Received 9 July 2022; revised 10 March 2023; accepted 29 March 2023. Date of publication 5 April 2023; date of current version 14 April 2023.
The review of this article was arranged by Editor M. J. Kumar.

Digital Object Identifier 10.1109/JEDS.2023.3264814

Formulation of Ground States for 2DEG at Rough Surfaces and Application to Nonlinear Model of Surface Roughness Scattering in nMOSFETs

KEI SUMITA (Graduate Student Member, IEEE), MIN-SOO KANG, KASIDIT TOPRASERTPONG (Member, IEEE), MITSURU TAKENAKA (Member, IEEE), AND SHINICHI TAKAGI (Senior Member, IEEE)

Department of Electrical Engineering and Information Systems, The University of Tokyo, Tokyo 113-8656, Japan

CORRESPONDING AUTHOR: K. SUMITA (e-mail: sumita@mosfet.t.u-tokyo.ac.jp)

This work was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI, Japan, under Grant 17H06148, Grant 21J10272, and Grant 22H00208; and in part by the New Energy and Industrial Technology Development Organization (NEDO), Japan, under Project JPNP16007.

ABSTRACT Electron mobility in extremely-thin-body (ETB) nanosheet channels and at cryogenic temperature is known to be dominated by surface roughness scattering. However, the conventional model of surface roughness scattering lacks accuracy because it requires the use of excessive roughness parameters to represent the experimental results. One of the main difficulties for the surface roughness scattering model is that the higher-order perturbations should be accurately included in the model because the surface roughness scattering is a strongly nonlinear phenomenon. Therefore, in this study, the formulation of ground states of two-dimensional electron gas (2DEG) at rough surfaces is derived by introducing a concept of the space-averaged perturbation Hamiltonian. This revised formulation of 2DEG at rough surfaces is different from the conventional solution for 2DEG at the flat surface. The space-averaged perturbation Hamiltonian is invisible in the linearized perturbation system, while its effect is significant in the system with the nonlinear perturbation energy. We combine the revised 2DEG formulation with a nonlinear model of surface roughness scattering and calculate the 2DEG mobility of the bulk Si and ETB Si-on-insulator (SOI) nMOSFETs. As a result, the experimental mobility of bulk and ETB SOI nMOSFETs is well explained in a wide temperature range of 4.2 to 300 K by using the roughness parameters experimentally obtained by transmission electron microscopy (TEM), which also supports the understanding of mobility at cryogenic temperature. The revised nonlinear model reveals that surface roughness scattering under the present model is 13 times stronger than that predicted by the conventional linear model.

INDEX TERMS Metal-oxide-semiconductor field-effect transistor (MOSFET), mobility, nanosheet, surface roughness scattering.

I. INTRODUCTION

An extremely-thin-body (ETB) nanosheet channel is one of the most promising structures for the future advanced technology nodes of complementary metal-oxide-semiconductor (CMOS) devices because of its superior immunity to short channel effects and compatibility with the three-dimensional stacking process [1], [2], [3]. Here, thickness scaling of the nanosheet channel is necessary to improve gate controllability, while mobility degradation is a severe challenge for channel thickness scaling [4], [5], [6], [7], [8]. Therefore, the high-mobility materials robust to surface roughness scattering in ETB channels have recently gained great interest. In parallel, CMOS circuits at cryogenic temperature have recently attracted strong interest as the control circuits of the quantum computer [9]. Since the dominant scattering mechanism of electrons in ETB channels and at

cryogenic temperature is surface roughness scattering [10], [11], [12], [13], [14], [15], [16], the quantitative understanding of surface roughness scattering becomes increasingly important. However, the conventional model of surface roughness scattering [10], [11], [12], [13], [14], [15], [16] has the problem that the surface roughness is overestimated in simulations to represent the experiments in comparison with the roughness parameters measured by transmission electron microscopy (TEM) [17], [18], [19]. As a result, it is difficult with the conventional model of surface roughness scattering to predict and assess the mobility in ETB channels for future channel materials alternative to Si.

Particularly, in the conventional Prange-Nee model of surface roughness scattering, the linearized perturbation Hamiltonian is employed to simplify the calculation [10], [11], [12], [13], [14], [15], [16]. However, the actual physical picture of surface roughness scattering is strongly nonlinear and asymmetric phenomena, which cannot be quantitatively modeled by such a linear model with only the first-order perturbation. Here, the nonlinearity means that the matrix element is not proportional to the amount of roughness. This fact leads to the excessively large roughness assumed to represent the experimental mobility. On the other hand, a nonlinear model of surface roughness scattering has been recently proposed [20], [21], [22], [23], which can include the higher order perturbations. This model has explained the experimental results with reasonable roughness parameters. These previous studies [20], [21], [22], [23] are valuable as the first introduction of the nonlinear model. However, their formulations are still insufficient because the nonlinear perturbation effect on wavefunctions of confined electrons, which does not occur for the linear perturbation, is neglected. Therefore, it is more difficult to represent the surface-roughness (SR)-limited mobility accurately for electronic systems with large scattering potentials and high kinetic energies of electrons, such as ETB III-V channels with the light effective mass. For example, the kinetic energy term in Hamiltonian, which is large for ETB III-V channels, is unreasonably neglected in the previous nonlinear model [20], [21], [22], [23]. In fact, under a nonlinear perturbation model, where the space-averaged perturbation Hamiltonian is not zero, the steady state of electrons is also affected by the perturbation Hamiltonian.

Therefore, in this study, we present a new formulation of surface roughness scattering of the two-dimensional electron gas (2DEG) at metal-oxide-semiconductor (MOS) surfaces, where the nonlinear model is applied to 2DEG with revised ground states under the rough surfaces. Compared to a model based on first-principles calculations [24], our model within the framework of effective mass approximation and Fermi's golden rule has advantages in terms of the computational cost and the convenience. The SR-limited mobility of bulk and Si-On-Insulator (SOI) n-channel metal-oxide-semiconductor field-effect transistors (nMOSFETs) with Si/SiO_2 MOS interfaces is calculated by using our revised nonlinear model and compared with the experimental mobility of Si nMOSFETs, evaluated in a wide temperature range of 4.2 – 300 K. It is found that our model can represent the experimental mobility of the bulk and SOI nMOSFETs by using the same and realistic roughness parameters, which are in good agreements with TEM analysis. Also, by comparing calculated and experimental mobility at cryogenic temperature, we have suggested the possibilities that the current model of the screening effect on surface roughness scattering is insufficient and/or that the conduction of electrons in the Δ_4 valley of Si can be dominant by that in the low-mobility tail states.

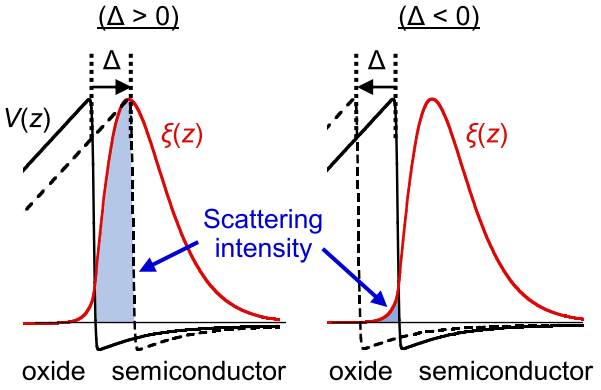

FIGURE 1. Schematic view of band diagrams to illustrate surface roughness scattering. An electron is strongly scattered by the oxide potential for positive Δ, while scattering is weaker for negative Δ.

II. THEORY AND SIMULATION METHOD

A. GROUND STATES OF 2DEG AT ROUGH SURFACE

The schematic view of surface roughness scattering is shown in Fig. 1. Electrons are scattered more strongly by the oxide potential, when the height of roughness Δ is positive. Therefore, surface roughness scattering is regarded as a strongly nonlinear and asymmetric scattering process with respect to Δ. In order to calculate the mobility accurately, such nonlinearity needs to be included in the model, indicating that the higher-order perturbation should be included in the model of surface roughness scattering. Here, the Prange-Nee model has been most widely used among the previous studies of surface roughness scattering [10], [11], [12], [13], [14], [15], [16]. In the Prange-Nee model, however, only the first-order perturbation Hamiltonian is considered into account, and the direct scattering by the oxide potential is neglected. Therefore, the Prange-Nee model is lacking in accuracy.

Recently, the need and first construction of nonlinear modeling in surface roughness scattering has been reported [20], [21], [22], [23]. However, as will be discussed in the next Section II-B, the previous model is still insufficient in its original form. This is mainly attributed to the fact that the nonlinear effect on the ground states of the wavefunction is fully neglected. This effect becomes more serious as the nonlinearity of perturbation becomes stronger, for example in III-V and ETB channels. This problem can be solved

by considering the nonlinear effect on the ground states of 2DEG. In fact, under a nonlinear perturbation model, where the space-averaged perturbation Hamiltonian is not zero, the steady ground states of electrons are also affected by the perturbation Hamiltonian. In this section, therefore, we derive the ground states of 2DEG with a rough MOS surface. In the next section, it will be shown that our formulation of the ground states is necessary for the nonlinear model of surface roughness scattering.

Since the mean free path of electrons (few tens nm) is much longer than the correlation length of surface roughness ($\sim$ 1 nm), conducting electrons should feel the effective potential with the averaged perturbation. Therefore, the wavefunction of 2DEG, $\xi(z)$, is reasonably approximated as uniform on the conduction plane. Then, the energy in the z direction, $E_{z,i}$, and the in-plane averaged energy, $\langle E_{z,i}\rangle_r$, in subband i can be expressed by

$$E_{z,i}(r) = \int_{-\infty}^{\infty} \xi_i^*(z) H_{\Delta(r)} \xi_i(z) dz, \tag{1}$$

$$\langle E_{z,i}\rangle_r = \int_{-\infty}^{\infty} \xi_i^*(z) \langle H_\Delta\rangle_r \xi_i(z) dz, \tag{2}$$

where $\langle\cdots\rangle_r$ is an average operator on the plane as

$$\langle X(z,\Delta)\rangle_r = \int_{-\infty}^{\infty} X(z,\Delta) f_1(\Delta) d\Delta. \tag{3}$$

Here, $X(z,\Delta)$ is a roughness-dependent physical quantity, and $f_1(\Delta)$ is the probability distribution function of roughness Δ. H_Δ is only the r-dependent term in (1), which is the Hamiltonian of a channel with roughness $\Delta(r)$ as

$$H_\Delta = -\frac{\hbar^2}{2}\frac{d}{dz}\left(\frac{1}{m_z(z,\Delta(r))}\frac{d}{dz}\right) + V(z,\Delta(r)), \tag{4}$$

where $m_z(z,\Delta)$ is a confinement effective mass at point z, $V(z,\Delta)$ is a potential, and $\hbar$ is the reduced Plank constant. When the wavefunction $\xi_i(z)$ is the eigenstate of $\langle H_\Delta\rangle_r$, $\langle E_{z,i}\rangle_r$ corresponding to the total energy in the system is minimized. Therefore, the energy operator of the ground state of Hamiltonian is $H_{\rm gnd} = \langle H_\Delta\rangle_r$, which is the total energy including the perturbation. It should be noted that $H_{\rm gnd} = \langle H_\Delta\rangle_r$ is only a pseudo-steady state for electrons that are conducting and scattered. The concept of the averaged Hamiltonian is also verified by the fact that the mean free path of electrons is much longer than the correlation length of surface roughness.

