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

Sishir Bhowmick,[1] Khairul Alam,[2,*] and Quazi Deen Mohd Khosru[1]

[1]Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka-1000
[2]Department of Electrical and Electronic Engineering
East-West University, Dhaka-1212, Bangladesh

Abstract

The effects, and their physics, of source-drain underlaps on the performance of silicon nanowire on insulator transistors are studied using three dimensional self-consistent Poisson-Schrodinger quantum simulation. Voltage control tunnel barrier is the device transport physics. The off current, the on/off current ratio, and the inverse subthreshold slope improve, and the on current degrades with underlap. The physics behind this behavior is the modulation of tunnel barrier with underlap. Underlap primarily affects the tunneling component of current and hence the improvement in the subthreshold regime of the transistors. About 50% contribution to the gate capacitance comes from the fringing electric fields emanating from the gate metal to the source and to the drain. The gate capacitance reduces with underlap, which can reduce the intrinsic switching delay and can 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. The optimal device design with underlap can be facilitated by appropriate choice of bias range and gate metal work function so that source-channel flat band condition is achieve in the on-state.

*Electronic address: kalam@ewubd.edu

I. INTRODUCTION

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 (SiNWs) 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], and ultra-thin body or FinFETs with undoped channels and bias dependent effective channel lengths have been proposed as the substitutes 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 study 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 (SS), the gate capacitance, the intrinsic switching delay τ_S, and the intrinsic cut-off frequency f_T, for a top gate silicon nanowire on insulator transistor using a self-consistent quantum simulation between Poisson and Schrodinger equations. The off current, the on/off current ratio, and the inverse subthreshold slope improve and the on current degrades with source-drain underlap. The physics behind this behavior is the modulation of a tunnel barrier by the source-drain underlap. The source-drain underlap reduces 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 the intrinsic cut-off frequency and the increase of switching delay.

II. 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-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-ex}. The nanowire under the gate region and the underlaps L_u between 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 SiNW has a square cross-section of 5×5 nm^2 and a band gap E_g value of 1.38 eV. The substrate oxide, the gate oxide, and the extended oxide are assumed to be 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 -V_{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-sub} value of 5 nm, and the t_{ox-ex} value of 5 nm are assumed for Poisson solver so that the fringing electric fields are treated correctly.

III. SIMULATION MODEL

The simulation model uses a self-consistent solution between three dimensional (3D) Poisson equation and effective mass Schrodinger equation. The 3D Poisson equation in cartesian coordinates is

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

where ϵ_o 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 and is able to handle both equal and unequal grid spacing. The kernel is stored in sparse matrix form and solved using standard Newton-Rapshon method. 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 equation in 3D cartesian coordinates is

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

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 Plank's constant. The nanowire is grown in $\langle 100\rangle$ 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 equation for charge density and current calculations. The open boundary condition in transport direction (x) is included in Schrodinger 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

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

where U_i is the potential energy at the ith cross-section (layer) obtained from Poisson solver and I is the identity matrix. The full Green's function at the first layer is calculated from

$$G_{1,1}=\left(EI-D_1-\Sigma_S-t_{1,2}g_{2,2}t_{2,1}\right)^{-1}, \tag{4}$$

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 decimation method and Ref. [16] has a detail discussion. The rest $\{2,\ldots,N_x\}$ block diagonal elements of the full Green's function are calculated from

$$G_{i,i}=g_{i,i}+g_{i,i}t_{i,i-1}G_{i-1,i-1}t_{i-1,i}g_{i,i}. \tag{5}$$

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

$$G_{i,1}=g_{i,i}t_{i,i-1}G_{i-1,1} \tag{6}$$

and the left connected spectral function from

$$A^L_{i,i} = G_{i,1}\Gamma_{1,1}G^\dagger_{i,1}, \tag{7}$$

where $\Gamma_{1,1} = i\left(\Sigma_S - \Sigma_S^\dagger\right)$ is the broadening function. The charge density at each cross-section is calculated from

$$\rho_{i,i} = (2e)\int \frac{dE}{2\pi}\mathrm{diag}\left\{f_S A^L_{i,i} + f_D\left[A_{i,i} - A^L_{i,i}\right]\right\}, \tag{8}$$

