. Thus, Eqs. 3.2 and 3.3 with those measured parameters and the assumption that $T_C = T_L$ describe our DD model in this study.

Bude's DD model used in our study is the same as the DD model except that $\upsilon_{sat} = 2.2 \cdot 10^7$ cm/s and β =1 to allow carrier velocity to overshoot [11]. Note, however, that we test Bude's DD model in a 10 nm double-gate MOSFET instead of the longer channel bulk MOSFETs in [11], and that inheriting all the assumptions employed in the DD model, the model only imitates the non-local effects by parameterizing the normal DD model.

To eliminate the local transport assumption, the ET model includes higher moment equations to describe the behavior of carrier energy based on the carrier temperature approach [44]. In 1D, we get the current equation as [42]

$$J_{nx} = q\mu_n\left(T_C\right) n\mathscr{E}_x + q\left[\mu_n\left(T_C\right)\frac{d}{dx}\left(\frac{k_B T_C}{q} n\right) + n\frac{k_B T_C}{q}\frac{\partial \mu_n\left(T_C\right)}{\partial T_n}\frac{dT_C}{dx}\right] \tag{3.4}$$

with

$$\mu_n\left(T_C\right) = \mu_0\left(1 + \left[\frac{F_{\dim}}{2}\frac{k_B}{q}\frac{\mu_0}{\upsilon_{sat}^2 \tau_E}\left(T_C - T_0\right)\right]^2\right)^{-1/2}, \tag{3.5}$$

where μ_0 is again the low-field mobility, τ_E is an energy relaxation time, υ_{sat} is saturation velocity ($1.0{\cdot}10^7$ cm/s), and $F_{\rm dim}$ is a degree of freedom. The carrier temperature T_C is no longer constant but an unknown that involves the energy balance and the energy flow equations, which are

$$\frac{dS}{dx} = \frac{1}{q} J_{nx} \mathscr{E}_x - \frac{F_{\rm dim}}{2} n \frac{k_B}{q} \frac{T_C - T_0}{\tau_E}, \tag{3.6}$$

and

$$S = -\frac{5}{2} \frac{k_B T_C}{q} \left[\frac{J_{nx}}{q} + C_Q \mu_n n \frac{d}{dx} \left(\frac{k_B T_C}{q} \right) \right], \tag{3.7}$$

where $C_Q = 1.0$. Then, the ET model in our study is described by Eqs. 3.4 $\sim$ 3.7 with $F_{dim} = 2$ for the 2D momentum space due to the vertical quantum confinement. Among the parameters in the model, τ_E is the most influential to the on-current next to μ_0. Hence, we change μ_0 from 0 to 500 cm^2/V·s for $\tau_E = 0.1$ ps and 0.3 ps.

3.3 Benchmarking the Conventional Models against the Ballistic Results

Figure 3.1 shows the comparison of the I–V characteristics of the macroscopic models and the internal velocity profiles against the ballistic limit at $\mu_0 = 100$ and 300 cm^2/V·s. The ballistic model in this study is described in detail in [1]. In our 10 nm DGFET, the ET results show abnormal behavior at $\mu_0 = 300$ cm^2/V·s; the currents exceed the ballistic limit regardless of τ_E and the velocity profiles are unreasonable, especially near the source barrier (at the top of the barrier, the velocity should not exceed the thermal injection velocity, which is $1.8 \cdot 10^7$ cm/s this case). The DD model and Bude's DD model behave as described in [11], but unlike in the longer channel bulk MOSFETs in [11], the velocity overshoot of Bude's model is less pronounced compared to the injection velocity at the top of the barrier.

For further investigation, we produce $\rm I_{ON}$ vs. μ_0 and υ_{inj} vs. μ_0 plots as shown in Fig. 3.2. The on-current of the DD model saturates at about 50% of the ballistic limit because its injection velocity approaches $1.0 \cdot 10^7$ cm/s, which is about 50% of the ballistic limit ($1.8 \cdot 10^7$ cm/s). Hence, the DD model underestimates $\rm I_{ON}$ as

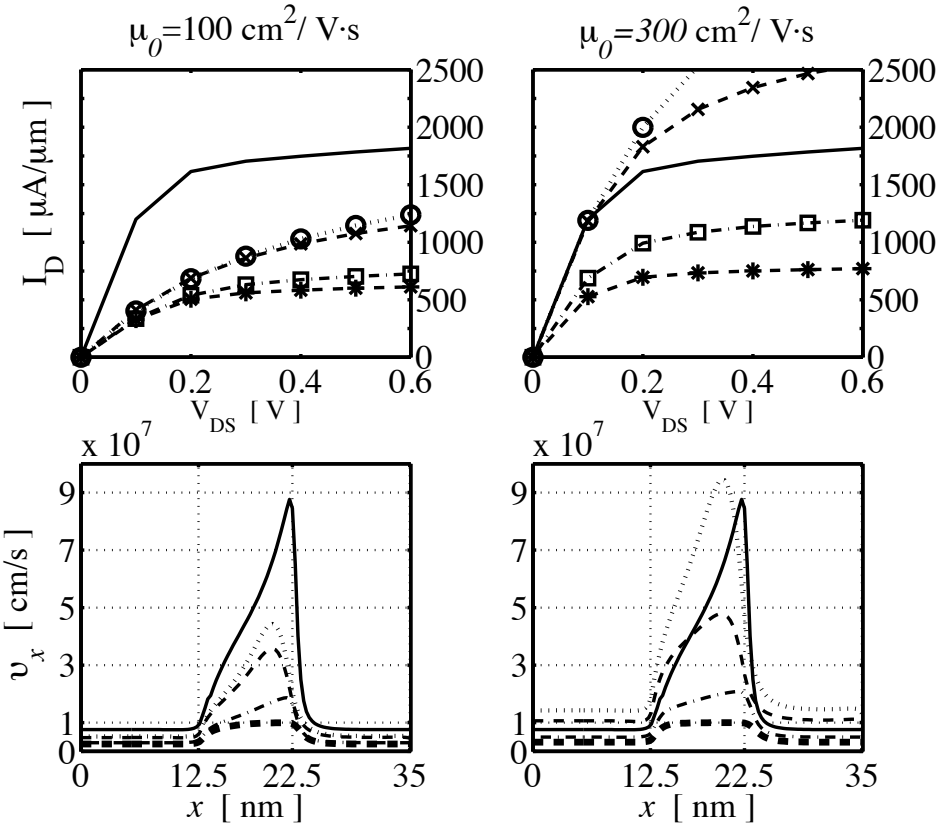

Fig. 3.1. Top: I–V curves of the ballistic BTE (solid lines), ET (o with $\tau_E = 0.3$ ps, x with 0.1 ps), Bude's DD (square), DD (*). Bottom: Velocity profiles of the ballistic BTE (solid), ET (dotted with $\tau_E = 0.3$ ps, dashed with 0.1 ps), Bude's DD (dash-dotted), DD (thick dotted).

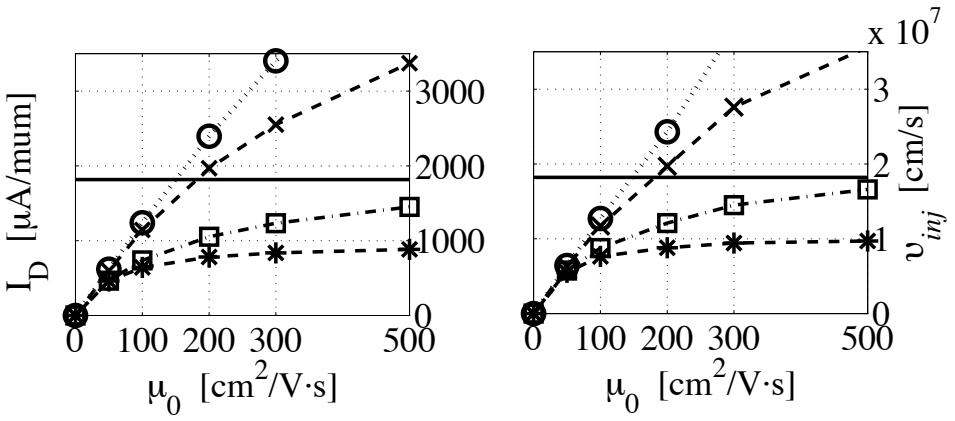

Fig. 3.2. I_{ON} vs. μ_0 and υ_{inj} vs. μ_0 plots of ET (o with $\tau_E = 0.3$ ps, x with 0.1 ps), Bude's DD (square), DD (*). The horizontal lines represent the ballistic limit (solid).

transport becomes more ballistic. Bude's DD model seems to approach the ballistic limit gracefully because υ_{inj} approaches $2.2 \cdot 10^7$ cm/s, which is close to $1.8 \cdot 10^7$ cm/s. However, its maximum velocity is limited by $2.2 \cdot 10^7$ cm/s although υ_{inj} is approaching it with increasing μ_0, and eventually the Bude's DD model will display no velocity overshoot in the ballistic limit. This is not physical because its maximum velocity should approach the maximum velocity of the ballistic limit. The ET model substantially overestimates I_{ON} because its injection velocity goes beyond the ballistic limit even when $\mu_0 < 200$ cm^2/V·s. Even with more energy relaxation ($\tau_E = 0.1$ ps), it exceeds the physical limit. As mentioned in the previous section, we can see that the sensitivity of I_{ON} to μ_0 is dominant over that to τ_E.

3.4 Discussion

The carrier distribution in the near-ballistic regime shows two distinctive features in its shape: the asymmetry at the top of the barrier associated with the current limiting process, and the development of ballistic peaks that accounts for velocity overshoot and the behavior of high energy carriers [1]. Note that the asymmetry at the top of the barrier (i.e., the current limiting process) cannot be captured by any

lower order polynomial approximation of the whole carrier distribution employed in conventional models [39] and that the development of ballistic peaks (i.e., velocity overshoot and high energy carrier effects) is neither a diffusion of high temperature carriers nor a drift but a convective motion [1]. In addition, in nanoscale MOSFETs, to capture the injection limit correctly, degenerate statistics of quantum confinement carriers should be considered as well [1].

The distinctive features of nanoscale carrier transport are not taken into account in conventional macroscopic models (not only the models tested in this study but also other models such as energy balance or hydrodynamic models). All conventional macroscopic models fail to describe the asymmetric distribution function at the top of the barrier [1] [39], and thus are unable to capture the thermal injection limit accurately (i.e., on-current). Also, they do not capture the development of ballistic peaks, thus are unable to describe velocity overshoot and the behavior of high energy carriers correctly. In other words, since the hierarchy closures of the macroscopic moment equations are associated with the behavior of carrier velocities (e.g., average velocity, and average energy) [44], without describing the development of ballistic peaks correctly, the hierarchy cannot be closed correctly, thus conventional models may not work in the near ballistic regime.

The DD model underestimates the ballistic on-current due to its incorrect limit on the carrier velocity and shows no velocity overshoot due to its local transport assumption. Despite its graceful near-ballistic on-current behavior, Bude's DD model substantially underestimates velocity overshoot in the near ballistic regime due to the same local transport assumption employed in the DD model. The ET model considerably overestimates the on-current and fails to capture the behavior of carrier energy correctly (regardless of the different choice of τ_E) unless μ_0 is used as a fitting parameter. It is because in addition to its incapability of imposing thermal injection limit, the ET model describes velocity overshoot with carrier temperature (broadening of carrier distribution) instead of the development of ballistic peaks. The incorrect behavior of high energy carriers in the model causes the injection velocity

to exceed the ballistic injection limit. With the same argument, we can also criticize the hydrodynamic model because it describes velocity overshoot with drift energy without limiting the injection velocity. However, if the source/drain series resistance dominates over the channel resistance, the above discrepancy will not appear as a serious problem in I–V because it reduces the difference in current through negative feedback.

3.5 Summary

We have benchmarked conventional macroscopic transport models against the ballistic limit in a nanoscale MOSFET and demonstrated that none of them is capable of providing reliable predictions for nanotransistors operating near the ballistic limit. Also, we have explained the reason for the failure and identified the challenges to a macroscopic model valid from diffusive to ballistic limit as:

- To impose thermal injection limit by capturing the asymmetry of carrier distribution at the top of the barrier
- To properly describe the velocity overshoot and the behavior of carrier energy associated with the development of ballistic peaks
- To include degenerate carrier statistics and quantum confinement if it affects the injection limit.

4. A DRIFT-DIFFUSION EQUATION FOR BALLISTIC TRANSPORT IN NANOSCALE FETS

We develop a drift-diffusion equation that describes ballistic transport in nanoscale FETs. We develop the directed-moment approach, in which we treat injection from different contacts separately, and describe each injection with a set of extended McKelvey's one-flux equations that include new hierarchy closure approximations appropriate for high-field ballistic transport and degenerate carrier statistics. Then, we re-express the extended one-flux equations in a drift-diffusion form with a properly defined Einstein relationship. The results obtained for nanoscale FETs show an excellent agreement with the solution of the ballistic Boltzmann transport equation with no fitting parameters. These results show that a macroscopic transport model based on the moments of the Boltzmann transport equation can describe ballistic transport.

4.1 Introduction

As transistors are scaled down to their ultimate limit, carrier transport may approach the ballistic limit. This imposes a great challenge on predictive assessments of device performance, especially the on-current of transistors, because: 1) commonly used macroscopic transport models assuming collision-dominated transport are expected to lose their validity near the ballistic limit [39] [47], and 2) computational burdens of reliable first principle simulators (Monte Carlo simulators [9] or full Boltzmann solvers [48]) preclude them from routine use for extensive design studies. A macroscopic model derived from the moments of the Boltzmann transport equation (BTE) capable of describing carrier transport from the diffusive to the ballistic limit would be of great interest. Because conventional moment-based macroscopic models (e.g., drift-diffusion, or energy transport models) fail in the ballistic limit [39] [47],

we turn our attention to McKelvey's one-flux method [27] [28], whose usefulness in qualitatively describing quasi-ballistic transport was demonstrated in the scattering theory [21]. But, the one-flux method is unable to describe carrier acceleration in a high-field region (e.g., the channel or the collector of a transistor) [49]. In this chapter, we develop the directed-moment approach by extending McKelvey's one-flux method and consequently a drift-diffusion equation to describe pure ballistic transport in nanoscale FETs.

We solve the ballistic drift-diffusion equation for the model MOSFET shown in Fig. 2.1 and compare the results to the ballistic BTE solution obtained in [1]. The model device is a 10 nm channel-length, double-gate MOSFET with an ultra thin body thickness of 1.5 nm. The gate oxides are 1.5 nm and both top and bottom gates have a mid-gap workfunction. The strong quantum confinement across the thin body provides two simplifying assumptions: 1) the one subband model for which we assume all carriers are accommodated in the lowest subband, and 2) the quasi-two-dimensional (2D) simulation in which we solve one-dimensional (1D) transport along the body and the 1D Schrödinger equation across the body self-consistently with the 2D Poisson equation [1].

We also test the ballistic drift-diffusion equation for the model coaxial gate carbon nanotube (CNT) FET shown in Fig. 4.1, which was examined in [2]. The geometry and the bandstructure of the CNTFET, which are different from those of the above planar MOSFET, are considered in the simulation. The channel length is 10 nm and the diameter of the tube is 1 nm. The insulator is 1 nm thick with the relative dielectric constant $\kappa = 4$ and the gate work function is 4.23 eV. The source/drain extension regions are doped at 10^7 cm^{-1}. Thanks to the strong quantum confinement around the circumference of the tube, we also apply the quasi-2D approach where we solve 1D transport along the tube and the 1D Schrödinger equation around the tube, and obtain a volume charge density to put into a 3D Poisson equation with radial symmetry.

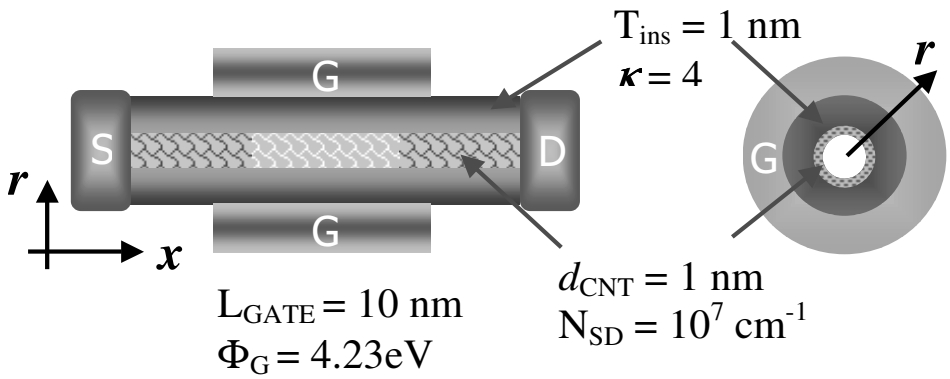

Fig. 4.1. Schematic of the model CNTFET.

The chapter is organized as follows. In Section 4.2, we examine the assumptions employed in the one-flux method. In Section 4.3, we present the directed-moment approach and the set of extended one-flux equations that can describe ballistic transport. In Section 4.4, we convert the extended one-flux equations into a drift-diffusion equation and solve in the ballistic limit for our model devices, and compare the results against the solutions of ballistic BTE in Section 4.5. Finally, we conclude in Section 4.6.

4.2 McKelvey's One-Flux Method

Figure 4.2 illustrates the idea of McKelvey's one-flux method [27] [28]. The two flux densities, $J^+(x)$ and $J^-(x+dx)$ (defined positively), incident on a semiconductor slab with thickness dx transmit or reflect with the backscattering probabilities per length, ξ and ξ', respectively, contributing to the outward fluxes $J^-(x)$ and $J^+(x+dx)$. This is described by the one-flux equations [27] [28] , which are given as

$$\frac{dJ^+}{dx} = -\xi J^+ + \xi' J^- \tag{4.1}$$

$$\frac{dJ^-}{dx} = -\xi J^+ + \xi' J^-, \tag{4.2}$$

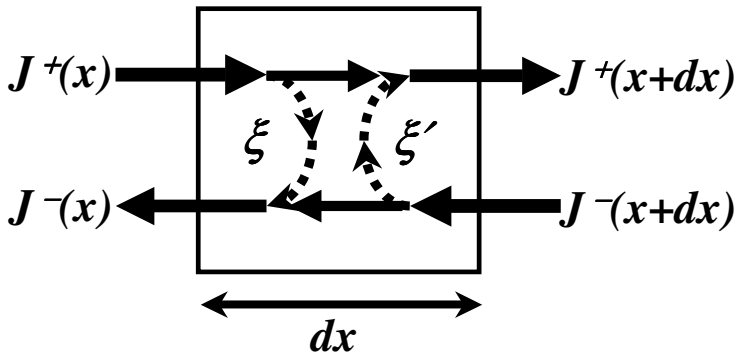

Fig. 4.2. Illustration of McKelvey's one-flux method. The two fluxes, $J^+(x)$ and $J^-(x+dx)$ are incident on a semiconductor slab with thickness dx, and transmit or backscatter inside with the backscattering probabilities per length, ξ and ξ', respectively.

where the backscattering probabilities per length are

$$\xi = \begin{cases} \xi_0 + \frac{q\mathscr{E}_x}{k_B T_L} & \mathscr{E}_x > 0 \\ \xi_0 & \mathscr{E}_x < 0 \end{cases} \tag{4.3}$$

$$\xi' = \begin{cases} \xi_0 & \mathscr{E}_x > 0 \\ \xi_0 - \frac{q\mathscr{E}_x}{k_B T_L} & \mathscr{E}_x < 0 \end{cases} . \tag{4.4}$$

The low-field backscattering probability per length ξ_0 is associated with the low-field mobility μ_0 through Shockley's relation [28], and T_L is the lattice temperature. See Appendix B.1 for a brief derivation of Eqs. 4.1 $\sim$ 4.4. The backscattering coefficients ξ and ξ' consist of two terms: 1) backscattering by actual scattering (ξ_0) and 2) backscattering by reflection due to an opposing electric field ($\mathscr{E}_x$) as shown in Fig. 4.3. In Eqs. 4.3 and 4.4, the effect of actual scattering is assumed to be symmetric under a low-field condition. By specifying the incoming fluxes at the boundaries, we can solve Eqs. 4.1 $\sim$ 4.4 for $J^\pm$. However, to solve them self-consistently with the Poisson equation, we need to know the carrier densities $n^\pm$ associated with $J^\pm$, which can be obtained by specifying the average velocities of $\pm$ streams, $\langle v_x \rangle^\pm$, because

$$J^\pm = n^\pm \langle v_x \rangle^\pm . \tag{4.5}$$

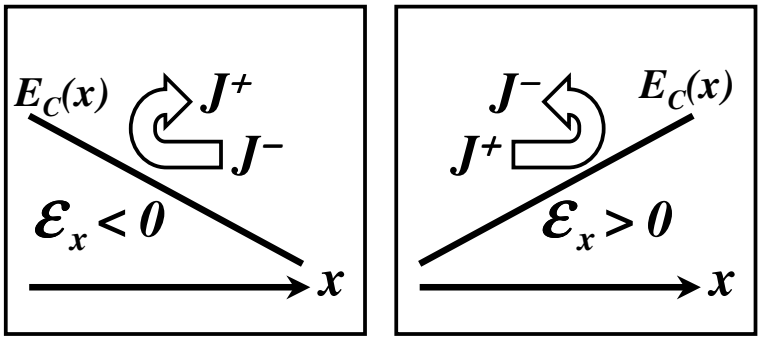

Fig. 4.3. Illustration of the effects of the opposing electric field.

As in conventional moment equations, the average velocities are obtained in two ways: 1) by solving the next order moment equations or 2) by employing hierarchy closure approximations [44]. The one-flux method closes the hierarchy by assuming $\langle v_x \rangle^{\pm}$ to be the average velocity of hemi-Maxwellian, i.e.,

$$\langle v_x \rangle^{\pm} \cong v_T = \sqrt{\frac{2k_B T_L}{\pi m^*}}. \tag{4.6}$$

Thanks to its flux description, the one-flux method successfully describes carrier transport from the diffusive to the ballistic limit when there is negligible acceleration, e.g., transport across a thin base [50]. However, the one-flux equations are derived from the zeroth moments of the BTE [49], in which the number of carriers in streams do not change due to acceleration. Thus, the field terms describe backscattering only and are associated with hemi-Maxwellian distributions in nondegenerate conditions (see Appendix B.1). The effect of acceleration such as velocity overshoot should be cast into hierarchy closure approximations on $\langle v_x \rangle^{\pm}$, which is missing in Eq. 4.6. In conclusion, the one-flux method is valid under nondegenerate conditions with negligible acceleration.

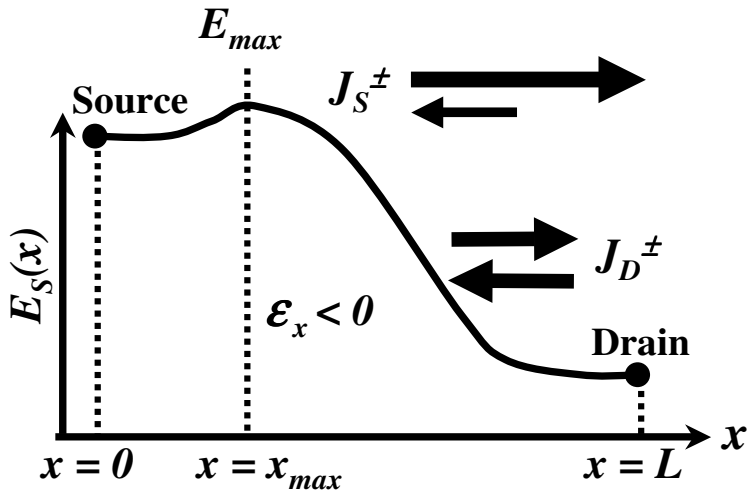

Fig. 4.4. Types of transport in a nanoscale MOSFET under bias. $E_{\max}$ is the maximum of the subband energy $E_S(x)$. In the channel region after the source barrier, J_S^+ is accelerated by the electric field whereas J_D^- is decelerated and then backscattered by the barrier.

4.3 Directed Moment Approach and Extended One-Flux Method for Ballistic MOSFETs

4.3.1 Separating injections, new closure approximations, and degenerate statistics

Figure 4.4 shows types of transport in a nanoscale MOSFET. In the channel region after the source barrier, the + stream from the source (J_S^+) is accelerated by the electric field whereas the injection from the drain contact (J_D^-) is decelerated and backscattered. Different populations experience different types of transport in the same region. Therefore, it is natural to treat the source-injected fluxes separately from the drain-injected fluxes and then describe each injection with the one-flux equations as illustrated in Fig. 4.5. In other words, we solve one set of flux equations for the source-injected carriers and another for the drain-injected carriers. The total carrier density and the net current are the sum of the quantities obtained for each injection. The idea of separating injections is based on the linearity of the ballistic

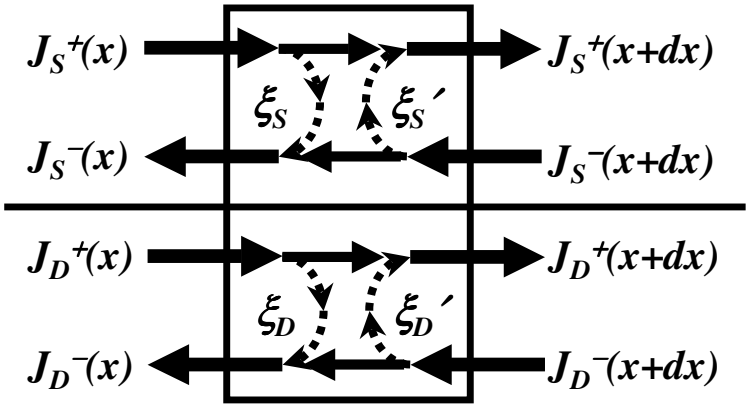

Fig. 4.5. Illustration of separation of carrier injections.

BTE; we solve the BTE for the source injection and the drain injection separately and then obtain the final distribution as a superposition of the two. This approach enables us to apply different macroscopic approximations (scattering parameters and hierarchy closure assumptions) to the sets of flux equations that describe different populations. It is obvious that we can use the original one-flux equations for streams under deceleration (e.g., drain-injected fluxes under bias), but new approximations are required to describe streams under acceleration. Another modification required for a nanoscale MOSFET is to include degenerate carrier statistics because they may affect the on-current [1]. In the ballistic limit, we can implement degenerate statistics into each injection separately.

