Modulated Exchange and Gate Oxide Thickness Effects on Surface Roughness Limited Mobility

Shoji Wakahara
Process Technology Development Division, Renesas Technology Corp.,
4-1 Mizuhara, Itami-shi, Hyogo 664-0005, Japan

***Abstract*—Many-body effects are modulated by the roughness of the Si-SiO_2 interface of MOSFET, and act like scattering potentials. The theory of surface roughness scattering has considered only the modulation of the Hartree potential. We reformulate the theory to incorporate many-body effects, and show that exchange decreases mobility about 35 % with thick gate oxide. We also calculate the dependence of mobility on the oxide thickness. This dependence is influenced by exchange significantly.**

I. Introduction

The electron mobility in the silicon inversion layer is a key feature of the performance of MOSFET. The electron mobility is determined mainly by scatterings from charged centers, phonons, and the roughness of the Si-SiO_2 interface. At high gate voltage the surface roughness scattering is dominant, then the understanding of it is very important. However, although the previous theory of the surface roughness was derived from deep physical insights, they involve approximations whose validity is not clear[1], [2], [3], [4], [5], [6]. In this work, we reformulate the theory of the surface roughness by a field theoretical approach, and derive two corrections to the previous theory, one of which is the modulation of exchange by the surface roughness, and the other is the effect of the gate oxide thickness. These corrections are important for quantitative and qualitative discussions about the surface roughness and the device scaling.

II. Hamiltonian

We consider an n-channel inversion layer of a MOS structure fabricated on a (001) silicon surface and use the coordinate where z-axis is perpendicular to the Si-SiO_2 interface. We denote the asperity of the Si-SiO_2 interface at $\mathbf{r} = (x, y)$ by $\Delta(\mathbf{r})$. The region $z > \Delta(\mathbf{r})$ is occupied by silicon, and $-t_{ox} < z < \Delta(\mathbf{r})$ is occupied by SiO_2. The gate electrode lies at $z < -t_{ox}$.

The silicon conduction band has six energy minima. The lowest subband in energy comes from the two energy minima located along the [001] axis in the Brillouin zone. For simplicity we consider the low electron densities and low temperatures where all electrons reside in the lowest subband, and ignore intervalley scatterings. In this case we only need to count the two energy minima located along the [001] axis in the Brillouin zone. We distinguish these minima by index υ, and the spin by s.

The Hamiltonian $\mathcal{H}$ is decomposed as

$$\mathcal{H} = \mathcal{T} + \mathcal{U} + \mathcal{V} + \mathcal{W}, \tag{1}$$

where $\mathcal{T}$, $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ represent the the kinetic part, the electron-electron interaction, the interaction with the depletion charge, and the image potential, respectively. $\mathcal{H}$ is divided as

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_R, \tag{2}$$

where $\mathcal{H}_0$ is $\mathcal{H}$ of the smooth Si-SiO_2 interface, and $\mathcal{H}_R$ is the deviation of $\mathcal{H}$ from $\mathcal{H}_0$ due to the surface roughness. Similarly, $\mathcal{T}$, $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ are divided into those of the smooth interface and the deviations due to the surface roughness. We define $\mathcal{T}_0$, $\mathcal{U}_0$, $\mathcal{V}_0$, and $\mathcal{W}_0$, as $\mathcal{T}$, $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ of the smooth Si-SiO_2 interface, respectively. We refer to the deviations of $\mathcal{T}$, $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ as $\mathcal{T}_R$, $\mathcal{U}_R$, $\mathcal{V}_R$, and $\mathcal{W}_R$, respectively.

In the effective-mass approximation

$$\mathcal{T} = \sum_{\upsilon s} \int d\mathbf{x}\, \Psi^{\dagger}_{\upsilon s}(\mathbf{x}) \left[-\frac{\hbar^2}{2} \boldsymbol{\nabla}^{\dagger} \mathbf{m}^{-1} \boldsymbol{\nabla} \right] \Psi_{\upsilon s}(\mathbf{x}), \tag{3}$$

where $\Psi_{\upsilon s}(\mathbf{x})$ is the field operator, $\mathbf{x} = (\mathbf{r}, z)$, $\mathbf{m}$ is the mass matrix, $\boldsymbol{\nabla}$ is the nabla operator, and $\boldsymbol{\nabla}^{\dagger}$ is its conjugate. In our coordinate $\mathbf{m} = [m_{\alpha\beta}]$ is a diagonal matrix such as $m_{xx} = m_{yy} = m_t$, and $m_{zz} = m_l$.

