Channel effective mass and interfacial effects in Si and SiGe metal-oxide-semiconductor field effect transistor: A charge control model study

Yifei Zhang and Jasprit Singh
Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, Michigan 48109-2122

(Received 20 October 1997; accepted for publication 31 December 1997)

We present results of a numerical formalism developed to address the band structure and charge control problem in n- and p-type silicon and silicon-germanium metal-oxide-semiconductor field effect transistors. We focus on the following issues: (i) the dependence of the in-plane carrier effective mass on sheet charge density and germanium content; (ii) the fraction of charge near the interface and the evaluation of the interface roughness matrix element. Results are compared to existing models. For n-type structure, the effective mass approximation and deformation potential theory is used to describe the electron states. However, for p-type structure, a six-band $\mathbf{k}\cdot\mathbf{p}$ Kohn–Luttinger formulation is used to describe the hole states due to the strong coupling of heavy-hole, light-hole, and split-off bands. This allows us to examine the influence of the coupling of the heavy-hole, light-hole, and the split-off bands. © *1998 American Institute of Physics.* [S0021-8979(98)00108-X]

I. INTRODUCTION

The metal-oxide-semiconductor field effect transistor (MOSFET) technology based on $Si–SiO_2$ technology is the most important device technology for modern microelectronics. Both n-type MOSFET (NMOS) and p-type MOSFET (PMOS) are an integral part of this technology. In spite of the importance of the MOS there are a number of unresolved issues for both n-type and p-type devices. These include: (i) Interface roughness scattering and its role in transport: This important issue is presently addressed through the use of very simplistic models for the envelope functions representing the two-dimensional interface charge (ii) Present PMOS charge control models do not include the coupling of the heavy-hole (HH) and light-hole (LH) bands and ignore the split-off (SO) band. Since the split-off energy in Si is only 44 meV it is important to verify if this approximation is justified (iii) An important manifestation of the simplistic nature of the modeling of MOS physics is the significant disagreement in the calculated and measured velocity-field relations.[1]

With the development of lattice-mismatched heteroepitaxy, it is possible to fabricate semiconductor devices with strained layer components. The heteroepitaxial system of $Si_{1-x}Ge_x$ layers grown on Si substrates is of great technological interest for fabricating semiconductor devices. Silicon based Si–SiGe heterojunction bipolar transistors (HBTs) have shown remarkable performance. The combined effects of strain and alloying are found to produce a monotonic increase in hole mobility in the theoretical study on the transport properties of pseudomorphic p-type material.[2] Furthermore, it has been shown that the strain in the channel of PMOS remarkably improves device performance which is related to higher mobility and lower carrier mass.[3]

In this article we will examine the following issues: (i) the importance of including the influence of the coupling between the heavy-hole, light-hole and split-off bands on the charge control studies of PMOS. It is important to note that while a four-band $\mathbf{k}\cdot\mathbf{p}$ method has been used to study the pseudomorphic p-type modulation doped field effect transistor (MODFET),[4] no such studies have been done for the MOS structure. (ii) The variation of the effective mass in the channel for both NMOS and PMOS as a function of gate bias (or sheet charge density) and Ge content; and (iii) the interface roughness matrix elements and the validity of the currently used interface roughness limited mobility model.

In current charge control models to address PMOS one self-consistently solves Schrödinger equation and Poisson equation by using an effective mass approximation method.[5] In the effective mass approximation method, the heavy-hole and light-hole bands are treated separately and usually the effect of SO band is neglected. However, as we know, the valence bands are very nonparabolic and anisotropic in the k space, which is mainly due to the strong coupling effect of heavy-hole, light-hole, and split-off hole states. The spin-orbit energy splitting between split-off and heavy-hole and light-hole is comparable to the energy splitting between heavy-hole and light-hole under a large quantum confinement and strain. As a result, it is necessary to use a six-band $\mathbf{k}\cdot\mathbf{p}$ method in order to give an accurate representation of the strongly coupled states.

The outline of the remainder of this article is as follows. In Sec. II, the modeling formalism for the n- and p-type devices is described. In Sec. III, results of the formalism are presented. These include the results for n- and p-type MOSFET. Conclusion are made in Sec. IV.

II. MODELING FORMALISM

Our model gives the potential profile in a MOS structure by solving the Schrödinger equation and Poisson equation self-consistently. The Schrödinger equation yields the confined charge terms in the Poisson equation which, in turn,

 © 1998 American Institute of Physics

determines the potential profile which is fed back into the Schrödinger equation. An iterative process brings Poisson equations into convergence.

A. Charge control model for NMOS

The one-dimensional Poisson equation can be written as

$$\frac{d^2}{dz^2} E_c(z) = -\frac{\rho(z)}{\epsilon(z)}, \tag{1}$$

where E_c is the conduction band profile in the device, ρ is the total charge density, and ϵ is the dielectric constant which can be changed in different regions of the device to account for different material parameters across interfaces. The total charge density ρ is the sum of the doping charge, the free charge including the hole and electron, and the quantum-confined charge. This can be written as

$$\rho(z) = q\left(N_d^*(z) - N_a^*(z) - n_{\text{free}}(z) + p_{\text{free}}(z) - \sum_i n_i \psi_i^*(z)\psi_i(z) \right), \tag{2}$$

where N_a^* and N_d^* are the effective doping concentrations, n_{free} and p_{free} are the free carrier concentrations, and the sum is over i two-dimensionally confined subbands of which normalized envelope functions are ψ_i and in which the occupation is n_i (p_i for the hole case). The effective acceptor concentration, N_a^*, can be written as

$$N_a^* = N_a\left(\frac{1}{1 + 2e^{(E_d - E_f)/k_B T}} \right). \tag{3}$$

The solution of the Poisson equation is performed by a vectorized Newton's method.

