Ballistic metal-oxide-semiconductor field effect transistor

Kenji Natori
Institute of Applied Physics, University of Tsukuba, Tsukuba, Ibaraki 305, Japan

(Received 14 March 1994; accepted for publication 6 July 1994)

Experiments on ultra-small metal-oxide-semiconductor field effect transistors (MOSFETs) less than 100 nm have been widely reported recently. The frequency of carrier scattering events in these ultra-small devices is diminished, so that further suppression of carrier scattering may bring these devices close to the regime of ballistic transport. Carrier scattering is suppressed by constructing their channel regions with intrinsic Si and also by low temperature operation. This article proposes the ballistic transport of carriers in MOSFETs, and presents the current-voltage characteristics of the ballistic n-channel MOSFET. The current is expressed with the elementary parameters without depending on the carrier mobility. It is independent of the channel length and is proportional to the channel width. The current value saturates as the drain voltage is increased and the triode and the pentode operation are specified as in the conventional MOSFET. Similar current-voltage characteristics in the ballistic transport regime are also investigated for the p-channel MOSFET, the dual gate ultra-thin silicon on insulator MOSFET, and the high electron mobility transistor device. The obtained current gives the maximum current limitation of each field effect transistor geometry. The current control mechanism of ballistic MOSFETs is discussed. The current value is governed by the product of the carrier density near the source edge in the channel, and the velocity with which carriers are injected from the source into the channel. Influence of optical phonon emission to the transport is discussed. It is suggested that if the device is operated with relatively low carrier density at low temperatures, and if the scattering processes other than the optical phonon emission are suppressed so as to attain the ballistic transport, the optical phonon emission is also suppressed and ballistic transport is sustained. A convenient figure of merit to show the ballisticity of carrier transport in an experimental MOSFET is proposed. Its value is estimated for some examples of the recent ultra-small MOSFET experiment. The proposed current voltage characteristics are evaluated for a dual gate silicon on insulator MOSFET geometry. The result is compared with the recently reported elaborate Monte Carlo simulation with satisfactory agreement.

List of main symbols

$\hbar$	$h/2\pi$ where h is Planck's constant.
q	Unit charge.
k	Boltzmann constant.
T	Temperature.
m_x, m_y, m_z	Carrier effective mass along the coordinate axes.
m_t	Electron transverse effective mass.
m	Carrier effective mass. (Except where otherwise specified.)
$\epsilon_s, \epsilon_{ox}, \epsilon_{AlGaAs}$	Permittivities associated with Si, SiO_2, and AlGaAs, respectively.
I	Transistor current.
E_{max}	Maximum value of carrier potential energy in the channel.
x_{max}	Value of x where E_{max} is realized.
$U(x,z)$	Potential energy of a carrier.
$\Psi(x,y,z)$, $\varphi_{n_z}(x,z)$	Wave functions of the carrier in the channel.
E	Energy of a carrier.
$E_{n_z}(x)$	Energy level of n_zth mode in the z direction.
W	Transistor channel width.
$p(x)$	Momentum value in the WKB approximation.
k_m	Wave number of electronic wave around x_{max}.
n_x, n_y	Quantum numbers associated with the motion along the x and y axes, respectively.
$\vec{v}, \overleftarrow{v}$	Velocities of carriers propagating from the source towards the drain and from the drain towards the source, respectively.
$\overrightarrow{D(E)}, \overleftarrow{D(E)}$	One-dimensional density of state functions associated with carriers propagating from the source towards the drain and from the drain towards the source, respectively.
$f(\phi_F, E)$	Fermi distribution function with the Fermi energy ϕ_F.
$T(E)$	Transmission coefficient of the subchannel.
ϕ_{FS}, ϕ_{FD}	Fermi energies of the source and the drain, respectively.
V_D, V_G	The drain voltage and the gate voltage measured from the source level, respectively.
Q	Carrier charge density.
F_0	Electric field underneath the channel carrier layer.
F	Average electric field that carriers suf-

 0021-8979/94/76(8)/4879/12/$6.00

	fer from at the oxide-semiconductor interface.
t_i	Intrinsic Si layer thickness.
χ_{MS}	Difference between the metal work function and the electron affinity of Si.
ϕ_s	Conduction band minimum energy of Si at the MOS interface.
t_{ox}	Oxide layer thickness.
C_{eff}	Effective metal-oxide-semiconductor structure capacitance.
$\rho(z)$	Carrier charge distribution in the z direction.
Z_{n_z}	Mean separation from the oxide-semiconductor interface of the carrier in the n_z state.
V_t	Threshold voltage.
M_v	Effective lowest level degeneracy modified to include the effect of the population in upper levels.
I_{sat}	Saturation current.
G	Transconductance.
d	n^+-AlGaAs layer thickness.
v_{inj}	Injection velocity.
v_f	Largest velocity of the filled carrier states propagating towards drain at x_{max}.
$\Delta\phi(x)$	Channel potential variation in the x direction.
$v(x)$	Carrier mean velocity at x in the channel.
$n(x)$	Channel carrier density variation in the x direction.
C_{ox}	Oxide layer capacitance.
r	An index that gives the degree of ballisticity of the carrier transport in an experimental device.

I. INTRODUCTION

Recently, Si metal-oxide-semiconductor field effect transistors (MOSFETs), of around 100 nm geometry, have been widely investigated and the experimental data are being accumulated.[1–8] These ultra-small devices are studied in view of future application to ultra-large scale integrated circuits (ULSIs) in the beginning of the 21st century. These devices are expected to show ultra-high performance based on a new transport mechanism different from that of conventional devices (called macroscopic devices hereafter). In conventional MOSFETs larger than submicrometer range, analysis based on the method of traditional electronics is fully powerful and effective. When the electric field inside is not excessively large, the carrier mean velocity is proportional to the electric field with the proportionality constant called the carrier mobility. The value of the carrier mobility is the result of complex scattering processes due to various scatterers and it is impossible to give it in a simple expression or calculation. In smaller devices where the electric field inside intensifies, carrier velocity saturates and is limited to a maximum value of around 10^7 cm/s. It is believed that the carrier kinetic energy is truncated by optical phonon scattering when the device size is larger than the associated scattering length.

When the device size is less than 100 nm, velocity overshoot phenomena have been reported in the Si MOSFETs.[1,9,10] A carrier velocity, exceeding the saturation velocity, is observed when the device size is too small for carriers to attain the equilibrium velocity. Detailed numerical simulation also confirms considerable enhancement of carrier velocity near the drain edge.[11]

In future applications, ultra-small devices may need to be operated in low temperature environments, as will be discussed later. Electronic transport in ultra-small devices at low temperatures needs to be treated with quantum mechanical considerations. In conventional macroscopic devices where the device size is far larger than the inelastic scattering length, carriers undergo a large number of inelastic scattering processes which result in the diffusive carrier motion well described by the carrier mobility. When the device size is reduced to the mesoscopic range, less than the so-called phase coherence length L_ϕ, the interference effect of the electronic wave has to be taken into account, although the electronic motion is still diffusive due to multiple elastic scattering. If the device size is so small such that the number of scattering events that a carrier encounters in the course from source to drain becomes very small, and also that the number of carriers that perform the switching action becomes very small, statistical fluctuations of random scattering directly bring about the fluctuations in the device characteristics. Electrical characteristics of devices, although fabricated through the same processing steps, may fluctuate from device to device and also from time to time in the same device. A highly integrated system of today consists of a huge number of precisely controlled component devices. If the component device shows a large indefinite fluctuation, the circuit technology to constitute the system will confront serious difficulty in controlling the system function. If the device size is further decreased to less than the scattering length, ballistic carrier motion dominates the transport and the statistical fluctuations will be almost eliminated. A practical solution for future highly integrated systems may be the direct device-size reduction to this ballistic transport range, rather than to struggle with the random fluctuations in the mesoscopic range. The carrier scattering length can be controlled by varying the fabrication process and the operational conditions. The elimination of the fluctuations may be realized either by elimination of scattering events, or conversely increasing the frequency of scattering. The latter approach is less attractive due to the degradation of device performance.