We assume that the electron states of $\mathcal{H}_0$ are acceptably described by a single-particle wave function, $L_x^{-1/2} L_y^{-1/2} e^{i\mathbf{k}\cdot\mathbf{r}} \zeta_i(z)$, where $\mathbf{k}$ is the wavenumber, i is the subband index, and L_x and L_y are the sample sizes of the x- and the y-direction, respectively. $\zeta_0(z)$, $\zeta_1(z)$, $\zeta_2(z)$, $\cdots$ form a complete orthonormal set. Moreover, we assume that the potential barrier of SiO_2 is too high for electrons to penetrate SiO_2. Then, $\zeta_i(z) = 0 \quad \text{for} \quad z \le 0$.

With the surface roughness we must consider wave functions which are 0 at $z < \Delta(\mathbf{r})$, and the Hilbert space which consists of these functions. $\{L_x^{-1/2} L_y^{-1/2} e^{i\mathbf{k}\cdot\mathbf{r}} \zeta_i(z - \Delta(\mathbf{r})) \,|\, i = 0, 1, 2, \cdots\}$ is a complete orthonormal set in this Hilbert space. Therefore, $\Psi_{\upsilon s}$ can be represented as

$$\Psi_{\upsilon s}(\mathbf{x}) = \frac{1}{\sqrt{L_x L_y}} \sum_{\upsilon s} \sum_{\mathbf{k}\, i} e^{i\mathbf{k}\cdot\mathbf{r}} \zeta_i(z - \Delta(\mathbf{r}))\, c_{\upsilon s\mathbf{k}i}, \tag{4}$$

where $c_{\upsilon si\mathbf{k}}$ and $c^{\dagger}_{\upsilon si\mathbf{k}}$ are an annihilation and creation operator, respectively. By substituting Eq. (4) to Eq. (3), we obtain

$$\mathcal{T}_0 = \sum_{\upsilon s} \sum_{\mathbf{k}} \sum_{i j} t_{0ij}(\mathbf{k})\, c^{\dagger}_{\upsilon si\mathbf{k}}\, c_{\upsilon sj\mathbf{k}}, \tag{5}$$

and

$$\mathcal{T}_R = \frac{1}{L_x L_y} \sum_{\upsilon s} \sum_{\mathbf{k}\mathbf{q}} \sum_{i j} \widetilde{\Delta}(\mathbf{q})\, t_{Rij}(\mathbf{k} + \mathbf{q}, \mathbf{k})\, c^{\dagger}_{\upsilon si\mathbf{k}+\mathbf{q}}\, c_{\upsilon sj\mathbf{k}}, \tag{6}$$

978-1-4244-1753-7/08/$25.00 ©2008 IEEE

to the first order of $\Delta(\mathbf{r})$, where

$$t_{0ij}(\mathbf{k}) = \frac{\hbar^2 k^2}{2m_t}\delta_{ij} - \frac{\hbar^2}{2m_l}\int dz\,\zeta_i^*(z)\frac{d^2}{dz^2}\zeta_j(z), \quad (7)$$

$$t_{Rij}(\mathbf{k},\mathbf{k}') = \frac{\hbar^2(k'^2-k^2)}{2m_t}\int dz\,\zeta_i^*(z)\frac{d}{dz}\zeta_j(z), \quad (8)$$

and $\widetilde{\Delta}(\mathbf{k})$ is the Fourier transform of $\Delta(\mathbf{r})$.

The electron-electron interaction is given by

$$\mathcal{U} = \frac{e^2}{2}\sum_{\upsilon\upsilon'}\sum_{ss'}\int d\mathbf{x}\,d\mathbf{x}'\,\Psi^\dagger_{\upsilon's'}(\mathbf{x}')\Psi^\dagger_{\upsilon s}(\mathbf{x}) \times\psi(\mathbf{x};\mathbf{x}')\Psi_{\upsilon s}(\mathbf{x})\Psi_{\upsilon's'}(\mathbf{x}'), \quad (9)$$

where $\psi(\mathbf{x};\mathbf{x}')$ is the electrostatic potential, which is obtained by solving the Poisson's equation,