In the calculation of the charge density, one must determine the quantized two-dimensional charge and the free-carrier charge. The quantized two-dimensional charge can be calculated from the eigen energy levels obtained by solving the Schrödinger equation. The free-carrier charge density can be written as

$$n_{\text{free}}(z) = N_c F_{1/2}\left(\frac{E_f - E_c(z)}{k_B T} \right), \tag{4}$$

where N_c is the material effective density of states and $F_{1/2}$ is a half-order Fermi integral. The reason we use Fermi–Dirac statistics instead of Boltzmann statistics, which have been used by many authors in the past, to determine the free-carrier concentrations is that in heavily doped cases, the bands are near degenerate or degenerate, and Boltzmann statistics will overestimate the free-carrier concentrations. To calculate the Fermi integral fast and accurately, a look-up table is used. Note that in Eq. (4), we use a three-dimensional effective density of states to obtain the free-carrier charge. However, to distinguish two-dimensional carriers with three-dimensional ones in the quantum well region, a cut-off value, which usually is set to be the lower boundary value, is used. For those carriers whose energy is below this cut-off value, they are treated as two-dimensional gas. For those whose energy is above this cut-off value, they are thus treated as free carriers.

In order to determine the two-dimensionally confined charge profile, one must solve the Schrödinger equation for the subband envelope functions and their occupations. For the n-type MOSFET, the one-band Schrödinger equation can be used. The Schrödinger equation must, in general, be written in three dimensions, and the effective mass will be, in general, a tensor. In the electron case the z-dependent part of the Schrödinger equation is strictly separable from the in-plane part of the equation which gives us extended Bloch-like states. The one-dimensional (z-dependent) Schrödinger equation can be written using the perpendicular part of the effective mass tensor as follows:

$$\frac{d^2}{dz^2} \psi_n(z) + \frac{2m_w}{\hbar^2}[E_n - V(z)]\psi_n(z) = 0, \tag{5}$$

where m_w represents electron effective mass along the quantum confinement direction. This method is also referred as effective mass approximation method. For bulk silicon, the band edge of the conduction band is at X point. Corresponding to this X point there are six equivalent valleys. Under quantum confinement or strain effect, the sixth degeneracy of X point is decomposed into a second and forth degeneracy. For electrons in the longitudinal valleys (along $\mathbf{k_z}$ direction), m_w is equal to m_t, electron longitudinal effective mass. Whereas for electrons in the transverse valleys (in-plane directions), m_w is equal to m_t, electron transverse effective mass. Under the quantum confinement, the two longitudinal valleys will move down with respect to the four transverse valleys. This splitting is related to the difference of longitudinal and transverse effective mass. Since the longitudinal effective mass for silicon is much larger than the transverse effective mass, the electron in the longitudinal valleys will occupy the lowest subband. In contrast, under the strain effect, the two longitudinal valleys will move up with respect to the four transverse valleys. According to the deformation potential theorem, we can write the change of energy of states along the (100) direction in $\mathbf{k}$ space as,

$$\delta E^{(100)} = \Xi_d^{(100)}(\epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz}) + \Xi_u^{(100)}\epsilon_{xx}, \tag{6}$$

$$\delta E^{(010)} = \Xi_d^{(100)}(\epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz}) + \Xi_u^{(100)}\epsilon_{yy}, \tag{7}$$

$$\delta E^{(001)} = \Xi_d^{(100)}(\epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz}) + \Xi_u^{(100)}\epsilon_{zz}, \tag{8}$$

where $\Xi_d^{(100)}$ and $\Xi_u^{(100)}$ represent the dilation and uniaxial deformation potentials[6] for the conduction band $(\bar{1}00)$ valleys, and $\epsilon_{ij}(i,j = x,y,z)$ represents the deformation coefficient which can be related to strain coefficient. In our case, we examine the growth of strained layer along the (001) direction, as noted earlier,

$$\epsilon_{xx} = \epsilon_{yy} = \epsilon, \qquad \epsilon_{zz} = -\frac{2c_{12}}{c_{11}}\epsilon, \tag{9}$$

where c_{12} and c_{11} are the elastic stiffness and ϵ can be expressed in terms of the bulk lattice constant of the substrate material, a_S, and the layer material, a_L,

$$\epsilon=\frac{a_S}{a_L}-1. \tag{10}$$

For $Si_{1-x}Ge_x$-Si system, the lattice constant of the channel material, $Si_{1-x}Ge_x$, is larger than that of Si. This results in a negative value for ϵ.

Once the Schrödinger equation has been solved for the envelope functions and the subband energy levels, it is straightforward to calculate the subband occupations. Since the subband density of states is constant with energy, the occupation comes from the first-order Fermi integral which is analytically integrable. In the electron case, we can write the occupation as

$$n_i=\frac{g_i m_i}{\pi\hbar^2}\ln\left[1+\exp\left(\frac{E_i-E_f}{k_BT}\right)\right], \tag{11}$$

where m_i and g_i are the in-plane effective mass and the degeneracy of ith subband, respectively. In our case, for electrons in the longitudinal valleys, we have

$$m_i=m_t, \quad g_i=2. \tag{12}$$

Whereas for those in the transverse valleys, we have

$$m_i=(m_t m_l)^{1/2}, \quad g_i=4. \tag{13}$$

The reason we do not use m_l as in-plane effective mass for electrons in the transverse valleys is because the in-plane mass tensor in this case has different values along x and y direction, which suggests a elipsordal shape for the constant energy surface.

