Quantum confinement effects and source-to-drain tunneling in ultra-scaled double-gate silicon n-MOSFETs*

Jiang Xiang-Wei(姜向伟)† and Li Shu-Shen(李树深)

State Key Laboratory of Superlattices and Microstructures, Institute of Semiconductors, Chinese Academy of Sciences, P. O. Box 912, *Beijing* 100083, *China*

(Received 8 September 2011; revised manuscript received 10 November 2011)

By using the linear combination of bulk band (LCBB) method incorporated with the top of the barrier splitting (TBS) model, we present a comprehensive study on the quantum confinement effects and the source-to-drain tunneling in the ultra-scaled double-gate (DG) metal–oxide–semiconductor field-effect transistors (MOSFETs). A critical body thickness value of 5 nm is found, below which severe valley splittings among different X valleys for the occupied charge density and the current contributions occur in ultra-thin silicon body structures. It is also found that the tunneling current could be nearly 100% with an ultra-scaled channel length. Different from the previous simulation results, it is found that the source-to-drain tunneling could be effectively suppressed in the ultra-thin body thickness (2.0 nm and below) by the quantum confinement and the tunneling could be suppressed down to below 5% when the channel length approaches 16 nm regardless of the body thickness.

Keywords: quantum confinement, tunneling, metal–oxide–semiconductor field-effect transistors, linear combination of bulk band

PACS: 73.23.Ad, 73.40.Qv **DOI:** 10.1088/1674-1056/21/2/027304

1. Introduction

According to the international technology roadmap for semiconductors (ITRS) 2008 update,[1] it is expected that a field-effect transistor (FET) with L_G of 10 nm is in commercial mass production in 2015. In such an ultra-scaled region, new device materials and structures have emerged to overcome the obstacles such as severe short-channel effects.[2,3] Among the emerged novel device structures, ultra-thin body (UTB) double-gate (DG) metal–oxide–semiconductor field-effect transistors (MOSFETs) are especially promising thanks to their superior short-channel effect immunity.[4−6] In the ultra-scaled DG FETs, the body thickness needs to be shrunk to several nanometers to effectively suppress the short-channel effects. In that case, the quantum confinement effect could be rather strong. In other cases where the channel length needs to be scaled down to sub 10 nm, the source-to-drain direct tunneling could emerge to degrade the device performance. So, it is worth while to find out how the quantum confinement and the source-to-drain tunneling work in the nanometer region and how they are related to each other. To examine those quantum mechanical effects, proper simulation approaches or models incorporating the quantum mechanical nature become indispensable.

Over the past 20 years, many methods have been developed to incorporate the quantum mechanical effects into the device simulation.[7−9] The first method proposed includes some quantum mechanical effective potentials in the drift equation based on the gradients of the charge density.[10,11] Such gradient terms make the charge density smooth near the Si/SiO2 interface, hence incorporating some of the quantum mechanical effects. The so-called quantum Poisson drift equation or quantum hydrodynamic model has been extensively used for device simulations. However, as the device sizes shrink further, it is realized that the quantum mechanical wave functions need to be calculated explicitly.

There are many ways to solve the Poisson–Schrödinger equation depending on the problem to be studied and the computational costs. One way is to calculate the local densities of states, then apply the Boltzmann transport equation based on such

*Project supported by the National Basic Research Program of China (Grant No. G2009CB929300) and the National Natural Science Foundation of China (Grant Nos. 60821061 and 60776061).

†Corresponding author. E-mail: xwjiang@semi.ac.cn

© 2012 Chinese Physical Society and IOP Publishing Ltd

densities of states.[12,13] However, in the ballistic size region, the use of the Boltzmann equation itself is questionable. In such the cases, the direct solution to the open boundary condition scattering states based on the Schrödinger equation is necessary. For example, this has been done for the one-dimensional (1D) cases using the non-equilibrium Green's function approach based on the tight-binding model.[14] The use of Green's function also provides a way to incorporate the inelastic scattering processes into the formalism. Recently, three-dimensional (3D) device models with hundreds of thousands of atoms have been simulated based on the calculation of scattering states using the tight-binding model.[15,16] But thousands of computer processors are needed for such the direct 3D simulations. There are also effective mass calculations for two-dimensional (2D) systems using the scattering state approach.[17] They involve the solutions of linear equations in the dimension of the number of real space grid points. However, the direct simulation of the 3D device model based on the quantum mechanical transport equation remains to be nontrivial. Thus, it will be very helpful to find a faster way to do the simulations. Recently, a top of the barrier splitting (TBS) model[18] is presented to simulate the transport problem using the stationary electron eigen-states that are solved by using the linear combination of bulk band (LCBB) method[19] based on the empirical pseudopotential. Incorporated with LCBB, the TBS enables one to carry out the simulation of million-atom device structures with a personal computer.