$$\nabla\cdot\left\{\varepsilon_i + (\varepsilon_s-\varepsilon_i)\,\theta\big(z-\Delta(\mathbf{r})\big)\right\}\nabla\psi(\mathbf{x};\mathbf{x}') = -\delta(\mathbf{x}-\mathbf{x}'), \quad (10)$$

on the condition that the potential is constant at the surface of the gate electrode, $\psi(\mathbf{x};\mathbf{x}')|_{z=-t_{ox}} = 0$, where ε_s and ε_i are the dielectric constants of silicon and silicon oxide, respectively. Eq. (10) can be rewritten as

$$\psi(\mathbf{x};\mathbf{x}') = \psi_0(\mathbf{r}-\mathbf{r}',z,z') - \int d\mathbf{x}''\left\{\nabla''\psi_0(\mathbf{r}-\mathbf{r}'',z,z'')\right\}\cdot\mathbf{P}(\mathbf{x}'';\mathbf{x}'), \quad (11)$$

where $\psi_0(\mathbf{r}-\mathbf{r}',z,z')$ is $\psi(\mathbf{x};\mathbf{x}')$ in the absence of the surface roughness, ∇'' is the nabla operator with respect to $\mathbf{x}''$, and

$$\mathbf{P}(\mathbf{x};\mathbf{x}') = (\varepsilon_s-\varepsilon_i)\left\{\theta\big(z-\Delta(\mathbf{r})\big)-\theta(z)\right\}\nabla\psi(\mathbf{x};\mathbf{x}'). \quad (12)$$

$\mathbf{P}(\mathbf{x};\mathbf{x}')$ can be regarded as the polarization due to the surface roughness. $\psi_0(\mathbf{x};\mathbf{x}')$ was obtained exactly[7]. We use $\psi_0(\mathbf{x};\mathbf{x}')$ in place of $\psi(\mathbf{x};\mathbf{x}')$ in Eq. (12), and substitute it to Eq. (11) to obtain $\psi(\mathbf{x};\mathbf{x}')$ to the first order of $\Delta(\mathbf{r})$. The calculation is performed in the momentum space, and leads to the Fourier transform of $\psi(\mathbf{x};\mathbf{x}')$,

$$\widetilde{\psi}(\mathbf{q},z;\mathbf{q}',z') = \int d\mathbf{r}\,d\mathbf{r}'\,e^{-i(\mathbf{q}\cdot\mathbf{r}-\mathbf{q}'\cdot\mathbf{r}')}\psi(\mathbf{r},z;\mathbf{r}',z') = (2\pi)^2\delta(\mathbf{q}-\mathbf{q}')\widetilde{\psi}_0(q,z,z') + \widetilde{\Delta}(\mathbf{q}-\mathbf{q}')\widetilde{\psi}_R(q,z;q',z'), \quad (13)$$

where $\widetilde{\psi}_0(q,z,z')$ is the Fourier transform of $\psi_0(\mathbf{r},z,z')$,

$$\widetilde{\psi}_0(q,z,z') = \frac{1}{2\varepsilon_s q}\left\{e^{-q|z-z'|} + \frac{\eta-e^{-2qt_{ox}}}{1-\eta e^{-2qt_{ox}}}e^{-q(z+z')}\right\}, \quad (14)$$

and

$$\widetilde{\psi}_R(\mathbf{q},z;\mathbf{q}',z') = (\varepsilon_i-\varepsilon_s)\left\{\mathbf{q}\cdot\mathbf{q}'\,\widetilde{\psi}_0(q,z,0)\widetilde{\psi}_0(q',0,z') + \frac{\varepsilon_s}{\varepsilon_i}\widetilde{\beta}_0(q,0,z)\widetilde{\beta}_0(q',0,z')\right\}. \quad (15)$$

to the first order of $\Delta(\mathbf{r})$. Here

$$\widetilde{\beta}_0(q,z,z') = \frac{1}{2\varepsilon_s}\left\{\mathrm{sgn}(z-z')\,e^{-q|z-z'|} + \frac{\eta-e^{-2qt_{ox}}}{1-\eta e^{-2qt_{ox}}}e^{-q(z+z')}\right\}, \quad (16)$$

and

$$\eta = \frac{\varepsilon_s-\varepsilon_i}{\varepsilon_s+\varepsilon_i}. \quad (17)$$