B. Charge control model for PMOS

To calculate the confined charges in the valence subbands, a six-band $\mathbf{k}\cdot\mathbf{p}$ method is used to represent the interaction among the top three valence bands which are heavy-hole, light-hole, and split-off hole bands. In recent study of the quantization effect in inversion layer of p-type MOSFET, Hu *et al.*[5] utilized the effective mass approximation to treat the hole bands separately by solving the one-dimensional Schrödinger equation. However, due to the double degeneracy of top valence bands at the zone center and the small spin-orbit splitting between the split-off band and heavy-hole and light-hole bands, the valence bands are strongly coupled to each other to demonstrate an anisotropic and nonparabolic band structure. Consequently, in order to give an accurate representation of the strongly coupled states, we use a six-band $\mathbf{k}\cdot\mathbf{p}$ method, introduced by Luttinger and Kohn,[7] to solve a six by six matrix which has included the interaction of HH, LH, and SO valence bands. The matrix equation, with the Luttinger and Kohn phases, if of the form:

$$-\begin{bmatrix} H_{hh} & b & c & 0 & ib/\sqrt{2} & -i\sqrt{2}c \\ b^* & H_{lh} & 0 & c & -iq & i\sqrt{3}b/\sqrt{2} \\ c^* & 0 & H_{lh} & -b & -i\sqrt{3}b^*/\sqrt{2} & -iq \\ 0 & c^* & -b^* & H_{hh} & -i\sqrt{2}c^* & -ib^*/\sqrt{2} \\ -ib^*/\sqrt{2} & iq & i\sqrt{3}b/\sqrt{2} & i\sqrt{2}c & H_{so} & 0 \\ i\sqrt{2}c^* & -i\sqrt{3}b^*/\sqrt{2} & iq & ib/\sqrt{2} & 0 & H_{so} \end{bmatrix} \begin{bmatrix} a_n(\mathbf{k}_\parallel,z)\cdot\phi_{3/2,3/2} \\ b_n(\mathbf{k}_\parallel,z)\cdot\phi_{3/2,1/2} \\ c_n(\mathbf{k}_\parallel,z)\cdot\phi_{3/2,-1/2} \\ d_n(\mathbf{k}_\parallel,z)\cdot\phi_{3/2,-3/2} \\ e_n(\mathbf{k}_\parallel,z)\cdot\phi_{1/2,1/2} \\ f_n(\mathbf{k}_\parallel,z)\cdot\phi_{1/2,-1/2} \end{bmatrix} = E_n \begin{bmatrix} a_n(\mathbf{k}_\parallel,z)\cdot\phi_{3/2,3/2} \\ b_n(\mathbf{k}_\parallel,z)\cdot\phi_{3/2,1/2} \\ c_n(\mathbf{k}_\parallel,z)\cdot\phi_{3/2,-1/2} \\ d_n(\mathbf{k}_\parallel,z)\cdot\phi_{3/2,-3/2} \\ e_n(\mathbf{k}_\parallel,z)\cdot\phi_{1/2,1/2} \\ f_n(\mathbf{k}_\parallel,z)\cdot\phi_{1/2,-1/2} \end{bmatrix}. \tag{14}$$

The elements in the matrix are given by

$$H_{hh}=\frac{\hbar^2}{2m_0}\left[(k_x^2+k_y^2)(\gamma_1+\gamma_2)-(\gamma_1-2\gamma_2)\frac{\partial^2}{\partial z^2}\right]-V(z)+\delta_{hh}, \tag{15}$$

$$H_{lh}=\frac{\hbar^2}{2m_0}\left[(k_x^2+k_y^2)(\gamma_1-\gamma_2)-(\gamma_1+2\gamma_2)\frac{\partial^2}{\partial z^2}\right]-V(z)+\delta_{lh}, \tag{16}$$

$$H_{so}=\frac{\hbar^2\gamma_1}{2m_0}\left(k_x^2+k_y^2-\frac{\partial^2}{\partial z^2}\right)-V(z)+\delta_{so}+\Delta_0, \tag{17}$$

$$c=\frac{\sqrt{3}\hbar^2}{2m_0}\left[\gamma_2(k_x^2-k_y^2)-2i\gamma_3k_xk_y\right], \tag{18}$$

$$b=-\frac{\sqrt{3}\hbar^2}{2m_0}(k_x-ik_y)\gamma_3\frac{\partial}{\partial z}, \tag{19}$$

TABLE I. Material parameters. Alloy parameters are obtained by linear interpolation.

Material parameters	Symbol	Units	Si value	Ref.	Ge value	Ref.
Electron mass	m_1	m_0	0.98	9	1.64	9
	m_1	m_0	0.19	9	0.082	9
Valence-band structure	γ_1	⋯	4.22	10	13.35	10
	γ_2	⋯	0.39	10	4.25	10
	γ_3	⋯	1.44	10	5.69	10
Spin-orbit splitting	Δ_0	eV	0.044	11	0.282	12
Deformation potential	a	eV	2.1	13	2.0	13
	b	eV	−1.5	9	−2.2	9
Lattice constant	a_0	Å	5.4309	9	5.6461	9
Dielectric constant	ϵ	⋯	11.9	9	16.0	9
Elastic stiffness	c_{11}	10^{11} dyn cm^{-2}	16.56	14	12.853	15
	c_{12}	10^{11} dyn cm^{-2}	6.39	14	4.826	15

$$q=\frac{\sqrt{2}\hbar^2\gamma_2}{2m_0}\left(k_x^2+k_y^2+2\frac{\partial^2}{\partial z^2}\right). \tag{20}$$