In conclusion, we will extend McKelvey's one-flux method by: 1) treating carriers injected from the source and drain separately, 2) introducing new hierarchy closure approximations for the streams under acceleration, and 3) including degenerate carrier statistics. We name this approach as the directed moment approach since it is based on the partial moments of the BTE associated with the directions of carrier flow.

4.3.2 Ballistic one-flux equations with degeneracy for a nanoscale MOSFET

In the ballistic limit ($\xi_0 = 0$) with degenerate carrier statistics, the backscattering coefficients become (see Appendix B.2 for a derivation.)

$$\xi = \begin{cases} +\frac{q\mathscr{E}_x}{k_B T_L} F^{+}_{\deg}(\eta_+) & \mathscr{E}_x > 0 \\ 0 & \mathscr{E}_x < 0 \end{cases} \tag{4.7}$$

$$\xi' = \begin{cases} 0 & \mathscr{E}_x > 0 \\ -\frac{q\mathscr{E}_x}{k_B T_L} F^{-}_{\deg}(\eta_-) & \mathscr{E}_x < 0 \end{cases} . \tag{4.8}$$

For 2D electrons in the lowest subband of our model device, the degeneracy factors are

$$F^{\pm}_{\deg}(\eta_\pm) = \frac{\mathscr{F}_{-1/2}(\eta_\pm)}{\mathscr{F}_{1/2}(\eta_\pm)}, \tag{4.9}$$

where $\mathscr{F}_{-1/2}$ and $\mathscr{F}_{1/2}$ are Fermi-Dirac integrals of order -1/2 and 1/2, and the normalized energies are

$$\eta_\pm(x) = (\mu_\pm - E_S(x))/k_B T_L, \tag{4.10}$$

in which μ_+ and μ_- are the Fermi-levels associated with + and − stream respectively, and $E_S(x)$ is the lowest subband energy of our model device. The Fermi-levels depend on whether a stream comes from the source or from the drain contact [1].

For the source injection, we solve Eqs. 4.1 and 4.2 with the backscattering coefficients in Eqs. 4.7 and 4.8. The normalized energies are

$$\eta_+(x) = \eta_S(x) = (\mu_S - E_S(x))/k_B T_L \tag{4.11}$$

$$\eta_-(x) = \eta_S(x_{\max}) = (\mu_S - E_S(x_{\max}))/k_B T_L, \tag{4.12}$$

where μ_S is the Fermi-level of the source contact and $x_{\max}$ is the position of the top of the source barrier. In the region $x_{\max} < x < L$ where $\mathscr{E}_x < 0$ in Fig. 4.4, $\eta_-(x)$ can be any finite value because there is no source injected negative stream in the ballistic

limit. Thus, Eq. 4.12 causes no error in the ballistic limit. In a similar way, we solve Eqs. 4.1 and 4.2 for the drain injection with

$$\eta_{+}(x) = \eta_{D}(x_{\max}) = (\mu_{D} - E_{S}(x_{\max}))/k_{B}T_{L} \tag{4.13}$$

$$\eta_{-}(x) = \eta_{D}(x) = (\mu_{D} - E_{S}(x))/k_{B}T_{L}, \tag{4.14}$$

where μ_D is the Fermi-level of the drain contact.

To solve Eqs. 4.1 and 4.2, we need to specify: 1) hierarchy closure assumptions on the average velocities of source and drain injections, $\langle \upsilon_x \rangle_S^{\pm}$ and $\langle \upsilon_x \rangle_D^{\pm}$, and 2) boundary conditions for source and drain injections.

4.3.3 Ballistic hierarchy closure approximations for the streams under acceleration

Figure 2.7 depicts the development of ballistic peaks along the channel of the model device [1], and Fig. 4.6 shows the corresponding average velocity profiles under bias. Figure 2.7 implies that in the ballistic limit, the distribution function approaches a Dirac delta function, and can be written as (assuming parabolic band structure)

$$f(x, \vec{p}) = g(x) \cdot \delta\left(\vec{p} - \hat{x}\, p_{x\max}\right), \tag{4.15}$$

$$p_{x\max} = \sqrt{2m_t^{*}(E_{\max} - E_S(x))}, \tag{4.16}$$

where $E_{\max}$ is the maximum of $E_S(x)$ (Fig. 4.4) and m_t^* is the transverse effective mass of ellipsoidal valleys of Si. As shown in Fig. 2.7, in the limit where the δ-peak approximation applies, the distribution in Eq. 4.15 yields

$$\langle \upsilon_x \rangle = \sqrt{\langle \upsilon_x^2 \rangle} = ... = \sqrt[l]{\langle \upsilon_x^l \rangle} = \frac{p_{x\max}}{m_t^{*}}, \tag{4.17}$$

where l is the order of moments. Consequently, Eq. 4.17 allows us to terminate the hierarchy of macroscopic moments. This is analogous to Baraff's maximum anisotropy closure for the spherical harmonics expansion of the distribution function [19].

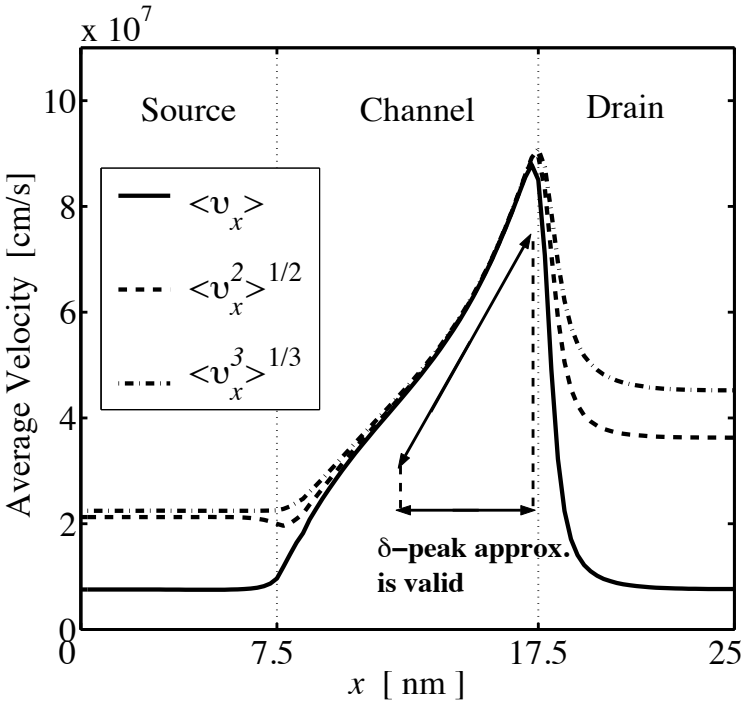

Fig. 4.6. The profiles of average velocities at the first (solid), second (dashed), and the third order moments (dash-dotted) in the ballistic limit. In the channel region where the ballistic peak develops as shown in Fig. 2.7, Eq. 4.17 is a valid approximation.

We build ballistic closures for $\langle \upsilon_x \rangle_S^{\pm}$ and $\langle \upsilon_x \rangle_D^{\pm}$ based on the above argument. In $0 < x < x_{\max}$, only J_D^- experiences acceleration and in $x_{\max} < x < L$, only J_S^+ does. Hence, the following closure assumptions can apply,

$$\langle \upsilon_x \rangle_S^+ = \begin{cases} \tilde{\upsilon}_T \left(\eta_S \left(x \right) \right) & 0 < x < x_{\max} \\ \sqrt{\tilde{\upsilon}_T^2 \left(\eta_S \left(x_{\max} \right) \right) + \frac{2 \left(E_{\max} - E_S(x) \right)}{m_t^*}} & x_{\max} < x < L \end{cases} \tag{4.18}$$

$$\langle \upsilon_x \rangle_S^- = \tilde{\upsilon}_T \left(\eta_S \left(x \right) \right) \qquad 0 < x < x_{\max} \tag{4.19}$$

$$\langle \upsilon_x \rangle_D^+ = \tilde{\upsilon}_T \left(\eta_D \left(x \right) \right) \qquad x_{\max} < x < L \tag{4.20}$$

$$\langle \upsilon_x \rangle_D^- = \begin{cases} \sqrt{\tilde{\upsilon}_T^2 \left(\eta_D \left(x_{\max} \right) \right) + \frac{2 \left(E_{\max} - E_S(x) \right)}{m_t^*}} & 0 < x < x_{\max} \\ \tilde{\upsilon}_T \left(\eta_D \left(x \right) \right) & x_{\max} < x < L \end{cases} . \tag{4.21}$$

The degenerate thermal velocity $\tilde{\upsilon}_T \left(\eta \right)$ for a two-dimensional electron gas (2DEG) is given as [33]

$$\tilde{\upsilon}_T \left(\eta \right) = \sqrt{\frac{2 k_B T_L}{\pi m_t^*}} \frac{\mathscr{F}_{1/2} \left(\eta \right)}{\ln \left(1 + e^{\eta} \right)}. \tag{4.22}$$

There are two things worth noting in Eqs. 4.18 $\sim$ 4.21. First, the closures in Eqs. 4.18 and 4.21 reduce to near-equilibrium closures at $x_{\max}$ (because $E_{\max} - E_S \left(x \right) = 0$) capturing the injection limit properly, but become high-field ballistic closures satisfying Eq. 4.17 where the accelerated carriers develop to a δ-peak away from $x_{\max}$. Second, $\langle \upsilon_x \rangle_S^-$ in $x_{\max} < x < L$ and $\langle \upsilon_x \rangle_D^+$ in $x_{\max} < x < L$ are not specified because in the ballistic limit, J_S^- and J_D^+ do not exist in the respective regions.

4.3.4 Boundary conditions

To solve Eqs. 4.1 and 4.2 self-consistently with the Poisson equation, the carrier densities and flux densities of the streams coming into the device should be specified. Those quantities are directly obtained from the distributions given at the contacts. For the source injection, the boundary distributions of the incoming fluxes are

$$f_S \left(x = 0, p_x > 0 \right) = \frac{1}{1 + \exp \left[\left(E \left(x = 0 \right) - \mu_S \right) / k_B T_L \right]} \tag{4.23}$$

at the source contact, and

$$f_S\left(x = L, p_x < 0\right) = 0 \tag{4.24}$$

at the drain contact if perfect absorbing contacts are assumed [1]. In Eqs. 4.23 and 4.24, the total energy of a 2D electron in the x-y plane (due to the vertical confinement in z in our model device) is

$$E\left(x\right) = E_S\left(x\right) + \frac{p_x^2 + p_y^2}{2m_t^*}. \tag{4.25}$$

For the drain injection, the boundary distribution given at the drain contact is

$$f_D\left(x = L, p_x < 0\right) = \frac{1}{1 + \exp\left[\left(E\left(x = L\right) - \mu_D\right)/k_B T_L\right]}, \tag{4.26}$$

and at the source contact, it is

$$f_D\left(x = 0, p_x > 0\right) = 0. \tag{4.27}$$

Then, integrating the above boundary distributions, we obtain the boundary conditions, which are

$$n_S^+\left(0\right) = \frac{m_t^* k_B T_L}{2\pi\hbar^2} \log\left[1 + \exp\left(\eta_S\left(0\right)\right)\right], \tag{4.28}$$

$$J_S^+\left(0\right) = n_S^+\left(0\right)\, \tilde{v}_T\left(\eta_S\left(0\right)\right), \tag{4.29}$$

and

$$J_S^-\left(L\right) = 0 \tag{4.30}$$

for the source injection, and

$$n_D^-\left(L\right) = \frac{m_t^* k_B T_L}{2\pi\hbar^2} \log\left[1 + \exp\left(\eta_D\left(L\right)\right)\right], \tag{4.31}$$

$$J_D^-\left(L\right) = n_D^-\left(L\right)\, \tilde{v}_T\left(\eta_D\left(L\right)\right), \tag{4.32}$$

and

$$J_D^+\left(0\right) = 0 \tag{4.33}$$

for the drain injection, respectively.

In principle, we solve Eqs. 4.1 and 4.2 with Eqs. 4.28 $\sim$ 4.30 for the source injection, and with Eqs. 4.31 $\sim$ 4.33 for the drain injection. However, we convert the extended

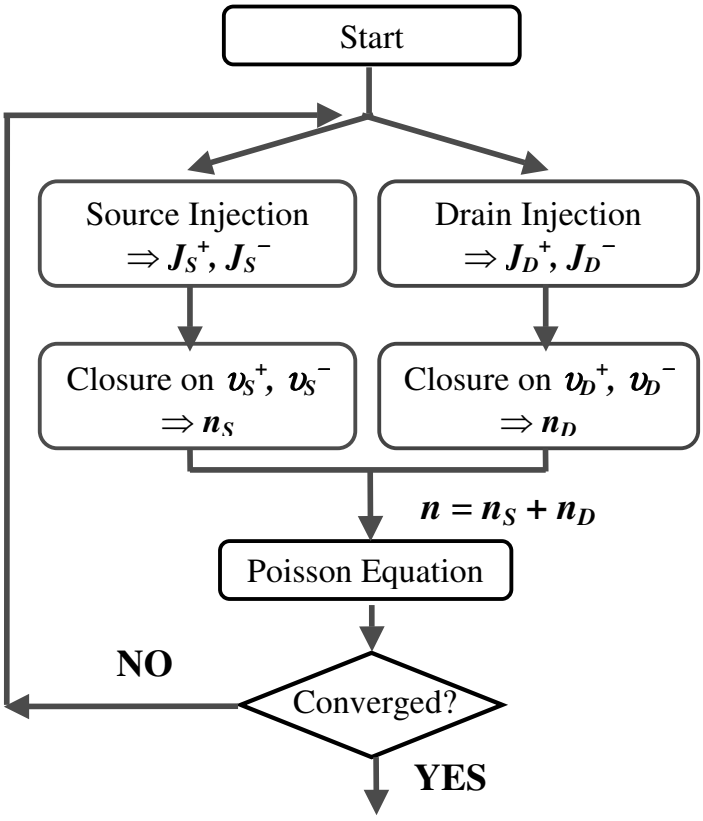

Fig. 4.7. A flow chart for the self-consistent simulation of the ballistic drift-diffusion model.

one-flux equations and the boundary conditions into a drift-diffusion form and solve numerically using the Scharfetter-Gummel discretization method [51]. The procedure is illustrated in the flow chart in Fig. 4.7.

4.4 A Drift-Diffusion Equation for Ballistic MOSFETs

4.4.1 Conversion into a drift-diffusion equation

By subtracting and adding Eqs. 4.1 and 4.2, we obtain

$$\frac{dJ}{dx} = 0 \tag{4.34}$$

$$J = -\frac{\xi - \xi'}{\xi + \xi'}\upsilon_T N - \frac{\upsilon_T}{\xi + \xi'}\frac{dN}{dx}, \tag{4.35}$$

where the quantities J and N are defined as

$$J \equiv J^+ - J^- \ [1/cm^2 \cdot s], \tag{4.36}$$

and

$$N \equiv \left(J^+ + J^-\right) \Big/ \upsilon_T \ [cm^{-3}]. \tag{4.37}$$

The thermal velocity υ_T given in Eq. 4.37 just defines the unit of N. Note that J denotes the net flux density but that N is not the actual carrier density but simply represents the sum of $J^{\pm}$ in the unit of carrier density. Thus, Eq. 4.35 is not restricted to the carriers moving at a fixed velocity υ_T. Using Eqs. 4.3 and 4.4 with the implementation of degenerate statistics, Eq. 4.35 can be expressed as a drift-diffusion equation,

$$J = -\mu^* \mathscr{E}_x N - D^* \frac{dN}{dx}. \tag{4.38}$$

The equivalent mobility μ^* and diffusivity D^* are defined as

$$\mu^* \equiv \frac{\upsilon_T}{2\xi_0 \frac{k_B T_L/q}{F_{\deg}} + |\mathscr{E}_x|} \ [cm^2/V \cdot s], \tag{4.39}$$

and

$$D^* \equiv \frac{k_B T_L}{q F_{\deg}} \mu^* \ [cm^2/s], \tag{4.40}$$

where

$$F_{\deg} = \begin{cases} F^+_{\deg}(\eta_+) & \mathscr{E}_x > 0 \\ F^-_{\deg}(\eta_-) & \mathscr{E}_x < 0 \end{cases}. \tag{4.41}$$

Although Eq. 4.38 is in a drift-diffusion form, it describes spatial variation of flux densities due to transmission and reflection as in the original one-flux equations. In the near-equilibrium diffusive limit, Eq. 4.35 reduces to a drift-diffusion equation and the equivalent mobility and diffusivity become the low-field mobility and corresponding diffusivity [28]. However, when $\xi_0 = 0$, Eqs. 4.38 $\sim$ 4.41 describe ballistic transport. Equation 4.40 is an equivalent Einstein relation under degenerate conditions, which originates from the equilibrium Fermi-Dirac distribution associated with

field backscattering (see Appendix B.2). For degenerate bulk semiconductors, the Einstein relation is [52]

$$\frac{D}{\mu} \equiv \frac{k_B T_L}{q} \left(\frac{n}{dn/d\eta} \right)^{-1}, \tag{4.42}$$

which would yield the degeneracy factor for the 2D carriers in our model device as

$$F_{\text{deg}} = \frac{n}{dn/d\eta} = \frac{\mathscr{F}_{-1}(\eta)}{\mathscr{F}_0(\eta)} = \frac{e^{\eta}}{(1+e^{\eta})\ln(1+e^{\eta})}, \tag{4.43}$$

where n is carrier density. However, the degeneracy factors in Eq. 4.9 imply that

$$F^{\pm}_{\text{deg}} = \frac{\mathscr{F}_{-1/2}(\eta_{\pm})}{\mathscr{F}_{1/2}(\eta_{\pm})} = \frac{J^{\pm}}{dJ^{\pm}/d\eta_{\pm}} \tag{4.44}$$

Equation 4.44 reduces to Eq. 4.43 when the degeneracy associated with the average velocities is independent of position (or $\eta_{\pm}$), which is true in uniform bulk semiconductors.

4.4.2 Boundary conditions

Using Eqs. 4.36 and 4.37, the boundary conditions for Eqs. 4.34 and 4.38 are expressed as

$$N(0)\left(\upsilon_T - \mu^* \mathscr{E}_x(0)\right) - D^* \left.\frac{dN}{dx}\right|_0 = 2J^+(0) \tag{4.45}$$

at the source contact, and

$$N(L)\left(\upsilon_T + \mu^* \mathscr{E}_x(L)\right) + D^* \left.\frac{dN}{dx}\right|_L = 2J^-(L) \tag{4.46}$$

at the drain contact, where $J^+(0)$ and $J^-(L)$ are given in Eqs. 4.29 and 4.30 for the source injection and Eqs. 4.32 and 4.33 for the drain injection. We solve Eqs. 4.34 and 4.38 for the source injection using the Scharfetter-Gummel discretization technique [51] with the boundary conditions given by Eqs. 4.45 and 4.46 to obtain $N_S(x)$ and $J_S(x)$ or $J_S^{\pm}(x)$. The discretization of the equations and the boundary conditions are described in Appendix B.3. Then, using the closure assumptions in Eqs. 4.18 and 4.19, we get the carrier densities of the source injection, $n_S^{\pm}$ using Eq. 4.5. Simi-

larly, we solve Eqs. 4.34 and 4.38 for the drain injection using Eqs. 4.45 and 4.46 to obtain $J_D^{\pm}$, and then $n_D^{\pm}$. The total carrier density is then,

$$n = n_S^+ + n_S^- + n_D^+ + n_D^-, \tag{4.47}$$

and the total flux net density is

$$J = \left(J_S^+ - J_S^-\right) + \left(J_D^+ - J_D^-\right). \tag{4.48}$$

4.5 Results

The I_D-V_{DS} characteristics and injection quantities vs. drain bias (injection velocity and injection charge density at the top of the barrier) of the model nanoscale MOSFET are plotted in the left column of Fig. 4.8. They are in an excellent agreement with the solution of the ballistic BTE [1] showing that the new closure approximations in Eqs. 4.18 $\sim$ 4.21 do not cause error in capturing the injection limit. Then, we plot profiles of internal quantities (the first subband profile, internal velocity, and charge density vs. position) in the right column of Fig. 4.8. Also, they show an excellent overall agreement with the solution of the ballistic BTE except the tolerable errors in the low-field region after the barrier where the δ-peak assumption in Eq. 4.17 is not valid.

In Fig. 4.9, similar plots are obtained for the model CNTFET in Fig. 4.1. We make proper modifications to the bandstructure and the transverse mode calculation in the Fermi-Dirac integrals as shown in Appendix B.4 so that it can be compared to the ballistic BTE result obtained in [2]. They also show good agreement with the results of the ballistic BTE although the mismatch caused by the assumption in Eq. 4.17 is more pronounced than in the MOSFET case.

These results show that a macroscopic transport model based on the moments of the BTE can describe carrier transport in the ballistic limit for nanoscale FETs.

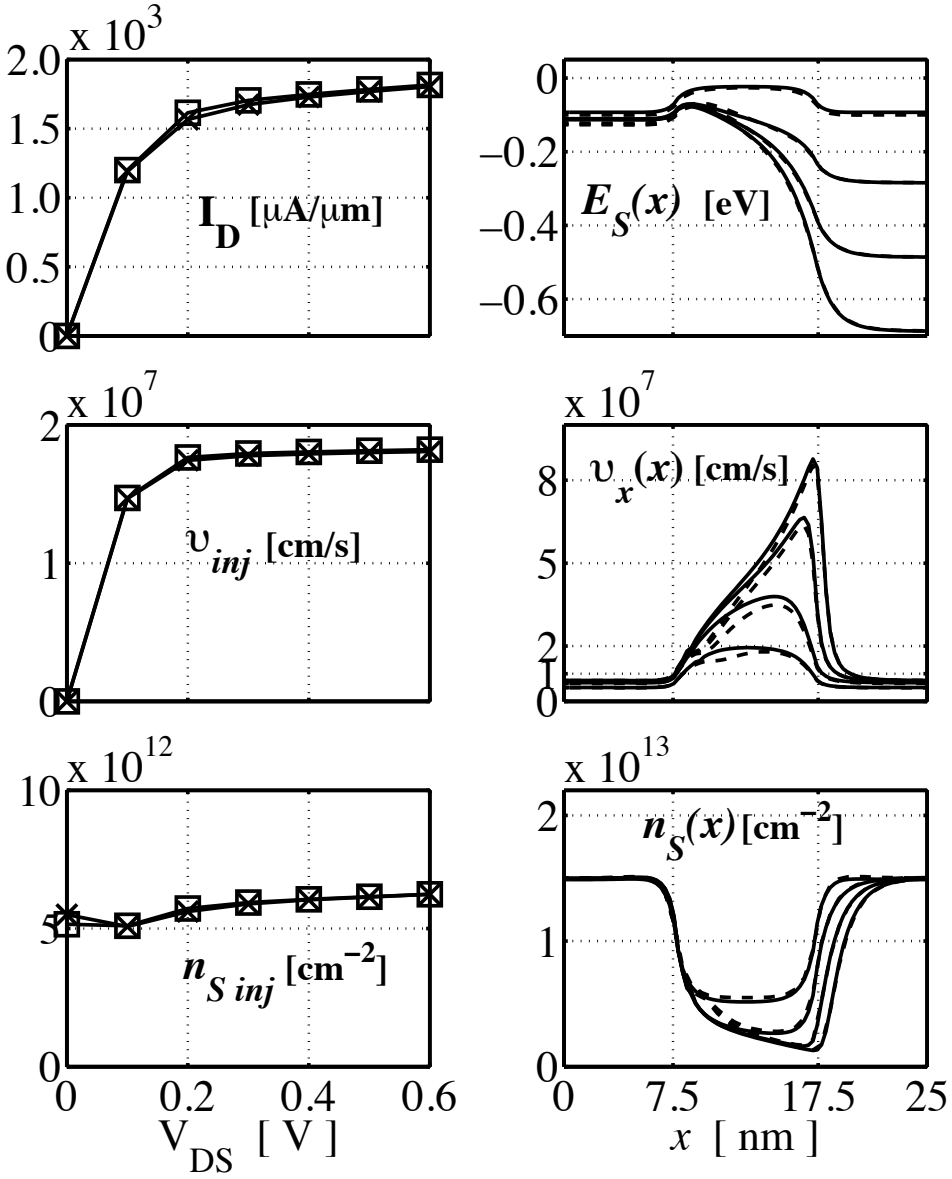

Fig. 4.8. Results for the model MOSFET. The I–V characteristics and injection quantities vs. V_{DS} (left column), and the profiles of the first subband energy, carrier velocity, and the charge density along the position (right column). The results of the ballistic drift-diffusion equation (x or dashed lines) accurately match the results of the ballistic BTE from [1] (square or solid lines).

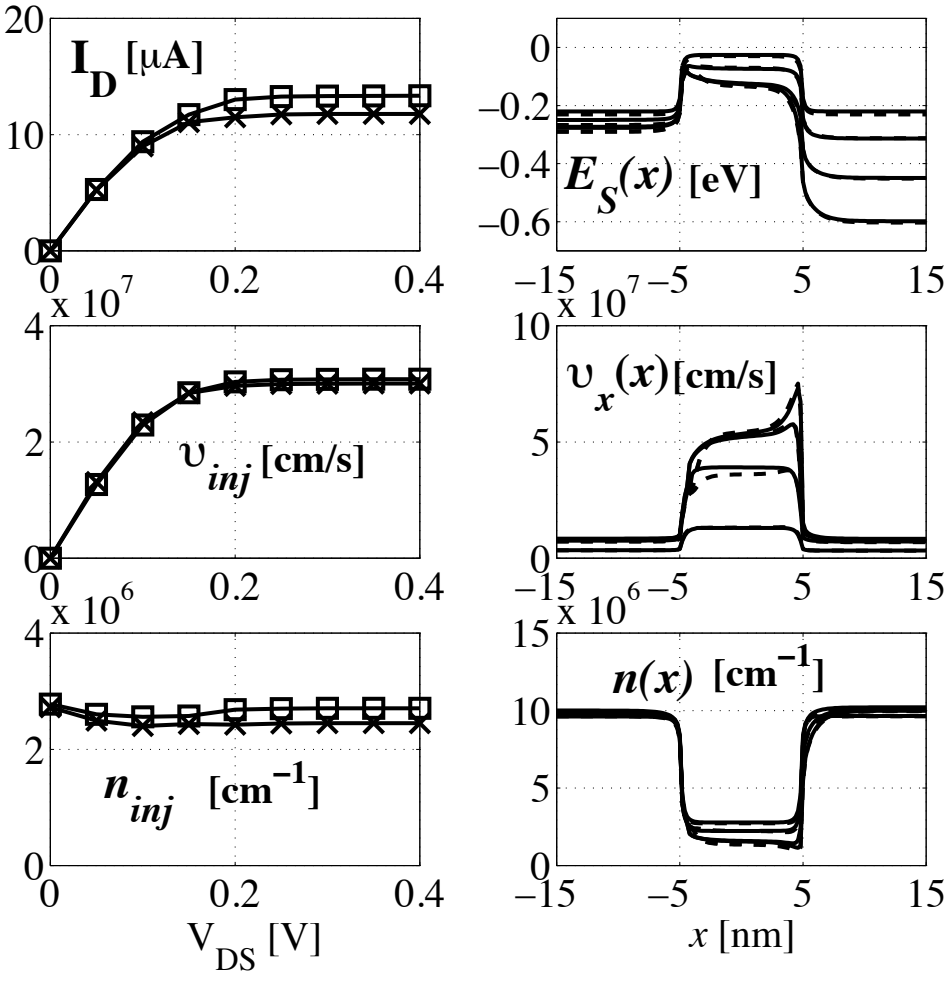

Fig. 4.9. Results for the model CNTFET. The I–V characteristics and injection quantities vs. V_{DS} (left column), and the profiles of the first subband energy, carrier velocity, and the charge density along the position (right column). The results of the ballistic drift-diffusion equation (x or dashed lines) well match the results of the ballistic BTE from [2] (square or solid lines).