By substituting Eq. (4) and (13) to Eq. (9), we get

$$\mathcal{U}_0 = \frac{1}{2}\frac{1}{L_xL_y}\sum_{\upsilon\upsilon'}\sum_{ss'}\sum_{ii'}\sum_{jj'}\sum_{\mathbf{k}\mathbf{k}'}\sum_{\mathbf{q}}u_{0i'ij'j}(\mathbf{q}) \times c^\dagger_{\upsilon's'j'\mathbf{k}'-\mathbf{q}}c^\dagger_{\upsilon si'\mathbf{k}+\mathbf{q}}c_{\upsilon si\mathbf{k}}c_{\upsilon's'j\mathbf{k}'}, \quad (18)$$

$$\mathcal{U}_R = \frac{1}{2}\left(\frac{1}{L_xL_y}\right)^2\sum_{\upsilon\upsilon'}\sum_{ss'}\sum_{ii'}\sum_{jj'}\sum_{\mathbf{k}\mathbf{k}'}\sum_{\mathbf{q}\mathbf{q}'}\widetilde{\Delta}(\mathbf{q}-\mathbf{q}') \times u_{Ri'ij'j}(\mathbf{q},\mathbf{q}')\,c^\dagger_{\upsilon's'j'\mathbf{k}'-\mathbf{q}'}c^\dagger_{\upsilon si'\mathbf{k}+\mathbf{q}}c_{\upsilon si\mathbf{k}}c_{\upsilon sj\mathbf{k}'}, \quad (19)$$

where

$$u_{0i'ij'j}(\mathbf{q}) = e^2\int dz\,dz'\,\zeta^*_{i'}(z)\zeta_i(z)\zeta^*_{j'}(z')\zeta_j(z')\widetilde{\psi}_0(\mathbf{q},z,z'), \quad (20)$$

and

$$u_{R\upsilon\upsilon'i'ij'j}(\mathbf{q},\mathbf{q}') = e^2\int dz\,dz'\,\zeta^*_{i'}(z)\zeta_i(z)\zeta^*_{j'}(z')\zeta_j(z') \times\left\{-\widetilde{\beta}_0(\mathbf{q}',z,z')-\widetilde{\beta}_0(-\mathbf{q},z',z)+\widetilde{\psi}_R(\mathbf{q},z;\mathbf{q}',z')\right\}, \quad (21)$$

to the first order of $\Delta(\mathbf{r})$.

The interaction with the depletion charge is described by

$$\mathcal{V} = \sum_{\upsilon s}\int d\mathbf{r}\,dz\,\Psi^\dagger_{\upsilon s}(\mathbf{r},z)V(\mathbf{r},z)\Psi_{\upsilon s}(\mathbf{r},z), \quad (22)$$

where $V(\mathbf{r},z)$ is the potential due to the depletion charge,

$$V(\mathbf{r},z) = e^2N_A\int d\mathbf{r}'\,dz'\,\psi(\mathbf{r},z;\mathbf{r}',z') \times\theta\big(z'-\Delta(\mathbf{r}')\big)\theta\big(z_d+\Delta(\mathbf{r}')-z'\big), \quad (23)$$

where N_A is the acceptor concentration, and z_d is the depletion layer width. We expand Eq. (22) to the first order $\Delta(\mathbf{r})$ by using Eq. (4), (13), and (23). The results are

$$\mathcal{V}_0 = \sum_{\upsilon s}\sum_{ij}\sum_{\mathbf{k}}v_{0ji}c^\dagger_{\upsilon sj\mathbf{k}}c_{\upsilon si\mathbf{k}}, \quad (24)$$

and

$$\mathcal{V}_R = \frac{1}{L_xL_y}\sum_{\upsilon s}\sum_{ij}\sum_{\mathbf{k}\mathbf{q}}\widetilde{\Delta}(\mathbf{q})\,v_{Rji}(\mathbf{q})\,c^\dagger_{\upsilon sj\mathbf{k}+\mathbf{q}}c_{\upsilon si\mathbf{k}}, \quad (25)$$

where

$$v_{0ji} = \int dz\,\zeta^*_j(z)\zeta_i(z)V_0(z), \quad (26)$$

and

$$v_{Rji}(q) = \frac{e^2N_{depl}}{\varepsilon_s}\left\{\delta_{ij} + \eta\frac{1+e^{-2qt_{ox}}}{1-\eta e^{-2qt_{ox}}}\int dz\,\zeta^*_j(z)\zeta_i(z)e^{-qz}\right\}. \quad (27)$$

for $qz_d \gg 1$. Here $N_{depl} = N_Az_d$, and $V_0(z)$ is $V(\mathbf{r},z)$ in the absence of the surface roughness.

