

# **Solving the Effective Mass Schrödinger Equation in State-of-the Art Devices**

Dragica Vasileska

Ira A.Fulton School of Engineering, Department of Electrical Engineering

Arizona State University, Tempe, AZ 85287-5706, USA

e-mail: [vasileska@asu.edu](mailto:vasileska@asu.edu)

## **1 Introduction**

It has been known for many years that carriers in the inversion layer of a Si MOSFET are confined by the barrier between the semiconductor-oxide interface on one side and the band-banding of the conduction band on the other side. Since the average thickness of the inversion layer is comparable to the de Broglie wavelength of the electrons, this confinement is sufficient to produce quantization in the direction normal to the oxide-semiconductor interface. The space quantization effect is very important in determining the number of carriers in the inversion layer for devices with very high substrate doping, representative of the state-of-the art technology. In principle, the solution of even the equilibrium problem in these structures requires self-consistent solution of the Poisson and the Schrödinger equation using any available band-structure method. This is a difficult and time consuming problem, however. In many practical situations, it is sufficient to utilize the band-structure solvers to get the proper band-edge effective masses, which are then used in the time-independent Schrödinger equation for stationary potentials. One such tool that has been successfully utilized in the calculation of the energy level structure in simple MOS or dual-gate capacitor structures is SCHRED that has been developed at Arizona State University and Purdue University and is currently residing on the Purdue NanoHUB [1].

## 2 Description of SCHRED

SCHRED is a one-dimensional solver that solves self-consistently the 1D Poisson

$$\frac{\partial}{\partial z} \left[ \epsilon(z) \frac{\partial \varphi}{\partial z} \right] = -e \left[ N_D^+(z) - N_A^-(z) + p(z) - n(z) \right], \quad (1)$$

and the 1D one-electron effective-mass Schrödinger equation

$$\left[ -\frac{\hbar^2}{2m_i^\perp} \frac{\partial^2}{\partial z^2} + V_{eff}(z) \right] \psi_{ij}(z) = E_{ij} \psi_{ij}(z). \quad (2)$$

In Eqs. (1) and (2),  $\varphi(z)$  is the electrostatic potential,  $\epsilon(z)$  is the spatially dependent dielectric constant,  $N_D^+(z)$  and  $N_A^-(z)$  are the ionized donor and acceptor concentrations,  $n(z)$  and  $p(z)$  are the electron and hole densities,  $V_{eff}(z)$  is the effective potential energy term,  $m_i^\perp$  is the effective mass normal to the semiconductor-oxide interface of the  $i$ -th valley, and  $E_{ij}$  and  $\psi_{ij}(z)$  are the energy level and the corresponding wavefunction of the electrons residing in the  $j$ -th subband from the  $i$ -th valley. The effective potential energy term  $V_{eff}(z)$  in the 1D Schrödinger equation equals the sum of the Hartree [ $V_H = -e\varphi(z)$ ], image [ $V_{im}(z)$ ] and exchange-correlation [ $V_{exc}(z)$ ] terms. The Hartree term represents the solution of the 1D Poisson equation. The image term, which arises because of the different dielectric constants of the semiconductor and the oxide, is calculated from

$$V_{im}(z) = \frac{e^2}{16\pi\epsilon_{sc}z} \frac{\epsilon_{sc} - \epsilon_{ox}}{\epsilon_{sc} + \epsilon_{ox}}, \quad (3)$$

where  $\epsilon_{sc}$  and  $\epsilon_{ox}$  are the semiconductor and the oxide dielectric constants, respectively. Note that in the present version of SCHRED, the image term is implicitly included via the boundary conditions on the potential since the sheet electron density is no longer an input parameter but the gate voltage. That means that some of the experiments regarding the influence of the image term only on the subband separation can not be performed with the present version of SCHRED. When evaluating the exchange-correlation corrections to the chemical potential, we have relied on the validity of the density functional theory and the local density approximation outlined below.

## 2.1 Inclusion of Exchange-Correlation Corrections

In silicon inversion layers, due to the large effective mass, many-body effects such as exchange and correlation can play an important role. For example, Stern [2] has calculated that the exchange energy is comparable to or even larger than the energy separation between subbands calculated in the Hartree approximation. In general, the exchange energy is the contribution to the overall energy of the electron gas that arises from the correlation between two electrons whose positions are reversed, or exchanged [3]. In other words, as a consequence of the Pauli exclusion principle, the electrons with equal spin tend to avoid each other (exchange repulsion) so that each electron is surrounded by an *exchange hole*. The presence of the exchange hole indicates that the mean separation between electrons with equal spin is larger than it would be without the Pauli principle. The existence of the exchange hole reduces the overall Coulomb repulsion which explains the reduction in the ground-state energy of the system.

According to the Hartree-Fock theory, electrons with different spin do not avoid each other, since the states are chosen to satisfy the exchange principle, but they do not include Coulomb correlations [4]. In reality, there exists an additional correlation, which leads to the so-called *Coulomb hole*. To treat these effects one has to go beyond the Hartree-Fock theory. Therefore, if one writes the exact ground state energy of the system as

$$E = E^{HF} + E_{corr} = E_{kin}^{HF} + E_{exc}^{HF} + E_{corr} , \quad (4)$$

it is obvious that the correlation energy represents the correction to the ground-state energy of the system beyond the Hartree-Fock approximation. Therefore, the correlation energy is not a quantity with physical significance; it merely represents the error incurred in making a fairly crude first-order approximation. Since an exact calculation of  $E_{corr}$  is generally not possible, one of the main tasks of the many-body theory is to obtain a good estimates for  $E_{corr}$ .

In a series of three papers Hohenberg and Kohn [5], Kohn and Sham [6], and Sham and Kohn [7] have laid the foundations for a “new” theory of electronic structure. The theory represents a systematic extension of the Thomas-Fermi ideas, and is capable in principle of providing exact answers. The theory is based on two theorems which center on the particle density as a fundamental variable for the description of any many-body system. The first theorem states that the total ground-state energy  $E$  of any many-body system is a functional of the one-particle density  $n(\mathbf{r})$ . In this context, different many-body systems differ only by the local potential felt by the electrons.

Furthermore, splitting off from the total energy the explicit interaction with the external potential  $V_{ext}(\mathbf{r})$ , the theorem also states that the rest is a universal functional of  $n(\mathbf{r})$ , i.e. independent of the external potential. Thus, if

$$E[n] = F[n] + \int d^3r V_{ext}(\mathbf{r})n(\mathbf{r}) , \quad (5)$$

then the functional  $F$  depends only on  $n$  and not on  $V_{ext}(\mathbf{r})$ . The second theorem states that for any system (any external potential) the functional  $E[n]$  for the total energy has a minimum equal to the ground-state energy at the physical ground-state density of the system. These theorems, although rather abstract in nature were of immense importance to the rapid development of density-functional theory. It is customary to extract from  $F[n]$  the classical Coulomb energy and write

$$G[n] = F[n] - \frac{1}{8\pi\epsilon_0} \int d^3r \int d^3r' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (6)$$

In this notation, the energy functional becomes

$$E[n] = G[n] + \int d^3r V_{ext}(\mathbf{r})n(\mathbf{r}) + \frac{1}{8\pi\epsilon_0} \int d^3r \int d^3r' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} . \quad (7)$$

The stationary functional  $E[n]$  allows in principle a much simpler determination of the ground-state energy  $E$  and density  $n(\mathbf{r})$  than the conventional Rayleigh-Ritz method [8]. The functional  $G[n]$  is further divided into two parts

$$G[n] = T_s[n] + E_{xc}[n] , \quad (8)$$

where  $T_s[n]$  is the kinetic energy of a non-interacting electron gas of density  $n(\mathbf{r})$  in its ground state and  $E_{xc}[n]$  represents the exchange and correlation energy. With these new quantities, we can write

$$E[n] = T_s[n] + \int d^3r V_{ext}(\mathbf{r})n(\mathbf{r}) + \frac{1}{8\pi\epsilon_0} \int d^3r \int d^3r' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + E_{xc}[n] . \quad (9)$$

The energy functional given in Eq. (9) has to be minimized with respect to the electron density  $n(\mathbf{r})$  subject only to the normalization condition

$$N = \int d^3r n(\mathbf{r}) , \quad (10)$$

where  $N$  is the total number of electrons in the system under consideration. The standard method of taking care of the constraint given in Eq. (10) is to make the variational principle read

$$\delta(E - \mu N) = 0 , \quad (11)$$

where  $\mu$  is a Lagrange multiplier. Carrying out the variation, one obtains the Euler condition

$$\mu = \frac{\delta T_s[n]}{\delta n(\mathbf{r})} + V_{ext}(\mathbf{r}) + \frac{1}{4\pi\epsilon_0} \int d^3r' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + V_{xc}(\mathbf{r}) , \quad (12)$$

where

$$V_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})} \quad (13)$$

is the exchange-correlation potential. The variational derivative of the kinetic energy  $\delta T_s[n]/\delta n(\mathbf{r})$  is then replaced with the kinetic operator  $-\hbar^2 \nabla^2 / 2m^*$ . At this point, Kohn and Sham [6,7] make a crucial observation that the Euler Eq. (12) is the Euler equation of non-interacting particles subject to the effective external potential  $V_{eff}(\mathbf{r})$ , given by

$$V_{eff}(\mathbf{r}) = V_{ext}(\mathbf{r}) + \frac{1}{4\pi\epsilon_0} \int d^3r' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + V_{xc}(\mathbf{r}) = V_H(\mathbf{r}) + V_{xc}(\mathbf{r}) , \quad (14)$$

where the Hartree potential is obtained from the solution of the corresponding Poisson equation. This scheme allows one to construct an equivalent one-particle formulation of the complicated many-body problem at hand.

The exchange-correlation energy  $E_{xc}[n]$  is in general an unknown functional of the electron density. However, for slowly varying density, one can make the local-density approximation (LDA)

$$E_{xc}[n] \approx \int d^3r \epsilon_{xc}(n(\mathbf{r})) n(\mathbf{r}) , \quad (15)$$

where  $\epsilon_{xc}(n(\mathbf{r}))$  is the exchange and correlation energy per electron of a uniform electron gas with

density  $n_0 = n(\mathbf{r})$ . The original idea for this approximation comes from Slater [9]. Using LDA, we find

$$V_{xc}(\mathbf{r}) \approx \frac{d(n\varepsilon_{xc}(n))}{dn} = \varepsilon_{xc}(n) + n \frac{d\varepsilon_{xc}(n)}{dn}, \quad (16)$$

which, according to the Seitz theorem [10], is equivalent to the definition of the chemical potential. Therefore,  $V_{xc}(\mathbf{r})$  can be interpreted as the exchange-correlation contribution to the chemical potential of a homogeneous electron gas of density  $n_0$  equal to the local electron density  $n(\mathbf{r})$  of the inhomogeneous system. As pointed out by Kohn and Vashishta [11], the LDA works surprisingly well in calculating the electronic structure of confined electronic systems where the electron density

$$n(\mathbf{r}) = n(z) = \sum_i N_i |\psi_i^2(z)| \quad (17)$$

is not slowly varying in space. The exchange-correlation potential  $V_{xc}$  for LDA has been parametrized by many authors. A standard parametrized form due to Hedin and Lundqvist [12,13,14,15], used also in SCHRED, is

$$V_{xc}(z) = -\frac{e^2}{4\pi^2\varepsilon_0} [3\pi^2 n(z)]^{1/3} \left[ 1 + 0.7734x \ln \left( 1 + \frac{1}{x} \right) \right], \quad (18)$$

where  $x = x(z) = r_s/21$ ,  $r_s = r_s(z) = [4\pi b^3 n(z)/3]^{-1/3}$  and  $b = 4\pi\varepsilon_{sc}\hbar^2/m^*e^2$ . The first term on the RHS of Eq. (18) is the exchange energy correction due to the attractive interaction between other electrons and the Fermi hole resulting from the displaced charge. The second term represents the correlation energy correction to the chemical potential  $\mu$ . Using this parametrized expression for  $V_{xc}(z)$ , one calculates the electronic subband wavefunctions and the corresponding subband energies by solving the so-called Hohenberg-Kohn-Sham (HKS) equation, which is formally the same as the Schrödinger equation in which, as already noted, one takes  $V_{eff}(z) = V_H(z) + V_{xc}(z) + V_{im}(z)$ . In such a calculation, one obtains not only the total energy and the electron density, but also the eigenvalues of the KS equations [16]. For silicon inversion layers, by analogy to the spin-density formalism [17], the exchange-correlation correction to the chemical

potential is different for the unprimed and primed valleys, and it depends only on the volume density of electrons in the unprimed and primed subbands.

The extension of this formalism for nonzero-temperatures was formally set up by Mermin [18], and the finite temperature exchange-correlation functions that enter the Kohn-Sham-Mermin formulation were calculated by Gupta *et al.* [19,20]. The finite-temperature exchange correction to the chemical potential calculated by Gupta *et al.* is

$$\frac{V_{exc}(\mathbf{r}, T)}{V_{exc}(\mathbf{r}, T = 0)} \approx \frac{2}{3} \left( \frac{T_F}{T} \right), \quad (19)$$

where the Fermi temperature  $T_F$  is defined in terms of the zero-temperature variables  $k_F = [3\pi^2 n(\mathbf{r})]^{1/3}$ ,  $E_F = \hbar^2 k_F^2 / 2m^*$  and  $k_B T_F = E_F$ . The result given in Eq. (19) is valid for Boltzmann statistics. In the Debye limit, the corresponding correlation energy correction is given by

$$V_{corr}(\mathbf{r}, T) = -\frac{e^2}{8\pi\epsilon_0} \sqrt{\frac{e^2 n(\mathbf{r})}{\epsilon_0 k_B T}}. \quad (20)$$

From the results given in Eqs. (19) and (20) it is obvious that the correlation contributions can still be important for temperatures where the exchange contribution has become vanishingly small.

A comparison of the calculated self-consistent potentials for (100) *p*-type Si with  $N_a = 2.8 \times 10^{15} \text{ cm}^{-3}$  (corresponding to  $N_{depl} = 2.02 \times 10^{11} \text{ cm}^{-2}$ ),  $N_s = 4 \times 10^{12} \text{ cm}^{-2}$  and  $T = 0 \text{ K}$ , with (thick lines) and without (thin lines) the inclusion of the exchange-correlation correction to  $V_{eff}(z)$ , is given in Ref. [21] and also shown on Figure 1. We observe that the subband energies are lowered considerably by the exchange-correlation effect. The energy of the ground subband is lowered by 35 meV, whereas the energy of the first excited subband is lowered by 20 meV, which is in agreement with results obtained by Vinter [22]. Since the inclusion of the exchange-correlation effects increases the subband separation, this many-body correction leads to an increase of the carrier concentration at which the occupation of the second subband begins. We also find that, in contrast to the image term which tends to increase the spatial extent of the wavefunctions, the exchange-correlation term tends to compress the wavefunctions.

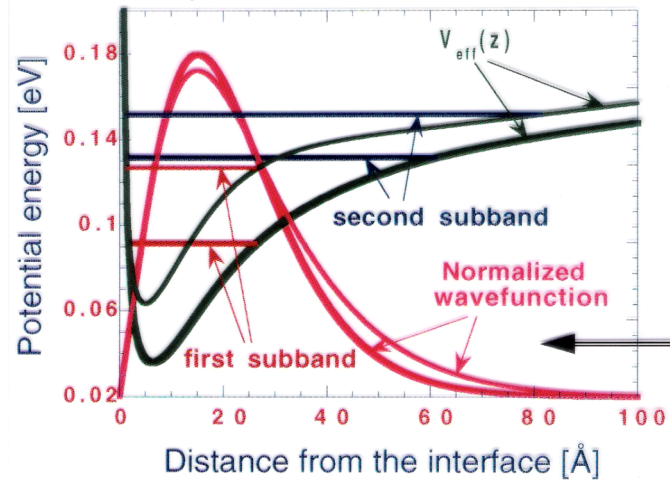


Figure 1 Calculated potential energy profile, subband structure and normalized wavefunctions for (100) *p*-type Si with  $N_a = 2.8 \times 10^{15} \text{ cm}^{-3}$ ,  $N_s = 4 \times 10^{12} \text{ cm}^{-2}$  and interface-trap density  $N_{it} = 9.5 \times 10^{10} \text{ cm}^{-2}$ . The thick (thin) lines correspond to the case when the exchange-correlation effect is included (omitted) in the simulation

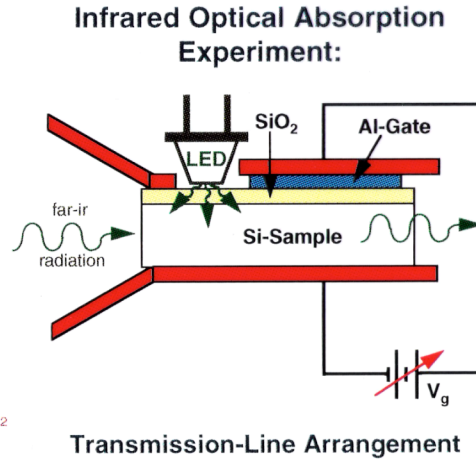
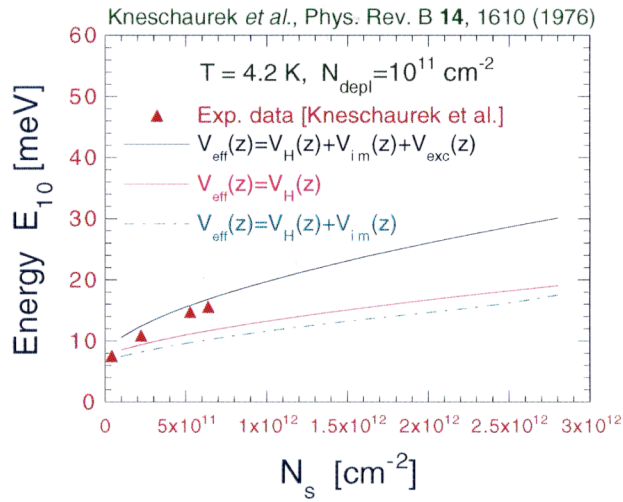


Figure 2. Density dependence of the separation of the  $\epsilon_1$  and  $\epsilon_0$  subbands in a (100) *p*-type silicon inversion layer. The filled triangles represent the infrared absorption measurements.

In Figure 2 we compare our simulation results for the energy spacing between the lowest two subbands (subbands  $\epsilon_0$  and  $\epsilon_1$  from the unprimed ladder) with the infrared absorption measurements of Kneschaurek *et al.* [23] on a *p*-type Si(100) at  $T = 4.2$  K. The doping



concentration is  $N_a - N_d = (2 \pm 0.2) \times 10^{15} \text{ cm}^{-3}$ . The experimental data shown in the figure represent the *dark sweep* spectroscopy results. For these experimental conditions, the depletion layer length and the depletion charge density do not reach their thermal equilibrium values and the measured value of the experimentally relevant effective depletion charge density is  $N_{depl}^* = (1.0 \pm 0.1) \times 10^{11} \text{ cm}^{-2}$ . To be in agreement with the experimental conditions, we take  $N_{depl} = 1 \times 10^{11} \text{ cm}^{-2}$ . We find that the experimental data show a faster increase in the level splitting than the Hartree theory (with and without the image term). The inclusion of the exchange-correlation correction to the chemical potential in the Kohn-Sham equation significantly improves the situation, especially at higher inversion charge densities (In the space-charge layer, the condition of slowly varying potential translates into  $N_s \gg N_{depl}^*$ ). It is believed that the so-called *excitonlike* and *depolarization* corrections nearly cancel each other except at very high electron concentrations [24,25]. (The exciton shift is the interaction of the excited electron with the hole in the ground state, analogous to the exciton associated with the valence-to-conduction band transition. The depolarization shift is a plasmon shift of the transition caused by the screening response of the electron gas.) Our simulation results for the energy spacing are in excellent agreement with those obtained by Ando [26].

In principle, the finite-temperature extension of the density-functional theory presented in this section is obtained by using the finite-temperature expressions for the exchange and correlation corrections to the chemical potential and through the change of the occupancies of various subbands. However, Das Sarma and Vinter [27,28] have shown that neglecting any temperature dependence of the exchange-correlation potential, but retaining its implicit temperature dependence through the electron density  $n(z)$ , which is calculated at finite temperature, leads to results which are in very good agreement with the measured subband separations, especially the ones for the unprimed ladder. To check this argument, we calculated the subband separation of a *p*-type Si with effective depletion charge density  $N_{depl}^* = 6 \times 10^{10} \text{ cm}^{-2}$  and (100) orientation of the surface at  $T=300$  K using first the parametrized expression given in Eq. (18) and then the finite-temperature results for the exchange-correlation corrections to the chemical potential given in Eqs. (19) and (20). The difference in the calculated subband energies for various inversion charge densities  $N_s$  was found to be always less than 5%, even for the worst case. The simulation results for the subband separations  $\varepsilon_{10}$  and  $\varepsilon_{1'0}$ , and various inversion charge densities, for the same sample, are shown in

Figure 3 and

Figure 4, respectively. The filled triangles in both figures represent the room-temperature infrared resonant absorption measurements due to Schäffler *et al.* [29]. It is believed that the net correction to the subband separation due to depolarization and exciton-like shifts is less than 4%. A total of 10(5+5) and 5(3+2) subbands was used in these simulations. We observe that the use of 10 instead of 5 subbands leads to the increase in the subband separations in both cases throughout the whole range of  $N_s$ . However, this increase is more pronounced for the primed ladder of subbands. For comparison, in both figures we also give the Hartree results for the subband separation. We see that the Hartree approximation becomes a better approximation for the subband energy difference at

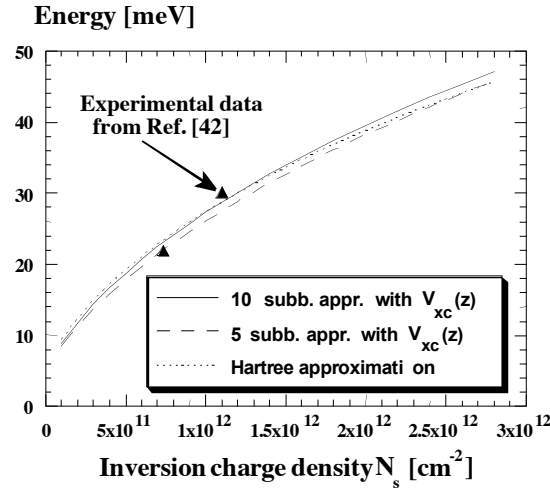


Figure 3. Subband energy difference  $\varepsilon_{10}$  vs. inversion charge density at  $T=300$  K.

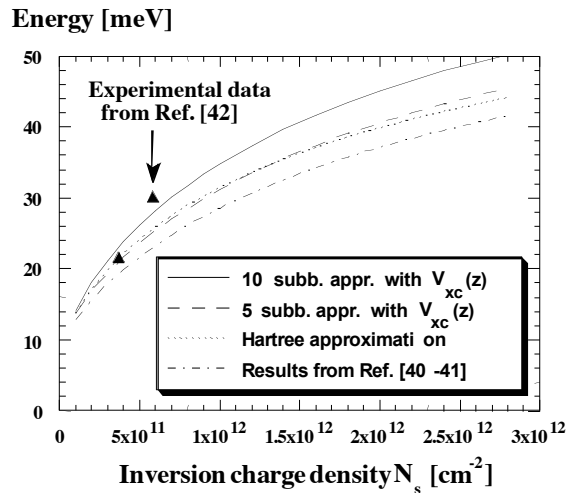


Figure 4. Subband energy difference  $\varepsilon_{1'0}$  vs. inversion charge density at  $T=300$  K.

elevated temperatures due to the decrease of the exchange energy correction to the chemical potential. Our simulation results for the subband energy difference for the unprimed ladder of subbands are in agreement with Ref. [27,28]. However, for the primed ladder of subbands, we are in better agreement with the experimental data compared to the results of Das Sarma and Vinter. The major difference comes from the fact that they use 5 instead of 10 subbands as well as the conductivity instead of the density-of states mass.

## 2.2 Other Simulator Details

In all the calculations presented here, we assume that the  $SiO_2/Si$  interface is parallel to the [100] plane. For this particular case, the six equivalent minima of the bulk silicon conduction band split into two sets of subbands (Figure 5).

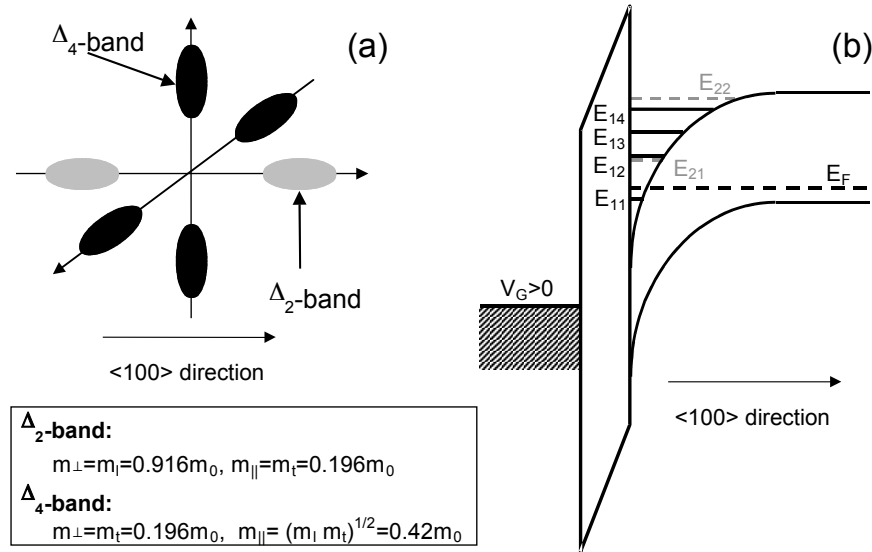


Figure 5. (a) Constant energy surfaces in Si together with the description of the  $\Delta_2$ - and  $\Delta_4$ -bands. Also shown are the appropriate transverse and in-plane masses for the two equivalent bands. (b) Schematics of the band-bending in MOS capacitors. Also shown are the energy levels belonging to the unprimed and primed ladder of subbands corresponding to the  $\Delta_2$ - and  $\Delta_4$ -bands, respectively. The first index describes the band ( $=1$  for the  $\Delta_2$ -band, and  $=4$  for the  $\Delta_4$ -band), whereas the second one refers to the appropriate energy level within the band.

The first set ( $\Delta_2$ -band) consists of the two equivalent valleys with in-plane effective mass  $m_{\parallel}=0.19m_0$  and perpendicular effective mass  $m_{\perp}=0.91m_0$ . The second set ( $\Delta_4$ -band) consists of the

four equivalent valleys with  $m^{\parallel}=0.42m_0$  and  $m^{\perp}=0.19m_0$ . The energy levels associated with the  $\Delta_2$ -band comprise the so-called *unprimed* ladder of subbands, whereas those associated with the  $\Delta_4$ -band comprise the *primed* ladder of subbands.

The self-consistent solution of the 1D Schrödinger-Poisson problem is obtained in the following way: We start with some initial guess for the electrostatic potential and use it to solve the 1D Schrödinger equation numerically [8]. After we determine the eigenfunctions and the eigenvalues that characterize the electrons in the inversion layer, the inversion layer electron density appearing in the 1D Poisson equation is obtained by summing over all subbands to get

$$n(z) = \sum_{i,j} N_{ij} \psi_{ij}^2(z) = \sum_{i,j} g_i \frac{m_i^{\parallel} k_B T}{\pi \hbar^2} \ln \left[ 1 + \exp \left( \frac{E_F - E_{ij}}{k_B T} \right) \right] \psi_{ij}^2(z). \quad (21)$$

In Eq. (21),  $E_F$  is the Fermi level,  $k_B$  is the Boltzmann constant,  $T$  is the temperature,  $m_i^{\parallel}$  is the in-plane effective mass of the  $i$ -th band,  $N_{ij}$  is the sheet-charge density corresponding to the  $j$ -th subband from the  $i$ -th band, and  $g_i$  ( $g_1=2$  for the  $\Delta_2$ -band, and  $g_2=4$  for the  $\Delta_4$ -band) is the band degeneracy. It is important to note that the inversion layer electrons are treated quantum-mechanically only when confined by the surface-field. If otherwise, or if we relied on the validity of the classical description of the inversion layer electron density, we skip the solution of the 1D Schrödinger equation and use

$$n(z) = N_C F_{1/2} \left[ \frac{E_F - E_C(z)}{k_B T} \right], \quad (22)$$

where  $N_C$  is the effective density of states of the conduction band. For holes, which are always treated classically for p-type substrates, we use

$$p(z) = N_V F_{1/2} \left[ \frac{E_C(z) - E_G - E_F}{k_B T} \right], \quad (23)$$

where  $N_V$  is the effective density of states of the valence band, and  $E_G$  is the semiconductor bandgap. For the evaluation of the Fermi-Dirac integrals, which appear in Eqs. (22) and (23), we use the analytical approximation due to Bednarczyk and Bednarczyk [30]

$$F_{1/2}(x) = \left[ e^{-x} + 3\sqrt{\pi}/4v^{3/8} \right]^1, \quad (24)$$

where

$$v(x) = x^4 + 50 + 33.6x \left\{ 1 - 0.68 \exp \left[ -0.17(x+1)^2 \right] \right\}. \quad (25)$$

The poly-silicon gates are modeled as heavily-doped single-crystal silicon. Both the electrons and holes are treated classically and assuming general Fermi-Dirac statistics, valid for degenerate semiconductors.

After we update the electron and hole concentrations in the semiconductor and/or the poly-silicon gates, we resolve numerically the 1D Poisson equation for the electrostatic potential using finite-difference discretization scheme and LU decomposition method. We then solve the 1D Schrödinger equation to find the updated values for the electron density at each mesh point, and repeat the above-described procedure until self-consistent solution is found. It is important to note that the potential energy profile for the next iteration is obtained by using fixed-convergence factor scheme for the first 10 iterations and the extrapolated convergence-factor scheme thereafter. The error criterion for the convergence of the self-consistent field iterations is that the absolute value of the difference between the input and output potentials at each mesh point is less than 0.01 mV.

At self-consistency, i.e. once we determine the self-consistent results for the variation of the charge distribution on the semiconductor side of the MOS capacitor as a function of the gate voltage  $V_G$ , we proceed with the calculation of the total gate capacitance  $C_{tot}$ . We determine  $C_{tot}$  by differentiating the total induced charge density in the channel with respect to  $V_G$ . In contrast to some previous studies [31], where  $C_{inv}$  was approximated with  $\epsilon_{sc} / \langle z \rangle_{av}$ , where  $\langle z \rangle_{av}$  is the centroid of the electron density distribution, here we calculate the inversion layer capacitance by differentiating the total sheet charge density

$$N_s = \sum_{i,j} N_{ij} \quad (26)$$

with respect to the surface potential [32]. The depletion-layer capacitance  $C_{depl}$  and the poly-gate capacitance  $C_{poly}$  are evaluated in an analogous manner.

### 2.3 Some Sample Simulation Results Obtained with SCHRED

To demonstrate the existence of the two physical origins of the inversion layer capacitance  $C_{inv}$ , discussed in Ref. [33], in Figure 6 we show the variation of  $C_{inv}$  with inversion charge density  $N_s$  in the channel of a MOS capacitor with substrate doping  $N_A = 5 \times 10^{17} \text{ cm}^{-3}$ , oxide thickness  $t_{ox} = 4 \text{ nm}$  and metal gates. Exchange-correlation and image contributions to the effective potential

energy term  $V_{eff}(z)$ , appearing in the 1D Schrödinger equation, have not been included in these simulations.

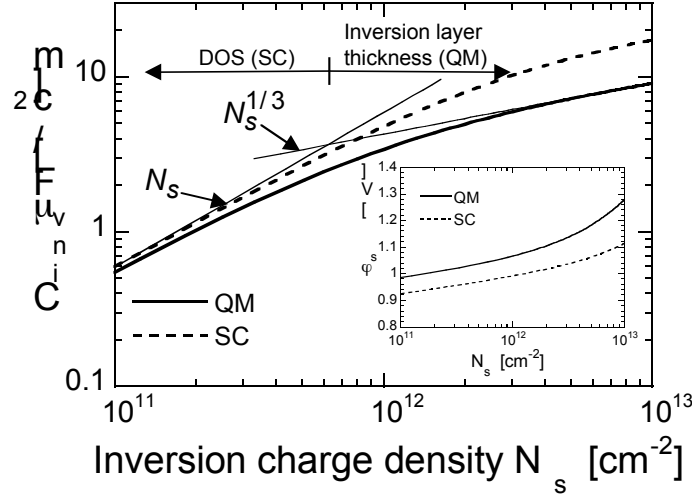


Figure 6. Variation of the inversion layer capacitance with inversion charge density at  $T=300$  K. In the inset we show the self-consistent results for the variation of the surface potential with  $N_s$  when using both SC and QM description of the electron density in the inversion layer. The surface potential is calculated using  $\varphi_s = (E_F - E_i)_{sc/ox} - (E_F - E_i)_{bulk}$ , where  $E_i$  is the intrinsic energy level.

The pronounced double-slope behavior of the quantum-mechanically calculated  $C_{inv}$  comes from the fact that the total inversion layer capacitance can be represented as a series capacitance of two contributions. The first contribution is classical and comes from the finite density of states, i.e. due to the fact that a finite change in the surface potential is always necessary to increase  $N_s$  (inset of Figure 6), which, in turn, leads to finite value for  $C_{inv}$ . This term dominates at low values of  $N_s$  (low gate voltages). The second contribution to  $C_{inv}$  is due to the finite inversion layer thickness, which effectively increases the oxide thickness in terms of the total gate capacitance, thus providing an additional capacitance component. This term dominates at large gate voltages, where the inversion charge density  $N_s$  significantly influences the band bending and leads to a steeper rise of the conduction band near the  $SiO_2/Si$  interface.

In Figure 7, we show simulated  $C_{tot}$  to oxide capacitance  $C_{ox}$  for metal/ $p$ -substrate and  $n^+$ -poly/ $p$ -substrate MOS capacitors, as a function of the physical oxide thickness  $t_{ox}$  and the doping of the polysilicon gates  $N_D$ , assuming  $V_G=3$  V. The high value for  $V_G$ , used here, may overestimate

the severity of the bias dependent attenuation for thinner oxides, but a consistent value for  $V_G$  is useful for the purpose of tabulating the simulated results. The results shown clearly demonstrate that classical charge model and Maxwell-Boltzmann (non-degenerate) statistics are clearly inadequate for oxide thickness below 10 nm. Even use of Fermi-Dirac statistics in the classical charge description can lead to significant errors in the estimate of the total gate capacitance for devices with metal gates and oxide thickness less than 5 nm, due to the higher surface fields and, therefore, pronounced quantum-mechanical size-quantization effect in the channel. For example, the classical model that uses Maxwell-Boltzmann (Fermi-Dirac) statistics predicts that, for the device with  $t_{ox} = 1$  nm,  $C_{tot}/C_{ox} = 0.983$  (0.882). On the other hand, the quantum-mechanical model predicts that  $C_{tot}/C_{ox} = 0.795$ , which leads to relative error of 23.65 (10.94) %. As previously noted, the depletion of the poly-silicon gates will further degrade the total gate capacitance.

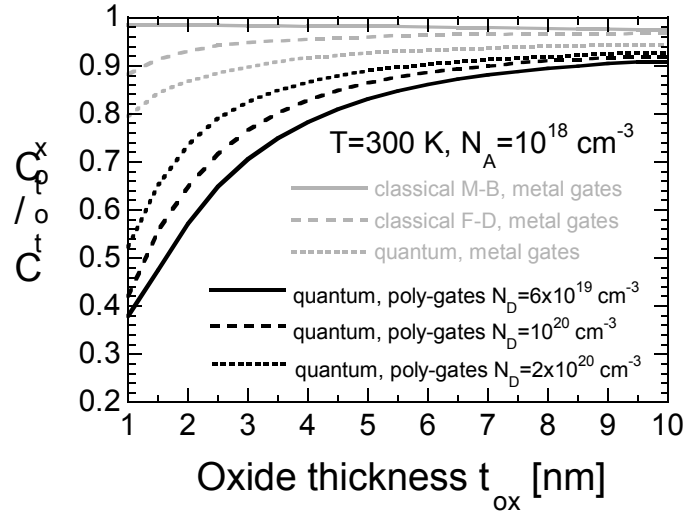


Figure 7. Simulated  $C_{tot}$  to oxide capacitance  $C_{ox}$  for metal/ $p$ -substrate and  $n^+$ -poly/ $p$ -substrate MOS capacitors, as a function of the physical oxide thickness  $t_{ox}$  and the doping of the polysilicon gates  $N_D$ . We use  $V_G = 3 \text{ V}$ .

The linear region threshold voltage shift between the QM and SC predictions for a device with  $N_A = 5 \times 10^{17} \text{ cm}^{-3}$  and  $t_{ox} = 4 \text{ nm}$  as a function of the doping of the poly-silicon gate is shown in Figure 8. The threshold voltage  $V_{th}$  equals the gate voltage for which  $Q_{inv} = 10^{-3} Q_{depl}$ . As expected, the QM description of the charge in the channel increases  $V_{th}$  and the shift in the threshold voltage is about 74 mV. This is due to the fact that the QM picture differs from the SC one in two ways: First, the energy spectrum is not continuous, but consists of discrete energy levels which, in turn, reduces

the DOS function. Second, the energy of the ground subband from the unprimed ladder of subbands does not coincide with the bottom of the conduction band and the energy difference  $\Delta E = E_{11} - E_C$  increases with increasing substrate doping. The depletion of the poly-silicon gate, due to insufficient doping, further increases the threshold voltage. The additional shift in the threshold voltage due to the inclusion of the poly-gate depletion can be as large as 68 mV for  $N_D = 10^{19} \text{ cm}^{-3}$ , and drops down to about 18 mV for  $N_D = 2 \times 10^{20} \text{ cm}^{-3}$ .

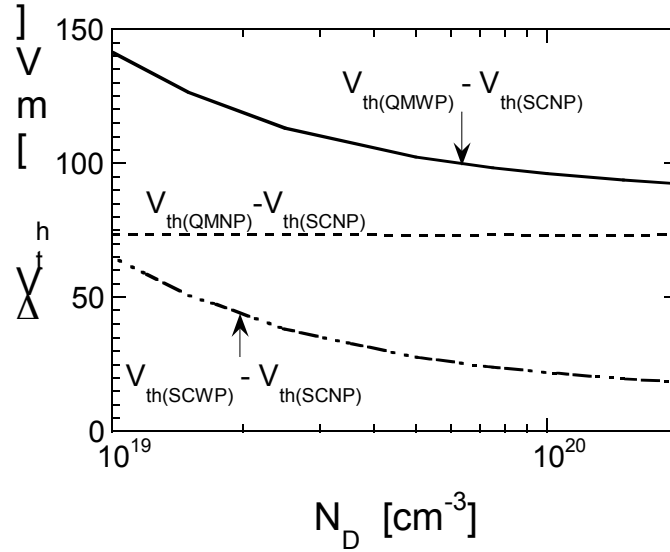


Figure 8. Linear region threshold voltage shift between the QM and the SC predictions versus  $N_D$ . We use  $N_A = 5 \times 10^{17} \text{ cm}^{-3}$  and  $t_{ox} = 4 \text{ nm}$ .

The linear region threshold voltage shift for the device with  $t_{ox} = 4 \text{ nm}$ ,  $N_D = 10^{20} \text{ cm}^{-3}$ , and different substrate doping is shown in Figure 9. Also shown in this figure are the van Dort *et al.* [34] experimental data for a device with metal gates and oxide thickness  $t_{ox} = 14 \text{ nm}$ . Very close agreement between the experimentally derived threshold voltage shifts and our simulation results for the device with 14 nm thick oxide can be observed. A major difference from the results shown in Figure 8 is that the inclusion of both the QM effects in the channel and poly-gate depletion leads to strong dependence of the threshold voltage shift upon the substrate doping  $N_A$ . For example, for a device with  $t_{ox} = 4 \text{ nm}$ ,  $N_A = 10^{18} \text{ cm}^{-3}$  and  $N_D = 10^{20} \text{ cm}^{-3}$ , the inclusion of the quantum-mechanical space-quantization effect leads to a threshold voltage shift of about 106 mV. The addition of poly-gate depletion leads to a further shift in the threshold voltage of about 34 mV. This observation, together with the results shown in Figure 9, suggests that both a QM description of the



charge density distribution in the channel and poly-gate depletion must be accounted for if accurate results for the threshold voltage are desired.

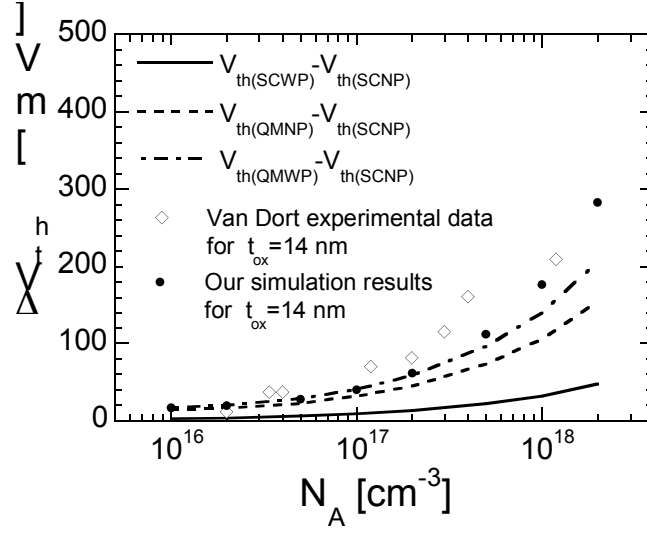


Figure 9. Linear region threshold voltage shift between the QM and the SC predictions versus  $N_A$ .

### 3 Modification of the Effective Mass Schrödinger Equation for Heterostructures

Note that in a solid in general, the momentum space is periodic and the true wavefunction is approximately the product of a periodic Bloch function and an envelope function. The Schrödinger equation can be used to study the evolution of the envelope wavefunction for an electron in the conduction band, provided that the effective mass  $m^*$  is used in the Hamiltonian. When the Schrödinger equation is applied to semiconductors in the effective mass approximation, the potential  $V(r)$  is assumed to be only the electrostatic potential, since the effect of the periodic crystal potential is accounted for by the effective mass itself. Such model can be used for relatively low energies close to the bottom of the conduction band, where a parabolic dispersion relation is a good approximation. In semiconductors, some of the most interesting applications of the Schrödinger equation involve spatially varying material compositions and heterojunctions. The effective mass approximation can still be used with some caution. Since the effective mass is a property of a bulk, it is not well defined in the neighborhood of a sharp material transition. In the hypothesis of slow material composition variations in space, one can adopt the Schrödinger equation with a spatially varying effective mass, taken to be the mass of a bulk with the local material properties. However, it can be shown that the Hamiltonian operator is no longer Hermitian

for varying mass. A widely used Hermitian form brings the effective mass inside the differential operator as

$$-\frac{\hbar^2}{2} \nabla \cdot \left( \frac{1}{m^*} \nabla \psi \right) . \quad (27)$$

This approach is extended to abrupt heterojunctions, as long as the materials on the two sides have similar properties and bandstructure, as in the case of the GaAs/AlGaAs system in a certain range of the Al concentration. One has to keep in mind that very close to the heterojunctions the effective mass Schrödinger equation provides a reasonable mathematical connection between the two regions, but the physical quantities are not necessarily well defined. For instance, in the case of a narrow potential barrier obtained by using a thin layer of AlGaAs surrounded by GaAs, it is not clear at all what effective mass should be used for the AlGaAs, since such a region cannot be certainly approximated by a bulk. Even more difficult is to treat the case when there is a transition between direct and indirect bandgap materials (example, GaAs and AlGaAs with large Al concentration).

Assuming a uniform mesh size  $\Delta x$ , the Hamiltonian of the Schrödinger equation can be discretized in 1-D by introducing midpoints in the mesh intervals on the two sides of the generic grid point  $i$ . First, we evaluate the outer derivative at point  $i$  with centered finite differences, using quantities defined at points  $(i - 1/2)$  and  $(i + 1/2)$

$$-\frac{\hbar^2}{2} \frac{\partial}{\partial x} \left[ \frac{1}{m^*} \frac{\partial \psi}{\partial x} \right] \approx -\frac{\hbar^2}{2\Delta x} \left[ \left( \frac{1}{m^*} \frac{\partial \psi}{\partial x} \right)_{i+1/2} - \left( \frac{1}{m^*} \frac{\partial \psi}{\partial x} \right)_{i-1/2} \right] \quad (28)$$

and then the derivatives defined on the midpoints are also evaluated with centered differences using quantities on the grid points

$$-\frac{\hbar^2}{2\Delta x^2} \left[ \frac{\psi(i+1) - \psi(i)}{m^*(i+1/2)} - \frac{\psi(i) - \psi(i-1)}{m^*(i-1/2)} \right] \quad (29)$$

The effective mass is the only quantity which must be known at the midpoints. If an abrupt heterojunction is located at point  $i$ , the abrupt change in effective mass is treated without ambiguity. It can be shown that the box integration procedure yields the same result.

Another hermitian Hamiltonian operator proposed for variable mass has the form

$$-\frac{\hbar^2}{4} \left[ \frac{1}{m^*} \nabla^2 \psi + \nabla^2 \left( \frac{1}{m^*} \psi \right) \right] \quad (30)$$

which is the linear combination of two non-Hermitian operators. It is instructive to compare the two formulations. In 1-D, the operators can be rewritten as follows

$$-\frac{\hbar^2}{2} \frac{\partial}{\partial x} \left[ \frac{1}{m^*} \frac{\partial \psi}{\partial x} \right] = -\frac{\hbar^2}{2} \left[ \frac{1}{m^*} \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial \psi}{\partial x} \frac{\partial}{\partial x} \left( \frac{1}{m^*} \right) \right], \quad (31)$$

$$-\frac{\hbar^2}{4} \left[ \frac{1}{m^*} \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2}{\partial x^2} \left( \frac{1}{m^*} \psi \right) \right] = -\frac{\hbar^2}{2} \left[ \frac{1}{m^*} \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial \psi}{\partial x} \frac{\partial}{\partial x} \left( \frac{1}{m^*} \right) + \frac{1}{2} \psi \frac{\partial^2}{\partial x^2} \left( \frac{1}{m^*} \right) \right]. \quad (32)$$

The second operator has an additional term involving the second order derivative of the effective mass. For smoothly varying mass, the two approaches are approximately equivalent. If one were to use the form on the right hand side of Eq. (32) for discretization of the operator, it is easy to see that a direct application of finite differences is awkward. The proper procedure is to apply box integration to the interval  $[i - 1/2; i + 1/2]$

$$-\frac{\hbar^2}{2} \int_{i-1/2}^{i+1/2} dx \left[ \frac{1}{m^*} \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial \psi}{\partial x} \frac{\partial}{\partial x} \left( \frac{1}{m^*} \right) \right]. \quad (33)$$

Integration by parts of the first term yields

$$-\frac{\hbar^2}{2} \left[ \left[ \frac{1}{m^*} \frac{\partial \psi}{\partial x} \right]_{i-1/2}^{i+1/2} - \int_{i-1/2}^{i+1/2} dx \frac{\partial \psi}{\partial x} \frac{\partial}{\partial x} \left( \frac{1}{m^*} \right) + \int_{i-1/2}^{i+1/2} dx \frac{\partial \psi}{\partial x} \frac{\partial}{\partial x} \left( \frac{1}{m^*} \right) \right]. \quad (34)$$

The two integrals cancel, and if the result is divided by the integration length  $\Delta x$ , we recover Eq. (29).

## 4 References

- 
- 1 [http://nanohub.org/Online\\_Simulation\\_on\\_the\\_nanoHUB](http://nanohub.org/Online_Simulation_on_the_nanoHUB).
  - 2 F. Stern, Phys. Rev. Lett. 30, 278 (1973).
  - 3 D. K. Ferry, *Semiconductors*, MacMillan, New York (1991).
  - 4 N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, W. B. Saunders Co., New York, (1976).

- 
- 5 P. Hohenberg and W. Kohn, Phys. Rev. 136, B864 (1964).
  - 6 W. Kohn and L. J. Sham, Phys. Rev. 140, A1133 (1965).
  - 7 L. J. Sham and W. Kohn, Phys. Rev. 145, 561 (1966).
  - 8 S. E. Koonin and D. C. Meredith, *Computational Physics*, Addison-Wesley Publishing Company, New York (1990).
  - 9 J. C. Slater, Phys. Rev. 81, 385 (1951).
  - 10 G. D. Mahan, *Many-Particle Physics*, Plenum, New York (1981).
  - 11 W. Kohn and P. Vashishta, in *Theory of the Inhomogeneous Electron Gas*, Ed. by S. Lundqvist and N. H. March, Plenum Press, New York (1983).
  - 12 L. Hedin and B. I. Lundqvist, J. Phys. C 4, 2064 (1971).
  - 13 O. Gunnarsson and B. I. Lundqvist, Phys. Rev. B 13, 4274 (1976).
  - 14 I. K. Marmorkos and S. Das Sarma, Phys. Rev. B 48, 1544 (1993).
  - 15 F. Stern and S. Das Sarma, Phys. Rev. B 30, 840 (1984).
  - 16 E. Gawlinski, T. Dzurak and R. A. Tahir-Kheli, J. Appl. Phys. 72, 3562 (1992).
  - 17 U. von Barth, in *Electron Correlations in Solids, Molecules and Atoms*, Ed. by J. T. Devreese and F. Brosens, Plenum Press, New York (1983).
  - 18 N. D. Mermin, Phys. Rev. 137, A1441 (1965).
  - 19 U. Gupta and A. K. Rajagopal, Phys. Rev. A 21, 2064 (1980).
  - 20 U. Gupta and A. K. Rajagopal, Phys. Rev. A 22, 2792 (1980).
  - 21 D. Vasileska, P. Bordone, T. Eldridge and D. K. Ferry, J. Vac. Sci. Tech. B 13, 1841 (1995).
  - 22 B. Vinter, Phys. Rev. B 15, 3947 (1977).
  - 23 P. Kneschaurek, A. Kamgar and J. F. Koch, Phys. Rev. B 14, 1610 (1976).
  - 24 T. Ando, Z. Physik B 26, 263 (1977).
  - 25 W. A. Bloss, J. Appl. Phys. 66, 3639 (1989).
  - 26 T. Ando, Phys. Rev. B 13, 3468 (1976).

- 
- 27 S. Das Sarma and B. Vinter, Phys. Rev. B 23, 6832 (1981).
  - 28 S. Das Sarma and B. Vinter, Phys. Rev. B 26, 960 (1982).
  - 29 F. Schäffler and F. Koch, Solid State Comm. 37, 365 (1981).
  - 30 J.S. Blakemore, Solid-State Electron. 25, 1067 (1982).
  - 31 S. Luryi, Appl. Phys. Lett. 52, 501 (1988).
  - 32 D. Vasileska, D. K. Schroder and D. K. Ferry, IEEE Trans. Electron Devices 44, 584 (1997).
  - 33 S. Takagi and A. Toriumi, IEEE Trans. Electron Devices 42, 2125 (1995).
  - 34 M.J. van Dort, P.H. Woerlee , A.J. Walker, C.A.H. Juffermans, and H. Lifka, IEEE Trans. Electron Devices 39, 932 (1992).