When the perturbation has a linear relationship to the height of roughness as $H_\Delta = H_0 + H^1 \times \Delta$, where H^1 is the first-order perturbation, the averaged Hamiltonian $\langle H_\Delta\rangle_r$ amounts to H_0. This is implicitly assumed in the Prange-Nee model. However, $\langle H_\Delta\rangle_r \neq H_0$ for the nonlinear perturbation with respect to surface roughness. Here, $\langle H_\Delta\rangle_r$ can be derived by

$$\langle H_\Delta\rangle_r = -\frac{\hbar^2}{2}\left\langle\frac{1}{m_z}\right\rangle_r\frac{d^2}{dz^2} + \frac{\hbar^2}{2}\left\langle\frac{1}{m_z^2}\frac{dm_z}{dz}\frac{d}{dz}\right\rangle_r + \langle V(z,\Delta)\rangle_r. \tag{5}$$

Effective mass and potential abruptly change at a semiconductor/oxide interface. Therefore, all $\langle X(z,\Delta)\rangle_r$ terms in (5) near the interface are different from $X(z,0)$, while $\langle X(z,\Delta)\rangle_r$ far from the interface is almost equal to $X(z,0)$. In the numerical calculation of $\langle\cdots\rangle_r$, the physical quantity is smoothly connected at the interface by the sigmoid function $S(z)$ as

$$V(z,\Delta) = V_q(z,\Delta) + V_{ox}(1 - S(z-\Delta) + S(z-T_B)), \tag{6}$$

$$m_z(z,\Delta) = m_{ox} - (m_{ox} - m_{sz}) \times (S(z-\Delta) - S(z-T_B)), \tag{7}$$

$$S(z) = \frac{1}{1+\exp(-z/z_0)}, \tag{8}$$

where z_0 and the mesh size is 0.02 nm, which is small enough for the calculation results to converge because the impact of the shape of the Sigmoid function is negligible on the average quantities $\langle\cdots\rangle_r$. Also, V_q and V_{ox} are the electrostatic potential and the energy barrier between a semiconductor and a gate oxide, respectively, m_{ox} and m_{sz} are the confinement effective mass of oxide and semiconductor, respectively, and T_B is the channel thickness. The potential V_q should be calculated for each Δ by solving the Poisson equation. In order to calculate the in-plane average $\langle\cdots\rangle_r$ in (3) and (5), the probability distribution function $f_1(\Delta)$ should be accurately modeled to fit the real MOS interface roughness. The autocorrelation of the roughness is generally modeled by the Gaussian or exponential formulation [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23]. On the other hand, $f_1(\Delta)$ is actually complex due to the atomic discreteness. However, $f_1(\Delta)$ is also modeled by the Gaussian or exponential formulation to simplify the calculation in this study as

$$f_1(\Delta) = \begin{cases} \frac{1}{\Delta_{\rm rms}\sqrt{2\pi}}\exp\left(-\frac{\Delta^2}{2\Delta_{\rm rms}^2}\right) & \ldots \text{(Gaussian)} \\ \frac{1}{\sqrt{2}\Delta_{\rm rms}}\exp\left(-\frac{\sqrt{2}|\Delta|}{\Delta_{\rm rms}}\right) & \ldots \text{(exponential)}, \end{cases} \tag{9}$$

where $\Delta_{\rm rms}$ is a root-mean-square (RMS) value of surface roughness. In this study, the exponential formulation of $f_1(\Delta)$ is used.

The examples of the potential and the wavefunction of the lowest subband are shown in Fig. 2. Here, the black and red curves correspond to them under the completely flat and rough interface, respectively. $\Delta_{\rm rms}$ of 0.2 nm is used for all calculations. The apparent potential of $\langle V(z)\rangle_r$ has a less steep slope at the interface because it is the averaged one. The wavefunction of the Δ_2 valley in the bulk (100) Si/SiO_2 interface in Fig. 2(a) moves away from the interface by taking into account the roughness, to minimize the perturbation energy. On the other hand, the change in the wavefunction of the Δ_4 valley in Fig. 2(b) by roughness is much smaller because the wavefunction spreads more widely due to lighter m_{sz}. For the 5-nm-thick $InAs/Al_2O_3$ interface in Fig. 2(c), the penetration of $\xi(z)$ into the oxide is large because of the very light m_{sz}. This large penetration makes it more difficult to calculate the matrix element of surface

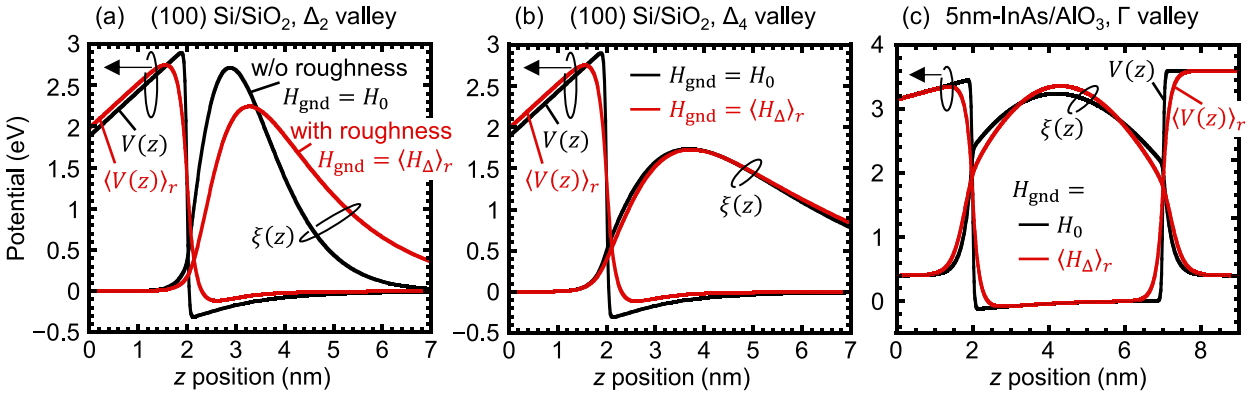

FIGURE 2. Potentials and wave functions of the lowest subband with flat and rough surfaces, where Hamiltonian corresponds to H_0 and $\langle H_\Delta \rangle_r$, respectively. (a) Δ_2 valley electron and (b) Δ_4 valley electron in a (100) Si/SiO_2 interface under effective electric field of 0.9 MV/cm. (c) Γ valley electron in a 5-nm-thick $InAs/Al_2O_3$ interface under N_s of 5×10^{12} cm^{-2}. Here, Δ_{rms} of 0.2 nm is assumed for all the calculations.

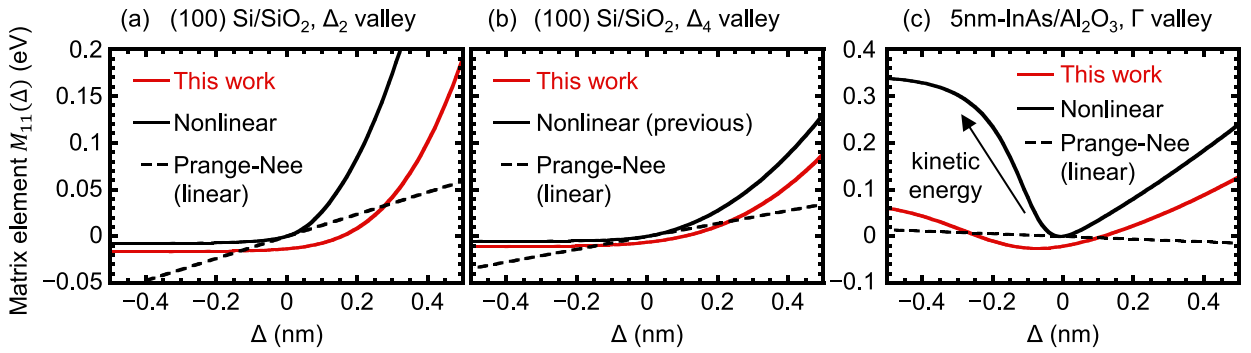

FIGURE 3. Matrix element $M_{ij}(\Delta)$ of intra-subband scattering in the lowest subband as $i = j = 1$, calculated by the Prange-Nee (linear) model, the previous nonlinear model and our model. (a) Δ_2 valley and (b) Δ_4 valley in the (100) Si/SiO_2 interface under effective electric field of 0.9 MV/cm. (c) Γ valley in the 5-nm-thick $InAs/Al_2O_3$ interface under N_s of 5×10^{12} cm^{-2}. Here, Δ_{rms} of 0.2 nm is assumed for all the calculations.

roughness scattering for III-V in the Prange-Nee and the previous nonlinear models.

B. MODELING METHOD OF SURFACE ROUGHNESS SCATTERING

In this section, the calculation method of the matrix element and the necessity of the averaged Hamiltonian including the effect of surface roughness on the ground states are described. The matrix element is the most important parameter to determine the mobility limited by surface roughness scattering. The matrix element from subband i to j is expressed by

$$M_{ij}(\Delta) = \int_{-\infty}^{\infty} \xi_j^*(z)\left[H_\Delta - H_{\rm gnd}\right]\xi_i(z)dz, \tag{10}$$

where the ground state of Hamiltonian is

$$H_{\rm gnd} = H_0 \ldots \text{(at a flat interface)}, \tag{11}$$

$$H_{\rm gnd} = \langle H_\Delta \rangle_r \ldots \text{(at a rough interface)}. \tag{12}$$

The (generalized) Prange-Nee model is commonly used in the previous studies, where the linear relationship of $M_{ij}(\Delta) \propto \Delta$ is assumed [10], [11], [12], [13], [14], [15], [16]. Therefore, the matrix element in the generalized Prange-Nee model is derived by

$$M_{ij}^{\rm GPN}(\Delta) = \left.\frac{dM_{ij}(\Delta)}{d\Delta}\right|_{\Delta=0} \times \Delta. \tag{13}$$

Although the expression of (13) is very simple and easy to calculate, the Prange-Nee model lacks accuracy for the strongly nonlinear and asymmetric nature of surface roughness scattering, shown in Fig. 1.

In the previously proposed nonlinear model [20], [21], [22], [23], $M_{ij}(\Delta)$ in (10) was numerically calculated for all Δ values by using the formulation of $H_{\rm gnd}$ by (11). Then, any higher order perturbations can be included in the nonlinear model. On the other hand, in our model, $M_{ij}(\Delta)$ is calculated for all Δ by using the expression of $H_{\rm gnd}$ by (12), which includes the roughness effect on the ground state of Hamiltonian.