where e 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 $A_{i,i} = \sqrt{-1}(G_{i,i} - G^\dagger_{i,i})$. 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 the initial guess of the potential profile. Anderson mixing [17] scheme is used for convergence acceleration. Once the convergence is achieved, the coherent drain current is calculated from

$$I_D = \frac{2e}{h}\int dE\; T(E)\left(f_S - f_D\right), \tag{9}$$

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

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

IV. 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 and a band gap value of 1.38 eV. A 20 nm doped source-drain extension (L$_{ex}$) with a uniform doping concentration value of $1.5{\times}10^{19}$/cm^3 is assumed in our simulation. The nanowire is model 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 $sp^3d^5s^*$ orbital basis, Zheng *et al.* [19] shows that the bulk masses are quite similar to the confinement masses for wire thickness greater than 3 nm. Poisson solver uses an extension of dielectric $t_{ox-ex} = 5$ 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.

The simulated $\log I_D - V_{GS}$ plots for six different values of underlap are shown in Fig. 2. The off current as well as the on current reduces with the increase of underlap. For a change of L_u from 0 to 13 nm, the off current reduces from $2.5 \times 10^{-5} \mu$A to $3.0 \times 10^{-8} \mu$A and the on current reduces from 3.8 μA to 0.15 μA. While the on current reduces by about one order of magnitude, the off current reduces by almost three orders of magnitude.

To understand the physics of off-state and on-state current reduction with underlap, we plot, in Fig. 3, the band profiles superimposed on the energy distribution of current for two different values of underlap, 0 nm and 5 nm. The source Fermi level is set to 0 eV and the drain Fermi level to -0.5 eV. Both the off-state and the on-state currents have the thermal and the tunneling components. For $L_u = 0$ nm, the tunneling component of the off-state current is $2.4 \times 10^{-5} \mu$A and the thermal component is $6.5 \times 10^{-7} \mu$A. These values are $3.4 \times 10^{-7} \mu$A and $3.3 \times 10^{-8} \mu$A, respectively, for $L_u = 5$ nm in the off-state. The on-state current for zero underlap has 2.99 μA tunnel component and 0.83 μA thermal component. These values for 5 nm underlap device are 0.36 μA and 0.15 μA, respectively. The potential barrier length 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 affect the tunneling current and the effect of underlap is higher in the off-state.

The off current, the on current, the on/off current ratio, and the inverse subthreshold slope (SS) are plotted in Fig. 4 as a function of source-drain underlap L_u. 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. This is because the effects of underlap in the off-state is more significant than its effects in the on-state. If an underlap value of 5 nm is assumed as an optimal design (as the on current and the SS do not change significantly after $L_u = 5$ nm), then the inverse subthreshold slopes improves from 81 to 75 mV/dec, the off current changes from $2.5 \times 10^{-5} \mu$A to $3.7 \times 10^{-7} \mu$A, the on/off current ratio improves from 1.5×10^5

to 1.4×10^6, and the on current degrades from 3.8 μA to 0.51 μA. 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 exponential fit to the simulated data for the off current and the on current is $I_{off}(L_u) = 2.396 \times 10^{-5} \exp[-1.259L_u] + 6.776 \times 10^{-7} \exp[-0.2847L_u]$ and $I_{on}(L_u) = 2.289 \exp[-0.8361L_u] + 0.9542 \exp[-0.1431L_u]$, and are shown in Fig. 4 (a) and (b) as the dotted lines. A polynomial fit to second order of the on/off current ratio and the inverse subthreshold slope results in $I_{on-off}(L_u) = 1.933 \times 10^4 L_u^2 + 1.136 \times 10^5 L_u + 2.08 \times 10^5$ and $SS(L_u) = 0.06534L_u^2 - 1.633L_u + 80.71$. The polynomials are shown in Fig. 4 (c) and (d) as the dotted lines.