When $qz_d \lesssim 1$, we need an additional term to describe $v_{Rji}(q)$. The surface roughness modulates the depletion charge.

This is reflected in $V(\mathbf{r}, z)$ by introducing $\Delta(\mathbf{r})$-dependent z_d. It can be shown that this introduces a term proportional to $\exp(-qz_d)$ to $v_{Rji}(q)$. When q is small, this term cannot be ignored. However, $v_{Rji}(q)$ with small q may contribute to the conductivity little because small q corresponds forward scattering. Then we have omitted this additional term.

The image potential part of the Hamiltonian is obtained by

$$\mathcal{W} = \sum_{\upsilon s} \int d\mathbf{r}\, dz\, \Psi^{\dagger}_{\upsilon s}(\mathbf{r}, z)\, W(\mathbf{r}, z)\, \Psi_{\upsilon s}(\mathbf{r}, z), \tag{28}$$

where $W(\mathbf{r}, z)$ is the image potential acting on the electron at $(\mathbf{r}, z)$, and obtained by subtracting the contribution of the self-field from $\psi(\mathbf{x}; \mathbf{x})$, and multiplying $\frac{1}{2}$. The calculation is similar to those for $\mathcal{U}$ and $\mathcal{V}$, and leads to

$$\mathcal{W}_0 = \sum_{\upsilon s} \sum_{ij} \sum_{\mathbf{k}} w_{0ji}\, c^{\dagger}_{\upsilon sj\mathbf{k}} c_{\upsilon si\mathbf{k}}, \tag{29}$$

$$\mathcal{W}_R = \frac{1}{L_x L_y} \sum_{\upsilon} \sum_{\mathbf{k}\mathbf{q}} \widetilde{\Delta}(\mathbf{q})\, w_{Rji}(\mathbf{q})\, c^{\dagger}_{\upsilon sj\mathbf{k}+\mathbf{q}} c_{\upsilon si\mathbf{k}}, \tag{30}$$

where

$$w_{0ji} = \frac{e^2}{8\pi\varepsilon_s} \int dz\, \zeta^*_j(z)\, \zeta_i(z) \int_0^{+\infty} dk\, \frac{\eta - e^{-2kt_{ox}}}{1 - \eta e^{-2kt_{ox}}} e^{-2kz}, \tag{31}$$

and

$$w_{Rji}(\mathbf{q}) = -\frac{e^2}{2\varepsilon_s} \int dz\, \zeta^*_j(z)\, \zeta_i(z) \int \frac{d\mathbf{k}}{(2\pi)^2} \left\{ \frac{\eta - e^{-2kt_{ox}}}{1 - \eta e^{-2kt_{ox}}} e^{-2kz} - \varepsilon_s \widetilde{\psi}_R(\mathbf{k} + \mathbf{q}, z; \mathbf{k}, z) \right\}. \tag{32}$$

III. Mobility

The electron conductivity σ is calculated by the Kubo formula, and it can be shown that

$$\sigma \approx \frac{e^2 g_v}{\pi\hbar m_t^2 L_x L_y} \sum_{\mathbf{p}\mathbf{p}'} \mathbf{p} \cdot \mathbf{p}' \left[\operatorname{Im} G(\mathbf{p}, \mathbf{p}', 0) \right]^2, \tag{33}$$

where $G(\mathbf{p}, \mathbf{p}', \omega)$ is the retarded Green's function, $\mathbf{p}$ and $\mathbf{p}'$ are electron momenta parallel to the Si-SiO_2 interface, and g_v is the valley degeneracy which is 2 for the lowest subband.

Unfortunately $\mathcal{H}$ is so complex that $G(\mathbf{p}, \mathbf{p}', \omega)$ cannot be calculated exactly, so we use the perturbation theory in terms of $\mathcal{U}_0$, $\mathcal{U}_R$, $\mathcal{V}_R$, and $\mathcal{W}_R$. The perturbation calculation in terms of $\mathcal{U}_0$ gives the expansion of σ in powers of $r_s = (\pi N_s a_B^2)^{-1/2}$, where N_s is the electron sheet density and a_B is the Bohr radius. On the other hand, the perturbation calculation in terms of $\mathcal{U}_R$, $\mathcal{W}_R$, and $\mathcal{V}_R$ gives the expansion by $\Delta(\mathbf{r})$.

$G(\mathbf{p}, \mathbf{p}', \omega)$ is obtained by analytic continuation from Matsubara Green's function $G(\mathbf{p}, \mathbf{p}', i\omega)$. $G(\mathbf{p}, \mathbf{p}', i\omega)$ can be expressed in the form,

$$G(\mathbf{p}, \mathbf{p}', i\omega) = \delta_{\mathbf{p}\mathbf{p}'} G^{(0)}(\mathbf{p}, i\omega) + \frac{1}{L_x L_y} \sum_{\mathbf{q}} G^{(0)}(\mathbf{p}, i\omega)\, \widetilde{\Delta}(\mathbf{q}) \times \Sigma(\mathbf{p}, \mathbf{p} - \mathbf{q}, i\omega)\, G(\mathbf{p} - \mathbf{q}, \mathbf{p}', i\omega), \tag{34}$$

where $G^{(0)}(\mathbf{p}, i\omega)$ is the Green's function in the absence of the surface roughness. To the lowest order of $\widetilde{\Delta}(\mathbf{q})$

$$\Sigma(\mathbf{p}, \mathbf{p} - \mathbf{q}, i\omega) = \Sigma_T(\mathbf{p}, \mathbf{p} - \mathbf{q}) + \Sigma_V(\mathbf{q}) + \Sigma_W(\mathbf{q}) + \Sigma_H(\mathbf{q}) + \Sigma_x(\mathbf{p}, \mathbf{p} - \mathbf{q}), \tag{35}$$

where Σ_T, Σ_V, and Σ_W are the self-energies due to $\mathcal{T}_R$, $\mathcal{V}_R$, and $\mathcal{W}_R$, respectively. Both Σ_H and Σ_x are the self-energies due to $\mathcal{U}_R$, and described by Feynman diagrams shown in Fig. 1 and 2, respectively. Wavy lines in these figures correspond to $u_R(\mathbf{q}, \mathbf{q}')$, and cross symbols represent $\widetilde{\Delta}(\mathbf{q})$ (Fig. 3). The surface roughness breaks the translational symmetry parallel to the Si-SiO_2 interface, and $\widetilde{\Delta}(\mathbf{q})$ gives the momentum $\mathbf{q}$. If we replace $u_R(\mathbf{q}, \mathbf{q}')$ by $u_0(\mathbf{q})$ in these figures, we obtain the familiar diagrams which describe the Hartree and exchange terms in $G^{(0)}(\mathbf{p}, i\omega)$. So we can regard Σ_H and Σ_x as the modulations of the Hartree and exchange terms by the surface roughness, respectively.

The exchange term is of a higher order in r_s than the Hartree term. Then exchange becomes relatively small at large N_s, and this is the reason why the electronic states are well described by the Hartree approximation in the absence of the surface roughness. In contrast to this, Σ_H and Σ_x are of the same order in r_s. This difference comes from the q-dependence of $u_0(\mathbf{q})$ and $u_R(\mathbf{q}, \mathbf{q}')$. $u_0(\mathbf{q})$ is approximately proportional to the inverse of q. On the other hand, $u_R(\mathbf{q}, \mathbf{q}')$ is of the zeroth order in q and q'. The previous theory uses the Hartree approximation for the electron-electron interaction from the start, so involves Σ_H, but doesn't count Σ_x. However, because Σ_H and Σ_x are of the same order in r_s, there is no reason to omit Σ_x.