How small do we need to reduce the Si MOSFET to meet the ballistic transport? It is a difficult question to answer. According to the data of Takagi *et al.*,[12] carrier mobility in an n-channel MOSFET (nMOSFET) on (100) Si surface amounts to around 3500 cm^2/V s at 77 K for an effective normal field of E_{eff}=0.3 MV/cm. We can estimate from this value that the scattering length is over 80 nm in this condition. Higher normal field brings about the mobility degradation due to surface roughness scattering. This implies that the ballistic transport, which has been thought of as an unreal dream in Si devices, becomes a reality when the size is

reduced to less than this value with the reduced normal field. Even at room temperature, an elaborate Monte Carlo simulation of the silicon on insulator (SOI) MOSFET performed by the IBM group[13] suggests that the carrier transport will become near ballistic if the size is reduced to around 30 nm.

Ultra-small MOSFETs are usually analyzed by the macroscopic device theory based on diffusive transport. Carrier mobility and velocity saturation have been the central concept there, partly modified by the introduction of velocity overshoot. However, down-sizing has come to such a level that the device characteristics in some cases are better understood when based on the ballistic point of view. The high performance limitation of ultra-small devices is better understood by analysis of ballistic transport. A comparison of actual device performance with that of ideal ballistic device will give insight into the effect of carrier scattering during device operation.

For the future development of high density ULSIs, Si devices will still be dominant due to the advantage of technological know-how. Si MOSFETs will still maintain the position of the most widely used switching device. In the application of switching devices to highly integrated systems, a low leakage current level will become the most important point to note. A small amount of leakage current at an individual device level will be easily summed up to an amount that exceeds the permitted power consumption of the total system. Switching of the device is usually performed by the application of a bias voltage, but the bias amplitude has to be reduced as the device size is scaled down so as to prevent device breakdown. In short, a sharp switching characteristic, steep gradient of current-voltage characteristic is increasingly important in ultra-small devices for high integration. In the case of MOSFETs, this gradient is characterized by the so-called S-factor. It is restricted by the Boltzmann (Fermi) distribution of carrier energy and the magnitude is controlled by the operating temperature. Practically, the low temperature operation is the only solution to improve its value. A method of device size scaling simultaneously reducing the operating temperature is also proposed.[14] Tendency towards low temperature operation seems inevitable in future high density ULSIs that consist of ultra-small switching devices.

MOSFET scaling theory usually demands that the depletion layer width in the Si should be reduced proportionally as the device size is scaled down. This implies the increase of impurity concentration of the device substrate, which brings about the carrier mobility degradation as well as the *pn* junction leakage current increase. On the contrary, for improvement of device characteristics, it is desirable to decrease the impurity concentration around the channel region. If it is reduced to the intrinsic level (i.e., containing few impurities), carrier scattering due to charged impurities will be greatly reduced. Low temperature operation greatly suppresses the phonon scattering probabilities. The surface roughness scattering and the electron-electron scattering may still persist, but the former of these two is not so serious if the electric field normal to the surface can be reduced.[12] The influence of electron-electron scattering to the current will also be insignificant if the energy of a carrier is approximated by a quadratic form of its momentum. This is because the conservation of total momentum in electron-electron scattering implies the conservation of total current in parabolic bands when the interband scattering is neglected. Thus, the applicability of ballistic transport is greatly enhanced in these intrinsic channel MOSFETs operated at low temperatures. The effect of the optical phonon scattering will be discussed separately.

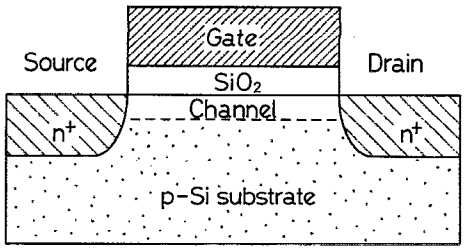


FIG. 1. The cross section of a bulk-type *n*MOSFET with an intrinsic channel region.

Some examples of these intrinsic channel MOSFETs are shown in Figs. 1 and 2. Figure 1 is the structure of the bulk type MOSFET where only the channel region is made intrinsic. Epitaxial growth technology may be necessary. Figure 2 is an ultra-thin film SOI MOSFET structure where carriers can be confined in the channel by the band discontinuity at the Si-SiO_2 interface. An intrinsic channel MOSFET structure can be constructed without depending on a depletion layer to separate the channel region from the substrate. Figure 2(a) is the normal single gate MOSFET structure, while Fig. 2(b) is the dual gate MOSFET structure with the improved gate electrode controllability.

We have already discussed the current-voltage (I-V) characteristics of SOI MOSFET in the ballistic mode.[15] In this article, we will expand on those results and propose a general expression for the I-V characteristics of a ballistic

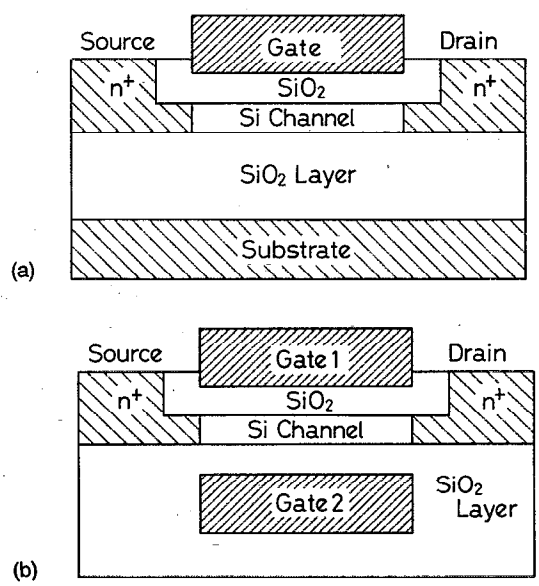


FIG. 2. The cross section of the ultrathin SOI *n*MOSFET where the channel region is made of intrinsic silicon: (a) the single gate MOSFET, (b) the dual gate MOSFET.

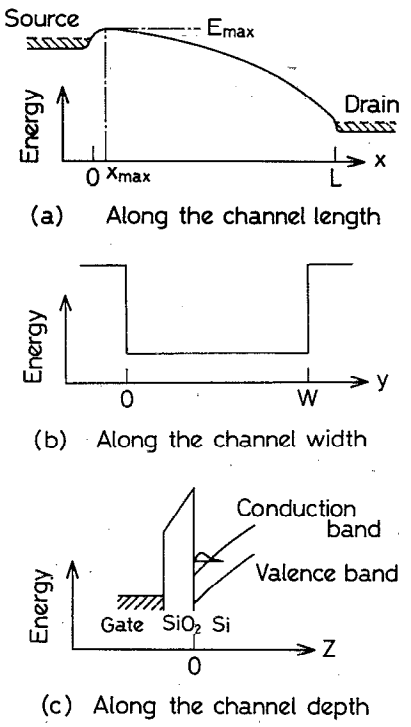


FIG. 3. Schematic distribution of electronic potential energy in a MOSFET structure: (a) the potential distribution along the channel length; (b) the potential distribution along the MOSFET width; (c) the electronic energy distribution along the channel thickness direction.

field effect transistor (FET) and also discuss the current control mechanism of these ballistic devices. In Section II, *I-V* characteristics of an *n*MOSFET are derived and Sec. III will give a generalization of the result. Section IV is devoted to the discussion on the current control mechanism and Sec. V deals with the analysis of experimental results. Section VI is a brief summary.