4.6 Summary

We derived and solved a drift-diffusion equation in the ballistic limit for nanoscale FETs. The equations are developed from the directed-moment approach, which extends McKelvey's one-flux method by: 1) treating carrier injection from the source and drain separately, 2) introducing new hierarchy closure approximations for the streams under acceleration, and 3) including degenerate carrier statistics involving a properly-defined Einstein relation. The results show that a moment-based macroscopic transport model can describe ballistic transport in excellent agreement with the solution of the ballistic BTE with no fitting parameters.

The development of the ballistic macroscopic equation was relatively simple because the hierarchy closure approximations were independent of scattering. When scattering is present, however, this simplification should be re-examined. The scattering and the accelerating electric field bring in complicated transport because the strength of scattering depends on the carriers' kinetic energy, which again depends on the interplay of the accelerating field and the impeding scattering. Describing these phenomena in an acceptable error range still remains as the most difficult challenge in developing a macroscopic model valid in the high-field quasi-ballistic regime.

5. A DRIFT-DIFFUSION EQUATION FOR QUASI-BALLISTIC TRANSPORT IN NANOSCALE MOSFETS

We develop a drift-diffusion type macroscopic transport model that can describe carrier transport from the diffusive limit to the ballistic limit in a nanoscale MOSFET. We also establish physical upper and lower limits for the current-voltage (I–V) characteristics and average carrier velocities. Because the range between these two limits is narrow, they can be used to assess the validity of various transport models. The I–V characteristics and the channel velocity profiles obtained for a model nanoscale MOSFET fall between the two limits, showing that our new drift-diffusion model can describe quasi-ballistic transport within a tolerable amount of error as well as diffusive and ballistic transport. Finally, we discuss the limitations of the model.

5.1 Introduction

As transistors are scaled down to their ultimate limit, carrier transport may approach the ballistic limit, where commonly used macroscopic transport models may lose their validity [12] [39] [47]. Thus, it is of interest to develop a new, reliable macroscopic transport model that can routinely assess various nanoscale transistor design options. Figure 2.7 shows the ballistic distribution function at different locations along the channel direction under high gate and drain bias in a nanoscale double-gate MOSFET depicted in Fig. 2.1. The corresponding conduction energy band profile is illustrated in Fig. 5.1. The overall distribution is highly distorted from that of collision-dominated transport. Note that there are two distinct populations that experience different transport in the same region. The carriers injected from the

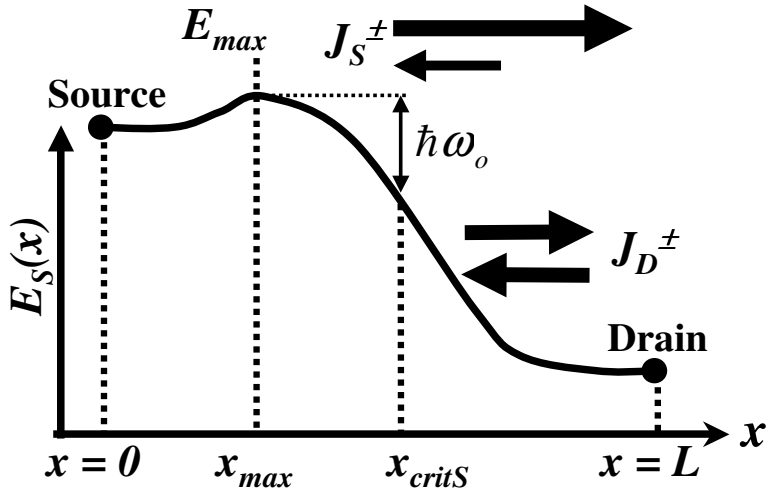

Fig. 5.1. Illustration of carrier transport along the lowest subband profile $E_S(x)$ in a nanoscale MOSFET. $E_{\max}$ and $x_{\max}$ denote the energy and the location of the top of the source-to-channel barrier. At $x = x_{critS}$, we assume optical phonon scattering is activated for the source stream. There are four types of carrier flows: 1) J_S^+, the source-injected + stream, 2) J_S^-, the source-injected − stream, 3) J_D^+, the drain-injected + stream, and 4) J_D^-, the drain-injected − stream, which experience different transport respectively.

source cross the source-channel barrier showing equilibrium hemi-Fermi-Dirac distribution and then get accelerated in the channel developing into a ballistic peak. The carriers from the drain are decelerated by the strong channel electric field undergoing near-equilibrium transport and are essentially confined in the drain region. As discussed in Chapter 2, the asymmetric but near-equilibrium shape distribution at the top of the barrier sets the thermal injection limit that primarily determines the on-current. The ballistic peak in the channel causes velocity to overshoot and affects the injection limit indirectly through two-dimensional (2D) electrostatics.

A new transport model for nanoscale transistors should be consistent with the physical picture described above. The challenges are [53]:

1. The two-population picture—we need two sets of transport equations: one for the source-injected carriers and the other for the drain-injected carriers.

2. Asymmetry in the distribution—especially at the top of the barrier

3. The ballistic peak and velocity overshoot—this requires us to develop different hierarchy closure approximations for different streams.

4. Degenerate carrier statistics must be treated—especially at the top of the barrier.

In the previous chapter, we developed a ballistic drift-diffusion model by extending McKelvey's one-flux model [27] [28] resolving the above challenges, and demonstrated that it matches the results of the ballistic Boltzmann transport equation (BTE) [53]. In this chapter, we will add scattering to the ballistic DD model so that it can cover the entire regime of quasi-ballistic transport between the diffusive and the ballistic limits.

The chapter is organized as follows. In Section 5.2, we review the theory of the previously developed ballistic drift-diffusion model, and then implement a simple scattering model into it. We also identify physical upper and lower limits of transport. In Section 5.3, we show the results obtained for the model nanoscale MOSFET and compare them to the physical limits established in Section 5.2. In Section 5.4, we discuss the assumptions employed in our model and corresponding limitations. Finally, we summarize in Section 5.5.

5.2 Theory

5.2.1 Review of theory

We start from our previously developed ballistic drift-diffusion model in Chap. 4. The model equations in the form of a conventional drift-diffusion equation are

$$\frac{dJ}{dx} = 0, \tag{5.1}$$

and

$$J = q\mu^* \mathscr{E}_x N + qD^* \frac{dN}{dx}. \tag{5.2}$$

The unknowns are the net current J and the equivalent carrier density N (not the actual carrier density), which are expressed in terms of flux densities J^+ and J^- as

$$J \equiv -q\left(J^+ - J^-\right), \tag{5.3}$$

and

$$N \equiv \left(J^+ + J^-\right)\Big/ \upsilon_T. \tag{5.4}$$

Note that the actual carrier density is

$$n\left(x\right) \equiv J^+\left(x\right)\Big/ \upsilon^+\left(x\right) + J^-\left(x\right)\Big/ \upsilon^-\left(x\right), \tag{5.5}$$

where $\upsilon^+(x)$ and $\upsilon^-(x)$ are the average velocities of the positive and negative fluxes. The equivalent mobility μ^* and the equivalent diffusivity D^* are defined as

$$\mu^* \equiv \frac{\upsilon_T}{\mathscr{E}_{scatt} + |\mathscr{E}_x|}, \tag{5.6}$$

and

$$D^* \equiv \frac{k_B T_L}{q F_{\text{deg}}}\mu^*. \tag{5.7}$$

The scattering field $\mathscr{E}_{scatt}$ accounts for the field-free backscattering effect, given as

$$\mathscr{E}_{scatt} = 2\xi_0 \frac{k_B T_L/q}{F_{\text{deg}}}, \tag{5.8}$$

where ξ_0 is the field-free backscattering probability per length, and F_{deg} is a degeneracy factor that approaches unity for nondegenerate conditions. Equation 5.7 is a modified Einstein relation. Equations 5.1 and 5.2 reduce to the conventional drift diffusion model when $\mathscr{E}_{scatt} >> |\mathscr{E}_x|$ [44].

We solve Eqs. 5.1 and 5.2 in the same way as in the ballistic drift-diffusion equation (see the flow chart in Fig. 4.7) except that we include the scattering term. In the previous chapter, scattering was set to zero and a hierarchy closure approximation to describe the ballistic peaks was constructed based on Baraff's maximum anisotropy principle [19]. The challenge in this chapter is to include effects of scattering not only in the scattering term but also in the hierarchy closure approximations.

5.2.2 Addition of scattering

Here, we implement a simple scattering into the scattering term in Eq. 5.2. In determining the on-current and average carrier velocity, momentum relaxation scattering processes play the most important role. Low-energy acoustic phonon scattering and surface roughness scattering are the most effective in the low-field region around the top of the barrier. For this, we specify a constant mean-free-path λ_{el}, which is related to the low-field mobility as shown in Appendix C.1. We are interested in the intrinsic device operation ignoring source/drain series resistances, but a doping dependent $\lambda_{el}(x)$ model can be readily implemented. The next important scattering is energy dissipation processes in the high-field region. Inelastic optical phonon scattering becomes active for the source-injected stream after carriers gain energy higher than a phonon quantum $\hbar\omega_o$ in a certain distance away from the barrier (see Fig. 5.1). We ignore phonon absorption scattering and velocity saturation because our focus is on the overshoot velocity in the high-field region that is too short to cause complete energy relaxation. For this, we specify a constant mean-free-path λ_o, and introduce a critical location x_{critS} for the source-injected stream beyond which optical phonon emission is activated. Thus, in the region $0 < x < x_{critS}$ in Fig. 5.1, the field-free backscattering for the source-injected stream is mainly due to the isotropic elastic scattering, that is

$$\xi_0 = \lambda_{el}^{-1} \qquad \text{if } E < \hbar\omega_o \ , \tag{5.9}$$

where E is average carrier energy. On the other hand, in $x_{critS} < x < L$, we combine the two mechanisms according to Matthiessen's rule as

$$\xi_0 = \lambda_{el}^{-1} + \lambda_o^{-1} \qquad \text{if } E > \hbar\omega_o \ . \tag{5.10}$$

To avoid a discontinuity, we connect Eq. 5.9 to Eq. 5.10 as

$$\xi_0 = \frac{1}{\lambda_{el}} + \frac{1}{\lambda_{in}(E)} \tag{5.11}$$

assuming the following energy-dependent inelastic mean-free-path

$$\frac{1}{\lambda_{in}(E)} = \frac{1}{\lambda_o}\left[1 - \exp\left(-\frac{E - \hbar\omega_o}{E_o}\right)\right] H(E - \hbar\omega_o), \tag{5.12}$$

where $H(\bullet)$ is the unit step function, λ_o is the high-energy optical phonon mean-free-path in silicon, and the energy E_o determines how fast $1/\lambda_{in}(E)$ reaches $1/\lambda_o$ from zero. We model scattering for the drain-injected carriers in the same way to achieve symmetry but under high drain bias it is negligible.

Although the functional form of Eq. 5.12 is somewhat arbitrary, we do not expect it to generate a large error. In the ballistic drift-diffusion model in Chap. 4, we made hierarchy closure approximations exact only at the top of the barrier and the high-field region near the drain. In between, the interpolation between the two values introduced a tolerable error. A scattering model also need not be exact in between. Two things are worth noting: 1) the elastic scattering primarily determines the thermal injection velocity (or the current) while the inelastic scattering affects the velocity overshoot directly but the thermal injection indirectly, and 2) λ_{el}, λ_o, $\hbar\omega_o$, and E_o are the parameters to specify in our model.

5.2.3 Quasi-ballistic hierarchy closure approximations

Building a hierarchy closure model inevitably involves an assumption on the shape of the carrier distribution. Conventional transport models assume collision-dominated transport and use a thermal distribution. In our ballistic drift-diffusion model, we adopted the idea of superimposing on a thermal distribution a delta function that describes the ballistic peak, which followed Baraff's approach to spherical harmonics expansion solution of the Boltzmann equation [19]. The ballistic case was simple because the distribution is well approximated by the superposition of a thermal distribution and a delta function. In the presence of scattering, however, the simple ballistic closure model is no longer valid because scattering redistributes carrier energy and momentum resulting in a distribution that deviates from the simple superposition. Now, we build a quasi-ballistic closure model that can describe carriers that are removed from the ballistic peak due to scattering as well as those that remain in it.

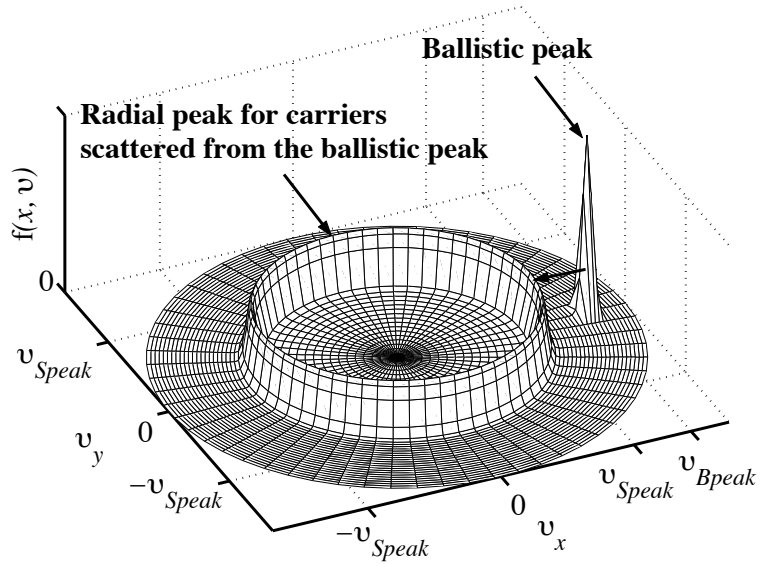

Fig. 5.2. A carrier distribution assumed to obtain the quasi-ballistic hierarchy closure model in Eqs. 5.23 $\sim$ 5.26. The point δ-peak represents the ballistic peak and the radial δ-peak describes energetic carriers scattered isotropically from the ballistic peak.

We assume that for a stream under acceleration, the following distribution (depicted in Fig. 5.2) is superimposed on the thermal distribution,

$$f\left(x,\vec{\upsilon}\right)=n_B\left(x\right)\,\delta\left(\vec{\upsilon}-\vec{\upsilon}_{B\,peak}\right)+\left[n_{total}\left(x\right)-n_B\left(x\right)\right]\delta\left(\vec{\upsilon}-\vec{\upsilon}_{S\,peak}\right). \tag{5.13}$$

The first delta function represents the carriers that remain ballistic at position x, whose density is $n_B(x)$ and peak velocity in the transport (x-) direction is located at

$$\vec{\upsilon}_{B\,peak}=\hat{x}\,\left.\frac{\partial E}{\partial p_x}\right|_{E=E_{\max}-E_S(x)}=\hat{x}\sqrt{\frac{2\left[E_{\max}-E_S\left(x\right)\right]}{m^*}}, \tag{5.14}$$

where $E_{\max}$ is the energy at the top of the barrier (see Fig. 5.1). Note that the potential energy difference $E_{\max}-E_S(x)$ becomes the kinetic energy of the ballistic carriers. Here, we assume a parabolic band for simplicity, but nonparabolicity can be easily implemented.

The second term represents carriers that are isotropically scattered from the ballistic peak. The density is $n_{total}(x)-n_B(x)$ where $n_{total}(x)$ is the total density of the

stream under acceleration. In this case, we assume that the isotropically scattered carriers form a delta peak distribution in the radial (r-) direction, and that if their kinetic energy is higher than $\hbar\omega_o$, they start to lose energy to phonon emission. Thus, the peak velocity in the second delta function is modeled as

$$\vec{v}_{S\,peak} = \hat{r}\ \left.\frac{\partial E}{\partial p_r}\right|_{E=E_{\max}-E_S(x)} = \hat{r}\sqrt{\frac{2\left[E_{\max}-E_S\left(x\right)\right]}{m^*}} \tag{5.15}$$

for $E_{\max} - E_S(x) < \hbar\omega_o$, and

$$\vec{v}_{S\,peak} = \hat{r}\sqrt{\frac{2\left[E_{\max}-E_S\left(x\right)-\Delta E_L\right]}{m^*}}H\left(E_{\max}-E_S\left(x\right)-\Delta E_L\right) \tag{5.16}$$

for $E_{\max} - E_S(x) \geq \hbar\omega_o$. In Eq. 5.16, the energy loss ΔE_L due to inelastic scattering is assumed to be

$$\Delta E_L = \left|\int_{x_{critS}}^{x}\frac{dx'}{\lambda_{in}\left[E=E_{\max}-E_S\left(x'\right)\right]}\right|\hbar\omega_o \tag{5.17}$$

for source-injected carriers in $x_{critS} < x < L$. The absolute value of the integral in Eq. 5.17 denotes the average number of inelastic scattering events along the distance $|x - x_{crit\,S}|$. The energy dissipation model in Eqs. 5.15 $\sim$ 5.17 is similar to that in [54], which is shown to be a good approximation in impact ionization and hot electron injection simulation.

From Eq. 5.13, we construct a hierarchy closure model including scattering as

$$\langle\upsilon\rangle^{\pm} = \sqrt{\tilde{v}_T^2\left(x=x_{\max}\right)+\frac{1}{4}\upsilon_{S\,peak}^2\left[1-B\left(x\right)\right]+\upsilon_{B\,peak}^2B\left(x\right)} \tag{5.18}$$

for a stream under acceleration (positive for $x_{\max} < x < L$ and negative for $0 < x < x_{\max}$), and

$$\langle\upsilon\rangle^{\mp} = \sqrt{\tilde{v}_T^2\left(x=x_{\max}\right)+\frac{1}{4}\upsilon_{S\,peak}^2\left[1-B\left(x\right)\right]} \tag{5.19}$$

for the corresponding backscattered stream. The derivation of Eqs. 5.18 and 5.19 is presented in Appendix C.2. In Eqs. 5.18 and 5.19, $\tilde{v}_T\left(x=x_{\max}\right)$ is the thermal injection velocity at the top of the barrier, υ_{Bpeak} and υ_{Speak} are the magnitudes of peak velocities in Eqs. 5.15 and 5.16 respectively. The ballisticity factor $B(x)$ is the

ratio of the ballistic carriers to the total carriers in the stream under acceleration at location x in the device, that is $B(x) = n_B(x)/n_{total}(x)$, $0 \leq B(x) \leq 1$. As in the simple ballistic closure model [53], the thermal velocity term dominates at the top of the barrier over the other term(s) in the square root of Eqs. 5.15 and 5.16 capturing thermal injection limit properly. Away from the barrier, velocity overshoot occurs and the second (x-directed velocity of isotropically scattered energetic carriers) and the third term (velocity of ballistic carriers in Eq. 5.18) are dominant over the thermal velocity.

Now the remaining problem is how to model $B(x)$, which is not known unless we solve the BTE. We model $B(x)$ as

$$B\left(x\right) = \exp\left(-\xi_0 \left|x - x_{\max}\right|\right), \tag{5.20}$$

which is the solution of the phenomenological equation

$$\frac{dn_B}{dx} = -\xi_0 n_B \tag{5.21}$$

with

$$n_B\left(x = x_{\max}\right) = n_{total}\left(x_{\max}\right) \tag{5.22}$$

Note that Eqs. 5.20 ∼5.22 are approximately valid only at the top of the barrier (i.e., in the low-field region), but we assume that they are adequate away from the barrier, which will be shown to work. Also note in Eq. 5.20 that: 1) $0 \leq B(x) \leq 1$, which is consistent with the definition, $B(x) = n_B(x)/n_{total}(x)$, 2) $B(x)$=1 at $x = x_{\max}$, which implies that carriers at the injection point are yet to experience a scattering event, 3) $B\left(x\right) \rightarrow 0$ when $\xi_0 \left|x - x_{\max}\right| \gg 1$, which means that no carriers survive scattering after they travel many mean-free-paths after the barrier.

5.2.4 Complete set of hierarchy closure models

We present the complete set of closures for $v_S^{\pm}$ and $v_D^{\pm}$. We assign the quasi-ballistic closure model in Eq. 5.18 to the streams under acceleration (J_D^- in $0 < x < x_{\max}$ and

J_S^+ in $x_{\max} < x < L$) and the closure in Eq. 5.19 to the corresponding backscattered streams (J_D^+ in $0 < x < x_{\max}$ and J_S^- in $x_{\max} < x < L$). To the streams that undergo thermionic emission against a potential barrier (J_S^+ in $0 < x < x_{\max}$ and J_D^- in $x_{\max} < x < L$) and their backscattered counterparts (J_S^- in $0 < x < x_{\max}$ and J_D^+ in $x_{\max} < x < L$), we assign the thermal velocity of hemi-Fermi-Dirac. Hence, the following closure approximations apply:

$$\upsilon_S^+ = \begin{cases} \tilde{\upsilon}_T\left(\eta_S\left(x\right)\right) & 0 < x < x_{\max} \\ \sqrt{\tilde{\upsilon}_T^2\left(\eta_S\left(x_{\max}\right)\right) + \frac{1}{4}\upsilon_{S\,peak}^2\left[1 - B\left(x\right)\right] + \upsilon_{B\,peak}^2 B\left(x\right)} & x_{\max} < x < L \end{cases}, \tag{5.23}$$

$$\upsilon_S^- = \begin{cases} \tilde{\upsilon}_T\left(\eta_S\left(x\right)\right) & 0 < x < x_{\max} \\ \sqrt{\tilde{\upsilon}_T^2\left(\eta_S\left(x_{\max}\right)\right) + \frac{1}{4}\upsilon_{S\,peak}^2\left[1 - B\left(x\right)\right]} & x_{\max} < x < L \end{cases}, \tag{5.24}$$

$$\upsilon_D^+ = \begin{cases} \sqrt{\tilde{\upsilon}_T^2\left(\eta_D\left(x_{\max}\right)\right) + \frac{1}{4}\upsilon_{S\,peak}^2\left[1 - B\left(x\right)\right]} & 0 < x < x_{\max} \\ \tilde{\upsilon}_T\left(\eta_D\left(x\right)\right) & x_{\max} < x < L \end{cases}, \tag{5.25}$$

and

$$\upsilon_D^- = \begin{cases} \sqrt{\tilde{\upsilon}_T^2\left(\eta_D\left(x_{\max}\right)\right) + \frac{1}{4}\upsilon_{S\,peak}^2\left[1 - B\left(x\right)\right] + \upsilon_{B\,peak}^2 B\left(x\right)} & 0 < x < x_{\max} \\ \tilde{\upsilon}_T\left(\eta_D\left(x\right)\right) & x_{\max} < x < L \end{cases}. \tag{5.26}$$

The degenerate thermal velocity $\tilde{\upsilon}_T\left(\eta\right)$ for a two-dimensional electron gas (2DEG) is [33]

$$\tilde{\upsilon}_T\left(\eta\right) = \sqrt{\frac{2k_B T_L}{\pi m^*}} \frac{\mathscr{F}_{1/2}\left(\eta\right)}{\ln\left(1 + e^{\eta}\right)}, \tag{5.27}$$

where

$$\eta\left(x\right) = \left[\mu\left(x\right) - E_S\left(x\right)\right]/k_B T_L. \tag{5.28}$$

Note, however, that the Fermi-level in Eq. 5.28 is position dependent. In the ballistic limit, the Fermi-level at the top of the barrier is that of source contact μ_S for the positive stream. With scattering, there is no well-defined Fermi-level at the injection point but it should be between μ_S (the upper limit) and the Fermi-level that follows the linear potential drop in the source region as in the diffusive limit (the lower limit), that is

$$\mu\left(x\right) = \mu_S + \left.\frac{dE_S\left(x\right)}{dx}\right|_{linear} x. \tag{5.29}$$

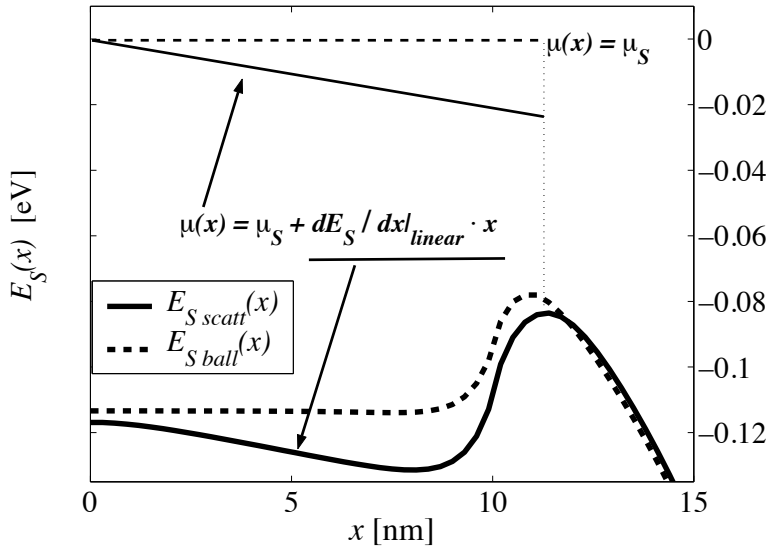

Fig. 5.3. Upper and lower limits of the Fermi-level for the source-injected carriers in $0 < x < x_{\max}$. In the ballistic limit, $\mu(x) = \mu_S$ (dashed), but in the diffusive limit, $\mu(x)$ follows the potential drop $dE_S(x)/dx$ (solid),which is approximately linear. We assume that the Fermi-level of the source-injected carriers follows Eq. 5.29

This is illustrated in Fig. 5.3. Using μ_S with scattering overestimates the degeneracy and the injection velocity while using Eq. 5.29 may underestimate them. Except in the regime where the ballistic carriers are predominant, i.e., λ_{el} is much larger than the size of a device, Eq. 5.29 is a good approximation. But we assume that Eq. 5.29 can also be used up to the ballistic limit because when scattering diminishes, $\mu(x)$ in Eq. 5.29 reduces to μ_S as the potential drop in the source region vanishes.