If intersubbands scattering is ignored, the result becomes simple.

$$\Sigma_H(\mathbf{q}) + \Sigma_{\mathcal{V}}(\mathbf{q}) + \Sigma_{\mathcal{W}}(\mathbf{q}) = \hbar^{-1} \varepsilon^{-1}(\mathbf{q}) \left\{ u_R(\mathbf{q}, 0)\, N_s + v_R(\mathbf{q}) + w_R(\mathbf{q}) \right\}, \tag{36}$$

and

$$\Sigma_x(\mathbf{p} + \mathbf{q}, \mathbf{p}) = -\frac{1}{\hbar L_x L_y} \sum_{|\mathbf{p}+\mathbf{q}'| < k_F} u_R(\mathbf{q}', \mathbf{q}' - \mathbf{q}), \tag{37}$$

where $\varepsilon^{-1}(\mathbf{q})$ is the dielectric function, and k_F is the Fermi wavenumber. We have omitted suffixes for the subbands because we are considering only the lowest subband. The life time τ is given by

$$\frac{1}{\tau} = 2\pi\hbar \int \frac{d\mathbf{q}}{(2\pi)^2} \frac{1}{L_x L_y} \left\langle \widetilde{\Delta}(\mathbf{q})\, \widetilde{\Delta}(-\mathbf{q}) \right\rangle_{sr} \Sigma(\mathbf{k}, \mathbf{k} + \mathbf{q})\, \Sigma(\mathbf{k} + \mathbf{q}, \mathbf{k}) \times \delta\left(E_{\mathbf{k}+\mathbf{q}} - E_F\right)\left(1 - \cos\theta_{\mathbf{k}\,\mathbf{k}+\mathbf{q}}\right), \tag{38}$$

where $\theta_{\mathbf{k}\,\mathbf{k}+\mathbf{q}}$ is the angle between $\mathbf{k}$ and $\mathbf{k}$+$\mathbf{q}$, and $\langle \cdots \rangle_{sr}$ stands for the average over the ensemble of the surface roughness. We assume the form of the autocorrelation function of $\Delta(\mathbf{r})$ as $\langle \Delta(\mathbf{r})\, \Delta(\mathbf{r}') \rangle_{sr} = \Delta^2 e^{-|\mathbf{r}-\mathbf{r}'|^2/\Lambda^2}$. We calculate $\varepsilon(\mathbf{q})$ by the random-phase approximation.

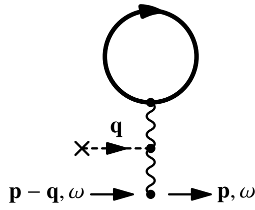

Fig. 1. diagram for $\widetilde{\Delta}(\mathbf{q})\,\Sigma_H(\mathbf{q})$

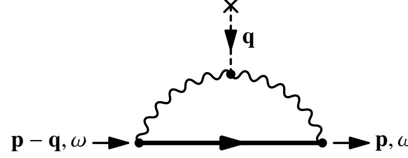

Fig. 2. diagram for $\widetilde{\Delta}(\mathbf{q})\,\Sigma_x(\mathbf{p}, \mathbf{p}-\mathbf{q})$

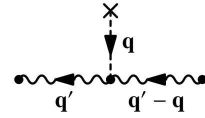

Fig. 3. diagram for $\widetilde{\Delta}(\mathbf{q})\,u_R(\mathbf{q}', \mathbf{q}'-\mathbf{q})$

IV. Results and Discussion

In Fig. 4 we show calculated surface roughness limited mobilities as functions of N_s for $t_{ox} = 32\,\text{nm}$ and $1\,\text{nm}$. We adopt $\Lambda = 1.5\,\text{nm}$ and $\Delta = 0.35\,\text{nm}$ which were extracted from comparisons between calculations and experimental data of Hall conductivities by Goodnick *et al.*[8]. The previous theory ignores the finiteness of t_{ox} assuming that $k_F t_{ox} \gg 1$, but this assumption does not hold for $t_{ox} \lesssim 2\,\text{nm}$. We consider the gate electrode explicitly when solving the Poisson equation, Eq. (10), then can discuss the t_{ox}-dependence. To evaluate the effect of Σ_x, we compare the mobilities with and without Σ_x. μ_H is the mobility without Σ_x, and μ_{Hx} is the one which involves both Σ_H and Σ_x. When $t_{ox} = 32\,\text{nm}$, Σ_x decreases the mobility by about 35 %, and this ratio is almost insensitive to N_s. This implies that Σ_H and Σ_x is of the same order in r_s.

The decrease of t_{ox} emphasizes the effect of the gate electrode. The electron induces the positive charge at the interface of the gate electrode. This charge decreases Σ_x in magnitude, and changes the sign of it at the extremely thin gate oxide. So μ_{Hx} is less than μ_H at $t_{ox} = 1\,\text{nm}$. The finiteness of t_{ox} affects also Σ_H, and μ_H decreases with the decrease of t_{ox}. The smaller N_s is, the more significantly μ_H decreases, because the finite t_{ox} effect is considerable only at $k_F t_{ox} \lesssim 1$, and small N_s corresponds to small k_F.

In reality, impurity scattering is dominant at small N_s. We calculate the mobility limited by coulomb scattering from acceptors in silicon substrate. The total mobility is calculated by Matthiessen's rule with surface roughness limited and coulomb scattering limited mobilities. In Fig. 5 we show calculated total mobilities as functions of N_s. Impurity scattering has little dependence on t_{ox}, so it smokes out the t_{ox}-dependence of μ_H and μ_{Hx}. The decrease of t_{ox} from 32 nm to 1 nm decreases the total mobility by less than 10 % with Σ_x, but without Σ_x the decrease of the mobility remains significant even with impurity scattering.

In conclusion, we have reformulated the theory of the surface roughness by using a field theoretical approach. We have found the effects of the modulation of exchange and the finiteness of the gate oxide thickness. These were ignored by the previous theory, but give the considerable effects on mobilities.

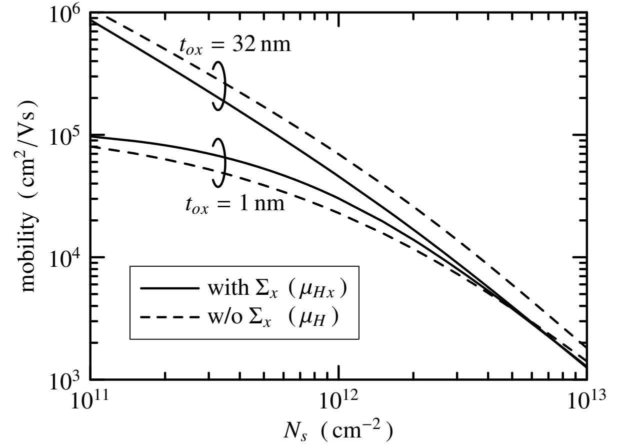

Fig. 4. Calculated surface roughness limited electron mobilities with or without Σ_x.

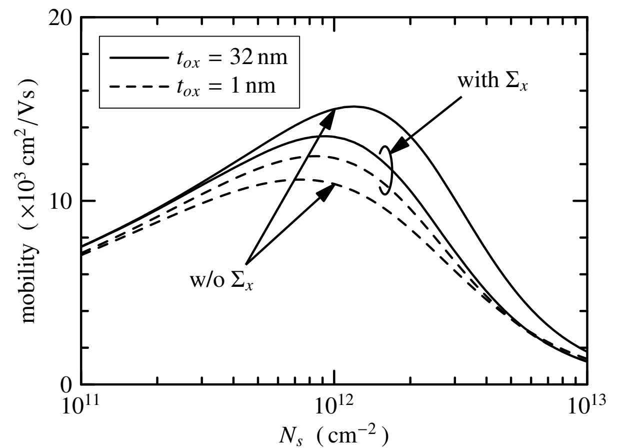

Fig. 5. Calculated total electron mobilities with or without Σ_x. $N_A = 10^{17}\,\text{cm}^{-3}$. Surface roughness scattering and coulomb scattering from acceptors are included.

References

[1] R. E. Prange and T.-W. Nee, *Phys. Rev.*, vol. 168, pp. 779–786, 1968.

[2] Y. C. Cheng and E. A. Sullivan, *J. Appl. Phys.*, vol. 44, pp. 923–925, 1973.

[3] A. V. Chaplik and M. V. Entin, *Sov. Phys.—JETP*, vol. 28, pp. 514–517, 1969.

[4] T. Ando, *J. Phys. Soc. Jpn.*, vol. 43, pp. 1616–1626, 1977.

[5] S. Mori and T. Ando, *Phys. Rev. B*, vol. 19, pp. 6433–6441, 1979.

[6] T. Ando, A. B. Fowler, and F. Stern, *Rev. Mod. Phys.*, vol. 54, pp. 437–672, 1982.

[7] A. V. Chaplik, *Sov. Phys.—JETP*, vol. 35, pp. 395–398, 1972.

[8] S. M. Goodnick, D. K. Ferry, and C. W. Wilmsen, *Phys. Rev. B*, vol. 32, pp. 8171–8186, 1985.