New Solution to High-Field Transport in Semiconductors:

II. Velocity Saturation and Ballistic Transmission

Kenji Natori*

Institute of Applied Physics, University of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan

High-field transport in semiconductor diodes at room temperature is analyzed in the reflection-transmission regime. The pseudo-one-dimensional Boltzmann equation with a constant electric field is transformed into a pair of carrier flux equations. They are analytically solved neither with the relaxation time approximation nor with the perturbation expansion. The carrier energy relaxation due to optical phonon emission is essential in high-field transport. The current- and velocity-field characteristics are closely related to flux transmission through a specific layer, in which the elastic scattering is dominant and the optical phonon emission is absent. If the transmission coefficient is much less than unity, the proportionality of the current to the field results as the Ohm's law in high-field range. The current and velocity tend to saturate when the coefficient approaches unity (ballistic transmission). This result provides simple insight into transport in nanoscale devices.

* E-mail address: natori@esys.tsukuba.ac.jp

1. Introduction

In semiconductor device theory, high-field transport has been a crucial issue that dominates device performance. A half century ago, Ryder[1)] and Shockley[2)] investigated the current in germanium and silicon and indicated the saturation of carrier velocity at a high field. Since then, high-field transport has been studied through various approaches: initially by theoretical analysis, and then by Monte Carlo simulation, along with experimental investigations. Now devices are on the nanoscale, and the electric field inside them is intensified. Clarification of high-field transport is crucial for control as well as proper understanding of device operation and performance.

The drift current density I under a constant electric field E is usually described by Ohm's law:[3)]

$$I = \frac{nq^2\tau}{m}E. \qquad (1)$$

Here, q, m, and n are the charge, the effective mass, and the density of a carrier, respectively. τ is the relaxation time. Equation (1) is often divided into $I = qnv$ and $v = (q\tau/m)E$, where v is the drift velocity. Equation (1) for a homogeneous bulk system, considering the carrier distribution in momentum space, is derived from the Boltzmann transport equation (BTE).[4)] The derivation is usually based on two major assumptions. One is the phenomenological relaxation time approximation (RTA).[4-6)] The collision integral of BTE is approximated by a simple expression:

$$\left(\frac{\partial f_{\mathbf{k}}}{\partial t}\right)_{coll.} = -\frac{1}{\tau}\left(f_{\mathbf{k}} - f^{0}{}_{\mathbf{k}}\right), \qquad (2)$$

where $f_{\mathbf{k}}$ ($f^{0}{}_{\mathbf{k}}$) is the distribution function with the applied field (at thermal equilibrium). This

assumes that the deviated distribution of the carrier relaxes to thermal equilibrium in relaxation time τ [the same τ as in eq. (1), the averaged value over the Fermi surface when the carrier distribution is considered]. The RTA is generally thought to be effective when the deviation is small. Details of the scattering process, such as whether the scattering is elastic or inelastic, or the complicated course of energy relaxation satisfying conservation rules, are all neglected. The other inevitable assumption is the perturbation expansion of current density I in powers of the field E. The linear eq. (1) is derived for a sufficiently low field. The higher-order terms in E are thought to represent the hot carrier effects[7,8] in high fields.

Velocity saturation (actually, current density saturation) shown in Ryder's data[1] is a characteristic phenomenon in high-field transport[8]. It was extensively investigated by Shockley [2] and his successors, usually using the balance-of-energy equation of hot carriers and assuming RTA. When an electric field is sufficiently large, the optical phonon emission constitutes the dominant process of energy relaxation, and carriers emit optical phonons as soon as the kinetic energy gained from the field exceeds the optical phonon energy (which is denoted by ε^* throughout this paper) and reduce their velocity to almost zero. The increase in mean carrier velocity is truncated at a value of $\sqrt{\varepsilon^*/2m}$ independent of the field. At first, Shockley assumed an infinitely strong interaction between the electron and the optical phonon, but later the relative contribution of various scattering mechanisms to the energy loss was extensively studied. Reik and Risken[9] derived saturation velocity dependent on both deformation potentials of the acoustic and the optical phonon scattering by an orthodox analysis of BTE. The complexity of the problem, however, forced an approximate discussion based on the system's uniformity and the RTA, and the physical mechanism for the transition from the lower-field E-dependence to the high-field velocity saturation was not clear. The

saturation velocity in *n*-silicon remains at 10^7 cm/s even at a high field of 1.3×10^5 V/cm[10]. However, the mean carrier velocity during acceleration by the field within a short optical phonon scattering time of 5×10^{-14} s exceeds 2×10^7 cm/s. Clarification of the transport physics based on the kinematical mechanism is desired.

In our previous paper,[11] referred to as Part I hereinafter, the high-field transport without energy relaxation was primarily analyzed. In sufficiently low electric fields, the carrier energy fully relaxes along the channel. The drift-diffusion current model is proved, and the linear *I*-*E* relationship in eq. (1) is verified. In the higher-field range of $E \geq 500$ V/cm for silicon, the energy relaxation cannot catch up with the energy gain due to the field. The kinetic energy of carriers accumulates along the current path, and the quasi-equilibrium collapses. The energy-dependent relaxation time produces a deviation from the linear *I*-*E* relationship in eq. (1), although the experimental curve still maintains linearity. Clarification of the mechanism of linearity at high fields is required.

As a limiting case, Part I analyzed high-field transport with elastic scattering and without energy relaxation as Model 1. The electric current density through the channel region with length *L*, which intervened between the source and the drain electrode, and to which the electric field *E* was applied, was analyzed. The transmission coefficient of carriers from the source to the drain was derived as

$$\overline{T} = \frac{\tau_c}{\tau_c + \Delta t^{\varepsilon}_{0\to L}}, \qquad (3)$$

where τ_c is the backscattering time. The value of τ_c is related to the backscattering probability *B* and the three-dimensional elastic scattering time τ_e derived from the low-field mobility, $\mu = q\tau_e / m$, as $\tau_c = 1/B = 2\tau_e$. Here $\Delta t^{\varepsilon}_{x_1\to x_2}$ stands for the traverse time of a carrier, with

kinetic energy ε at x=0, from x=x_1 to x=x_2 accelerated by the electric field E in the absence of scattering. Namely, $\Delta t^{\varepsilon}_{x_1 \to x_2}$ was evaluated by the constant acceleration motion of a carrier as

$$\Delta t^{\varepsilon}_{x_1 \to x_2} = \frac{\sqrt{2m}\left(\sqrt{\varepsilon + qEx_2} - \sqrt{\varepsilon + qEx_1}\right)}{qE}. \qquad (4)$$

For a wide field range that satisfies $\tau_c << \Delta t^{\varepsilon}_{0 \to L}$, the electric current was expressed as $I \propto \sqrt{E/L}$. In contrast, for exceedingly large fields corresponding to $\Delta t^{\varepsilon}_{0 \to L} << \tau_c$, the transmission coefficient approaches unity (ballistic transmission), and the electric current tends to saturate.

This paper analyzes high-field transport by combining the elastic scattering in Model 1 with energy relaxation due to the optical phonon emission[12,13] and clarifies the mechanism of the linear *I*-*E* relationship as well as the physics of velocity saturation. Similar to Part I, the pseudo-one-dimensional BTE is transformed into a pair of flux equations, and the transport in the reflection-transmission regime is discussed. The two major assumptions in eq. (1), RTA and perturbation expansion, are removed, and a consistent solution from the low-field linear region toward the high-field velocity saturation is derived.

Due to the acceleration by the electric field, the carrier motion in our system is not uniform along the path. Its velocity increases as it moves along. In contrast, the time flow is always uniform. The system's behavior is characterized by the ratio of the traverse time to the scattering time of carriers. The description of the solution is very complicated when expressed by spatial position x as in eq. (4), but it reduces to a simple description when expressed by the time parameter $\Delta t^{\varepsilon}_{x_1 \to x_2}$, as we see in eq. (3). Henceforth, the mathematical evolution in terms of the spatial coordinate x may be complicated, but the final result is rewritten with the time parameter to simplify the expression and to clarify the physical meaning.

The purpose of this theoretical analysis is to identify the primary mechanisms in the high-field transport and to develop the underlying physics. The analysis is considered valid if the qualitative features as well as the rough magnitude of *I-V* characteristics are well reproduced. This standpoint is contrasted to the numerical simulation, where detailed mechanisms of transport are considered in a large-scale calculation, and numerical accuracy is pursued. The two approaches complement each other, and both are required in the development of nanoscale electronics. In §2, we analyze carrier transport combining the elastic scattering and the optical phonon emission, and the current density and the carrier velocity are derived. In §3, the physics of the high-field transport is discussed in detail, including Ohm's law and velocity saturation. The conclusion is given in §4.

2. Model 2: Elastic Scattering Combined with Optical Phonon Emission

2.1 Incorporation of energy relaxation into Model 1

We analyze high-field transport through a semiconductor body, referred to as a channel, intervening between the source and drain electrodes, similarly to in Part I. A silicon channel with a low doping concentration n_0=2.5×10^{14} cm^{-3} is assumed. A drain bias V is applied with respect to the source electrode, and a constant electric field $E = V / L$ is assumed within a channel of length L. In contrast to Part I, we investigate the case with an energy relaxation of a fixed amount ε* due to optical phonon emission, in addition to the energy-conserving elastic scattering. We assume that the acoustic phonon scattering is included in the elastic scattering. Although silicon has a multivalley band structure and optical phonon scattering is allowed as intervalley scattering, there is no problem in effectively considering the energy relaxation process in our single-valley framework. In Part I, the value of the backscattering probability B

is estimated as 2.5×10^{12} s^{-1} from the low field mobility, 1430 $cm^2V^{-1}s^{-1}$ [14], of high-purity silicon. Now we also introduce the transition probability per unit time due to optical phonon emission, denoted by *D*. The optical phonon scattering probability is estimated [15] to be roughly similar to that of acoustic phonon scattering for silicon, and here we assume for simplicity that $D = B = 2.5\times10^{12}$ s^{-1}. The optical phonon energy ε* is 63 meV for silicon[14] and is 2.4 times larger than the thermal energy (26 meV at room temperature). Therefore, the transition to higher-energy states by absorption of an optical phonon is presumed to be suppressed and can be neglected when compared with the emission. The mean number of optical phonons is estimated as $\langle N\rangle = 1/[\exp(0.063/0.026)-1] = 0.097$. The absorption probability, which is proportional to $\langle N\rangle$, is only 9% of the emission probability proportional to $(\langle N\rangle + 1)$, so we neglect it. A similar model has been investigated by Baraff.[16]

As is shown in the potential profile of Fig. 1, a carrier with kinetic energy ε, which we assume is less than ε*, is injected from the source to the channel. The energy level of the injected carrier is hereafter designated as the incident energy level. When the carrier stays within the region $0 \le x \le x_0 = (\varepsilon^* - \varepsilon)/qE$, the kinetic energy is smaller than ε* and the energy relaxation due to optical phonon emission is inhibited. The theory of elastic scattering in Part I is applied to carriers within the region. This region is denoted hereafter as the initial elastic zone. Beyond the point $x = x_0$, optical phonon emission is allowed and carriers are exposed to energy relaxation to the lower energy level $(\varepsilon - \varepsilon^*)$. The energy dispersion of the optical phonon is neglected, and so it is not necessary to consider momentum conservation. The lower energy level at $(\varepsilon - \varepsilon^*)$ is denoted as the first relaxed level hereafter. Carriers in the first relaxed level are only allowed to populate the region $x \ge x_0$. Energy relaxation due

to optical phonon emission is inhibited for the carrier in the first relaxed level within the region $x_0 \le x \le x_0 + x_1$, where $x_1 = \varepsilon^* / qE$. We call this region the first relaxed elastic zone. Beyond $x = x_0 + x_1$, carriers are allowed to relax their energy and undergo a transition to the energy level $(\varepsilon - 2\varepsilon^*)$; this level is denoted hereafter as the second relaxed level. Further relaxed levels with lower energy values are similarly defined. The values of x_0 and x_1 for a thermal carrier in silicon are estimated to be 0.37 and 0.63 μm, respectively, for $E = 1000$ V/cm.

Recently, the ballistic or the quasi-ballistic transport in nanoscale devices[17-20] has been frequently analyzed in the reflection-transmission regime. The concept of the "*kT* layer"[21-23] proposed in the analysis is effective for understanding the underlying physics. This concept is contrasted to the "elastic zone" just introduced.