The scattering probability in the wavenumber space can be described by the power spectrum density of matrix elements, according to the Fermi's golden rule. The power spectrum density is the Fourier transform of the autocorrelation according to the Wiener–Khinchin theorem. Here, the autocorrelation of the surface roughness, $C_\Delta(r)$, is generally modeled by [10], [11], [12], [13], [14], [15], [16], [17], [18],

[19], [20], [21], [22], [23]

$$C_\Delta(r) = \Delta_{\rm rms}^2 \exp\left[-2\left(\frac{|r|}{\sqrt{2}\Lambda}\right)^n\right] \tag{14}$$

where $n = 1$ and 2 correspond to the exponential and Gaussian forms, respectively. Λ is the correlation length of surface roughness. In this study, the exponential formulation ($n = 1$) is used for $C_\Delta(r)$ because of the good agreement with the TEM analysis [17], [18], [19], [25]. The autocorrelation of the matrix element, $C_M(r)$, is derived by [20], [21], [22], [23], [26], [27], [28]

$$C_M(r) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} M_{ij}(\Delta_1)M_{ij}(\Delta_2)f_2(\Delta_1, \Delta_2, \rho)d\Delta_1 d\Delta_2, \tag{15}$$

$$f_2(\Delta_1, \Delta_2, \rho) = \frac{1}{\sqrt{1-\rho^2}} f_1(\Delta_1)f_1(\Delta_2) \times \exp\left[-\frac{\left(\rho\Delta_1'\right)^2 - 2\rho\Delta_1'\Delta_2' + \left(\rho\Delta_2'\right)^2}{2\left(1-\rho^2\right)}\right], \tag{16}$$

where $f_2(\Delta_1, \Delta_2, \rho)$ is the two-variate joint probability distribution function, which means the probability that the roughness has Δ_1 and Δ_2 at a distance r. $\rho(r) = C_\Delta(r)/C_\Delta(0)$ is the Pearson's product-moment correlation coefficient. Also, Δ_i' for $i = 1,\ 2$ is expressed by

$$\Delta_i' = \Phi^{-1}\left[\int_{-\infty}^{\Delta_i} f_1(\Delta)d\Delta\right], \tag{17}$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left(-\frac{1}{2}t^2\right)dt. \tag{18}$$

Here, $\Delta_i' = \Delta_i/\Delta_{\rm rms}$ for the Gaussian formulation, while $\Delta_i' \neq \Delta_i/\Delta_{\rm rms}$ for the exponential formulation in (17). It is worth pointing out that both the Gaussian or exponential formulation can be independently applied to $C_\Delta(r)$ and $f_2(\Delta_1, \Delta_2, \rho)$. In this study, the exponential formulations of $f_1(\Delta)$, $C_\Delta(r)$, and $f_2(\Delta_1, \Delta_2, \rho)$ are simply used. The power spectrum density of the matrix element is derived by the two-dimensional Fourier transform of $C_M(r)$ as

$$\left|M_{ij}(q)\right|^2 = 2\pi\int_0^{\infty} C_M(r)J_0(qr)rdr, \tag{19}$$

where q is the wavenumber and $J_0(qr)$ is the zero-order Bessel function [20]. It should be noted that when $C_M(r \to \infty) \neq 0$, the integral in (19) diverges to infinity and is very oscillatory as $|M_{ij}(q)|^2 \to \pm\infty$. However, in the previous nonlinear model [20], [21], [22], [23], since $C_M(r \to \infty) \neq 0$, the calculation of (19) might not have been correctly done, and the DC component in (19) might be ignored. Here, since $\lim_{r\to\infty} f_2(\Delta_1, \Delta_2, \rho) = f_1(\Delta_1)f_1(\Delta_2)$ in (15), $C_M(r \to \infty)$ is

$$C_M(r \to \infty) = \left(\int_{-\infty}^{\infty} M_{ij}(\Delta)f_1(\Delta)d\Delta\right)^2 = \left\langle M_{ij}(\Delta)\right\rangle_r^2. \tag{20}$$

Therefore, $\langle M_{ij}(\Delta)\rangle_r = 0$ is a necessary condition for the matrix element model. It should be noted that this condition is not visible in the linear Prange-Nee model because it always holds. While the condition of $\langle M_{ij}(\Delta)\rangle_r = 0$ is only derived from the convergence condition in (19), $\langle M_{ij}(\Delta)\rangle_r$ has also the important physical meaning of being the expected value of the perturbation energy. Thus, $\langle M_{ij}(\Delta)\rangle_r$ should be zero, when applied to the perturbation theory. On the other hand, one may consider that $M_{ij}(\Delta)$ in (15) can be replaced by $M_{ij}(\Delta) - \langle M_{ij}(\Delta)\rangle_r$ because, statistically, the autocorrelation function is determined by subtracting the average of the original function. Of course, it is possible to calculate $|M_{ij}(q)|^2$ by replacing $M_{ij}(\Delta) \to (M_{ij}(\Delta) - \langle M_{ij}(\Delta)\rangle_r)$. However, we believe that this correction lacks the physical adequacy of the perturbation.

The matrix elements calculated by the Prange-Nee model [14], the previous nonlinear model [20] and our model are shown in Fig. 3. Here, $M_{ij}(\Delta)$ of the lowest subband as $i = j = 1$ is shown. The values of $M_{11}(\Delta)$ calculated by the Prange-Nee model are so different from those calculated by the nonlinear ones. As clearly seen here, the space-averaged perturbation energy, $\langle M_{11}(\Delta)\rangle_r$, is positive in the previous nonlinear model, in spite of the fact that $\langle M_{ij}(\Delta)\rangle_r = 0$ is the necessary condition as discussed above. On the other hand, $M_{ij}(\Delta) \neq 0$ at $\Delta = 0$ in our model because the Hamiltonian at the geometrical center of $\Delta = 0$, H_0, is different from the perturbative average of the Hamiltonian. As a result, $\langle M_{ij}(\Delta)\rangle_r$ in our model is strictly equal to zero, as proven in the following.

$$\begin{aligned}\left\langle M_{ij}(\Delta)\right\rangle_r &= \left\langle\xi_j\right|\langle H_\Delta\rangle_r - \left\langle H_{\rm gnd}\right\rangle_r\left|\xi_i\right\rangle \\ &= \left\langle\xi_j\right|\langle H_\Delta\rangle_r - \langle\langle H_\Delta\rangle_r\rangle_r\left|\xi_i\right\rangle = 0.\end{aligned} \tag{21}$$

A characteristic example of the quite nonlinear $M_{ij}(\Delta)$ is shown in Fig. 3(c) for a 5-nm-thick InAs/Al_2O_3 interface, where m_{sz} is very light. The matrix element calculated by the generalized Prange-Nee model in Fig. 3(c) is not realistic because the derivative of $M_{ij}(\Delta)$ at $\Delta = 0$ is coincidentally very small due to a large contribution of the kinetic energy term. In the previous nonlinear model, the amount of the matrix element in $\Delta < 0$ is also quite large and positive because the kinetic energy term of Hamiltonian is very large for InAs due to light m_{sz}. Therefore, such kinetic energy of the first term in (4) was unreasonably omitted in the calculation of the matrix element in the previous studies [20], [21], [22], [23], which is significant for III-V with light m_{sz}. However, this way undermines the physical understanding of surface roughness scattering. On the other hand, since our model can include higher-order perturbations and the kinetic energy term in Hamiltonian, it can be applied to materials with light m_{sz} such as III-V semiconductors.

The autocorrelation of the matrix element, $C_M(r)$, for electrons in the Δ_2 valley at the (100) Si/SiO_2 interface, calculated by the previous nonlinear [20] and our models, is shown in Fig. 4. $C_M(r \to \infty) \neq 0$ for the previous nonlinear model because $\langle M_{ij}(\Delta)\rangle_r^2 \neq 0$ as shown in Fig. 3(a). This nonzero $C_M(r \to \infty)$ in the previous nonlinear model causes the divergence of $|M_{ij}(q)|^2$ in (19). Therefore, the averaged Hamiltonian $\langle H_\Delta\rangle_r$ should be used for the ground states at rough surfaces. It is confirmed, on the other hand,

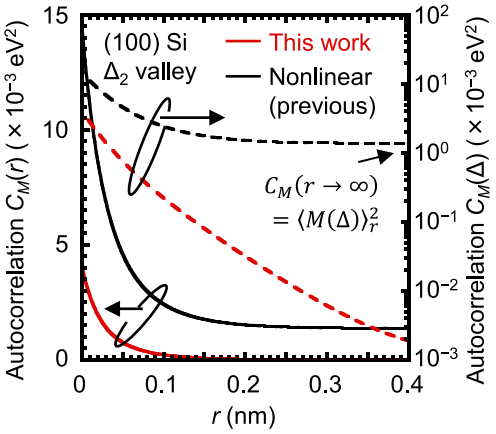

FIGURE 4. Autocorrelation of matrix elements $C_M(r)$ in the linear and log scale for Δ_2 valley electrons in the (100) Si/SiO_2 interface calculated by the previous nonlinear model and our model. Here, the ground states of Hamiltonian are H_0 and $\langle H_\Delta \rangle_r$ in the previous nonlinear model and our model, respectively.

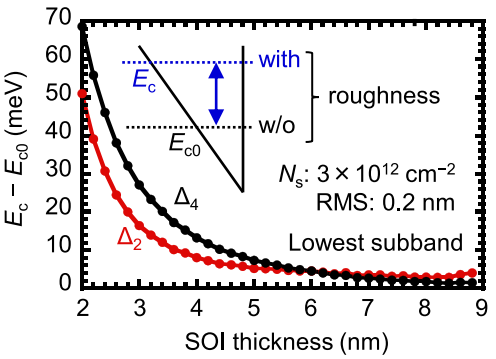

FIGURE 5. Difference of the quantization energy between the flat and rough surfaces for Δ_2 and Δ_4 valley electrons in the bulk (100) Si/SiO_2 interface under N_s of 3×10^{12} cm^{-2} and Δ_{rms} of 0.2 nm.

for our model that $C_M(r \to \infty) = 0$. The energy difference of the lowest subband between H_0 and $\langle H_\Delta \rangle_r$, $(E_c - E_{c0})$, is shown in Fig. 5. Here, $(E_c - E_{c0})$ is the additional energy in the system caused by the existence of the nonlinear perturbation due to surface roughness. This additional increase in the quantization energy can appear as an additional threshold voltage shift from the expected value at the flat interface, which was actually observed experimentally in Si nMOSFETs [4].

C. SURFACE ROUGHNESS AT TOP AND BOTTOM INTERFACE

In the previous section, we consider only the roughness of the one-side interface in ETB channels. However, in the real ETB channels, there is surface roughness at the top and bottom interfaces. In the previous studies [16], [20], the matrix elements introduced by the roughness of the top and bottom interfaces are added when the roughness at the top and bottom interfaces are uncorrelated. However, it should be noted that, in our model, the matrix elements cannot be simply added even under the assumption that the roughness at the top and bottom interfaces are uncorrelated. This is because the existence of the roughness at both interfaces itself modulates the wavefunction. As similar to the discussion in the previous section, the matrix element is

$$M_{ij}(\Delta_t, \Delta_b) = \int_{-\infty}^{\infty} \xi_j^*(z)\left[H_{\Delta_t,\Delta_b} - H_{\rm gnd}\right]\xi_i(z)dz, \tag{22}$$

$$H_{\Delta_t,\Delta_b} = -\frac{\hbar^2}{2}\frac{d}{dz}\left(\frac{1}{m_z(z)}\frac{d}{dz}\right) + V(z), \tag{23}$$

$$V(z) = V_q(z) + V_{ox}(1 - S(z - \Delta_t) + S(z - T_B + \Delta_b)), \tag{24}$$

$$m_z(z) = m_{ox} - (m_{ox} - m_{sz}) \times (S(z - \Delta_t) - S(z - T_B + \Delta_b)), \tag{25}$$

where Δ_t and Δ_b are the height of roughness at the top and bottom interfaces, respectively. Here, $H_{\rm gnd}$ is the averaged Hamiltonian $\langle H_{\Delta_t,\Delta_b} \rangle_r$, where the average operator is

$$\langle X(\Delta_t, \Delta_b)\rangle_r = \iint X(\Delta_t, \Delta_b) f_2(\Delta_t, \Delta_b, \rho_{tb}) d\Delta_t d\Delta_b. \tag{26}$$

Also, ρ_{tb} is the Pearson's product-moment correlation coefficient between the top and bottom interfaces. In this study, $\rho_{tb} = 0$ is reasonably assumed to simplify the calculation. When $\rho_{tb} = 0$, f_2 can be simplified as $f_2(\Delta_t, \Delta_b, 0) = f_1(\Delta_t) \times f_1(\Delta_b)$. Finally, the autocorrelation of $M_{ij}(\Delta_t, \Delta_b)$ is

$$C_M(r) = \iiiint M_{ij}(\Delta_{t1}, \Delta_{b1}) M_{ij}(\Delta_{t2}, \Delta_{b2}) \times f_4(\Delta_{t1}, \Delta_{t2}, \Delta_{b1}, \Delta_{b2}, \rho_{tb}) d\Delta_{t1} d\Delta_{t2} d\Delta_{b1} d\Delta_{b2}, \tag{27}$$

where the integration range is $(-\infty, \infty)$ in (26) and (27), and $f_4(\cdots)$ is the four-variate joint probability distribution function. When the top and bottom interfaces are uncorrelated as $\rho_{tb} = 0$, $f_4(\cdots)$ can be simplified by

$$f_4(\Delta_{t1}, \Delta_{t2}, \Delta_{b1}, \Delta_{b2}, \rho_{tb} = 0) = f_2(\Delta_{t1}, \Delta_{t2}, \rho_t) \times f_2(\Delta_{b1}, \Delta_{b2}, \rho_b), \tag{28}$$

where $\rho_t(r)$ and $\rho_b(r)$ are the Pearson's product-moment correlation coefficient of the top and bottom interfaces, respectively.