5.2.5 Four limits of carrier transport

The ballistic limit vs. the diffusive limit

The ballistic limit occurs when $\xi_0 \to 0$ and consequently $B(x) \to 1$. Then, Eq. 5.2 becomes the ballistic drift-diffusion equation, and the closure model in Eqs. 5.23 $\sim$ 5.26 reduces to the ballistic closure model in [53]. The diffusive limit is obtained if we let the velocities of the accelerated streams be limited by the thermal injection

velocity. Then, except the degenerate statistics, our transport model becomes the original one-flux equation, which is equivalent to conventional drift-diffusion models. This limit is achieved only when both momentum and energy are completely relaxed, which rarely occurs in nanoscale MOSFETs. It is clear that I–V characteristics and velocity profiles of any physical model should fall between these two limits. However, the range is too wide in practice to assess the validity of various transport models.

The maximum anisotropy limit vs. the complete isotropy limit

The ballisticity factor is bounded above and below as $0 \leq B(x) \leq 1$ regardless of scattering strength. Consequently, the closure approximations are also bounded above and below. For example, the velocities of the source-injected stream in $x_{\max} < x < L$ in Eqs. 5.23 and 5.24 are bounded as

$$\sqrt{\tilde{v}_T^2\left(\eta_S\left(x_{\max}\right)\right) + \frac{1}{4}v_{S\,peak}^2} \leq v_S^+ \leq \sqrt{\tilde{v}_T^2\left(\eta_S\left(x_{\max}\right)\right) + v_{B\,peak}^2} \tag{5.30}$$

$$\sqrt{\tilde{v}_T^2\left(\eta_S\left(x_{\max}\right)\right) + \frac{1}{4}v_{S\,peak}^2} \geq v_S^- \geq \tilde{v}_T\left(\eta_S\left(x_{\max}\right)\right). \tag{5.31}$$

The left boundaries in the above inequalities correspond to $B(x) = 0$, and the right boundaries to $B(x) = 1$. This allows us to establish two more physically meaningful limits in between the ballistic limit and the diffusive limit; results of any physical model should fall between these two limits.

The upper limit is attained assuming $B(x) = 1$ although we keep the scattering term in Eq. 5.6 (i.e., $\mathscr{E}_{scatt} \neq 0$). This means that the distribution for the closure approximation in Eq. 5.13 consists of the ballistic peak alone (see Fig. 5.2) but that we have a backscattered stream of thermal velocity. This sets upper limits on overshoot velocity and consequently on current in the quasi-ballistic regime because Eq. 5.30 assigns the maximum velocity $\sqrt{\tilde{v}_T^2 + v_{B\,peak}^2}$ to v_S^+ (a stream under acceleration) but Eq. 5.31 the thermal injection velocity $\tilde{v}_T$ to v_S^- (the backscattered counterpart). No transport model that exceeds this upper limit is physical. We call this upper limit the

maximum anisotropy limit since maximum difference between v_S^+ and v_S^- is achieved. Note that when $\mathscr{E}_{scatt} \to 0$, this upper limit reaches the ballistic limit.

We obtain the lower limit assuming $B(x) = 0$. This implies that the distribution in Eq. 5.13 consists of the radial peak alone (see Fig. 5.2) at any x as all carriers in the ballistic peak are scattered out. This sets lower limits on carrier velocity and current because as in Eqs. 5.30 and 5.31, v_S^+ takes its minimum $\sqrt{\tilde{v}_T^2 + v_{S\,peak}^2/4}$ and v_S^- gains the same magnitude of velocity, which is its maximum. No transport model that produces current and velocity below this limit can be physical. This lower limit will be called the complete isotropy limit since $v_S^+ = v_S^-$. Note that when the energy of scattered carriers is completely relaxed (i.e., $v_{Speak} = 0$, thus $v_S^+ = v_S^- = \tilde{v}_T$), it reaches the diffusive limit.

We will demonstrate that the range between these two limits is narrow enough in practice to assess the validity of various transport models.

5.3 Results

5.3.1 I–V characteristics and velocity profiles

Figure 5.4 shows the results obtained for the model device in Fig. 2.1. The on-current (Fig. 5.4a) and the injection velocity (Fig. 5.4b) vary gradually from the diffusive to the ballistic limit as λ_{el} increases. The overshoot peak velocity (Fig. 5.4d) increases showing reasonable amount of velocity overshoot but does not approach the velocity of the ballistic limit, so does the injection carrier density (Fig. 5.4c), which is affected by velocity overshoot through electrostatics [1]. The limitation of the peak overshoot velocity as λ_{el} increases is due to the inelastic scattering, which is not present in the pure ballistic limit. The parameters for the inelastic isotropic scattering are fixed at bulk values, i.e., $\lambda_o = 6.5$ nm, and $\hbar\omega_o = 63$ meV [55].

The inelastic scattering model of a bulk Si might not be appropriate for the 2D confined carriers in the model device. Note, however, that our transport model can accommodate a more rigorous model, and that the current of a MOSFET is mainly

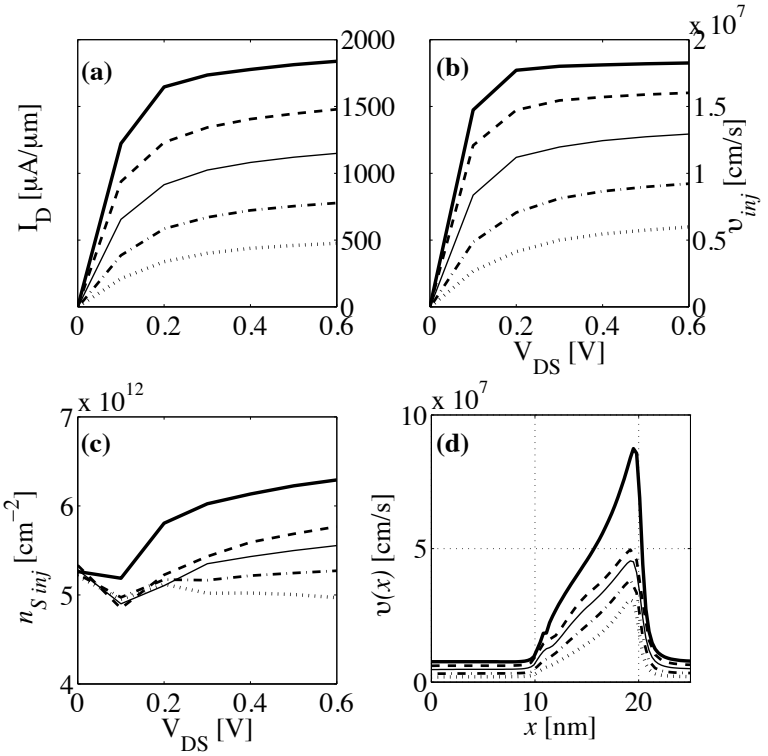

Fig. 5.4. Plots of (a) I–V, (b) v_{inj} vs. V_{DS}, (c) n_{inj} vs. V_{DS}, and (d) $v(x)$ profile of our transport model at different λ_{el}: 1) ballistic (thick solid), 2) λ_{el} =100 nm (dashed), 3) 20 nm (thin solid), 4) 7 nm (dash-dotted), and 5) 3 nm (dotted)

determined by scattering at the injection point as long as we have reasonable amount of velocity overshoot in the channel. This is demonstrated in Fig. 5.5. The I–V and the velocity profiles with and without inelastic scattering ($\lambda_o = \infty$) do not differ much even when λ_{el} = 20 nm (twice of the channel length) because elastic scattering into transverse modes also relaxes the longitudinal energy of carriers. It implies that a more rigorous inelastic scattering model does not have a large room for the overall accuracy improvement in MOSFET simulation. But it is expected to be critical in carbon nanotube or nanowire devices because they have no transverse modes and inelastic scattering is the only mechanism that can relax longitudinal carrier energy. Overall, the results indicate that our transport model is able to describe carrier transport from the diffusive to the ballistic limit.

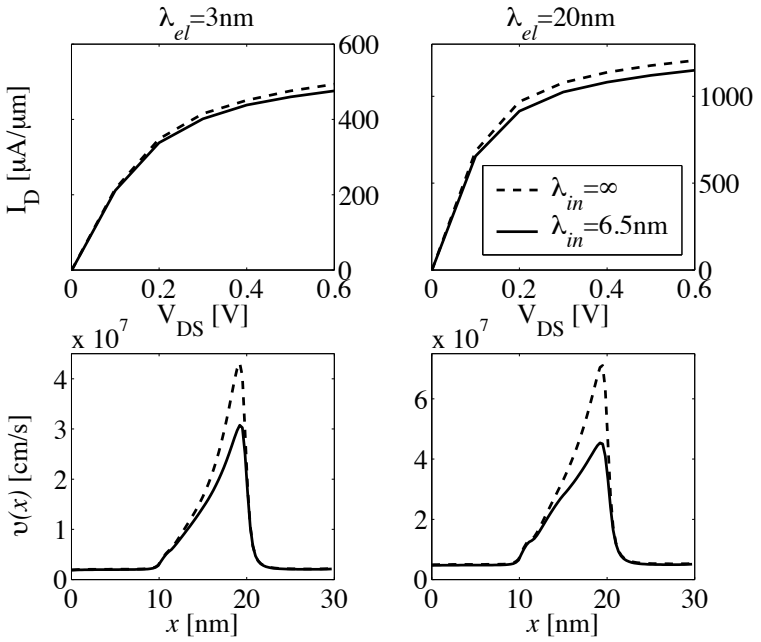

Fig. 5.5. Effect of energy relaxation mean-free-path λ_{in} on I–V and $\upsilon(x)$ for λ_{el}=3 nm (left column) and 20 nm (right column). The dashed lines represent no energy relaxation and the solid lines are for λ_{in}=6.5 nm.

5.3.2 Comparison against the physical limits

We compare the results of our transport model against the physical limits mentioned above: 1) the maximum anisotropy limit, 2) the complete isotropy limit, and 3) the diffusive limit. The I–V and the velocity profiles are compared in Figs. 5.6 and 5.7 respectively. The diffusive limit is obtained by limiting the carrier velocity to the degenerate thermal injection velocity. Thus, it is equivalent to a drift-diffusion equation with saturation velocity higher than $1.07 \cdot 10^7$ cm/s. In the near ballistic regime (λ_{el} =100 nm), the I–V and the velocity profile of our model are very close to the maximum anisotropy limit. As scattering becomes stronger, the quasi-ballistic model gets closer to the complete isotropy limit, which is the lower limit. The difference between these limits is not very pronounced in the I–V characteristics, but the velocity profiles show clear difference, especially the considerable underestimation of velocity overshoot in the diffusive limit. Also, it is shown that for λ_{el} from 3 nm to 10

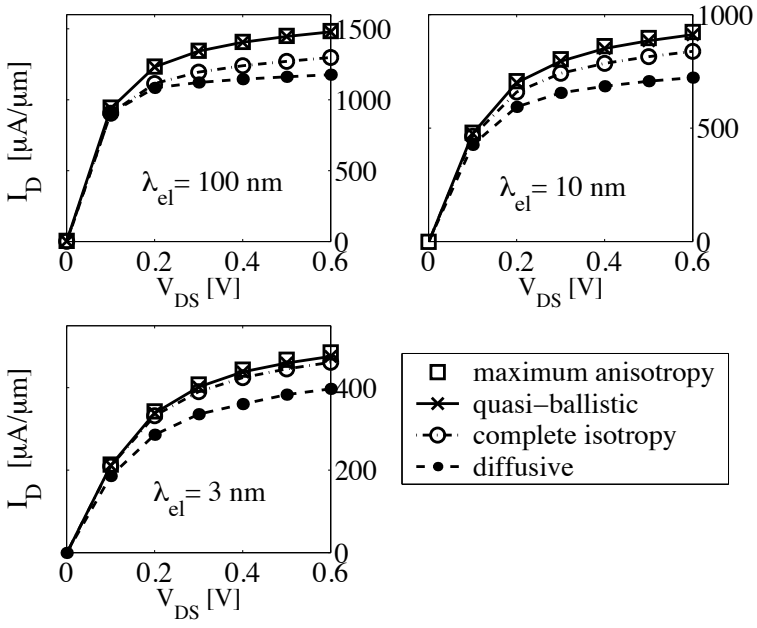

Fig. 5.6. I–V comparison of the quasi-ballistic drift-diffusion model (x) to the three physical limits: 1) the maximum anisotropy (squares), 2) the complete isotropy (circles), and 3) the diffusive limit (dots) for different λ_{el}.

nm, which is the quasi-ballistic regime, the range between the maximum anisotropy limit and the complete isotropy limit is narrow enough to assess the validity of transport models.

5.3.3 Comparison to the scattering theory

The ratio of the injection velocity to the thermal velocity of the positive source stream under high drain bias is a measure of how close transport is to the ballistic limit [23]. This ratio is related to the channel transmission **T** as [23]

$$\frac{\upsilon_{inj}}{\tilde{\upsilon}_T^+} = \frac{\mathrm{T}}{2-\mathrm{T}}, \tag{5.32}$$

where

$$\mathrm{T} = \frac{\lambda_{el}}{\ell + \lambda_{el}}. \tag{5.33}$$

The length ℓ denotes the length of critical region where the potential energy drops by initial average carrier kinetic energy at the top of the barrier, which is approximately

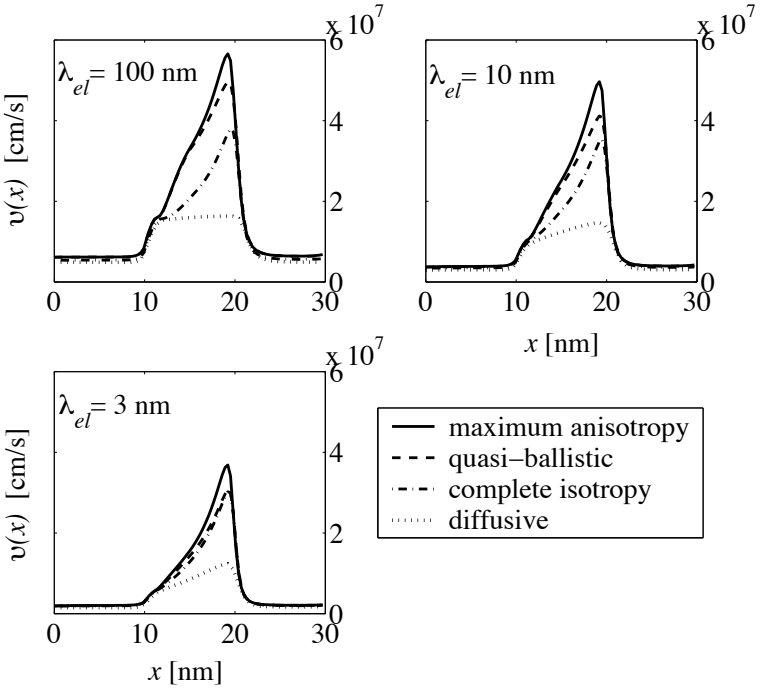

Fig. 5.7. $v(x)$ comparison of the quasi-ballistic drift-diffusion model (dashed) to the three physical limits: 1) the maximum anisotropy (solid), 2) the complete isotropy (dash-dotted), and 3) the diffusive limit (dotted) for different λ_{el}.

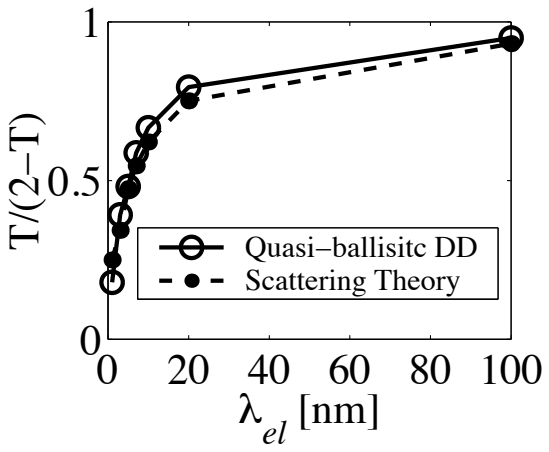

Fig. 5.8. Comparison of $\upsilon_{inj}/\tilde{\upsilon}_T^+$ vs. λ_{el} extracted from the quasi-ballistic drift-diffusion model (circles) to that from the scattering theory (dots)

the difference between the Fermi-energy and the subband energy [22]. This energy is extracted from the subband energy profiles. In this region we can let the mean-free-path be our elastic scattering mean-free-path λ_{el}. Figure 5.8 compares the ratio obtained directly from the simulation of our transport model to that obtained using Eqs. 5.32 and 5.33. Reasonable agreement shows that our model is consistent with the scattering theory.

5.4 Discussion

The shrinking of the device feature sizes has increased the importance of the contacts and the injection point (e.g., the source-channel barrier), which used to be obscured by heavy scattering. The former was discovered in mesoscopic physics [8] and the latter in the scattering theory [21]. These two aspects, however, are not separate but share the same insight that current is the transmission of a carrier flow [23]. Thus, McKelvey's one-flux method [27] [28] became a better choice for transport in nanoscale transistors than conventional approaches (e.g., drift-diffusion and energy transport).

We extended the one-flux method to accommodate off-equilibrium high field transport by: 1) including the two-population nature of transport, and 2) adopting the idea of superimposing delta functions in hierarchy closure approximations as in Baraff's maximum anisotropy principle. Note that our model consists only of the zeroth moment equations of the BTE [49] [53], and that the hierarchy closure model depends on specific scattering mechanisms. Considering the highly distorted asymmetric distribution in quasi-ballistic transport, it is not the addition of higher moments that would improve accuracy of describing high-field quasi-ballistic transport but an appropriate closure model that describes how scattering redistributes momentum and energy of ballistic carriers. The results obtained for our model device demonstrate that our approach is able to describe transport from the diffusive to the ballistic limit.

In our approach, injection velocity calculation is based on the theory of 1D ballistic conductors [56] [8], in which we know how to populate the states because carriers follow only one path from source to drain in 1D. In 2D, although we can derive a set of 2D flux equations, we do not know how carriers fill the states unless we follow every carrier trajectory in 2D as illustrated in Fig. 5.9, that is, 2D injection velocity cannot be obtained using a macroscopic approach. The possibility of this limitation is very briefly stated in [24] without explanation. Thus, our model is limited to 1D transport, which is not applicable to a 2D structure such as bulk MOSFETs. This implies that in order to do device simulation using 2D macroscopic transport models, we cannot help but continue to use conventional models such as the DD and the ET/HD models. But, users should be aware of their limitations. In nanoscale structures such as carbon nanotube FETs, silicon nanowires or FinFETs, and ultra thin body SOI MOSFETs, transport in their internal part is essentially a 1D problem. Thus, we expect our model to be useful if the associated phonon structures become available and their interaction with strongly quantum-confined electrons can be modeled macroscopically. Although BJTs are 1D devices, we will explain in the next chapter that our approach is not applicable to them because long collector regions involve high-field scattering-dominant transport such as velocity saturation.

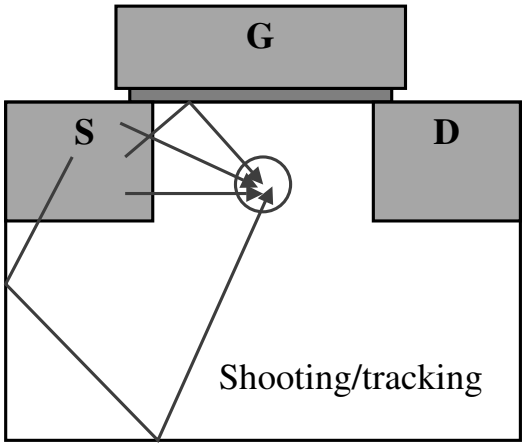

Fig. 5.9. Illustration on why the macroscopic flux model cannot be extended into 2D. The circle represents a spatial location in a bulk MOSFET and the arrows show how carriers have traveled to populate the states at that location.

5.5 Summary

In this chapter, we develop a drift-diffusion type 1D macroscopic transport model that can describe quasi-ballistic transport in a nanoscale MOSFET based on the physical understanding we have established in the previous chapters. We also establish the maximum anisotropy and the complete isotropy limits that can assess the validity of various transport models. The results obtained for our model nanoscale MOSFET fall between the two limits, showing that the quasi-ballistic drift-diffusion model can describe quasi-ballistic transport within a tolerable amount of error. Although the model is limited to 1D transport, it has some potential to be used for simulating nanoscale FETs.

6. AN EXAMINATION OF MOMENT-BASED MACROSCOPIC TRANSPORT MODELS FOR THE SIMULATION OF HIGH-SPEED SILICON-BASED BIPOLAR TRANSISTORS

In this chapter, we examine the applicability of macroscopic transport models to the simulation of high-speed Si-based bipolar transistors, where nonequilibrium, quasi-ballistic transport dominates. We first inspect their fundamental limitations in the case of a diffusive thin base followed by high-field collector, and then in the case of a graded base with a collector by benchmarking them against Monte Carlo simulations. We show that in the diffusive thin base case, moment-based macroscopic models fail to describe base and collector transport simultaneously because of abnormal diffusion behavior triggered by collector velocity overshoot. In the graded base case, however, we find that the unreasonable base diffusion is relaxed due to the drift component, and that advanced macroscopic models such as energy transport can describe the nonequilibrium, quasi-ballistic transport approximately well. Moreover, we confirm that the fundamental limitations do not matter much for calculating transit time or cut-off frequency in practical graded-base SiGe BJTs thanks to collector being the bottleneck of transport as in [14]. Finally, we discuss the essence of the issues on macroscopic modeling of nonequilibrium quasi-ballistic transport.

6.1 Introduction

As the device technology progresses, carrier transport in bipolar transistors (BJTs) has become nonequilibrium and quasi-ballistic [57] [14]. In this regime, moment-based macroscopic transport models such as the drift-diffusion (DD) and the energy

transport/hydrodynamic models (ET/HD) are expected to fail in principle [39] [47]. In practice, however, successful examples of moment-based models were reported in describing quasi-ballistic transport in BJTs [13] [14] [58] [59]. There are two branches of research regarding the matter: 1) transport across a diffusive thin base [50] [58] [60] [61] and 2) transport across a graded base followed by a high-field collector region [13] [14] [59]. In the low-field thin base case, the discussion is mainly about the DD model or the one-flux model, whereas the latter discusses the ET/HD models as well. In the thin base study, the following were recognized: 1) that the DD model deviates from the Monte Carlo (MC) simulation as the base width reduces [60] [58], and 2) that by modeling carrier velocity to capture the exit velocity discussed in [60] [61], the DD model can match the MC simulation when the collector transport is replaced by a proper boundary condition [58]. In the graded base study, it was found that ET/HD model can successfully describe quasi-ballistic transport in both base and collector region if the transport parameters are carefully calibrated to MC simulation [13] [14] [59].

In this chapter, we clarify the apparent contradiction between the principle and the practice by investigating transport in an idealized device, in which we can eliminate unnecessary complexity but keep the essence of the problem. This chapter is organized as follows. In Section 6.2, we review the physics of quasi-ballistic transport and briefly discuss the issues in the derivation of moment-based transport equations. In Section 6.3, we benchmark macroscopic transport models against MC simulations. In Section 6.4, we address issues on whether a moment-based macroscopic model could describe quasi-ballistic transport in general. Finally, we conclude in Section 6.5.

6.2 Quasi-Ballistic Transport and Moment-based Macroscopic Models

6.2.1 Quasi-ballistic transport in BJTs

Radio frequency (RF) and analog operations are the main application of high-speed BJTs, where the device figure of merit is carrier transit time directly determined

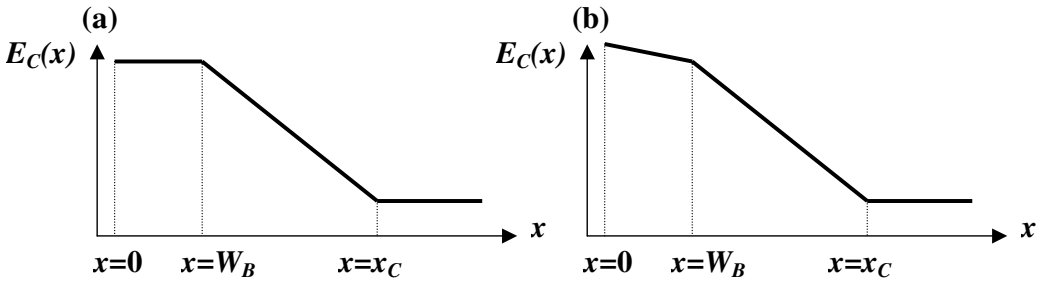

Fig. 6.1. Model BJT conduction band profiles: device (a) consists of a diffusive base and a high-field collector, and device (b) of a graded base with a field 10^4 V/cm and a high-field collector. The base is doped at $N_A = 5 \cdot 10^{18}$ cm^{-3}, the collector is at $N_D = 10^{16}$ cm^{-3}, and the subcollector at $N_D = 10^{19}$ cm^{-3} in both cases. The collector field is 10^5 V/cm unless specified.

by average velocity of minority carriers, which undergo nonequilibrium, quasi-ballistic transport [57]. Thus, we focus on the velocity profiles resulting from quasi-ballistic transport in the base and collector using two idealized device structures shown in Fig. 6.1 , which represent the base (one is diffusive and the other graded), the collector, and the sub-collector region of an *npn* silicon BJT. In the idealized structure we assume the following simplification. First, a piecewise constant frozen field (non-self-consistent) is assumed because: 1) the conduction band profile is close to a realistic profile in the base and in the beginning of the collector, and 2) the device operates under low-current so that the electrostatics is decoupled from electron transport in the base and the base-collector junction. Second, we can ignore the multiple valley transport effects in the collector because of the large separation of satellite conduction valleys from the main valley in silicon. Despite a deviation in the deep collector, it is sufficient to capture the essence of quasi-ballistic transport and to study the fundamental behavior of various macroscopic transport models.

We review the quasi-ballistic transport in the diffusive thin base followed by a high-field collector by examining the distributions obtained from MC simulation [3] (solid lines in Fig. 6.2).