2.2 The incident energy level

The general formalism of analysis based on BTE, as well as the phase space that consists of the (*x*,*k*) plane, is introduced in Part I. The state of a carrier in $0 \le x \le x_0$, where only elastic scattering is allowed, is described by the theory derived therein. Setting $L = x_0$, eqs. (8a) and (8b) in Part I are rewritten using eq. (4) as

$$F(x) = \frac{\left(\tau_c + \Delta t^{\varepsilon}_{x \to x_0}\right) F_0 + \Delta t^{\varepsilon}_{0 \to x} G(x_0)}{\tau_c + \Delta t^{\varepsilon}_{0 \to x_0}} \qquad (5a)$$

$$G(x) = \frac{\Delta t^{\varepsilon}_{x \to x_0} F_0 + \left(\tau_c + \Delta t^{\varepsilon}_{0 \to x}\right) G(x_0)}{\tau_c + \Delta t^{\varepsilon}_{0 \to x_0}}. \qquad (5b)$$

$$(0 \le x \le x_0)$$

Here, $F(x)$ and $G(x)$ divided by Planck's constant h denote the positive-velocity and the negative-velocity fluxes, respectively. In the region $x_0 \le x$, we add to the BTE the term

representing the energy relaxation due to optical phonon emission. The transition from state (x,k) is allowed to states (x,k') and $(x,-k')$ in the first relaxed level, where $(\hbar k')^2/2m = qEx + \varepsilon - \varepsilon^*$, with equal transition probability D, emitting an optical phonon. The frequency of transition is proportional to the product of the transition probability, the distribution $f(x,k)$, and the empty probability of the destination. The empty probability, however, is set to unity as discussed in Appendix B of Part I. Since the energy relaxation removes a carrier from the incident energy level, the BTE of Model 1, i.e., eq. (3) in Part I, is modified to

$$\frac{q}{\hbar}E\frac{\partial f(x,k)}{\partial k} + \frac{\hbar k}{m}\frac{\partial f(x,k)}{\partial x} + B\left[f(x,k) - f(x,-k)\right] + 2Df(x,k) = 0. \qquad (6)$$

The term $2Df(x,k)$ designates the removal of carriers in proportion to $f(x,k)$. The fluxes, $F(x)$ and $G(x)$, in the incident energy level are related to $f(x,k)$ in the same way as in Part I.

$$f(x,|k|) \equiv F(x)\delta\left(\frac{\hbar^2 k^2}{2m} - qEx - \varepsilon\right) \qquad (7b)$$

$$f(x,-|k|) \equiv G(x)\delta\left(\frac{\hbar^2 k^2}{2m} - qEx - \varepsilon\right). \qquad (7b)$$

Substituting eqs. (7a) and (7b) into eq. (6) and integrating over k, the equations for $F(x)$ and $G(x)$ in this case are derived as

$$\sqrt{\frac{2(qEx+\varepsilon)}{m}}\frac{d}{dx}\begin{pmatrix}F(x)\\G(x)\end{pmatrix} = \begin{pmatrix}-(B+2D) & B\\ -B & (B+2D)\end{pmatrix}\begin{pmatrix}F(x)\\G(x)\end{pmatrix} \qquad (x_0 \le x), \qquad (8)$$

where a matrix representation is used. Diagonalizing the matrix, we see that the solution consists of the terms expressed as $\exp\left[\pm 2\sqrt{2m(B+D)D(qEx+\varepsilon)}/qE\right]$. We solve the equation by setting $F(x_0)$ and $G(L)$ as boundary values. The $F(x)$ and $G(x)$ obtained are linear in $F(x_0)$ and $G(L)$. We introduce an averaged scattering time $\tau_{ave} \equiv 1/\sqrt{(B+D)D}$

for convenience. Then the prefactor of $F(x_0)$ in the expression of $F(x)$ is transformed to a linear combination of the terms $\exp\left(-2\Delta t^{\varepsilon}_{x\to L}/\tau_{ave}\right)$ and $\exp\left(2\Delta t^{\varepsilon}_{x\to L}/\tau_{ave}\right)$. (Quantities are converted to the simple expression in time parameters.) Similarly, the prefactor of $G(L)$ is also a linear combination of the terms $\exp\left(2\Delta t^{\varepsilon}_{x_0\to x}/\tau_{ave}\right)$ and $\exp\left(-2\Delta t^{\varepsilon}_{x_0\to x}/\tau_{ave}\right)$. Let us compare the magnitudes of these terms. The traverse times $\Delta t^{\varepsilon}_{x\to L}$ and $\Delta t^{\varepsilon}_{x_0\to x}$ are usually much larger than the scattering time, and so the magnitude of the argument of these exponential factors is much larger than unity. For an applied field of $E=10^2-10^4$ V/cm, for example, $2\sqrt{2m(B+D)D\varepsilon^*}/qE=305-3.05$. Equation (4) implies that these exponential factors with a positive argument are much larger than those with a negative argument, except for the case $x\sim L$ or $x\sim x_0$. In addition, the exponential factors with a positive argument themselves rapidly decrease as x increases from x_0 or decreases from L due to the reduction of the traverse time. A similar discussion is possible for the expression of $G(x)$. Sorting out the dominant terms and dividing them by the factor of the denominator (not explicitly shown), we finally obtain the expressions for $F(x)$ and $G(x)$ as

$$F(x)\cong\exp\left[-\frac{2\Delta t^{\varepsilon}_{x_0\to x}}{\tau_{ave}}\right]F(x_0)+\left(\sqrt{1+\frac{D}{B}}-\sqrt{\frac{D}{B}}\right)^2\exp\left[-\frac{2\Delta t^{\varepsilon}_{x\to L}}{\tau_{ave}}\right]G(L)\quad(9a)$$

$$G(x)\cong\left(\sqrt{1+\frac{D}{B}}-\sqrt{\frac{D}{B}}\right)^2\exp\left[-\frac{2\Delta t^{\varepsilon}_{x_0\to x}}{\tau_{ave}}\right]F(x_0)+\exp\left[-\frac{2\Delta t^{\varepsilon}_{x\to L}}{\tau_{ave}}\right]G(L).\quad(9b)$$

$$(x_0\le x\le L,\quad \tau_{ave}<\Delta t^{\varepsilon}_{x_0\to L})$$

2.3 Carrier transport accompanying energy relaxation

In the region $0\le x\le x_0$, eqs. (5a) and (5b), together with $\overline{T}$ in eq. (3) where $L=x_0$, yield

$$F(x_0) = \overline{T}F_0 + \left(1-\overline{T}\right)G(x_0) \qquad (10a)$$
$$G(0) = \left(1-\overline{T}\right)F_0 + \overline{T}G(x_0). \qquad (10b)$$

In the region $x_0 \le x \le L$, eqs. (9a) and (9b) are applicable. Using eqs. (10a) and (10b), as well as the expression for $G(x_0)$ derived from eq. (9b), we obtain the expressions for $F(x_0)$, $G(0)$, and $G(x_0)$ as a linear combination of F_0 and $G(L) \equiv G_L$.

$$F(x_0) = \frac{\overline{T}F_0 + (1-\overline{T})\beta G_L}{1-(1-\overline{T})\alpha} \qquad (11a)$$
$$G(0) = \frac{\left[\left(1-\overline{T}\right)-\left(1-2\overline{T}\right)\alpha\right]F_0 + \overline{T}\beta G_L}{1-(1-\overline{T})\alpha} \qquad (11b)$$
$$G(x_0) = \alpha F(x_0) + \beta G_L = \frac{\alpha\overline{T}F_0 + \beta G_L}{1-(1-\overline{T})\alpha}, \qquad (11c)$$

where

$$\alpha \equiv \left(\sqrt{1+\frac{D}{B}} - \sqrt{\frac{D}{B}}\right)^2 \qquad (12a)$$
$$\beta \equiv \exp\left\{\frac{2\sqrt{2m(B+D)D}\left[\sqrt{\varepsilon^*} - \sqrt{qE(L-x_0)+\varepsilon^*}\right]}{qE}\right\} = \exp\left(-2\frac{\Delta t^{\varepsilon}_{x_0\to L}}{\tau_{ave}}\right). \qquad (12b)$$
$$(x_0 \le x \le L, \quad \tau_{ave} < \Delta t^{\varepsilon}_{x_0\to L})$$

As for the flux injection at the boundary of the channel, we have $G_L = F_0 \exp\left(-qV/k_BT\right)$, as is suggested by eq. (B13) of Part I. If the bias V between the source and drain is sufficiently large, i.e., $qV >> k_BT$, we have $G_L << F_0$, and we can neglect the terms proportional to G_L in eqs. (11a)-(11c). Equation (11c) indicates that α represents the carrier reflection from the region $x > x_0$. The value of α depends on the trade-off between the elastic backscattering and the energy relaxation within $x_0 \le x \le L$, and eq. (12a) indicates that α is independent of E. The value amounts to 0.17 for our choice of $B = D$.

Equation (11) in Part I suggests that the net electric current I_ε from source to drain for the incident energy ε is provided by

$$I_\varepsilon = \frac{q}{h}\left[F_0 - G(0)\right] = \frac{q}{h}\frac{\overline{T}(1-\alpha)F_0 - \overline{T}\beta G_L}{1-(1-\overline{T})\alpha} = \frac{q}{h}\tilde{T}\left(F_0 - \frac{\beta}{1-\alpha}G_L\right), \qquad (13)$$

where the modified transmission coefficient $\tilde{T}$ is defined by

$$\tilde{T} \equiv \frac{\overline{T}(1-\alpha)}{1-(1-\overline{T})\alpha} = \frac{qE(1-\alpha)}{qE + \sqrt{2mB^2}\left(\sqrt{\varepsilon^*} - \sqrt{\varepsilon}\right)(1-\alpha)} = \frac{(1-\alpha)}{1+(\Delta t^{\varepsilon}_{0\to x_0}/\tau_c)(1-\alpha)}. \qquad (14)$$
$$(L > x_0)$$

Let us derive the expressions for $F(x)$ and $G(x)$ for the incident energy level assuming that $L >> x_0$ and G_L is neglected. In the case of $G_L = 0$, eqs. (5a) and (5b) are rearranged with the use of eqs. (11c) and (13) to

$$F(x) = I_\varepsilon \frac{h}{q}\left(\frac{1}{1-\alpha} + \frac{\Delta t^{\varepsilon}_{x\to x_0}}{\tau_c}\right) \quad (15a)$$

$$G(x) = I_\varepsilon \frac{h}{q}\left(\frac{\alpha}{1-\alpha} + \frac{\Delta t^{\varepsilon}_{x\to x_0}}{\tau_c}\right). \quad (15b)$$
$$(0 \le x \le x_0, \quad x_0 << L)$$

Similarly, we can transform eqs. (9a) and (9b) with the use of eqs. (13) and (14) and derive $F(x)$ and $G(x)$ in the region $x_0 \le x$ as

$$F(x) = I_\varepsilon \frac{h}{q}\frac{1}{1-\alpha}\exp\left(-\frac{2\Delta t^{\varepsilon}_{x_0\to x}}{\tau_{ave}}\right) \qquad (16a)$$

$$G(x) = I_\varepsilon \frac{h}{q}\frac{\alpha}{1-\alpha}\exp\left(-\frac{2\Delta t^{\varepsilon}_{x_0\to x}}{\tau_{ave}}\right). \qquad (16b)$$
$$(x_0 \le x, \quad x_0 << L, \quad \tau_{ave} < \Delta t^{\varepsilon}_{x_0\to L})$$

The total current density *I* of the system, considering the contribution of G_L, is obtained by substituting eq. (12) in Part I, which is reprinted as

$$\frac{F_0}{h}\times \quad \to \quad \int d\varepsilon \frac{n_0}{\sqrt{2m\pi k_B T}}\exp\left(-\frac{\varepsilon}{k_B T}\right)\times,$$

and also performing a similar substitution for G_L/h in eq. (13). Here the carrier density at

the entrance to the channel is assumed to be the same as that in the body of the channel, n_0. The range of integration over ε is limited to less than ε*, as pointed out in §2.1. For $\varepsilon \geq \varepsilon^*$, the initial elastic zone vanishes, and eq. (14) does not make sense. The contribution of carriers with energy larger than ε* cannot be taken into account in our calculation. The ratio of the incident flux with the larger energy to the total incident flux is $\exp\left(-\varepsilon^* / k_B T\right)$, as suggested by eq. (B13) in Part I, and is less than 9%. We neglect that part, since our purpose is to clarify the dominant mechanism of transport rather than to achieve numerical accuracy. Thus, for $L > x_0$, we obtain

$$I = \frac{qn_0}{\sqrt{2mk_B T\pi}}\left[1 - \frac{\beta}{1-\alpha}\exp\left(-\frac{qV}{k_B T}\right)\right]\int_0^{\varepsilon^*}\exp\left(-\frac{\varepsilon}{k_B T}\right)\tilde{T}\, d\varepsilon. \qquad (qEL >> \varepsilon^*) \qquad (17)$$