Although the expression of $C_M(r)$ by (27) is a strict one, it is not realistic to directly calculate (27) because it is extremely time-consuming. Therefore, we approximated the autocorrelation $C_M(r)$ by

$$C_M(r) \cong C_M\left(r\middle|\Delta_{{\rm rms},t} = 0\right) + C_M\left(r\middle|\Delta_{{\rm rms},b} = 0\right), \tag{29}$$

where $\Delta_{{\rm rms},t}$ and $\Delta_{{\rm rms},b}$ are RMS of the surface roughness at the top and bottom interfaces, respectively. When RMS at the one side interface is taken to be zero, $C_M(r)$ can be calculated by (15), which makes the calculation cost realistic. However, it is worth to note that the approximation of $C_M(r)$ by (29) underestimates the real value of $C_M(r)$.

D. SCREENING EFFECT

The matrix element $M_{ij}^{scr}(q)$ including the screening effect is delivered by using the scalar dielectric function $\epsilon_D(q)$ as [16], [29]

$$M_{ij}^{scr}(q) = M_{ij}(q)/\epsilon_D(q), \tag{30}$$

$$\epsilon_D(q) = 1 + \sum_i \frac{e^2}{2\varepsilon_s q} F_{i,i}(q)\Pi_{i,i}(q), \tag{31}$$

where ε_s is the dielectric constant of a semiconductor. The definition of the form factor $F_{i,i}(q)$ and the polarization factor $\Pi_{i,i}(q)$ are same as in [29].

E. MOMENTUM RELAXATION TIME

Solving the Boltzmann transport equation (BTE) for anisotropic valleys is a complex problem. Therefore, in this study, we use an approximated solution of the linearized BTE under a weak electric field in isotropic valleys. The momentum relaxation time, $\tau_{sr,i}(\boldsymbol{k})$, in subband i is [14], [16]

$$\frac{1}{\tau_{sr,i}(k_i)} = \sum_j \frac{g_{v,j}}{2\pi\hbar} \int \left|M_{\mathrm{ij}}(\boldsymbol{k_i} - \boldsymbol{k_j})\right|^2 \times \left[1 - \frac{\tau_{sr,j}(k_j)v_{x,j}(k_j)}{\tau_{sr,i}(k_i)v_{x,i}(k_i)}\right] dk_j, \tag{32}$$

where $g_{v,j}$ is the degeneracy of subband j. Also, $v_{x,i}$ and $\boldsymbol{k_i}$ are the velocity of x direction and the wavenumber of subband i, respectively. In an isotropic valley, the momentum relaxation time at energy E can be obtained by solving the following equations [20], [29]

$$\tau_{sr}(E) = A^{-1}k(E), \tag{33}$$

$$\tau_{sr}(E) = \begin{bmatrix} \tau_{sr,1}(E) \\ \tau_{sr,2}(E) \\ \vdots \\ \tau_{sr,v}(E) \end{bmatrix}, \quad k(E) = \begin{bmatrix} k_1(E) \\ k_2(E) \\ \vdots \\ k_v(E) \end{bmatrix}, \tag{34}$$

where the matrix A is expressed by

$$\begin{aligned} A_{ii} = {} & \frac{g_{v,i}m_{d,i}}{2\pi\hbar^3}[1 + 2\alpha_i(E - E_{ci})] \\ & \times \int_0^{2\pi} \left|M_{ii}(\boldsymbol{k_i} - \boldsymbol{k_i'})\right|^2 (k_i - k_i\cos\theta)d\theta \\ & + \frac{1}{2\pi\hbar^3} \sum_{j\neq i} g_{vj}k_i m_{d,j}\left[1 + 2\alpha_j(E - E_{cj})\right] \\ & \times \int_0^{2\pi} \left|M_{ij}(\boldsymbol{k_i} - \boldsymbol{k_j'})\right|^2 d\theta, \end{aligned} \tag{35}$$

$$\begin{aligned} A_{ij} = {} & -\frac{g_{v,j}}{2\pi\hbar^3}k_j' m_{d,i}[1 + 2\alpha_i(E - E_{ci})] \\ & \times \int_0^{2\pi} \left|M_{ij}(\boldsymbol{k_i} - \boldsymbol{k_j'})\right|^2 \cos\theta d\theta \quad (i \neq j), \end{aligned} \tag{36}$$

$$\boldsymbol{k_i} = \begin{pmatrix} k_i \\ 0 \end{pmatrix}, \boldsymbol{k_j'} = \begin{pmatrix} k_j'\cos\theta \\ k_j'\sin\theta \end{pmatrix}. \tag{37}$$

Here, $\boldsymbol{k_i}$ and $\boldsymbol{k_j'}$ are the wavenumber of the initial state in subband i and the final state in subband j, respectively. $k_i = |\boldsymbol{k_i}|$. $m_{d,i}$ is the density-of-state effective mass of subband i. α_i and E_{ci} are the nonparabolicity factor and the ground energy of subband i, respectively.

F. PHONON SCATTERING

The contribution of phonon scattering is pronounced for a thick channel and low carrier density at room temperature [11], [30], [31]. Also, phonon scattering in (100) SOI cannot be ignored even in the 3-nm-thick channels [4], [25], [30]. The momentum relaxation time of intravalley scattering by acoustic phonon for electrons in subband i, $\tau_{ph,i}^{ac}$, is [11], [16], [32]

$$\frac{1}{\tau_{ph,i}^{ac}(E)} = \sum_j \frac{g_{v,i}m_{d,i}D_{ac}^2 k_B T}{\hbar^3\rho_m s_l V_{i,j}} U(E - E_{cj}), \tag{38}$$

$$V_{i,j} = \left(\int_{-\infty}^{\infty} |\xi_j(z)|^2 |\xi_i(z)|^2 dz\right)^{-1}, \tag{39}$$

where $U(E)$ is the step function, k_B is the Boltzmann constant, T is temperature, ρ_m is the mass density of the crystal, s_l is the longitudinal sound velocity, and D_{ac} is the deformation potential of acoustic phonon.

The momentum relaxation time of intravalley scattering by surface optical (SO) phonon for electrons in subband i, $\tau_{ph,i}^{sop}$, is [11], [33], [34]

$$\begin{aligned} \frac{1}{\tau_{ph,i}^{sop}(E)} = {} & \sum_{j,s} \frac{m_{d,i}}{\pi\hbar^3}\left(N_{SO,s} + \frac{1}{2} \mp \frac{1}{2}\right) \\ & \times \frac{\left(1 - f(E \pm E_{SO,s})\right)}{1 - f(E)} \times U(E - E_{cj} \pm E_{SO,s}) \\ & \times \int_0^{\pi} d\theta \left|M_{SO,s}^{(i,j)}(q)\right|^2 \\ & \times \left[1 - \frac{\tau_{ph,j}^{sop}(E \pm E_{SO,s})k_j(E \pm E_{SO,s})}{\tau_{ph,i}^{sop}(E)k_i(E)}\cos\theta\right], \end{aligned} \tag{40}$$

where $s = 1, 2$ denotes one of the two SO phonon modes. $N_{SO,s}$ and $E_{SO,s}$ are the number and the energy of SO phonon of mode s, respectively. $f(E)$ is the Fermi-Dirac distribution function. $M_{SO,s}^{(i,j)}(q)$ is the matrix element in [11], where $q = |\boldsymbol{k_i}(E) - \boldsymbol{k_j}(E \pm E_{SO,s})|$. $\tau_{ph,j}^{sop}(E \pm E_{SO,s})$ in (40) is approximated to be equal to $\tau_{ph,i}^{sop}(E)$ to simplify the calculation.

The momentum relaxation time of intervalley phonon scattering for electrons in subband i, $\tau_{ph,i}^{inter}$, is [32]

$$\begin{aligned} \frac{1}{\tau_{ph,i}^{inter}(E)} = {} & \sum_{j,k} \frac{g_{v,j}m_{d,j}D_k^2}{\hbar\rho_m E_k V_{i,j}}\left(N_k + \frac{1}{2} \mp \frac{1}{2}\right) \\ & \times \frac{(1 - f(E \pm E_k))}{1 - f(E)} \times U(E - E_{cj} \pm E_k), \end{aligned} \tag{41}$$

where, D_k, E_k, and N_k are the deformation potential, the energy, and the number of the k-th intervalley phonon, respectively. In this study, scattering between the two Δ_2 valleys is considered as the intervalley scattering by g-phonon.

TABLE 1. Parameters of (100) Si/SiO_2 interface used in the simulation.

Name	Symbol	Value	Note
Effective mass (unit: m_0 = 9.11×10^{-31} kg)	m_t	0.19	Transverse
	m_l	0.916	Longitudinal
	m_{ox}	0.3	SiO_2
Band offset	V_{ox}	3.3 eV	
Acceptor concentration	N_A	2×10^{16} cm^{-3}	
Nonparabolicity	α	0.27 eV^{-1}	
Mass density	ρ	2329 kg/m^3	
Sound velocity	s_l	9037 m/s	
Deformation potential	D_{ac}	9.5 eV	Intravalley
	D_k (f)	7×10^8 eV/cm	f-phonon
	D_k (g)	9.5×10^8 eV/cm	g-phonon
Phonon energy	E_k (f)	59 meV	f-phonon
	E_k (g)	63 meV	g-phonon
	$E_{TO,1}$	57.2 meV	SO-phonon
	$E_{TO,2}$	140.7 meV	SO-phonon

G. MOBILITY

The total momentum relaxation time, τ_i, and the mobility, $\mu_{total,i}$, for electrons in subband i are [20]

$$\frac{1}{\tau_i(E)} = \frac{1}{\tau_{sr,i}(E)} + \frac{1}{\tau^{ac}_{ph,i}(E)} + \frac{1}{\tau^{sop}_{ph,i}(E)} + \frac{1}{\tau^{inter}_{ph,i}(E)}, \quad (42)$$

$$\mu_{total,i} = \frac{eg_{v,i}}{N_{s,i}\pi\hbar^2}\left(\frac{m_{d,i}}{m_{x,i}}\right)^2 \int_{-\infty}^{\infty} \tau_i(E)\left|\frac{\partial f(E)}{\partial E}\right| \times \frac{(E-E_{ci})\times[1+\alpha_i(E-E_{ci})]}{1+2\alpha_i(E-E_{ci})}dE, \quad (43)$$

where $m_{x,i}$ and $N_{s,i}$ are the conduction effective mass and the areal density of electrons in subband i, respectively. The total mobility is

$$\mu_{total} = \frac{\sum_i N_{s,i}\mu_{total,i}}{\sum_i N_{s,i}}. \quad (44)$$

The parameters used for the mobility calculation are summarized in Table 1 [11], [32], [35]. Here, the tunneling effective mass is used as m_{ox} of SiO_2 [35].