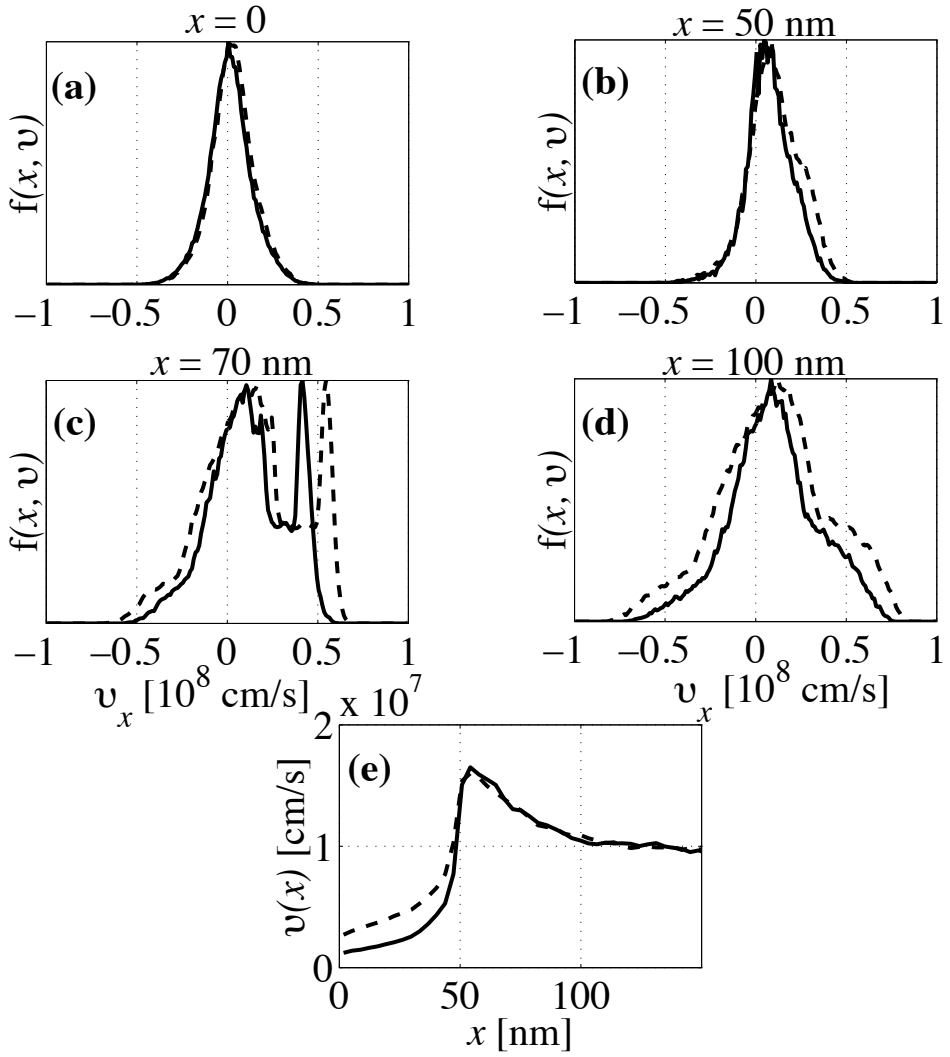

Fig. 6.2. Carrier distributions obtained from MC simulation [3] for a diffusive base (solid) and for a graded base (dashed) with the base width $W_B = 50$ nm at: (a) $x = 0$ (beginning of the base), (b) $x = 50$ nm (end of the base), (c) $x = 70$ nm (overshoot region), and (d) $x = 100$ nm (velocity saturation region). The average carrier velocity is shown in (e).

At the beginning of the base (Fig. 6.2a), the positive portion of the velocity distribution is a thermal equilibrium distribution injected from the left contact, and the negative portion is from backscattering because the injection from the right contact is suppressed due to the barrier. As carriers transport through the base, the negative portion reduces and the positive portion grows because the high-field collector acts as an absorbing contact. Note, however, that the positive portion is shifted toward higher velocity (Fig. 6.2b) as scattering in the base favors carriers with higher velocity in the transport direction [61]. Thus, the exit velocity at the base end becomes about 1.15 times the thermionic emission velocity $\upsilon_T = \sqrt{2k_BT_L/\pi m^*}(= 1.06 \cdot 10^7$ cm/s) although it reduces to υ_T if the base width W_B is less than 5 nm [61].

Once carriers exit the base, they start to get accelerated by the high field in the beginning of the collector region. In this region, the carriers can gain kinetic energy in the direction of the field within a mean free path of scattering, i.e., before they undergo heavy scattering events. Consequently, the distribution displays a prominent ballistic peak (Fig. 6.2c), which causes velocity to overshoot as shown in Fig. 6.2e. Note that the maximum overshoot is slightly away from the base-collector junction or the location of low-to-high field transition where the ballistic peak just starts developing.

As the carrier kinetic energy further increases down the collector beyond the onset of optical phonon emission, heavy scattering prompts the carriers in the ballistic peak to scatter out into transverse directions. As a result, the ballistic peak disappears but the distribution gets broadened as shown in Fig. 6.2d. Interestingly, under a field higher than a certain critical value, the average velocity of the broadened distribution remains the same although the width, i.e., average energy, increases with the field. This is known as velocity saturation. In the graded base case, the only difference is that carrier distribution is shifted more towards the right and more broadened than in the diffusive base. This causes the base velocity to exceed the velocity produced in the diffusive base case for the same base width.

6.2.2 Moment-based macroscopic models

Since the detailed derivation of macroscopic models is widely available [44], we just remind readers of a few general features associated with the derivation. The moment-based macroscopic transport models are derived by averaging the Boltzmann equation over carrier momentum weighted with polynomials of carrier momentum. The derivation inevitably involves two key approximations: 1) the hierarchy closure approximation that terminates the infinite chain of moment equations at lower order moments, and 2) the scattering approximation that expresses the effect of scattering in terms of measurable transport parameters such as the low-field mobility, and variables in the moment equations.

In principle, high order effects such as velocity overshoot associated with the development of ballistic peaks should be taken care of in the closure approximation [53]. However, in practice, the closure approximation assumes a carrier distribution that can be fitted with lower order polynomials [39]. The scattering approximation in the current equation, which is the most important since it determines the current or velocity, casts most of the transport effects into a mobility model built to describe high-field collision-dominated transport. Hence, the moment-based macroscopic models are theoretically incapable of describing quasi-ballistic transport. Speaking in another perspective, to reproduce the velocity profile of the MC simulation in the base and collector region, a macroscopic transport model should capture 1) the thermionic emission velocity in the diffusive base, 2) the overshoot velocity due to the ballistic peak in the beginning of the collector, and 3) saturation velocity in the deep collector region, not to mention the smooth transition in between.

Note that these velocities and the smooth transition are the results of not only the random motion due to scattering but also deterministic motion between scattering events due to external forces. However, the latter is missing in macroscopic moment equations because of the assumptions of collision-dominated transport. Nonetheless, there are examples of successful application of moment-based macroscopic models to

quasi-ballistic transport in high-speed BJTs [14] [13] [59]. In the next section, we show that the above features of the moment-based equations lead to an unavoidable limitations. Then, we resolve this apparent contradiction between the principle and the practice.

6.3 Benchmarking Macroscopic Models against Monte Carlo Simulation

6.3.1 Macroscopic models in test

We benchmark the following macroscopic models against MC simulation: 1) the drift-diffusion, 2) the one-flux, and 3) the energy transport models.

The drift-diffusion model

The drift-diffusion model (DD) for 1D transport in x-direction is given as

$$J_n = q\mu_n n\mathscr{E}_x + qD_n\frac{dn}{dx}, \tag{6.1}$$

and the field dependent mobility is

$$\mu_n\left(\mathscr{E}_x\right) = \mu_0\left(x\right)\Big/\sqrt{1+\left(\frac{\mu_0\left(x\right)\mathscr{E}_x}{\upsilon_{sat}}\right)^2}, \tag{6.2}$$

which is related to the diffusion coefficient through the Einstein relation

$$\frac{D_n}{\mu_n} = \frac{k_B T_L}{q}, \tag{6.3}$$

where T_L is lattice temperature. Note that the *field-independent* low-field mobility $\mu_0(x)$ is position-dependent, which is determined by doping concentration and the type of carriers at x as proposed by Klaassen, et al. [62], and the saturation velocity is υ_{sat}= 1.07·10^7 cm/s.

The one-flux model

The equations of the one-flux model are [27] [28] [53]

$$\frac{dJ^+}{dx} = -\xi J^+ + \xi' J^- \tag{6.4}$$

$$\frac{dJ^-}{dx} = -\xi J^+ + \xi' J^-, \tag{6.5}$$

where the backscattering probabilities per lengths are

$$\xi = \begin{cases} \xi_0 + \frac{q\mathscr{E}_x}{k_B T_L} & \mathscr{E}_x > 0 \\ \xi_0 & \mathscr{E}_x < 0 \end{cases}, \tag{6.6}$$

and

$$\xi' = \begin{cases} \xi_0 & \mathscr{E}_x > 0 \\ \xi_0 - \frac{q\mathscr{E}_x}{k_B T_L} & \mathscr{E}_x < 0 \end{cases}. \tag{6.7}$$

The field-independent backscattering probability per length ξ_0 is associated with the low-field mobility $\mu_0(x)$ through Shockley-Einstein relation [21] [28], that is

$$D_0 = \frac{\upsilon_T}{2\xi_0} = \frac{k_B T_L}{q}\mu_0. \tag{6.8}$$

The net current and carrier density are

$$J_n = -q\left(J^+ - J^-\right), \tag{6.9}$$

and

$$n = \left(J^+ + J^-\right)\Big/\upsilon_T. \tag{6.10}$$

The one-flux model can be converted to a form of drift-diffusion equation [53], which is

$$J_n = q\mu^* n\mathscr{E}_x + \mu^* k_B T_L \frac{dn}{dx}, \tag{6.11}$$

where the equivalent mobility μ^* is

$$\mu^* = \frac{\upsilon_T}{\upsilon_T/\mu_0 + |\mathscr{E}_x|}. \tag{6.12}$$

The energy transport model

The energy transport model (ET) used in this chapter adopts the current equation of Blotekjaer's approach [63] [64]

$$J_n = q\mu_n\left[n\mathscr{E}_x + \frac{k_B T_n}{q}\frac{dn}{dx} + f_n^{td} n\frac{d}{dx}\left(\frac{k_B T_n}{q}\right)\right], \tag{6.13}$$

where the temperature dependent mobility is given as [65]

$$\mu_n(T_n) = \frac{\mu_0(x)}{\left[1 + \left(\mu_0 \frac{3k_B(T_n - T_L)}{2q^2 \tau_{en} v_{sat}^2}\right)^{\beta_n}\right]^{1/\beta_n}}, \tag{6.14}$$

and the energy equations are

$$\frac{dS_n}{dx} = J_n \mathscr{E}_x - \frac{3}{2} n k_B \frac{T_n - T_L}{\tau_{en}} \tag{6.15}$$

$$S_n = -\frac{5r_n}{2}\left(\frac{k_B T_n}{q} J_n + f_n^{hf} \hat{\kappa}_n \frac{dT_n}{dx}\right). \tag{6.16}$$

where the thermal conductivity is

$$\hat{\kappa}_n = \frac{k_B^2}{q} n \mu_n T_n. \tag{6.17}$$

The nomenclature for the new quantities and parameters is as follows.

T_n is electron temperature

S_n is electron energy flux density

f_n^{td} is a thermal diffusion factor ($0 \leq f_n^{td} \leq 1$)

τ_{en} is electron energy relaxation time

β_n is the exponent in the electron mobility

f_n^{hf} is a heat flux factor ($0 \leq f_n^{hf} \leq 1$) This model is one of many energy balance or hydrodynamic models. As it will be shown, however, we can identify the fundamental behavior of the ET model and consequently the fundamental limitations, which are common to all of the kind.

6.3.2 Benchmark for a diffusive thin base followed by a high-field collector

We benchmark the aforementioned transport models against MC simulation for the diffusive thin base followed by a high-field collector.

The drift-diffusion models

We benchmark two DD models: 1) one (denoted as DD1) adopts the field-dependent mobility model in Eq. 6.2, which is widely used in device simulation,

and 2) the other (denoted as DD2) lets the mobility μ_n be the field-independent low-field mobility $\mu_0(x)$ to clarify the diffusion problem in a thin base. The calibration is done simply by using the electron minority carrier mobility in the p-type base (∼240 $cm^2/V{\cdot}s$ for $N_A = 5 \cdot 10^{18}$ cm^{-3}), and the majority carrier mobility elsewhere (∼1155 $cm^2/V{\cdot}s$ for $N_D = 10^{16}$ cm^{-3} in the collector, and ∼100 $cm^2/V{\cdot}s$ at $N_D = 10^{19}$ cm^{-3} in the subcollector) according to [62].

The velocity profiles of the drift-diffusion models are shown in Fig 6.3a. The solid line shows the velocity of DD1 and the dash-dotted line that of DD2. The solid line matches the MC result in the base while the dashed line overestimates the velocity towards the base end, although their diffusion coefficients are the same due to a negligible electric field in the base. The mismatch between DD1 and DD2 arises from the difference in the exit velocity v_{ext} , the average velocity at the base end.

Figures 6.3b and 6.3c show the diffusion and drift components of the two models providing the idea of what determines v_{ext}. In either case, the diffusion current is dominant in the base while the drift current is prevailing in the high field collector. Note that the drift velocity in the high-field collector is imposed at the right end of the base as if it were a boundary condition for a diffusion equation in the thin base with the collector replaced by an absorbing contact [61] [50]. Thus, despite the same diffusion coefficients, the velocity profiles of the two DD models are different because the effective boundary conditions are dissimilar. In the high field collector, v_{ext} of DD1 is limited at the saturation velocity $v_{sat} = 1.07 \cdot 10^7$ cm/s for Si, which is close to the physical exit velocity 1.15 v_T. Hence, the base velocity profile can be close to the MC result [60]. However, for DD2, $v_{ext} = \mu_0 \cdot \mathscr{E}_x \sim 10^8$ cm/s (μ_0=1155 $cm^2/V{\cdot}s$ and $\mathscr{E}_x = 10^5$ V/cm in the collector), which is about nine times higher than 1.15 v_T. Consequently, it generates a steeper gradient of carrier density and overestimates the velocity in the base.

Figure 6.4 summarizes the behavior of the DD models for different W_B. It shows that thanks to $v_{sat} \sim 1.15\, v_T$ in Si, DD1 behaves as good as MC simulation in the calculation of the collector current (Fig. 6.4a) and the base transit time τ_B (Fig. 6.4b).

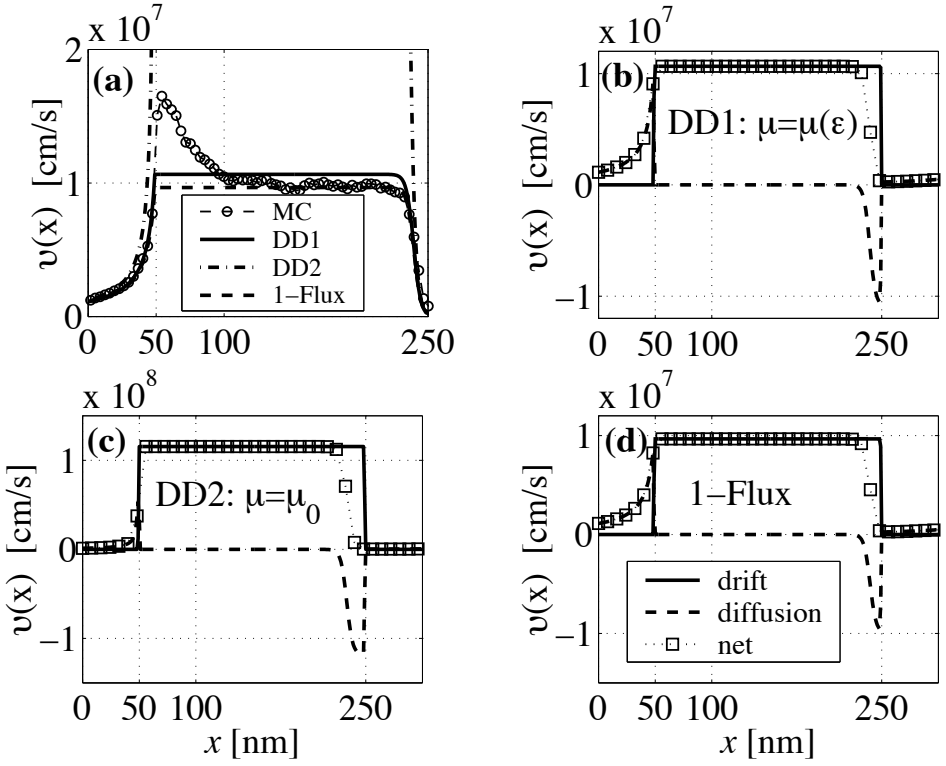

Fig. 6.3. (a) $v(x)$ of DD1 (solid), DD2 (dash-dotted), and one-flux models (dashed) is compared to MC simulation (circles). Velocity components of DD1 is shown in (b), of DD2 in (c), and of one-flux in (d); the solid lines are the drift component, the dashed are the diffusion, and the squares are the net velocity.

However, if $\upsilon_{sat} < 0.5\upsilon_T$ as in GaAs, DD1 is expected to fail. The DD2 model displays the classical diffusion behavior exceeding the ballistic limit showing that its current is inversely proportional to W_B (Fig. 6.4a) and τ_B is proportional to ${W_B}^2$ (Fig. 6.4b). The unreasonable exit velocity of DD2 causes this discrepancy as shown in Fig. 6.4c. Karmarkovic, et al. showed that the above shortcomings of the DD2 model could be overcome; with the collector replaced by a perfect sink, their modified DD model matched the MC simulation [58]. They imposed a boundary condition at the end of the diffusive base as

$$J_n\left(W_B\right) = qn\left(W_B\right)\upsilon_{ext} = qn\left(W_B\right) \cdot 1.15\,\upsilon_T, \tag{6.18}$$

and obtained

$$\upsilon\left(x\right) = 1.15\,\upsilon_T \Big/ \left[1 + \frac{3}{4}\frac{W_B}{\lambda}\left(1 - \frac{x}{W_B}\right)\right] \tag{6.19}$$

for $0 < x < W_B$, where λ is a mean-free-path obtained from $D_n = \frac{4}{3}\upsilon_T\lambda$. But, note that this modification is valid only when the current is exclusively diffusion current.

The one-flux model

To calibrate the one-flux model, we simply extract the field-independent backscattering probability ξ_0 from Eq. 6.8. Figure 6.3 shows the velocity profile of the one-flux model, which has been used to study the diffusive thin base transport with a collector replaced by an absorbing contact [50]. As reported in [50], it displays an excellent match in the diffusive base thanks to the flux approach and the notion of the thermionic emission velocity υ_T in the model. As shown in Fig. 6.3d, the equivalent drift velocity obtained from the field term in Eq. 6.11 imposes υ_T as an exit velocity on the diffusive base. It matches the saturation velocity in the deep collector region, but again it is because υ_T is remarkably close to υ_{sat} in a bulk Si at room temperature. Note that υ_{sat} is not modeled in the one-flux model. Thus, in case of GaAs where $\upsilon_{sat} < 0.5\upsilon_T$, the one-flux model expects to either overestimate the base transit time or underestimate the collector transit time depending on which velocity to choose. Moreover, it cannot describe velocity overshoot in the beginning of the collector. For

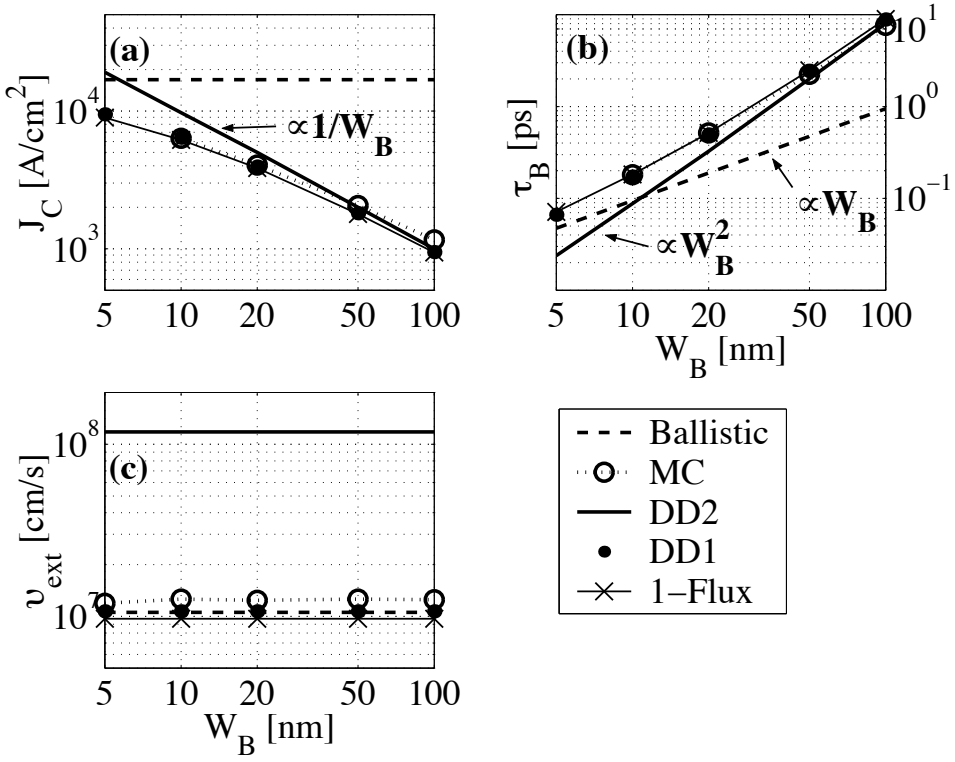

Fig. 6.4. (a) Collector current vs. W_B, (b) τ_B vs. W_B, and (c) υ_{ext} vs. W_B of DD1 (dots), DD2 (solid), and one-flux (-x-) are compared to MC simulation (circles) and the ballistic limit (dashed).

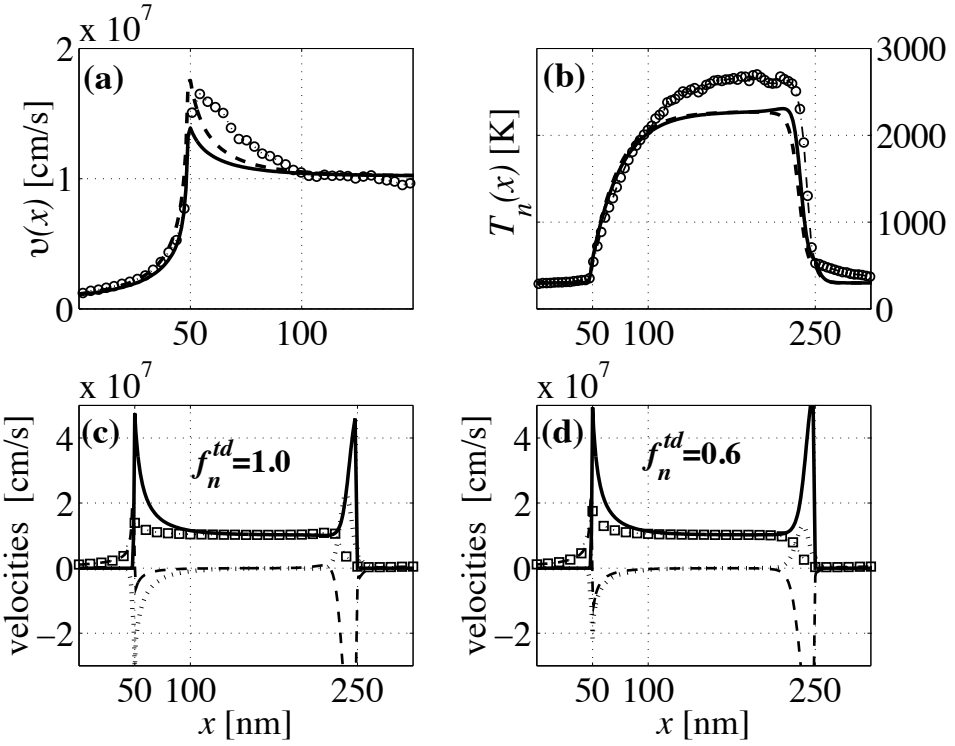

Fig. 6.5. (a) $\upsilon(x)$ and (b) $T_n(x)$ of ET with f_n^{td}=1.0 (solid) and 0.6 (dashed) are compared to MC simulation (circle). (c) Velocity components of ET with f_n^{td}=1.0 and (d) with f_n^{td}=0.6 are shown; the drift components are in solid lines, the diffusion in dashed lines, the thermal diffusion in dotted lines, and the net velocities are in squares.

different W_B, the collector current and the base transit time of the one-flux model are also compared to MC simulation in Figs. 6.4a and 6.4b, respectively.

The energy transport model

It is very difficult to assess the transport models known as energy transport models not only because there are numerous versions of the kind but also because they have a set of parameters that need to be carefully optimized. However, we identify the fundamental behavior common to all ET models using the model presented in Eqs. 6.13 $\sim$ 6.17, and consequently recognize their unavoidable limitations.

Figures 6.5a and 6.5b show the comparison of average velocity and electron energy in absolute temperature between the ET model and MC simulation. We first calibrate the parameters in the energy equations, Eqs. 6.15 and 6.16, so that it can match the energy profile of MC simulation, especially in the base and in the beginning of the collector where the slope of the carrier temperature has a sizable effect on the velocity through the thermal diffusion term in Eq. 6.13. Although the fitting is not perfectly done, the fundamental behavior of the ET model can be examined as explained later.

Before we move on, we would like to mention an abnormal behavior of the ET model that is not widely appreciated as unreasonable, and sometimes misunderstood as a thermodynamic phenomenon when it shows up in device simulation. That is the abnormal temperature rise in the diffusive thin base, which is usually attributed to thermal back diffusion. It is often argued that when carriers are heated in the high-field collector, they can diffuse back into the low-field base due to the temperature gradient and heat the base up. However, in reality, when carriers are scattered back into the base from the collector, they get decelerated and give up at least the same amount of energy to the collector-field as they have gained originally from it. Therefore, the argument that thermal back diffusion from the high-field collector can heat up the low-field base violates the very first law of thermodynamics. In this sense, the term thermal back diffusion does not reflect physics correctly. We choose parameters in the energy equations so that the abnormal temperature rise in the base is prevented as shown in Fig. 6.5b, but in practice, we often see this kind of abnormal behavior.