II. ANALYSIS OF Si *n*MOSFET

In ballistic transport, carriers are transferred from source to drain without being scattered. Remaining interactions other than scattering can be averaged and transformed to a mean field potential. Discussion of electronic states in the channel is greatly simplified and a quantum mechanical solution can be given if the mean field potential is identified. The potential energy of a carrier will be simply referred to as "potential" hereafter. The potential curve, which is designated by the spatial variation of conduction band energy minimum in Si, amounts to something schematically illustrated in Fig. 3. Along the x axis, it varies slowly from source potential to drain potential usually with a maximum point near source [$E_{\max}$ at $x_{\max}$ in Fig. 3(a)] and with a steeper gradient toward the drain edge. Numerical simulation of the MOSFET usually gives this type of potential variation as a self-consistent solution of charge distribution and Poisson's equation. Ballistic transport may bring about a serious deviation from that obtained by the classical calculation. The probable potential variation along the channel will be discussed later. Here we assume only basic features that are evident; the channel potential varies slowly from high source level to the low drain level with a maximum in between ($E_{\max}$ at $x_{\max}$). Along the channel width direction denoted by the y axis, carriers are confined within the channel by steep barriers at the edges. The potential profile is well approximated by a square potential well if the channel width W is large compared with the range over which the potential varies. Along the channel depth direction denoted by the z axis, carriers are confined at the MOS interface by a sharp triangle-like potential well and the electronic states are expressed by discrete energy levels. Thus, the electronic states in the channel are provided by the solution of the following effective mass equation,

$$\left(-\frac{\hbar^2}{2m_x}\frac{\partial^2}{\partial x^2}-\frac{\hbar^2}{2m_y}\frac{\partial^2}{\partial y^2}-\frac{\hbar^2}{2m_z}\frac{\partial^2}{\partial z^2}+U(x,z)\right)\psi(x,y,z)$$
$$=E\Psi(x,y,z) \quad 0\leqslant y\leqslant W \tag{1}$$

with the boundary condition,

$$\Psi(x,0,z)=\Psi(x,W,z)=0,$$

if the barrier heights at channel edges in the y direction are sufficiently large. $U(x,z)$ is the potential curve illustrated in Figs. 3(a) and 3(c). Parabolic bands are assumed and the crystal principal axes are parallel to the coordinate axes. An *n*MOSFET on the (100) surface is examined. Since the potential variation along the x direction is slowly varying compared with that along the z direction, the adiabatic approximation is applied. With use of the Wentzel–Kramers–Brillouin (WKB) approximation for the solution along the x direction, electronic states with the energy above $E_{\max}$ are expressed as

$$\Psi(x,y,z)=\frac{A}{\sqrt{p(x)}}\exp\left(\frac{i}{\hbar}\int^x p(x)dx\right)$$
$$\cdot\sqrt{\frac{2}{W}}\sin\left(\frac{n_y\pi}{W}y\right)\cdot\varphi_{n_z}(x,z), \tag{2}$$

where A is a constant and $\varphi_{n_z}(x,z)$ is the solution of

$$\left(-\frac{\hbar^2}{2m_z}\frac{\partial^2}{\partial z^2}+U(x,z)\right)\varphi_{n_z}(x,z)=E_{n_z}(x)\varphi_{n_z}(x,z), \tag{3}$$

and

$$p(x)=\sqrt{2m_x\left[E-\frac{\hbar^2}{2m_y}\left(\frac{n_y\pi}{W}\right)^2-E_{n_z}(x)\right]}. \tag{4}$$

A solution propagating towards the drain is shown in Eq. (2). Equation (3) defines the energy level $E_{n_z}(x)$ associated with the n_zth mode in the z direction as a function of x. Electronic states in the quantum well along the y direction show discrete levels specified with the quantum number n_y. Equation (2) implies that electronic wave modes in the channel consist of a number of one-dimensional wave subchannels or subbands, each specified with a pair of quantum numbers n_y and n_z. Wave reflection at the drain edge is neglected. Around the energy maximum point at $x_{\max}$, Eq. (2) is further approximated by a plane wave expression

$$\Psi(x,y,z)=\frac{A'}{\sqrt{\hbar k_m}}\exp(ik_m x)\cdot\sqrt{\frac{2}{W}}\sin\left(\frac{n_y\pi}{W}y\right)\cdot\varphi_{n_z}(x,z) \tag{5}$$

with

$$k_m=\frac{1}{\hbar}\sqrt{2m_x\left[E-\frac{\hbar^2}{2m_y}\left(\frac{n_y\pi}{W}\right)^2-E_{n_z}(x_{\max})\right]}, \tag{6}$$

and A' is another constant.

The solution of Eq. (1) satisfies the current continuity condition and the MOSFET current evaluated at any point in the channel gives the same result. The current I can be evaluated with use of the method proposed in Landauer's formula,[16] because it is expressed as a sum of many one-dimensional subchannel components. Each subchannel current component flowing in one direction is given by the product of the unit charge, the number of carriers flowing into the subchannel per unit time, the transmission coefficient of the subchannel, and the probability that the destination is empty, all integrated over the carrier energy. The number of carriers flowing into the subchannel is further expressed by the product of the input carrier group velocity, the density of states, and the probability that the state is occupied by the carrier. The probability of carrier occupancy is given by the Fermi distribution function with the source Fermi level on the source side, and that with the drain Fermi level on the drain side of the subchannel. Both current directions, the one from the source to the drain and that in the opposite direction, should be considered. Thus, it is expressed as

$$I=q\sum_{\text{valley}}\sum_{n_y}\sum_{n_z}\int\{\vec{v}\overrightarrow{D(E)}f(\phi_{FS},E)[1-f(\phi_{FD},E)]-\vec{v}\overrightarrow{D(E)}f(\phi_{FD},E)[1-f(\phi_{FS},E)]\}T(E)dE, \tag{7}$$

where $\vec{v}$ is the carrier group velocity of a one-dimensional electronic wave propagating towards the drain through a certain subchannel, and $\overrightarrow{D(E)}$ is the density of states for that electronic wave, both evaluated near the source side edge of the subchannel. These quantities are expressed as $\vec{v}=dE/dp(x)=p(x)/m_x$ and $\overrightarrow{D(E)}=\{\pi\hbar[dE/dp(x)]\}^{-1}=m_x/\pi\hbar p(x)$ in the WKB approximation with use of Eq. (4), and the product gives a constant $(\pi\hbar)^{-1}$. $\vec{v}$ and $\overrightarrow{D(E)}$ are the group velocity and the density of states for carriers propagating from the drain to the source, both evaluated near the drain side edge of the subchannel. The product reduces to the same constant value similarly. $f(\phi_F,E)$ is the Fermi distribution function with the Fermi energy ϕ_F

$$f(\phi_F,E)=\left[1+\exp\left(\frac{E-\phi_F}{kT}\right)\right]^{-1}.$$

k is the Boltzmann constant and T is the temperature. The source and drain regions are assumed to be ideal reservoirs with the Fermi energies ϕ_{FS} and ϕ_{FD}, respectively. They feed carriers in thermal equilibrium to the channel and also absorb carriers from the channel without reflection. Deviation from this ideal model will be discussed elsewhere. $T(E)$ is the transmission coefficient of the subchannel at energy E.

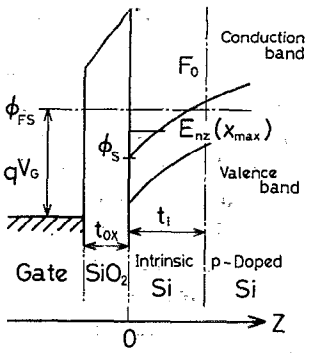


FIG. 4. The electronic energy band structure at x_{max} along the channel depth direction.