Here γ_1, γ_2, and γ_3 are the Luttinger parameters[8] for the given material. The symbol Δ_0 represents the spin-orbit splitting energy. The terms like $\phi_{3/2,3/2}$ represent the pure angular momentum states from the solution to the hydrogen-atom problem with spin-orbit interactions included. Equation (14) is solved for the p-type MOSFET to determine the subband energy levels and envelope functions. The strain-induced band splitting is given by

$$\delta_{\mathrm{hh}}=\left[2a\left(\frac{c_{11}-c_{12}}{c_{11}}\right)+b\left(\frac{c_{11}+2c_{12}}{c_{11}}\right)\right]\epsilon, \tag{21}$$

$$\delta_{\mathrm{lh}}=\left[2a\left(\frac{c_{11}-c_{12}}{c_{11}}\right)-b\left(\frac{c_{11}+2c_{12}}{c_{11}}\right)\right]\epsilon, \tag{22}$$

$$\delta_{\mathrm{so}}=2a\left(\frac{c_{11}-c_{12}}{c_{11}}\right), \tag{23}$$

where a, b, and c are deformation potentials. It is important to note that the problem is quite complex due to the fact that the resulting hole bands are very nonparabolic and anisotropic. Because of this, the Schödinger equation must be solved all through $\mathbf{k}_{\parallel}$ space. A consequence of the bands being nonparabolic and anisotropic is that the density of states are not a constant function of energy. Hence, we must get the density of states numerically. We can write D_n, the density of states in subband n, as,

$$D_n=\frac{2}{(2\pi)^2}\int_0^{2\pi}\frac{k}{|\nabla_k E|}\,d\theta. \tag{24}$$

The integration is carried out over $\mathbf{k}_{\parallel}$ space. This expression is then numerically multiplied by the Fermi distribution function and integrated over energy to obtain the subband occupations.

III. RESULTS AND DISCUSSION

The material parameters used in this work are given in Table I.[9–15] Material parameters for the alloy are not well documented in the literature over the wide range of composition used in our work. In the absence of such data, alloy material parameters are obtained by linear interpolation of the Si and Ge values. The elemental values are weighted by the mole fractions of the respective alloy contents. For the NMOS, the channel layer remains silicon-like up to large Ge content, as a result, we will use the values of effective masses of Si over the range of Ge composition used in our work.

The evaluation of our formalism for n-type and p-type MOSFETs yields the band profiles, the charge-density profile, and the energies and occupations of the two-dimensional subbands. This information can be obtained for any given layer structure. It is useful to have this kind of information available when designing optimized MOSFET structures.

In Fig. 1 we show a typical result of simulations for carrier density profiles in n- and p-type devices. This simulation was done for 300 K device operation. The oxide thickness and the background doping density are set to be 100 Å and 10^{16} cm^{-3}, respectively for all the simulations otherwise indicated. The figure shows the normalized quantized charge density profiles in the typical n- and p-type MOSFET devices. The sheet charge density is 4.5×10^{12} cm^{-2} for both NMOS and PMOS devices. From the one hump in the shape of the quantized charge density profile shown in the figure, it can be seen that the subbands states are highly spatially localized in the channel under the SiO_2–Si interface for both NMOS and PMOS. The peak of the quantized charge density

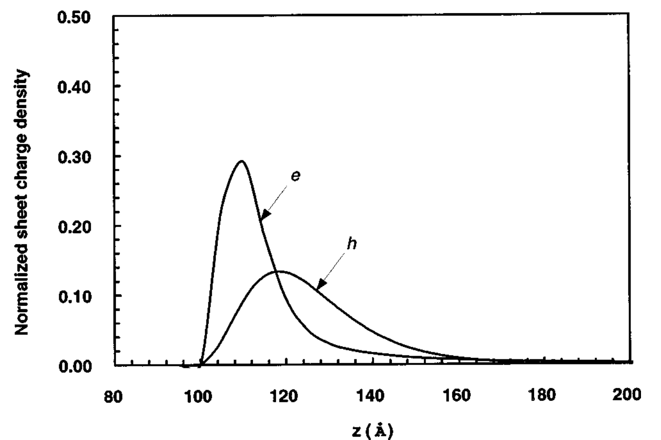
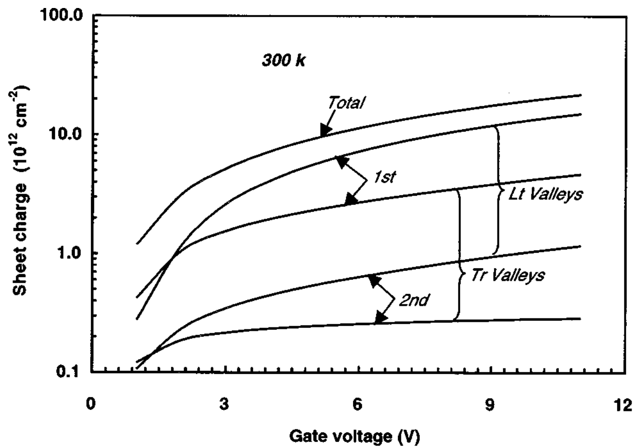
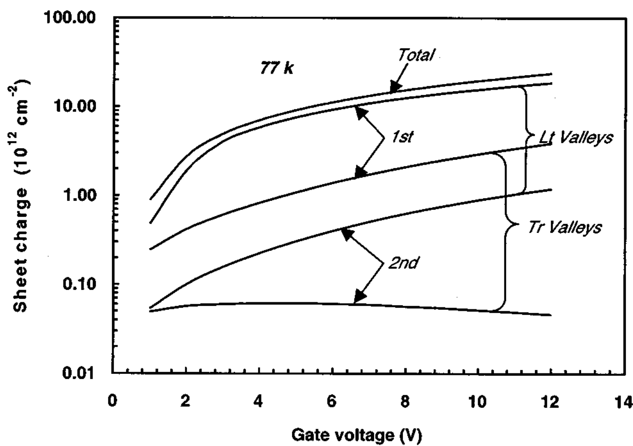


FIG. 1. Normalized quantized charge density profiles of typical n- and p-type MOSFETs at 300 K. The oxide thickness is 100 Å and the background doping density is 10^{16} cm^{-3}.