Let us examine some plots of numerical examples depicted for silicon. Figure 2 shows plots of the transmission coefficient $\tilde{T}$ as a function of electric field *E*. Equations (3) and (14) are employed for $L \leq x_0$ and $L \geq x_0$, respectively. We compare this figure with Fig. 4 in Part I. The remarkable *L*-dependence in the low-field region is unified to a curve independent of *L* in the high-field region. The curve for a sufficiently large *E* saturates at a value less than unity. Equation (3) suggests that the value of $\bar{T}$ for the initial elastic zone approaches unity for a field so large that $\Delta t^{\varepsilon}_{0\to x_0} << \tau_c$, indicating the total or ballistic transmission through the zone in $0 \leq x \leq x_0$. The reflection from the region $x > x_0$ constitutes the backward injection $G(x_0) = \alpha F_0$ at x_0, and degrades the combined transmission $\tilde{T}$ to saturate at a value of (1−α). Figure 3 shows the distribution of $n_\varepsilon(x)$ in the incident energy level shown in arbitrary units. It is evaluated using eqs. (15a), (15b), (16a), and (16b). Within the region $0 \leq x \leq x_0$ where only elastic scattering occurs, $n_\varepsilon(x)$ decreases as *x* increases, as discussed in Part I. When we proceed into the region $x_0 \leq x$, the

optical phonon emission rapidly decreases the carrier density in the incident energy level, and $n_\varepsilon(x)$ vanishes within several tens of nanometers, resulting in transitions to the first relaxed level.

Figure 4 is the total current density derived in eq. (17). The carrier density in the entrance part of the channel is assumed to be 2.5×10^{14} cm^{-3}, similar to that in Fig. 5 in Part I, but the behavior is greatly changed. The marked dependence on *L* is eliminated except for a weak symptom in the low-*E* region. The fade-out of the dependence on *L* is explained as follows. Whereas the current of Fig. 5 in Part I is controlled by backscattering within a wide region of $0-L$, the current in this case is proportional to the modified transmission $\tilde{T}$, which consists of the transmission $\overline{T}$ through the initial elastic zone between 0 and $x_0=(\varepsilon^*-\varepsilon)/qE$ and the backward injection αF_0, both of which are independent of *L*. Of course, the carrier injection at the drain edge proportional to $\exp(-qEL/k_BT)$ depends on *L*, but it is neglected for the practical case of $qEL >> k_BT$. For this high *E* value, eq. (17) is rewritten in the form

$$I=\frac{q^2n_0}{2mB}E\,\gamma(E), \qquad (qEL >> k_BT, \quad L > x_0), \qquad (18)$$

where

$$\gamma(E)=\frac{1}{\sqrt{\pi k_BT}}\int_0^{\varepsilon^*}\frac{\exp(-\varepsilon/k_BT)}{\frac{qE}{\sqrt{2mB^2}}\frac{1}{1-\alpha}+\sqrt{\varepsilon^*}-\sqrt{\varepsilon}}d\varepsilon. \qquad (19)$$

For $qE << \sqrt{2mB^2\varepsilon^*}(1-\alpha)$ (~10^4 eV/cm), the function γ(E) shows only a weak dependence on *E*, as $\gamma(E)=1.54$ for $E=10$ V/cm, 1.15 for 10^2 V/cm, 0.70 for 10^3 V/cm, and 0.21 for 10^4 V/cm. Insofar as the dependence of γ(E) on *E* is neglected, eq. (18) indicates that Ohm's law is restored; we thus call the field region the Ohm's law range.

We see a slight dependence on *L* remaining in the region of smallest *E* in Fig. 4. In this

region, where both E and L are small and $qEL < k_B T$, L is smaller than $x_{0,}$ and the transmission $\tilde{T}$ is effectively reduced to $\bar{T}$ in Part I. The lowest-order term of I in E derives from the $\bar{T}$ value of eq. (10) in Part I. When the small $\sqrt{2\varepsilon}$-term in the denominator can be neglected, the L-independent eq. (14) in Part I results. However, if the term persists, the weak dependence of I on L remains, as we see in Fig. 4. For a sufficiently high E region, the current shows saturation. This corresponds to the saturation of the transmission coefficient in Fig. 2 and is due to ballistic transmission through the initial elastic zone. The current density is expressed as

$$I \approx q\frac{n_0}{2}\sqrt{\frac{2k_B T}{\pi m}}(1-\alpha). \qquad \left(\sqrt{2mB^2\varepsilon^*}(1-\alpha) << qE, \quad x_0 < L\right), \qquad (20)$$

in contrast to eq. (18) in Part I. The entire current injected from the source to the channel transmits into the region $x > x_0$, but the part with the ratio α is reinjected into the initial elastic zone as a result of competition between elastic backscattering and optical phonon emission in that region, so the current density is reduced by a factor of $(1-\alpha)$.

2.4 The first relaxed level

We next analyze flux distribution at the first relaxed level for the case where both the channel length L and the applied bias V are sufficiently large and the carrier injection G_L from the drain is neglected. In the region of $x_0 \le x \le x_0 + x_1$, carriers in the first relaxed level are due to elastic scattering, but the energy relaxation by optical phonon emission is inhibited. On the other hand, the inflow of carriers from the incident energy level by optical phonon emission needs be considered. The BTE for carriers in this level is described as

$$\frac{q}{\hbar}E\frac{\partial f(x,k)}{\partial k}+\frac{\hbar k}{m}\frac{\partial f(x,k)}{\partial x}+B\left[f(x,k)-f(x,-k)\right]-D\left[f_0(x+x_0,k')+f_0(x+x_0,-k')\right]=0, \quad (21)$$

where $f_0(x,k')$ is the distribution function of the incident energy level, and $(\hbar k')^2/2m=(\hbar k)^2/2m+\varepsilon^*$. The term proportional to *D* denotes the inflow of carriers from the incident energy level. In discussing the first relaxed level, we move the origin of the *x*-axis from the original source edge to the point x_0 on the original scale. The functions $F(x)$ and $G(x)$ of the incident energy level with the new origin are designated $F_0(x)$ and $G_0(x)$, respectively. These functions are basically given by eqs. (9a) and (9b), respectively, except that the traverse times [defined in eq. (4)] $\Delta t^{\varepsilon}_{x_0\to x}$ and $\Delta t^{\varepsilon}_{x\to L}$ are replaced by $\Delta t^{\varepsilon^*}_{0\to x}$ and $\Delta t^{\varepsilon^*}_{x\to(L-x_0)}$, respectively. The distribution function $f_0(x+x_0,k')$ [$f_0(x+x_0,-k')$] requires the substitution of $F_0(x)$ [$G_0(x)$] for $F(x)$ [$G(x)$] in eq. (7a) [eq. (7b)]. The distribution function of the first relaxed level is expressed by $F(x)$ and $G(x)$ for the level

$$f(x,|k|)\equiv F(x)\delta(\frac{\hbar^2k^2}{2m}-qEx) \qquad (22a)$$

$$f(x,-|k|)\equiv G(x)\delta(\frac{\hbar^2k^2}{2m}-qEx). \qquad (22b)$$

We substitute eqs. (22a) and (22b), and also $f_0(x+x_0,k')$ and $f_0(x+x_0,-k')$, into eq. (21), rearrange it, and then integrate it over *k* to eliminate the δ-function factor. We finally obtain a pair of differential equations for *F*(*x*) and *G*(*x*) in the first relaxed level, as shown in eqs. (A1a) and (A1b) in the Appendix. The point $x=0$ is the turning point at which the negative-velocity carrier changes to a positive-velocity carrier, and we put $G(0)=F(0)$. For the other boundary condition for the flux equation, we use $G(x_1)$ as a given value. Figure 3 shows that $F(x)$ and $G(x)$ in the corresponding region of the incident energy level $(x_0\le x)$ quickly decay within several tens of nanometers, and we designate the decay length

as Δx, which is usually much less than x_1. Within the region $\Delta x < x \le x_1$ which constitutes most of $0 \le x \le x_1$, the functions $f_0(x+x_0,k')$ and $f_0(x+x_0,-k')$ in eq. (21) almost vanish. Equation (21) is effectively reduced to eq. (3) in Part I, and similar solutions to eqs. (15a) and (15b) are derived. Specifically, the Appendix shows the derivation, and eqs. (A7a) and (A7b) are rewritten as

$$F(x) = \frac{I_\varepsilon h}{q}\left(\frac{1}{1-\alpha} + \frac{\Delta t^0_{x\to x_1}}{\tau_c}\right) \qquad (23a)$$

$$G(x) = \frac{I_\varepsilon h}{q}\left(\frac{\alpha}{1-\alpha} + \frac{\Delta t^0_{x\to x_1}}{\tau_c}\right). \qquad (23b)$$

$$(\Delta x < x \le x_1, \quad x_0 + x_1 << L)$$

Here, $\Delta t^0_{x\to x_1}$ is the traverse time for a carrier in the first relaxed level to move from x to x_1. For the very narrow region $0 \le x \le \Delta x$, we resort to the original eqs. (A1a) and (A1b). The solution is given as eqs. (A3a)—(A3c), but the analytic integration of the expression is not available.

The carrier state in $x_1 \le x$, where elastic scattering and optical phonon emission need to be considered, is controlled by BTE in eq. (6), since the incident energy level is assumed to be empty. The flux equation eq. (7) leads to eqs. (A9a) and (A9b) to yield

$$F(x) = I_\varepsilon \frac{h}{q}\frac{1}{1-\alpha}\exp\left(-\frac{2\Delta t^0_{x_1\to x}}{t_{ave}}\right) \qquad (24a)$$

$$G(x) = I_\varepsilon \frac{h}{q}\frac{\alpha}{1-\alpha}\exp\left(-\frac{2\Delta t^0_{x_1\to x}}{t_{ave}}\right), \qquad (24b)$$

$$(x_1 \le x, \quad x_0 + x_1 << L, \quad \tau_{ave} < \Delta t^0_{x_1\to(L-x_0)})$$

which are similar to eqs. (16a) and (16b), because eqs. (23a) and (23b) are similar to eqs. (15a) and (15b). The carrier density distribution in the first relaxed level is similar to that shown in Fig. 3. In view of the behavior of carriers in the first relaxed level, we readily

conclude that the carrier state in the second relaxed level is described by eqs. (23a), (23b), (24a), and (24b) with appropriate modifications. Furthermore, all higher relaxed levels, excluding the ones close to the drain, are basically described by these equations with appropriate modifications. In the sense that almost the same unit is repeated along the channel and that averaging over the unit yields a uniform distribution throughout the bulk, we can say that the carrier state is uniform on average, although it varies microscopically and is distributed in each relaxed level just as we see in the carrier density of Fig. 3.

2.5 Mean carrier density and mean carrier velocity

Similar to the case in Fig. 3, in eqs. (23a), (23b), (24a), and (24b) the carrier density distribution is not microscopically uniform in the bulk, and neither is the carrier velocity distribution. However, the bulk part consists of repetition of the same unit structure with a period x_1 for each value of incident energy ε, and the unit structure is represented by that in the region $0 \le x \le x_1$ of the first relaxed level. We evaluate the mean carrier density in the bulk part. First we evaluate the total carrier number within the region that corresponds to the first relaxed elastic zone, i.e., $0 \le x \le x_1$, per unit area of cross section. This number consists of the contribution N_{i} from the incident energy level, estimated by eqs. (16a) and (16b), and also the contribution N_{r} from the first relaxed level derived from the solution in the level. N_{i} is obtained by integrating eq. (7) in Part I with $F(x)$ and $G(x)$, which are given respectively by eqs. (16a) and (16b) over the region $x_0 \le x \le x_0 + x_1$. The evaluation is straightforward by first shifting the origin of the incident energy level to that of the first relaxed level, and then changing the variable of integration from x to $z^* = \sqrt{2m(qEx + \varepsilon^*)}/qE$, as

$$N_i = \frac{\sqrt{m}}{\sqrt{2}h}\int_{x_0}^{x_0+x_1} dx\,\frac{F(x)+G(x)}{\sqrt{qEx+\varepsilon}} = \frac{I_\varepsilon}{q}\frac{1}{2D}\left\{1-\exp\left[\frac{2\Delta t^0_{0\to x_1}}{\tau_{ave}}\left(1-\sqrt{2}\right)\right]\right\}. \qquad (25)$$

The expression of N_r, although a little complicated, is derived in eq. (A11). The mean carrier density within the first relaxed elastic zone for a given I_ε is $(N_i + N_r)/x_1$. The second relaxed elastic zone is located beyond the first relaxed elastic zone, and the bulk region is a continuous series of elastic zones for successively higher relaxed levels. The values of N_i and N_r for higher relaxed levels are all common. The dependence of eqs. (25) and (A11) on ε is limited to the factor I_ε. To obtain the mean carrier density $\langle n\rangle$ in the bulk region, we need to sum the contributions from each I_ε. Since N_i and N_r are linear in I_ε, the carrier density is obtained by replacing I_ε in the expression for $(N_i + N_r)/x_1$ by the total current density I:

$$\langle n\rangle = \frac{I}{q}\sqrt{\frac{2m}{\varepsilon^*}}\sqrt{\frac{B+D}{D}}\left\{\frac{D}{B+D}\left[\frac{1}{2}\left(\frac{\Delta t^0_{0\to x_1}}{\tau_{ave}}\right)^{-1}+1\right]+\frac{B}{B+D}\frac{\Delta t^0_{0\to x_1}}{\tau_{ave}}\right.$$
$$\left.-\left[\frac{D}{2(B+D)}\left(\frac{\Delta t^0_{0\to x_1}}{\tau_{ave}}\right)^{-1}+\frac{\left(1-\sqrt{2}\right)B+D}{B+D}\right]\exp\left[\frac{2\Delta t^0_{0\to x_1}}{\tau_{ave}}\left(1-\sqrt{2}\right)\right]\right\}. \qquad (26)$$

Next we derive the mean carrier velocity. Denoting the summed density of the carrier for both the incident energy level and the first relaxed level at x by $n(x)$, the drift time for a carrier to traverse from $x=0$ to x_1 opposing multiple scatterings is given by

$$\int_0^{x_1}\left[\frac{I_\varepsilon}{qn(x)}\right]^{-1}dx = \frac{q}{I_\varepsilon}(N_i+N_r) \qquad (27).$$

The mean drift velocity $\langle v\rangle$ within the region $(0-x_1)$ is given by $\langle v\rangle = x_1/[(q/I_\varepsilon)(N_i+N_r)]$, and eqs. (25) and (A11) yield

$$\langle v\rangle = \sqrt{\frac{\varepsilon^*}{2m}}\sqrt{\frac{D}{B+D}}\left\{\frac{D}{B+D}\left[\frac{1}{2}\left(\frac{\Delta t^0_{0\to x_1}}{\tau_{ave}}\right)^{-1}+1\right]+\frac{B}{B+D}\frac{\Delta t^0_{0\to x_1}}{\tau_{ave}}\right.$$
$$\left.-\left[\frac{D}{2(B+D)}\left(\frac{\Delta t^0_{0\to x_1}}{\tau_{ave}}\right)^{-1}+\frac{\left(1-\sqrt{2}\right)B+D}{B+D}\right]\exp\left[\frac{2\Delta t^0_{0\to x_1}}{\tau_{ave}}\left(1-\sqrt{2}\right)\right]\right\}^{-1}. \qquad (28)$$

Notice that $\langle v\rangle$ is independent of the incident energy ε. Then we have $I = q\langle n\rangle\langle v\rangle$. Among the many terms in eqs. (26) and (28), the exponential terms reflect the influence of $F(x)$ and $G(x)$ at $x = x_0 + x_1$ in the incident energy level as inferred from eqs. (16a) and (16b), and they are sufficiently small when $\Delta t^0_{0\to x_1}/\tau_{ave}$ is large. The linear term in $\Delta t^0_{0\to x_1}/\tau_{ave}$ in the braces of eq. (28) is due to the term $\Delta t^0_{x\to x_1}/\tau_c$ in eqs. (23a) and (23b); this term represents the effect of backscattering and is responsible for the characteristic decrease of carrier distribution along the channel.

In the Ohm's law range where $1 << \sqrt{2mB^2\varepsilon^*}(1-\alpha)/qE = 2\left(\Delta t^0_{0\to x_1}/\tau_{ave}\right)(1-\sqrt{D/(B+D)})$, we can assume that $1 << \Delta t^0_{0\to x_1}/\tau_{ave}$, and the dominant contributions to $\langle n\rangle$ and $\langle v\rangle$ come from the linear term of $\Delta t^0_{0\to x_1}/\tau_{ave}$ in the braces of eqs. (26) and (28), insofar as the exponentially decaying terms are neglected. Thus we have

$$\langle n\rangle \approx \frac{I}{q}\sqrt{\frac{2m}{\varepsilon^*}}\sqrt{\frac{B+D}{D}}\left(\frac{B}{B+D}\frac{\Delta t^0_{0\to x_1}}{\tau_{ave}}\right) = \frac{I}{q}\frac{2mB}{qE} = n_0\gamma(E) \approx n_0, \qquad (29)$$

considering eq. (18), and we can see that $\langle n\rangle \approx n_0$ if we neglect the dependence of $\gamma(E)$ on E. Mean carrier velocity also is reduced to

$$\langle v\rangle \approx \sqrt{\frac{\varepsilon^*}{2m}}\sqrt{\frac{D}{B+D}}\left(\frac{B}{B+D}\frac{\Delta t^0_{0\to x_1}}{\tau_{ave}}\right)^{-1} = \frac{qE}{2mB}. \qquad (30)$$

Notice that these expressions are in agreement with the homogeneous result of the

full-energy-relaxation model discussed in Appendix A in Part I if we consider the relations $B \sim (1/2\tau)$ and $(v_+ - v_-)/2 = (q\tau/m)E$. The common term proportional to $\Delta t^{\varepsilon}_{x\to x_0}/\tau_c$ in eqs. (15a) and (15b), or that proportional to $\Delta t^{0}_{0\to x_1}/\tau_c$ in eqs. (23a) and (23b), makes a dominant contribution to the carrier density and also to the drift velocity in the Ohm's law range. We can say that the positive- and negative-velocity components contribute almost equally, and that their slight difference yields the net current in the initial elastic zone as well as in the bulk region. Such a situation is similar to what was pointed out in the full-energy-relaxation model in Appendix A of Part I. However, the distribution is not uniform along the channel and shows a spatial periodicity with a period x_1 due to the transition of carriers to higher relaxed levels. In the opposite limit, where $\Delta t^{0}_{0\to x_1}/\tau_{ave} << 1$ (or $\Delta t^{0}_{0\to x_1}/\tau_c << 1$) and $\sqrt{2m(B+D)D\varepsilon^*} << qE$, eq. (28) is reduced to $\langle v\rangle \approx (\sqrt{2}+1)\sqrt{D/(B+D)}\sqrt{\varepsilon^*/2m}$, independent of the field. This result is the velocity saturation predicted by eq. (28) and corresponds to the case where the transmission coefficient through the elastic zone of each relaxed level amounts to unity (ballistic transmission), as suggested by eq. (3) with ε= 0.

Plotted in Fig. 5 are the normalized mean carrier density $\langle n\rangle/n_0$ derived from eq. (26), the mean drift velocity in eq. (28), the normalized carrier flux $\langle n\rangle\langle v\rangle/n_0 = I/qn_0$, and $\gamma(E)$ in eq. (19) as functions of *E*. Within the Ohm's law range, $\gamma(E)$ slowly varies as a function of *E* and deviates from unity. This causes the deviation of $\langle n\rangle/n_0$ from unity as seen in the figure and yields a deficit or excess of the mean charge density compared with the background charge density $-qn_0$ in the bulk region, resulting in a violation of charge neutrality. The proportionality of the normalized carrier flux I/qn_0 to *E* in the figure seems slightly damaged due to the presence of the *E*-dependent $\gamma(E)$. The mean carrier velocity

$\langle v \rangle$ shows a clearer proportionality to the electric field in the region $E \leq 10^4$ V/cm.

In the Ohm's law range, $qE << \sqrt{2mB^2\varepsilon^*}(1-\alpha) \sim 10^4$ V/cm, the decay length Δx in the incident energy level is much smaller than x_1 of the first relaxed level, and a similar relation is verified with respect to the pair of the first and second relaxed levels; this pattern continues toward higher relaxed levels by turns. After the carrier in the incident energy level has completely relaxed to the first relaxed level, the carrier in that relaxed level begins to relax to the second relaxed level. A similar situation is realized regarding the first and second relaxed levels, and so on toward higher relaxed levels. Carriers are distributed in two energy levels at most (two-level distribution), and a procedure for deriving the mean carrier velocity in eq. (28) as the bulk value is guaranteed. If the value of Δx exceeds x_1, however, the procedure for higher relaxed levels in the bulk breaks down. Within the second relaxed elastic zone beyond $x = (x_0 + x_1)$ in Fig. 1, for example, the flux in the first relaxed level is no longer controlled by eq. (6), which describes the scattering inside the level and the transition to the second relaxed level, but is also disturbed by the transition from the incident energy level where the residual carrier is distributed. Deep in the bulk, carriers are distributed in many energy levels within the same spatial region (multilevel distribution). In such a situation, the value $\langle n \rangle$ is larger than that obtained from eq. (26) due to contribution from the increased number of energy levels. Equation (28) for the two-level distribution predicts an overestimated value since the distribution is modified. However, the mechanism of velocity saturation, identified as the ballistic transmission through the elastic zone, still works in the multi-level distribution. We can roughly estimate the validity range of the two-level distribution by comparing the modulus of argument of the exponential factor in eqs. (24a) and (24b) at $x = 2x_1$ to unity. Thus, eq. (28) is shown to be valid for $E < 5 \times 10^4$ V/cm.

3. Discussion

In Part I, we have analyzed carrier transport only with elastic scattering as Model 1. Within a wide electric field range of $V >> k_B T / q$, Model 1 anticipates that $I \propto \sqrt{E / L}$, in contrast to Ohm's law. If the field is not excessively large and $\tau_c << \Delta t^{\varepsilon}_{0 \to L}$ is satisfied, I is proportional to the transmission probability $\overline{T} \approx \tau_c / \Delta t^{\varepsilon}_{0 \to L} = qE / \left[\sqrt{2mB^2} \left(\sqrt{\varepsilon + qEL} - \sqrt{\varepsilon} \right) \right]$ from source to drain, but the probability is reduced to be proportional to $\sqrt{E / L}$ if the incident kinetic energy ε, which is on the order of $k_B T$, is sufficiently less than qEL. In Model 2, it is interesting that the *I-E* characteristics analogous to Ohm's law are restored in the same field range again, insofar as we neglect the weak dependence of the factor γ(*E*) on *E*. The optical phonon emission is suppressed within the initial elastic zone where the kinetic energy is less than the optical phonon energy. Carriers that have survived the backscattering and traversed this zone are exposed to optical phonon emission and immediately relax to the first relaxed level. Because they never return to the source, they eventually constitute part of the drain current. We can say the effective channel length L is reduced to the width x_0 of this zone. Since the kinetic energy of the carrier at $x = x_0$ is ε*, the transmission probability is reduced to $\overline{T} \approx \tau_c / \Delta t^{\varepsilon}_{0 \to x_0} = qE / \left[\sqrt{2mB^2} \left(\sqrt{\varepsilon^*} - \sqrt{\varepsilon} \right) \right]$; this expression is proportional to E without dependence on the total L. After integration over ε, the proportionality of the current density to E is maintained. The factor $\sqrt{E / L}$ is reduced to being proportional to E, because the zone width x_0 is inversely proportional to E. The magnitude of the current density is dominated by elastic scattering within the initial elastic zone, as we see in its inverse proportionality to B in eq. (18). We note that the expression is similar to eq. (1). The parameter τ in eq. (1) is the relaxation time of the distribution function due to acoustic phonon

scattering. Here, B is the backscattering probability due to the elastic scattering, but the mechanism of transport is completely different, as is clear from the discussion above and in Appendix A of Part I. Such a mechanism of current control also works even if the scattering is not purely elastic but includes a slight energy relaxation, insofar as the kinetic energy accumulates toward the drain. Note that part of the backscattering is actually caused by optical phonon emission. The current component injected at the source edge with energy larger than ε* is partly backscattered to the source by optical phonon emission and causes the net current to decrease. The contribution is small, as was pointed out before, and we neglect it in this analysis.

At sufficiently high fields, the observed current density is known to saturate. Model 1 yields the current saturation at such a high field that the traverse time $\Delta t^{\varepsilon}_{0\to L}$ becomes sufficiently smaller than τ_c and that the total ballistic transmission occurs through the channel. Then the current density at saturation is equal to the injected flux from source to channel multiplied by the carrier charge, but the field required is exceedingly high unless the channel length L is very small. In Model 2, on the other hand, the effective channel length is reduced to the width of the initial elastic zone, and the current density saturates when the transmission coefficient through this zone amounts to unity. The initial elastic zone is thin, and so the required electric field for saturation is reduced to a realistic order of magnitude 10^4 V/cm. Part of the flux, after it has been transferred beyond the initial elastic zone, is backscattered into the zone again, surviving the optical phonon emission. The ratio α of backscattering is irrelevant to the applied field, and the value of the saturated current is reduced by a factor of $(1-\alpha)$ compared with the original injected flux. The saturated current is eventually provided by eq. (20). The current saturation is based on the total transmission of the injected flux

through the initial elastic zone and is not a consequence of the carrier velocity truncation caused by optical phonon emission. Truncation always works in Model 2, but the current increases as the field increases when the transmission is sufficiently small and tends to saturate if the transmission approaches unity.

The carrier velocity in the semiconductor bulk has been measured using the time-of-flight technique[24]. The magnitude of the velocity is confirmed to increase with an increase in the applied field and tends toward saturation at the highest field. This theory anticipates that the velocity of the carrier injected from the source with fixed energy periodically oscillates in the bulk as the carrier is successively transmitted to higher relaxed levels. Within each elastic zone, it varies similarly as in Fig. 3 in Part I and does not remain constant. The mean velocity $\langle v \rangle$ averaged over the period is uniform throughout the bulk region and is given by eq. (28). As shown in Fig. 5, $\langle v \rangle$ is clearly proportional to the electric field and tends to saturate toward the high-field region over 10^4 V/cm. The overall flux distribution within the elastic zone of each relaxed level is represented by the simple expression of eqs. (23a) and (23b), where the second terms in the brackets on the right-hand side are predominant in the Ohm's law range. The carrier density sharply decreases toward the drain within the elastic zone, similar to Fig. 3. The smooth transmission is hindered by the perpetual backscattering, and the carrier flow becomes stagnant. As E is increased, transmission through the zone, given by eq. (16) in Part I, is increased and the stagnancy of the carrier is relaxed. The mean velocity $\langle v \rangle$ in eq. (28) is reduced to that in eq. (30), yielding the proportionality to E.

In contrast, in a sufficiently high field where $\Delta t^0_{0\to x_1} / \tau_c << 1$, the transmission coefficient through the elastic zone tends toward unity as eq. (3) applied to the zone indicates, and the carrier velocity saturates. More precisely, the carrier transport in an energy level is controlled

by both backscatterings: one within the elastic zone, and the other back into the elastic zone from the region beyond. The backscattering probability within the elastic zone is naturally proportional to the traverse time of a carrier across the zone. The traverse time decreases as E is increased, leading to the proportionality of I to E. The backscattering probability loses its dependence on E when $\Delta t^0_{0\to x_1} / \tau_c << 1$ is attained in a sufficiently high field. The region beyond the elastic zone has no specific size, contrary to the elastic zone. The scattering effect therein depends on the tradeoff between the backscattering time and the energy relaxation time, showing no dependence on E. Thus the proportionality to the field in the low-field region turns into independence of E in the higher-field region. On the other hand, the carrier distribution scheme in an extremely high field changes from the two-level to the multi-level distribution as pointed out previously. The qualitative mechanism of velocity saturation depicted here is also valid in the multi-level distribution. However, the saturation velocity in eq. (28), evaluated in the two-level distribution, is overestimated, as previously discussed. The rigorous evaluation of saturation velocity in the multi-level distribution is outside the scope of this work. The conventional theory of velocity saturation is based on the balance-of-energy equation and predicts that the carrier velocity will saturate if the optical phonon scattering ever dominates the energy relaxation. However, our theory indicates that the velocity is proportional to E if the transmission is small, even if the optical phonon emission is dominant. The mechanism differs between the conventional and proposed theories.

The current-controlling mechanism discussed so far implies that the current density is dominated by a very thin zone at the interface to the source electrode. In actuality, however, the feedback control from the bulk part regulates the current as follows. The bulk of the

channel consists of the repetition of the same unit structure of length x_1, which is similar to the structure in the initial elastic zone. Once the current is set in the initial elastic zone, the flux distribution in the first relaxed level is determined as in eqs. (23a) and (23b), as well as in eqs. (24a) and (24b). The flux in the second relaxed level, and also those in the higher relaxed levels, are determined similarly. However, the resultant carrier charge distribution may not be consistent with the original constant field distribution. For the carrier injection described by substitution (12) in Part I, for example, a carrier charge excess ($E < 200$ V/cm) or deficit ($E > 200$ V/cm) results, as the plot of $\langle n \rangle / n_0$ in Fig. 5 suggests. The carrier charge excess (deficit) pushes up (pulls down) the potential profile of the original constant-field curve in the bulk, which modifies the barrier height at the source-channel junction, decreases (increases) the carrier injection, and compensates for the excess (deficit) of the charge. The potential profile in the thin interface region is modified by the feedback from the bulk part. At room temperature, a change in the barrier height of only 18 meV modulates the effective carrier density at the entrance of the channel by a factor of 2 and modifies the current density by the same factor. The steady-state potential profile actually realized is obtained by the self-consistent solution of the coupled system of the BTE and the Poisson equation. Does the self-consistent solution support the constant field? The solution demands the optimization of the system's electrostatic energy, which is realized by the carrier charge distribution that tends to overlap and cancel the dopant charge distribution to eliminate the space charge in the bulk. The feedback works toward $\langle n \rangle = n_0$. For carriers with a specific energy ε, the charge distribution shows a repetition of the unit structure with a short period x_1, and the repetition begins at $x=x_0$ ($x_0 = (\varepsilon^* - \varepsilon)/qE$). The position of each unit region shifts as the value of the incident energy ε is varied, and the charge distribution tends to be leveled off by integration

over ε, as a simple estimation can easily verify. Then the dopant charge distribution cancels the average carrier charge through the feedback, promoting the constant field distribution.

If the feedback controls the current injection and fully neutralizes the average charge in the bulk, the current density is described by $qn_0\langle v\rangle$ because $\langle n\rangle = n_0$. Two current curves, $qn_0\langle v\rangle$ (solid line) and $q\langle n\rangle\langle v\rangle$ (dotted line), are compared in Fig. 6. In contrast to the languishing curve $q\langle n\rangle\langle v\rangle$, the curve $qn_0\langle v\rangle$ shows a clearer proportionality to *E*. The ultimate proportionality in the field range $qE << \sqrt{2mB^2\varepsilon^*}(1-\alpha)$ (Ohm's law range) is due to the feedback from the bulk region. We call this the high-field Ohm's law, since the mechanism supporting the proportionality is distinct from that in the low field.

In Fig. 6 we check the agreement between the estimated current density as well as the carrier velocity and the corresponding experimental data to see if the dominant mechanism of transport is correctly captured. The curve with the empty circles shows Ryder's experimental data [1] on the high-field transport of *n*-type silicon; although these data are a half-century old, they remain reliable. They are reproduced from his paper, according to which the measurement was obtained at 298 K, and the ordinate scale of his figure suggests that the sample's carrier density was identified as 2.6×10^{14} cm^{-3} by the low-field mobility measurement. A high-purity sample was investigated in accordance with our set of parameters: $n_0 = 2.5\times10^{14}$ cm^{-3} and $B = 2.5\times10^{12}$ s^{-1}. As for the value of parameters used for theoretical estimation, the value *B* given by the low-field mobility offers no other choice, but we tentatively use the value $D = 1.25\times10^{12}$ s^{-1}. This choice does not affect the curve in the Ohm's law region controlled by *B*, but it may improve the part of the curve close to saturation. The line with solid triangles shows the drift velocity measured by the time-of-flight measurement at 300 K reported by Canali *et al.*[25], and the dashed solid line shows the

averaged carrier velocity in eq. (28). The same parameters are also used for the velocity curve. The agreement between our curves and the experimental data is satisfactory. The ~10% disagreement is modest if we consider that our simple theory does not include the flux component with incident energy larger than ε*. The discrepancy in the highest-field region is attributed to the fact that the multi-level distribution is outside the scope of our theory. The overall features of the experimental data are well reproduced regardless of the use of pseudo-one-dimensional BTE, which neglects various secondary effects. The primary mechanism of transport is correctly captured in the simple theory.

One may suspect that the selection of B to fit the low-field mobility automatically implies agreement with the current- or the velocity-field curve in the linear region in Fig. 6. However, this suggestion may not be correct. The low-field linearity in eq. (1) was guaranteed only in the region $E \leq 490\,\mathrm{V/cm}$, and may not apply in higher fields, as discussed in Part I. The above agreement indicates a new fact that the linearity is guaranteed also in the high-field range where $E \geq 500\,\mathrm{V/cm}$, and that the linearity is controlled by the same low-field parameter B despite the difference in the physical mechanism.

In our analysis, the model's simplicity is essential for the elucidation of complicated transport physics. We briefly discuss the model's validity. As detailed in Part I, our analysis uses the pseudo-one-dimensional model, which is based on the assumption that the longitudinal kinetic energy and transverse kinetic energy are separately conserved on average in elastic scattering. In actuality, however, an energy exchange occurs between the two energy components. The extreme case of close exchange is roughly estimated by assuming that one-third of the kinetic energy gain from the longitudinal field is used for the acceleration of the longitudinal motion on average. In Model 1, such an effect can be evaluated by

substituting (E/3) for E in the basic eqs. (5a) and (5b) of Part I. In Model 2 in addition, the values of x_0 and x_1 remain unaltered regardless of their dependence on E, because the total kinetic energy controls the optical phonon emission. Thus, the current $I \propto \sqrt{E/x_0}$ is reduced to $I \propto \sqrt{E/3x_0}$ and the relevant curves in Figs. (4)–(6) require a slight horizontal shift of $\ln(\sqrt{3})$. Since this is the extreme case, one can conclude that the qualitative features of the result are not damaged by neglect of the three-dimensional effect.

This analysis neglects carrier transition due to optical phonon absorption. The carrier injection from the drain edge is also neglected. Due to neglect of these secondary effects, this analysis does not clarify the carrier distribution close to the drain edge. However, the primary features of transport discussed here are valid, except when an extremely short channel device is concerned. The analysis employs a parabolic energy band and neglects the energy dependence of the scattering probabilities B and D. The premise is justified when the transport is described by the two-level distribution model, and the effective energy distribution is confined to a narrow region with width less than ~ε*. A highly doped semiconductor with degenerate carriers, typically with $n > 10^{18} - 10^{19}$ cm^{-3} in silicon[26], is also excluded, since Maxwell-Boltzmann statistics are used.

The effect of the inelastic electron-electron (EE) scattering is neglected in our analysis. According to Pines and Bohm[27], EE interaction consists of two components: a long-range component associated with the collective plasma excitations (plasmons) and a short-range, single-particle component. As for the effect of electron-plasmon interaction, Fischetti[28] pointed out that the break-even point between Landau damping and collisional damping occurs in the range $10^{17} \le n \le 10^{18}$ cm^{-3} in n-type silicon, and that the electron-plasmon interaction does not affect the mobility directly in the lower-density region where Landau

damping dominates. He also concluded that the short-range EE scattering does not remove momentum from the electron ensemble, so that it can affect the current density only indirectly by modifying the distribution function. The distribution function is increased by short-range EE scattering in the high-energy tail of the energy distribution[29]. However, the increased magnitude is less than 10^{-5} of the peak distribution when the carrier density is 10^{17} cm^{-3}, for example. We can see that the electric field distribution is dominated by the main body of the carrier charge and is not affected by such a low-level increase. To summarize, the effect of EE interaction can be neglected for n-silicon when the carrier density is less than $10^{17} - 10^{18}$ cm^{-3}. As for the validity range associated with the field strength, the electric field must be within $5\times10^{2} \le E \le 5\times10^{4}$ V/cm in our theory for n-silicon. The lower-field region is controlled by the conventional full-energy-relaxation model, and the higher-field region is dominated by the multi-level distribution.