Ballistic transport implies that $T(E)=1$ at energies allowed for propagation along the channel, i.e., for $E > E_{n_z}(x_{\max})$. Contribution from the evanescent mode waves is neglected. When the channel width W is not excessively small, summation over n_y in Eq. (7) is replaced by integration over energy with the one-dimensional density of state function being multiplied. Since $\phi_{FD}=\phi_{FS}-qV_D$, where V_D is the applied drain voltage, Eq. (7) is transformed to

$$I=W\frac{\sqrt{2}q(kT)^{3/2}}{\pi^2\hbar^2}\sum_{\text{valley}}\sum_{n_z}\sqrt{m_y}\left[F_{1/2}\left(\frac{\phi_{FS}-E_{n_z}(x_{\max})}{kT}\right)-F_{1/2}\left(\frac{\phi_{FS}-E_{n_z}(x_{\max})-qV_D}{kT}\right)\right], \tag{8}$$

where $F_{1/2}(u)$ is the Fermi–Dirac integral[17] and

$$F_{1/2}(u)=\int_0^\infty\frac{\sqrt{y}}{1+\exp(y-u)}dy. \tag{9}$$

The charge density around $x_{\max}$ is evaluated in a similar way. Electronic states of subbands propagating from source to drain are filled up to the energy of the source Fermi level, while those states propagating from drain to source are filled up to the drain Fermi level energy. Utilizing the two-dimensional density of states function as well as the Fermi distribution function, we obtain the carrier charge density

$$|Q|=\frac{qkT}{2\pi\hbar^2}\sum_{\text{valley}}\sum_{n_z}\sqrt{m_x m_y}\ln\times\left\{\left[1+\exp\left(\frac{\phi_{FS}-E_{n_z}(x_{\max})}{kT}\right)\right]\times\left[1+\exp\left(\frac{\phi_{FS}-E_{n_z}(x_{\max})-qV_D}{kT}\right)\right]\right\}. \tag{10}$$

The charge density of the MOS junction is usually expressed with the MOS capacitance and the threshold voltage, and a similar expression is explored. Since the lateral component of the field vanishes around $x_{\max}$, the electric field there is normal to the interface. The electronic potential distribution there is schematically illustrated for a bulk MOSFET in Fig. 4. The width of the oxide layer is denoted

with t_{ox}, and ϕ_s shows the conduction band minimum energy at the MOS interface. Inversion carriers are localized only near the MOS interface and the electric field underneath this carrier layer, which is still almost normal to the interface, is denoted by F_0. Carriers are pressed to the interface by F_0 as well as by the field generated by themselves. We assume that the inversion charge layer is wholly included within the intrinsic Si region of thickness t_i, and F_0 is defined, for example, as the normal electric field at the interface between the intrinsic Si and the doped Si. Taking account of the fact that the electric field in the oxide layer is $(|Q|+\epsilon_s F_0)/\epsilon_{ox}$, where ϵ_s and ϵ_{ox} are permittivities associated with Si and SiO_2, we come to the expression of the gate voltage V_G measured from the source Fermi level ϕ_{FS},

$$qV_G=\chi_{MS}+q\frac{t_{ox}}{\epsilon_{ox}}(\epsilon_s F_0+|Q|)+[\phi_{FS}-E_{n_z}(x_{max})]+[E_{n_z}(x_{max})-\phi_s]. \tag{11}$$

χ_{MS} is the difference between the metal work function and the Si electron affinity measured from the conduction band minimum. With the use of Eq. (10), we come to an expression

$$\frac{d|Q|}{dV_G}=\left[q\sum_{\text{valley}}\sum_{n_z}\alpha_{n_z}\left(1-\frac{t_{ox}}{\epsilon_{ox}}\epsilon_s\frac{dF_0}{dV_G}\right)\right]\Bigg/\left[1+\sum_{\text{valley}}\sum_{n_z}\alpha_{n_z}\left(\frac{qt_{ox}}{\epsilon_{ox}}+\frac{d[E_{n_z}(x_{max})-\phi_s]}{d|Q|}\right)\right], \tag{12}$$

where

$$\alpha_{n_z}=\frac{q}{2\pi\hbar^2}\sqrt{m_x m_y}\{f[\phi_{FS},E_{n_z}(x_{max})]+f[\phi_{FS}-qV_D,E_{n_z}(x_{max})]\}. \tag{13}$$

F_0 is a function of ϕ_s as well as other terminal voltages like those at the source, the drain, and the substrate. Small variation of the gate voltage δV_G brings about the potential variation at $z=t_i$, which is less than $|q\delta V_G|$, and induces the variation of F_0, which is less than $|\delta V_G|$/(depletion layer width). Therefore, on the right-hand side of Eq. (12),

$$\left|\frac{t_{ox}}{\epsilon_{ox}}\epsilon_s\frac{dF_0}{dV_G}\right|\ll\frac{\epsilon_s}{\epsilon_{ox}}\frac{t_{ox}}{\text{depletion layer width}}\ll 1$$

holds good in MOSFET geometries where $t_{ox}\ll$depletion layer width. When V_G is small enough, $[\phi_{FS}-E_{n_z}(x_{max})]$ is small and $|Q|$ goes extremely small. In this case $|q\delta V_G|\sim|\delta\phi_s|\sim|\delta E_{n_z}(x_{max})|$ and $|Q|$ varies exponentially with V_G. This corresponds to the weak inversion region of the MOS junction. When V_G is large so that $[\phi_{FS}-E_{n_z}(x_{max})]$ takes appreciable values, $|Q|$ becomes large and the onset of strong inversion occurs. (A common expression "inversion" is used although the channel region is intrinsic Si.) Usually, α_{n_z} becomes so large that the denominator on the right-hand side of Eq. (12) is far larger than unity, and the MOS junction capacitance $C_{eff}=(d|Q|/dV_G)$ is expressed as

$$C_{eff}=\frac{d|Q|}{dV_G}=\left[\frac{t_{ox}}{\epsilon_{ox}}+\frac{1}{q}\left(\sum_{\text{valley}}\sum_{n_z}\alpha_{n_z}\frac{d[E_{n_z}(x_{max})-\phi_s]}{d|Q|}\right)\Bigg/\left(\sum_{\text{valley}}\sum_{n_z}\alpha_{n_z}\right)\right]^{-1}. \tag{14}$$

When the (t_{ox}/ϵ_{ox}) term dominates in the right-hand side of the above expression, C_{eff} is approximated by the oxide capacitance and the deviation is small. For example, we assume the carrier is confined within a triangular potential at the MOS interface. The electric field F that defines the triangular well is evaluated by averaging the field strength that the carrier suffers from. With use of the inversion charge density $\rho(z)$

$$F=\int_0^{t_i}\left(F_0-\frac{1}{\epsilon_s}\int_z^{t_i}\rho(z)dz\right)\rho(z)dz\Bigg/\int_0^{t_i}\rho(z)dz=F_0+\frac{|Q|}{2\epsilon_s}. \tag{15}$$

The electronic states in the triangular potential are solved[18] with the Airy function and the energy level is approximated as

$$E_{n_z}(x_{max})-\phi_s=\left(\frac{\hbar^2}{2m_z}\right)^{1/3}\left[\frac{3}{2}\pi q\left(n_z+\frac{3}{4}\right)\right]^{2/3}F^{2/3}=\tfrac{3}{2}qFZ_{n_z}\quad n_z=0,1,2,... \tag{16}$$

where

$$Z_{n_z}=\int_0^{t_i}z|\varphi_{n_z}(x_{max},z)|^2\,dz\Bigg/\int_0^{t_i}|\varphi_{n_z}(x_{max},z)|^2\,dz. \tag{17}$$

Equations (14)–(17) lead to

$$C_{eff}=\left(\frac{t_{ox}}{\epsilon_{ox}}+\frac{\langle Z_{n_z}\rangle}{2\epsilon_s}\right)^{-1} \tag{18}$$

with

$$\langle Z_{n_z}\rangle=\frac{\Sigma_{\text{valley}}\Sigma_{n_z}\alpha_{n_z}Z_{n_z}}{\Sigma_{\text{valley}}\Sigma_{n_z}\alpha_{n_z}}. \tag{19}$$

In an ultra-thin single gate SOI MOSFET, where carriers are confined between the two Si-SiO_2 interfaces by the energy band discontinuities, a slight modification to the discussion is necessary. In this case F_0 is conveniently defined as the normal electric field in the Si layer at $z=t_i$, where t_i is the width of the intrinsic Si layer. A similar discussion leads to $|(t_{ox}/\epsilon_{ox})\epsilon_s(dF_0/dV_G)|\ll 1$ and therefore to Eq. (14). The actual expression for C_{eff} depends on the detailed electronic structure in the Si layer. If we assume the square-well wave function $\varphi_{n_z}(x_{max},z)=\sqrt{(2/t_i)}\sin(n_z\pi z/t_i)$, and if only the lowest level is populated, C_{eff} is evaluated as

$$C_{\text{eff}}=\left[\frac{t_{\text{ox}}}{\epsilon_{\text{ox}}}+\frac{t_i}{\epsilon_s}\left(\frac{1}{3}+\frac{5}{8\pi^2}\right)\right]^{-1}. \tag{20}$$

Generally, in strong inversion, variation of C_{eff} with the variation of V_G is small because the second term on the right-hand side of Eq. (14) depends only weakly on V_G, and C_{eff} is approximately constant irrespective of the value of V_G. Thus $|Q|$ is expressed with an appropriate constant value V_t as

$$|Q|=C_{\text{eff}}(V_G-V_t). \tag{21}$$

The onset of strong inversion from the weak inversion region occurs rather sharply as V_G is increased.