(a)



(b)

FIG. 2. (a) The total sheet charge density and occupation of the first and second subbands of transverse and longitudinal valleys vs gate bias at 300 K for a NMOS with 100 Å oxide thickness. (b) The total sheet charge density and occupation of the first and second subbands of transverse and longitudinal valleys vs gate bias at 77 K for a NMOS with 100 Å oxide thickness.

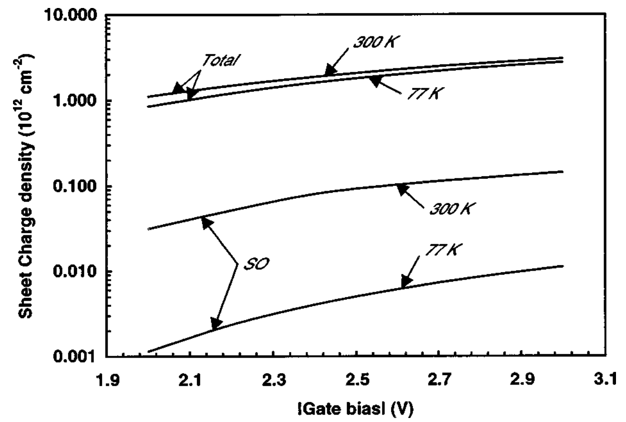


FIG. 3. The sheet charge density and occupation of split-off band vs gate bias at 300 and 77 K for a PMOS with 100 Å oxide thickness.

for NMOS (about 10 Å from the interface) is closer to the SiO_2–Si interface than that for PMOS. This indicates that the interfacial roughness effect, which will be presented later in this section, will be stronger for the NMOS than for the PMOS.

The channel effective mass, also called in-plane effective mass, is important for transport parallel to the layers. In our study, the channel effective mass is defined as follows;

$$m_{\mathrm{eff}}=\frac{\Sigma_i n_i m_i}{\Sigma_i n_i}, \tag{25}$$

where m_i and n_i are the density of states effective mass and the occupation of ith subband, respectively. An examination of the conduction band states for NMOS device shows that several of the lower subbands have a lighter in-plane effective mass. These subbands are formed by the electrons in longitudinal valleys which have larger z-directional effective mass, m_l and smaller in-plane effective mass, m_t. As one goes to the higher sheet charge (higher gate bias), the energy difference between the ground state due to electrons in the longitudinal valleys and the ground state due to electrons in the transverse valleys tends to increase. Thus we expect that the in-plane electron mass will become lighter and eventually saturate. In contrast, an examination of the valence band states in quantum wells shows that the top of the valence band has a relatively lighter mass. As one goes to the higher lying subbands, the mass tends to increase. Thus we expect that the in-plane hole mass will be lighter for lower sheet charge and for lower temperatures.

In Fig. 2(a) and 2(b) we show the sheet charge density as a function of gate bias for NMOS at 300 and 77 K. Also shown are the electron density in the first and second subbands of transverse and longitudinal valleys. For NMOS, we find that the electron density in the first and second subbands of the transverse and longitudinal valleys increases as the gate bias increases at 300 K. At low gate bias, the charge distribution in the subbands of transverse valleys exceeds the subbands of longitudinal valleys. This is because at low gate bias, the energy splitting between transverse valleys and longitudinal valleys caused by the quantum confinement is not large enough to balance the difference of degeneracy and effective mass. However, as gate bias increases, the energy splitting becomes larger and larger due to the increased quantum confinement and eventually it will balance out all the difference and the longitudinal valleys will become more occupied by the electrons. As seen from the figure, the ratio of occupation of the first of subband of the longitudinal valleys to the first subband of the transverse valleys increases with gate bias and approaches a value of 3. From this ratio, we can approximate the channel effective mass to be $\sim 0.25 m_0$. At low temperature, this ratio increases to a value of 7 then decreases slowly to a value of 5.

In Fig. 3 we show the sheet charge density as a function of gate bias at 300 K and at 77 K for PMOS. The hole density in SO band and the total sheet charge density in the channel with respect to the gate bias are shown. We find for the PMOS that as the gate bias increases, the occupation number of split-off hole states and the total sheet charge density increase. The ratio of the hole density in SO band to the total sheet charge density increases with gate bias. This indicates a stronger coupling between split-off hole and

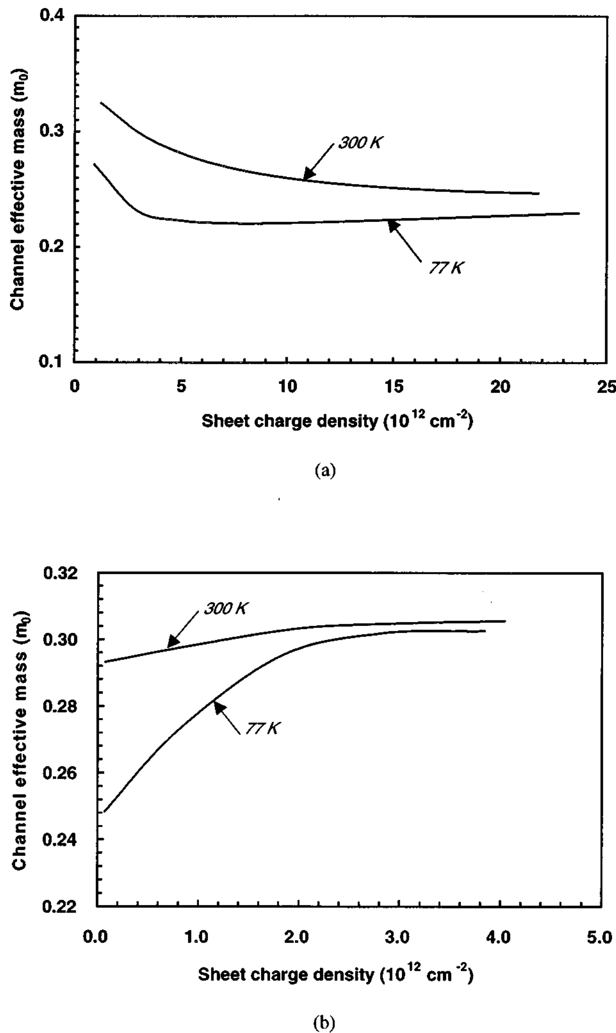


FIG. 4. (a) In-plane electron effective mass vs sheet charge density at 300 and 77 K in a NMOS. (b) In-plane hole effective mass vs sheet charge density at 300 and 77 K in a PMOS.