III. SIMULATION RESULTS

A. MOBILITY OF BULK SI nMOSFETs

The influence of the subband number included in the mobility calculation is shown in Fig. 6. It is found that when more than 3-subbands are included, the mobility becomes saturated. Therefore, 5-subbands are considered in each valley for all the calculations in this study. The $E_{\rm eff}$ dependence of SR-limited mobility, μ_{sr}, is shown in Fig. 7. The experimental mobility at T = 300, 77, and 4.2 K is also shown in Figs. 7(a), (b) and (c), respectively [36]. Here, the experimental mobility at 4.2 K in Fig. 7(c) is measured in this study for the identical samples. Although the $E_{\rm eff}$ universality of μ_{sr} for different acceptor concentrations (N_A) is not simply supported by the nonlinear model because of its complicated formulation, the present simulation results show the good $E_{\rm eff}$ universality of μ_{sr} as shown in Fig. 7(a). The

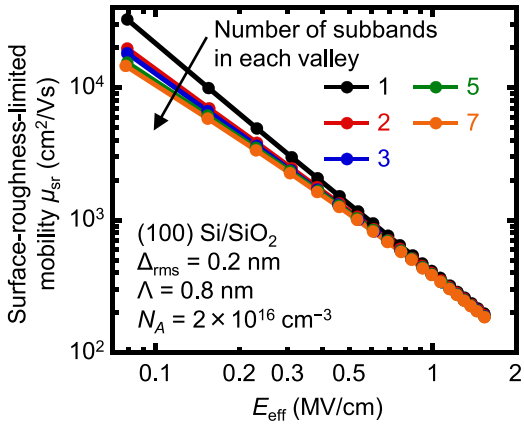

FIGURE 6. SR-limited mobility in the bulk (100) Si/SiO_2 nMOSFET as a parameter of the number of subbands included in the simulation. Δ_{rms} of 0.2 nm and Λ of 0.8 nm are used.

constant mobility of $\mu_{ni} = 2\times10^4$ cm^2/Vs limited by neutral impurity scattering is included to fit with the experimental mobility at 4.2 K in Fig. 7(c). It should be noted that only the Δ_2 valley is considered in Fig. 7(c) to eliminate the effect of the unreasonably large mobility of the Δ_4 valley electrons due to the screening effect as discussed later in Fig. 8. The total mobility is in excellent agreement with the experimental one at 77 and 4.2 K by using the same roughness parameters of $\Delta_{\rm rms}$ = 0.2 nm and Λ = 0.8 nm. On the other hand, the Prange-Nee model needs to represent the experimental results by using $\Delta_{\rm rms}$ = 0.5 nm and Λ = 0.8 nm. The roughness parameters used in the simulation and evaluated by TEM are summarized in Table 2. Here, the exponential formulation in (9) and (16) is used in this study because the roughness spectra measured by TEM are closer to the exponential ones [18], [25]. The roughness parameters in our model are certainly in good agreement with the TEM analysis [17], [18], [19]. On the other hand, it should be pointed out that $\Delta_{\rm rms}$ evaluated by TEM might not be too accurate because of the projection effect [17], [18], [19] and the existence of discrete atomic images [25]. TEM images are obtained as averaged ones along the transmission direction of electrons and, thus, the $\Delta_{\rm rms}$ value by TEM is apparent and smaller than the real value, which is called the projection effect. In addition, there is a lower limit on $\Delta_{\rm rms}$ evaluated by TEM due to the existence of discrete atomic images. The lower limit of measurable $\Delta_{\rm rms}$ is roughly $a/4$ $\sim$ 0.14 nm, where a is a lattice constant. These two effects make it difficult to evaluate the real small $\Delta_{\rm rms}$ by TEM. However, we would believe that the present $\Delta_{\rm rms}$ of 0.2 nm is very realistic because the $\Delta_{\rm rms}$ of initial substrates evaluated by atomic force microscopy is typically less than 0.2 nm [37], [38], and the thermal oxidation process might not degrade the interface roughness so much [7], [39].

However, it is still questionable the validity of our mobility model at cryogenic temperature. The calculated SR-limited mobility at 300 and 4.2 K with and without the screening effect is shown in Fig. 8(a). Also, the scalar dielectric

TABLE 2. Roughness parameters of (100) Si/SiO_2.

Method	Δ_{rms} (nm)	Λ (nm)	Formulation method			Reference
			$C_\Delta(r)$	$f_1(\Delta)$	$f_2(\Delta_1,\Delta_2,\rho)$	
TEM	0.14–0.18	0.7–2.2	Exponential	-	-	[17]
	0.14–0.18	0.6–1.5	Gaussian	-	-	[17]
	0.2	0.9	Exponential	-	-	[18], [19]
	0.2	1.5	Gaussian	-	-	[18], [19]
Simulation by Prange-Nee (linear) model	0.51	1.0	Gaussian	-	-	[11]
	0.6	2.12	Exponential	-	-	[12]
	0.55	1.3	Gaussian	-	-	[13]
	0.47	1.3	Exponential	-	-	[14]
	0.65	1.4	Exponential	-	-	[20]
	0.5	0.8	Exponential	-	-	This work
Simulation by nonlinear model	0.21	1.4	Exponential	-	Gaussian	[20]
Our revised nonlinear model	0.2	0.8	Exponential	Exponential	Exponential	This work

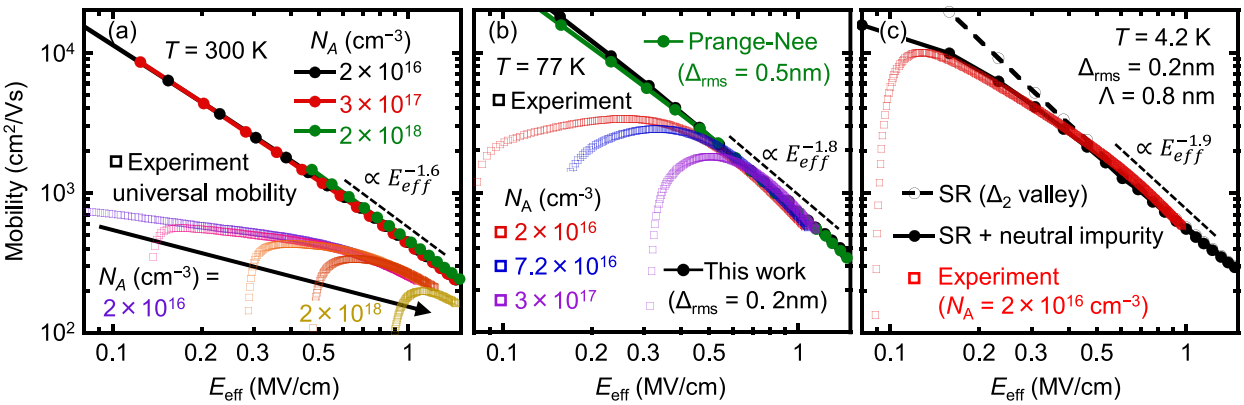

FIGURE 7. Effective electric field dependence of the SR-limited mobility for the bulk (100) Si/SiO_2 nMOSFETs at (a) 300 K, (b) 77 K, and (c) 4.2 K. The experimental mobility was extracted by using the identical samples for the same N_A. (a) The N_A dependence of μ_{sr} is shown. (b) Δ_{rms} of 0.5 nm and 0.2 nm are used in the simulation of the Prange-Nee and our models, respectively. (c) Only the Δ_2 valley is considered. The constant mobility of 2 × 10^4 cm^2/Vs limited by neutral impurity scattering is included to fit with the experimental mobility. Here, Λ of 0.8 nm is used for all the calculations.

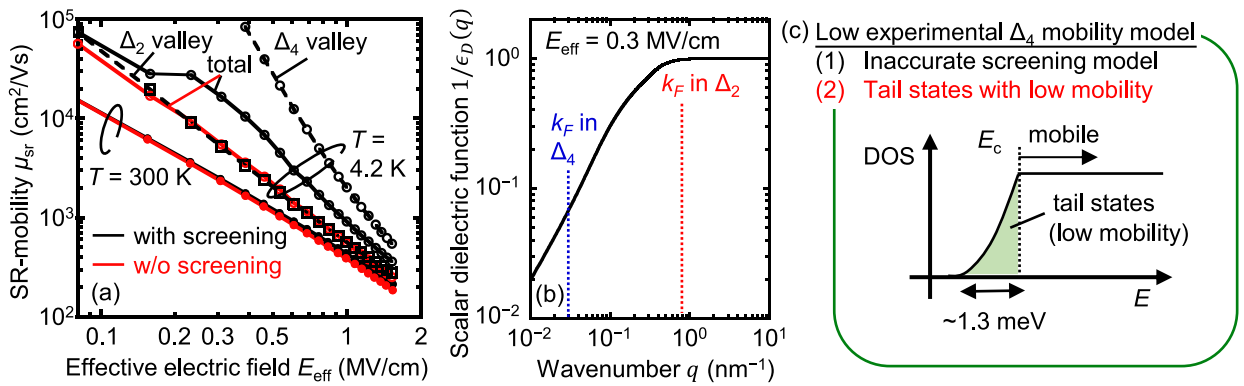

FIGURE 8. (a) Simulated SR-limited mobility at 300 and 4.2 K with and without the screening effect. (b) Scalar dielectric function $1/\epsilon_D(q)$ under E_{eff} of 0.3 MV/cm at 4.2 K. The Fermi wavenumber k_F in the Δ_2 and Δ_4 valleys is also shown. (c) Possible interpretations of the low experimental mobility of the Δ_4 valley electrons at 4.2 K. The energy width of the tail states has been experimentally estimated as 1.3 meV for the Δ_2 valley [40].

function $\epsilon_D(q)$, which is the screening term in (30), is shown in Fig. 8(b). The screening effect is negligible at 300 K and for the Δ_2 valley electrons at 4.2 K because $\epsilon_D(q) \cong 1$ for the large wavenumber q. On the other hand, the mobility of the Δ_4 valley electrons is extremely large at 4.2 K due to the strong screening effect, which is far from the experimental mobility in Fig. 7(c). There are two possible interpretations for the low experimental mobility of the Δ_4 valley electrons

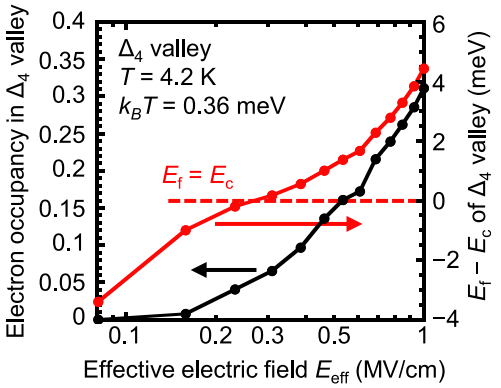

FIGURE 9. Effective electric field dependence of the electron occupancy and the Fermi level ($E_f - E_c$) in the Δ_4 valley at 4.2 K.