Note that the on/off current ratio of 1.5×10^5 without source-drain underlap is already a decent value and the on current goes below 1.0 μA for a source-drain underlap value of 5 nm or higher. This is because of the particular device and the particular bias range (0 to 0.5 V) that have been used in this simulation study. A 10 nm gate length introduces sufficient tunnel barrier to reduce the tunneling leakage current in the off-state and to obtain a decent value of on/off current ratio. Simulations (results are not shown here) for a 5 nm gate length device with the same gate bias range (0 to 0.5 V) and drain bias fixed to 0.5 V shows that the on/off current ratio without underlap is 2.6×10^2 and it is 3.7×10^4 for 5 nm underlap. The on current with underlap goes below 1.0 μA because the significant contribution to the on current is the tunneling component (see discussion regarding Fig. 3). This can be improved by extending the bias range or engineering the gate metal work function so that a flat band between the source Fermi level and the channel potential under the gate is obtained in the on-state. The nanowire has a band gap value of 1.38 eV. 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 (0.7 eV). By extending the gate bias range from 0 to 0.7 V with drain bias fixed to 0.5 V, simulations (results are not shown here) show that the on current for 10 nm gate device changes from 22.6 μA to 11.3 μA for a change of underlap value from 0 to 5 nm.

Next we study the effects of source-drain underlap on the gate capacitance C_g, the intrinsic switching delay τ_S, and the intrinsic cut-off frequency f_T. For this, the gate capacitance is

calculated from

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

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 on-state current I_{ON} is the drain current corresponds to $V_{DS} = V_{GS} = 0.5$ V, V_{DD} is 0.5 V in this study, and 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. 5. Here C_b corresponds to the contribution from the fluxes emanating from the bottom surface of the gate metal and 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 in on-state is 51%. The value of C_b ranges from 45% to 51% over the entire range of gate voltage and the rest (almost 50%) comes from the fringing fields.

In Fig. 6, we plot the on-state gate capacitance and its different components (C_b, C_s, and C_d), the on-state transconductance, the on-state switching delay, and the on-state 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, 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. We observe improvement in the switching delay and cut-off frequency when the bias range is extended to achieve the source-channel flat band in the on-state. Simulations for the same 10 nm gate device with the gate bias swing from 0.0 to 0.7 V and drain bias fixed to 0.5 V show that

the on-state τ_S increases from 0.0563 to 0.093 pico second and the on-state f_T reduces from 8.106 to 5.0495 THz for L_u change from 0 to 5 nm.

V. CONCLUSION

A three dimensional quantum simulation is performed for silicon nanowire on insulator transistors to see the effects of source-drain underlap on the device performance and to understand the physics of the effects. The underlap primarily affect 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 bias range and gate metal work function combined with the source-drain underlap can improve the on-state current as well as the switching performance that can facilitate the optimal device design.

---

[1] Y. Wu, Y. Cui, L. Huynh, C. J. Barrelet, D. C. Bell, and C. M. Lieber, Nano Lett. **4**, 433 (2004).

[2] Y. Cui, Z. Zhong, D. Wang, W. U. Wang, and C. M. Lieber, Nano Lett. **3**, 149 (2003).

[3] H. C. Lin and C. J. Su, IEEE Trans. Nanotechnol. **6**, 206 (2007).

[4] S. M. Koo, M. D. Edelstein, Q. Li, C. A. Richter, and E. M. Vogel, Nanotechnology **16**, 1482 (2005).

[5] J. Wang, A. Rahman, A. Ghosh, G. Klimeck, and M. Lundstrom, Appl. Phys. Lett. **86**, 093113 (2005).

[6] Y. Huang, X. Duan, Y. Cui, L. J. Lauhon, K. H. Kim, and C. M. Lieber, Science **294**, 1313 (2001).

[7] Y. Cui, Q. Wei, H. Park, and C. M. Lieber, Science **293**, 1289 (2001).

[8] F. Boeuf and *et al.*, Tech. Dig. -Int. Electron Devices Meet. **1**, 637 (2001).

[9] R. Gusmeroli and *et al.*, 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).

[12] K. Alam and R. Lake, Appl. Phys. Lett. **87**, 073104 (2005).

[13] K. Alam and R. K. Lake, J. Appl. Phys. **98**, 064307 (2005).

[14] M. Shin, IEEE Transections on Nanotechnology **6**, 230 (2007).

[15] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, J. Appl. Phys. **81**, 7845 (1997).

[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).