For the given set of parameters in the energy equations, the current equation, Eq. 6.13, is calibrated using the field-independent low-field mobility $\mu_0(x)$ as in the DD models and with two different thermal back diffusion parameters, one for f_n^{td}=1.0 and the other for 0.6. As shown in Fig. 6.5a, both cases present velocity overshoot *at* the beginning of the collector region ($x = W_B$) and captures velocity saturation in the deep collector. For f_n^{td}=1.0, the base velocity profile in Fig. 6.5a is in good agreement with MC simulation but overshoot in the collector is largely underestimated. For f_n^{td}=0.6, maximum overshoot is comparable to that of MC simulation, but the location

of the peak occurs at $x = W_B$, slightly off the peak of the MC result. Consequently, the exit velocity at $x = W_B$ exceeds the physical limit and the base velocity is overestimated although it is expected to produce a collector transit time closer to that of MC simulation than for f_n^{td}=1.0. The reason for this behavior is that the maximum overshoot velocity of the ET model is always placed at the location of low-to-high field transition while that of MC is shifted toward the collector region as explained in section 6.2.

Figures 6.5c and 6.5d show the velocity components of the ET model, one for f_n^{td}=1.0 and the other for f_n^{td}=0.6 respectively. The velocity is obtained from Eq. 6.13 as

$$\upsilon(x) = \frac{J_n}{-qn} = -\mu_n(T_n)\left[\mathscr{E}_x + \frac{k_B T_n}{q}\frac{d\ln n}{dx} + f_n^{td}\frac{d}{dx}\left(\frac{k_B T_n}{q}\right)\right]. \tag{6.20}$$

The first term is the drift velocity υ_{drft}, the second is the diffusion velocity υ_{diff}, and the third is the thermal diffusion velocity υ_{Tdiff}, respectively. In the base ($0 < x < W_B$), it is exclusively diffusive because $\mathscr{E}_x \cong 0$ and $T_n \cong T_L$. Thus, the ET model behaves in the similar way as the DD except the difference in the exit velocity. From Figs. 6.5c and 6.5d, we can see that υ_{ext} is mainly determined by υ_{drft} and υ_{Tdiff} because the contribution from υ_{diff} is not large enough; the logarithmic derivative of carrier density is relatively small at $x = W_B$. In the high-field collector region, υ_{drft} always shows two positive peaks, one at $x = W_B$ and the other at the end of the collector, $x = x_c$. In between, we get $\upsilon_{drft} \cong \upsilon_{sat}$ as carriers are in high energy. The overall shape of υ_{drft} can be understood from the temperature profile and the mobility in the collector region as shown in Fig. 6.6. At $x = W_B$, T_n starts rising but is close to T_L, thus $\mu_n(T_n) \cong \mu_0$, which is the maximum value. Since $\upsilon_{drft} = -\mu_n \cdot \mathscr{E}_x$, it takes its maximum at $x = W_B$. We can understand the peak at $x = x_c$ accordingly. Notice that in a realistic self-consistent energy band profile, the collector field is decreasing toward the collector end, thus the drift velocity will take its maximum right at the low-to-high field transition.

The thermal diffusion velocity υ_{Tdiff} takes a negative peak at $x = W_B$ because T_n starts rising quickly, but has a positive peak near $x = x_c$ because it starts falling.

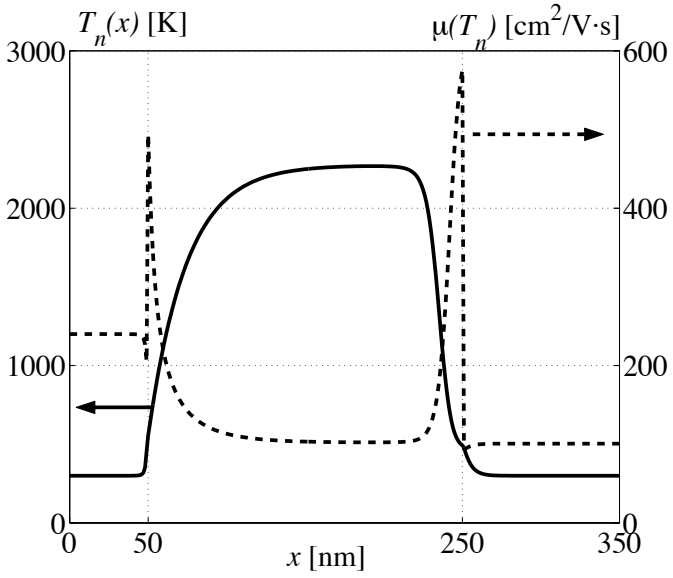

Fig. 6.6. Behavior of temperature dependent mobility (dashed) of ET model is shown with the temperature profile (solid)

Hence, the thermal diffusion component counteracts the drift velocity at $x = W_B$ but adds to it at $x = x_c$. The thermal diffusion coefficient f_n^{td} directly controls the height of the peaks as shown in Figs. 6.5c and 6.5d. The diffusion velocity v_{diff} displays two negative peaks at $x = W_B$ and $x = x_c$. Its counteraction against the overshoot is negligible at $x = W_B$ due to a mild logarithmic gradient of carrier density. But the negative peak at $x = x_c$ almost cancels the drift velocity because carrier density starts increasing quickly in the low-field sub-collector region.

From the behavior of each velocity component, we can see that if we optimize the model for the base transport, we underestimate the velocity overshoot in the collector, but that we overestimate the base velocity profile if we optimize it for the collector transport. Note that the behavior of these velocity components is inherent in any ET/HD model regardless of choice of parameters or mobility models since the diffusion is mainly determined by the exit velocity as in the DD models. The origin of the problem that the location of maximum overshoot velocity occurs at the end of the base is that in ET/HD models velocity overshoot is imitated by the

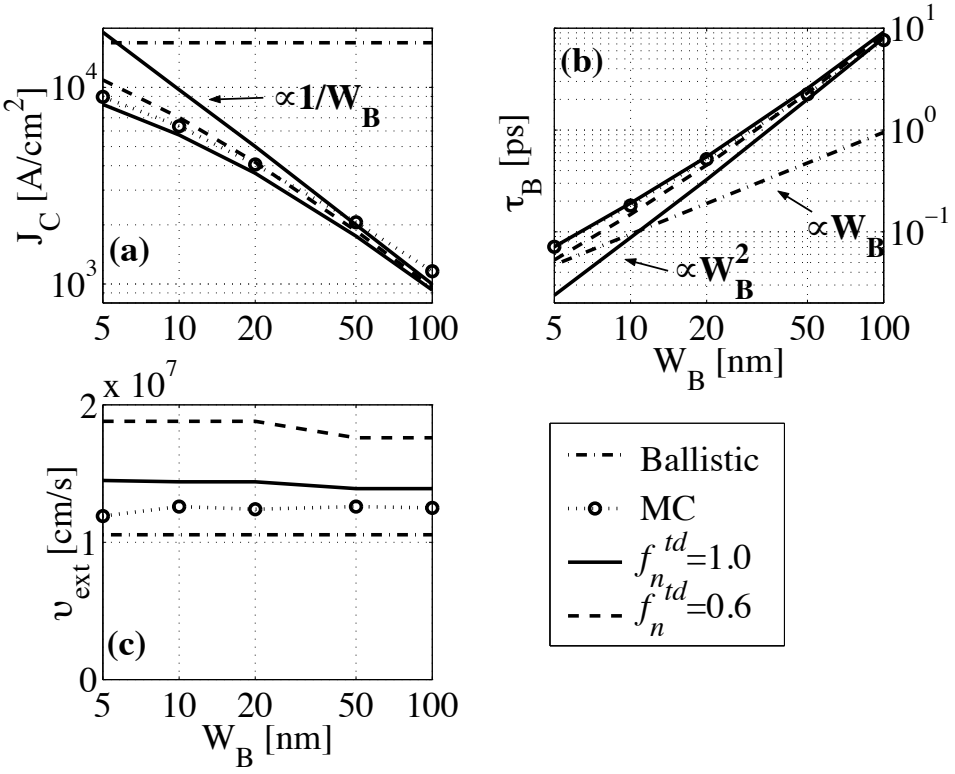

Fig. 6.7. (a) Collector current vs. W_B, (b) τ_B vs. W_B, and (c) υ_{ext} vs. W_B of ET model with f_n^{td}=1.0 (solid) and 0.6 (dashed) are compared to MC simulation (circles) and the ballistic limit (dash-dotted).

scattering approximation in the drift and the diffusion terms in Eq. 6.13 instead of being taken care of in the hierarchy closure approximation. This problem becomes more pronounced as the width of the base reduces or the collector field increases as shown in Figs. 6.7 and 6.8, respectively. In Figs. 6.7a and 6.7b, at the cost of velocity overshoot the ET model with f_n^{td}=1.0 yields reasonable collector current and base transit time for different W_B. However, if we increase the overshoot velocity reducing f_n^{td} to 0.6, the collector current and the base transit time start to deviate from the MC results as W_B decreases. The deviation will be extended to longer W_B if the minority carrier mobility in the base improves. Figure 6.8 shows how the collector field affects the base velocity profile.

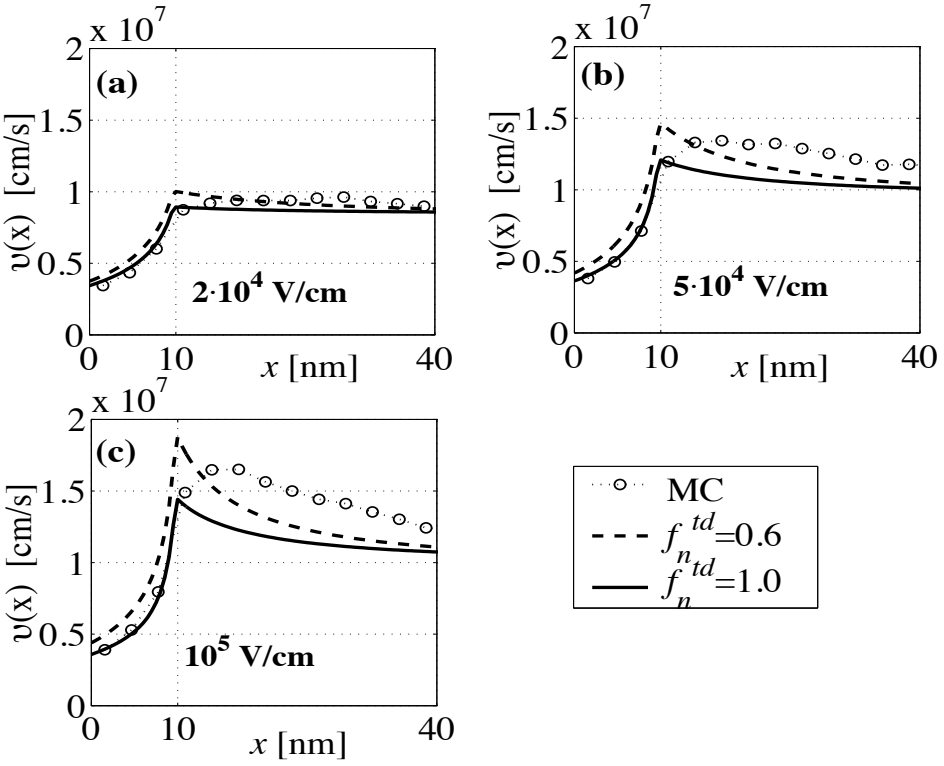

Fig. 6.8. The dependence of velocity in the base is shown for different values of collector field: (a) $2{\cdot}10^4$, (b) $5{\cdot}10^4$, and (c) 10^5 V/cm. The solid lines are for f_n^{td}=1.0 and the dashed lines for 0.6.

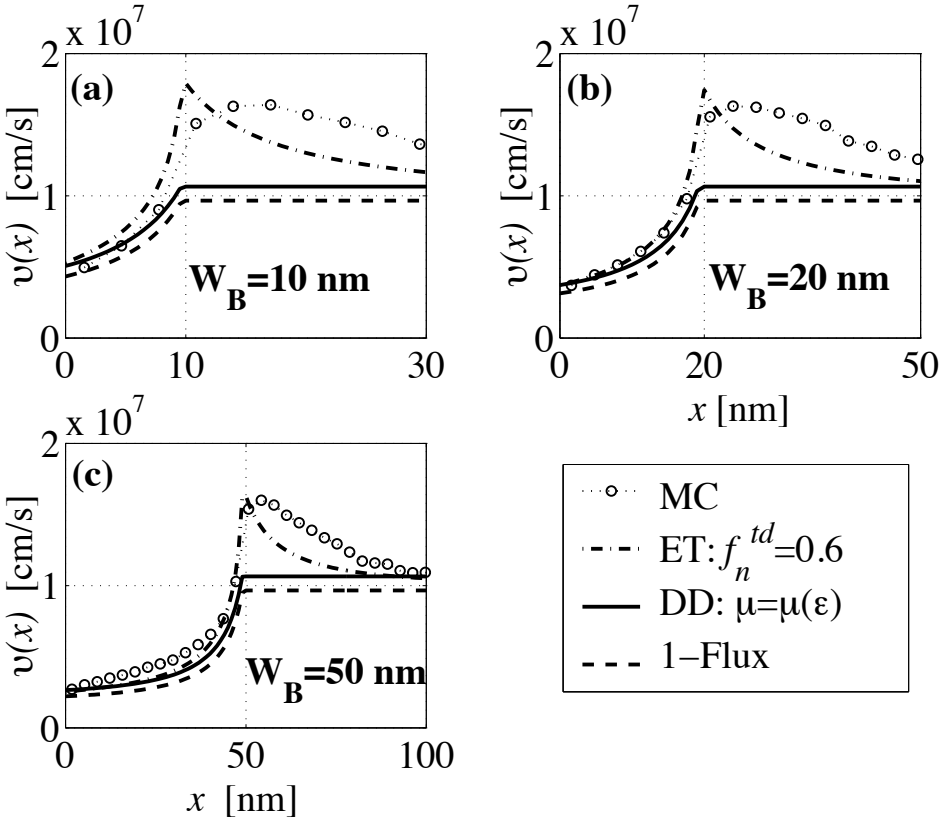

Fig. 6.9. For a base graded with a field of 10^4 V/cm and a collector with a field of 10^5 V/cm, we compare $\upsilon(x)$ of ET with f_n^{td}=0.6 (dash-dotted), DD (solid), and one-flux (dashed) are compared to MC simulation (circles) for W_B =: (a) 10 nm, (b) 20 nm, and (c) 50 nm.

6.3.3 Benchmark for a graded thin base followed by a high-field collector

The graded base case is more complicated to benchmark for a wide range of base width than in the thin base case. First, how or how much the base is graded increases the complexity of the problem. Second, as the drift motion becomes important in the base, the parameters in the mobility model in Eq. 6.14 start to play a role, increasing the complexity associated with parameter tuning. Therefore, our purpose of benchmarking for a graded base is to identify fundamental features of macroscopic moment equations instead of exploring all combinations of parameters for better fitting.

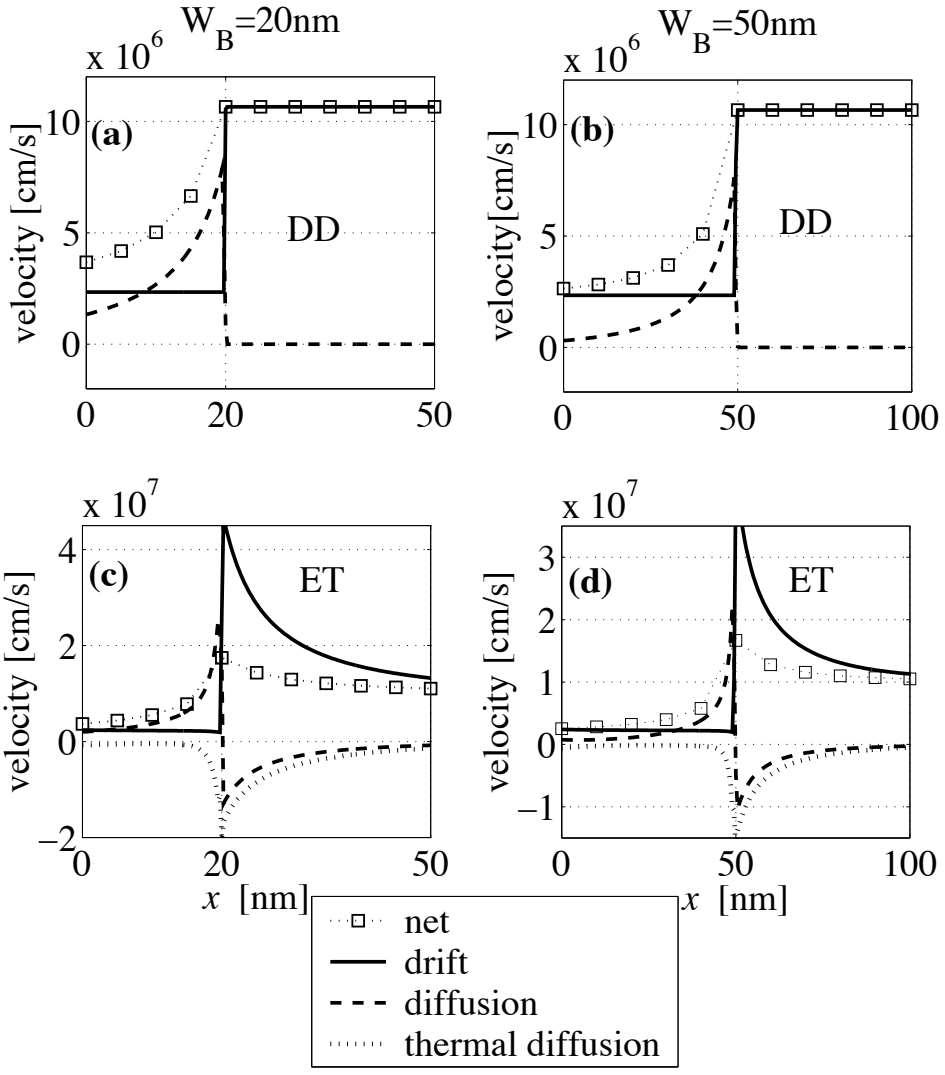

Fig. 6.10. Velocity components of DD and ET models are shown for a graded base with $W_B = 20$ nm (left column) and 50 nm (right column).

Figure 6.9 compares the velocity profiles of the drift-diffusion, the one-flux, and the energy transport models against MC simulation for the graded base device in Fig. 6.1b. For W_B = 10, 20, and 50 nm, the base electric field is fixed at 10^4 V/cm to clarify the difference between drift and diffusion. The one-flux equation obviously underestimates the base velocity in all cases. The reason is that at 10^4 V/cm, the equivalent mobility μ^* in Eq. 6.12 becomes smaller than the mobility of the DD model in Eq. 6.2 since the field term in the one-flux model represents backscattering not drift motion [53].

For W_B=10 and 20 nm, the DD model shows a good agreement in the base with MC simulation except near the end of the base because the velocity is limited at v_{sat}. However, for W_B=50 nm, disagreement is shown in the middle of the base. We can explain this by comparing the velocity components of the DD model for W_B=20 nm to those for W_B=50 nm in Figs. 6.10a and 6.10b. Note that the drift velocity is constant in the base due to the constant electric field. Thus, the increase of the velocity towards the end of the base is attributed to the diffusion velocity. For W_B=10 and 20 nm, the rise of v_{diff} happens to be close to the MC profile but for 50 nm, v_{diff} rises slower than the MC result showing mismatch in the middle of the base. Note that the rise of the velocity in the graded base is partly from the gradual shift of the distribution toward the base end as carriers get accelerated (see Fig. 6.2) but in the DD model, it is described as a pure diffusive motion.

This problem is inherited in the ET model as well since its current is also described as drift and diffusion. The velocity of the ET model is compared to the MC in Fig. 6.9, and Figs. 6.10c and 6.10d show the velocity components of the ET model for W_B=20 nm and 50 nm, respectively. Notice that even in the graded base case, the maximum overshoot velocity is located exactly at the low-to-high field transition for the same reason as in the diffusive base case. As W_B reduces, the effect of the exit velocity on the diffusion increases, and consequently the velocity in the base rises faster (see Fig. 6.9). However, since the drift component exists, it is less pronounced than in the diffusive base. Now the parameters in the mobility model also start to play a role

through the drift component, providing more flexibility for fitting. Thus, as shown in Fig. 6.9b, the base velocity of the ET model can be quite close to the MC result without sacrificing the overshoot velocity in the collector. Although we do not try in this work, we argue that the matching of collector velocity can be improved because the current in the collector is exclusively drift and consequently it can be controlled by the parameters in the mobility model.

Bearing in mind the above observation on macroscopic models, we turn our attention to the work by Jungemann, et al. for the simulation of a SiGe HBT with W_B=24 nm and base grading steeper than 10^4 V/cm [14], and clarify the apparent contradiction between the principle and the practice regarding the applicability of macroscopic models to the simulation of high-speed BJTs. In [14], they observed that the DD model underestimates the cutoff frequency of the SiGe BJT only by 10% despite the underestimation of base velocity and that the HD model is in good agreement with MC simulation. Now we confirm their observation through our model device with a graded base. The base field is given as $3 \cdot 10^5$ V/cm to emulate the steeper base grading in their device. We choose W_B=20 nm and W_C=200 nm, which are close to their values.

Figure 6.11 shows the velocity profile (left column) and the transit time at position x (right column) of the DD, the ET, and the MC in different x-axis scales. The transit time $\tau(x)$ is calculated as

$$\tau(x) = \int_0^x \frac{1}{\upsilon(x')} W(x')\, dx' \tag{6.21}$$

for $0 < x < x_c = W_B + W_C$, where the weighting function $W(x)$ is

$$W(x) = 1 \tag{6.22}$$

for $0 < x < W_B$, and

$$W(x) = 1 - \frac{x - W_B}{W_C} \tag{6.23}$$

for $W_B < x < W_B + W_C$ [66]. Thus, the base transit time $\tau_B = \tau(x = W_B)$, and, the collector transit time $\tau_C = \tau(x = W_B + W_C) - \tau(x = W_B)$.

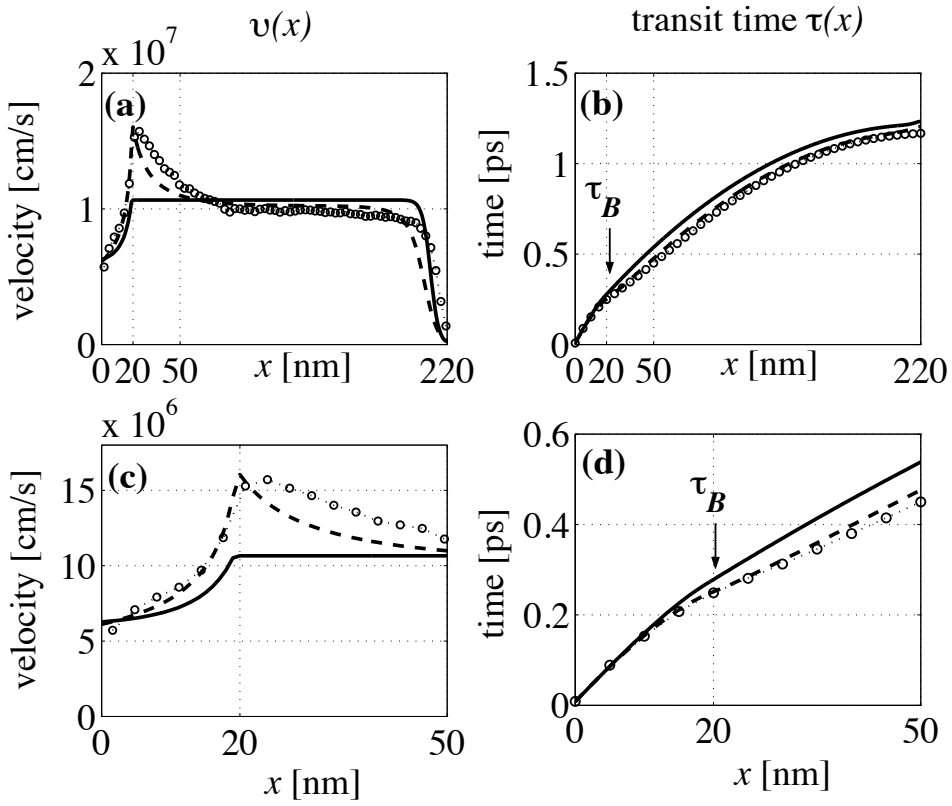

Fig. 6.11. The velocity profiles and transit times at x of DD (solid), ET (dashed), and MC (circles) are shown for different x-axis scale, obtained for a graded base structure in Fig. 6.1 with $W_B = 20$ nm, the base field of $3 \cdot 10^4$ V/cm, W_C=200 nm, and the collector field of 10^5 V/cm.

From the velocity profile, we can clearly see that the ET model matches the MC result well in the graded base because the velocity is not limited, but that the DD model underestimates average velocity in the base and in the beginning of the collector overestimating $\tau(x)$ around that region as shown in Fig. 6.11d. However, the overall transit time of the DD is within 10% of the MC because the collector transit time becomes the bottleneck because $W_C = 10W_B$, as pointed out in [14]. Also despite the underestimation of the collector overshoot of the ET model, the overall $\tau(x)$ of the ET is very close to that of MC. In addition to the reduction of abnormal diffusion thanks to the choice of a graded base, the detailed position dependent structure of average velocity and the fundamental limitations of macroscopic models are buried in the integral process in Eq. 6.21, which brings in physics other than transport such as geometric design of a device, i.e., the length of the collector in this case. This makes macroscopic models more applicable in practical situations.

6.4 Discussion

We will discuss the fundamental limitations of the macroscopic models in transport theory perspective and then in practical device simulation perspective.

As we have mentioned previously, in moment-based macroscopic transport models, high order transport effects such as velocity overshoot are taken care of not in the hierarchy closure approximation but in the mobility model in the current equation with (or without) the next order moment equations, whose order is still too low to describe the high order effects. Through the benchmarking study, we identify the fundamental limitations of the moment-based models in describing quasi-ballistic transport in high-speed BJTs. For a diffusive base, the base transport should be compromised to describe collector transport and vice versa. For a graded base, the drift motion helps relax the trade-off between the base and the collector transport, but the range of the effectiveness is limited.