4. Conclusions

A semiconductor system equipped with elastic scattering, as well as inelastic scattering due to optical phonons with a comparatively large energy ε*, is analyzed. The acoustic phonon scattering is counted in the elastic scattering, as in conventional analysis. In silicon, ε*=63 meV, which is much larger than the thermal energy, and the inelastic scattering is dominated by energy relaxation due to optical phonon emission. The source electrode injects thermal carriers into the channel. While they pass through the initial elastic zone, where the kinetic energy of the carrier is less than the optical phonon energy ε*, the transport is controlled by the elastic scattering. Beyond the elastic zone, the carrier energy relaxes to the first relaxed energy level, emitting an optical phonon. Within the first relaxed energy level,

the kinetic energy of the carrier is less than ε* at first (the first relaxed elastic zone), and then it relaxes to the lower energy level beyond the zone. The carrier energy relaxes along a cascade of energy levels in the bulk of the semiconductor. The current-voltage characteristics are closely related to the carrier transmission through these elastic zones. In the electric field range in which the transmission coefficient is much less than unity, the proportionality of the current to the electric field E results, similar to the conventional Ohm's law in eq. (1), but the mechanism of proportionality is distinct from that in eq. (1). The proportionality arises because the elastic zone has a finite width inversely proportional to E, and the transmission coefficient is inversely proportional to the square root of the width. In contrast, eq. (1) is for a homogeneous borderless system.

In the higher-field range where the transmission coefficient approaches unity, the current density tends to saturate. The saturation of the current density and of the carrier velocity is understood as the ballistic transmission of carriers through these elastic zones within the bulk of the channel. The averaged carrier velocity increases in proportion to the field when the transmission coefficient is much less than unity, and it tends to saturate when the coefficient approaches unity. The current density is basically provided by the product of the carrier velocity, the carrier charge, and the carrier density equal to the doping concentration of the bulk semiconductor. The carrier injection from the source electrode to the channel is controlled by the electrostatic feedback from the semiconductor bulk to minimize the system's electrostatic energy.

Acknowledgement

The author thanks A. Natori for her continuous support throughout the investigation.

Appendix: Flux States in the First Relaxed Level

The flux state within the region $0 \le x \le x_1$ in the first relaxed level is derived from the following pair of flux equations.

$$\sqrt{\frac{2}{m}qEx}\frac{dF(x)}{dx} + B\left[F(x) - G(x)\right] - D\sqrt{\frac{qEx}{qEx+\varepsilon^*}}\left[F_0(x) + G_0(x)\right] = 0, \qquad \text{(A1}a\text{)}$$

$$-\sqrt{\frac{2}{m}qEx}\frac{dG(x)}{dx} + B\left[G(x) - F(x)\right] - D\sqrt{\frac{qEx}{qEx+\varepsilon^*}}\left[F_0(x) + G_0(x)\right] = 0. \qquad \text{(A1}b\text{)}$$

Changing the variable from x to $z \equiv \sqrt{2mx/qE}$ and then designating $F(x)$ and $G(x)$ as $\Phi(z)$ and $\Gamma(z)$, respectively, eqs. (A1a) and (A1b) are converted to a simple form:

$$\frac{d\Phi(z)}{dz} + B\left[\Phi(z) - \Gamma(z)\right] - DH(z) = 0, \qquad \text{(A2}a\text{)}$$

$$-\frac{d\Gamma(z)}{dz} + B\left[\Gamma(z) - \Phi(z)\right] - DH(z) = 0, \qquad \text{(A2}b\text{)}$$

where, according to the definitions of $F_0(x)$ and $G_0(x)$ in §2.3, we have

$$\begin{aligned} H(z) &= \frac{qEz}{\sqrt{(qEz)^2 + 2m\varepsilon^*}}\left[F_0\left(\frac{qEz^2}{2m}\right) + G_0\left(\frac{qEz^2}{2m}\right)\right] \\ &= \frac{(1+\alpha)qEz}{\sqrt{(qEz)^2 + 2m\varepsilon^*}}\left(F_0(0)\exp\left\{\frac{2\sqrt{(B+D)D}\left[\sqrt{2m\varepsilon^*} - \sqrt{(qEz)^2 + 2m\varepsilon^*}\right]}{qE}\right\}\right. \\ &\left. + G_0(L - x_0)\exp\left\{\frac{2\sqrt{(B+D)D}\left[\sqrt{(qEz)^2 + 2m\varepsilon^*} - \sqrt{2mqE(L-x_0) + 2m\varepsilon^*}\right]}{qE}\right\}\right). \end{aligned} \qquad \text{(A2}c\text{)}$$

The point $x = 0$ ($z = 0$) is the turning point of a carrier where the negative-velocity carrier changes to a positive-velocity carrier, and we set $\Phi(0) = \Gamma(0)\ (= F(0))$. As the other boundary condition for eq. (A2), we use $\Gamma(z_1) = G(x_1)$ where $z_1 = \sqrt{2mx_1/qE}$. Summing the result of eqs. (A2a) and (A2b) yields a differential equation of $\left[\Phi(z) - \Gamma(z)\right]$, and the

solution is obtained by integrating the result. Similarly, taking the difference in the results of eqs. (A2a) and (A2b), we have a differential equation for $[\Phi(z)+\Gamma(z)]$. After solving this equation by substituting the previous solution and then rearranging the result by integration by parts, we obtain

$$\Phi(z) = D\left[\int_0^z dz\, H(z)\right](1-2Bz) + D\left[\int_0^{z_1} dz\, H(z)\right](1+2Bz_1) - 2BD\left[\int_z^{z_1} dz\, zH(z)\right] + \Gamma(z_1) \quad \text{(A3}a\text{)}$$

$$\Gamma(z) = -D\left[\int_0^z dz\, H(z)\right](1+2Bz) + D\left[\int_0^{z_1} dz\, H(z)\right](1+2Bz_1) - 2BD\left[\int_z^{z_1} dz\, zH(z)\right] + \Gamma(z_1) \quad \text{(A3}b\text{)}$$

$$F(0) = D\left[\int_0^{z_1} dz\, H(z)\right](1+2Bz_1) - 2BD\left[\int_0^{z_1} dz\, zH(z)\right] + \Gamma(z_1). \quad \text{(A3}c\text{)}$$

$$(0 \le z \le z_1)$$

First we briefly discuss the flux state in $x_1 \le x$. The carrier density in the incident energy level is negligible within this region, as Fig. 3 (in $x_0 + x_1 \le x$) suggests, and the inflow from this level to the first relaxed level may be neglected. The optical phonon emission occurs in the first relaxed level and causes the outflow of carriers to the second relaxed level. Therefore, the carrier dynamics within this region are described by eq. (6), and the solution is basically provided by eqs. (9a) and (9b) appropriately modified. Since the origin here is at the point x_0 from the source edge, we substitute x_1, 0, and $(L-x_0)$ for x_0, ε, and L, respectively. Equation (9) indicates that $G(x_1) = \alpha F(x_1)$ when $G(L\text{-}x_0)$ can be neglected. Now we return to the carrier state in $0 \le x \le x_1$. According to the definition and discussion of Δx in §2.4, the function $H(z)$ in eqs. (A3a)—(A3c) almost vanishes within the region of z corresponding to $\Delta x < x \le x_1$. The integration of $H(z)$ from 0 to a value of z larger than $\sqrt{2m\Delta x/qE}$ is equivalent to the integration from 0 to z_1, because the integration from the value of z to z_1 almost vanishes. Equations (A3a) and (A3b) are reduced to

$$F(x) = 2D\left[\int_0^{z_1} dz\, H(z)\right]\left[1+\frac{\sqrt{2mB^2}}{qE}\left(\sqrt{\varepsilon^*}-\sqrt{qEx}\right)\right]+G(x_1) \qquad \text{(A4}a\text{)}$$

$$G(x) = 2D\left[\int_0^{z_1} dz\, H(z)\right]\frac{\sqrt{2mB^2}}{qE}\left(\sqrt{\varepsilon^*}-\sqrt{qEx}\right)+G(x_1) \qquad \text{(A4}b\text{)}$$

$$(\Delta x < x \le x_1).$$

Substituting $x = x_1$ in eq. (A4a) and using $G(x_1) = \alpha F(x_1)$, we obtain

$$G(x_1) = \frac{2\alpha D}{1-\alpha}\int_0^{z_1} dz\, H(z). \qquad \text{(A5)}$$

Some integrals in eqs. (A3a)—(A3c) can be calculated by changing the variable from z to $z^* = \sqrt{z^2 + 2m\varepsilon^*/(qE)^2}$. Using $H(z)$ in eq. (A2c) where $F_0(x)$ and $G_0(x)$, respectively, are derived from eqs. (16a) and (16b) by substituting 0 and ε* for x_0 and ε (moving the origin to x_0), we can evaluate the integral

$$D\int_0^{z_1} dz\, H(z) = \frac{1}{2}\frac{I_\varepsilon h}{q}\left\{1-\exp\left[\frac{2\sqrt{2m(B+D)D\varepsilon^*}}{qE}\left(1-\sqrt{2}\right)\right]\right\} \approx \frac{1}{2}\frac{I_\varepsilon h}{q}, \qquad \text{(A6)}$$

except in the case of an excessively high field. Substituting these values into eqs. (A4a) and (A4b), we obtain

$$F(x) = \frac{I_\varepsilon h}{q}\left[\frac{1}{1-\alpha}+\frac{\sqrt{2mB^2}}{qE}\left(\sqrt{\varepsilon^*}-\sqrt{qEx}\right)\right] \qquad \text{(A7}a\text{)}$$

$$G(x) = \frac{I_\varepsilon h}{q}\left[\frac{\alpha}{1-\alpha}+\frac{\sqrt{2mB^2}}{qE}\left(\sqrt{\varepsilon^*}-\sqrt{qEx}\right)\right]. \qquad \text{(A7}b\text{)}$$

$$(\Delta x < x \le x_1)$$

For the narrow region $0 \le x \le \Delta x$, we return to the original eqs. (A3a) and (A3b). However, one indefinite integral in these equations is not analytically evaluated, and explicit expressions for $F(x)$ and $G(x)$ are unavailable. The value of $F(0)\,[= G(0)]$ is given by eq. (A3c), where the first integral on the right-hand side is given in eq. (A6). The second

integral is evaluated by changing the variable from z to $y^* = \sqrt{z^2 - 2m\varepsilon^*/(qE)^2} - \sqrt{2m\varepsilon^*/(qE)^2}$ for a realistic value of the electric field satisfying $2\sqrt{2m(B+D)D\varepsilon^*}/qE >> 1$. Thus, after some manipulations we obtain the approximation

$$F(0) \approx \frac{I_\varepsilon h}{q}\left\{\left[\frac{1+\alpha}{2(1-\alpha)} + \frac{\sqrt{2mB^2\varepsilon^*}}{qE}\right] - \frac{\sqrt{\pi}B}{2\sqrt{2(B+D)D}}\left[\frac{2\sqrt{2m(B+D)D\varepsilon^*}}{qE}\right]^{1/2}\right\}. \quad \text{(A8)}$$

The solution in the region $x_1 \le x$ is basically given by eqs. (9a) and (9b) appropriately modified, as mentioned. Terms proportional to $G(L)$ are neglected. $F(x_0)$ is replaced by $F(x_1)$, and $F(x_1)$ results by substituting $x = x_1 = \varepsilon^*/qE$ into eq. (A7a). Thus,

$$F(x) = I_\varepsilon \frac{h}{q}\frac{1}{1-\alpha}\exp\left[\frac{2\sqrt{2m(B+D)D}\left(\sqrt{qEx_1} - \sqrt{qEx}\right)}{qE}\right] \qquad \text{(A9}a\text{)}$$

$$G(x) = I_\varepsilon \frac{h}{q}\frac{\alpha}{1-\alpha}\exp\left[\frac{2\sqrt{2m(B+D)D}\left(\sqrt{qEx_1} - \sqrt{qEx}\right)}{qE}\right]. \qquad \text{(A9}b\text{)}$$

$$(x_1 \le x)$$

Lastly, we evaluate the carrier number N_r within the region $0 \le x \le x_1$ per unit cross section. The integration of an expression similar to eq. (7) in Part I is simplified by changing the variable from x to z. Substituting Φ(z) and Γ(z) in eqs. (A3a) and (A3b) for $F(x)$ and $G(x)$ into eq. (7) in Part I, and transforming the integral by integration by parts, we obtain

$$N_r = \frac{1}{h}\int_0^{z_1} dz\left[\Phi(z) + \Gamma(z)\right] = \frac{2}{h}\left\{D\left[\int_0^{z_1} dz\, H(z)\right](1 + 2Bz_1)z_1 - 2BD\int_0^{z_1} dz\, z\int_0^{z} dz\, H(z) - 2BD\int_0^{z_1} dz\, z^2 H(z) + z_1\Gamma(z_1)\right\}. \quad \text{(A10)}$$

The first integral on the right-hand side is given in eq. (A6). The second and third integrals on the right-hand side also are evaluated by changing the variable of integration from z to z*. Finally we obtain,

$$N_r = \frac{I_\varepsilon}{qD}\left\{\frac{\Delta t^0_{0\to x_1}}{\tau_{ave}} + \frac{B}{B+D}\left[\left(\frac{\Delta t^0_{0\to x_1}}{\tau_{ave}}\right)^2 - \frac{\Delta t^0_{0\to x_1}}{\tau_{ave}} - \frac{1}{2}\right]\right.$$
$$\left. - \left[\frac{\Delta t^0_{0\to x_1}}{\tau_{ave}} - \frac{B}{B+D}\left(\frac{1}{2} + \sqrt{2}\,\frac{\Delta t^0_{0\to x_1}}{\tau_{ave}}\right)\right]\exp\left[\frac{2\Delta t^0_{0\to x_1}}{\tau_{ave}}\left(1-\sqrt{2}\right)\right]\right\}. \qquad \text{(A11)}$$

1) E. J. Ryder: Phys. Rev. **90** (1953) 766.

2) W. Shockley: Bell Syst. Tech. J. **30** (1951) 990.

3) F. J. Blatt: in *Solid State Physics*, ed. F. Seitz and D. Turnbull (Academic, New York, 1957) vol. 4.

4) J. M. Ziman: *Principles of the Theory of Solids* (Cambridge University Press, Cambridge, U.K., 1972) 2nd ed.

5) W. Shockley: *Electrons and Holes in Semiconductors* (Van Nostrand, Princeton, NJ, 1950).

6) N. W. Ashcroft and N. D. Mermin: *Solid State Physics* (Saunders College, Philadelphia, PA, 1976).

7) K. Seeger: *Semiconductor Physics* (Springer, Berlin, 1989).

8) E. M. Conwell: in *High Field Transport in Semiconductors*, ed. F. Seitz, D. Turnbull, and H. Ehrenreich (Academic, New York, 1967).

9) H. G. Reik and H. Risken: Phys. Rev. **126** (1962) 1737.

10) P. M. Smith, M. Inoue, and J. Frei: Appl. Phys. Lett. **37** (1980) 797.

11) K. Natori: submitted to Jpn. J. Appl. Phys.

12) K. Natori and T. Shimizu: Ext. Abstr. Solid State Devices and Materials, 2006, p. 348.

13) K. Natori: Phys. Status Solidi C **5** (2008) 111.

14) For example, M. Levinshtein, S. Runyanstev, and M. Shur: *Handbook Series on Semiconductor Parameters* (World Scientific, Singapore, 1996) Vol. 1, p. 1.

15) C. Jacoboni and L. Reggiani: Rev. Mod. Phys. **55** (1983) 645.

16) G. A. Baraff: Phys. Rev. **128** (1962) 2507.

17) M. Lundstrom: IEEE Electron Device Lett. **18** (1997) 361.

18) K. Natori: IEICE Trans. Electron. **E84-C** (2001) 1029.

19) K. Natori: Appl. Surf. Sci. **254** (2008) 6194.

20) S. Datta, F. Assad, and M. S. Lundstrom: Superlattices Microstruct. **23** (1998) 771.

21) F. Berz: Solid-State Electron. **28** (1985) 1007.

22) M. Lundstrom and Z. Ren: IEEE Trans. Electron Devices **49** (2002) 133.

23) R. Clerc, P. Palestri, and L. Selmi: IEEE Trans. Electron Devices **53** (2006) 1634.

24) C. Canali, C. Jacoboni, F. Nava, G. Ottaviani, and A. Quaranta: Phys. Rev. B **12** (1975) 2265.

25) C. Canali, G. Ottaviani, and A. A. Quaranta: J. Phys. Chem. Solids **32** (1971) 1707.

26) Y. Taur and T. H. Ning: *Fundamentals of Modern VLSI Devices* (Cambridge University Press, Cambridge, U.K., 1998).

27) D. Pines and D. Bohm: Phys. Rev. **85** (1952) 338.

28) M. V. Fischetti: Phys. Rev. B **44** (1991) 5527.

29) P. A. Childs and C. C. C. Leung: Electron. Lett. **31** (1995) 139.

Figure Captions

Fig. 1. Potential profile and carrier transport where the elastic scattering and the energy relaxation due to optical phonon emission are considered (Model 2). Along the potential profile, the carrier energy relaxes stepwise by an amount equal to the optical phonon energy ε* toward the drain. The elastic zone denotes regions where the kinetic energy of the carrier is less than ε* and the optical phonon emission is suppressed.

Fig. 2. Transmission coefficient from source to drain in Model 2 as a function of the applied field for various values of L.

Fig. 3. Carrier density $n_\varepsilon(x)$ in the incident energy level along the channel. The position x is normalized by x_0. For $x < x_0$, the kinetic energy of the carrier is less than ε* and the optical phonon emission is suppressed. Distribution of the carrier is caused by elastic scattering. The position x_1 in the first relaxed level is at $(1 + x_1 / x_0)$ in this figure.

Fig. 4. Current density flowing from the source to the drain in Model 2 as a function of the applied field for various values of L.

Fig. 5. Mean carrier velocity $\langle v \rangle$ in the bulk, $I/(qn_0)$, $\gamma(E)$, and $\langle n \rangle / n_0$, as a function of the applied field. $\gamma(E)$ is maintained close to unity in the Ohm's law range. The deviation of $\langle n \rangle / n_0$ from unity implies a violation of charge neutrality within the bulk region.

Fig. 6. Current density $I = q\langle n \rangle \langle v \rangle$ without charge neutrality and the modified current density $qn_0 \langle v \rangle$, where the charge neutrality within the bulk is recovered, compared with the experimental data reported by Ryder [1]. The averaged velocity $\langle v \rangle$ also is compared with the experimental drift velocity reported by Canali *et al* [25].

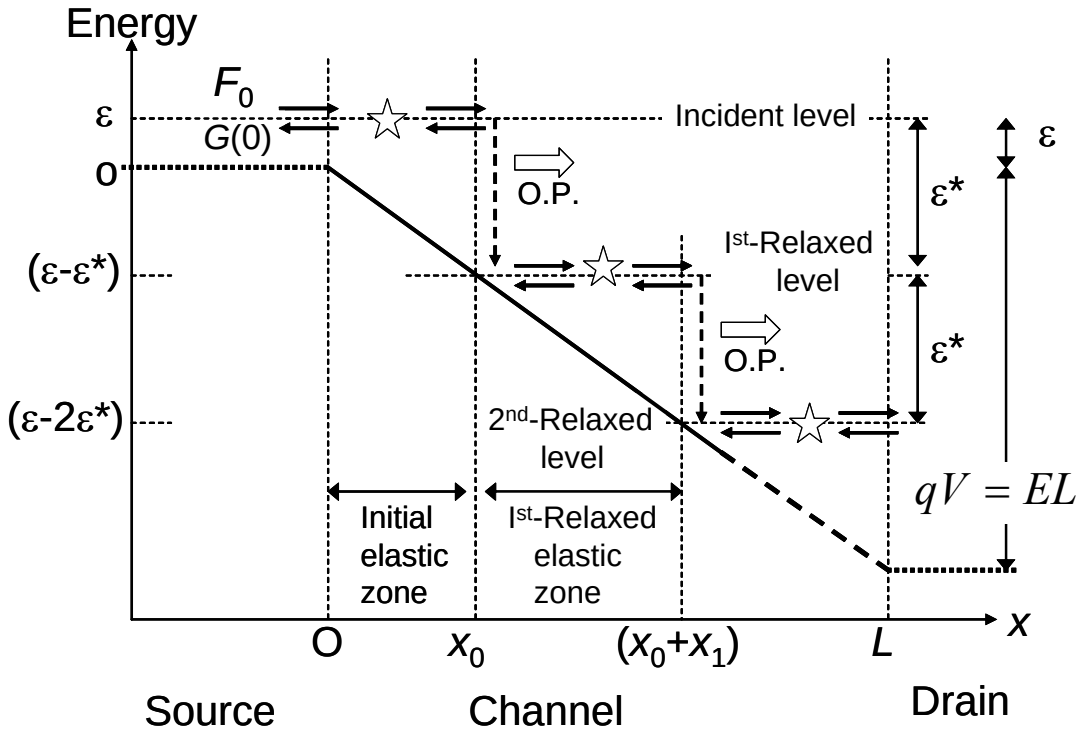
Energy
F_0
ε
$G(0)$
o
(ε-ε*)
(ε-2ε*)
Incident level
O.P.
Iˢᵗ-Relaxed level
O.P.
2ⁿᵈ-Relaxed level
ε*
ε*
ε
$qV = EL$
Initial elastic zone
Iˢᵗ-Relaxed elastic zone
O
x_0
(x_0+x_1)
L
x
Source
Channel
Drain

Fig. 1

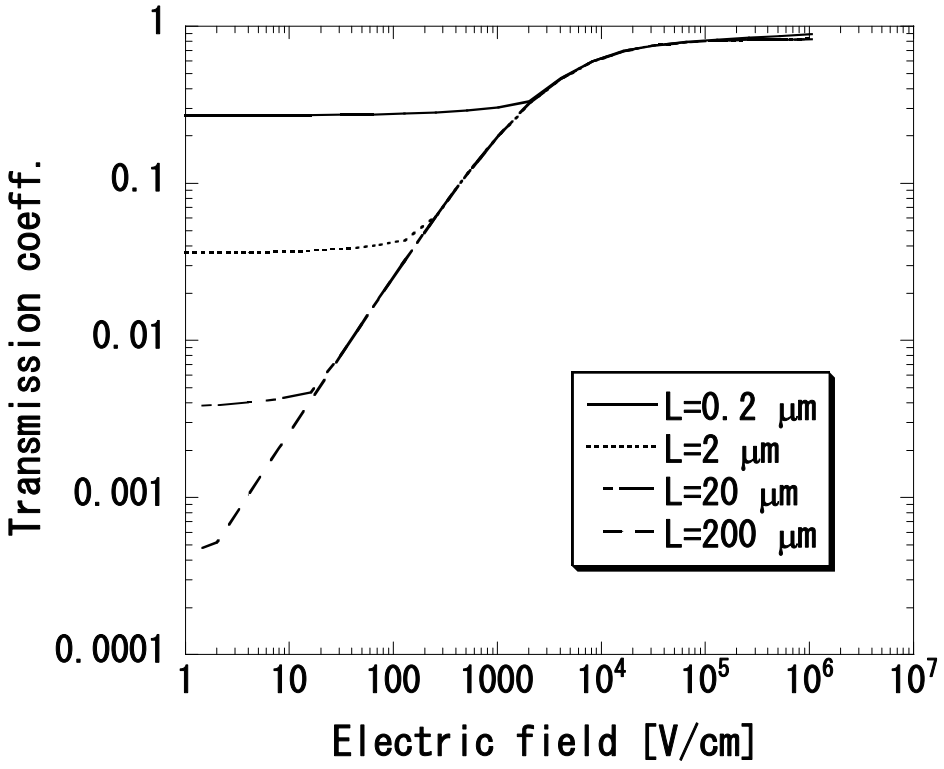
Transmission coeff.
1
0.1
0.01
0.001
0.0001
1
10
100
1000
10^4
10^5
10^6
10^7
Electric field [V/cm]
L=0.2 μm
L=2 μm
L=20 μm
L=200 μm

Fig. 2

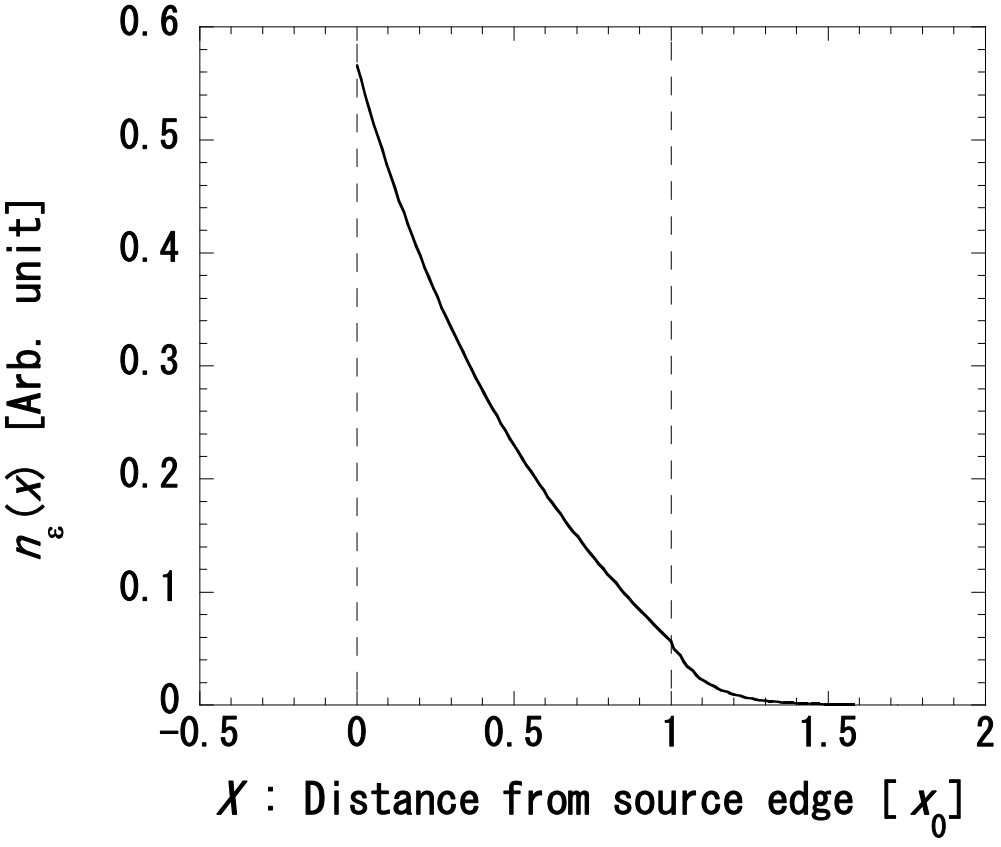
0.6
0.5
0.4
0.3
0.2
0.1
0
$n_\varepsilon(x)$ [Arb. unit]
-0.5
0
0.5
1
1.5
2
X : Distance from source edge [x_0]

Fig. 3

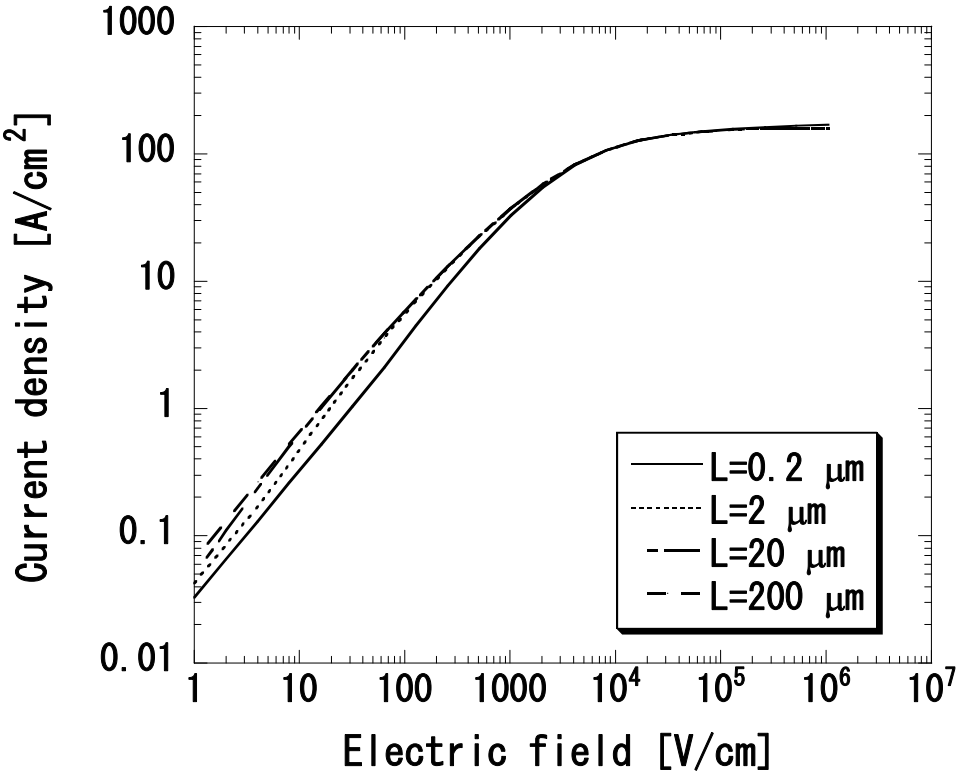
1000
100
10
1
0.1
0.01
Current density [A/cm^2]
1
10
100
1000
10^4
10^5
10^6
10^7
Electric field [V/cm]
L=0.2 μm
L=2 μm
L=20 μm
L=200 μm

Fig. 4

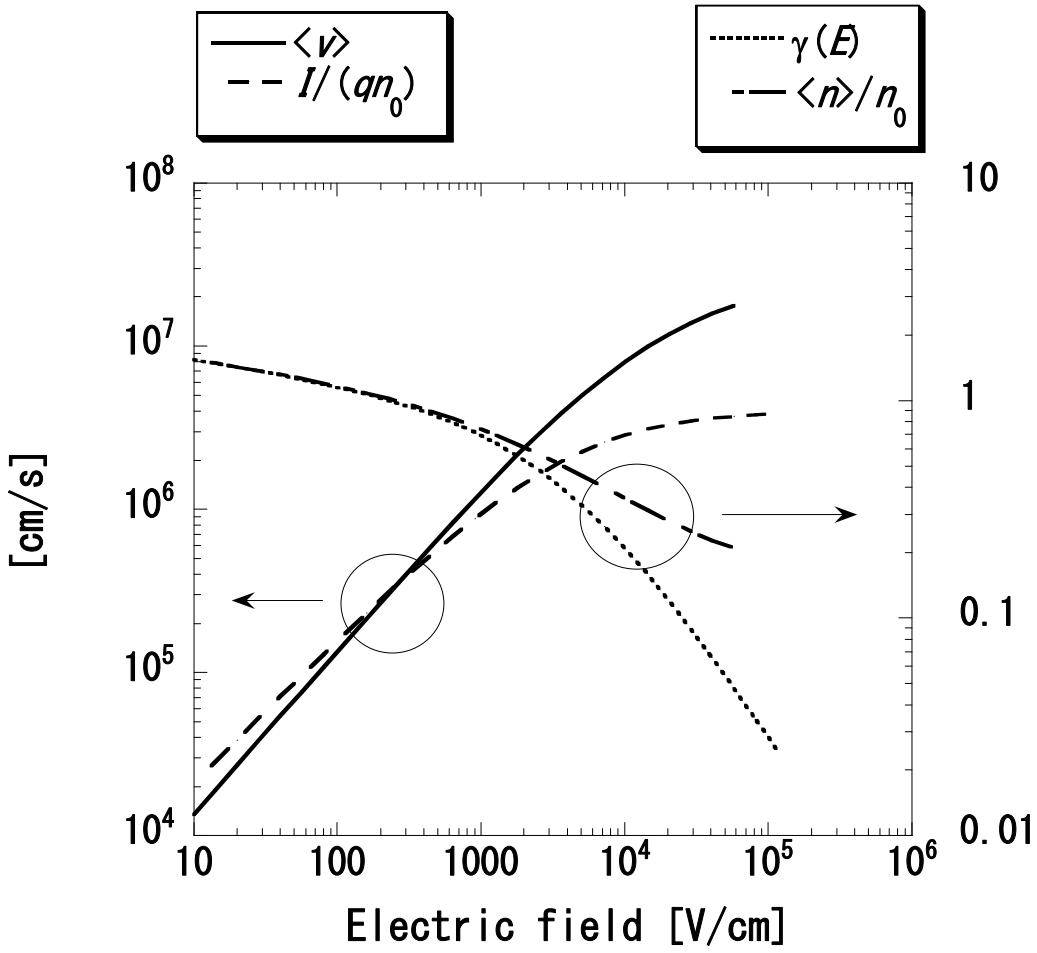
<v>
I/(qn0)
γ(E)
<n>/n0
[cm/s]
10^8
10^7
10^6
10^5
10^4
10
1
0.1
0.01
10
100
1000
10^4
10^5
10^6
Electric field [V/cm]

Fig. 5

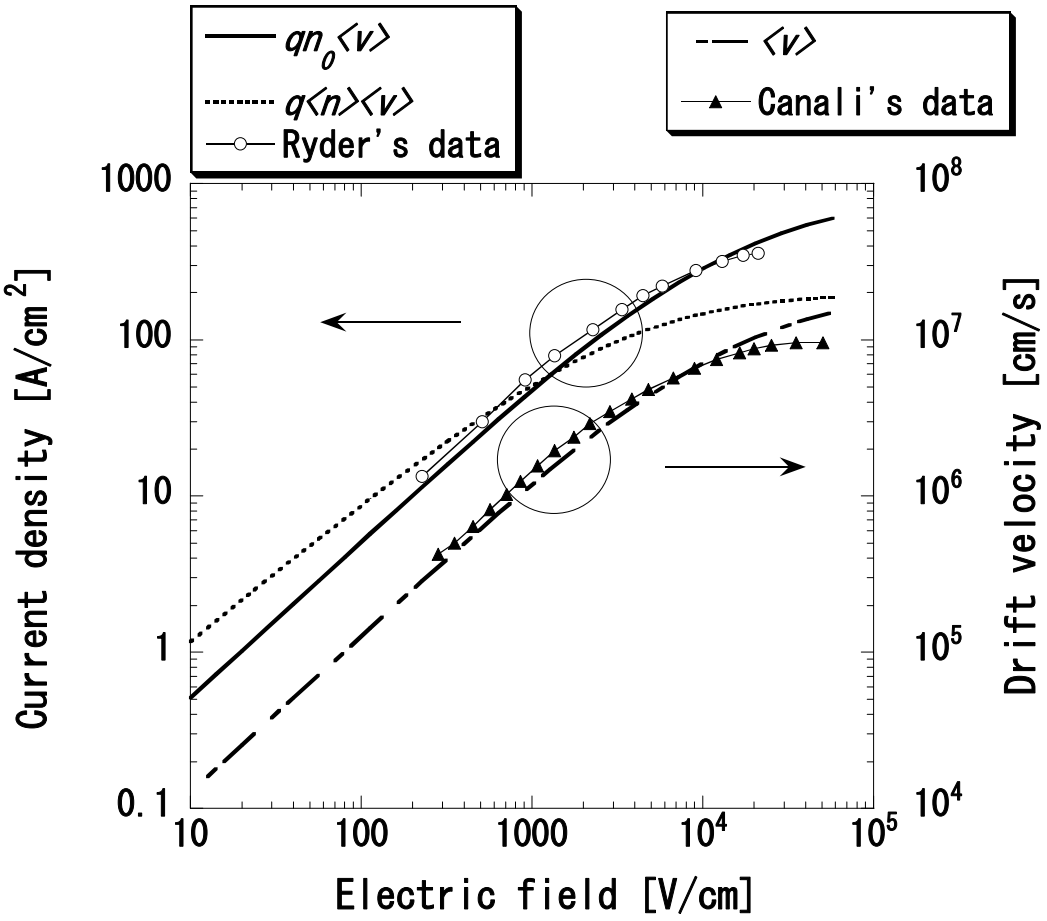
$qn_0\langle v\rangle$
$q\langle n\rangle\langle v\rangle$
Ryder's data
$\langle v\rangle$
Canali's data
Current density [A/cm^2]
Drift velocity [cm/s]
Electric field [V/cm]
1000
100
10
1
0.1
10^8
10^7
10^6
10^5
10^4
10
100
1000
10^4
10^5

Fig. 6