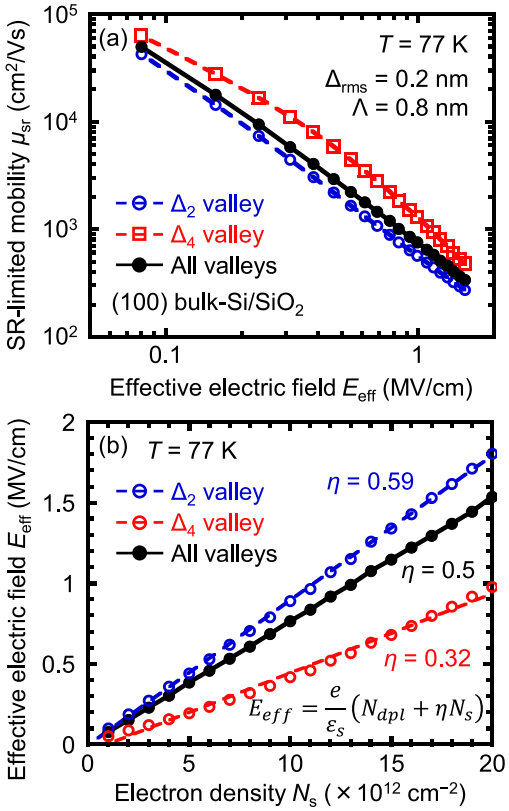

FIGURE 10. (a) Effective electric field dependence of the SR-limited mobility and (b) the N_s dependence of E_{eff} for each valley of the bulk (100) Si/SiO_2 nMOSFET at 77 K.

as shown in Fig. 8(c). First, the model of the screening effect on surface roughness scattering is inaccurate. In the current screening model, the potential change is induced by the change of the electron distribution functions due to the perturbations. On the other hand, such a potential change by the screening effect also makes the ground state Hamiltonian $\langle H_\Delta \rangle_r$ closer to H_0, resulting in the enhancement of the amount of the matrix element, as shown in Fig. 3. This is because the unscreened wavefunction, which is the eigenstate of $\langle H_\Delta \rangle_r$, moves away from the interface to decrease the scattering by surface roughness. However, the establishment of more accurate screening modeling of the ground state is still a challenge.

The second interpretation, which is more likely, is the significant decrease in the mobility due to the occupation of the electrons in the tail states with low mobility. The electron occupancy and the Fermi level $E_f - E_c$ in the Δ_4 valley are shown in Fig. 9. The Fermi energy of the Δ_4 valley is less than 1 meV even for $E_{\text{eff}} < 0.5$ MV/cm, while the energy width of the tail states has been estimated to be $\sim$ 1.3 meV [40]. Therefore, all the electrons can be occupied in the tail states at 4.2 K, where the mobility should be low because the tail states can originate from the Lifshitz tail [41] and are different from electronics sates in a standard extended band. The energy width of 1.3 meV [40] is estimated for the Δ_2 valley, whereas the similar width can be also expected for the Δ_4 valley.

The SR-limited mobility of electrons in each valley at 77 K is shown in Fig. 10(a). μ_{sr} of the Δ_4 valley electrons is higher than μ_{sr} of the Δ_2 valley electrons, even though the conduction mass m_x is $0.19m_0$ and $0.315m_0$ for the Δ_2 and Δ_4 valleys, respectively, where m_0 is the free electron mass. Here, the temperature is taken to be 77 K to simplify the discussion by suppressing the electron occupancy of higher subbands. The high μ_{sr} in the Δ_4 valley is attributed to the two screening mechanisms. First, the screening effect in (30) is strong for the Δ_4 valley electrons because the Fermi wavenumber is very low. Second, the z-direction electric field is also screened by the Δ_2 valley electrons. The effective electric field of each valley is shown in Fig. 10(b). The effective electric field at each subband i, $E_{eff,i}$, and the total effective electric field, E_{eff}, are determined by

$$E_{eff,i} = \int_{-\infty}^{\infty} \mathcal{E}(z)|\xi_i(z)|^2 dz, \tag{45}$$

$$E_{\text{eff}} = \left[\sum_i N_{s,i} E_{eff,i}\right] / N_s, \tag{46}$$

where $\mathcal{E}(z)$ is the electric field. E_{eff} in the Δ_2 valley is much higher than E_{eff} in the Δ_4 valley because the spread width of the wavefunction of Δ_2 valley electrons is narrower due to the heavy m_{sz} and the electron occupancy is higher due to the lower quantization energy. Therefore, μ_{sr} in the Δ_4 valley can be higher than μ_{sr} in the Δ_2 valley at low temperature because of lower E_{eff} for electrons in the Δ_4 valley. Experimentally, the total effective electric field can be given by N_s and the coefficient η as follows.

$$E_{\text{eff}} = \frac{e}{\varepsilon_s}\left(N_{dpl} + \eta N_s\right). \tag{47}$$

where eN_{dpl} is the charge density of the depletion layer. The coefficient η, determined by the slope of E_{eff}, amounts to 0.59, 0.32, and 0.5 for the Δ_2, Δ_4, and total valleys, respectively, whereas the experimental η of the total valley is known to be 0.5 for (100) Si [36]. Also, η for the Δ_4 valley is smaller in the low N_s region than 0.32. It should be noted that η of the total valley defined by (46) is always

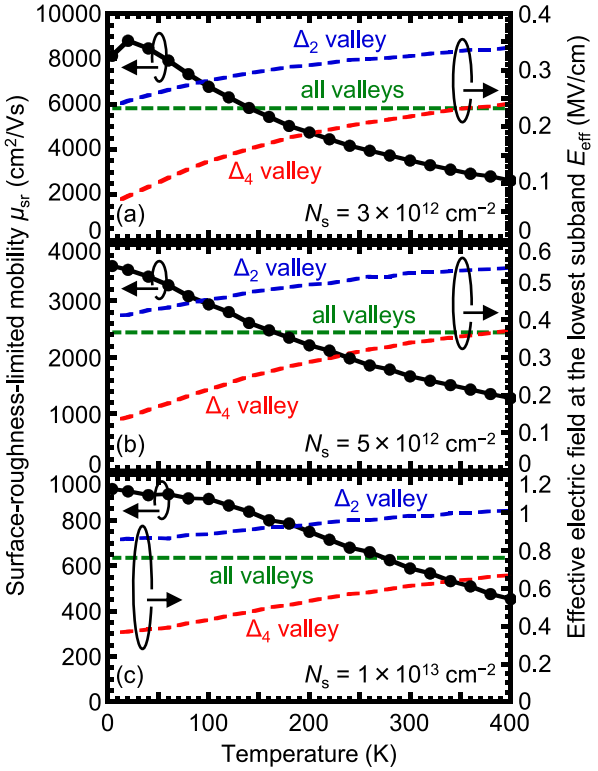

FIGURE 11. Temperature dependence of the SR-limited mobility in all the valleys and $E_{eff,i}$ in the lowest subband for N_s of (a) 3×10^{12}, (b) 5×10^{12}, and (c) 1×10^{13} cm^{-2}. Here, the screening effect is ignored.

0.5 in theory [16]. Therefore, in the bulk nMOSFETs with the multi-valley electron occupation, the existence of the valley with heavy m_x does not simply result in mobility reduction. On the other hand, in ETB channels, μ_{sr} in valleys with light m_{sz} is severely degraded as discussed in the next Section III-B.

The temperature dependence of μ_{sr} of electrons in all the valleys and $E_{eff,i}$ of electrons in the lowest subband is shown in Fig. 11. Here, the screening effect is temporary ignored because of the too strong influence on electrons in the Δ_4 valley at cryogenic temperature. However, it is not a bad approximation, as shown in Fig. 8(a). While μ_{sr} is independent of temperature in the single-valley and single-subband conditions (data is not shown), μ_{sr} for the multi-valley structures is actually dependent on temperature. As temperature decreases, the electron occupancy in the subbands with more widely-spread wavefunctions decreases because of the larger quantization energy, leading to the reduction in $E_{eff,i}$ in the lowest subband of the Δ_2 valley. As a result, the total mobility increases with a decrease in temperature. Since the screening effect becomes also stronger at lower temperatures, the tendency for μ_{sr} to increase at low temperature still remains unchanged.

B. MOBILITY OF ETB SOI nMOSFETs

The SOI thickness dependence of the phonon-limited, the SR-limited, and the total mobility under $N_s = 3 \times 10^{12}$ cm^{-2} is shown in Fig. 12. The experimental mobility [4] is well

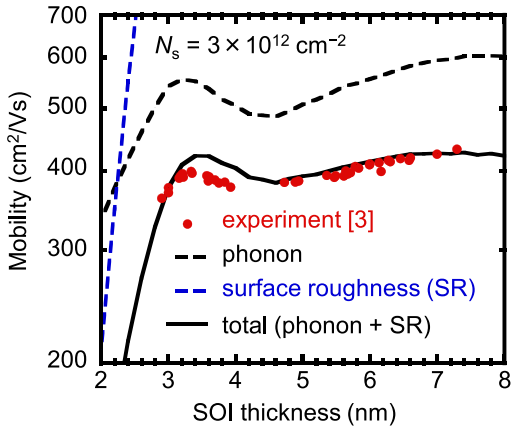

FIGURE 12. SOI thickness dependence of the phonon-limited, SR-limited, and total mobility under Δ_{rms} of 0.2 nm, Λ of 0.8 nm, and N_s of 3×10^{12} cm^{-2}.

explained by the calculated values. Here, the validity of our model is also supported by the fact that the same roughness parameters are used in Figs. 7 and 10. The SOI thickness dependence of μ_{sr} under $N_s = 3 \times 10^{12}$ cm^{-2} is shown for various $\Delta_{\rm rms}$ in Fig. 13(a). The general trend is similar among various $\Delta_{\rm rms}$. For the channel thickness of less than 4 nm, which is narrower than the spread width of the wavefunction of electrons in the Δ_2 valley, as shown in Fig. 2(a), μ_{sr} rapidly drops in proportion to T_B^r with $r = 5 \sim 6$. This strong T_B dependence in ETB channels is well known as thickness fluctuation scattering expressed by [42]

$$\mu_{sr} \propto \frac{m_{sz}^2}{m_x \Delta_{\rm rms}^2 \Lambda^2 E_{\rm eff}^2} T_B^6. \tag{48}$$

In the linear model with the infinite energy barrier, μ_{sr} is proportional to T_B^6. On the other hand, in our model with a finite energy barrier, the factor $r = d(\ln \mu_{sr})/d(\ln T_B)$ is lower than 6, as shown in Fig. 13(b). The $\Delta_{\rm rms}$ and Λ dependence of μ_{sr} for bulk (100) Si/SiO$_2$ nMOSFETs under $N_s = 5 \times 10^{12}$ cm^{-2} is shown in Figs. 14(a) and (b), respectively. Here, μ_{sr} is proportional to Λ^{-2} in our model, as similar to that in the linear model in (45). On the other hand, μ_{sr} is proportional to $\Delta_{\rm rms}^{-2.7}$ and $\Delta_{\rm rms}^{-2}$ in our model and the Prange-Nee model, respectively. As a result, surface roughness scattering has a 13 times stronger effect in our model than that predicted in the conventional Prange-Nee model. Therefore, there is little room for improvement of μ_{sr} in Si nMOSFETs because the $\Delta_{\rm rms}$ value of 0.2 nm at Si MOS interfaces is sufficiently small. The $E_{\rm eff}$ dependence of μ_{sr} for the 2–6-nm-thick SOI is shown in Fig. 15. In the lower $E_{\rm eff}$ region, μ_{sr} drops more significantly with channel thickness scaling, attributed to the wider spread width of the wavefunction with lower $E_{\rm eff}$. μ_{sr} increases very slightly with channel thickness scaling in the high $E_{\rm eff}$ region because of the reduction of the electron occupancy in the Δ_4 valley.

The comparison of μ_{sr} between the single-gate and double-gate structures under $N_s = 3 \times 10^{12}$ cm^{-2} is shown in Fig. 16. The mobility is slightly lower for the double-gate structure because the electron occupancy in the Δ_2 valley

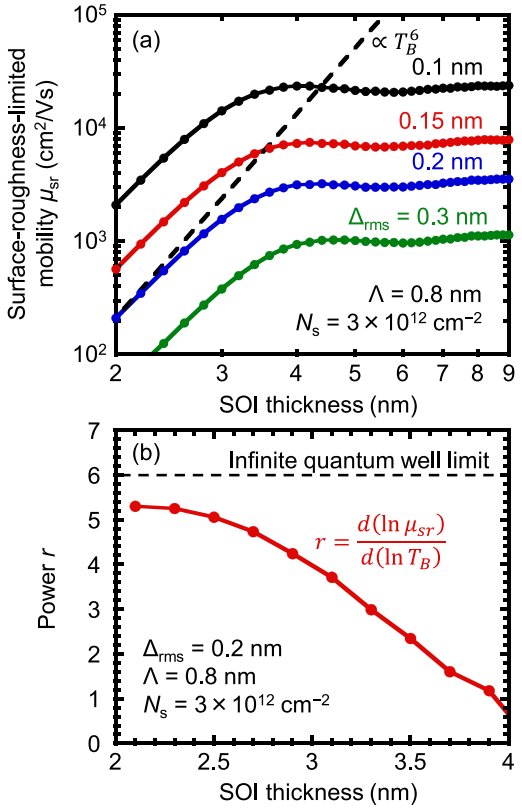

FIGURE 13. SOI thickness dependence of (a) SR-limited mobility for various Δ_{rms} and (b) $r = d(\ln \mu_{sr})/d(\ln T_B)$ under N_s of 3 × 10^{12} cm^{-2} and Λ of 0.8 nm.

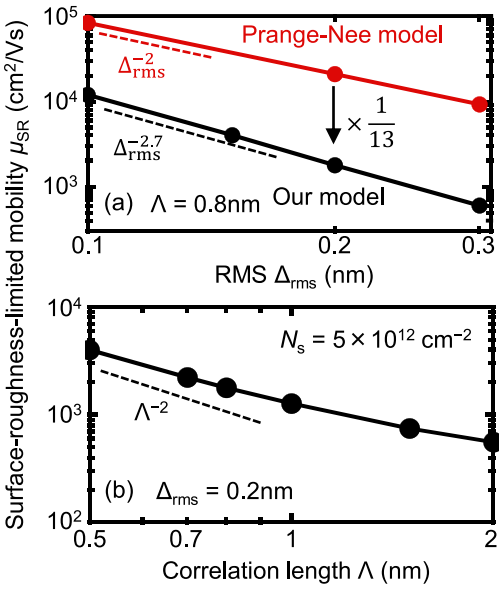

FIGURE 14. (a) Δ_{rms} and (b) Λ dependence of the SR-limited mobility for the bulk (100) Si/SiO_2 nMOSFETs under N_s of 5 × 10^{12} cm^{-2}.

with light m_x is lower. On the other hand, the mobility is almost the same between the single-gate and double-gate structures in the ETB channels less than 3 nm because the electron occupancy in the Δ_2 and Δ_4 valleys is determined by the channel thickness. It should be noted that the

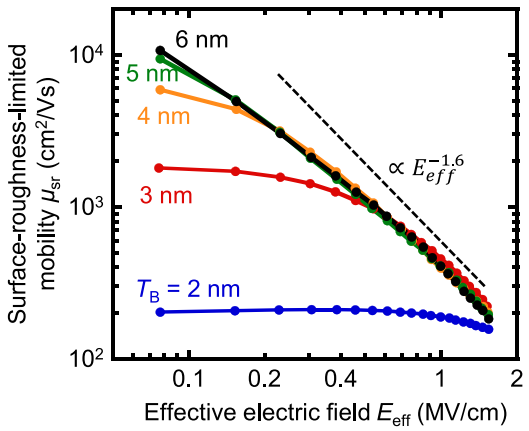

FIGURE 15. Effective electric field dependence of the SR-limited mobility for 2–6-nm-thick SOI nMOSFETs.

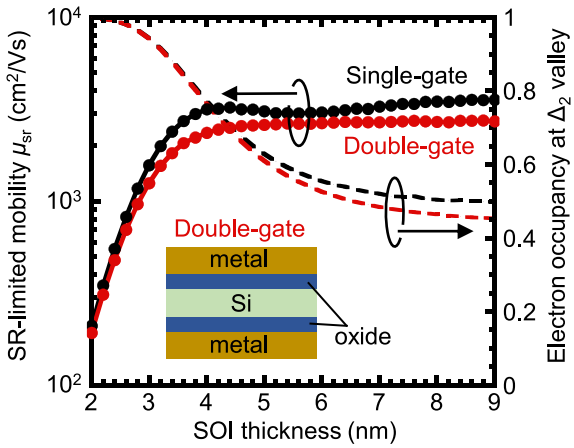

FIGURE 16. SR-limited mobility and electron occupancy in the Δ_2 valley for (100) SOI nMOSFETs with the single-gate and double-gate structures under Δ_{rms} of 0.2 nm, Λ of 0.8 nm and N_s of 3 × 10^{12} cm^{-2}.

formulation of the screening effect by the scalar dielectric function in (30) is not accurate for the double-gate structure, and the tensorial dielectric function should be used [22], [43]. However, the scalar dielectric function is used in this study to reduce the calculation cost. Fig. 17 shows μ_{sr} of Δ_2 and Δ_4 valley electrons. The channel thickness at which μ_{sr} starts to drop is thicker for the Δ_4 valley than the Δ_2 valley because of the wider wavefunction of Δ_4 valley electrons. Therefore, the utilization of the anisotropic valley with the combination of light m_x and heavy m_{sz} is important to suppress the influence of surface roughness scattering in ETB channels, as seen in (48). It should be recalled here that there is little room for improvement of μ_{sr} in Si nMOSFETs because the Δ_{rms} value of 0.2 nm at Si MOS interfaces is sufficiently small. Therefore, the introduction of other materials with the anisotropic valley is very important for the future CMOS technology nodes.

IV. CONCLUSION

Since surface roughness scattering is a strongly nonlinear phenomenon, higher-order perturbations must be taken into account to quantitatively predict mobility. We have proposed

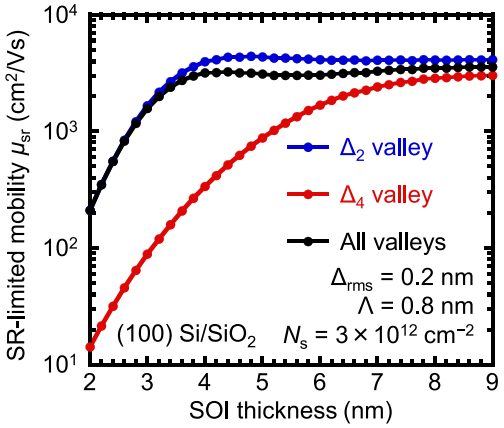

FIGURE 17. SR-limited mobility of the Δ_2, Δ_4 and all valleys for (100) SOI nMOSFETs under Δ_{rms} of 0.2 nm, Λ of 0.8 nm and N_s of 3 × 10^{12} cm^{-2}.

the new surface roughness scattering model on a basis of the re-consideration of the ground states for 2DEG with rough MOS surfaces and a combination with the nonlinear surface roughness scattering probability. This revised nonlinear model of surface roughness scattering can represent the experimental mobilities well by using the realistic roughness parameters obtained from the TEM analysis. It has been revealed that surface roughness scattering has a much stronger influence than that predicted by the Prange-Nee model. As a result, there is little room to improve μ_{sr} in Si MOSFETs because the Δ_{rms} value of 0.2 nm at Si MOS interfaces is sufficiently small. Thus, the introduction of high mobility materials with anisotropic valleys is much important. Then, the quantitative nonlinear model in this study allows us to predict mobility in ETB nanosheet channels and to provide an assessment for future channel materials in the advanced CMOS technology.

REFERENCES

[1] J. P. Colinge, "Multi-gate SOI MOSFETs," *Microelectron. Eng.*, vol. 84, nos. 9–10, pp. 2071–2076, Sep./Oct. 2007, doi: 10.1016/j.mee.2007.04.038.

[2] S. Barraud et al., "7-levels-stacked nanosheet GAA transistors for high performance computing," in *Proc. IEEE Symp. VLSI Technol.*, Jun. 2020, pp. 1–2, doi: 10.1109/VLSITechnology18217.2020.9265025.

[3] C.-E. Tsai et al., "Highly stacked 8 $Ge_{0.9}Sn_{0.1}$ nanosheet pFETs with ultrathin bodies (~3nm) and thick bodies (~30nm) featuring the respective record I_{ON}/I_{OFF} of 1.4x10^7 and record I_{ON} of 92μA at V_{OV}=V_{DS}= -0.5V by CVD epitaxy and dry etching," in *IEDM Tech. Dig.*, Dec. 2021, pp. 1–4, doi: 10.1109/IEDM19574.2021.9720660.

[4] K. Uchida, H. Watanabe, A. Kinoshita, J. Koga, T. Numata, and S. Takagi, "Experimental study on carrier transport mechanism in ultrathin-body SOI nand p-MOSFETs with SOI thickness less than 5 nm," in *IEDM Tech. Dig.*, Dec. 2002, pp. 47–50, doi: 10.1109/IEDM.2002.1175776.

[5] M. Schmidt, M. C. Lemme, H. D. B. Gottlob, F. Driussi, L. Selmi, and H. Kurz, "Mobility extraction in SOI MOSFETs with sub 1 nm body thickness," *Solid-State Electron.*, vol. 53, no. 12, pp. 1246–1251, Dec. 2009, doi: 10.1016/j.sse.2009.09.017.

[6] K.-W. Jo, C.-M. Lim, W.-K. Kim, K. Toprasertpong, M. Takenaka, and S. Takagi, "Strain and surface orientation engineering in extremely-thin body Ge and SiGe-on-insulator MOSFETs fabricated by Ge condensation," in *IEDM Tech. Dig.*, Dec. 2019, pp. 673–676, doi: 10.1109/IEDM19573.2019.8993595.

[7] K. Sumita, K. Toprasertpong, M. Takenaka, and S. Takagi, "Subband engineering by combination of channel thickness scaling and (111) surface orientation in InAs-On-insulator nMOSFETs," in *IEDM Tech. Dig.*, Dec. 2020, pp. 27–30, doi: 10.1109/IEDM13553.2020.9371995.

[8] K. Sumita, K. Toprasertpong, M. Takenaka, and S. Takagi, "Proposal and experimental demonstration of ultrathin-body (111) InAs-on-insulator nMOSFETs with L valley conduction," *IEEE Trans. Electron Devices.*, vol. 68, no. 4, pp. 2003–2009, Apr. 2021, doi: 10.1109/TED.2021.3049455.

[9] B. Patra et al., "Cryo-CMOS circuits and systems for quantum computing applications," *IEEE J. Solid-State Circuits*, vol. 53, no. 1, pp. 309–321, Jan. 2018, doi: 10.1109/JSSC.2017.2737549.

[10] T. Ando, A. B. Fowler, and F. Stern, "Electronic properties of two-dimensional systems," *Rev. Mod. Phys.*, vol. 54, no. 2, pp. 437–672, Apr. 1982, doi: 10.1103/RevModPhys.54.437.

[11] D. Esseni, A. Abramo, L. Selmi, and E. Sangiorgi, "Physically based modeling of low field electron mobility in ultrathin single- and double-gate SOI n-MOSFETs," *IEEE Trans. Electron Devices*, vol. 50, no. 12, pp. 2445–2455, Dec. 2003, doi: 10.1109/TED.2003.819256.

[12] T. Low et al., "Modeling study of the impact of surface roughness on silicon and germanium UTB MOSFETs," *IEEE Trans. Electron Devices*, vol. 52, no. 11, pp. 2430–2439, Nov. 2005, doi: 10.1109/TED.2005.857188.

[13] T. Ishihara, K. Uchida, J. Koga, and S.-I. Takagi, "Unified roughness scattering model incorporating scattering component induced by thickness fluctuations in silicon-on-insulator metal–oxide–semiconductor field-effect transistors," *Jpn. J. Appl. Phys.*, vol. 45, no. 4B, pp. 3125–3132, Apr. 2006, doi: 10.1143/JJAP.45.3125.

[14] S. Jin, M. V. Fischetti, and T.-W. Tang, "Modeling of surface-roughness scattering in ultrathin-body SOI MOSFETs," *IEEE Trans. Electron Devices*, vol. 54, no. 9, pp. 2191–2203, Sep. 2007, doi: 10.1109/TED.2007.902712.

[15] S. Jin, M. V. Fischetti, and T.-W. Tang, "Modeling of electron mobility in gated silicon nanowires at room temperature: Surface roughness scattering, dielectric screening, and band nonparabolicity," *J. Appl. Phys.*, vol. 102, no. 8, Oct. 2007, Art. no. 83715, doi: 10.1063/1.2802586.

[16] D. Esseni, P. Palestri, and L. Selmi, *Nanoscale MOS Transistors: Semi-Classical Transport and Applications*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[17] S. M. Goodnick, D. K. Ferry, C. W. Wilmsen, Z. Liliental, D. Fathy, and O. L. Krivanek, "Surface roughness at the Si(100)-SiO_2 interface," *Phys. Rev. B, Condens. Matter*, vol. 32, no. 12, pp. 8171–8186, Dec. 1986, doi: 10.1103/PhysRevB.32.8171.

[18] Y. Zhao, H. Matsumoto, T. Sato, S. Koyama, M. Takenaka, and S. Takagi, "A novel characterization scheme of Si/SiO_2 interface roughness for surface roughness scattering-limited mobilities of electrons and holes in unstrained- and strained-Si MOSFETs," *IEEE Trans. Electron Devices*, vol. 57, no. 9, pp. 2057–2066, Sep. 2010, doi: 10.1109/TED.2010.2052394.

[19] Y. Zhao, H. Matsumoto, T. Sato, S. Koyama, M. Takenaka, and S. Takagi, "Effect of sample thickness on SiO_2/Si interface roughness characterization through transmission electron microscope measurements in strained-Si MOSFETs," *J. Electrochem. Soc.*, vol. 159, no. 1, pp. H57–H60, Dec. 2011, doi: 10.1149/2.055201jes.

[20] D. Lizzit, D. Esseni, P. Palestri, and L. Selmi, "A new formulation for surface roughness limited mobility in bulk and ultra-thin-body metal–oxide–semiconductor transistors," *J. Appl. Phys.*, vol. 116, no. 22, Dec. 2014, Art. no. 223702, doi: 10.1063/1.4903768.

[21] D. Lizzit, O. Badami, R. Specogna, and D. Esseni, "Improved surface-roughness scattering and mobility models for multi-gate FETs with arbitrary cross-section and biasing scheme," *J. Appl. Phys.*, vol. 121, no. 24, Jun. 2017, Art. no. 245301, doi: 10.1063/1.4986644.

[22] O. Badami et al., "An improved surface roughness scattering model for bulk, thin-body, and quantum-well MOSFETs," *IEEE Trans. Electron Devices*, vol. 63, no. 6, pp. 2306–2312, Jun. 2016, doi: 10.1109/TED.2016.2554613.

[23] O. Badami, D. Lizzit, R. Specogna, and D. Esseni, "Surface roughness limited mobility in multi-gate FETs with arbitrary cross-section," in *IEDM Tech. Dig.*, Dec. 2017, pp. 1–4, doi: 10.1109/IEDM.2016.7838551.

[24] M. H. Evans, X.-G. Zhang, J. D. Joannopoulos, and S. T. Pantelides, "First-principles mobility calculations and atomic-scale interface roughness in nanoscale structures," *Phys. Rev. Lett.*, vol. 95, no. 10, Sep. 2005, Art. no. 106802, doi: 10.1103/PhysRevLett.95.106802.

[25] K. Sumita, C.-T. Chen, K. Toprasertpong, M. Takenaka, and S. Takagi, "Optimum design of channel material and surface orientation for extremely-thin-body nMOSFETs under new modeling of surface roughness scattering," in *IEDM Tech. Dig.*, Dec. 2021, pp. 406–409, doi: 10.1109/IEDM19574.2021.9720715.

[26] S. Yue, T. B. M. J. Ouarda, and B. Bobée, "A review of bivariate gamma distributions for hydrological application," *J. Hydrol.*, vol. 246, nos. 1–4, pp. 1–18, Jun. 2001, doi: 10.1016/S0022-1694(01)00374-2.

[27] J. Duan, J. Xiang, L. Zhou, X. Wang, X. Ma, and W. Wang, "Electron mobility in silicon nanowires using nonlinear surface roughness scattering model," *Jpn. J. Appl. Phys.*, vol. 59, no. 3, Feb. 2020, Art. no. 34002, doi: 10.35848/1347-4065/ab7722.

[28] X. Wang et al., "Physically based evaluation of effect of buried oxide on surface roughness scattering limited hole mobility in ultra-thin GeOI MOSFETs," *IEEE Trans. Electron Devices*, vol. 64, no. 6, pp. 2611–2616, Jun. 2017, doi: 10.1109/TED.2017.2688489.

[29] D. Esseni and A. Abramo, "Modeling of electron mobility degradation by remote Coulomb scattering in ultrathin oxide MOSFETs," *IEEE Trans. Electron Devices*, vol. 50, no. 7, pp. 1665–1674, Jul. 2003, doi: 10.1109/TED.2003.814973.

[30] S.-I. Takagi, J. Koga, and A. Toriumi, "Mobility enhancement of SOI MOSFETs due to subband modulation in ultrathin SOI films," *Jpn. J. Appl. Phys.*, vol. 37, no. 3S, pp. 1289–1294, Mar. 1998, doi: 10.1143/JJAP.37.1289.

[31] S. Reggiani, E. Gnani, A. Gnudi, M. Rudan, and G. Baccarani, "Low-field electron mobility model for ultrathin-body SOI and double-gate MOSFETs with extremely small silicon thicknesses," *IEEE Trans. Electron Devices*, vol. 54, no. 9, pp. 2204–2212, Sep. 2007, doi: 10.1109/TED.2007.902899.

[32] S. Takagi, J. L. Hoyt, J. J. Welser, and J. F. Gibbons, "Comparative study of phonon-limited mobility of two-dimensional electrons in strained and unstrained Si metal–oxide–semiconductor field-effect transistors," *J. Appl. Phys.*, vol. 80, no. 3, pp. 1567–1577, Apr. 1996, doi: 10.1063/1.362953.

[33] M. V. Fischetti, D. A. Neumayer, and E. A. Cartier, "Effective electron mobility in Si inversion layers in metal–oxide–semiconductor systems with a high-κ insulator: The role of remote phonon scattering," *J. Appl. Phys.*, vol. 90, no. 9, pp. 4587–4608, Nov. 2001, doi: 10.1063/1.1405826.

[34] M. V. Fischetti and S. E. Laux, "Monte Carlo study of electron transport in silicon inversion layers," *Phys. Rev. B, Condens. Matter*, vol. 48, no. 4, pp. 2244–2274, Jul. 1993, doi: 10.1103/PhysRevB.48.2244.

[35] B. Brar, G. D. Wilk, and A. C. Seabaugh, "Direct extraction of the electron tunneling effective mass in ultrathin SiO_2," *Appl. Phys. Lett.*, vol. 69, no. 18, pp. 2728–2730, Aug. 1996, doi: 10.1063/1.117692.

[36] S. Takagi, A. Toriumi, M. Iwase, and H. Tango, "On the universality of inversion layer mobility in Si MOSFET's: Part I-effects of substrate impurity concentration," *IEEE Trans. Electron Devices*, vol. 41, no. 12, pp. 2357–2362, Dec. 1994, doi: 10.1109/16.337449.

[37] A. Pirovano, A. L. Lacaita, G. Ghidini, and G. Tallarida, "On the correlation between surface roughness and inversion layer mobility in Si-MOSFETs," *IEEE Electron Device Lett.*, vol. 21, no. 1, pp. 34–36, Jan. 2000, doi: 10.1109/55.817444.

[38] O. Bonno, S. Barraud, D. Mariolle, and F. Andrieu, "Effect of strain on the electron effective mobility in biaxially strained silicon inversion layers: An experimental and theoretical analysis via atomic force microscopy measurements and Kubo-Greenwood mobility calculations," *J. Appl. Phys.*, vol. 103, no. 6, Mar. 2008, Art. no. 63715, doi: 10.1063/1.2896589.

[39] T. Yamanaka, S. J. Fang, H.-C. Lin, J. P. Snyder, and C. R. Helms, "Correlation between inversion layer mobility and surface roughness measured by AFM," *IEEE Electron Device Lett.*, vol. 17, no. 4, pp. 178–180, Apr. 1996, doi: 10.1109/55.485166.

[40] M.-S. Kang, K. Toprasertpong, M. Takenaka, H. Oka, T. Mori, and S. Takagi, "Verification of influence of tail states and interface states on sub-threshold swing of Si n-channel MOSFETs over a temperature range of 4–300 K," *Jpn. J. Appl. Phys.*, vol. 61, Feb. 2022, Art. no. SC1032, doi: 10.35848/1347-4065/ac4444.

[41] I. M. Lifshitz, "The energy spectrum of disordered systems," *Adv. Phys.*, vol. 13, no. 52, pp. 483–536, 1964, doi: 10.1080/00018736400101061.

[42] D. Esseni, "On the modeling of surface roughness limited mobility in SOI MOSFETs and its correlation to the transistor effective field," *IEEE Trans. Electron Devices*, vol. 51, no. 3, pp. 394–401, Mar. 2004, doi: 10.1109/TED.2003.822344.

[43] P. Toniutti, D. Esseni, and P. Palestri, "Failure of the scalar dielectric function approach for the screening modeling in double-gate SOI MOSFETs and in FinFETs," *IEEE Trans. Electron Devices*, vol. 57, no. 11, pp. 3074–3083, Nov. 2010, doi: 10.1109/TED.2010.2068990.