heavy-hole and light-hole as the gate bias increases. The fraction of holes in the split-off state increases from 3% to 5% for 300 K device operation and from 0.1% to 0.4% for 77 K operation as the gate bias goes from -2 to -3 V.

Our studies for PMOS suggest that at low temperatures, the contribution of sheet charge from the SO bands is fairly small even up to a large gate bias and thus a four-band model will be sufficient. However, for room-temperature device operation, the contribution from split-off band is not negligible. As the device temperature increases, the importance of the split-off band becomes even more important. Thus it is important to include the SO band for higher temperature studies of the PMOS charge control models.

The results for the in-plane effective mass for NMOS and PMOS are shown in Fig. 4(a) and 4(b), respectively. The simulations were done for 300 and 77 K device operation. As shown in Fig. 4(a), the channel effective mass for NMOS decreases with sheet charge density and approaches a value equal to $\sim 0.25 m_0$. At low temperature, as seen from the figure, the channel effective mass decreases much faster to a saturated value, $0.22 m_0$. When the sheet charge density increases above 5×10^{12}, the channel effective mass begins to increase slightly. This is due to the slightly increased charge redistribution in the first subband of the transverse valleys as noted in our discussion of Fig. 2.

As shown in Fig. 4(b), for PMOS the channel effective mass increases with sheet charge density. At low temperature, this effect is more pronounced. Note in our model, the effective mass is calculated explicitly from the valence bandstructure. This is achieved by solving the six-band Schrödinger equation self-consistently with the charge control model instead of *a priori* as is the case in models which do not include the coupling of the heavy-hole and light-hole bands.[5] We see that in-plane hole mass changes by about 25% at 77 K as the sheet charge increases from 1×10^{11} to 5×10^{12} cm^{-2}. The change for room temperature operation is not significant. The variation of hole masses with occupation has been reported in p-type MODFETs,[16,17] but there is no such experimental report on the MOSFET.

The effects of varying strain in the channel of an NMOS and a PMOS through addition of Ge are shown in Fig. 5(a) and 5(b). The left-most point on the curve is for a strain-free MOSFET with Ge concentration equal to zero. In each device, the SiGe well width is chosen to be 100 Å. As can be seen from the figure for NMOS, increasing the strain produces an increase in the channel effective mass. This is due to the increased splitting of the transverse valleys and longitudinal valleys caused by the strain in the pseudomorphic layer. This splitting raise the longitudinal valleys above the transverse valleys in energy. The magnitude of the splitting is given by Eq. (7) and Eq. (8). In the channel direction of the device, the transverse valleys are much heavier than the longitudinal valleys. Thus, the channel effective mass can be much higher than the normal electron effective mass.

In contrast, the effects of varying strain in the channel of a PMOS is much more complicated due to the nonparabolic and anisotropic nature of the valence bands. Figure 5(b) shows the channel effective hole mass as a function of Ge concentration. Unlike the case for the NMOS, increasing the strain yields a decrease in the channel effective mass. This is due to the increased splitting of the light-hole and heavy-hole subbands. Increasing strain bring about larger band splittings which cause the masses to become lighter. Thus, we can see significant decrease in hole effective mass which is brought about by the decoupling of the light-hole and heavy-hole states due to the strain in the system. At room temperature, for NMOS, our simulation gives channel electron effective mass as $0.22 m_0$ for the lattice-matched channel and $0.43 m_0$ for the strained channel with 40% Ge concentration. For PMOS, our simulation gives the channel hole effective mass as $0.31 m_0$ for the lattice-matched channel and $0.21 m_0$ for the strained channel with 40% Ge concentration. At lower temperature, the strain effect is even more pronounced.

In CMOS circuit design it is important to have a balanced performance for PMOS and NMOS in a n-well CMOS process. However since the mobility of electrons is almost as three times as that of holes, to balance the discrepancy, a usual approach is to increase the area of PMOS so that the channel current is comparable to that of NMOS. This leads to a symmetric voltage transient characteristic and balanced

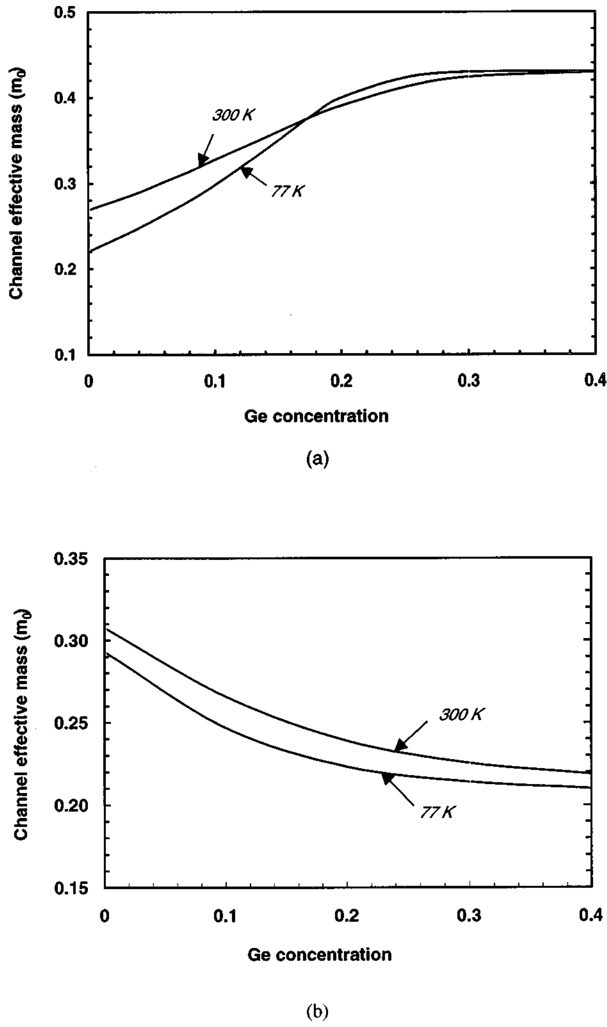


FIG. 5. (a) Channel electron effective mass as a function of Ge composition in the channel. The zero Ge device is lattice matched to the Si substrate. The width of strained channel is chosen to be 100 Å. The gate voltage is 4 V. (b) Channel hole effective mass as a function of Ge composition in the channel. The zero Ge device is lattice matched to the Si substrate. The width of strained channel is chosen to be 100 Å. The gate voltage is −4 V.

noise margins. Our study indicates that addition of Ge in the channel can increase the electron channel effective mass and decrease the hole channel effective mass and thus balance the transport performance of electrons and holes. Results of Fig. 5 suggest that one way to optimize digital circuits based on CMOS logic would be to add a small fraction of Ge (about 10%) in the NMOS and PMOS which may balance the NMOS and PMOS performance.

The performance of a MOSFET depends not only on the carrier effective mass and mobility at the active region but also on the quality of the interface formed by silicon and silicon dioxide. An important effect that has been studied extensively[18,19] is the surface scattering. It has been shown[18] that the surface roughness is the major scattering effect for a MOSFET device operating at higher sheet charge concentration ($\sim 10^{12}$ cm^{-2}). Due to the complexity of the problem a simple model has been proposed by Ando.[19] Various forms of this approach have been used in literature. The interface roughness potential is written as

$$\delta U(r,z) = U(z+\Delta(r)) - U(z) \sim \Delta(r)\,\frac{\delta}{\delta z}\,U = e\widetilde{E}(z)\Delta(r). \tag{26}$$

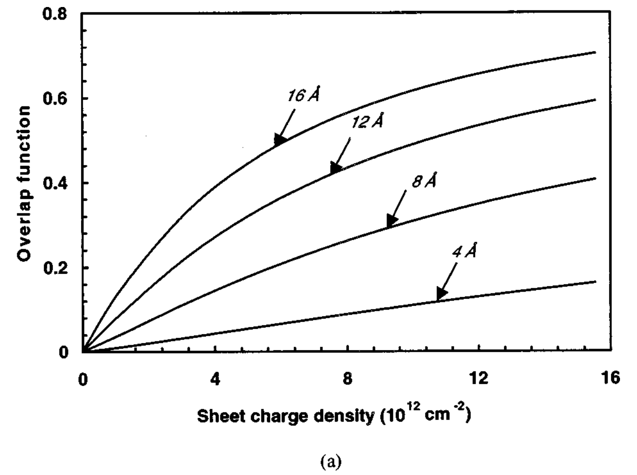
Here $\Delta(r)$ represents the position dependent height of the interface islands, $\widetilde{E}(z)$ represents the perpendicular electric field in the channel.

Assuming the maximum height of the islands to be Δ it is then shown that the z-direction matrix element for interface roughness scattering takes the form (using Gauss' law)

$$\langle \psi | \delta U | \psi \rangle \sim \int \psi^* e\widetilde{E}(z)\Delta\psi dz = e^2\Delta\,\frac{n_{\mathrm{depl}}+n_s/2}{\epsilon_s}. \tag{27}$$

Thus the matrix element is found to be proportional to n_s and the height of the islands. Note that in this simple form the scattering has no direct dependence on the carrier mass. The carrier mass dependence only comes from the in-plane integral of the density of states.

To study the accuracy of the model given above we evaluate the weighted z-direction overlap function



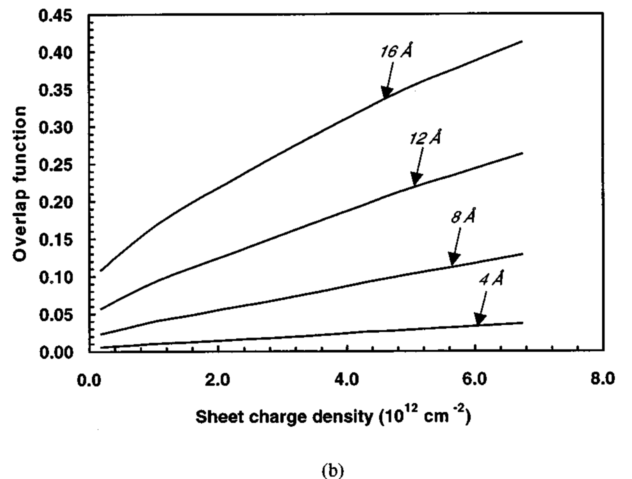


FIG. 6. (a) The weighted overlap function of conduction subbands vs sheet charge density at 300 K. Results are shown for a number of values of the parameter Δ. (b) The weighted overlap function of valence subbands vs sheet charge density at 300 K. Results are shown for a number of values of the parameter Δ.

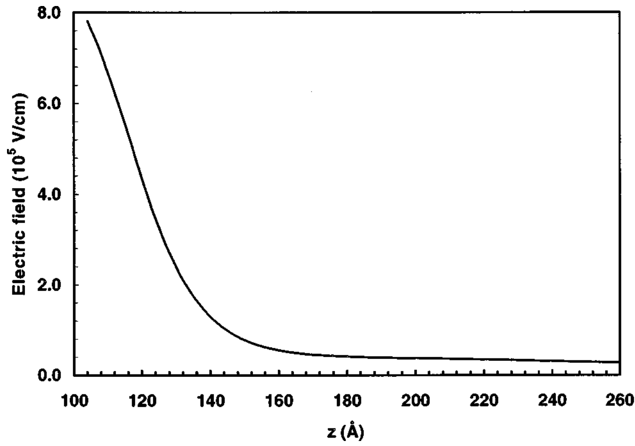


FIG. 7. The field profile of PMOS at 300 K. Result is shown for gate bias −4 V. The SiO_2–Si interface is at $z=100$ Å.

$$M(\Delta)=\frac{\Sigma_i \int_{z=0}^{z=\Delta} n_i \langle \psi_i | \psi_i \rangle dz}{\Sigma_i n_i}, \tag{28}$$

where Δ is the distance from the interface and ψ_i is the envelope function for the subband i. The Si–SiO_2 interface is assumed to be located at $z=0$. Our results are shown in Fig. 6(a) and 6(b) for NMOS and PMOS device. Several values of Δ are used to evaluated on the integrand. Both figures show that the probability or the fraction of the ground state from the interface is neither linear in the sheet charge density nor in the Δ. We also find that the z-directional matrix element calculated from the exact model is about an order of magnitude greater than the value given by the existing approximation. For example, we find that for $\Delta=4$ Å and $n_s=6\times10^{12}$ cm^{-2}, we get a value of 18.4 meV from Eq. (27) for the z-direction integral of the interface roughness scattering. However, from our numerical formalism we find a value of 0.21 and 0.14 eV for NMOS and PMOS, respectively. This can partly be understood on the basis of the variations of the field $E(z)$ in the channel. Figure 7 shows that under high gate bias, the magnitude of the electric field at the interface is about an order of magnitude higher (e.g., for $n_s=5\times10^{12}$ cm^{-2}, field at the interface is 8×10^5 V/cm and 10^5 V/cm 40 Å away from the interface) than that in the depletion region.

An important outcome of our finding is that since the matrix element is much larger than what one expects from simplisitic models, it is possible that Born approximation may not be valid in treating carrier transport limited by interface roughness. Detailed implications of this on mobility in MOS structures will be explored in a separate paper.

IV. CONCLUSION

In summary we have presented a charge control model to examine charge control and interface effects in NMOS and PMOS. We find the strain in the channel created by adding small fraction of Ge increases the channel effective mass of electrons and decreases the channel effective mass of holes which leads to a better balanced performance of NMOS and PMOS. We shows that the channel effective mass decreases with the sheet charge density due to the splitting of the transverse and longitudinal valleys and the difference of effective masses for NMOS and in contrast, we show that the channel effective mass for PMOS increases with sheet charge density due to the occupation of higher subbands which have relatively larger effective mass. We find that at low temperature the hole mass has an appreciable variation with gate bias. The percentage of the electron density in the first subband of longitudinal valleys is found to be ~75% of the total sheet charge density at 300 K and ~80% at 77 K under high gate bias. The hole density in the split-off band is found to be ~5% at 300 K. More importantly, we find that the matrix element for interface roughness scattering has significant departure from the linear model currently being used.

ACKNOWLEDGMENTS

This work was funded by US Air Force (Grant No. AFOSR-91-0349) and the US Army Research Office.

[1] D. Simitsky, F. Assaderaghi, C. Hu, and J. Bokor, IEEE Electron Device Lett. **18**, 54 (1997).
[2] J. M. Hinckley and J. Singh, Phys. Rev. B **41**, 2912 (1990).
[3] D. K. Nayak, J. C. S. Woo, J. S. Park, K. L. Wang, and K. P. MacWilliams, IEEE Electron Device Lett. **12**, 154 (1991).
[4] M. Jaffe and J. Singh, J. Appl. Phys. **65**, 329 (1989).
[5] C.-Y. Hu, S. Banerjee, K. Sadra, B. G. Streetman, and R. Sivan, IEEE Electron Device Lett. **17**, 276 (1996).
[6] J. Singh, *Physics of Semiconductors and their Heterostructures* (McGraw–Hill, New York, 1993).
[7] J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955).
[8] J. M. Luttinger, Phys. Rev. **102**, 1030 (1956).
[9] J. Singh, Ref. 6, Appendix L.
[10] P. Lawaetz, Phys. Rev. B **4**, 3460 (1971).
[11] L. M. Roth, B. Lax, and S. Zwerdling, Phys. Rev. **114**, 90 (1959).
[12] S. H. Groves, C. R. Pidgeon, and J. Femleib, Phys. Rev. Lett. **17**, 643 (1966).
[13] J. D. Wiley, Solid State Commun. **8**, 1865 (1970).
[14] H. J. McSkimin and P. Andreatch, J. Appl. Phys. **35**, 2161 (1964).
[15] H. J. McSkimin and P. Andreatch, J. Appl. Phys. **34**, 651 (1963).
[16] M. Jaffe, J. E. Oh, J. Pamulapati, J. Singh, and P. Bhattacharya, Appl. Phys. Lett. **54**, 2345 (1989).
[17] E. D. Jones, R. M. Biefeld, J. F. Klem, and S. K. Lyo, in *Proceedings of International Symposium on GaAs and Related Compounds*, 1989, p. 435.
[18] D. K. Ferry, International Electron Devices Meeting **84**, 605 (1984).
[19] T. Ando, J. Phys. Soc. Jpn. **51**, 3900 (1982).