[18] J. Wang, A. Rahman, A. Ghosh, G. Klimeck, and M. Lundstrom, IEEE Trans. Electron Dev. **52**, 1589 (2005).

[19] Y. Zheng, C. Rivas, R. Lake, K. Alam, T. B. Boykin, and G. Klimeck, IEEE Trans. Electron Dev. **52**, 1097 (2005).

Figure Captions

Fig. 1: Cross-sectional view and coordinates of the silicon nanowire on insulator transistor used in our simulation. Here gate length $L_g = 10$ nm and $L_{ex} = 20$ nm. For Poisson solver, $t_{ox-ex} = t_{ox-sub} = 5$ nm.

Fig. 2: Simulated log I_D vs V_{GS} characteristics for six different values of source-drain underlap. The drain bias is fixed to $V_{DS} = 0.5$ V.

Fig. 3: Conduction band and valence band profiles superimposed on the energy distribution of current for two different values of underlap, both in (a) off-state and (b) on-state. The source Fermi level is at 0 eV and the drain Fermi level is at -0.5 eV.

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

Fig. 5: 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.

Fig. 6: The on-state (a) gate capacitance and its different components, (b) transconductance, (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.

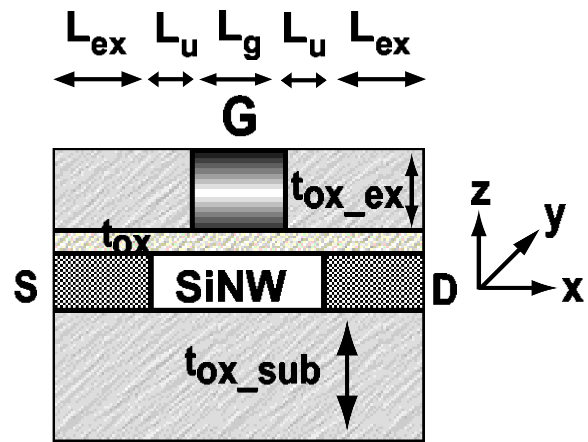

FIG. 1:

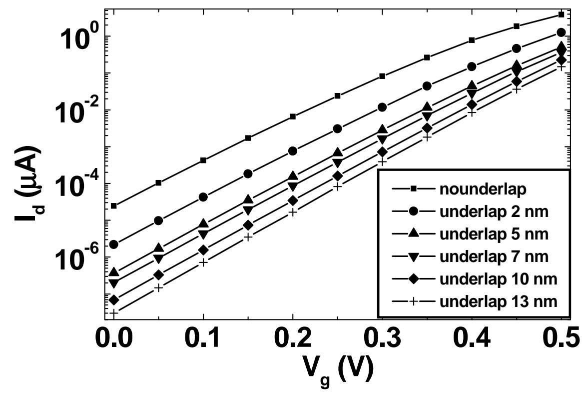

FIG. 2:

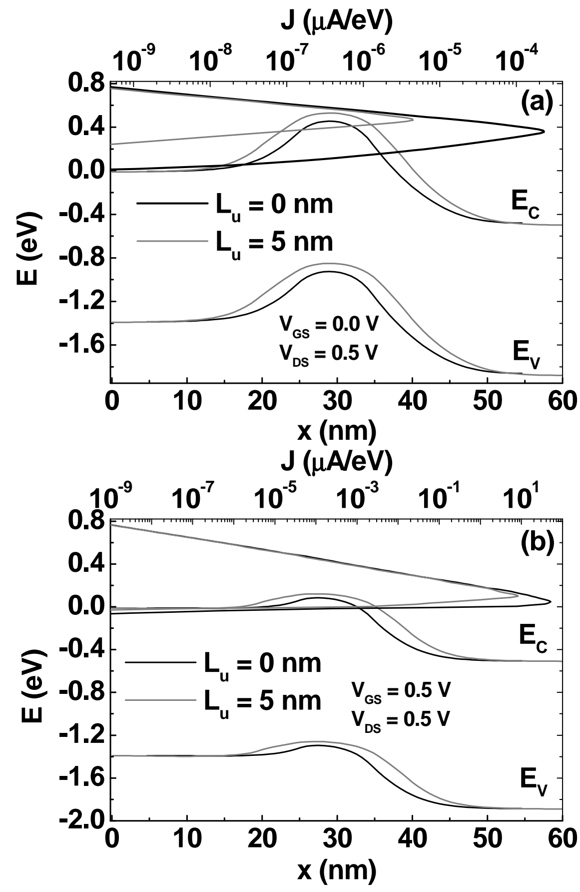

FIG. 3:

