NANO-MICRO LETTERS

Vol. 2, No. 2
83-88 (2010)

Effects of source-drain underlaps on the performance of silicon nanowire on insulator transistors

Sishir Bhowmick[1] and Khairul Alam[2,*]

The effects of source-drain underlaps on the performance of a top gate silicon nanowire on insulator transistor are studied using a three dimensional (3D) self-consistent Poisson-Schrodinger quantum simulation. Voltage-controlled tunnel barrier is the device transport physics. The off current, the on/off current ratio, and the inverse subthreshold slope are improved while the on current is degraded with underlap. The physics behind this behavior is the modulation of a tunnel barrier with underlap. The underlap primarily affects the tunneling component of drain current. About 50% contribution to the gate capacitance comes from the fringing electric fields emanating from the gate metal to the source and drain. The gate capacitance reduces with underlap, which should reduce the intrinsic switching delay and increase the intrinsic cut-off frequency. However, both the on current and the transconductance reduce with underlap, and the consequence is the increase of delay and the reduction of cut-off frequency.

Keywords: Silicon nanowire; Insulator transistors; Source-drain

Citation: Sishir Bhowmick and Khairul Alam, "Effects of source-drain underlaps on the performance of silicon nanowire on insulator transistors", Nano-Micro Lett. 2, 83-88 (2010). doi: 10.5101/nml.v2i2.p83-88

Scaling the transistor sizes has made significant improvement in the cost effectiveness and performance of integrated circuit over the last few decades. The bulk CMOS technology is rapidly approaching the scaling limit and alternate materials or device structures are essential for future electronics. One dimensional nanostructures such as the carbon nanotubes and silicon nanowires are the attractive materials for future nanoelectronics because their electronic properties can be controlled in a predictable manner. Controlled growth of silicon nanowires down to 3 nm diameter [1], their applications as field-effect transistors (FETs) [2-5], logic gates [6] and sensors [7] have been demonstrated.

When the transistors are scaled to nanometer regime, the device performance degrades mainly due to the short channel effects. The scaling of bulk silicon MOSFETs has been facilitated by introducing the device structures with source-drain underlaps [8]. However, large underlaps are required for optimal performance of bulk MOSFETs [9]. The ultra-thin body or FinFETs with undoped channels and bias dependent effective channel lengths have been proposed for optimal device performance [10,11]. Source-drain underlaps have been used to improve the device performance for carbon nanotube transistors [12,13] and silicon nanowire field-effect transistors (SiNWFETs) [14]. Shin uses multiple gates SiNWFETs and studies the subthreshold behaviors with source-drain underlaps [14].

In this paper, we study the effects of source-drain underlaps on device performance, namely the off current, the on current, the inverse subthreshold slope, S, the gate capacitance, C_g, the intrinsic switching delay, τ_S, and the intrinsic cut-off frequency, f_T, of a top gate silicon nanowire on insulator transistor by self-consistently solving the Poisson's and Schrodinger's equations. The off current, the on/off current ratio, and the inverse subthreshold slope are improved while the on current is degraded with source-drain underlaps. The physics behind this behavior is

[1]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka-1000, Bangladesh
[2]Department of Electrical and Electronic Engineering, East West University, Dhaka-1212, Bangladesh
*Corresponding author. Email: kalam@ewubd.edu

the modulation of a tunnel barrier by the source-drain underlap. The source-drain underlaps reduce the gate capacitance that should improve the switching performance of the device. However, the transconductance and the on current degrade with underlap and the consequence is the reduction of intrinsic cut-off frequency and increase of switching delay.

DEVICE STRUCTURE

Details of the device shown in Fig. 1 are as follows. The silicon nanowire is placed on a thick oxide layer $t_{ox\text{-}sub}$. The gate oxide t_{ox} is grown on the nanowire. A gate metal of length L_g is deposited on gate oxide and the exposed regions on both sides of the gate metal are covered by oxide $t_{ox\text{-}ex}$. The nanowire under the gate region and the underlaps L_u between the gate and the n-type doped source and drain extension L_{ex} are undoped. The gate length L_g is 10 nm and the gate oxide thickness t_{ox} is 1 nm. The silicon nanowire used in our study has a square cross-section of 5 × 5 nm^2. The substrate oxide, the gate oxide, and the extended oxide are SiO_2 with a dielectric constant value of 3.9. The source Fermi level is set to zero (0) and the drain Fermi level to $-qV_{DS}$. The gate metal is assumed to have the same work function value as the nanowire has. The L_{ex} value of 20 nm, the $t_{ox\text{-}sub}$ value of 5 nm, and the $t_{ox\text{-}ex}$ value of 5 nm are used for Poisson solver so that the fringing electric fields are treated correctly.