Thus, to describe nonequilibrium, quasi-ballistic transport in general with moment-based macroscopic models, we need to: 1) terminate hierarchy at the current equation since the next one or two higher order moment equations do not improve much the accuracy on the high order effects (see the ballistic peak in the distribution in Fig. 6.2) and then 2) apply such a hierarchy closure approximation to the current equation that it can capture the fundamental carrier motions such as thermionic emission velocity in a low-field base, overshoot velocity in the beginning of the high-field collector, saturation velocity deep in a high-field collector, and the smooth transition in between.

Recall that these velocities and the smooth transition result from how deterministic carrier motion that follows Newton's law is distorted by random motion due to scattering. In this sense, the hierarchy closure problem in the moment-based equations is similar to that of the theory of impact ionization, where Shockley's lucky electron model [67] describes how a carrier survives scattering in a low-field bulk, Wolf's theory does that in a high-field bulk [68] and Baraff's maximum anisotropy principle [19] connects both extremes [69].

Using the above recipe, we extended the one-flux model with a hierarchy closure approximation inspired by Baraff's theory [19], and had a full success for ballistic transport [53] and a limited success for quasi-ballistic transport in a nanoscale MOSFET (Chap. 5), where velocity saturation is not likely to occur. It was possible because the error margin for a velocity profile of nanoscale MOSFETs is not very tight; as long as the model captures the accurate source injection velocity, it is enough for a transport model to describe approximately a certain amount of velocity overshoot, which indirectly affects the on-current, the figure of merits of logic devices.

In high-speed BJTs, on the other hand, the error margin for velocity profile is much tighter for a wider region in a device because the velocity profile in the entire region is of more interest than in MOSFETs. Moreover, there is no understanding that can provide a quantitative macroscopic description of how carrier velocity changes from the thermionic velocity, to overshoot velocity, then to saturation velocity. Note that Baraff's theory does not provide the required understanding but a

quantitative circumvention of the closure problem of a spherical harmonics expansion of the Boltzmann equation, which is a microscopic description. Thus, we expect that no moment-based macroscopic transport model can describe nonequilibrium, quasi-ballistic transport in general for a wide range of design options, although for a graded base, advanced models such as ET and HD can match MC simulation for a certain range of base width of practical interest.

In the perspective of practical device simulation, we observed that the above fundamental limitations of macroscopic models may not lead to an unreasonable estimation of circuit design figure of merits such as transit time and cutoff frequency of practical SiGe BJTs. It is because: 1) the graded base reduces the abnormal base diffusion effect, and 2) the collector becomes the bottleneck of transport through the design options demanded by physics other than pure transport such as collector breakdown. This is likely to be true in the simulation of III-V HBTs, but the problem is too complicated to be distilled down to a simple one because they usually operate at high-current level for optimum performance [70] and have complicated collector structure due to the effect of multi-valley transport such as velocity modulation [71]. Thus, it is not easy to show a persuasive quantitative assessment on whether advanced macroscopic models are applicable to the simulation of high-speed III-V HBTs. However, we believe it is important to be aware of those limitations in mind when exploring a wide range of device design options using macroscopic transport models.

6.5 Summary

We investigated nonequilibrium, quasi-ballistic transport in high-speed BJTs by benchmarking against MC simulation various macroscopic transport models derived from the moments of the Boltzmann equation. We identified their fundamental limitations, and discussed the essence of the problem in relation to the theory of impact ionization explaining why no macroscopic models can overcome those limitations. But we also demonstrated that those limitations may not lead to an unreasonable

estimation of circuit performance of practical high-speed HBTs, and consequently resolved the apparent contradiction in related existing research.

7. SUMMARY AND FUTURE WORK

7.1 Summary

In Chapter 1, we defined quasi-ballistic transport, and pointed out that in this transport regime, the effects of deterministic carrier motion on determining important physical quantities required in device design and circuit simulation are missing in conventional macroscopic transport models. Further, we implied that this may lead to unreliable performance assessments of nanoscale MOSFETs and ultrahigh-speed BJTs as the scale or speed of devices is challenged to the edge.

In Chapter 2, we explored pure ballistic transport for a nanoscale MOSFET, where current and carrier density are solely determined by deterministic carrier motion between contacts not by scattering. We solved the Boltzmann equation in the ballistic limit, obtained ballistic carrier distribution, confirmed the essential physics in [22] as reflected in the shape of the distribution, and identified the possible limitations of moment-based macroscopic models in describing quasi-ballistic transport. This work is published as [1].

In Chapter 3, we performed a benchmarking study of conventional macroscopic models, the drift-diffusion models including Bude's modification [11], and a version of the energy transport models for *near*-ballistic transport in a nanoscale MOSFET assuming source/drain series resistance is not a limiting factor. We identified the limitations discussed in Chapter 2 and suggested a way to develop a new model that does not suffer the limitations. We published the previous version of benchmarking study for an *n*-*i*-*n* structure as [12] and this version for a nanoscale MOSFET as [47].

In Chapter 4, based on the understanding developed in the previous chapters, we formulated a drift-diffusion type macroscopic model that accurately describes the ballistic transport in a nanoscale MOSFET. We developed the directed-moment ap-

proach by extending McKelvey's flux method, in which we separate injections from different contacts, and describe each injection with flux equations derived from the partial moments or directed moments of the BTE, which adopt new hierarchy closure approximations valid in the ballistic limit and thus overcome the aforementioned limitations. This work is published as [53].

In Chapter 5, we implemented a simple scattering model into the ballistic drift-diffusion equation developed in Chapter 4 and formulated a quasi-ballistic drift-diffusion model for a nanoscale MOSFET where velocity saturation is not likely to occur. We demonstrated that the model can describe carrier transport from the diffusive limit to the ballistic limit as the strength of scattering decreases. Also, we learned that not only the scattering but also the hierarchy closure approximations should depend on dominant scattering mechanisms, which is determined by the type of material, bandstructure, and device structure since the effects of deterministic carrier motion depends on those factors as well. However, we found that this approach is limited to 1D. Although the quasi-ballistic DD model is limited to 1D and the scattering model is yet to be refined, it shows a promise to the simulation of carbon nanotube FETs and silicon nanowire FETs, whose intrinsic part is essentially 1D. This work is submitted to Semiconductor Research Corporation (SRC) as a deliverable report.

In Chapter 6, we studied quasi-ballistic transport in high-speed silicon based BJTs, identified fundamental limitations of the drift-diffusion, the one-flux, and the energy transport models, and resolved the issue on whether macroscopic transport models are applicable to the simulation of high-speed BJTs. We explained in relation to the theory of impact ionization that no macroscopic model based on moments of the BTE can overcome these limitations, but showed that despite the fundamental limitations in transport theory perspective, conventional macroscopic models can be used to assess transit time and cutoff frequency of practical SiGe BJTs. However, we believe that those limitations should be born in mind in exploring a wide range of design options since the design window may touch an area in the design space where the limitations could lead to an unreasonable performance assessment.

All in all, in this thesis:

1. We developed a thorough understanding of the physics of quasi-ballistic transport in nanoscale FETs and high-speed BJTs.

2. Based on this understanding, we identified the fundamental limitations of widely used macroscopic transport models such as the drift-diffusion and the energy transport models, and learned that the the fundamental limitations can lead to unreasonable assessment of device and circuit performance.

3. Then, we built a new macroscopic transport model that does understand quasi-ballistic transport and the ballistic limit for the first time.

4. But we also found that the model has its own limitations as a general-purpose replacement for conventional models. Although it is not a comprehensive, quantitative macroscopic description of quasi-ballistic transport, it should prove useful for small field-effect transistors.

5. Finally, we conclude that conventional macroscopic models will continue to be used unless device functionality takes a full advantage of quantum effects, but that the users should be aware of their fundamental limitations as they may lead to incorrect prediction of device performance.

7.2 Future Work

The possibility of developing a general, comprehensive macroscopic transport model valid from the diffusive to the ballistic limit seems remote. There is, however, some future work that would make this study more complete.

7.2.1 Scattering in nanoscale FETs

As mentioned in Chapter 5, the quasi-ballistic drift-diffusion model shows a promise for simulating nanoscale FETs whose internal structure is essentially 1D. However,

the scattering and hierarchy closure approximations depend on the characteristics of dominant scattering mechanisms, which is not fully known yet for strongly quantum-confined degenerate carriers. Thus, to complete the quasi-ballistic model for nanoscale FETs, we need to understand the dominant scattering mechanisms and how they can be translated into macroscopic quantities such as low-field mobility.

7.2.2 Quasi-ballistic transport in high-speed III-V HBTs

As an extension of Chapter 6, it would be of interest to do a similar study for high-speed III-V HBTs since there are two things that make the problem different from that of SiGe BJTs: 1) that the thermionic emission velocity is much higher than the saturation velocity, thus 2) that unlike the velocity overshoot resulting from the ballistic peak in SiGe BJTs, the collector velocity overshoot and collector design of III-V HBTs are more or less associated with extending the transport of carriers in the Γ-valley into the deep collector [57]. As a result, advanced III-V HBTs are often operated at high current levels for optimum high-speed performance [70]. This requires self-consistent simulation with the Poisson equation and we may have to examine multi-valley balance equations [13] to complete the assessment of macroscopic models.

LIST OF REFERENCES

LIST OF REFERENCES

[1] Jung-Hoon Rhew, Zhibin Ren, and M. S. Lundstrom. A numerical study of ballistic transport in a nanoscale MOSFET. *Solid-State Electronics*, 46(11):1899–1906, November 2002.

[2] Jing Guo, Mark Lundstrom, and Supriyo Datta. Performance projections for ballistic carbon nanotube field-effect transistors. *Appl. Phys. Lett.*, 80(17):3192–3194, April 2002.

[3] M. Stettler, A. Das, and M. S. Lundstrom. *S-DEMON 3.1 user's manual*. School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, 1997.

[4] S. Thompson, N. Anand, and et al. M. Armstrong. A 90 nm logic technology featuring 50 nm strained silicon channel transistors, 7 layers of Cu interconnects, low k ILD, and 1 μm^2 SRAM cell. In *IEDM 2002. Technical Digest*, pages 61–64. IEEE, December 2002.

[5] H. Jackel, U. Hammer, J. Ruiz, I. Schnyder, V. Schwarz, A. Gaspar, D. Huber, A. Rohner, and A. Huber. High speed InP-based HBTs and OEICs. In *IEDM 2002. Technical Digest*, pages 83–86. IEEE, December 2002.

[6] International Technology Roadmap for Semiconductors, 2002.

[7] M. Lundstrom, R. W. Dutton, D. K. Ferry, and K. Hess. Foreword (in Special Issue on Computational Electronics). *IEEE Trans. on Electron Dev.*, 47(18):1778, October 2000.

[8] S. Datta. *Electronic Transport in Mesoscopic Systems*. Cambridge University Press, Cambridge, U.K, 1995.

[9] S. E. Laux and M. V. Fischetti. Monte-Carlo simulation of submicrometer Si n-MOSFETs at 77 and 300 K. *IEEE Electron Dev. Lett.*, 9(9):467–469, September 1988.

[10] Kausar Banoo. *Direct solution of the Boltzmann transport equation in nanoscale Si devices*. PhD thesis, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, December 2000.

[11] J. D. Bude. MOSFET modeling into the ballistic regime. In *2000 International Conference on Simulation Semiconductor Processes and Devices, SISPAD 2000*, pages 23–26. IEEE, September 2000.

[12] Kausar Banoo, Jung-Hoon Rhew, M. S. Lundstrom, Chi-Wang Shu, and J. W. Jerome. Simulating quasi-ballistic transport in Si nanotransistors. *VLSI Design*, 13(1-4):5–13, 2001.

[13] K. Tamura and K. Tomizawa. Comparative study of balance equation models. *IEE Proc.: Circuits, Devices & Systems*, 146(5):285–290, October 1999.

[14] C. Jungemann, B. Neinhus, and B. Meinerzhagen. Comparative study of electron transit times evaluated by DD, HD, and MC device simulation for a SiGe HBT. *IEEE Trans. on Electron Dev.*, 48(10):2216–2220, October 2001.

[15] M. G. Ancona. Macroscopic description of quantum-mechanical tunneling. *Phys. Rev. B*, 42(2):1222–33, July 1990.

[16] D. K. Ferry, R. Akis, and D. Vasileska. Quantum effects in MOSFETs: use of an effective potential in 3D Monte Carlo simulation of ultra-short channel devices. In *IEDM 2000. Technical Digest*, pages 287–290. IEEE, December 2000.

[17] M. Trovato and L. Reggiani. Maximum entropy principle within a total energy scheme: Application to hot-carrier transport in semiconductors. *Phys. Rev. B*, 61(24):16667–16681, June 2000.

[18] C. Cercignani, I. M. Gamba, J. W. Jerome, and Chi-Wang Shu. A domain decomposition method: a simulation study. In *1998 Sixth International Workshop on Computational Electronics. Extended Abstracts (Cat. No.98EX116) 6th IWCE*, pages 174–177. IEEE.

[19] G. A. Baraff. Maximum anisotropy approximation for calculating electron distributions; application to high field transport in semiconductors. *Phys. Rev.*, 133(1A):A26–A33, January 1964.

[20] T. Grasser, Ting-Wei Tang, H. Kosina, and S. Selberherr. A review of hydrodynamic and energy-transport models for semiconductor device simulation. *Proc. of the IEEE*, 91(2):251–274, February 2003.

[21] M. S. Lundstrom. Elementary scattering theory of the Si MOSFET. *IEEE Electron Dev. Lett.*, 18(7):361–363, July 1997.

[22] M. S. Lundstrom and Z. Ren. Essential physics of carrier transport in nanoscale MOSFETs. *IEEE Trans. on Electron Dev.*, 49(1):133–141, January 2002.

[23] M. S. Lundstrom and Jung-Hoon Rhew. A landauer approach to nanoscale MOSFETs. *J. of Computational Electronics*, 1:481–489, 2002.

[24] Dietmar Schroeder. *Modeling of Interface Carrier Transport for Device Simulation*. Springer-Verlag, Wien New York, 1994.

[25] M. S. Shur. Low ballistic mobility in submicron HEMTs. *IEEE Electron Dev. Lett.*, 23(9):511–513, September 2002.

[26] J. Wang and M. S. Lundstrom. Ballistic transport in high electron mobility transistors. *IEEE Trans. on Electron Dev.*, 50(7):1604–1609, July 2003.

[27] J. P. McKelvey, R. L. Longini, and T. P. Brody. Alternative approach to the solution of added carrier transport problems in semiconductors. *Phys. Rev.*, 123(1):51–57, July 1961.

[28] W. Shockley. Diffusion and drift of minority carriers in semiconductors for comparable capture and scattering mean free paths. *Phys. Rev.*, 125(5):1570–1576, March 1962.

[29] M. S. Lundstrom. On the mobility versus drain current relation for a nanoscale MOSFET. *IEEE Electron Dev. Lett.*, 22(6):293–295, January 2001.

[30] G. Timp, J. Bude, K. K. Bourdelle, J. Garno, A. Ghetti A, H. Gossmann, M. Green, G. Forsyth, Y. Kim, R. Kleiman, F. Klemens, A. Kornblit, C. Lochstampfor, W. Mansfield, S. Moccio, T. Sorsch, D. M. Tennant, W. Timp, and R. Tung. The ballistic nano-transistor. In *IEDM 1999. Technical Digest*, pages 55–58. IEEE, December 1999.

[31] P. Hesto P, J. F. Pone, R. Castagne, and J. L. Pelouard. A dynamic Monte Carlo simulation of conduction in submicron GaAs devices at 77K. In H. L. Grubin; K. Hess; G. J. Iafrate; D. K. Ferry, editor, *Physics of Submicron Structures*, pages 101–106, New York, NY, USA, 1984. Plenum.

[32] H. U. Baranger and J. W. Wilkins. Ballistic structure in the electron distribution function of small semiconducting structures: general features and specific trends. *Phys. Rev. B*, 36(3):1487–1502, July 1987.

[33] F. Assad, Zhibin Ren, D. Vasileska, S. Datta, and M. S. Lundstrom. On the performance limits for Si MOSFETs: a theoretical study. *IEEE Trans. on Electron Dev.*, 47(1):232–240, January 2000.

[34] Zhibin Ren, R. Venugopal, S. Datta, M. S. Lundstrom, D. Jovanovic, and J. Fossum. The ballistic nanotransistor: a simulation study. In *IEDM 2000. Technical Digest*, pages 10–13. IEEE, December 2000.

[35] Y. Naveh and K. K. Likharev. Modeling of 10-nm-scale ballistic MOSFET's. *IEEE Electron Dev. Lett.*, 21(5):242–244, May 2000.

[36] Y. Taur and T. Ning. *Fundamentals of Modern VLSI Devices*. Cambridge University Press, Cambridge, U.K, 1998.

[37] F. Stern. Self-consistent results for n-type Si inversion layers. *Phys. Rev. B*, 5(12):4891–4899, June 1972.

[38] A.C. Smith, J. F. Janak, and R. B. Adler. *Electronic Conduction in Solids*. McGraw-Hill, New York, 1967.

[39] M. Nekovee, B. J. Geurts, H. M. J. Boots, and M. F. H. Schuurmans. Failure of extended-moment-equation approaches to describe ballistic transport in submicrometer structures. *Phys. Rev. B*, 45(12):6643–6651, March 1992.

[40] A. Lochtefeld and D. A. Antoniadis. On experimental determination of carrier velocity in deeply scaled NMOS: how close to the thermal limit? *IEEE Electron Dev. Lett.*, 22(2):95–97, February 2001.

[41] C. Jungemann, B. Neinhus, and B. Meinerzhagen. Spatial analysis of the electron transit time in a silicon/germanium heterojunction bipolar transistor by drift-diffusion, hydrodynamic, and full-band Monte Carlo device simulation. In *2000 International Conference on Simulation Semiconductor Processes and Devices, SISPAD 2000*, pages 42–45. IEEE, September 2000.

[42] D. Chen, E. C. Kan, U. Ravaioli, C-W. Shu, and R. W. Dutton. An improved energy transport model including nonparabolicity and non-Maxwellian distribution effects. *IEEE Electron Dev. Lett.*, 13(1):26–28, January 1992.

[43] K. Rim, J. L. Hoyt, and J. F. Gibbons. Transconductance enhancement in deep submicron strained Si n-MOSFETs.

[44] M. S. Lundstrom. *Fundamentals of Carrier Transport*. Cambridge University Press, Cambridge, U.K, 2000.

[45] D. M. Caughey and R. E. Thomas. Carrier mobilities in Silicon empirically related to doping and field. *Proc. of IEEE*, pages 2192–2193, December 1967.

[46] J. A. Cooper Jr. and D. F. Nelson. High-field drift velocity of electrons at the Si-SiO_2 interface as determined by a time-of-flight technique. *J. of Appl. Phys.*, 54(3):1445–1456, March 1983.

[47] J.-H. Rhew, Z. Ren, and M. S. Lundstrom. Benchmarking macroscopic transport models for nanotransistor TCAD. *J. of Computational Electronics*, 1:385–388, 2002.

[48] K. Banoo, M. S. Lundstrom, and R. K. Smith. Direct solution of the Boltzmann transport equation in nanoscale Si devices. In *2000 International Conference on Simulation Semiconductor Processes and Devices, SISPAD 2000*, pages 50–53. IEEE.

[49] M. A. Alam, M. A. Stettler, and M. S. Lundstrom. Formulation of the Boltzmann equation in terms of scattering matrices. *Solid-State Electronics*, 36(2):263–271, February 1993.

[50] S. Tanaka and M. S. Lundstrom. A flux-based study of carrier transport in thin-base diodes and transistors. *IEEE Trans. on Electron Dev.*, 42(10):1806–1815, October 1995.

[51] D. L. Scharfetter and H. K. Gummel. Large-signal analysis of a silicon Read diode oscillator. *IEEE Trans. on Electron Dev.*, ED-(1):64–77, January 1969.

[52] Neil W. Ashcroft and N. David Mermin. *Solid State Physics*. W. B. Saunders, Philadelphia, 1976.

[53] Jung-Hoon Rhew and M. S. Lundstrom. Drift-diffusion equation for ballistic transport in nanoscale metal-oxide-semiconductor field effect transistors. *J. of Appl. Phys.*, 92(9):5196–5202, November 2002.

[54] P. Hasler, A. G. Andreou, C. Diorio, B. A. Minch, and C. A. Mead. Impact ionization and hot-electron injection derived consistently from Boltzmann transport. *VLSI Design*, 8(1-4):455–461, 1998.

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

[56] K. Natori. Ballistic metal-oxide-semiconductor field effect transistor. *J. of Appl. Phys.*, 76(8):4879–4890, October 1994.

[57] T. Ishibashi. Nonequilibrium electron transport in HBTs. *IEEE Trans. on Electron Dev.*, 48(11):2595–2605, November 2001.

[58] J. P. Karamarkovic and N. D. Jankovic. Modification of drift-diffusion model for short base transport. *IEE Electronics Lett.*, 36(24):2047–2048, November 2000.

[59] Shin-Chi Lee and Ting-Wei Tang. Hydrodynamic modeling of silicon BJT with Monte Carlo calibrated transport coefficients. *IEICE Trans. on Electronics*, E75-C(2):189–193, February 1992.

[60] G. Baccarani, C. Jacoboni, and A. M. Mazzone. Current transport in narrow-base transistors. *Solid-State Electronics*, 20(1):5–10, January 1977.

[61] M. A. Stettler and M. S. Lundstrom. A microscopic study of transport in thin base silicon bipolar transistors. *IEEE Trans. on Electron Dev.*, 41(6):1027–1033, June 1994.

[62] D. B. M. Klaassen, J. W. Slotboom, and H. C. De Graaff. Unified apparent bandgap narrowing in n- and p-type silicon. *Solid-State Electronics*, 35(2):125–129, February 1992.

[63] K. Blotekjaer. Transport equations for electrons in two-valley semiconductors. *IEEE Trans. on Electron Dev.*, ED-17(1):38–47, January 1970.

[64] Y. Apanovich, E. Lyumkis, B. Polsky, A. Shur, and P. Blakey. Steady-state and transient analysis of submicron devices using energy balance and simplified hydrodynamic models. *IEEE Trans. on Computer-Aided Design of Integrated Circuits & Systems*, 13(6):702–711, June 1994.

[65] B. Meinerzhagen and W. L. Engl. The influence of the thermal equilibrium approximation on the accuracy of classical two-dimensional numerical modeling of silicon submicrometer MOS transistors. *IEEE Trans. on Electron Dev.*, 35(5):689–697, May 1988.

[66] S. E. Laux and W. Lee. Collector signal delay in the presence of velocity overshoot. *IEEE Electron Dev. Lett.*, 11(4):174–176, April 1990.

[67] William Shockley. Problems related to p-n junctions in silicon. *Solid-State Electronics*, 2(1):35–60, January 1961.

[68] P. A. Wolff. Theory of electron multiplication in silicon and germanium. *Phys. Rev.*, 95(6):1415–1420, September 1954.

[69] E. Zanoni and G. Meneghesso. Impact ionization in compound semiconductor devices. In H. S. Nalwa, editor, *Handbook of advanced electronic and photonic materials and devices*, volume 2, chapter 2, pages 67–131. Academic Press, San Diego, CA, 2001.

[70] M. Rohner, B. Willen, and H. Jackel. Sub-f_t gain resonance of InP/InGaAs-HBTs. *IEEE Trans. on Electron Dev.*, 49(2):213–220, February 2002.

[71] M. Rohner, B. Willen, and H. Jackel. Velocity modulation in III/V-HBTs. *IEEE Trans. on Electron Dev.*, 50(5):1205–1213, May 2003.

[72] S. Datta. *Quantum Phenomena*. Addison-Wesley, Reading, Mass., 1989.

[73] J. S. Blakemore. *Semiconductor Statistics*. Macmillian, New York, 1962.

[74] J. W. Mintmire and C. T. White. Universal density of states for carbon nanotubes. *Phys. Rev. Lett.*, 81(12):2506–2509, September 1998.

APPENDIX

APPENDIX A
ON THE VALIDITY OF THE SEMICLASSICAL APPROACH FOR 10 NM MOSFETS

The time-independent electron wave function implicitly assumed in our quasi-2D approach is [72]

$$\Phi(x,y,z) = \sqrt{k_x(0)/k_x(x)}\, e^{i\int dx' k_x(x')}\, e^{ik_y y} \cdot \phi(z), \tag{A.1}$$

where $k_x(x)$, the position-dependent wave number in the x-direction, is determined by

$$k_x(x) = F(x,E,k_y) = \sqrt{E - E_S(x) - \frac{\hbar^2 k_y^2}{2m_t^*}}. \tag{A.2}$$

Note that $k_x(x)$ is a function of well-defined quantum variables, E and k_y. Although it is expressed as $f(x,k_x,k_y)$ implicitly, the distribution of the Boltzmann transport equation is explicitly $f(x,E,k_y)$. Thus, no uncertainty relation applies to the pair of x and k_x. In other words, $k_x(x)$ is not spectral k_x (the Fourier transform pair of x) but a locally defined wave number. Consequently, average carrier velocity is not related to Δk_x, the uncertainty of spectral k_x, but to the average of local $k_x(x)$.

Equation A.1 is valid if the subband energy $E_S(x)$ varies slowly with position x so that quantum mechanical reflections can be negligible [72]. The WKB approximation in the x-dependent component of $\Phi(x,y,z)$ allows us to use the Boltzmann transport equation in describing transport along the x-direction at the cost of not capturing tunneling effects. Recent works [34] [35] have demonstrated that the assumption in Eq. A.1 is valid down in nanoscale MOSFETs to a channel length of approximately 10 nm. The reason is that quantum mechanical reflections exist in the regions that have little effects on macroscopic quantities such as current. However, it seems necessary to address the issue on the tunneling through the source-to-channel barrier.

Quantum calculation yields higher current than semiclassical calculation due to the tunneling if the same potential profile is given. However, self-consistent electrostatics makes little distinction between the macroscopic quantities of the two approaches; the charge density at the top of the barrier is mainly determined by gate electrostatics whether it is thermionic or tunneling charge. Hence, the barrier height for the quantum approach is higher than the semiclassical one resulting in small difference in the current [34]. However, below 10 nm, direct tunneling from source to drain disables the functionality of MOSFETs [34].

APPENDIX B
SUPPLEMENTAL MATERIAL FOR THE BALLISTIC DRIFT-DIFFUSION MODEL

B.1 Derivation of the One-Flux Equations

We present a brief but proper derivation of Eqs. 4.1 $\sim$ 4.4 from the Boltzmann transport equation (BTE). For 1D transport along the x-direction as in our model device, the steady-state BTE is

$$\upsilon_x \frac{\partial f}{\partial x} - q\mathscr{E}_x \frac{\partial f}{\partial p_x} = \left.\frac{\partial f}{\partial t}\right|_{coll} \tag{B.1}$$

where the carrier distribution as a function of x and $\vec{p} = (p_x, p_y, p_z)$ is $f = f(x, \vec{p})$. Equations 4.1 $\sim$ 4.4 are derived taking partial averages of Eq. B.1 over the two subdomains of the momentum space: 1) $\Omega_+ = \{\vec{p}\,|\, p_x > 0\}$ for + stream, and 2) $\Omega_- = \{\vec{p}\,|\, p_x < 0\}$ for − stream. The whole momentum space is $\Omega = \Omega_+ \cup \Omega_-$. The averages of the flux term of Eq. B.1 over $\Omega_\pm$ yield the spatial variations of $J^\pm$, i.e.,

$$\frac{1}{V}\sum_{\Omega_\pm} \upsilon_x \frac{\partial f}{\partial x} = \frac{d}{dx}\left(\frac{1}{V}\sum_{\Omega_\pm} \upsilon_x f\right) = \frac{d}{dx}\left(\pm J^\pm\right), \ \textit{Flux terms} \tag{B.2}$$

where V is a normalization volume and J^- is defined positively. Applying the chain rule, $\frac{\partial f}{\partial p_x} = \frac{\partial f}{\partial E}\cdot\frac{\partial E}{\partial p_x}$ and the identity, $\frac{\partial E}{\partial p_x} = \upsilon_x$, to the field term of Eq. B.1, we obtain

$$\frac{1}{V}\sum_{\Omega_\pm} q\mathscr{E}_x \frac{\partial f}{\partial p_x} = q\mathscr{E}_x\left(\frac{1}{V}\sum_{\Omega_\pm}(\upsilon_x f)\,\frac{1}{f}\frac{\partial f}{\partial E}\right), \tag{B.3}$$

where E is total carrier energy. Converting the sum over p_x into an integral, we obtain

$$\frac{1}{V}\sum_{\Omega_\pm} q\mathscr{E}_x \frac{\partial f}{\partial p_x} = \pm q\mathscr{E}_x \frac{1}{2\pi\hbar A}\sum_{p_y,p_z} f(p_x = 0, p_y, p_z), \tag{B.4}$$

where A is a normalization area of the y-z plane. Note that $f(0, p_y, p_z)$ represents stationary carriers in the x-direction, which are generated by the decelerating electric field (See Fig. 4.3). Hence, it is reasonable to assume that the distribution associated with a stream under deceleration remains in a near-equilibrium shape as in thermionic emission. In nondegenerate conditions, a hemi-Maxwellian, $f \sim \exp(-E/k_B T_L)$, can be assumed. Then, from Eq. B.4, we obtain

$$\frac{1}{V}\sum_{\Omega_\pm}(\upsilon_x f)\frac{1}{f}\frac{\partial f}{\partial E} = -\frac{1}{k_B T}\frac{1}{V}\sum_{\Omega_\pm}\upsilon_x f = \mp\frac{J^\pm}{k_B T}, \tag{B.5}$$

or

$$\frac{1}{V}\sum_{\Omega_\pm} q\mathscr{E}_x\frac{\partial f}{\partial p_x} = \mp\frac{q\mathscr{E}_x}{k_B T}J^\pm.\ Field\ terms \tag{B.6}$$

The averages of the scattering term of the BTE are expressed as

$$\frac{1}{V}\sum_{\Omega_\pm}\left.\frac{\partial f}{\partial t}\right|_{coll} = \pm\left(-\frac{n^+}{\tau^+} + \frac{n^-}{\tau^-}\right) \tag{B.7}$$

where $n^\pm$ are carrier densities in $\pm$ streams and $\tau^\pm$ are corresponding macroscopic relaxation times associated with backscattering. Although Eq. B.7 can be derived rigorously from the scattering integral of the BTE, the following phenomenological explanation verifies Eq. B.7. The averages of the BTE over $\Omega_\pm$ yield the rate equations for $n^\pm$. Thus, the backscattering of + stream decreases n^+ with the rate of n^+/τ^+ but the backscattering of $-$ stream increases n^+ with the rate of n^-/τ^-. Scattering that causes a carrier to remain in the same stream does not appear because it does not change the number of carriers. The sign $\pm$ in Eq. B.7 results from the continuity condition,

$$\frac{1}{V}\sum_{\Omega}\left.\frac{\partial f}{\partial t}\right|_{coll} = \frac{1}{V}\sum_{\Omega_+}\left.\frac{\partial f}{\partial t}\right|_{coll} + \frac{1}{V}\sum_{\Omega_-}\left.\frac{\partial f}{\partial t}\right|_{coll} = 0, \tag{B.8}$$

since scattering neither creates nor destroy carriers (we exclude explicit generation or recombination of carriers). To make Eq. B.7 compatible with the one-flux equations, we define scattering mean-free-paths for $\pm$ streams as

$$\lambda^\pm \equiv 1\Big/\xi^\pm = \langle\upsilon_x\rangle^\pm\,\tau^\pm. \tag{B.9}$$

Further, we assume that $\xi^+ \cong \xi^- \cong \xi_0$, which is valid in near-equilibrium. Then, using Eq. 4.5, Eq. B.7 becomes

$$\frac{1}{V}\sum_{\Omega_\pm}\left.\frac{\partial f}{\partial t}\right|_{coll} = \pm\left(-\xi_0 J^+ + \xi_0 J^-\right). \; Scattering\ terms \tag{B.10}$$

From Eqs. B.2, B.6, and B.10, we can obtain Eqs. 4.1 $\sim$ 4.4.

B.2 Derivation of the Degeneracy Factors for the Ballistic Drift-Diffusion Equation

We derive Eqs. 4.7 and 4.8. In degenerate semiconductors, the near equilibrium distribution f associated with streams under deceleration can be assumed to be

$$f\left(x,E\right) = \frac{1}{1+\exp\left(\left(E-\mu_\pm\right)/k_B T_L\right)}. \tag{B.11}$$

Using the property, $\frac{\partial f}{\partial E} = -\frac{\partial f}{\partial \mu_\pm}$, we can see that

$$\sum_{\Omega_\pm}\frac{\partial f}{\partial p_x} = \sum_{\Omega_\pm}\upsilon_x\frac{\partial f}{\partial E} = -\frac{\partial}{\partial \mu}\left(\sum_{\Omega_\pm}\upsilon_x f\right) = -\frac{\partial}{\partial \mu}\left(\pm J^\pm\right). \tag{B.12}$$

For the 2D electrons in our model device, $J^\pm = J_0\mathscr{F}_{1/2}\left(\eta_\pm\right)$ [33], where J_0 is bias independent flux density, $\eta_\pm(x) = \left(\mu_\pm - E_S(x)\right)/k_B T_L$. From Eqs. B.3, B.11, and B.12, we can show that

$$F^\pm_{\rm deg} = \frac{\mathscr{F}_{-1/2}\left(\eta_\pm\right)}{\mathscr{F}_{1/2}\left(\eta_\pm\right)}, \tag{B.13}$$

using the property of Fermi-Dirac integral [73],

$$\frac{\partial}{\partial \mu_\pm}\mathscr{F}_j\left(\eta_\pm\right) = \frac{1}{k_B T_L}\mathscr{F}_{j-1}\left(\eta_\pm\right). \tag{B.14}$$

B.3 Scharfetter-Gummel Discretization of the Ballistic Drift-Diffusion Equation

We show how to discretize Eq. 4.38 with the boundary conditions in Eqs. 4.45 and 4.46 using the Scharfetter-Gummel (SG) scheme [51]. Figure B.1 shows that in the SG scheme, the position x, the equivalent density N, and the potential V are

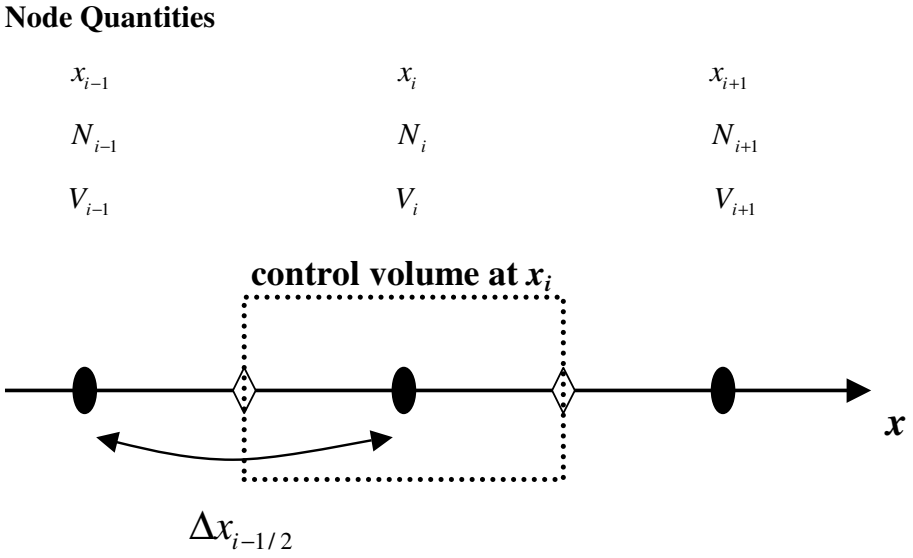

Fig. B.1. Node and edge quantities in the Scharfetter-Gummel discretization

defined at the nodes (or centers) of the control volumes indexed with $i-1$, i, $i+1$ (i=2, 3, ..., M−1, M is the total number of nodes) but the flux density J, electric field $\mathscr{E}_x$, the equivalent mobility μ^*, and the degeneracy factor $F_{\rm deg}$ are defined at the edges of the control volumes, which are indexed with $i-\frac{1}{2}$, $i+\frac{1}{2}$.

We first integrate Eq. 4.38 over the interval $[x_i, x_{i+1}]$ weighting with a function $W(x)$ that converts the right-hand-side of Eq. 4.38 into a total derivative with respect to x, which is given as

$$W(x) = \exp\left(\frac{q\mathscr{E}_x}{k_B T_L} x\right). \tag{B.15}$$

(We assume a nondegenerate condition, i.e., $F_{\rm deg} = 1$ but it can be readily included in the derivation by replacing $1/k_B T_L$ with $F_{\rm deg}/k_B T_L$.) Then, Eq. 4.38 becomes

$$J_{i+1/2} = -\mu^*_{i+1/2} \frac{k_B T_L}{q\Delta x_{i+1/2}} \left[B\left(-\frac{q\mathscr{E}_{i+1/2}}{k_B T_L}\Delta x_{i+1/2}\right) N_{i+1} - B\left(\frac{q\mathscr{E}_{i+1/2}}{k_B T_L}\Delta x_{i+1/2}\right) N_i \right], \tag{B.16}$$

where $B(\bullet)$ is the Bernoulli function, i.e., $B(x) = x/[\exp(x)-1]$ and $\Delta x_{i+1/2} = x_{i+1} - x_i$. Note that the edge quantities (e.g., $J_{i+1/2}$) are assumed to be constant over the interval $[x_i, x_{i+1}]$ in the integration. Similarly, we integrate Eq. 4.38 over the interval $[x_{i-1}, x_i]$ weighting with $W(x)$ and get

$$J_{i-1/2} = -\mu^*_{i-1/2} \frac{k_B T_L}{q\Delta x_{i-1/2}} \left[B\left(-\frac{q\mathscr{E}_{i-1/2}}{k_B T_L}\Delta x_{i-1/2}\right) N_i - B\left(\frac{q\mathscr{E}_{i-1/2}}{k_B T_L}\Delta x_{i-1/2}\right) N_{i-1} \right]. \tag{B.17}$$

Since the flux density satisfies the continuity equation in Eq. 4.34, Eqs. B.16 and B.17 should satisfy

$$J_{i+1/2} - J_{i-1/2} = 0. \tag{B.18}$$

From Eqs. B.16 $\sim$ B.18, we can set up a matrix equation for all N_i except at the boundaries as

$$C_1^{i-1/2} N_{i-1} - \left[C_1^{i+1/2} - C_2^{i-1/2}\right] N_i - C_2^{i+1/2} N_{i+1} = 0, \tag{B.19}$$

where

$$C_1^{i+1/2} = -\mu^*_{i+1/2} \frac{k_B T_L}{q\Delta x_{i+1/2}} B\left(\frac{q\mathscr{E}_{i+1/2}}{k_B T_L}\Delta x_{i+1/2}\right), \tag{B.20}$$

and

$$C_2^{i+1/2} = \mu^*_{i+1/2} \frac{k_B T_L}{q \Delta x_{i+1/2}} B\left(-\frac{q\mathscr{E}_{i+1/2}}{k_B T_L} \Delta x_{i+1/2}\right), \tag{B.21}$$

i=1, ..., M−1.

Discretizing the boundary conditions in Eqs. 4.45 and 4.46 using the same technique, we get

$$\left(1 - C_1^{1+1/2}\right) N_1 - C_2^{1+1/2} N_2 = 2J^+(0) \tag{B.22}$$

at $x = 0$, and

$$C_1^{\mathrm{M}-1/2} N_{\mathrm{M}-1} + \left(1 + C_2^{\mathrm{M}-1/2}\right) N_{\mathrm{M}} = 2J^-(L) \tag{B.23}$$

at $x = L$.

Using Eqs. B.19, B.22, and B.23, we obtain the following tridiagonal matrix equation for N_i (i=1, ..., M) as

$$\begin{bmatrix} 1 - C_1^{1+1/2} & C_2^{1+1/2} & & & & 0 \\ \ddots & \ddots & \ddots & & & \\ & C_1^{i-1/2} & -C_1^{i+1/2} + C_2^{i-1/2} & C_2^{i+1/2} & & \\ & & \ddots & \ddots & \ddots & \\ 0 & & & C_1^{\mathrm{M}-1/2} & 1 + C_2^{\mathrm{M}-1/2} & \end{bmatrix} \begin{bmatrix} N_1 \\ \vdots \\ N_i \\ \vdots \\ N_{\mathrm{M}} \end{bmatrix} = \begin{bmatrix} 2J^+(0) \\ \vdots \\ 0 \\ \vdots \\ 2J^-(L) \end{bmatrix}. \tag{B.24}$$

From N_i's, we can get $J_{i+1/2}$'s using Eq. B.16.

B.4 Modification of the Ballistic Drift-Diffusion Model for a Carbon Nanotube FET

Starting from the bandstructure, we derive modification of the ballistic drift-diffusion model for the model CNTFET in Fig. 4.1.

For the lowest subband $E_S(x)$ in the model CNTFET, there are two valleys of conduction that satisfy the following E-k relation [74]

$$E = A\sqrt{k^2 + k_R^2} - A|k_R| + E_S(x), \tag{B.25}$$

where $A = \frac{3}{2}bt$ and the length of C-C bonding $b \approx 1.44\mathring{A}$ and the bonding energy $t \approx 2.5$ eV. The confinement wave number k_R is expressed in terms of the tube diameter d as $k_R = 2\pi/3\pi d$ [74]. The corresponding density of states (DOS) is [2]

$$D(E) = \frac{2}{\pi}\frac{dk}{dE} = \frac{4}{3\pi bt}\frac{|E - E_S(x) + \Delta|}{\sqrt{(E - E_S(x) + \Delta)^2 - \Delta^2}}, \tag{B.26}$$

where $\Delta \equiv A|k_R|$. From the DOS, we get carrier density and current for the positive stream at the top of the source-to-channel barrier as

$$n^+ = \frac{2}{\pi}\int_0^\infty \frac{dk}{1 + \exp\left[\left(\sqrt{A^2k^2 + \Delta^2} - \Delta + E_S(x) - \mu_+\right)/k_BT_L\right]} \tag{B.27}$$

and

$$I^+ = q\frac{4k_BT_L}{\pi\hbar}\log\left[1 + \exp\left(\frac{\mu_+ - E_S(x)}{k_BT_L}\right)\right]. \tag{B.28}$$

where μ_+ is the Fermi-level for the positive stream, according to which the states are occupied.

Now we derive the degeneracy factor for the positive stream as in Eq. 4.9, the thermal injection velocity $\tilde{v}_T^+$ as in Eq. 4.22, and the overshoot peak velocity v_{peak}^+ as in Eq. 4.17. From Eq. 4.9, the degeneracy factor for the positive stream becomes

$$F_{\text{deg}}^+ = \frac{dI^+/d\zeta_+}{I^+} = \frac{\mathscr{F}_{-1}(\zeta_+)}{\log\left[1 + \exp(\zeta_+)\right]}, \tag{B.29}$$

where $\mathscr{F}_{-1}$ is the Fermi-Dirac integral of order -1 and the normalized energy ζ_+ is

$$\zeta_+ = \frac{\mu_+ - E_S(x)}{k_BT_L}. \tag{B.30}$$

The negative part can be obtained in the same way. From Eqs. B.27 and B.28, we get thermal injection velocity $\tilde{v}_T^+$ as

$$\tilde{v}_T^+ = v_F\frac{\log\left[1 + e^{\zeta_+}\right]}{\int_0^\infty \frac{1}{1+\exp\left(\sqrt{\zeta^2+\delta^2}-\delta-\zeta_+\right)}d\zeta}, \tag{B.31}$$

where the Fermi velocity of the nanotube $v_F = \frac{3}{2}tb/\hbar$, $\zeta = Ak/k_BT_L$, and $\delta = \Delta/k_BT_L$. The overshoot peak velocity at x is approximated as the group velocity at energy $E_{\max} - E_S(x)$, which is

$$v_{peak}^+(x) = \frac{1}{\hbar}\frac{\partial E}{\partial k} = v_F\frac{\sqrt{(E_{\max} - E_S(x) + \Delta)^2 - \Delta^2}}{|E_{\max} - E_S(x) + \Delta|}. \tag{B.32}$$

Then, the ballistic closure approximation on $\upsilon^+(x)$ in $x_{\max} < x < L$ becomes

$$\upsilon^+(x) = \sqrt{\tilde{\upsilon}_T^+(x = x_{\max})^2 + \upsilon_{peak}^+(x)^2}. \tag{B.33}$$

In this way, one can derive the rest of the closure approximations similar to in Eqs. 4.18 $\sim$ 4.21 for the model CNTFET.

APPENDIX C
SUPPLEMENTAL MATERIAL FOR THE QUASI-BALLISTIC DRIFT-DIFFUSION MODEL

C.1 Mean Free Path vs. Low-field Mobility

There are several versions of low-field mobility μ_0 to mean-free-path relation λ_{el}, but we adopt the relation that arises when Eq. 5.2 reduces to the drift-diffusion equation. First, we derive a relation between the associated diffusivity D_0 and the mean free path, which reduces to Shockley's relation in nondegenerate conditions [28],

$$D_0 = \frac{\upsilon_T}{2}\lambda. \tag{C.1}$$

Then, we relate μ_0 to λ_{el} using the modified Einstein relation for confined degenerate carriers in inhomogeneous semiconductors [53]

$$D_0 = \frac{k_B T_L}{q}\mu_0 \frac{\mathscr{F}_{1/2}}{\mathscr{F}_{-1/2}}, \tag{C.2}$$

where $\mathscr{F}_j$ is the Fermi-Dirac integral of order j.

In a low-field condition $\mathscr{E}_{scatt} \gg |\mathscr{E}_x| \approx 0$, we can ignore the field terms in Eqs. 5.2 and 4.39 get

$$J = q\frac{\upsilon_T}{2\xi_0}\frac{dN}{dx}, \tag{C.3}$$

which should be identical to the Fick's law

$$J = qD_0\frac{dn}{dx}. \tag{C.4}$$

Under this condition, inelastic scattering is not active and the distribution is almost symmetric. Thus, we get: 1) $\xi_0 = \lambda_{el}^{-1}$, 2) $N \approx 2J^+/\upsilon_T$ since $J^+ \approx J^- \gg -J/q$, and 3) $n \approx 2n^+$ since $n^+ \approx n^-$. Also for 2D confined carriers, the positive and the negative streams are given as $J^\pm = J_0\mathscr{F}_{1/2}(\eta_\pm)$ and the densities are $n^\pm = n_0\mathscr{F}_0(\eta_\pm)$,

Table C.1
λ_{el} to μ_0 relation for both degenerate and nondegenerate conditions

λ_{el} [nm]	1	3	5	7	10	20	100
$\mu_0^{\deg}$ [cm^2/Vs]	22	62	101	140	198	387	1881
μ_0^{non} [cm^2/Vs]	23	70	116	160	232	463	2317

where J_0 and n_0 are a bias independent flux and carrier density respectively [33]. From Eqs. C.3 and C.4 we obtain

$$J = q\frac{\lambda_{el}}{2}J_0\frac{d\mathscr{F}_{1/2}}{dx} = qD_0n_0\frac{d\mathscr{F}_0}{dx}. \tag{C.5}$$

Then, we obtain a modified Shockley relation as

$$D_0 = \frac{\upsilon_T}{2}\lambda_{el}\frac{\mathscr{F}_{-1/2}}{\mathscr{F}_{-1}}, \tag{C.6}$$

which becomes Eq. C.1 in nondegenerate conditions. We utilize the following facts that $J_0/n_0 = \upsilon_T$, and that from the property of derivatives of Fermi-Dirac integrals [73] and the chain rule, we have

$$\frac{d\mathscr{F}_{1/2}}{dx}\Big/\frac{d\mathscr{F}_0}{dx} = \frac{d\mathscr{F}_{1/2}}{d\eta}\Big/\frac{d\mathscr{F}_0}{d\eta} = \frac{\mathscr{F}_{-1/2}}{\mathscr{F}_{-1}}. \tag{C.7}$$

Now, eliminating D_0 in Eqs. C.2 and C.4, we obtain the relation between μ_0 and λ_{el} as

$$\mu_0 = \frac{q}{k_BT_L}\frac{\upsilon_T}{2}\frac{\mathscr{F}^2_{-1/2}}{\mathscr{F}_{-1}\mathscr{F}_{1/2}}\lambda_{el}. \tag{C.8}$$

Table C.1 shows the value of mobility extracted from the simulation results in section 5.4 for given λ_{el} in both degenerate (μ_0^{deg})and nondegenerate (μ_0^{non}) conditions. Note that in a 10 nm nanoscale MOSFET, at quite low values of μ_0 transport starts to become quasi-ballistic, i.e., $\lambda_{el} > \ell \approx 3$ nm, where ℓ is the length of the critical region under high drain bias [22].

C.2 Derivation of the Quasi-Ballistic Closure in Eqs. 5.23 $\sim$ 5.26

We derive Eqs. 5.23$\sim$ 5.26 for the source-injected carriers in the channel region shown in Fig. 5.1. We assume that $W_{xx}^{+} = \frac{1}{2}m^{*}n^{+}\left\langle \upsilon^{2}\right\rangle ^{+}$, the total kinetic energy density in the transport direction (x-direction) of the positive stream, has three components: 1) thermal component with the total density n^{+}

$$W_{xx}^{+}\Big|_{th} = \frac{1}{2}m^{*}n^{+}\tilde{\upsilon}_{T}^{2}\left(x = x_{\max}\right), \tag{C.9}$$

2) ballistic component in Eqs. 5.13 and 5.14 with the density n_{ball}^{+}

$$W_{xx}^{+}\Big|_{ball} = \frac{1}{2}m^{*}n_{ball}^{+}\,\upsilon_{B\,peak}^{2}, \tag{C.10}$$

and 3) scattered component in Eqs. 5.13, 5.15 and 5.16 with the density $n^{+} - n_{ball}^{+}$

$$W_{xx}^{+}\Big|_{scat} = \frac{1}{2}m^{*}\left(n^{+} - n_{ball}^{+}\right)\left\langle \left(\upsilon_{S\,peak}\vec{r}\cdot\vec{x}\right)^{2}\right\rangle _{\phi}^{+}. \tag{C.11}$$

The brackets in Eq. C.11 denote an average over angle ϕ in 2D momentum space associated with the positive half, which is

$$\left\langle \left(\upsilon_{S\,peak}\hat{r}\cdot\hat{x}\right)^{2}\right\rangle _{\phi}^{+} = \upsilon_{S\,peak}^{2}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cos^{2}\phi\,d\phi\Big/\int_{-\pi}^{\pi}d\phi\;=\;\frac{1}{4}\upsilon_{S\,peak}^{2}. \tag{C.12}$$

Note that the domain of the integral in the numerator is $[-\pi/2, \pi/2]$ for the positive half. From Eqs. C.9 $\sim$ C.12, we get

$$\left\langle \upsilon^{2}\right\rangle ^{+} = \tilde{\upsilon}_{T}^{2} + \frac{1}{4}\upsilon_{S\,peak}^{2}\left[1 - \frac{n_{ball}^{+}\left(x\right)}{n^{+}\left(x\right)}\right] + \upsilon_{B\,peak}^{2}\frac{n_{ball}^{+}\left(x\right)}{n^{+}\left(x\right)}. \tag{C.13}$$

Since $B(x) = n_{ball}^{+}(x)/n^{+}(x)$, and $\sqrt{\left\langle \upsilon^{2}\right\rangle ^{+}} \approx \left\langle \upsilon\right\rangle ^{+}$ if the peak distributions in Eq. 5.13 is dominant [53], Eq. C.13 reduces to Eq. 5.23

$$\left\langle \upsilon\right\rangle ^{+} = \sqrt{\tilde{\upsilon}_{T}^{2}\left(x = x_{\max}\right) + \frac{1}{4}\upsilon_{S\,peak}^{2}\left[1 - B\left(x\right)\right] + \upsilon_{B\,peak}^{2}B\left(x\right)}. \tag{C.14}$$

The corresponding backscattered stream has the same components except the ballistic part, thus we get Eq. 5.24

$$\left\langle \upsilon\right\rangle ^{-} = \sqrt{\tilde{\upsilon}_{T}^{2}\left(x = x_{\max}\right) + \frac{1}{4}\upsilon_{S\,peak}^{2}\left[1 - B\left(x\right)\right]}. \tag{C.15}$$

VITA

Jung-Hoon Rhew was born in Seoul, South Korea in 1971. He received his BS degree with magna cum laude in Electrical Engineering from POSTECH (Pohang University of Science and Technology), Pohang, Korea in 1994. After attending the graduate program at POSTECH for a year and finishing his military service, he came to the University of Washington, Seattle, WA in 1997 and received his MSEE degree in Electrical Engineering in 1999. From 1999, he pursued his PhD in Electrical Engineering at Purdue University, West Lafayette, IN and received the degree in 2003. Since September 2003, he has been working in the compact modeling group at Intel, Hillsboro, OR.