In the following, we will first sketch the LCBB-TBS model. Then we will give a thorough examination on the quantum confinement effects and the source-to-drain tunneling in nanoscaled DG MOSFETs. Finally, we will give a summary of our study.

2. Simulation approach

The fully quantum mechanical simulation of nanometer MOSFET devices relies on the iterative calculation of the following Schrödinger equation and Poisson equation:

$$\left(-\frac{1}{2}\nabla^2 + \boldsymbol{V}(\boldsymbol{r}) + \boldsymbol{V}_{\rm str}(\boldsymbol{r}) + e\phi(\boldsymbol{r})\right)\psi_i(\boldsymbol{r}) = \boldsymbol{E}_i\psi_i(\boldsymbol{r}), \tag{1}$$

$$\nabla\cdot[\varepsilon(\boldsymbol{r})\nabla\phi(\boldsymbol{r})] = -4\pi[\boldsymbol{p}(\boldsymbol{r}) - \boldsymbol{n}(\boldsymbol{r}) + \boldsymbol{N}_{\rm d}^{+}(\boldsymbol{r}) - \boldsymbol{N}_{\rm a}^{-}(\boldsymbol{r})], \tag{2}$$

where $V(r) = \sum_{\alpha,R} \upsilon_\alpha(r-R)$ is the total empirical pseudopotential of the silicon crystal,[20] $V_{\rm str}(r)$ is the quantum confinement potential representing the SiO_2 barrier layer and the potential barrier between the source and the drain for the artificial periodic boundary condition, $\phi(r)$ is the electrostatic potential to be solved by the Poisson equation, $\epsilon(r)$ is the position-dependent dielectric constant, $N_{\rm d}^{+}(r)$ and $N_{\rm a}^{-}(r)$ are the donor and the acceptor nuclei charges, respectively and will be treated as continuous charge densities in the current calculation, $p(r)$ is the hole charge density and will be treated semiclassically and $n(r)$ is the occupied electron charge density. A quantum transport model working on the electronic state occupation and the current flow connects the above two equations. The kernel equation (1) is solved by the EPM based LCBB method, while the transport is calculated by using the top of the barrier splitting model.

2.1. Linear combination of bulk band method

Unlike the effective-mass envelope-function approximation (EMA) and its multiband $k\cdot p$ generalization where the wave function of the nanostructure is expanded in terms of zone-center ($k=0$ or "Γ point") Bloch bands of the underlying periodic solid, the LCBB wavefunction $\psi(r)$ in Eq. (1) is expanded in terms of full-zone Bloch states of the constituent bulk solid[19,21]

$$\psi(r) = \frac{1}{\sqrt{N}}\sum_{n}^{N_{\rm B}}\sum_{k}^{N_k} C_{k,n}u_{k,n}(r)\,{\rm e}^{{\rm i}k\cdot r}, \tag{3}$$

where $N_{\rm B}$ and N_k are the numbers of bands and k points included in the expansion, respectively, $u_{k,n}(r)$ is the full zone periodic Bloch function and N is the number of primary cells in the system. The full-zone periodic Bloch function $u_{k,n}(r)$ is described by the plane wave functions as

$$u_{k,n}(r) = \frac{1}{\sqrt{V_0}}\sum_{G}^{N_G} A_{k,n}(G)\,{\rm e}^{{\rm i}G\cdot r}, \tag{4}$$

where N_G is the number of the reciprocal-lattice vector G within an energy cutoff and V_0 is the primary cell volume.

From the expansion of the wavefunction, we can find many merits of the LCBB method compared with the EMA and the $k\cdot p$ formalism:[19]

(i) Unlike the EMA and the $k\cdot p$ model, which can only describe the band structure near the Γ

point,[22,23] the LCBB can describe the whole band structure of the constituent materials.

(ii) The LCBB can describe the inter-valley coupling.

(iii) The LCBB is an atomistic method, which can be used to study the atomistic features as well as the strains of nanostructures and devices.

In real devices, the above features may be important for realistic simulations. For example, it is found that the EMA gives incorrect electron density of states (DOS) for the silicon inversion layer and the indirect bandgap of the bulk silicon presents real challenges for the $k\cdot p$/EMA models.[24] Besides the accuracy, nowadays the quantum device simulation demands highly efficient calculations that can handle real size nanodevices such as the 22-nm FinFET, which is expected to be in production in the near future. These devices usually consist hundreds of thousands to millions of atoms, where the first principle packages (such as the VASP[25]) cannot be used, while even the tight-binding (TB) approach is limited due to the demand for large computer resources. The LCBB, however, can be used to calculate million-atom systems even with a personal computer workstation and the calculated eigenenergy is within 10 meV of the directly calculated results using the full plane wave basis set.[19] The details of the LCBB calculation can be found in Ref. [19].

2.2. Top of barrier splitting model

Recently, a top of the barrier splitting model (TBS) has been developed by us to simulate the electronic transport problem in nanostructured devices under highly non-equilibrium conditions.[18] Based on the physical intuition of the ballistic and the tunneling currents, the TBS decomposes a stationary eigenstate into two injecting scattering states, then each scattering state is occupied according to the corresponding left or right Fermi level. Unlike the QFL model, which assumes a local equilibrium, the TBS model assumes an elastic transport from left and right. Thus, they are the two opposite limits of the approximations. In our previous work,[18] the TBS model has been verified in a one-dimensional case by a comparison with the directly calculated effective mass scattering states and in a three-dimensional case by a comparison with the non-equilibrium Green's function (NEGF) tight-binding model. With the efficient electronic structure calculation LCBB method,[19] the TBS provides a very efficient way to simulate million-atom non-equilibrium devices without the lose of accuracy.

Unlike the QFL model, the TBS model treats each electron eigenstate separately. In the one-dimensional case, to decompose the total wave function into two injecting parts $|\psi(x)|^2 = |\psi_{\rm L}(x)|^2 + |\psi_{\rm R}(x)|^2$, we assume in the TBS model that the total wave function can be decomposed into two equal parts at the top of the barrier (x_m)

$$|\psi_{\rm L}(x_m)|^2 = |\psi_{\rm R}(x_m)|^2 = \frac{1}{2}|\psi(x_m)|^2, \qquad (5)$$

where $V(x_m)$ is the potential barrier maximum.

In the ballistic cases where the electron eigenenergy is above the barrier height $V(x_m)$, we further assume that the current becomes ballistic after the electron passing the top of the barrier $V(x_m)$, which leads to

$$\upsilon(x_m)\frac{1}{2}|\psi(x_m)|^2 = \upsilon(x)|\psi_{\rm L}(x)|^2, \quad (x > x_m), \qquad (6)$$

and

$$\upsilon(x_m)\frac{1}{2}|\psi(x_m)|^2 = \upsilon(x)|\psi_{\rm R}(x)|^2, \quad (x < x_m), \qquad (7)$$

where $\upsilon(x) = \sqrt{2|E - V(x)|/m_x^*}$ is the ballistic velocity for eigenenergy E.

In the tunneling cases where the electron eigenenergy has two crossing points $(x_{\rm L}, x_{\rm R})$ with the barrier, we assume that the total wave function can still be decomposed to two equal parts at x'_m where $|\psi(x)|^2$ is minimum. We further assume that the left and the right injecting parts decay in the same rate for a given interval (x, x_m) or (x_m, x), where x is inside $(x_{\rm L}, x_{\rm R})$. This leads to

$$\frac{|\psi_{\rm L}(x)|^2}{|\psi_{\rm L}(x'_m)|^2} = \frac{|\psi_{\rm R}(x'_m)|^2}{|\psi_{\rm R}(x)|^2}, \qquad (8)$$

which is used to calculate the state decomposition inside the barrier. Outside the barrier, the tunneling state is also assumed to become ballistic. Finally, the occupied electron density and the total current are obtained as

$$\begin{aligned} n(x) &= \sum_i |\psi_i^{\rm L}(x)|^2 f_{\rm D}(E_i - E_{\rm F}^{\rm L}) \\ &\quad + |\psi_i^{\rm R}(x)|^2 f_{\rm D}(E_i - E_{\rm F}^{\rm R}), \qquad (9) \\ J_{\rm tot} &= \sum_i J_i^{\rm L} f_{\rm D}(E_i - E_{\rm F}^{\rm L}) \\ &\quad - J_i^{\rm R} f_{\rm D}(E_i - E_{\rm F}^{\rm R}), \qquad (10) \end{aligned}$$

where $f_{\rm D}(x)$ is the Fermi–Dirac distribution function and $J_i^{\rm L}$ and $J_i^{\rm R}$ are the current elements. For the ballistic states, $J_i^{\rm L} = J_i^{\rm R} = \upsilon(x'_m)|\psi_i(x'_m)|^2/2$,

while for the tunneling states, $J_i^{\rm R} = J_i^{\rm L} = |\psi_i^{\rm R}(x_{\rm L})| \cdot |\psi_i^{\rm L}(x_{\rm R})| v(x'_m)/2$.

For the three-dimensional model, the effective one-dimensional kinetic potential and the electrostatic potential are extracted from the LCBB calculation and then the three-dimensional wave functions are splitted according to the one-dimensional ansatz. A detailed description on the extension of the above one-dimensional TBS model to the three-dimensional model using the LCBB calculation can be found in Ref. [18].

3. Quantum confinement effects

As mentioned in the introduction, the body thickness can be scaled down to several nanometers when the double-gate structure is further downsized. In that case, the lateral confinement can be very strong. For a specific channel direction, the specific directional confinement might affect electrons from different energy valleys differently. This valley splitting can be important for the detailed device design. For example, it was pointed out recently that the occupied charge density transferring caused by the quantum confinement is related to the suppression of the electron mobility degradation in (100)-oriented double-gate ultra-thin body nMOSFETs.[26] In the following, we will examine the (100) channel double-gate structure shown in Fig. 1.

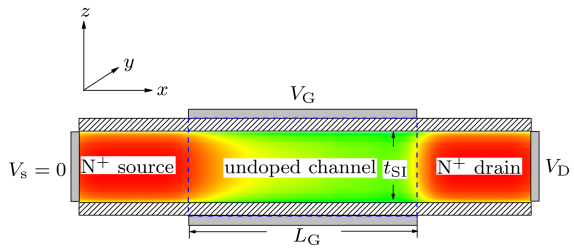


Fig. 1. (color online) Structure of double-gate MOSFETs.

The channel lengths of the devices are 20 nm, and the silicon body thicknesses varies from 2.0 nm to 8.0 nm. The source and the drain are highly n^+ doped, $N_{\rm D} = 1 \times 10^{20}$ cm^{-3}, while the channel is undoped. The oxide thickness is $t_{\rm OX} = 1.5$ nm. In our calculation, about 7000 k-points are chosen from six X valleys (two identical X100 valleys, two identical X010 valleys and two identical X001 valleys). In the following, we will refer to these six valleys as three X valleys (X100, X010, and X001). In our previous work, it is found that there is no coupling among these three valleys.[27] So we can separately calculate the contribution of each valley for the occupied charge density and the current.

Figure 2 shows the percentages of the contributions to the total charge density from X100, X010, and X001 valleys. From the figure, the valley splitting can be clearly found when the body thickness is below 5.0 nm where the contribution from X001 starts to dominate. When the body thickness is approaching 2.0 nm, the contributions from X100 and X010 become negligible. This is clearly the consequence of the lateral quantum confinement effect along the z direction. The quantum confinement mainly happens along the z direction, the electrons from X100 and X010 valleys are confined the most because they have smaller effective masses along the z (001) direction, $m_z^*(X100) = m_z^*(X010) = m_t^* = 0.19m_e$, while the electron from X001 valley is less confined due to its larger effective mass along the z (001) direction, $m_z^*(X001) = m_l^* = 0.98m_e$. As the body thickness increases from 2.0 nm to 8.0 nm, the quantum confinement becomes less important and the different valley contributions become comparable.

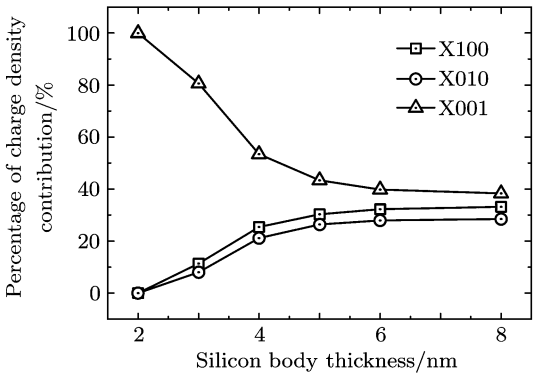


Fig. 2. Contributions from each valley for the occupied charge density.

More specifically, figure 3 plots the densities of states (DOS) from each valley for the body thicknesses of 2.0 nm, 4.0 nm and 6.0 nm. From the figure, it can be seen more clearly how the quantum confinement works. In thicker body cases, such as $t_{\rm SI} = 6.0$ nm, the densities of states from the three valleys are comparable, indicating comparable charge density contributions. As the body thickness down scales, the DOS is suppressed and shifted to a wider range. This is because the interval of the eigenenergies becomes larger due to the strong quantum confinement. When the body thickness approaches 2.0 nm, we can see that the valley splitting clearly occurs. The X100 and the X010

DOSs are shifted by 0.4 eV towards the higher energy due to the quantum confinement. This coincides with the occupied charge density contribution. As the DOS is suppressed by the quantum confinement, a higher Fermi level with respect to the conduction band edge is needed to maintain the charge neutrality. The corresponding Fermi level with respect to the conduction band edge is plotted in Fig. 4. The Fermi level is rather high in very thin body cases and then decreases and finally saturates when the body thickness is above 6.0 nm.

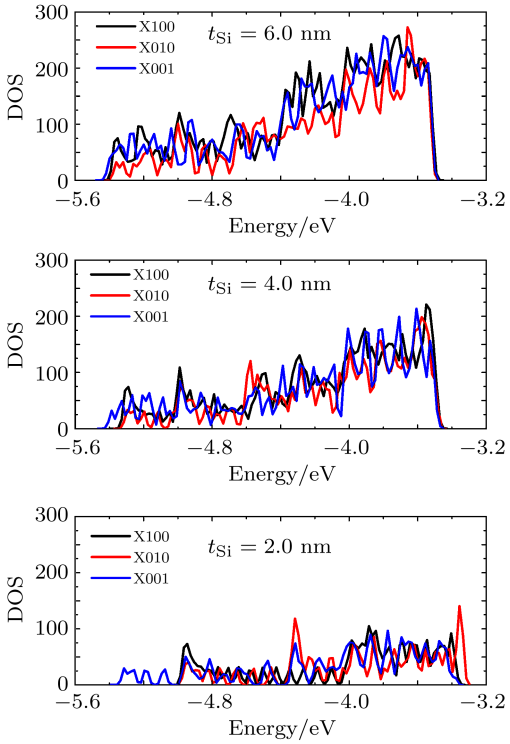


Fig. 3. (color online) Density of states (DOS) from different valleys.

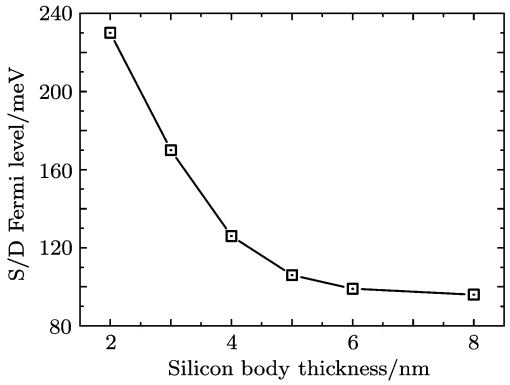


Fig. 4. Fermi level with respect to the conduction band edge.

Finally, figure 5 plots the current contributions from the three valleys in the small and the large gate voltage cases. Comparing the charge density contributions, we find two differences. The first difference is that the X001 contribution is significantly larger than the X100 and the X010 contributions, i.e., there is no comparable saturation similar to that in the charge density contributions. This is because unlike the charge density contribution, there is an additional factor needed to be counted for the current, which is the electron velocity or mobility and is related to the transport directional effective mass. The double effects of the quantum confinement and the higher mobility make X001 contribution to the current significantly larger than the other two. The factor of the transport directional effective mass also explains the second difference, the contribution of X010 is larger than that of X100. The double effects of the quantum confinement and the transport directional effective mass make the valley splitting in the current contribution larger than that in the charge density contribution.

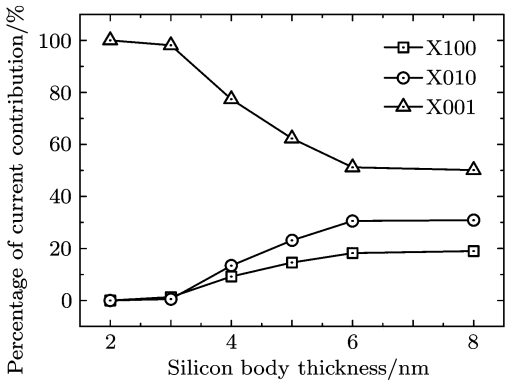


Fig. 5. Total current contributions from different valleys.

It should be noted that all the above evidences of the quantum confinement lead to a critical value of the body thickness, 5.0 nm. Bellow this value, the splitting among the three X valleys could be very severe and affects both the charge density and the current distributions. This should be carefully token into account in nanometer device designing.

4. Source-to-drain tunneling

Another quantum mechanical effect that has drawn much attention recently is the tunneling current, which includes the gate tunneling[28,29] and the source-to-drain direct tunneling.[30] This work focuses on the source-to-drain direct tunneling. As the dimensions of the MOSFETs scale down to sub 10 nm, this direct tunneling could be very important. It

is sure that the tunneling will degrade the device performance. But it is still not very clear whether the source-to-drain direct tunneling will finally set an ultimate scaling limit. Previously, this problem was studied by various approaches, such as the WKB approximation[31,32] and the density gradient corrections to the drift-diffusion equation.[33,34] However, as pointed by Ref. [30], those studies have been clouded by the approximations. In Ref. [30], Wang and Lundstrom presented a thorough study on the source-to-drain tunneling by using a 2D NEGF formalism and the BTE approach. From their results, Wang and Lundstrom pointed out that the source-to-drain tunneling does set an ultimate scaling limit but that limit is well below 10 nm. Due to the limitation of 2D simulation for 3D structures, the tunneling current was not directly calculated in Ref. [30] but extracted from the difference of the NEGF and the BTE results. So, a direct calculation of the tunneling current from a 3D model is still worth to carry out. In the following, we will present such a calculation.

The basic structure of the devices simulated is shown in Fig. 1. The channel length varies from 6.0 nm to 16.0 nm, while the body thickness varies from 2.0 nm to 4.0 nm. The gate work functions are modulated so that all the devices have an identical off current of 4.5×10^{-7} μA/μm.

Figure 6 shows the I–V characteristics of 2.0 nm body thickness devices with channel lengths of 8.0 nm, 12.0 nm and 16.0 nm. It can be seen that as the channel length shrinks, the device performance factors, such as subthreshold swing and on/off ratio, degrade. Besides the electrostatics degradation in the short channel, the source-to-drain tunneling is partially responsible for that. Figure 7 shows the subthreshold swing and the on/off ratio versus the channel length in very thin body ($t_{\rm SI}$ = 2.0 nm) devices. As can be seen from Fig. 7, the device performance degrades as the channel length scales down. As the channel length decreases from 16 nm to 6 nm, the subthreshold swing (SS) increases from the nearly ideal 67 mV/dec to 93 mV/dec. This is due to the significant source-to-drain direct tunneling in the ultra short channels, which can be seen from Fig. 8. We find that the percentage of the direct tunneling current could be nearly 100% of the total current. Such a source-to-drain direct tunneling significantly lifts the subthreshold current in the ultra short channel, which increases the SS value. However, we find that the subthreshold swing approaches 65 mV/dec when the channel length is near 16 nm. This means that around the 20-nm technology node, the source-to-drain tunneling does not affect much the device performance.

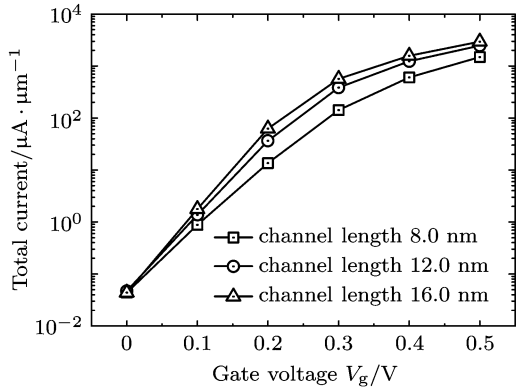


Fig. 6. I–V characteristics for the double-gate FETs with silicon thickness $t_{\rm SI}$ = 2.0 nm and channel length $L_{\rm G}$ = 8 nm, 12 nm, 16 nm.

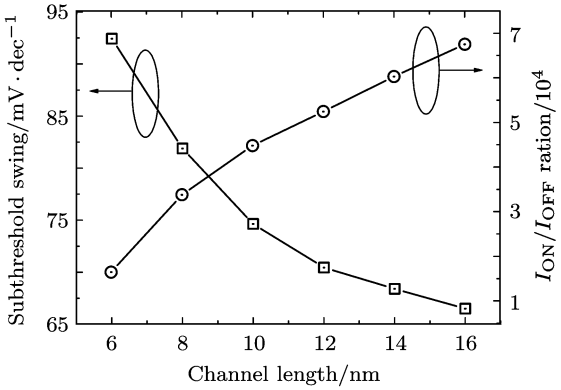


Fig. 7. Subthreshold swing and ON–OFF ratio as a function of the channel length. Device body thickness is set to be $t_{\rm SI}$ = 2.0 nm.

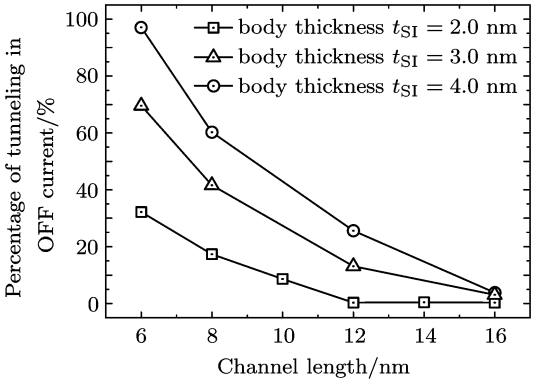


Fig. 8. Percentage of the tunneling current in the total current as a function of channel length and silicon body thickness.

As described above, our TBS-LCBB model enables us to separate the tunneling and the ballistic states and calculate the currents for each part.

In Fig. 8, we plot the percentage of the source-to-drain tunneling current in the total off current ($I_{\rm TUNN}/I_{\rm OFF}$) versus the channel length and the body thickness. Big differences compared with the previous 2D NEGF-BTE results[30] can be found. The source-to-drain tunneling is effectively suppressed to below 40% in the 2.0-nm thin body case and the tunneling drops down to well below 5% when the channel length approaches 16 nm regardless of the body thickness. We believe that the source-to-drain tunneling is suppressed in the thin body due to the quantum confinement effects. To see this more clearly, we define occupied densities of states for the ballistic and the tunneling states respectively as

$$D_{\rm B}^{\rm OC}(E) = D_{\rm B}(E)(f_{\rm D}(E-E_{\rm F}^{S}) - f_{\rm D}(E-E_{\rm F}^{\rm D})), \quad (11)$$

and

$$D_{\rm T}^{\rm OC}(E) = D_{\rm T}(E)(f_{\rm D}(E-E_{\rm F}^{S}) - f_{\rm D}(E-E_{\rm F}^{\rm D})), \quad (12)$$

where $D_{\rm B}(E)/D_{\rm T}(E)$ is the density of ballistic/tunneling states, and $f_{\rm D}(x)$ is the Fermi–Dirac distribution function under a given temperature. In the following, we will refer to $D_{\rm B}^{\rm OC}$ and $D_{\rm T}^{\rm OC}$ as ODOBS and ODOTS (short for occupied density of ballistic/tunneling states). As can be seen in Fig. 9, where the ODOBS and the ODOTS of the off device performance state are plotted, the ODOTS is suppressed while the ODOBS is raised as the body thickness varies from 4.0 nm to 2.0 nm. The changes of ODOTS and ODOBS coincide well with the source-to-drain tunneling suppression shown in Fig. 8. As explained in the previous section, the quantum confinement raises the electron energy levels and suppresses the total density of states in ultra-thin body thickness structures. In that case, the Fermi level is also raised to maintain the charge neutrality. That is what we see in Fig. 9. The source Fermi level moves towards the higher energy region, making the over-the-hill states (ballistic states) more contributive and the tunneling states less contributive. This effect suppresses the source-to-drain tunneling when the body thickness becomes thinner.

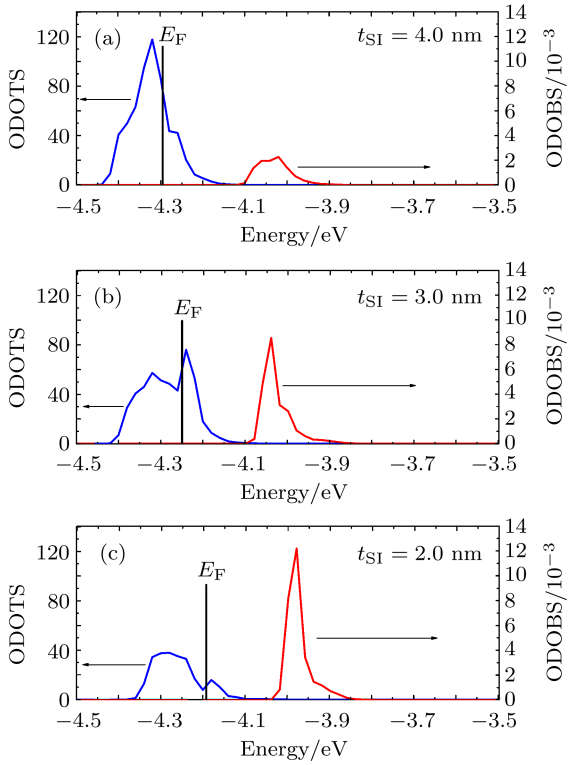


Fig. 9. (color online) Occupied density of states (DOS) for ballistic and tunneling.

5. Summary

In summary, we have presented a comprehensive study on the quantum confinement effect and the source-to-drain tunneling in ultra-scaled double-gate MOSFETs. It is found that due to the quantum confinement, three X valleys of silicon have different contributions to the occupied charge density and the current. It is explained by the fact that the electrons from the different valleys are differently confined, which leads to the valley splitting in both the occupied charge density and the current. A critical body thickness value of 5 nm is found, below which the quantum confinement effects are rather severe. This finding should be taken into account carefully in device designing. As for the source-to-drain tunneling, different from the previous 2D NEGF-BTE results, our 3D atomistic simulation results show that the source-to-drain tunneling can be effectively suppressed in the ultra-thin body (2.0 nm and below) and our results also show that the source-to-drain tunneling will not set an obstacle for the 20-nm technology node, because the tunneling contribution is well below 5% when the channel length approaches 16 nm regardless of the body thickness. For the the quantum confinement suppressed source-to-drain tunneling, we give an explanation based on the occupied density of ballistic/tunneling states. We find that the occupied density of the ballistic states is raised while that of the tunneling states is suppressed as the body thickness scales down. This explains why the tunneling current becomes less important in thinner body devices.

Since the industry is approaching the 10-nm region now, where the quantum mechanical effects are believed to change the device physics, this study could be useful for device design.

References

[1] online available: http://www.itrs.net

[2] Wong H S 2002 *IBM J. RES. & DEV.* **46** 2/3

[3] Haensch W, Nowak E J, Dennard R H, Solomon P M, Bryant A, Dokumaci O H, Kumar A, Wang X, Johnson J B and Fischetti M V 2006 *IBM J. RES. & DEV.* **50** 4/5

[4] Frank D J, Laux S E and Fischetti M V 1992 *IEDM Tech. Dig.* pp. 553–556

[5] Wong H S, Frank D J, Taur Y and Stork M C 1994 *IEDM Tech. Dig.* pp. 747–750

[6] Hisamoto D, Lee W C, Kedzierski J, Takeuchi H, Asano K, Kuo C, Anderson E, King T J, Bokor J and Hu C 2000 *IEEE Trans. Electron Dev.* **47** 2320

[7] Shin M 2007 *J. Appl. Phys.* **101** 024510

[8] Anantram M P and Svizhenko A 2007 *IEEE Trans. Electron Dev.* **54** 2100

[9] Sverdlov V, Ungersboeck E, Kosina H and Selberherr S 2008 *Mat. Sci. Eng.* **R58** 228

[10] Iafrate G J, Grubin H L and Ferry D K 1981 *J. Phys.* **42** 307

[11] Ferry D K and Zhou J R 1993 *Phys. Rev.* B **48** 7944

[12] Jin S, Fischetti M V and Tang T W 2008 *Appl. Phys. Lett.* **92** 082103

[13] Vasileska D, Ferry D K and Goodnick S M 2006 *Handbook of Theoretical and Computational Nanotechnology* 10, Chap. 1

[14] Lake R, Klimeck G, Bowen R C and Javanovic D 1997 *J. Appl. Phys.* **81** 7845

[15] Luisier M, Schenk A, Fichtner W and Klimeck G 2006 *Phys. Rev.* B **74** 205323

[16] Luisier M and Klimeck G 2009 *IEEE Electron Dev. Lett.* **30** 602

[17] Laux S E, Kumar A and Fischetti M V 2004 *J. Appl. Phys.* **95** 5545

[18] Jiang X W, Li S S, Xia J B and Wang L W 2011 *J. Appl. Phys.* **109** 054503

[19] Wang L W and Zunger A 1999 *Phys. Rev.* B **59** 15806

[20] Wang L W and Zunger A 1994 *J. Phys. Chem.* **98** 2158

[21] Wang L W, Franceschetti A and Zunger A 1997 *Phys. Rev. Lett.* **78** 2819

[22] Wood D M and Zunger A 1996 *Phys. Rev.* B **53** 7949

[23] Wang L W and Zunger A 1996 *Phys. Rev.* B **54** 11417

[24] Esseni D and Palestri P 2005 *Phys. Rev.* B **72** 165342

[25] VASP: *Vienna ab-initio Simulation Package*, http://cms.mpi.univie.ac.at/vasp/vasp/vasp.html

[26] Shimizu K, Saraya T and Hiramoto T 2010 *IEEE Electron Dev. Lett.* **31** 4

[27] Deng H X, Jiang X W, Li S S, Xia J B and Wang L W 2008 *J. Appl. Phys.* **103** 124507

[28] Wu T F, Zhang H M, Wang G Y and Hu H Y 2011 *Acta Phys. Sin.* **60** 027305 (in Chinese)

[29] Wu H Y, Zhang H M, Song J J and Hu H Y 2011 *Acta Phys. Sin.* **60** 097302 (in Chinese)

[30] Wang J and Lundstrom M 2002 International Electron Device Meeting (IEDM) p. 707 http: // ieeexplore.ieee.org / xpls/abs_all.jsp?arnumber= 1175936&tag=1

[31] Kawaura H, Sakamoto T and Baba T 2000 *Appl. Phys. Lett.* **76** 3810

[32] Chang L and Hu C 2000 *Supperlattices and Microstructures* **28** 351

[33] Schenk A and Wettstein A 2002 Technical Proceedings of the 2002 International Conference on *Modeling and Simulation of Microsystems* p. 552 http://www.nsti.org/procs/MSM2002/11/X21.03

[34] Asenov A, Brown A R and Watling J R 2002 Technical Proceedings of the 2002 International Conference on *Modeling and Simulation of Microsystems* p. 490 http://www.nsti.org/procs/MSM2002/10/M41.02