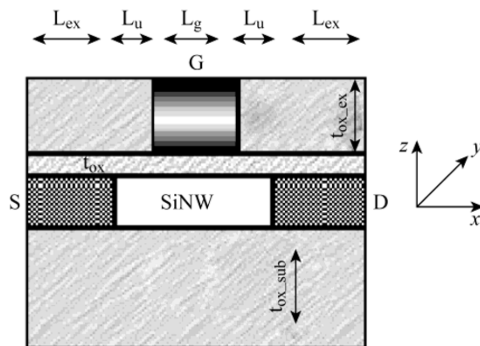

FIG. 1. The cross-sectional view and coordinates of the silicon nanowire on insulator transistor used in this study. Here gate length L_g = 10 nm and L_{ex} = 20 nm. For Poisson solver, $t_{ox\text{-}ex} = t_{ox\text{-}sub}$ = 5 nm.

SIMULATION MODEL

The simulation model uses a self-consistent solution between 3D Poisson's equation and effective mass Schrodinger's equation. The 3D Poisson's equation in cartesian coordinates is

$$\frac{\partial}{\partial x}\left(\varepsilon\frac{\partial V}{\partial x}\right)+\frac{\partial}{\partial y}\left(\varepsilon\frac{\partial V}{\partial y}\right)+\frac{\partial}{\partial z}\left(\varepsilon\frac{\partial V}{\partial z}\right)=-\frac{\rho}{\varepsilon_0}, \quad (1)$$

where ε_0 is the free space permittivity, ε is the relative dielectric constant, V is the 3D potential, and ρ is the charge density, which is non-zero in silicon nanowire only. Poisson kernel is created by discretizing Eq. (1) using finite difference. The normal component of electric field is set to zero at the source and drain ends and at the exposed surface of dielectric. Potential is fixed at the gate metal.

The Schrodinger's equation in 3D cartesian coordinates is

$$-\frac{\hbar^2}{2}\left[\frac{\partial}{\partial x}\left(\frac{1}{m_x}\frac{\partial\psi}{\partial x}\right)+\frac{\partial}{\partial y}\left(\frac{1}{m_y}\frac{\partial\psi}{\partial y}\right)+\frac{\partial}{\partial z}\left(\frac{1}{m_z}\frac{\partial\psi}{\partial z}\right)\right], \quad (2)$$
$$+U\psi = E\psi$$

where ψ is the wave function, m_x, m_y, and m_z are the effective masses in device coordinates, and $\hbar$ is the reduced Planck's constant. The nanowire is grown in <100> direction, which is device x coordinate in our study. Ballistic transport is assumed and recursive Green's function algorithm (RGFA) [15] is used to solve Schrodinger's equation for charge density and current calculations. The open boundary condition in transport direction x is included in Schrodinger's equation via self-energy matrices and hard-wall boundary condition is used in the transverse directions (y and z). For RGFA, the layer (cross-section) Hamiltonian and layer-to-layer coupling matrices are created by discretizing Eq. (2) using finite difference. With layer Hamiltonian H_i and layer-to-layer coupling matrix t, we create the right-connected Green function at each layer (cross-section) from

$$g_{i,i}=\left(EI-H_i-U_i-t_{i,i+1}g_{i+1,i+1}t_{i+1,i}\right)^{-1}, \quad (3)$$

where U_i is the potential energy at the ith cross-section (layer) obtained from Poisson solver and I is the identity matrix. We discretize Schrodinger's equation with equal grid spacing, and therefore, H_i is same at each cross-section and $t_{i,i+1} = t$ and $t_{i+1,i} = t^{\dagger}$. The full Green's function at the first layer is calculated from

$$G_{1,1}=\left(EI-H_1-U_1-\Sigma_S-t_{1,2}g_{2,2}t_{2,1}\right)^{-1}, \quad (4)$$

where $\Sigma_S = t_{1,0}g_{0,0}t_{0,1}$ is the self-energy matrix and $g_{0,0}$ is the surface Green's function. The surface Green's function is calculated from the decimation method and Ref. [16] has a detailed discussion. The rest {2, …, N_x} block diagonal elements of the full Green's function are calculated from

$$G_{i,i}=g_{i,i}+g_{i,i}t_{i,i-1}G_{i-1,i-1}t_{i-1,i}g_{i,i}\,. \quad (5)$$

We calculate the first column blocks of full Green's function from

$$G_{i,1}=g_{i,i}t_{i,i-1}G_{i-1,i} \quad (6)$$

and the left connected spectral function from

$$A^{L}_{i,i} = G_{i,1}\Gamma_{1,1}G^{+}_{i,1}, \quad (7)$$

where $\Gamma_{1,1}$ = -Im(Σ_S - Σ_S^+) is the broadening function. The charge density at each cross-section is calculated from

$$\rho_{i,i} = (2q)\int \frac{dE}{2\pi} diag\left\{ f_S A^{L}_{i,i} + f_D\left[A_{i,i} - A^{L}_{i,i}\right]\right\}, \quad (8)$$

where q is the electronic charge, f_S and f_D are the source and drain Fermi functions, respectively, and the full spectral function is obtained from $A_{i,i}$ = -Im$(G_{i,i} - G^{\dagger}_{i,i})$. The factor 2 at the beginning of right hand side of Eq. (8) includes spin degeneracy. Note that the charge density $\rho_{i,i}$ is a column vector of length $N_y \times N_z$ and is created by taking the diagonal elements of the matrix in the brace of the right hand side of Eq. (8).

The self-consistent loop is started with an initial guess of the potential profile. To generate the initial band profile, we calculate the conduction band position, E_{CS}, relative to the source Fermi level from charge neutrality. The band profile under the gate region is raised by $E_g/2$ (the channel is intrinsic) and that in the drain region is lowered by qV_{DS}. In other word, the initial profile is a step profile with E_{CS} in the source region, $E_{CS}+E_g/2$ under the gate region, and $E_{CS}-qV_{DS}$ in the drain region. Anderson mixing [17] scheme is used for convergence acceleration. Once the convergence is achieved, the coherent drain current is calculated from

$$I_D = \frac{2q}{h}\int dE T(E)\left[f_S(E) - f_D(E)\right], \quad (9)$$

where transmission coefficient $T(E)$ is calculated from [15]

$$T(E) = tr\left(\Gamma_{1,1}\left[A_{1,1} - G_{1,1}\Gamma_{1,1}G^{+}_{1,1}\right]\right). \quad (10)$$

The calculation is performed for each valley, and the charge density and drain current are obtained by taking sum over the valleys.

SIMULATION RESULTS

The silicon nanowire on insulator device used in our simulation is shown in Fig. 1. The channel consists of an undoped silicon nanowire of square cross-section of 5 × 5 nm^2. A 20 nm doped source-drain extension (L_{ex}) with a doping concentration value of 2×10^{19} cm^{-3} is assumed in our simulation. The nanowire is modeled using bulk effective mass parabolic band structure. Using the tight binding (TB) dispersion relation and the bulk effective mass model, Wang *et al.* [18] argued, using a semiclassical over the top of the barrier model, that the bulk effective mass model overestimates the threshold voltage for wire width < 3 nm and the on current for wire width < 5 nm. Using sp^{3}d^{5}s* orbital basis, Zheng *et al.* [19] show that the bulk masses are quite similar to the confinement masses for wire thickness greater than 3 nm. Shin [14,20] has used bulk effective masses to model silicon nanowire transistors of different gate structures. Poisson solver uses an extension of dielectric $t_{ox\text{-}ex}$ = 5 nm in the z-direction and equal the width of the nanowire on either side of the wire (y-direction) so that the fringing electric fields emanating from the gate metal are captured. In Fig. 1, the underlap can be varied in two ways: (a) by changing the nanowire length while the gate length is fixed and (b) by changing the gate length while the nanowire length is fixed. We follow method (a) to study the underlap effects on device performance. This is because the underlap as well as the gate length is changed in method (b), and it would be difficult to interpret whether the effect is due to underlap or due to gate length.

The simulated current-voltage characteristics for six different values of underlap are shown in Fig. 2.

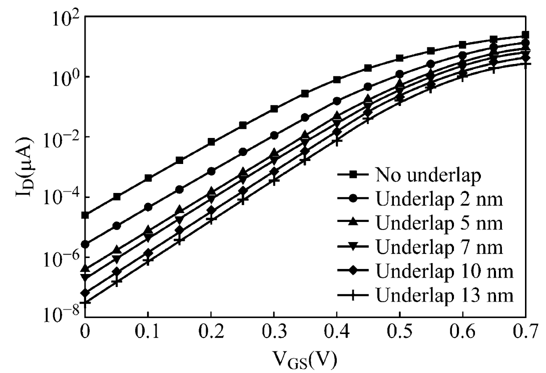

FIG. 2. The simulated current-voltage characteristics for six different values of source-drain underlap. The drain bias is fixed to 0.5 V.

The off current as well as the on current reduces with the increase of underlap. In our study, the off-state current is defined as the drain current at V_{GS} = 0 V and the on-state current is defined as the drain current at V_{GS} = 0.7 V. Our choice of on-state voltage of 0.7 V comes from the fact that, with the gate metal work function value equal to the nanowire, a flat band situation between the source Fermi level and the channel potential is obtained when the applied gate bias is about half of the band gap, which is 0.7 eV in our study. For a change of L_u from 0 to 13 nm, the off current reduces from 2.5×10^{-5} μA to 3.0×10^{-8} μA and the on current reduces from 22.6 μA to 2.95 μA. While the on current reduces by less than one order of magnitude, the off current reduces by almost three orders of magnitude.

To understand the physics of current reduction with underlap, we plot, in Fig. 3, the band profiles superimposed on

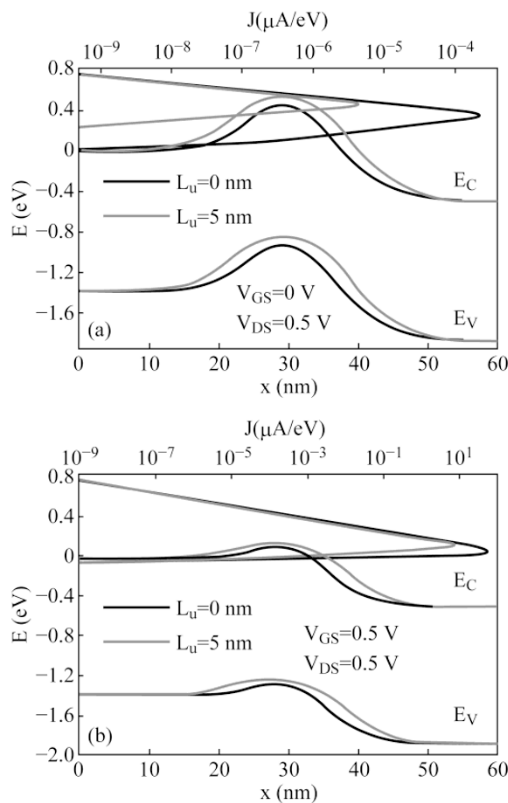

FIG. 3. Conduction band and valence band profiles superimposed on the energy distribution of current for two different values of underlap, 0 nm and 5 nm, at (a) V_{GS} = 0 V and (b) V_{GS} = 0.5 V. The source Fermi level is at 0 eV and the drain Fermi level is at -0.5 eV.

the energy distribution of current for two different values of underlap, 0 nm and 5 nm, at two different gate biases 0 and 0.5 V.

The source Fermi level is set to 0 eV and the drain Fermi level to -0.5 eV. The current has thermal and tunneling components, and we can calculate them respectively from the energy spectrum of current $J(E)=(2q/h)T(E)[f_s(E)-f_D(E)]$ (current per unit energy). The thermal component of drain current can be calculated by integrating $J(E)$ from the top of the conduction band to ∞. The tunneling current can be obtained by integrating $J(E)$ from -∞ to the top of the conduction band. At V_{GS} = 0 and L_u = 0 nm, the tunneling component of current is 2.4×10^{-5} μA and the thermal component is 6.5×10^{-7} μA. These values are 3.4×10^{-7} μA and 3.3×10^{-8} μA, respectively, for L_u = 5 nm. At V_{GS} = 0.5 V, the tunneling and thermal components are 2.99 μA and 0.83 μA, respectively for L_u = 0 and 0.36 μA and 0.15 μA, respectively for L_u = 5 nm. The potential barrier width as well as the height becomes larger with the increase of underlap. This reduces both the tunneling and the thermal components of current. The underlap primarily affects the tunneling current, and therefore, the underlap effect is more pronounced in the off-state.

The off current, the on current, the on/off current ratio, and the inverse subthreshold slope are plotted in Fig. 4 as a function of source-drain underlap L_u.

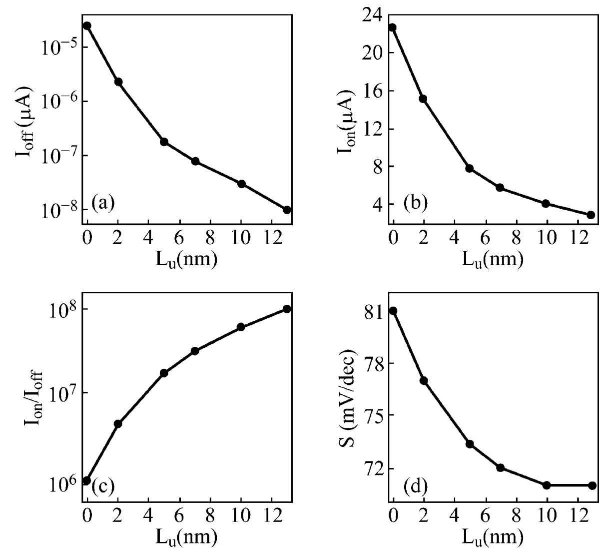

FIG. 4. The (a) off current, (b) on current, (c) on/off current ratio, and (d) inverse subthreshold slope versus underlap plots.

Both the on current and the inverse subthreshold slope reduce rapidly with L_u and then get almost saturated when L_u is about 6 nm. The off current and the on/off current ratio, on the other hand, do not show this behavior. If an underlap value of 5 nm is assumed as an optimal design (as the on current and the S do not change significantly after L_u = 5 nm), then the inverse subthreshold slope improves from 81 to 73.5 mV/dec, the off current improves from 2.5×10^{-5} μA to 3.7×10^{-7} μA, the on/off current ratio improves from 9.2×10^{5} to 1.7×10^{7}, and the on current degrades from 22.6 μA to 7.83 μA when the underlap changes from 0 to the optimal value. The improvement of off-state current and inverse subthreshold slope with gate underlap, and degradation of on-state current with gate underlap for gate-all-around and tri-gate silicon nanowire transistors have been reported by Shin [14]. The off-state current in both types of gate structure improves by four orders of magnitude or higher when the underlap is changed from 0 to 5 nm. The inverse subthreshold slope in their [14] gate-all-around transistors is ≈ 135 mV/dec at no underlap. This value improves to below 100 mV/dec at an underlap of 5 nm.

Note that the on/off current ratio of 9.2×10^{5} and the off-state current of 2.5×10^{-5} μA without source-drain underlap are already decent values and the role of underlap in improving device performance may not be pronounced. To highlight the role of underlap, we simulate the devices with 5 nm gate length and two underlap values 0 and 5 nm. The current-voltage characteristics are shown in Fig. 5.

For these devices, the on/off current ratio is 6.2×10^{2} without underlap and 2.2×10^{5} with 5 nm underlap. The off-state currents are 8.1×10^{-2} μA and 7.0×10^{-5} μA, respectively for 0 and 5 nm

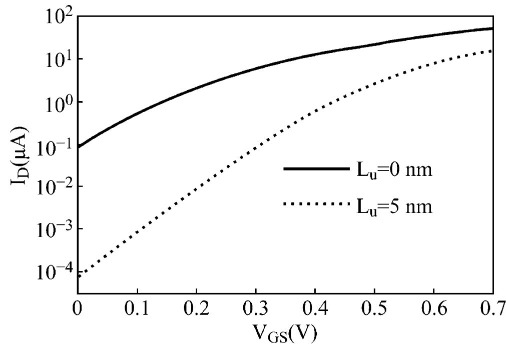

FIG. 5. The current-voltage characteristics of 5 nm gate length devices with two different values of underlap.

underlaps. A 10 nm gate length introduces sufficient tunnel barrier to reduce the tunneling leakage current without underlap and we get a decent value of on/off current ratio. However, the 5 nm gate length device has significantly high off-state current due to narrower tunnel length and the role of underlap is evident in improving device performance, especially the subthreshold performance.

Next we study the effects of source-drain underlaps on the gate capacitance, the intrinsic switching delay, and the intrinsic cut-off frequency. For this, the gate capacitance is calculated from

$$C_g = \iint dxdy \frac{\delta D_z}{\delta V_g} + \iint dydz \frac{\delta D_x}{\delta V_g}, \tag{11}$$

where, the first integral takes care of the electric fluxes emanating from the bottom surface of the gate metal and the second integral takes care of the fringing fields emanating from the two sides of the gate metal facing to the source and drain. The intrinsic switching delay is calculated from $\tau_S = C_g V_{DD}/I_{on}$ and the intrinsic cut-off frequency from $f_T = g_m/2\pi C_g$. The transconductance is calculated from $g_m = \partial I_D/\partial V_{GS}$.

The gate capacitance values and the percentage contribution of its different components versus gate bias are shown in Fig. 6. Here C_b is corresponding to the contribution from the fluxes emanating from the bottom surface of the gate metal and is evaluated by the first integral of Eq. (11), and C_s and C_d are the fringing field contributions emanating from the left side of the gate metal to the source, and from the right side of the gate metal to the drain, respectively, and are evaluated from the second integral of Eq. (11). The major contribution comes from C_b and its value ranges from 45% to 51%. The rest, which is almost 50% of the contribution of gate capacitance comes from the fringing fields.

In Fig. 7, we plot the gate capacitance and its different components (C_b, C_s, and C_d), the transconductance, the switching

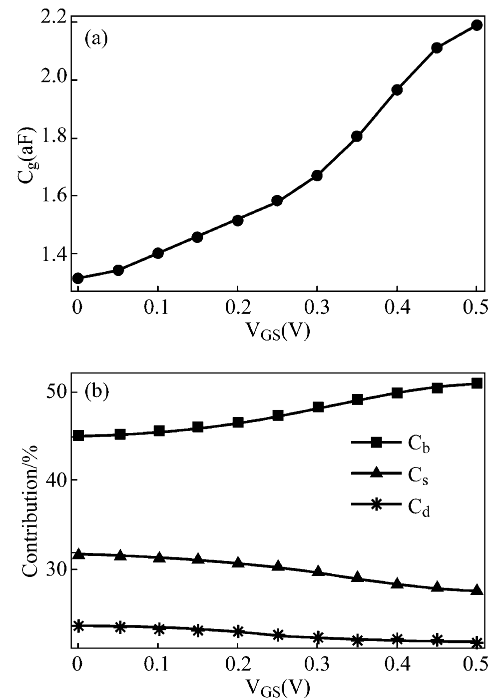

FIG. 6. The (a) gate capacitance and (b) the percentage contribution of its different components versus gate bias. The meanings of C_b, C_s, and C_d are described in the text. The drain bias is fixed to 0.5 V and the underlap value L_u = 0 nm.

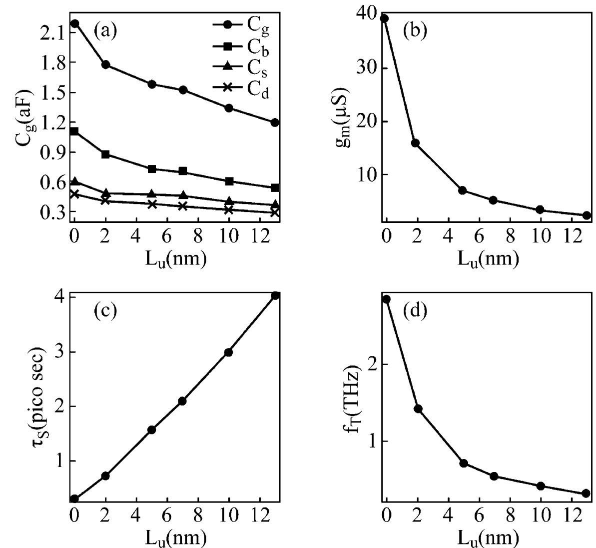

FIG. 7. The (a) gate capacitance and its different components, (b) trans-conductance, (c) intrinsic switching delay, and (d) intrinsic cut-off frequency versus underlap. The meanings of C_b, C_s, and C_d are described in the text.

delay, and the intrinsic cut-off frequency as a function of underlap. The gate capacitance reduces with underlap that should reduce the switching delay. However, the on current also reduces with underlap, and the combined effect is increase of the switching delay. The reduction of g_m with underlap should reduce f_T and the reduction of C_g with underlap should increase f_T. However, the reduction rate of g_m is higher and the consequence is the reduction of f_T. The gate capacitance, the

transconductance, and the cut-off frequency all have significant change with underlap up to 5 nm. After L_u = 5 nm, their changes are not large. However, the switching delay does not show this behavior. For a change of L_u from 0 to 5 nm, the τ_S increases from 0.286 to 1.557 pico second and the f_T reduces from 2.85 to 0.71 THz.

CONCLUSION

A three dimensional quantum simulation is performed for silicon nanowire on insulator transistors to study the effects of source-drain underlaps on device performance and to understand the physics of the effects. The underlap primarily affects the tunneling current and improves the short channel effects of the transistor at the cost of on current and the intrinsic switching performance. Appropriate choice of device structure combined with the source-drain underlaps can improve the device performance that can facilitate the optimal device design.

Received 21 March 2010; accepted 21 April 2010; published online 27 April February 2010.

References

1. Y. Wu, Y. Cui, L. Huynh, C. J. Barrelet, D. C. Bell and C. M. Lieber, Nano Lett. 4, 433 (2004). doi:10.1021/nl035162i
2. Y. Cui, Z. Zhong, D. Wang, W. U. Wang and C. M. Lieber, Nano Lett. 3, 149 (2003). doi:10.1021/nl025875l
3. H. C. Lin and C. J. Su, IEEE Trans. Nanotechnol. 6, 206 (2007). doi:10.1109/TNANO.2007.891828
4. S. M. Koo, M. D. Edelstein, Q. Li, C. A. Richter and E. M. Vogel, Nanotechnology 16, 1482 (2005). doi:10.1088/0957-4484/16/9/011
5. J. Wang, A. Rahman, A. Ghosh, G. Klimeck and M. Lundstrom, Appl. Phys. Lett. 86, 093113 (2005). doi:10.1063/1.1873055
6. Y. Huang, X. Duan, Y. Cui, L. J. Lauhon, K. H. Kim and C. M. Lieber, Science 294, 1313 (2001). doi:10.1126/science.1066192
7. Y. Cui, Q. Wei, H. Park and C. M. Lieber, Science 293, 1289 (2001). doi:10.1126/science.1062711
8. F. Boeuf, T. Skotnicki, S. Monfray, C. Julien, D. Dutartre, J. Martins, P. Mazoyer, R. Palla, B. Tavel, P. Ribot, E. Sondergard and M. Sanquer, Tech. Dig. Int. Electron Devices Meet. 1, 637 (2001).
9. R. Gusmeroli, A. S. Spinelli, A. Pirovano, A. L. Lacajta, F. Boeuf and T. Skotnicki, Tech. Dig. Int. Electron Devices Meet. 3, 225 (2003).
10. J. G. Fossum, M. M. Chowdhury, V. P. Trivedi, T. J. King, Y. K. Choi, J. An and B. Yu, Tech. Dig. Int. Electron Devices Meet. 3, 679 (2003).
11. V. P. Trivedi, J. G. Fossum and M. M. Chowdhury, IEEE Trans. Electron Devices 52, 56 (2005). doi:10.1109/TED.2004.841333
12. K. Alam and R. Lake, Appl. Phys. Lett. 87, 073104 (2005). doi:10.1063/1.2011788
13. K. Alam and R. K. Lake, J. Appl. Phys. 98, 064307 (2005). doi:10.1063/1.2060962
14. M. Shin, IEEE Trans. Nanotechnol. 6, 230 (2007). doi:10.1109/TNANO.2007.891819
15. R. Lake, G. Klimeck, R. C. Bowen and D. Jovanovic, J. Appl. Phys. 81, 7845 (1997). doi:10.1063/1.365394
16. M. P. L. Sancho, J. M. L. Sancho and J. Rubio, J. Phys. F 15, 851 (1985).
17. V. Eyert, J. Comput. Phys. 124, 271 (1996). doi:10.1006/jcph.1996.0059
18. J. Wang, A. Rahman, A. Ghosh, G. Klimeck and M. Lundstrom, IEEE Trans. Electron Dev. 52, 1589 (2005). doi:10.1109/TED.2005.850945
19. Y. Zheng, C. Rivas, R. Lake, K. Alam, T. B. Boykin and G. Klimeck, IEEE Trans. Electron Dev. 52, 1097 (2005). doi:10.1109/TED.2005.848077
20. M. Shin, IEEE Trans. Electron Dev. 55, 737 (2008). doi:10.1109/TED.2008.916149