In the electric quantum limit,[18] where only one n_z level is occupied, summations in Eqs. (8) and (10) reduce to a numerical factor of 2 expressing the contribution of each lowest level of the two valleys with the larger effective mass normal to the interface. Also, in realistic cases where the contribution of higher levels is not neglected, it is possible to approximate these summations with a numerical factor M_v if the fraction of higher levels is not large. M_v is so evaluated that $\Sigma_{\text{valley}}\Sigma_{n_z}[\ ]=M_v\times$(lowest level contribution, per valley if degenerate), and therefore $M_v\approx$(lowest valley degeneracy)$\times$(total population/population of lowest levels). In a heavily inverted n-channel (100) MOS junction at 77 K, for example, the fraction of population in the lowest levels is around 0.8[18] and we can put $M_v\approx 2.5$. Equations (8), (10), and (21) give a compact expression of I in terms of the terminal voltages

$$\begin{aligned}
&I=WI_0[F_{1/2}(u)-F_{1/2}(u-v_d)],\\
&u=\ln[\sqrt{(1+e^{v_d})^2+4e^{v_d}(e^{\rho}-1)}-(1+e^{v_d})]-\ln 2,\\
&v_d=\frac{qV_D}{kT},\quad I_0=\frac{\sqrt{2}q(kT)^{3/2}}{\pi^2\hbar^2}M_v\sqrt{m_t},\\
&\rho=\frac{2\pi\hbar^2C_{\text{eff}}(V_G-V_t)}{qkTm_tM_v},
\end{aligned} \tag{22}$$

where m_t is the transverse effective mass of an electron. This is the I-V characteristics of the nMOSFET on the (100) surface. The result is rigorous with $M_v=2$ in the electric quantum limit.

III. I-V CHARACTERISTICS OF BALLISTIC FETS

According to Eq. (22), the current value of a ballistic MOSFET is compactly expressed with elementary parameters like $m_t,\hbar,q$, etc. without depending on carrier mobility. The value is proportional to W and independent of channel length. The expression correctly reduces to $I=0$ when $V_D=0$ or $V_G=V_t$ (i.e., there are no carriers). For sufficiently large V_D, I becomes independent of V_D and it shows current saturation. The saturation current I_{sat} is

$$I_{\text{sat}}=WI_0F_{1/2}[\ln(e^{\rho}-1)]. \tag{23}$$

The Fermi–Dirac integral, Eq. (9), is usually simplified when carriers are degenerate in the conduction band, and here this corresponds to the situation where the electronic states propagating in one direction are occupied up to some Fermi level and the Fermi level measured from the subband energy minimum is far larger than kT. Around $x_{\max}$, for example, carrier states propagating from the source towards the drain are occupied from the lowest energy state to the ϕ_{FS}, and those states propagating from the drain towards the source are occupied up to $\phi_{FS}-qV_D$. If the energy separations from the lowest energy state to those Fermi energies are large compared with kT, the condition is satisfied. We call these situations as carriers towards the drain or towards the source are degenerate, respectively, hereafter. Now, if carriers towards drain are degenerate $F_{1/2}(u)$ is approximated by $(2/3)u^{3/2}$ and I_{sat} is simplified if $\rho\gg 1$,

$$I_{\text{sat}}=W\frac{8\hbar|Q|^{3/2}}{3m_t\sqrt{q\pi M_v}}=W\frac{8\hbar[C_{\text{eff}}(V_G-V_t)]^{3/2}}{3m_t\sqrt{q\pi M_v}}. \tag{24}$$

The transconductance is also simplified as

$$G_m=W\frac{4\hbar C_{\text{eff}}\sqrt{C_{\text{eff}}(V_G-V_t)}}{m_t\sqrt{q\pi M_v}}. \tag{25}$$

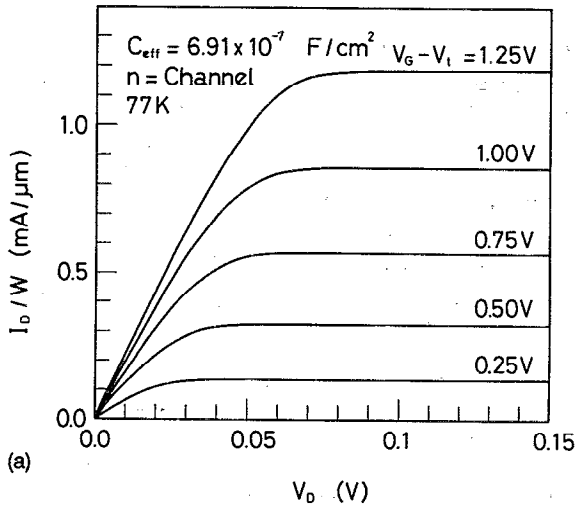
These expressions can be used as an approximation. However, note that it is difficult to attain $\rho\gg 1$ at room temperature.

An example of the I-V characteristics described by Eq. (22) is plotted in Fig. 5, where the value of C_{eff} is chosen to be that of 5 nm oxide. Figure 5(a) shows the I-V_D characteristics while Fig. 5(b) shows the I-V_G characteristics. Note that the transistor parameters included in Eq. (22), other than the applied voltages, are only W and C_{eff}. The current is proportional to W, and the carrier density is proportional to C_{eff}. Equation (22) gives I as a function of $|Q|$, and even if C_{eff} and V_G are varied the same value of $|Q|$ gives the same current.

The mechanism of current saturation is easily understood. As we see in Eq. (8), the current consists of two components with the opposite sign; one is due to carriers propagating towards the drain with the energy distribution characterized by ϕ_{FS}, and the other is due to carriers propagating toward the source characterized by the drain Fermi level ($\phi_{FS}-qV_D$). These two components exactly cancel each other when $V_D=0$, but the former component gradually dominates as V_D becomes large. When V_D becomes sufficiently large, the drain Fermi level goes far below the lowest energy level of the propagating state at $x_{\max}$. The latter term becomes negligibly small and the former term, which is characterized only by the source electrode, dominates giving rise to the saturation current. I_{sat} in Eq. (24) does not depend on T. This is because the carriers towards the drain are degenerate in this case. The current value depends on the energy separation between ϕ_{FS} and $E_{\max}$ and the details of energy distribution around ϕ_{FS} are neglected.

The I-V characteristics in Eq. (22) are generalized to include other ballistic FET structures. The p-channel MOSFET (pMOSFET) characteristics are also expressed by Eq. (22) if we modify I_0 and ρ to reflect the valence band structure of Si. That is,

$$I_0=\frac{\sqrt{2}q(kT)^{3/2}}{\pi^2\hbar^2}M_v\sqrt{m}, \tag{26}$$



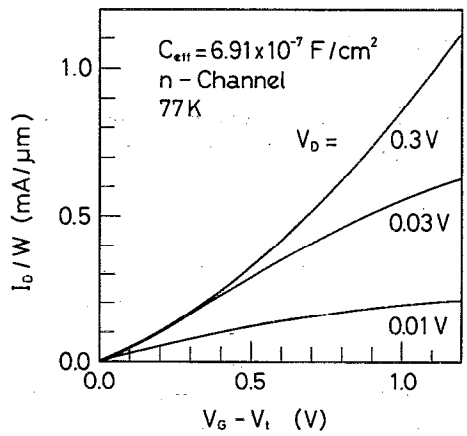


FIG. 5. Calculated examples of I-V characteristics of an nMOSFET on (100) plane. C_{eff} corresponds to effective 5 nm oxide thickness: (a) channel current per unit width vs drain voltage, parameter is the gate voltage; (b) channel current per unit width versus gate voltage, parameter is the drain voltage.

$$\rho=\frac{2\pi\hbar^2 C_{\text{eff}}(|V_G-V_t|)}{qkTmM_v}, \tag{27}$$

where m is the heavy hole effective mass. The absolute value of V_D should be used in evaluation of v_d. The valley degeneracy for evaluation of M_v is 1. An example of the I-V_D characteristics of a p-channel MOSFET is shown in Fig. 6.

The dual gate SOI MOSFET structure illustrated in Fig. 2(b) is also a promising candidate for the sub-100 nm transport device. The single gate structure in Fig. 2(a) may possibly suffer from the difficulty in threshold voltage control, because the short channel effect intensifies in these ultrasmall devices. The dual gate structure will offer ideal current controllability through the gate electrodes. We assume that gates 1 and 2 are symmetrically set to the channel and that an equal gate voltage V_G is applied. The I-V characteristics of the dual gate SOI nMOSFET on the (100) plane depend on the magnitude of t_i. When t_i is large enough, carriers are localized at the both Si-SiO_2 interfaces and the current is approximated by Eq. (22) with the modification of I_0 to

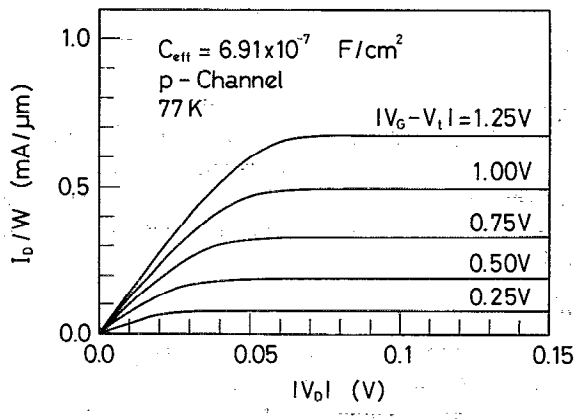


FIG. 6. I-V characteristics of a p-channel MOSFET. Channel current per unit width vs drain voltage with the gate voltage as the parameter. C_{eff} corresponds to effective 5 nm oxide thickness. M_v=1 is assumed.

$$I_0=\frac{2\sqrt{2}q(kT)^{3/2}}{\pi^2\hbar^2}M_v\sqrt{m_t}. \tag{28}$$

The expression of ρ is the same as in the single gate nMOSFET. When t_i is so small that carriers extend inside the SOI film, the discussion is complicated because M_v and C_{eff} depend on the details of the electronic structure.

Usually, I-V characteristics of the GaAs high electron mobility transistor (HEMT) device, illustrated in Fig. 7, are known to be expressed with a similar formula as in the MOSFET. The drain current of a ballistic HEMT device is also approximated by Eq. (22) with the modifications to

$$I_0=\frac{\sqrt{2}q(kT)^{3/2}}{\pi^2\hbar^2}M_v\sqrt{m}, \tag{29}$$

$$\rho=\frac{2\pi\hbar^2\epsilon_{\text{AlGaAs}}(V_G-V_t)}{qkTmM_v d}, \tag{30}$$

where the valley degeneracy for evaluation of M_v is chosen to be 1. m is the conduction band effective mass of GaAs ($0.067m_0$), and ϵ_{AlGaAs} and d are the permittivity and the thickness of the n^+-AlGaAs layer, respectively. This result applies to the case when the drain voltage is not so large and the carrier scattering to higher valleys can be neglected. When the intervalley scattering is dominant, the situation is more complex.

Equation (22) is derived by ignoring all scattering processes that prevent free carrier motion in the channel. There-

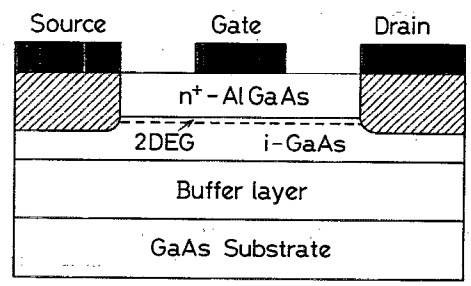


FIG. 7. Cross section of a HEMT structure.

TABLE I. Relative magnitude of I_{sat} for various FETs with $|Q|=5\times10^{12}$ q/cm^2 and $T=77$ K. Columns (1) and (2), respectively, show the results of Eqs. (23) and (24). The ratio of M_v is substituted for the ratio of the number of valleys including the lowest level. In the dual gate SOI nMOSFET, carriers are assumed to be localized near the MOS interfaces.

Kinds of FET	(1)	(2)
nMOSFET	1.00	1.00
pMOSFET	0.53	0.52
Dual gate SOI MOSFET	2.00	2.00
HEMT	3.93	4.01

fore, the current value gives the maximum available current in a given FET geometry. Actual devices with scattering will give less current in the same environment and Eq. (22) presents the maximum performance limitation of a FET structure. For comparison of various devices, the relative magnitude of I_{sat} with the same dimension is tabulated in Table I. Other devices, such as junction field effect transistors and static induction transistors in the ballistic transport regime are investigated in a similar way, but the Si nMOSFETs on (110) and (111) planes are slightly more complex because the effective mass Hamiltonian is not separable from the coordinate axes fixed in the device geometry.

IV. CURRENT CONTROL MECHANISM

In conventional MOSFETs, the drain current is governed by the carrier velocity in the channel. This carrier velocity is expressed as a product of the lateral electric field and the carrier mobility when the electric field is not as large. When the electric field intensifies, however, the carrier velocity saturates to a value around 10^7 cm/s presumably due to the optical phonon scattering. In the ballistic MOSFETs discussed here, carriers in the channel are free from scattering and all carriers propagating towards the drain, passing through the bottleneck around x_{max}, reach the drain without scattering back to the source, if the reflection at the drain edge is neglected. The channel current is governed by the current at the bottleneck. In current saturation, for example, the current injected from the source to the bottleneck dominates the total current, and the carrier velocity in the backward channel has little to do with the current value. If we define the mean carrier velocity at the bottleneck v_{inj} as

$$v_{inj}=\frac{I_{sat}}{WC_{eff}(V_G-V_t)}, \tag{31}$$

v_{inj} gives the mean carrier velocity injected from the source to the channel in ballistic MOSFETs, and we will call it the injection velocity. With use of Eq. (24),

$$v_{inj}=\frac{8\hbar\sqrt{|Q|}}{3m_t\sqrt{q\pi M_v}}=\frac{8\hbar\sqrt{C_{eff}(V_G-V_t)}}{3m_t\sqrt{q\pi M_v}} \tag{32}$$

when the carriers towards the drain are degenerate. This value is consistent with the fact that the propagating state towards the drain is occupied at x_{max} up to those with the velocity v_f

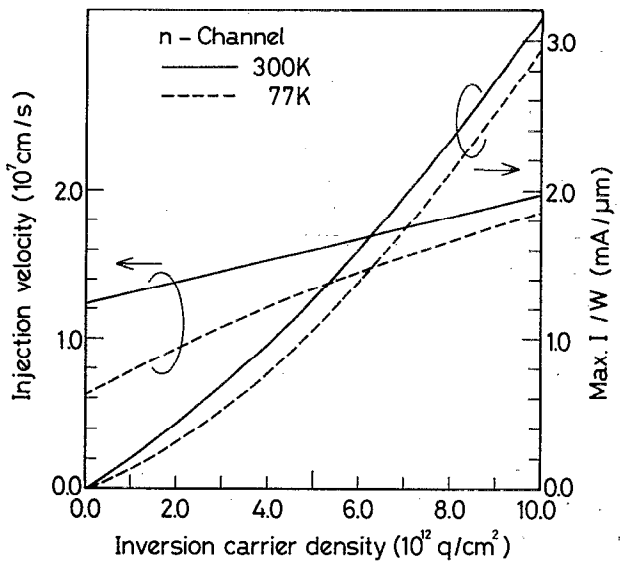


FIG. 8. Saturation current per unit width and the injection velocity of a nMOSFET on (100) plane as functions of inversion carrier density. The current value gives the maximum limit of the MOSFET current.

$$v_f=\frac{2\sqrt{\pi}\hbar\sqrt{|Q|}}{m_t\sqrt{qM_v}}. \tag{33}$$

The v_{inj} for ballistic MOSFETs is of the same order of magnitude as v_f. The value of v_{inj} and the nMOSFET current per unit width are plotted in Fig. 8 as functions of $|Q|$. Note that v_{inj} is around or slightly larger than the electronic saturation velocity in Si, the typical value of which is approximately 10^7 cm/s.

The saturated current, Eq. (24), is proportional to $(V_G-V_t)^{3/2}$ since the carrier mean velocity is proportional to $\sqrt{|Q|}$, as is the Fermi velocity. The fact that carriers are degenerate Fermi particles plays an important role here. This is in contrast with the classical MOSFET where the saturated current is proportional to $(V_G-V_t)^2$ or (V_G-V_t).

The ballistic current in the bulk material reminds us of the space charge limited current expressed by the 3/2 power law[19] of Langmuir's equation. The geometry of the MOSFET is different from those that are assumed in these bulk type current structures, but a similar simple discussion is attempted so that we may be able to gain some insight into the resultant potential variation along the channel. Suppose that the carrier transport is ballistic and the channel length is not as short. The channel potential variation along the x axis measured from the value at the source edge is denoted by $\Delta\phi(x)$ $[\Delta\phi(0)=0]$. The carrier mean velocity at x,$v(x)$, is related to $\Delta\phi(x)$ as

$$\tfrac{1}{2}mv(0)^2=\tfrac{1}{2}mv(x)^2+\Delta\phi(x), \tag{34}$$

where m is the carrier effective mass. The current continuity condition requires that

$$I=qWn(x)v(x). \tag{35}$$

I is the channel current and $n(x)$ is the carrier density at x. Adaptation of Poisson's equation is slightly more complex. In bulk current, the system is uniform in transverse direction and Poisson's equation is solved in the longitudinal direction. In MOSFET geometry, however, the potential variation transverse to the current direction is remarkable due to the gate voltage imposition and its variation along the channel is far more gentle. An approximate solution only considering the variation in the transverse direction (i.e., along the z axis) gives

$$V_G = -\gamma \frac{\Delta\phi(x)}{q} + \frac{qn(x)}{C_{ox}} + K, \tag{36}$$

where C_{ox} is the MOS oxide capacitance, and K and γ are constants. Elimination of $n(x)$ and $v(x)$ from Eqs. (34), (35), and (36) results in the equation of $\Delta\phi(x)$ whose unique solution satisfying the boundary condition is $\Delta\phi(x) \equiv 0$. This suggests that the potential, the charge density, as well as the carrier velocity is constant along the channel. All carriers injected from the source into the channel propagate along the channel with the same velocity without scattering or acceleration. However, when the channel length becomes extremely short so that the approximation in Eq. (36) becomes inappropriate due to the three dimensional nature of Poisson's equation, we will have a different result. In that case, the channel potential will be more influenced by the potential of the source or the drain electrode. Between the source potential level and the drain level which is lower by qV_D, there is a potential barrier in the channel controlled by the gate electrode. The potential profile is bell-shaped with a maximum in the channel. This potential maximum results in a bottleneck for the carrier transport, constituting a structure similar to the quantum point contact[20] between the two electrodes. Either side of the maximum is under the influence of the source or the drain electrode and is regarded as the extended region of these electrodes, respectively. In either case of the longer channel with the constant potential profile or the ultra-short channel with the bell-shaped potential profile, the assumption at the beginning of Sec. II is justified and the results obtained there are effective.

When the kinetic energy of a carrier is small, the carrier velocity is governed by acoustic phonon scattering, impurity scattering, and surface roughness scattering.[12] These scattering mechanisms will be moderated or removed if the device equipped with an intrinsic channel region is operated at low temperatures, and the normal electric field is relaxed. When the kinetic energy is large, however, optical phonon scattering (actually the optical phonon emission at low temperatures) becomes serious. Here we discuss the influence of optical phonon emission on the ballistic transport. Around x_{max} where carriers towards the drain are degenerate, the energy difference between the Fermi level and the lowest subband energy is expressed by $kT \ln(e^{\rho}-1)$ when the drain level is far below. If we set the electric field in the MOS oxide to be less than 2.5 MV/cm, so as to meet the request of long-term reliability of the device, the inversion charge $|Q|$ is less than 5.4×10^{12} q/cm^2 and the energy difference above is less than 54 meV. As optical phonon emission needs nearly 63 meV,[21] the influence of optical phonon emission around x_{max} to carrier transport is expected to be negligible. In ballistic MOSFETs with longer channels, the potential is constant throughout the channel and most carriers do not have enough kinetic energy to emit the optical phonon. Therefore, even if the optical phonon emission is present, the accidental event is suppressed to sustain the ballistic transport. Absence of the other scattering processes brings about the suppression of optical phonon emission. In ultra-short channel MOSFETs with the bell-shaped potential profile, optical phonon emission occurs on the potential slopes of the maximum point. However, those scattered carriers never go over the potential maximum back to the source electrode because their energy is already less than the lowest energy at the maximum point. Scattering occurs in the extended regions of the source and the drain electrode and the scattered carriers are eventually absorbed in these electrodes, respectively. Discussion so far suggests that the influence of optical phonon emission on transport of nMOSFET is quite small when the other scattering processes are suppressed and the induced carrier density is not so large.

V. DISCUSSION

As a figure of merit to show the high performance of MOSFETs, the transconductance value per unit width in current saturation is usually referred to. It correctly reflects the performance of a device, but it is not a convenient index for the analysis of MOSFET operation. High performance is brought about both by the high carrier concentration yielded by the large capacitance value, and the large carrier velocity due to the reduction of carrier scattering. Transconductance does not tell us which mechanism is predominant.

We can evaluate the mean carrier velocity v_{inj} of an experimental device by Eq. (31). Since $C_{eff}(V_G - V_t)$ usually gives the channel carrier density near the source edge, where the potential level is highest in the channel, the experimental v_{inj} value also gives the mean velocity with which carriers are injected from the source into the channel in a measured device. It will usually take smaller values than those derived in Sec. IV due to the presence of the backward flow of scattered carriers. However, it will approach the ballistic value as the scattering in the channel is eliminated. Therefore, the value of the experimental v_{inj} will serve as a figure of merit to give the carrier velocity contribution removing the effect of capacitance. A more convenient index, r, is given by the ratio of the experimental v_{inj} to the theoretical ballistic v_{inj}, or the ratio of the experimental I_{sat} to the theoretical ballistic I_{sat} that is,

$$r = \frac{v_{inj.exp}}{v_{inj.bal}} = \frac{I_{sat.exp}}{I_{sat.bal}}. \tag{37}$$

This factor will increase as the carrier scattering is reduced and will eventually reach unity, theoretically, when ballistic transport is realized. Normally it is less than or equal to unity, and some abnormal transport mechanism is dominant if it exceeds unity (for example, the lack of thermal equilibrium in the source electrode). Table II shows r factors evaluated from recent ultra-small MOSFET experiments.[1,5–7] Values of I and V_t are estimated from published data. Although

TABLE II. The index r showing the ballisticity is estimated for various experimental MOSFETs. In evaluation of $I_{sat.bal}$ of the pMOSFET, M_v is set to unity. No source and drain parasitic resistance is considered.

Data of experimental ultra-small MOSFET	Authors	Estimated value of index r
L=70 nm, t_{ox}=4.5 nm V_D=1 V, V_G-V_t=0.74 V 77 K, n-channel	a	0.92
L=90 nm, t_{ox}=3.5 nm V_D=1.5 V, V_G-V_t=0.56 V 85 K, n-channel	b	0.71
L=80 nm, t_{ox}=3.5 nm V_D=1.5 V, V_G-V_t=1.45 V 77 K, p-channel	c	0.32
L=100 nm, t_{ox}=4 nm V_D=2 V, V_G-V_t=1.7 V RT ?, p-channel	d	0.27

[a]See Ref. 1.
[b]See Ref. 5.
[c]See Ref. 6.
[d]See Ref. 7.

the estimation includes ambiguities, the overall characteristics are evident. The n-channel device shows a larger r value compared with that of the p-channel at the present stage, and the situation close to ballistic transport seems to be realized in the 70 nm example. A small (V_G-V_t) value is favorable and the increase of V_G seriously degrades the value of r. These results suggest that ballistic transport is highly probable in ultra-small MOSFETs, less than 50 nm, operated at low temperatures with small (V_G-V_t) values.

Frank *et al.* have performed an elaborate Monte Carlo simulation of a 30 nm dual gate SOI MOSFET.[13] Although it is at room temperature, the authors comment that carriers move nearly ballistically. Figure 9 shows a comparison of the present result with their simulation. The agreement is satisfactory. The present result has considered the series resistance of the narrow source and drain electrode regions leading to the channel. These resistance values are extracted from the published simulated structure. No adjustable parameters are included. Neglect of these series resistances leads to unexpected large current values. This fact re-confirms the importance of the effort to suppress the source and the drain parasitic resistances for improvement of ultra-small MOSFET performance.

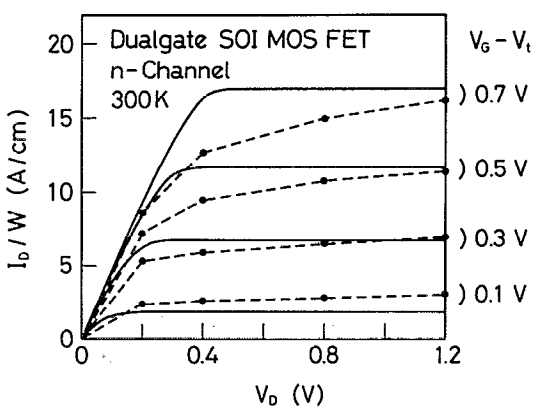


FIG. 9. Current-voltage characteristics of an ultra-small dual gate SOI MOSFET. Comparison of present result (solid line) with the Monte Carlo simulation by Frank *et al.* (dashed line). L=30 nm, t_{ox}=3 nm and the source and drain resistivity is 0.05 Ω mm. The present result assumes C_{eff} of 3.3 nm oxide and the source and drain series resistance of 9 mΩ each per 1 cm width. Carriers are assumed to be localized near the MOS interfaces.

VI. SUMMARY

MOSFETs in the ballistic transport mode are proposed and the expression of their I-V characteristics is derived. It shows the triode and the pentode operation as in the conventional MOSFETs. The current value is independent of the channel length and is proportional to the channel width. Carrier mobility is eliminated and the current is expressed with elementary parameters. The evaluated value gives the maximum limit of the current in a given MOSFET geometry. The result is originally derived for a bulk nMOSFET operated at low temperatures, but similar expressions are discussed to include p-channel MOSFETs, SOI MOSFETs, and also HEMT devices. The current control mechanism of ballistic MOSFETs is discussed. The ballistic current value is truncated by the carrier flux injected from the source electrode to the transport bottleneck point in the channel. It is the product of the carrier density at the bottleneck and the injection velocity which is the mean value of the group velocity of occupied states there. If the carriers towards the drain are degenerate, the injection velocity is of the same order of magnitude as the velocity with which a carrier at the source Fermi level passes the bottleneck point, and amounts to a value larger than 10^7 cm/s in high carrier densities. The effect of optical phonon emission is discussed. It is suggested that optical phonon emission will not bring about serious degradation of ballistic transport if the other scattering processes are removed and the channel carrier density is not as large. A convenient figure of merit to evaluate the ballisticity realized in an experimental device is proposed. The I-V characteristics of the ballistic MOSFETs proposed are compared with the result of an elaborate Monte Carlo simulation with satisfactory agreement.

ACKNOWLEDGMENTS

The author wishes to thank Dr. Yoshimi and Dr. Toriumi of Toshiba Research and Development Center, Professor A. Natori of the University of Electro-Communications, and Dr. M. V. Fischetti of IBM Research Division.

[1] G. A. Sai-Halasz, M. R. Wordeman, D. P. Kern, R. Rishton, and E. Ganin, IEEE Electron Device Lett. **9**, 464 (1988).
[2] A. Hartstein, Appl. Phys. Lett. **59**, 2028 (1991).
[3] T. Hashimoto, Y. Sudoh, H. Kurino, A. Narai, S. Yokoyama, Y. Horiike, and M. Koyanagi, Extended Abstracts of the International Conference on Solid State Devices and Materials, 1992, p. 490.
[4] A. Toriumi, M. Iwase, M. Takahashi, H. Niiyama, M. Fukumoto, S. Inaba, I. Mori, and M. Yoshimi, Extended Abstracts of the International Conference on Solid State Devices and Materials, 1992, p. 487.
[5] Y. Mii, S. Rishton, Y Taur, D. Kern, T. Lii, K. Lee, K. Jenkins, D. Quinlan,

T. Brown, Jr., D. Danner, F. Sewell, and M. Polcari, Symposium on VLSI Technology, Digest Technology Papers, 1993, p. 91.
[6] Y. Taur, S. Cohen, S. Wind, T. Lii, C. Hsu, D. Quinlan, C. Chang, D. Buchanan, P. Agnello, Y. Mii, C. Reeves, A. Acovic, and V. Kesan, IEDM Tech. Dig., p. 901 (1992).
[7] K. F. Lee, R. H. Yan, D. Y. Jeon, Y. O. Kim, D. M. Tennant, E. H. Westerwick, K. Early, G. M. Chin, M. D. Morris, R. W. Johnson, T. M. Liu, R. C. Kistler, A. M. Voshchenkov, R. G. Swartz, and A. Ourmazd, IEDM Tech. Dig., p. 1012 (1992).
[8] M. Ono, M. Saito, T. Yoshitomi, C. Fiegna, T. Ohguro, and H. Iwai, IEDM Tech. Dig., p. 119 (1993).
[9] S. Y. Chow, D. A. Antoniadis, and H. I. Smith, IEEE Electron Device Lett. **EDL-6**, 665 (1985).
[10] G. G. Shahidi, D. A. Antoniadis, and H. I. Smith, IEEE Electron Device Lett. **9**, 94 (1988).
[11] S. E. Laux and M. V. Fischetti, IEEE Electron Device Lett. **9**, 467 (1988).
[12] S. Takagi, M. Iwase, and A. Toriumi, IEDM Tech. Dig., p. 398 (1988).
[13] D. J. Frank, S. E. Laux, and M. V. Fischetti, IEDM Tech. Dig., p. 553 (1992).
[14] M. Yokoyama, T. Hidaka, Y. W. Yi, K. Masu, and K. Tsubouchi, Jpn. J. Appl. Phys. **32**, 419 (1993).
[15] K. Natori, Jpn. J. Appl. Phys. **33**, 554 (1994).
[16] R. Landauer, Philos. Mag. **21**, 863 (1970).
[17] J. S. Blakemore, Electron Commun. **29**, 131 (1952).
[18] F. Stern, Phys. Rev. B **5**, 4891 (1972).
[19] L. F. Eastman, R. Stall, D. Woodard, N. Dandekar, C. E. C. Wood, M. S. Shur, and K. Board, Electron. Lett. **16**, 524 (1980).
[20] B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel, and C. T. Foxon, Phys. Rev. Lett. **60**, 848 (1988).
[21] S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. (Wiley, New York, 1981), p. 851.