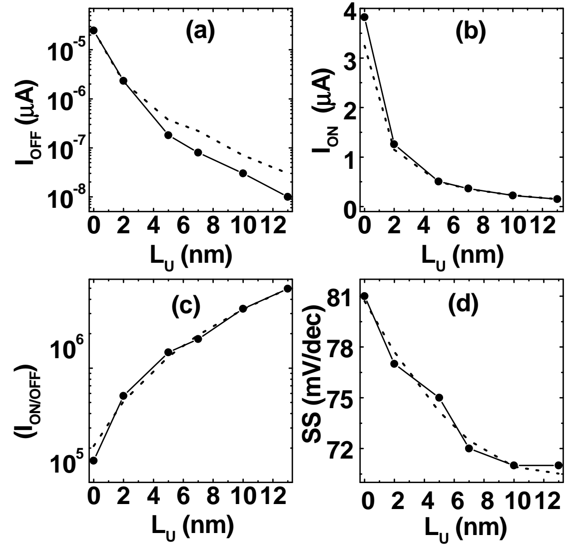

FIG. 4:

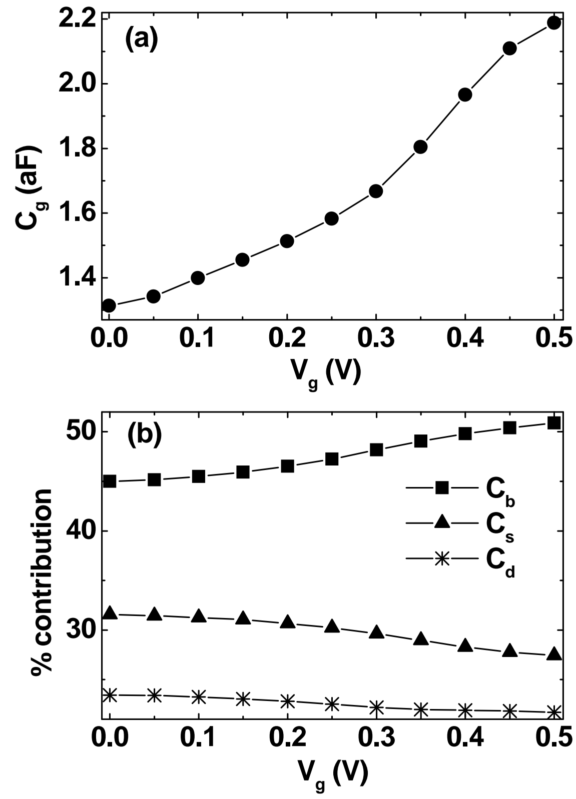

FIG. 5:

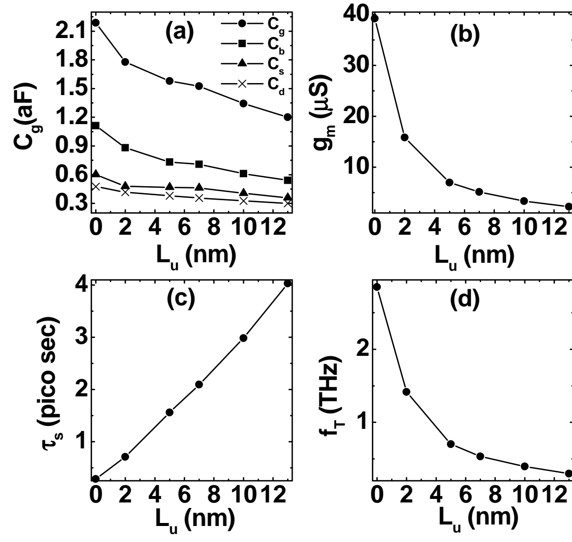

FIG. 6: