國立交通大學

電控工程研究所

碩 士 論 文

矽奈米元件模擬器之發展

Development of Device Simulator for Future Nano Technology

研 究 生：林珀瑞

指導教授：渡邊浩志　教授

中 華 民 國 一 百 零 一 年 九 月

矽奈米元件模擬器之發展
Development of Device Simulator for Future Nano Technology

研 究 生：林珀瑞　　　　　Student：Po-Jui Lin

指導教授：渡邊浩志　　　　Advisor：Hiroshi Watanabe

國 立 交 通 大 學
電 控 工 程 研 究 所
碩 士 論 文

A Thesis

Submitted to Institute of Electrical Control Engineering

College of Electrical and Computer Engineering

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Electrical Control Engineering

September 2012

Hsinchu, Taiwan, Republic of China

中華民國一百零一年九月

矽奈米元件模擬器之發展

研究生: 林珀瑞　　　　　　　　　　指導教授: 渡邊浩志 教授

國立交通大學 電控工程研究所 碩士班

摘　要

目前，半導體元件製程技術不斷的進步，當元件尺寸小到奈米等級時，如何對元件中的載子運動做數值上的分析就顯得十分重要。在奈米尺寸等級的元件中，載子的傳輸行為必然與目前被廣泛採用於 Technology Computer-Aided Design (TCAD)中的漂移－擴散電流模型(drift-diffusion model)有很大的不同。我們的目標是建立一組可以適用於奈米元件的新的載子運動模型(不同於漂移－擴散電流模型)，來描述奈米元件中的載子運動行為。在這篇論文中，我們會先回顧在元件模擬領域中會使用到的物理模型及常被使用的數值分析方法。透過歷史的回顧，我們可以對於現今的元件模擬器是如何運作的以及對它們在使用上的限制有更深入的了解。接著我們會以奈米元件的觀點來探討目前所採用之分析方法的優劣。

在載子傳輸的模型裡面，波茲曼方程式(Boltzmann transport equation)是最為普遍的描述方程。但是由於它同時包含了連續項與不連續項，導致我們無法直接求得其解。於是在漂移－擴散電流模型中使用了平均場近似(mean-field approximation)的假設來消除不連續項。在此篇論文，我們先回顧了半導體物理中基本的物理模型，接著討論不同的數值分析方法，如：有限差分離散法、Scharfetter-Gummel 離散法、牛頓迭代法及 Gummel 迭代法。此外，我們會探討目前半導體元件物理的優缺點，接著討論對於開發通用奈米元件模擬器(General-Purpose Nano Device Simulator)的關鍵因素。在我接續的博士論文中，則會持續的開發通用奈米元件模擬器。

Development of Device Simulator for Future Nano Technology

Student: Po-Jui Lin   Advisor: Prof. Hiroshi Watanabe

Institute of Electrical Control Engineering,

National Chiao Tung University

Abstract

The computational study of electron devices physics is becoming more important, as the device scaling is advanced to lift the curtain on nano-devices technologies. The carrier transport phenomenon in nano-meter scaled devices (nano-devices) must be quite different from the drift-diffusion model which has played a central role of the Technology Computer-Aided Design (TCAD). Our final goal is to establish new device physics (beyond the drift-diffusion model) for the fundamental study of the carrier transport in nano-devices. In this thesis, we will review the history of the basic formulation and the numerical approaches of the device physics so far, which is useful to foreknow the essential ability of today's device modeling, and briefly survey the perspective of the future device modeling.

In particularly, the Boltzmann transport equation (BTE) is the most general equation in the history of the carrier transport study. Since BTE involves the discontinuous term that causes the equation unsolvable, the drift-diffusion model is established within the mean-field approximation for successfully removing this discontinuity [1]. In this thesis, we will review the physical model and the numerical recipes, such as the finite difference discretization, Scharfetter-Gummel discretization [2], Newton's method, and Gummel's iteration method [3]. In addition, we will discuss about the advantages and disadvantages of today's device physics, and then to reveal a key point for General-Purpose Nano Device Simulator. In the PhD dissertation to be continued from this Master thesis, we will develop General-Purpose Device Simulator.

Acknowledgement

This thesis would not have been possible without the help of several people. To all of them, I wish to express my gratitude.

First and foremost, I would like to acknowledge and extend my heartfelt gratitude to my advisor—Professor Hiroshi Watanabe, for his vital encouragement and patient guidance, generous assistance and priceless advice, all of which have been of inestimable worth to the completion of my thesis.

Secondly, my special thanks go to all teachers who have helped and taught me immensely during the two years of my study in National Chiao Tung University. And I would also like to thank all the classmates and friends who have given me generous support and helpful advice during the past years. They have provided me great help and comprehensive supervision through the two years. I have benefited a great deal from their advice and suggestions.

Last but not least, my thanks would also go to my beloved parents and elder sister for their boundless love and whole-hearted support over all these past years.
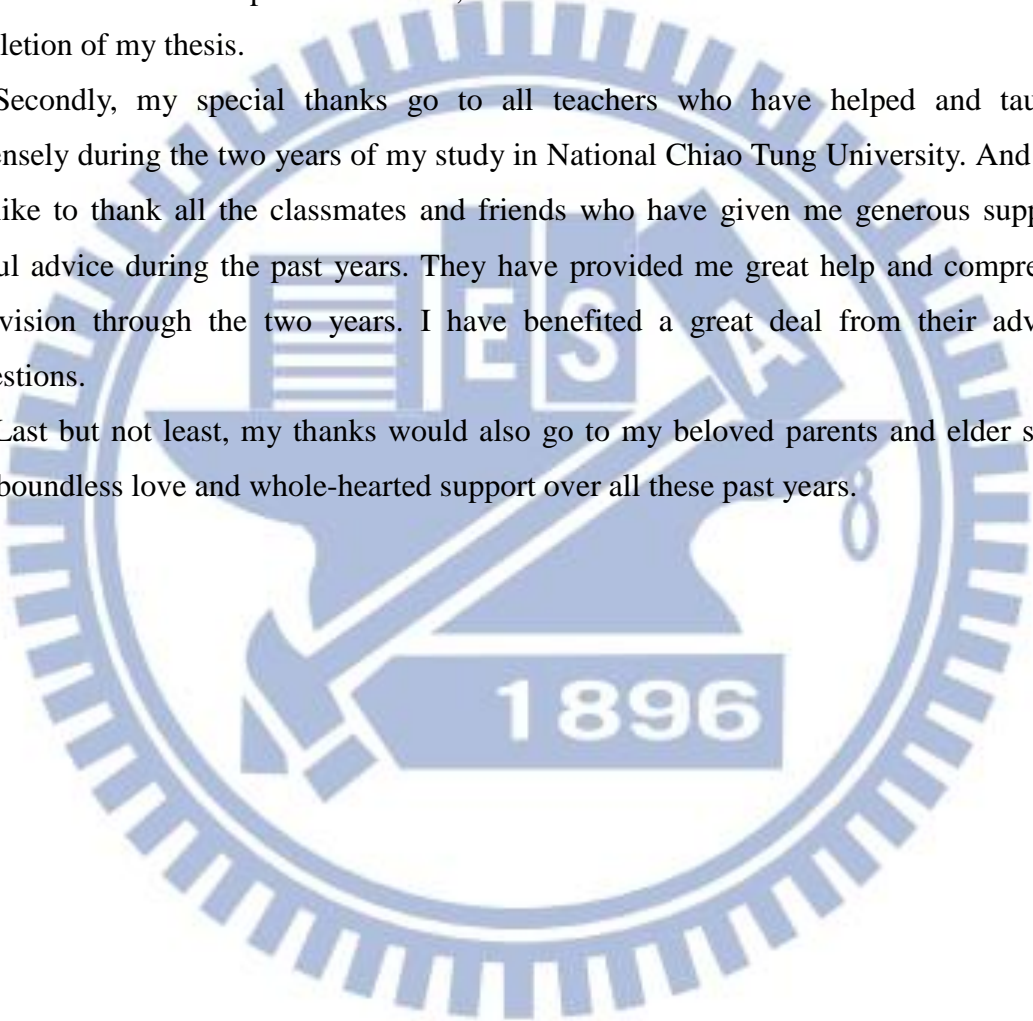
Table of Contents

List of Figures

Chapter 1.   Introduction

The first idea of the insulated-gate-field-effect transistor (IGFET) was invented by Lilienfeld [4] in his patent which can date back from 1926. However, the technology of these days was not mature enough to make sure the high-quality interface between the insulator and the semiconductor. In addition, the invention of bipolar transistor in 1947 [5] shifted most of research activities from IGFET to the potential-effect transistor (PET) [6]. This trend was changed since the metal-oxide-semiconductor filed-effect transistor (MOSFET) was firstly developed by Kahng and Atalla [7] in 1960. MOSFET has several advantages that hold well even in our day: 1) High speed switching, 2) Thermal stability (the current decrease with the temperature increase), and 3) Without offset voltage. After the invention of MOSFET by Kahng and Atalla [7], a rapid growth of the semiconductor industry started. The device dimension has been shrunk from the micrometer scale at its beginning to the present state-of-the-art, i.e., nano-meter scale. This is "BULK-to-NANO". Figure 1-1 shows the forecast for transistor physical gate length during 2011-2026 from the International Technology Roadmap for Semiconductor (ITRS) 2011 [8]. It is aimed that the physical gate length will reach 16nm, 10nm, and 6nm in 2017, 2020, and 2026, respectively. Here note that 10nm is just fit to the de Broglie Length (DBL), which is a spatial width of electron wave in silicon materials. Nowadays, the semiconductor device design has become complicated by the demand of more functions to the integrated circuits and the progressive innovations in process engineering. The trial-and-error routine of the device design, which causes the significant reduction of cost-performance, has become out of date. On the other hand, it is hopefully expected that Technology Computer-Aided Design (TCAD) will compensate this reduction of the cost-performance. However, most of the TCAD tools, nowadays, still need many empirically-determined fitting parameters. The number of empirical parameters is also increased as the device scaling is advanced; which causes TCAD itself complicated. Until 2017 (16nm), the transport mechanism shall become away from the drift-diffusion model annually and, as a result, further more fitting parameters shall be required to fit the measured data. Until 2020 (10nm), the transport mechanism shall become much closer to full-quantum and far away from the drift-diffusion and it would be a challenge to fit the measured data. After 2020 (beyond 10nm), the transport mechanism in nano-devices shall be fully quantum-mechanical and then completely different from the drift-diffusion, as illustrated in

Figure 1-2. We can no more find a slope on the current-voltage characteristic of MOSFET, which is necessary to determine the mobility, in nano-device characteristics. We can no more fit the calculation results assuming the drift-diffusion with the measured data. How to design LSI? We are facing a historical turning point of the device engineering. Most of the important innovations in TCAD were made in the 1960s. Firstly, the exponential dependence was the major problem on numerical computing during 1960s [9]. Gummel proposed a decoupled iterative method in 1964, for solving the three coupled partial differential equations: Poisson equation, the electron continuity equation and the hole continuity equation [3]. In 1969, Scharfetter and Gummel presented a discretization method for the hole and electron current continuity equations, which can greatly reduce the computational resource for the device simulation [2]. Throughout '70s' and '80s', subsequently, most activities were devoted to the physical phenomena caused by device scaling [10], such as: the velocity overshoot effect [11], the hot-electron effect [12], and the narrow gate width effect [13]. These newly coming effects caused the researchers of the day to develop three-dimensional simulation [14] and Monte Carlo method [15].

The drift-diffusion (DD) model, which is adopted for the current equations of electron and hole has reproduced the measured device operation quite well so far. Nevertheless, we have to note the following matters: 1) DD model assumes the bulk body of semiconductor substrate. 2) We are reaching the period of nano-devices, as shown in Figure 1-1. Accordingly, whole new physical models must be implemented to a 3D general-purpose device simulator for nano-devices; meanwhile keeping the robustness and using less computational resources is also demanded.
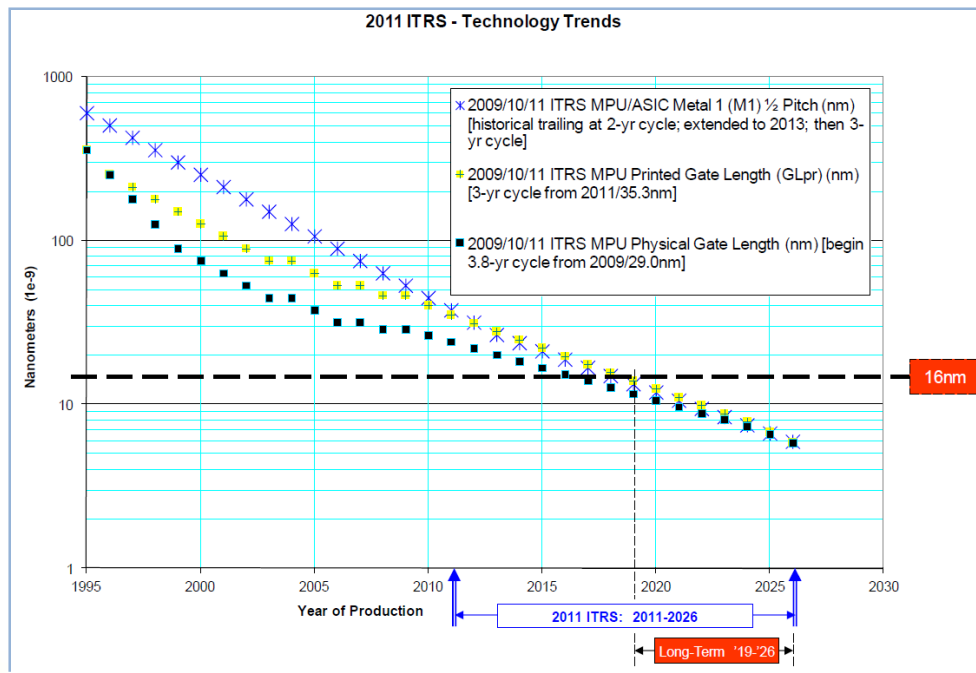
Figure 1-1   ITRS 2011 Roadmap: MPU/high-performance ASIC Half Pitch and Gate
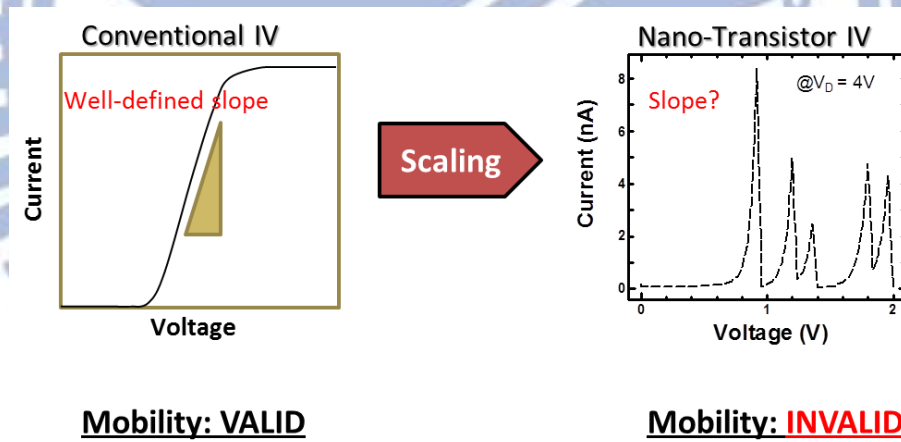
Length Trends

Figure 1-2   Turning Point of IV-characteristics (from Lecture Note of Prof. Watanabe)

Chapter 2.   Transport Theory

2.1 Boltzmann Transport Equation

Let us regard carriers as classical particles; which dynamics are completely determined by assigning their positions $\vec{r}$ and momentums $\vec{p}$ at time t. Here, it is not necessary to consider about the size of carriers/particles. Next, let us consider a small region in the phase space about the point (x, y, z, p_x, p_y, p_z), as shown in Figure 2-1, and the distribution function $f(\vec{r}, \vec{p}, t)$ which represents the probability to find a particle within this small region in phase space. The distribution function will change by df during a time period dt as follows:

$$df = f(x, y, z, p_x, p_y, p_z)$$

$$-f(x + v_x dt, y + v_y dt, z + v_z dt, p_x + F_x dt, p_y + F_y dt, p_z + F_z dt), \quad (1)$$

where $\vec{v}$ is the velocity of the considered carrier and $\vec{F}$ is the force field affecting on particle. Assuming the period dt is infinitesimally small and the distribution function is continuous in phase space, we can obtain:

$$\frac{df}{dt} + \vec{v} \cdot \nabla_r f + \vec{F} \cdot \nabla_p f = 0, \quad (2)$$

or using the k-representation:

$$\frac{df}{dt} + \vec{v} \cdot \nabla_r f + \frac{\vec{F}}{\hbar} \cdot \nabla_k f = 0. \quad (3)$$

Points in phase space express dynamical states. These points can enter into and be emitted out of a region while satisfying this equation just like a hydrodynamic flow. Generally speaking, we can involve scattering process into this equation. With the rate of change of distribution function due to scattering $\left(\frac{\partial f}{\partial t}\right)_{coll}$, the distribution function turns out to be satisfied to the equation:

$$\frac{df}{dt} + \vec{v} \cdot \nabla_r f + \frac{\vec{F}}{\hbar} \cdot \nabla_k f = \left(\frac{\partial f}{\partial t}\right)_{coll}. \quad (4)$$

The right-hand side is called "collision-term", which can break the continuity and cannot be defined uniquely. Here, it is noteworthy to say that the BTE is a strange equation. Firstly,

the left hand side should be continuous since it is written in terms of differentiation. Secondly, the right hand side can be discontinuous. In general, a discontinuous quantity cannot be equal to any continuous quantity. Accordingly, the treatment of the collision term has been a most sensitive issue in the history of modern physics. In terms of scattering theory, the contribution of scattering processes from a state k to another state k', and vice versa, to the collision term is written as:

$$S_{k' \to k} O_{k'} (1 - O_k) - S_{k \to k'} O_k (1 - Of_{k'}),$$ (5)

where the $S_{k' \to k}$, and $S_{k \to k'}$, represent the transition rates from k' to k and from k to k', respectively. Here O_k depicts the occupation rate of state k. According to the Fermi's Golden Rule [16], we can write:

$$S_{k \to k'} = \frac{2\pi}{\hbar} |\langle k|V|k' \rangle|^2 \delta(E_k - E_{k'}),$$ (6)

where V is the considered scattering potential, and E_k and $E_{k'}$ are the initial and final state energies of the particle that scatters.

We must take into account more cases of contributions to the collision term. Most important in engineering is the process in which the number of particles is changed. For example, in electron tunneling phenomena, an electron disappears in anode where the tunneling starts, while an electron appears in cathode where the tunneling ends. The tunneling of electron decreases the number of electrons in anode and increase the number of electron in cathode. As long as we assume the energy band in semiconductor, the pair-creation and pair-annihilation of electron and hole can appear; which is called Shockley-Read-Hall (SRH) process. Since the creation of hole means the annihilation of valence electron, the total number of electrons in conduction band and valence band is unchanged. However, once we formulated the basic equations of electrons and holes separately, we must consider that the number of electrons (negative carriers; i.e., conduction electron) and the number of holes (positive carrier; a lack of valence electron) are increased at the same moment by SRH process. Moreover, we should note that the initial and end points of tunneling phenomena are regarded as spatially different, while no position change occurs in SRH process. Auger process can also increase the number of carriers.
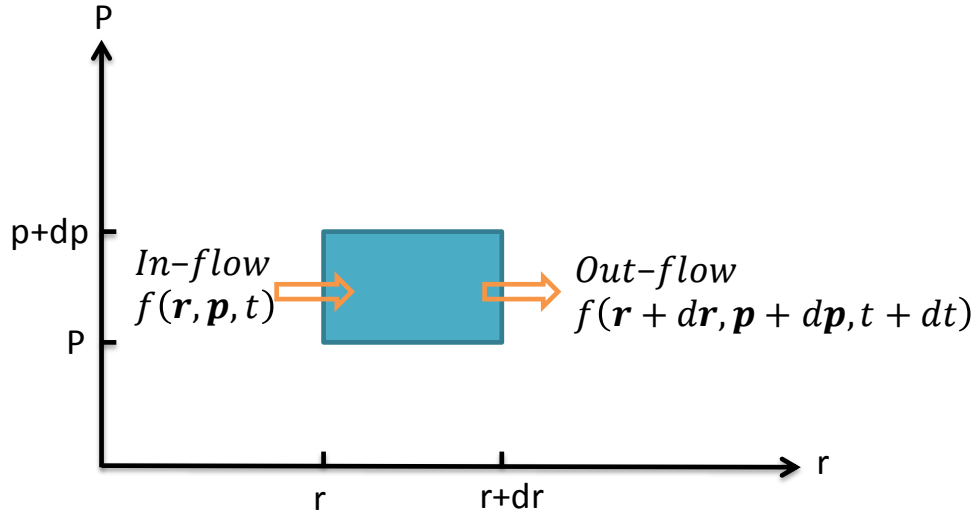
P

p+dp

In-flow
$f(\boldsymbol{r}, \boldsymbol{p}, t)$

P

Out-flow
$f(\boldsymbol{r} + d\boldsymbol{r}, \boldsymbol{p} + d\boldsymbol{p}, t + dt)$

r

r   r+dr

Figure 2-1   Carrier Transport in Phase Space

2.2 Relaxation-Time Approximation

The Boltzmann equation is very difficult to solve, because it has both the continuous terms (related to the differential of the distribution function) and the discontinuous term (the collision term). In order to solve it, we have to make some simplification. Within the first order approximation, the most well-known is the relaxation-time approximation (RTA). In this approximation, we can replace the collision term with $(f^0 - f)/\tau$, where f^0 is the distribution function at equilibrium condition and τ is the so-called relaxation time that specifies the time needed for the system to relax to its equilibrium state. Through a plenty of collisions, the distribution function shall be relaxed to the equilibrium state, i.e., equilibrated. In RTA, all the collisions are averaged and the essence of the collisions is involved to the relaxation time. We can no more discuss about each collision. In other words, the RTA gives us the view point of independent particle (with no collision). Here, we can distinguish three different types of relaxation times: 1) The collision time, which represents the average time to the next collision. This means that all primitive processes are averaged by the collision time, which is the same as the mean-free time in mean-field theory. The concept of this relaxation is valid with or without the conservation laws of energy and momentum. 2) The momentum relaxation time, which is the time needed for relaxing the momentum of a considered particle (individual from others within mean-field theory). Let us consider a particle with momentum higher than the average. The momentum of this particle is relaxed progressively with time.

This relaxation time is usually longer than the collision time, because it needs a plenty of collisions to relax the momentum of the considered particle. 3) The energy relaxation time, which is the time needed for relaxing the energy of considered particle. Let us consider a particle with energy higher than the average. The energy of this particle is relaxed progressively with time. This relaxation time is usually longer than the collision time, because it needs a plenty of collisions to relax the energy of considered particle. To consider all possibilities of the relaxations mentioned above from the view point of independent particle, we need to sum up all the inverse relaxation times to get the total relaxation time: $1/\tau_t = \sum_i 1/\tau_k$, where the subscript k represents the different relaxation view point, that is, k=1) collision, k=2) momentum relaxation and k=3) energy relaxation.

To see more about the meaning of relaxation time, we assume a condition without external field and within a homogeneous semiconductor material. The BTE is reduced to:

$$\frac{\partial f}{\partial t} = -\frac{f - f^0}{\tau_k}, where \nabla_r f = 0 \ and \ \vec{F} = 0. \tag{7}$$

Solving this differential equation with assuming the initial state of distribution function to be f^i, we can obtain:

$$f = f^0 + (f^i - f^0)e^{-t/\tau_k}. \tag{8}$$

Here note that this result is valid for any k, i.e. the classification of relaxation. And we assume the relaxation time is independent of the velocity, position, and time.

Chapter 3.   Drift-Diffusion model

3.1 Drift-Diffusion Model Derivation

We can derive the drift-diffusion current equations from the BTE easily. The general expression of current equation is $\vec{J} = e \int \vec{v} f(\vec{r}, \vec{v}, t) d^3 v$, where $e = -q$ for electrons and $e = q$ for holes, and q is the elementary charge. Firstly, let us consider a steady-state condition ($\frac{\partial f}{\partial t} = 0$), for simplicity, within 1D geometry. Using the relaxation-time approximation discussed above and the effective mass approximation ($\vec{p} = m^* \vec{v}$), the BTE is reduced to:

$$v_x \frac{\partial f}{\partial x} + \frac{eE_x}{m^*} \frac{\partial f}{\partial v_x} = \frac{f^0 - f}{\tau}, \tag{9}$$

where, E_x is the electric field and we have used $F_x = eE_x$. Here note that the arguments of the distribution function are: the velocity, $\vec{v}$, and the location, x. The time, t, has been eliminated at steady state.

Thereby, we can obtain:

$$f = f^0 - \left[\tau v_x \frac{\partial f}{\partial x} + \frac{\tau e E_x}{m^*} \frac{\partial f}{\partial v_x} \right]. \tag{10}$$

Substituting this into the current equation, we can obtain:

$$J(x) = e \int_{-\infty}^{+\infty} v_x f^0 dv_x - e \int_{-\infty}^{+\infty} \left[\tau v_x^2 \frac{\partial f}{\partial x} + \frac{v_x \tau e E_x}{m^*} \frac{\partial f}{\partial v_x} \right] dv_x. \tag{11}$$

Since the equilibrium distribution function f^0 is an even function, it is symmetrical in $\vec{v}$. The integral of the first term in the current equation must thereby be zero, that is, $v_x f^0$ is an odd function and then the integration must be zero over an even integration interval. Since the relaxation time is assumed to be independent of $\vec{v}$, we have the following expression:

$$J(x) = -e \int_{-\infty}^{+\infty} \tau v_x^2 \frac{\partial f(v_x, x)}{\partial x} dv_x - e \int_{-\infty}^{+\infty} \frac{v_x \tau e E_x}{m^*} \frac{\partial f(v_x, x)}{\partial v_x} dv_x$$

$$= -e\tau \frac{\partial}{\partial x} \int_{-\infty}^{+\infty} v_x^2 f(v_x, x) dv_x - \frac{e^2 \tau E_x}{m^*} \int_{-\infty}^{+\infty} v_x df(v_x, x). \tag{12}$$

The first term in the right hand side is rewritten as:

$$\int_{-\infty}^{+\infty} v_x{}^2 f(v_x, x) dv_x \rightarrow \langle v_x{}^2 \rangle c(x), \tag{13}$$

where $\langle v_x{}^2 \rangle$ is the average of the square of the velocity and $c(x)$ stands for the carrier concentration, $c(x) = \int_{-\infty}^{+\infty} f(v_x, x) dv_x$. According to the energy equipartition rule: $\frac{m^*}{2}\langle v_x{}^2 \rangle = \frac{1}{2} k_B T$ with T being the absolute temperature, the first term turns out to be:

$$
\begin{aligned}
-e\tau \frac{\partial}{\partial x}\{\langle v_x{}^2 \rangle c(x)\} &= -e\tau \langle v_x{}^2 \rangle \frac{\partial}{\partial x}\{c(x)\} \\
&= -\frac{e\tau}{m^*} k_B T \frac{\partial c(x)}{\partial x} \\
&= -\frac{e}{q} \mu k_B T \frac{\partial c(x)}{\partial x} \\
&= -eD \frac{d}{dx} c(x).
\end{aligned}
\tag{14}
$$

Here we assume: $\mu = q\tau/m^*$, i.e. the mobility, and the Einstein relation, $D/\mu = k_B T/q$, where $e = -q$ for electrons and $e = q$ for holes. It should be noted that T concerns the temperature of electron-hole gases in DD model with regardless of any lattice motion. This means DD model itself cannot involve any phonon process. For the second term of the right hand side, by using integration by parts, we can obtain:

$$-\frac{q^2 \tau E_x}{m^*} \int_{-\infty}^{+\infty} v_x df(v_x, x) = -\frac{q^2 \tau E_x}{m^*}\left[v_x f|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} f dv_x\right], \tag{15}$$

where $e^2 = q^2$. Assuming $f(+\infty, x) = f(-\infty, x) = 0$ and $\int_{-\infty}^{+\infty} f dv_x = c(x)$, the second term turns out to be:

$$-\frac{e^2 \tau E_x}{m^*} \int_{-\infty}^{+\infty} v_x df(v_x, x) \rightarrow q\mu E_x c(x). \tag{16}$$

Finally, we have the current density equation:

$$J(x) = q\mu E_x c(x) - eD \frac{d}{dx} c(x). \tag{17}$$

Substituting $e = -q$, $\mu = \mu_n$, and $D = D_n$ for electrons and $e = q$, $\mu = \mu_p$, and $D = D_p$ for holes, we can obtain the drift-diffusion equations for electrons and holes

$$J_n(x) = qn(x)\mu_n E_x + qD_n \frac{\partial n(x)}{\partial x}. \tag{18}$$

$$J_p(x) = qp(x)\mu_p E_x - qD_p \frac{\partial p(x)}{\partial x}. \tag{19}$$

3.2 Continuity Equation

Using the law of particle conservation, we derive the continuity equation from BTE:

$$\frac{df(\vec{r},\vec{k},t)}{dt} + \vec{v} \cdot \nabla_r f(\vec{r},\vec{k},t) + \frac{\vec{F}}{\hbar} \cdot \nabla_k f(\vec{r},\vec{k},t) = \left(\frac{\partial f(\vec{r},\vec{k},t)}{\partial t}\right)_{coll}. \tag{20}$$

Integrating (20) over all k space, we then obtain:

$$\iiint \frac{\partial f}{\partial t} d\vec{k} + \iiint \vec{v} \cdot \nabla_r f d\vec{k} + \iiint \frac{\vec{F}}{\hbar} \cdot \nabla_k f d\vec{k} = \iiint \left(\frac{\partial f}{\partial t}\right)_{coll} d\vec{k}. \tag{21}$$

The second term in the left hand side of (21) is reduced to:

$$\iiint \vec{v} \cdot \nabla_r f d\vec{k} = \iiint \nabla_r \cdot (\vec{v}f) \, d\vec{k} - \iiint f\nabla_r \cdot \vec{v} \, d\vec{k}. \tag{22}$$

Since $\vec{v}$ is explicitly independent of $\vec{r}$, this term $\nabla_r \cdot \vec{v}$ will be zero; which results in:

$$\iiint \vec{v} \cdot \nabla_r f \, d\vec{k} = \nabla_r \cdot \iiint \vec{v}f(\vec{r},\vec{k},t) \, d\vec{k} = \nabla_r \cdot \left(\frac{1}{e}\vec{J}\right), \tag{23}$$

where $\vec{J}$ is an electric current. Assuming that the distribution function f is zero while the wavenumber is very high, the third term in the left hand side of (21) turns out to be:

$$\iiint \frac{\vec{F}}{\hbar} \cdot \nabla_k f \, d\vec{k} = \frac{1}{\hbar} \iiint (F_x, F_y, F_z) \cdot \left(\frac{\partial f}{\partial k_x}, \frac{\partial f}{\partial k_y}, \frac{\partial f}{\partial k_x}\right) d\vec{k}$$
$$= \frac{1}{\hbar} \iiint \left(F_x \frac{\partial f}{\partial k_x} + F_y \frac{\partial f}{\partial k_y} + F_z \frac{\partial f}{\partial k_x}\right) dk_x dk_y dk_z = 0. \tag{24}$$

Here we also have used the integration by parts, in the similar manner to (15) and note that the force is independent of the momentum. The integration of the collision term, i.e., the right hand side of (21), can be a-priori replaced by the generation-recombination term as follows:

$$\iiint \left(\frac{\partial f}{\partial t}\right)_{coll} d\vec{k} = G - R. \tag{25}$$

Thereby, we obtain the following equation:

$$\frac{\partial}{\partial t} c(\vec{r}, t) + \frac{1}{e} \nabla_r \cdot \vec{J} = G - R, \tag{26}$$

where we have adopted:

$$\iiint f \, d\vec{k} \equiv c(\vec{r}, t). \tag{27}$$

Regarding that $c(x) = n(x)$, which is the electron density, for electrons, and $c(x) = p(x)$, which is the hole density, for holes, and note that $e = -q$ for electrons and $e = q$ for holes. From (26), we can have the current continuity equation for electrons and holes as follows:

$$\frac{\partial n}{\partial t} - \frac{1}{q} \nabla_r \cdot \vec{J}_n = G_n - R_n, \tag{28}$$

$$\frac{\partial p}{\partial t} + \frac{1}{q} \nabla_r \cdot \vec{J}_p = G_p - R_p, \tag{29}$$

Where G_n and G_p are the generation rates of electrons and holes, respectively, and R_n and R_p are the recombination rates of electrons and holes, respectively. $\vec{J}_n$ and $\vec{J}_p$ are the electric currents of electrons and holes, respectively, which are described by the drift-diffusion model.

As mentioned above, we replaced the collision term with the relaxation term, $(f^0 - f)/\tau$, within the first order approximation, i.e., RTA. Subsequently, we appended the generation-recombination terms as a correction to the relaxation term by hand. This correction is no more than a part of the second and higher order terms of the collision term, but it was carefully deduced to catch the physical intuition related to some deviations from the drift-diffusion model. However, as the device scaling is progressed, we shall be forced to involve more correction terms that may be discontinuous. The discontinuity is enhanced with the scaling, and then, once entering the period of nano-devices, the discontinuous terms must prevail over the continuous term. As long as the drift-diffusion model is adopted, we cannot find any solution in nano-devices.

3.3 Poisson Equation

We can derive the Poisson equation from the differential form of Gauss's Law,

$$\nabla \cdot \left(\epsilon \vec{E}\right) = \rho, \tag{30}$$

where ρ is the charge density and ϵ is the permittivity. With the definition of the electric field ($\vec{E}$):

$$\vec{E} = -\nabla \psi, \tag{31}$$

here ψ represents the static potential, we can combine these two equations and assuming ϵ is a constant to obtain

$$\nabla \cdot \nabla \psi = -\frac{\rho}{\epsilon}. \tag{32}$$

In a semiconductor material, we can express the charge density as

$$\rho = q(p - n + N_D^+ - N_A^-), \tag{33}$$

where N_D^+ is the ionized donor concentration and N_A^- is the ionized acceptor concentration.

By the above equations, the Poisson equation in a homogeneous semiconductor is given by,

$$\nabla^2 \psi = -\frac{q}{\epsilon}(p - n + N_D^+ - N_A^-). \tag{34}$$

In Cartesian form,

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} = -\frac{q}{\epsilon}(p - n + N_D^+ - N_A^-). \tag{35}$$

Chapter 4. Numerical Solutions

4.1 Concept of Quasi-Equilibrium

The Poisson equation, the continuity equations for electrons and holes with the drift-diffusion models are simultaneous equations that have three variables, ψ, n, and p.

$$\nabla^2 \psi = -\frac{q}{\epsilon}(p - n + N_D^+ - N_A^-). \tag{36}$$

$$\nabla \cdot (n\mu_n \nabla \psi - D_n \nabla n) = G_n - R_n. \tag{37}$$

$$\nabla \cdot \left(-p\mu_p \nabla \psi - D_p \nabla p\right) = G_p - R_p. \tag{38}$$

However, there are some numerical issues about using this set of variables, (ψ, n, p), to simulate semiconductor devices. One of them is the exponential dependency of carrier concentrations, n and p, on the electrostatic potential, ψ; which increases the risk of numerical errors in calculation results.

In terms of mass-action law, we can write the relation between the carrier concentrations and the Fermi level, E_F as below:

$$n = n_i \exp\left(\frac{E_F - E_i}{k_B T}\right), \tag{39}$$

$$p = n_i \exp\left(\frac{E_i - E_F}{k_B T}\right), \tag{40}$$

where n_i is the intrinsic density and E_i is the mid-gap energy [17]. Let us assume that the origin of the potential is the mid-gap energy in the p-type region across the p-n junction shown in Figure 4-1 and then we can write:

$$E_i = -q\psi. \tag{41}$$

As shown in the upper illustration in Figure 4-1, E_i is decreased by the build-in-potential multiplied by q across the p-n junction at equilibrium (with no bias). At the same moment, ψ is increased by the build-in-potential across the p-n junction, as shown in the bottom illustration. Next, let us introduce the Fermi level[1] as follows:

---

[1] Let us call both E_F and φ "Fermi level" in this thesis.

$$\varphi = -\frac{E_F}{q}. \tag{42}$$

Substituting (41) and (42) to (39) and (40), we can obtain:

$$n = n_i \exp\left(\frac{q(\psi - \varphi)}{k_B T}\right), \tag{43}$$

$$p = n_i \exp\left(\frac{q(\varphi - \psi)}{k_B T}\right). \tag{44}$$

It is therefore found that the densities of electrons and holes are equal to n_i while the Fermi level coincides to the mid-gap in the intrinsic semiconductor. Deviation of ψ from φ will increase the electron density and hole density by factors $\exp(q(\psi - \varphi)/k_B T)$ and $\exp(q(\varphi - \psi)/k_B T)$, respectively
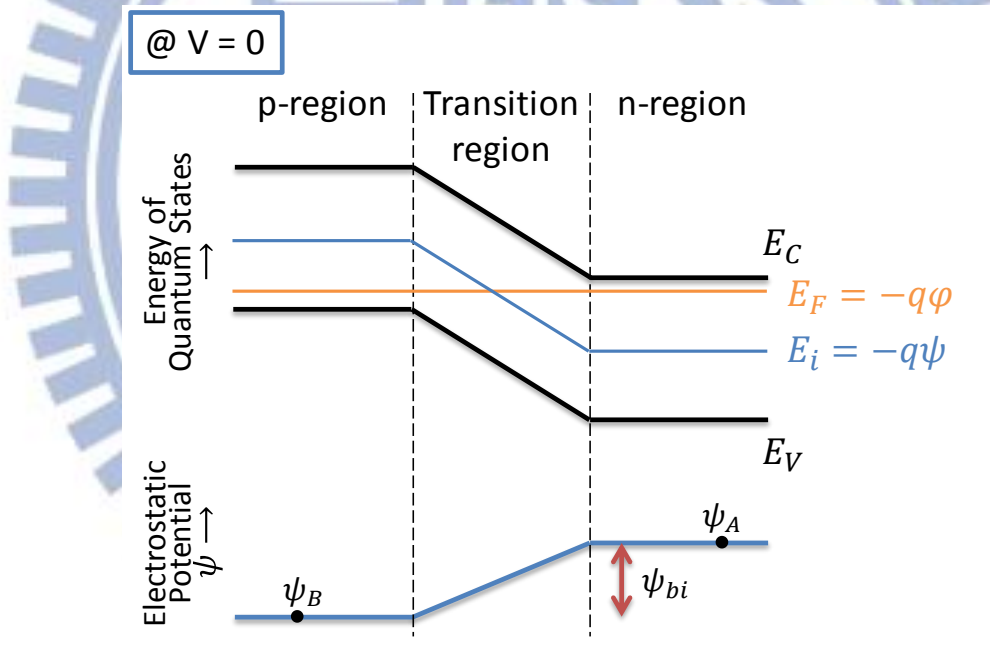
Figure 4-1   Energy Band Diagram and Electrostatic Potential of a p-n Junction under Equilibrium

When a forward voltage (V_F) is applied to a p-n junction, as shown in Figure 4-2, the equilibrium condition is broken and then the electron-hole system will seek for a secondary statistical balance, which is called the quasi-equilibrium state. In this event, we may define the quasi-Fermi level separately for the electron density and the hole density as follows:

$$n \equiv n_i \exp\left(\frac{E_{Fn} - E_i}{k_B T}\right), \tag{45}$$

$$p \equiv n_i \exp\left(\frac{E_i - E_{Fp}}{k_B T}\right), \tag{46}$$

or

$$n \equiv n_i \exp\left(\frac{q(\psi - \varphi_n)}{k_B T}\right), \tag{47}$$

$$p \equiv n_i \exp\left(\frac{q(\varphi_p - \psi)}{k_B T}\right), \tag{48}$$

where $E_{Fn} = -q\varphi_n$ and $E_{Fp} = -q\varphi_p$ are the quasi-Fermi levels for electrons[2] and holes[3], respectively.

From the above equations, we can derive the quasi-Fermi levels for electrons and holes as follows:

$$\varphi_n \equiv \psi - \frac{k_B T}{q} \ln\left(\frac{n}{n_i}\right), \tag{49}$$

$$\varphi_p \equiv \psi + \frac{k_B T}{q} \ln\left(\frac{p}{n_i}\right), \tag{50}$$

respectively. Figure 4-2 illustrates the E_{Fn} and E_{Fp} throughout the p-n junction under quasi-equilibrium condition with forward bias, V_F. It is noteworthy that the Fermi level E_F is defined under equilibrium condition, while $E_{Fn} - E_{Fp} = qV_F$ under quasi-equilibrium condition with V_F.

Using φ_n, φ_p, eq. (47), and the Einstein relationship, the current density equation for electron turns out to be:

---

[2] Let us call both E_{Fn} and φ_n "Fermi level" in this thesis.

[3] Let us call both E_{Fp} and φ_p "Fermi level" in this thesis.

$$\vec{J}_n = qn\mu_n\vec{E} + qD_n\nabla n = -qn\mu_n\nabla\psi + qD_n\nabla\left(n_i\exp\left(\frac{q(\psi-\varphi_n)}{k_BT}\right)\right)$$

$$= -qn\mu_n\nabla\psi + qD_n n_i\frac{q}{k_BT}\exp\left(\frac{q(\psi-\varphi_n)}{k_BT}\right)\nabla(\psi-\varphi_n)$$

$$= -qn\mu_n\nabla\psi + q\mu_n n\nabla(\psi-\varphi_n) = -q\mu_n n\nabla\varphi_n. \tag{51}$$

In similar way, we can derive:

$$\vec{J}_p = -q\mu_p p\nabla\varphi_p. \tag{52}$$

The equations (51) and (52) are quite similar to the current density in conductor:

$$\vec{J} = -q\mu n\nabla\psi. \tag{53}$$

Now we can use the electron and hole quasi-Fermi levels to define a new set of the equations. Here, ψ is the electrostatic potential that can be defined at any location of considered semiconductor material. Substituting (47) and (48) to (36), we can obtain:

$$\nabla^2\psi = -\frac{qn_i}{\epsilon}\left(\exp\left(\frac{\varphi_p-\psi}{V_T}\right) - \exp\left(\frac{\psi-\varphi_n}{V_T}\right) + N_D^+ - N_A^-\right). \tag{54}$$

here $V_T = k_BT/q$ represents the thermal voltage.

Substituting (47) to (37), we can obtain:

$$\nabla\cdot\left(\mu_n n_i\exp\left(\frac{\psi-\varphi_n}{V_T}\right)\nabla\varphi_n\right) = G_n - R_n. \tag{55}$$

Substituting (48) to (38), we can obtain:

$$\nabla\cdot\left(-\mu_p n_i\exp\left(\frac{\varphi_p-\psi}{V_T}\right)\nabla\varphi_p\right) = G_p - R_p. \tag{56}$$

Furthermore, we can change the choice of variables according to Slotboom [18] as follows:

$$\phi_n = \exp\left(\frac{-\varphi_n}{V_T}\right), \tag{57}$$

$$\phi_p = \exp\left(\frac{\varphi_p}{V_T}\right). \tag{58}$$

Substituting (57) and (58) into (54), we can obtain:

$$\nabla^2\psi = -\frac{qn_i}{\epsilon}\left(\phi_p\exp\left(\frac{-\psi}{V_T}\right) - \phi_n\exp\left(\frac{\psi}{V_T}\right) + N_D^+ - N_A^-\right). \tag{59}$$

Substituting (57) to (55), then we obtain:

$$\nabla\cdot\left(-n_iD_n\exp\left(\frac{\psi}{V_T}\right)\nabla\phi_n\right) = G_n - R_n. \tag{60}$$

Here we have used: $\mu_n = D_n/V_T$, i.e., the Eisnstein relationship.

Substituting (58) to (56), then we obtain:

$$\nabla\cdot\left(-n_iD_p\exp\left(\frac{-\psi}{V_T}\right)\nabla\phi_p\right) = G_p - R_p. \tag{61}$$

Here we have used: $\mu_n = D_n/V_T$, i.e., the Einstein relationship.

In general, these three sets of different variables are the mostly possible choices that are preferred for device simulation under different boundary conditions.
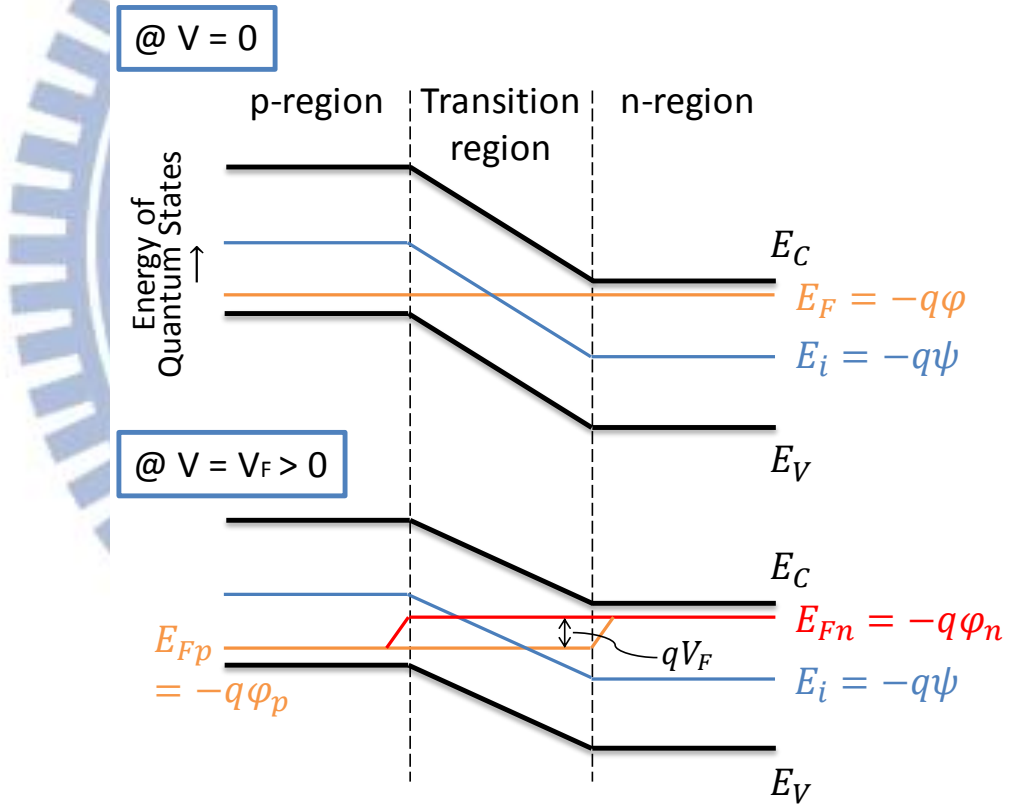

Figure 4-2   Energy Band Diagram under Equilibrium and Forward Biased

4.2 Discretization of the Poisson Equation

We can discretize the Poisson equation using a finite-difference scheme [19]. Considering Taylor's expansion in a case of one-dimension (1D),

$$f(x_i + \Delta x) = f(x_i) + f'(x_i)(\Delta x) + \frac{f''(x_i)}{2}(\Delta x)^2 + \cdots, \tag{62}$$

$$f(x_i - \Delta x) = f(x_i) - f'(x_i)(\Delta x) - \frac{f''(x_i)}{2}(\Delta x)^2 - \cdots, \tag{63}$$

where $f'(x_i)$ and $f''(x_i)$ represent the first and second derivative of the function $f(x)$ at point x_i, respectively. Subtracting (62) from (63) and ignoring the high order terms, we can approximate the first and second derivatives as

$$f'(x_i) \cong \frac{f(x_i + \Delta x) - f(x_i - \Delta x)}{2(\Delta x)}, \tag{64}$$

$$f''(x_i) \cong \frac{f(x_i + \Delta x) + f(x_i - \Delta x) - 2f(x_i)}{(\Delta x)^2}, \tag{65}$$

respectively. If a nonuniform rectangular mesh is adopted, i.e. $h_{i+1} \neq h_i$, as depicted in Figure 4-3, the approximations become

$$f'(x_i) \cong \frac{f(x_{i+1}) - f(x_{i-1})}{h_{i+1} + h_i}, \tag{66}$$

$$f''(x_i) \cong \frac{2}{h_{i+1} + h_i}\left[\frac{f(x_{i+1}) - f(x_i)}{h_{i+1}} + \frac{f(x_{i-1}) - f(x_i)}{h_i}\right], \tag{67}$$

where $h_i \equiv x_i - x_{i-1}$.

With the finite-difference discretization [19] in a case of three-dimension (3D), we can approximate the left hand side of (59) at a mesh point (i, j, l) as follows:

$$\left.\frac{\partial^2\psi}{\partial x^2}\right|_{i,j,l} \cong \frac{2}{h_{i+1} + h_i}\left[\frac{\psi_{i+1,j,l} - \psi_{i,j,l}}{h_{i+1}} + \frac{\psi_{i-1,j,l} - \psi_{i,j,l}}{h_i}\right], \tag{68}$$

$$\left.\frac{\partial^2\psi}{\partial y^2}\right|_{i,j,l} \cong \frac{2}{k_{j+1} + k_j}\left[\frac{\psi_{i,j+1,l} - \psi_{i,j,l}}{k_{j+1}} + \frac{\psi_{i,j-1,l} - \psi_{i,j,l}}{k_j}\right], \tag{69}$$

$$\left.\frac{\partial^2\psi}{\partial z^2}\right|_{i,j,l} \cong \frac{2}{w_{l+1} + w_l}\left[\frac{\psi_{i,j,l+1} - \psi_{i,j,l}}{w_{l+1}} + \frac{\psi_{i,j,l-1} - \psi_{i,j,l}}{w_l}\right], \tag{70}$$

where $k_j \equiv y_j - y_{j-1}$ and $w_l \equiv z_l - z_{l-1}$.

Thereby, the discretized Poisson equation turns out to be:

$$\frac{2}{h_{i+1} + h_i} \left[\frac{\psi_{i+1,j,l} - \psi_{i,j,l}}{h_{i+1}} + \frac{\psi_{i-1,j,l} - \psi_{i,j,l}}{h_i} \right]$$

$$+ \frac{2}{k_{j+1} + k_j} \left[\frac{\psi_{i,j+1,l} - \psi_{i,j,l}}{k_{j+1}} + \frac{\psi_{i,j-1,l} - \psi_{i,j,l}}{k_j} \right]$$

$$+ \frac{2}{w_{l+1} + w_l} \left[\frac{\psi_{i,j,l+1} - \psi_{i,j,l}}{w_{l+1}} + \frac{\psi_{i,j,l-1} - \psi_{i,j,l}}{w_l} \right]$$

$$= -\frac{q n_i}{\epsilon} \left(\phi_{p(i,j,l)} \exp\left(\frac{-\psi_{i,j,l}}{V_T} \right) - \phi_{n(i,j,l)} \exp\left(\frac{\psi_{i,j,l}}{V_T} \right) + N^+_{D(i,j,l)} \right.$$

$$\left. - N^-_{A(i,j,l)} \right). \tag{71}$$

In the above, we have assumed the permittivity is constant to discretize the Poisson equation. However, the permittivity is varying over several angstroms and more (around one nanometer) [22]. The constant permittivity assumption is invalid in General-Purpose Nano Device simulation. To follow up the permittivity change inside nano-devices, it is preferable to replace the left-hand side of (71) with the discretized surface integral of continuous flux under the finite-difference scheme [19].
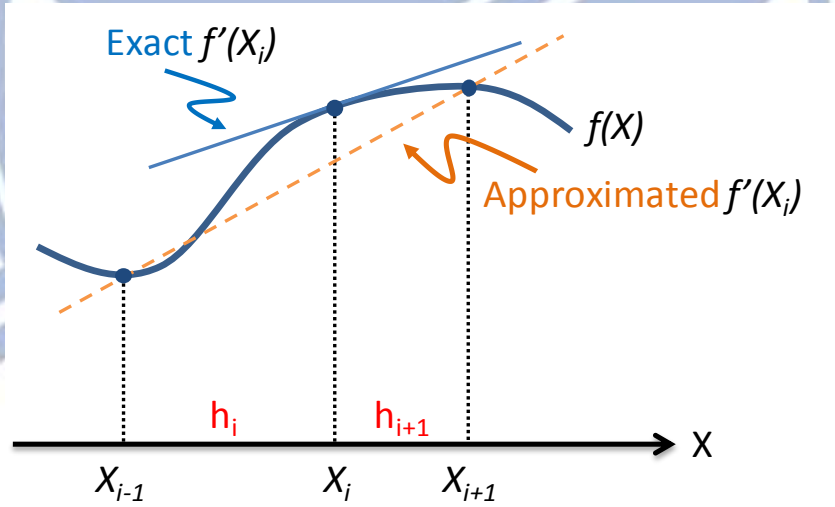
Figure 4-3   Graphical Illustration of Finite Difference Discretization

4.3 Control Volume

Generally, the continuous space coordinate can be discretized using the finite-difference scheme [19], in which discretization cell is called "control volume". Continuous matters can also be divided according to the set of control volumes. We must then carefully consider how the continuity of matters can be confirmed at each control volume.

Hydrodynamics simulation has been extensively adopted to study the dynamics of continuous matters. In semiconductor device physics, the drift-diffusion model is adopted to describe the continuity of the carrier transport. Moreover, the generation-recombination term is added, and can then break the continuity. Nevertheless, assuming that the elementary charge (causing the discontinuity) is too small to be detected in the carrier transport phenomena; we have treated the device simulation using hydrodynamic formulation with the control volume.

Here note that the drift-diffusion model is confined within the divergence of the continuity equation, as shown in (55) and (56). Let us consider the numerical fluxes that will inject to and ejected from a control volume, in which the divergence theorem is satisfied as follows:

$$\iint_s \nabla \cdot \vec{F} dA = \oint_{\partial s} \vec{F} \cdot \vec{n} ds,$$ (72)

where s represents the control volume, ∂s represents the boundary of the control volume, as illustrated in Figure 4-4, dA represents the volume element of the control volume, $\vec{F}$ is the flux vector related to the control volume, and ds represents the surface element on the boundary of the control volume. We can thereby replace the divergence of the vector function with the surface integral on the boundary of the control volume. This is the key to discretize Poisson equation and the continuity equations for electrons and holes [20].

We have three independent variables in the formulation, for example, potential, electron density, and hole density. First of all, these independent variables are put on each mesh point. The first-order differentiations of these variables are defined as the increment between adjoining mesh points; which may put on the middle point between these adjoining mesh points. It is noteworthy to say that these middle points are on the boundary of the control volume. Thereby, we can illustrate the control volume, as shown in Figure 4-4. Independent variables are put on mesh points, while the first differentiations of these variables are put on the boundary of the control volume that bounds the mesh point. The aforementioned surface integral is carried out on this boundary. The second differentiations are, on the other hand, defined as the increments between the midpoints on the boundary of the control volume. As a result, the second differentiations are put on mesh points that are centered in control volumes. If $\vec{F}$ is electric current, it is a first differentiation of carrier densities that are independent variables. If $\vec{F}$ is electric field, it is a first differentiation of potential that is an independent variable. The divergences of the electron current, hole current and electric field are therefore put on mesh points and are replaced by the surface integration around the boundary of the

control volume with continuous flux ($\vec{F}$). Poisson equation is, on the other hand, written in the second differentiation of potential (put on mesh point in control volume). The continuity equations are written in the second differentiation of carrier densities (put on mesh point in control volume). Therefore, the basic formulation of the device physics is written on mesh points (in control volume), while we use control volumes to perform the device simulation. As long as continuous matters satisfy the divergence theorem (72), we can thereby discretize the equations of the continuous matters on the boundary of control volume. *__As long as the electric field and the electric current are regarded as (continuous) fluxes, the finite volume method with control volume is valid to the device physics.__* For the simplicity, let us assume that the permittivity is distributed in the control volume on the two-dimensional rectangular mesh, as shown in Figure 4-4(b). Using (72), the integration of (71) results in:

$$(\epsilon_1 u_{i+1} + \epsilon_8 u_i)\frac{(\psi_{i,j+1} - \psi_{i,j})}{2v_i} + (\epsilon_7 v_i + \epsilon_6 v_{i+1})\frac{(\psi_{i-1,j} - \psi_{i,j})}{2u_i}$$

$$+ (\epsilon_4 u_{i+1} + \epsilon_5 u_i)\frac{(\psi_{i,j-1} - \psi_{i,j})}{2v_i} + (\epsilon_3 v_{i+1} + \epsilon_2 v_i)\frac{(\psi_{i+1,j} - \psi_{i,j})}{2u_{i+1}}$$

$$= -q n_i \left(\phi_{p(i,j)} \exp\left(\frac{-\psi_{i,j}}{V_T}\right) - \phi_{n(i,j)} \exp\left(\frac{\psi_{i,j}}{V_T}\right) + N^+_{D(i,j)} \right.$$

$$\left. - N^-_{A(i,j)} \right) \frac{(h_{i+1}+h_i)(k_{j+1}+k_j)}{4}.$$

$$\tag{73}$$

It should be noted that the control volume is consistent to the meshing by the finite difference scheme [19].

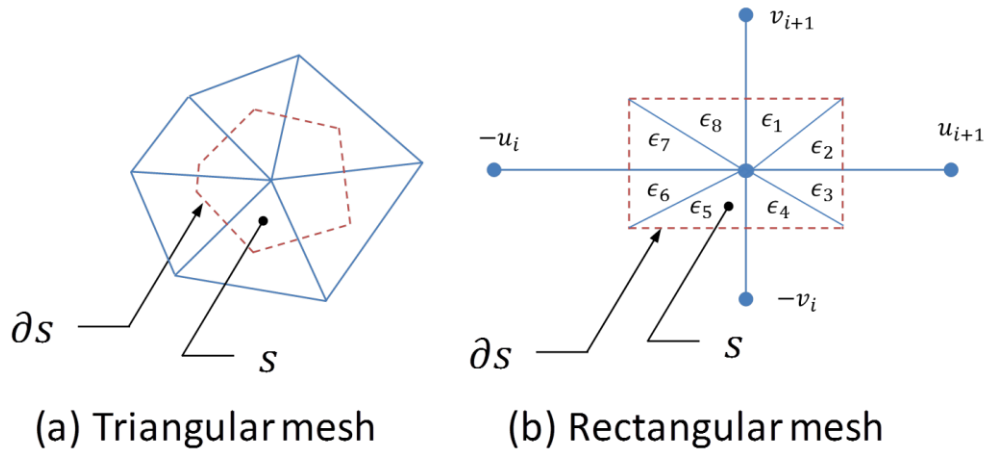(a) Triangular mesh   (b) Rectangular mesh

Figure 4-4   The Control Volume s and Boundary ∂s Defined by Perpendicular Bisectors of Mesh Grid in (a) Triangular Mesh and (b) Rectangular Mesh.

4.4 Scharfetter-Gummel Discretization for Continuity Equations

For the hole and electron current continuity equations, we can discretize them, similar to the above, i.e., by the finite-difference approach [19]. It is implicitly assumed that the same control volume is adopted as the discretization of Poisson equation. From Figure 4-5, we can approximate the electron current continuity equation (28) in x direction as:

$$\frac{dn_i}{dt} \cong G_{n(i)} - R_{n(i)} + \frac{1}{q}\frac{\vec{J}_{i+\frac{1}{2}} - \vec{J}_{i-\frac{1}{2}}}{\Delta x}. \tag{74}$$

Next issue is the meshing in real space. Here note that any linearly-gradual meshing gives rise to a significant problem, which is numerical instability owing to the exponential change of the carrier density to the potential. We can see this exponential dependence considering a 1D case of equation (60),

$$\frac{d}{dx}\left(-n_i D_n \exp\left(\frac{\psi}{V_T}\right)\frac{d\phi_n}{dx}\right) = G_n - R_n. \tag{75}$$

We can express it as

$$\frac{d^2\phi_n}{dx^2} + \frac{1}{V_T}\frac{d\psi}{dx}\frac{d\phi_n}{dx} = -\frac{G_n - R_n}{n_i D_n e^{\frac{\psi}{V_T}}}. \tag{76}$$

Assuming the electric field, $E_x = -\frac{d\psi}{dx}$ as a constant for simplicity, the solution of this ordinary differential equation satisfies the following form:

$$\phi_n = C_1 + C_2 e^{\frac{E_x}{V_T}} + g(x), \tag{77}$$

where $g(x)$ represents the particular solution and C_1 and C_2 are arbitrary numbers. It is obvious that ϕ_n is mainly an exponential function of the electric field.

In order to overcome this problem, Scharfetter and Gummel pointed out that the potential varies linearly across mesh points, while the carrier density varies exponentially. The carrier density equations are regarded as differential equations, in which $\vec{J}_n$, $\vec{J}_p$, μ_n, μ_p, and $\vec{E}$ are assumed constants between mesh points. In a case of 1D, we can thereby write the electron current density as follows:

$$J_{nx} = qD_n n_i e^{\frac{\psi}{V_T}}\frac{d\phi_n}{dx} = const., \quad x_i < x < x_{i+1}. \tag{78}$$

It can be written as

$$d\phi_n = \frac{J_{nx}}{qD_n n_i} e^{-\frac{\psi}{V_T}} dx, \quad x_i < x < x_{i+1}. \tag{79}$$

Substituting $e^{-\frac{\psi}{V_T}} dx = -\frac{V_T}{\frac{d\psi}{dx}} d\left(e^{-\frac{\psi}{V_T}}\right)$ into the above equation, then we obtain

$$d\phi_n = -\frac{J_{nx}}{qD_n n_i} \frac{V_T}{\frac{d\psi}{dx}} d\left(e^{-\frac{\psi}{V_T}}\right), \quad x_i < x < x_{i+1}. \tag{80}$$

Applying the finite-difference method [19] to the electric field, we can obtain:

$$E_x = -\frac{d\psi}{dx} = -\frac{\psi_{i+1} - \psi_i}{h_{i+1}}, \quad x_i < x < x_{i+1}. \tag{81}$$

That is,

$$d\phi_n = -\frac{J_{nx}}{qD_n n_i} \frac{V_T h_{i+1}}{\psi_{i+1} - \psi_i} d\left(e^{-\frac{\psi}{V_T}}\right), \quad x_i < x < x_{i+1}. \tag{82}$$

Integrating this from x_i to x_{i+1}, we obtain:

$$\phi_{n(i+1)} - \phi_{n(i)} = -\frac{J_{nx}}{qD_n n_i} \frac{V_T h_{i+1}}{\psi_{i+1} - \psi_i} \left(e^{-\frac{\psi_{i+1}}{V_T}} - e^{-\frac{\psi_i}{V_T}}\right). \tag{83}$$

Regarding the constant shown in (77) as determined at the middle point of x_i and x_{i+1}, we obtain:

$$J_{nx} \equiv J_{nx}|_{i+\frac{1}{2}} = -q\mu_n n_i \frac{\psi_{i+1} - \psi_i}{e^{-\frac{\psi_{i+1}}{V_T}} - e^{-\frac{\psi_i}{V_T}}} \frac{\phi_{n(i+1)} - \phi_{n(i)}}{h_{i+1}}. \tag{84}$$

This is a coarse-graining. Note that $\vec{J}_{ny}$, $\vec{J}_{nz}$, μ_n, $\vec{E}_y$ and $\vec{E}_z$ are assumed constants between mesh points in similar manner. Here we have used: $\mu_n = D_n/V_T$, i.e., the Eisnstein relationship. A similar analysis in the y and z direction result in

$$J_{ny}|_{j+\frac{1}{2}} = -q\mu_n n_i \frac{\psi_{j+1} - \psi_j}{e^{-\frac{\psi_{j+1}}{V_T}} - e^{-\frac{\psi_j}{V_T}}} \frac{\phi_{n(j+1)} - \phi_{n(j)}}{k_{j+1}}, \tag{85}$$

and,

$$J_{nz}|_{l+\frac{1}{2}} = -q\mu_n n_i \frac{\psi_{l+1} - \psi_l}{e^{-\frac{\psi_{l+1}}{V_T}} - e^{-\frac{\psi_l}{V_T}}} \frac{\phi_{n(l+1)} - \phi_{n(l)}}{w_{l+1}}. \tag{86}$$

By this way, we can discretize the steady-state electron continuity equation three-dimensionally within control volume as follows:

$$\frac{1}{q}\nabla \cdot \vec{J}_n \cong \frac{1}{q}\frac{2}{h_{i+1}+h_i}\left(J_{nx}\big|_{i+\frac{1}{2},j,l} - J_{nx}\big|_{i-\frac{1}{2},j,l}\right) + \frac{1}{q}\frac{2}{k_{j+1}+k_j}\left(J_{ny}\big|_{i,j+\frac{1}{2},l} - J_{ny}\big|_{i,j-\frac{1}{2},l}\right)$$

$$+\frac{1}{q}\frac{2}{w_{l+1}+w_l}\left(J_{nz}\big|_{i,j,l+\frac{1}{2}} - J_{nz}\big|_{i,j,l-\frac{1}{2}}\right) = G_{n(i,j,l)} - R_{n(i,j,l)}. \tag{87}$$

Here note that the first and second terms in each bracket correspond to the out-flow and in-flow, as shown in Figure 2-1. The factor "2" in the numerator before the bracket in each term in the right hand side comes from the control volume (See §0). Substituting (84)-(86) to (87), we can obtain the explicit expression of the steady-state three-dimensionally discretized continuity equation for electrons:

$$\frac{2n_i}{h_{i+1}+h_i}\left(-\mu_{n(i+\frac{1}{2},j,l)}\frac{\psi_{i+1,j,l}-\psi_{i,j,l}}{e^{-\frac{\psi_{i+1,j,l}}{V_T}}-e^{-\frac{\psi_{i,j,l}}{V_T}}}\frac{\phi_{n(i+1,j,l)}-\phi_{n(i,j,l)}}{h_{i+1}}\right.$$

$$\left.+\mu_{n(i-\frac{1}{2},j,l)}\frac{\psi_{i,j,l}-\psi_{i-1,j,l}}{e^{-\frac{\psi_{i,j,l}}{V_T}}-e^{-\frac{\psi_{i-1,j,l}}{V_T}}}\frac{\phi_{n(i,j,l)}-\phi_{n(i-1,j,l)}}{h_i}\right)$$

$$+\frac{2n_i}{k_{j+1}+k_j}\left(-\mu_{n(i,j+\frac{1}{2},l)}\frac{\psi_{i,j+1,l}-\psi_{i,j,l}}{e^{-\frac{\psi_{i,j+1,l}}{V_T}}-e^{-\frac{\psi_{i,j,l}}{V_T}}}\frac{\phi_{n(i,j+1,l)}-\phi_{n(i,j,l)}}{k_{j+1}}\right.$$

$$\left.+\mu_{n(i,j-\frac{1}{2},l)}\frac{\psi_{i,j,l}-\psi_{i,j-1,l}}{e^{-\frac{\psi_{i,j,l}}{V_T}}-e^{-\frac{\psi_{i,j-1,l}}{V_T}}}\frac{\phi_{n(i,j,l)}-\phi_{n(i,j-1,l)}}{k_j}\right)$$

$$+\frac{2n_i}{w_{l+1}+w_l}\left(-\mu_{n(i,j,l+\frac{1}{2})}\frac{\psi_{i,j,l+1}-\psi_{i,j,l}}{e^{-\frac{\psi_{i,j,l+1}}{V_T}}-e^{-\frac{\psi_{i,j,l}}{V_T}}}\frac{\phi_{n(i,j,l+1)}-\phi_{n(i,j,l)}}{w_{l+1}}\right.$$

$$\left.+\mu_{n(i,j,l-\frac{1}{2})}\frac{\psi_{i,j,l}-\psi_{i,j,l-1}}{e^{-\frac{\psi_{i,j,l}}{V_T}}-e^{-\frac{\psi_{i,j,l-1}}{V_T}}}\frac{\phi_{n(i,j,l)}-\phi_{n(i,j,l-1)}}{w_l}\right)$$

$$= G_{n(i,j,l)} - R_{n(i,j,l)}. \tag{88}$$

The variables determined at middle points are regarded as average values according to the coarse-graining. In a similar manner, we can derive the equation for hole current density corresponding to (84) as follows:

$$J_{px}\big|_{i+\frac{1}{2}} = -q\mu_p n_i \frac{\psi_{i+1}-\psi_i}{e^{\frac{\psi_{i+1}}{V_T}}-e^{\frac{\psi_i}{V_T}}}\frac{\phi_{p(i+1)}-\phi_{p(i)}}{h_{i+1}}. \tag{89}$$

Therefore, the steady-state three-dimensionally discretized continuity equation for holes turns out to be:

$$\frac{2n_i}{h_{i+1}+h_i}\left(-\mu_{p(i+\frac{1}{2},j,l)}\frac{\psi_{i+1,j,l}-\psi_{i,j,l}}{e^{\frac{\psi_{i+1,j,l}}{V_T}}-e^{\frac{\psi_{i,j,l}}{V_T}}}\frac{\phi_{p(i+1,j,l)}-\phi_{p(i,j,l)}}{h_{i+1}}\right.$$

$$\left.+\mu_{p(i-\frac{1}{2},j,l)}\frac{\psi_{i,j,l}-\psi_{i-1,j,l}}{e^{\frac{\psi_{i,j,l}}{V_T}}-e^{\frac{\psi_{i-1,j,l}}{V_T}}}\frac{\phi_{p(i,j,l)}-\phi_{p(i-1,j,l)}}{h_i}\right)$$

$$+\frac{2n_i}{k_{j+1}+k_j}\left(-\mu_{p(i,j+\frac{1}{2},l)}\frac{\psi_{i,j+1,l}-\psi_{i,j,l}}{e^{\frac{\psi_{i,j+1,l}}{V_T}}-e^{\frac{\psi_{i,j,l}}{V_T}}}\frac{\phi_{p(i,j+1,l)}-\phi_{p(i,j,l)}}{k_{j+1}}\right.$$

$$\left.+\mu_{p(i,j-\frac{1}{2},l)}\frac{\psi_{i,j,l}-\psi_{i,j-1,l}}{e^{\frac{\psi_{i,j,l}}{V_T}}-e^{\frac{\psi_{i,j-1,l}}{V_T}}}\frac{\phi_{p(i,j,l)}-\phi_{p(i,j-1,l)}}{k_j}\right)$$

$$+\frac{2n_i}{w_{l+1}+w_l}\left(-\mu_{p(i,j,l+\frac{1}{2})}\frac{\psi_{i,j,l+1}-\psi_{i,j,l}}{e^{\frac{\psi_{i,j,l+1}}{V_T}}-e^{\frac{\psi_{i,j,l}}{V_T}}}\frac{\phi_{p(i,j,l+1)}-\phi_{p(i,j,l)}}{w_{l+1}}\right.$$

$$\left.+\mu_{p(i,j,l-\frac{1}{2})}\frac{\psi_{i,j,l}-\psi_{i,j,l-1}}{e^{\frac{\psi_{i,j,l}}{V_T}}-e^{\frac{\psi_{i,j,l-1}}{V_T}}}\frac{\phi_{p(i,j,l)}-\phi_{p(i,j,l-1)}}{w_l}\right)$$

$$=G_{p(i,j,l)}-R_{p(i,j,l)}. \tag{90}$$

The essence of the Scharfetter-Gummel discretization is to replace the difference (that should be determined at meshing point) with the increment between the variables which are determined respectively on the boundary of the control volume that involves the corresponding meshing point. The variables are, on the other hand, averaged at middle points between the considered meshing point and its adjoining meshing points. This can be regarded as "Coarse-Graining in computation". The memory capacity needed to computation is therefore sensitive to the size of control volume.

It is noteworthy to say that the control volume is adopted to describe non-uniform permittivity in Poisson equation under the finite difference scheme [19]. Scharfetter-Gummel method is adopted to suppress the exponential change of carrier densities in the current continuity equations. Accordingly, the meshing must be consistent to solve these equations at the same moment.
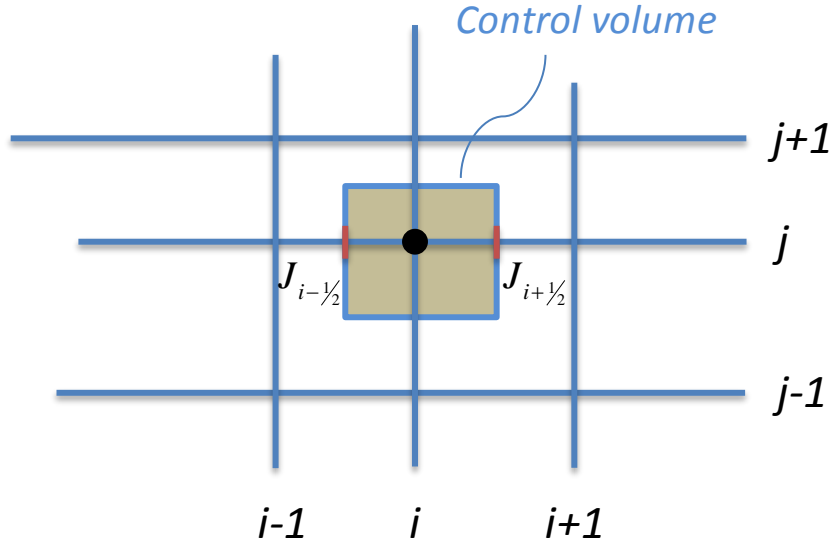
Control volume

$J_{i-\frac{1}{2}}$ $J_{i+\frac{1}{2}}$

j+1

j

j-1

i-1   *i*   *i+1*

Figure 4-5   Nonuniform Meshes for Discretization

4.5 Algorithms for Device Simulation

As mentioned above, the equations that we solve for the device simulation are summarized as:

$$f_1(\psi, n, p) = \nabla^2\psi + \frac{q}{\epsilon}(p - n + N_D^+ - N_A^-) = 0. \tag{91}$$

$$f_2(\psi, n, p) = \nabla \cdot \vec{J}_n - (G_n - R_n) = 0. \tag{92}$$

$$f_3(\psi, n, p) = \nabla \cdot \vec{J}_p - (G_p - R_p) = 0. \tag{93}$$

First of all, we substitute initial values of the independent variables, i.e. (ψ, n, p) to (91), (92), and (93), and then calculate $f_1, f_2,$ and f_3. Secondly, we adjust the values of (ψ, n, p) from the initial values according to an algorithm to make $f_1, f_2,$ and f_3 close to zero at the same moment. Subsequently, we adjust the values of (ψ, n, p) from those at the last step according to the algorithm to make $f_1, f_2,$ and f_3 further close to zero at the same moment. This process is continued until a predetermined convergence condition is satisfied. Finally, we solve the equations of (91), (92), and (93) after the convergence is reached, so as to obtain the converged values of the independent variables. We will briefly review two famous algorithms below.

4.5.1 Gummel's Iteration Method

Gummel's iteration method is a decoupled procedure for making (ψ, n, p) satisfy (91), (92), and (93) in turn. At each step, the equations are solved separately, as illustrated in Figure 4-6. Let us explain the solution flow from the i-th step at which the independent variables have the values of (ψ^i, n^i, p^i). Firstly, we solve Poisson equation while (n^i, p^i) is unchanged. The Poisson equation can then be regarded as a nonlinear equation with only one variable. We can solve it by other numerical methods. As a solution of Poisson equation, we obtain (ψ^{i+1}, n^i, p^i), where the potential is updated to ψ^{i+1}. Subsequently, we substitute (ψ^{i+1}, n^i, p^i) to f_2. (92) may not be satisfied before the convergence. In this event, we regard the electron density as a variable to solve (92) while (ψ^{i+1}, p^i) is unchanged. As a result, we obtain $(\psi^{i+1}, n^{i+1}, p^i)$. Next, we substitute $(\psi^{i+1}, n^{i+1}, p^i)$ to f_3. (93) may not be satisfied before the convergence. In this event, we regard the hole density as a variable to solve (93) while (ψ^{i+1}, n^{i+1}) is unchanged. As a result, we obtain $(\psi^{i+1}, n^{i+1}, p^{i+1})$. This procedure $(i \rightarrow i+1)$ will be iterated until the condition of the convergence that we can separately predetermine is satisfied.

Note that the computation would not be stable owing to the exponential change of carrier densities with regard to the potential, as mentioned in Sec. 4.1. Actually, we may use the independent variable set of $(\psi, \varphi_n, \varphi_p)$ in Gummel's Iteration method.

The convergence of Gummel's iteration scheme is critically sensitive to coupled phenomena of electron and hole. For example, the pair-creation of electron and hole updates G_n and G_p at the same moment, but nothing happens in (91) because of no change in total charge amount. This is an inherent weak point of this method. However, if the dimension of electron device becomes small in the level of nano-meters, the valence band and the conduction band can not be clearly distinguished. In this event, we should ignore the existence of hole, and then the weak point of this method can be automatically resolved. From this fact, we can image another solution to overcome the weak point in General-Purpose Nano-Device Simulator. We may replace (92) and (93) with fully-quantum-mechanical transport equation in which the existence of hole is not assumed. This will be left for Ph.D. thesis.
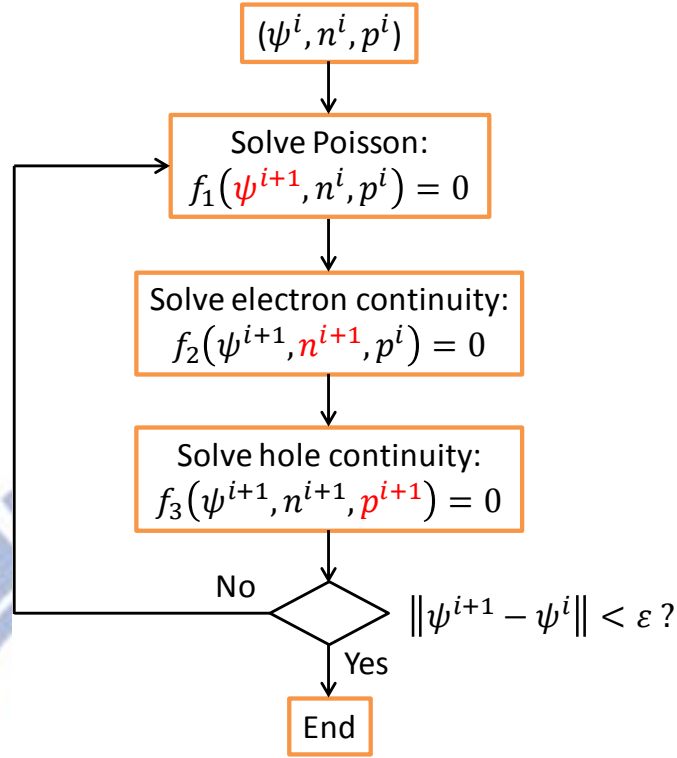
Figure 4-6   Gummel's Iteration Scheme

4.5.2 Newton's Method

The aforementioned weak point of Gummel's Iteration method can be easily resolved by Newton's method [21], which is a commonly used method for solving nonlinear equations.

First of all, we may point out that all of f_1, f_2, and f_3 are smoothly varying around the converged values of (ψ, n, p) to the change of these independent variables as long as the discontinuity (the correction of the collision term to the relaxation approximation) is moderate. That is, we can linearize $(91) - (93)$ around the converged values of the independent variables. This is the essential meaning of Newton's Method. Here let us consider the linearization at the $(i + 1)$-th step. The coupled equations of $(91) - (93)$ are divided into the term of i-th step and the linear deviation therefrom is as follows:

$$\begin{pmatrix} f_1(\psi^i, n^i, p^i) \\ f_2(\psi^i, n^i, p^i) \\ f_3(\psi^i, n^i, p^i) \end{pmatrix} + \begin{bmatrix} \dfrac{\partial f_1}{\partial \psi} & \dfrac{\partial f_1}{\partial n} & \dfrac{\partial f_1}{\partial p} \\ \dfrac{\partial f_2}{\partial \psi} & \dfrac{\partial f_2}{\partial n} & \dfrac{\partial f_2}{\partial p} \\ \dfrac{\partial f_3}{\partial \psi} & \dfrac{\partial f_3}{\partial n} & \dfrac{\partial f_3}{\partial p} \end{bmatrix} \begin{pmatrix} \Delta \psi^i \\ \Delta n^i \\ \Delta p^i \end{pmatrix} = 0. \tag{94}$$

28

The problem that we will solve is thereby turned out to be a simultaneous linear equation:

$$L(\vec{x}^{i+1} - \vec{x}^i) = -\vec{f}(\vec{x}^i), \tag{95}$$

where $\vec{f} = (f_1, f_2, f_3)^t$, $\vec{x}^i = (\psi^i, n^i, p^i)^t$, and the Jacobian matrix is:

$$L = \begin{bmatrix} \dfrac{\partial f_1}{\partial \psi} & \dfrac{\partial f_1}{\partial n} & \dfrac{\partial f_1}{\partial p} \\ \dfrac{\partial f_2}{\partial \psi} & \dfrac{\partial f_2}{\partial n} & \dfrac{\partial f_2}{\partial p} \\ \dfrac{\partial f_3}{\partial \psi} & \dfrac{\partial f_3}{\partial n} & \dfrac{\partial f_3}{\partial p} \end{bmatrix} \tag{96}$$

$$= \begin{bmatrix} \nabla^2 & -\dfrac{q}{\epsilon} & \dfrac{q}{\epsilon} \\ \nabla \cdot (-qn\mu_n\nabla) & \nabla \cdot [(-q\mu_n\nabla\psi + qD_n\nabla)] - \dfrac{\partial(G_n - R_n)}{\partial n} & -\dfrac{\partial(G_n - R_n)}{\partial p} \\ \nabla \cdot (-qp\mu_p\nabla) & -\dfrac{\partial(G_p - R_p)}{\partial p} & \nabla \cdot [(-q\mu_p\nabla\psi - qD_p\nabla)] - \dfrac{\partial(G_p - R_p)}{\partial p} \end{bmatrix}. \tag{97}$$

Here we assume that N_D^+ and N_A^- are constant with regard to any change of (ψ, n, p) for simplicity. However, the ionization rates of doping atoms are not constant. Actually, we need to take the differentiation of the ionization rates [22, 24, 25] into consideration. The brief algorithm is illustrated in Figure 4-7. Here we start with the last update of the vector $\vec{x}^i$, i.e., the last updated set of the independent variables. Next, we calculate the elements of the Jacobian matrix at the last update. Subsequently, we substitute the Jacobian matrix into the recursive formula:

$$\vec{x}^{i+1} = \vec{x}^i - L^{-1}(\vec{x}^i)\vec{f}(\vec{x}^i), \tag{98}$$

to obtain the next update, i.e., $\vec{x}^{i+1}$. Finally, it is noteworthy to say that Newton's iteration scheme is preferable when the couplings between equations are deeper, while more computer resource is required for calculating the Jacobian matrix.
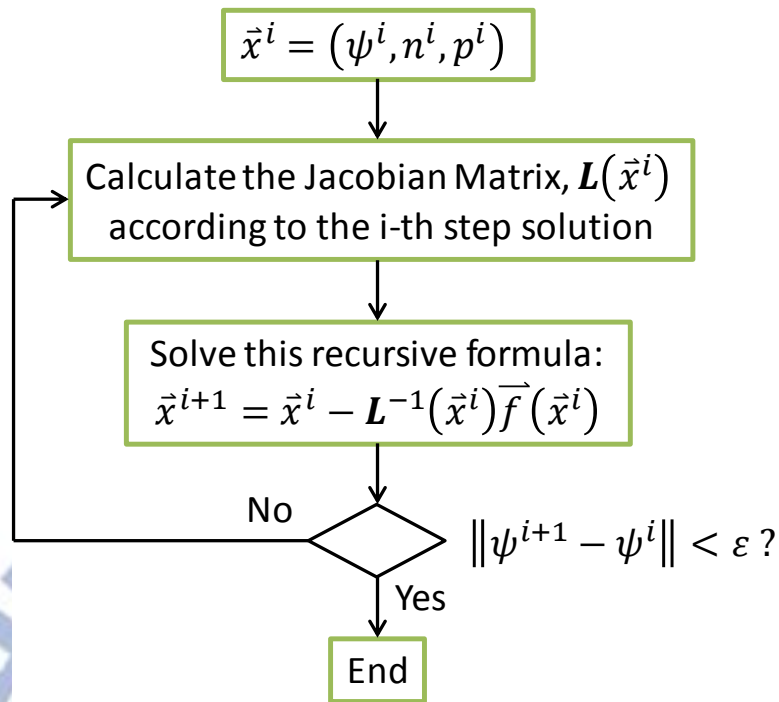
$$\vec{x}^i = (\psi^i, n^i, p^i)$$

Calculate the Jacobian Matrix, $\boldsymbol{L}(\vec{x}^i)$ according to the i-th step solution

Solve this recursive formula:
$$\vec{x}^{i+1} = \vec{x}^i - \boldsymbol{L}^{-1}(\vec{x}^i)\overrightarrow{f}(\vec{x}^i)$$

No

$$\left\|\psi^{i+1} - \psi^i\right\| < \varepsilon \ ?$$

Yes

End

Figure 4-7   Newton's Iteration Scheme

Chapter 5. Discussion and Conclusions

In this thesis, we briefly reviewed theoretical background of the device simulation. Firstly, we started it with the Boltzmann-transport equation (BTE), which is the most general expression of carrier transport equation in semiconductor devices. However, it is not easy to be solved, since continuous distribution function and discontinuous collision terms compete with each other in the BTE. The collision term is divided into the relaxation term (continuous) and its correction (generally discontinuous). From some naïve approximation of BTE, the relaxation term is left as the drift-diffusion model for describing the current continuity equation of electron and hole. The correction to the relaxation is left as the generation-recombination term in the current continuity equations of electron and hole. Poisson equation and these two current continuity equations are correlated to each other, so as to compose the basic formulation of today's device physics with three independent variables. It is noteworthy that this formulation is described in terms of differentials of continuous variables while ***the generation-recombination term can be expanded to involve more discontinuous physical processes***. In this sense, the current continuity equations can be regarded as "apparent" differential equations. In order to simulate this formulation, we need to discretize these equations. Poisson equation is discretized by the finite difference scheme [19]. The current continuity equations are discretized by the Scharfetter-Gummel discretization method for suppressing the numerical instability owing to the exponential dependence of carrier densities with regard to electrostatic potential. Subsequently, we briefly reviewed two iteration methods for solving the discretized Poisson equation and the discretized current continuity equations self-consistently. The Gummel iteration method, which was proposed by Gummel in 1964, solves these equations in turn. In this method, less computational resource is demanded, whereas a strong correlation between equations suppresses the convergence. For example, the pair-creation of electron-hole gives rise to the correlation between the continuity equations of electron and hole. The Newton iteration method solves all equations simultaneously. This method is evidently suitable while the correlation between equations is stronger. However, much more computational resource is demanded since Jacobian matrix is used.

As the dimension of devices shrinks into the range of de Broglie length (around 10 nm in silicon), the transport behavior of carrier inside the devices will approach quantum transport.

In other words, the continuity described by the drift-diffusion model and the discontinuity of fully quantum-mechanical process would compete each other as the device scaling is advanced. There are many methods have been proposed for the quantum correction, such as the Van Dort Model [26], the Hänsch model [27] and the density-gradient theory [28]. Nevertheless, these quantum correction methods are based on the drift-diffusion model, which should be unreliable in nano-devices. They can work as long as the discontinuity is not large, and then we can conclude that these methods are not valid to "MORE THAN MOORE". This invalidity comes from the origin of the drift-diffusion. Accordingly, in nano-devices, this concept would be no longer available, since the transport behavior is discontinuous. For example, we may take into account the tunneling process inside Nano-Devices[4]. In the electron continuity equation:

$$\frac{\partial n}{\partial t} - \frac{1}{q}\nabla_r \cdot \vec{J}_n = U_{tunnel}, \qquad (99)$$

the left hand side is continuous and the right hand side corresponds to this tunneling process which is discontinuous. These two terms are going to compete with each other as the device scaling is advanced. It should be noted that we can append more discontinuous physical processes into the right hand side. We can take into account more discontinuous physical processes. For the sake of Nano-Device Simulation, we may consider discontinuous "quantum hopping", as shown in Figure 5-1. Firstly, (a) illustrates the continuous transport for a comparison. A carrier can move from a control volume to a nearest neighbor one. Markovian process like this can describe the continuous transport in terms of discretization. On the other hand, (b) illustrates the discontinuous transport between two control volumes that are not adjacent. In this event, the transport is not Markovian since the distance of hopping is too long. For Nano Devices, non-Markovian process prevails; which breaks the assumption of today's device simulation.

In conclusion, for the future nano-technology, we will establish a new description for discontinuous carriers. This will be continued to Ph.D. Work.

---

[4] Gate tunneling and band-to-band tunneling are also taken into account as discontinues terms in today's reliability simulation.
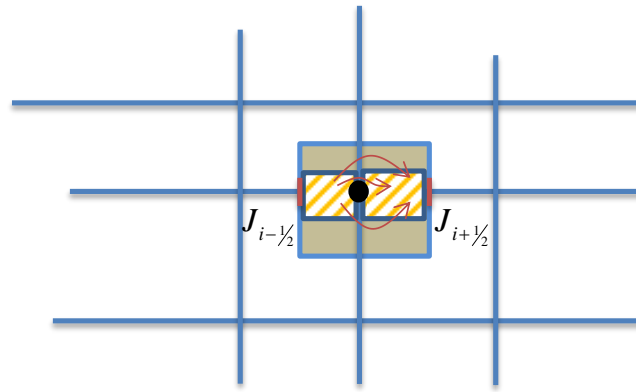
(a) Current flow
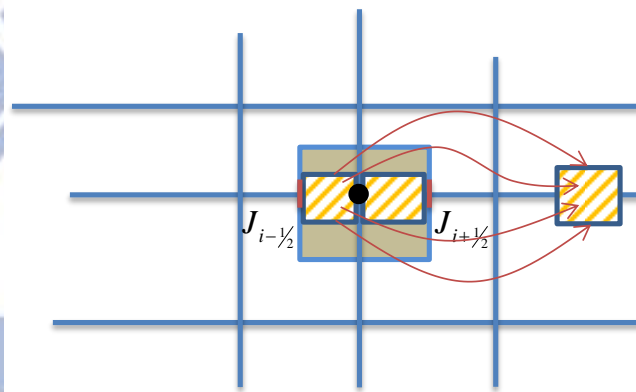
(b) Tunneling

Figure 5-1 (a) Continuous Transport and (b) Discontinuous Transport

Bibliographies

1. T. Grasser, T.-W. Tang, H. Kosina and S. Selberherr, "A review of hydrodynamic and energy-transport models for semiconductor device simulation," *Proc. IEEE*, vol. 91, no. 2, pp. 251-274, 2003.

2. D. L. Scharfetter and H. K. Gummel, "Large-signal analysis of a silicon Read diode oscillator," *IEEE Trans. Electron Devices*, vol. 16, no. 1, pp. 64-77, 1969.

3. H. K. Gummel, "A self-consistent iterative scheme for one-dimensional steady state transistor calculations," *IEEE Trans. Electron Devices*, vol. 11, no. 10, pp. 455-465, 1964.

4. J. E. Lilienfeld, "U.S. patents 1,745,175 (filed 1926, issued 1930), 1,877,140 (filed 1928, issued 1932), and 1,900,018 (filed 1928, issued 1933)."

5. W. Shockley, "The path to the conception of the junction transistor," *IEEE Trans. Electron Devices*, vol. 23, no. 7, pp. 597-620, 1976.

6. S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*, 3rd Ed., John Wiley & Sons, Hoboken, New Jersey, 2007.

7. D. Kahng and M. Atalla, "U.S. patents, 3206670 & 3102230 (1960)."

8. International SEMATECH, "Intl. Technology Roadmap for Semiconductor (ITRS)," 2011. http://www.itrs.net/

9. R. W. Dutton, "Modeling and simulation for VLSI," *Tech. Dig. IEEE IEDM*, pp. 2-7, 1986.

10. R. H. Dennard, F. H. Gaensslen, H.-N. YU, V. L. Rideout, E. Bassous and A. R. LeBlanc, "Design of ion-implanted MOSFET's with very small physical dimensions," *IEEE J. Solid State Circuits*, vol. 9, no. 5, pp. 256-268, 1974.

11. S. L. Teitel and J. W. Wilkins, "Ballistic transport and velocity overshoot in semiconductors: Part I—Uniform field effects," *IEEE Trans. Electron Devices*, vol. 30, no. 2, pp. 150-153, 1983.

12. K. K. Ng and G. W. Taylor, "Effects of hot-carrier trapping in n- and p-channel MOSFET's," *IEEE Trans. Electron Devices*, vol. 30, no. 8, pp. 871-876, 1983.

13. L. A. Akers, "Threshold voltage of a narrow-width MOSFET," *Electron. Lett.*, vol. 17, no. 1, pp. 49-51, 1981.

14. R. Klima, *Three-Dimensional Device Simulation with Minimos-NT*. Dissertation, Technischen Universität Wien, 2002.

http://www.iue.tuwien.ac.at/phd/klima

15. C. Jacoboni and L. Reggiani, "The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials," *Rev. Mod. Phys.*, vol. 55, no. 3, pp. 645-705, 1983.

16. L. I. Schiff, *Quantum Mechanics*, McGraw-Hill Inc., New York, 1955.

17. W. Shockley, *Electrons and Holes in Semiconductors*, D. Van Nostrand, Princeton, New Jersey, 1950.

18. J. W. Slotboom, "Computer-aided two-dimensional analysis of bipolar transistors," *IEEE Trans. Electron Devices*, vol. 20, no. 8, pp. 669-679, 1973.

19. D. Vasileska, S. M. Goodnick and G. Klimeck, *Computational Electronics: Semiclassical and Quantum Device Modeling and Simulation*, Taylor & Francis Group, Boca Raton, Florida, 2010.

20. R. E. Bank, D. J. Rose and W. Fichtner, "Numerical methods for semiconductor device simulation," *IEEE Trans. Electron Devices*, vol. 30, no. 9, pp. 1031-1041, 1983.

21. K. P. Chong and S. H. Żak, *An Introduction to Optimization*, 3rd Ed., John Wiley & Sons, Hoboken, New Jersey, 2008.

22. H. Watanabe, K. Kawabata, T. Ichikawa, "A tight binding method study of optimized Si-SiO$_2$ system", IEEE Trans. Electron Devices, vol. 57, no. 11, pp. 3084-3091, 2010.

23. US Patent 6,954,723, "Device simulation method, device simulation system and device simulation program" by H. Watanabe and K. Matsuzawa (Issued: Oct. 11, 2005).

24. H. Watanabe and S. Takagi, "Effects of incomplete ionization of impurities in poly-Si gate and band gap narrowing on direct tunneling gate leakage current", J. Appl. Phys., vol. 90, No. 3, pp. 1600-1607, 2001.

25. H. Watanabe, K. Matsuzawa, and S. Takagi, "Scaling effects on gate leakage current", IEEE Trans. Electron Devices, vol. 50, no. 8, pp. 1779-1784, 2003.

26. M. J. van Dort, P. H. Woerlee, A. J. Walker, C. A. H. Juffermans and H. Lifka, "Influence of high substrate doping levels on the threshold voltage and the mobility of deep-submicrometer MOSFETs," *IEEE Trans. Electron Devices*, vol. 39, no. 4, pp. 932-938, 1992.

27. W. Hänsch, Th. Vogelsang, R. Kircher and M. Orlowski, "Carrier transport near the Si/SiO2 interface of a MOSFET," *Solid State Electron.*, vol. 32, no. 10, pp. 839-849, 1989.

28. M. G. Ancona and G. J. Iafrate, "Quantum correction to the equation of state of an

electron gas in a semiconductor," *Physical Review B (Condensed Matter)*, vol. 39, no. 13, pp.9536-9540, 1989.