Available online at www.sciencedirect.com
ScienceDirect
Solid-State Electronics 50 (2006) 1489–1494
SOLID-STATE ELECTRONICS
www.elsevier.com/locate/sse

Gate current modeling of high-*k* stack nanoscale MOSFETs

Wei Wang [a,b], Ning Gu [a], J.P. Sun [c], P. Mazumder [c,*]

[a] *Department of Biomedical Engineering, Southeast University, Nanjing 210096, China*
[b] *College of Optoelectronic Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China*
[c] *Department of Electrical Engineering and Computer Science, The University of Michigan, USA*

Received 17 April 2005; accepted 8 August 2006
Available online 12 October 2006

The review of this paper was arranged by C. Tu

Abstract

A unified approach, particularly suitable for evaluation of high-*k* stack structures, is presented. This approach is based on fully self-consistent solutions to the Schrödinger and Poisson equations. Various structures and materials of high-*k* stacks of interest have been examined and compared to access the reduction of gate current in these structures. The present approach is capable of modeling high-*k* stack structures consisting of multiple layers of different dielectrics. The results of gate current and capacitance obtained from our model are in very good agreement with experimental data.
© 2006 Elsevier Ltd. All rights reserved.

Keywords: High-*k* structures; Nanoscale MOSFETs

1. Introduction

High-*k* gate stack structures as possible candidates to replace silicon dioxide layer for nanoscale MOSFETs have been of great interest very recently due to their promise in reduction of gate current in order to reduce standby power consumption of CMOS circuits and to alleviate scaling limits. When the device feature sizes reach nanoscale dimensions, a gate dielectric thickness of less than 2.0 nm would be required for sub-0.1 micrometer devices with power supply voltages of 1.0–1.5 V. The simultaneous requirements to keep the same or greater gate capacitance and lower gate current entail the same or greater equivalent oxide thickness (EOT) as well as a thicker physical insulating layer for reduction of gate tunneling current. The same or larger gate capacitance is necessary in order to maintain good gate control over the channel current and reduce the short channel effects. A number of high-*k* structures have been proposed and studied both experimentally and theoretically for a solution of this problem over the last few years [1–5]. While high-*k* dielectric materials have been demonstrated to be viable for storage capacitors [3], significant engineering of high-*k* stack structures and silicon dielectric interface are required due to complications associated with the applicability and compatibility of the materials and processing of the high-*k* stacks with silicon technologies.

From previous work of high-*k* stack structures of nanoscale MOSFETs, it has been found that, in order to maintain good interface quality between the high-*k* layer and silicon substrate, an interfacial oxide layer is present and a transition layer between high-*k* dielectric and silicon substrate may exist [2,4]. Furthermore, combinations of various high-*k* dielectric materials with different dielectric constant and other material properties need to be investigated for their effects on the gate current and capacitance. On the other hand, most of the theoretical work on the quantification of these effects make use of the WKB approximation, whereas there are few works addressing the multilayer

* Corresponding author. Tel.: +1 7347632107; fax: +1 7347631503.
E-mail address: mazum@eecs.umich.edu (P. Mazumder).

0038-1101/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.sse.2006.08.004

dielectric stack structures using self-consistent solutions to the Schrödinger–Poisson equations. In the WKB approximation, the 3-dimensional (3D) direct tunneling current components from the electrodes are evaluated based on a field approximation for the dispersion relationship, and the 2-dimensional (2D) current component from the inversion layer quantum well is calculated using the lifetime of the quasi-bound states in the quantum well. Although approximation models based on the WKB approximation can be validated by fitting parameters necessary to achieve agreement with experimental results, they are not very capable of modeling multilayer dielectric stack structures, since the material parameters, such as the bandgap, conduction band offset, dielectric constant, and the electron effective mass, in each layer differ from each other and cannot be specified accurately, making good approximations for the tunneling-related relationships difficult to attained. Moreover, models based on the WKB approximation are not capable of providing gate capacitance information needed simultaneously for the assessment of high-*k* stack structures. A unified model suitable for investigation of multiple dielectric high-*k* stack structures is therefore in need. To address this issue, we have developed an approach based on self-consistent solutions to the Schrödinger–Poisson equations. This approach is particularly suitable for structures consisting of multiple layers of different material parameters. It is conceptually simple, computationally efficient and very stable even for large device structures with multiple dielectric layers. It can evaluate gate current and capacitance simultaneously for a nanoscale MOSFET structure within a few hours on a modern PC and give very consistent results without having to use fitting parameters. The details of the numerical model is described in [6,7]. In this paper, we present our investigation of high-*k* stack structures using the present model to indicate its broad applicability.

In Section 2, the main formulations of the present model of gate current and capacitance for nanoscale MOSFETs is described briefly to provide overall information of its computational aspects. Section 3 presents gate current and capacitance for a number of high-*k* stack structures and the effects of gate current reduction for different combinations of high-*k* stack layers are examined and compared. A conclusion is given in Section 4.

2. Method of calculation

The present model is based on self-consistent solutions of the Schrödinger–Poisson equations previously developed for modeling quantum tunneling devices [7,8], such as the resonant tunneling diodes and transistors, with the modifications made for nanoscale MOSFETs in the *y*-direction as shown in Fig. 1, where a transition layer between the high-*k* dielectric and oxide is also included. The 3D gate current components are evaluated by a traveling wave calculation, sending a spectrum of electrons from the device electrodes, with their energy distributions governed by the Fermi statistics. The 2D gate current component is evaluated by a transmission calculation, with the quasi-bound states in the inversion layer quantum well fully included. The total gate current is then the sum of the 3D current components originated from the electrodes and 2D component originated from the inversion layer. Open boundary conditions, based on the quantum transmitting boundary method, are employed for all the current evaluations. We first calculate the self-consistent potential and charge distributions in the polysilicon and substrate regions from the coupled Poisson equation and Fermi distribution function. The electrostatic potential across the device in the *y*-direction is determined by the doping profile and mobile charges, as governed by the Poisson equation.

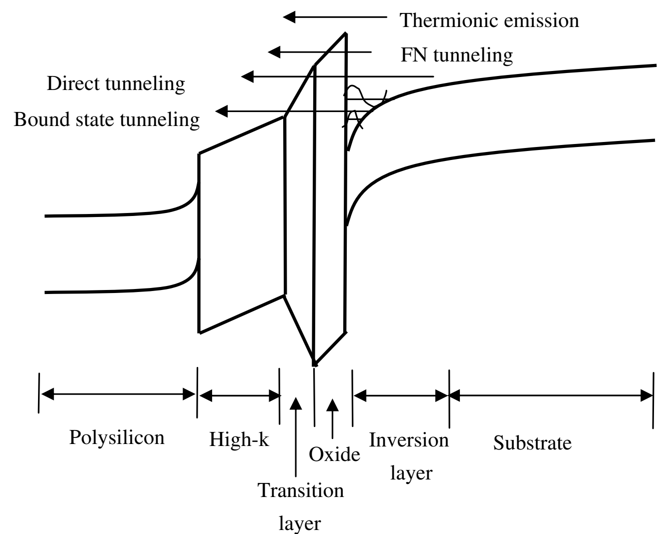

Fig. 1. Schematic band diagram of a high-*k* stack structure consisting of gate, high-*k*, transition, interfacial, and silicon layers, biased in the inversion region, showing various gate current components. All the current components in this work are calculated in a unified formulation using the same material parameters.

$$\frac{\partial^2 V(y)}{\partial y^2} = -\frac{q}{\varepsilon(y)}[N_D(y) - N_A(y) - n(y) + p(y)] \tag{1}$$

where $\varepsilon(y)$ is the dielectric constant of individual stack layers, and the electron concentration $n(y)$ in the equilibrium contact regions is given in terms of local Fermi levels,

$$n(y) = \int_0^\infty N(E)f(E)\,dE \tag{2}$$

where $N(E)$ is the density of states in the gate and substrate regions, respectively, and

$$f(E) = \left[1 + \exp\left(\frac{E - E_F(y)}{k_B T}\right)\right]^{-1} \tag{3}$$

is the Fermi–Dirac distribution function. The electrostatic potential $V(y)$ includes the device energy band offset at the silicon–oxide interfaces, the gate bias voltage applied across the gate and substrate, and the effects of device doping profiles as well as mobile charges. The Poisson equation and the Fermi integral lead to a pair of coupled nonlinear equations which are then solved using the Newton iterative

method with a sparse matrix technique. A similar treatment is used for the holes. For the charges inside the inversion layer quantum well formed under positive gate bias voltages, and for the gate current components through and/or above the ultrathin oxide barrier, quantum calculations are performed by directly solving the Schrödinger equation

$$-\frac{\hbar^2}{2}\frac{\partial}{\partial y}\left(\frac{1}{m^*(y)}\frac{\partial\psi(y)}{\partial y}\right)+E_C(y)\psi(y)=E\psi(y) \tag{4}$$

The self-consistent charge at the jth energy subband in the ith valley in the inversion layer is given by

$$n_{ij}(y)=\frac{m_i^* k_B T}{\pi\hbar^2}\ln\left[1+\exp\left(-\frac{E_{ij}-E_F}{k_B T}\right)\right]\times\left|\psi_{ij}(E_{ij},y)\right|^2 \tag{5}$$

where $\psi_{ij}(E_{ij}, y)$ is the electron wavefunction at the jth subband in the ith valley from the solution of the Schrödinger equation in the inversion layer. The gate current evaluation by the traveling wave calculations for the thermionic emission, FN tunneling, and direct tunneling through the oxide barrier, give

$$J_{3D}=-q\hbar\sum_k W(k)\mathrm{Im}\left(\psi_k^*(y)\frac{1}{m^*(y)}\frac{\partial\psi_k(y)}{\partial y}\right) \tag{6}$$

For the electron tunneling current from the inversion layer into the oxide and gate electrode, a transmission calculation is performed. Based on the wave functions calculated in the inversion layer, the transmission is given by

$$T_{ij}=\frac{\left|C_{tr,ij}\right|^2}{\left|A_{in,ij}\right|^2}\frac{k_{tr,ij}}{k_{in,ij}}\frac{m_{in,i}^*}{m_{tr,i}^*} \tag{7}$$

where $C_{tr,ij}$ is the transmitted wave amplitude in the ith valley and the jth subband, and $A_{in,ij}$ the incident wave amplitude in the corresponding subband and valley, respectively. The 2D gate current component originated from the subbands in the inversion layers is then

$$J_{2D}=\sum_{i,j}J_{i,j}=q\sum_{i,j}n_{eij}T_{ij}f_{ij} \tag{8}$$

where T_{ij} is the electron transmission probability, $f_{ij}=\frac{E_{ij}}{j\pi\hbar}$ the interface impact frequency, $n_{e,ij}$ the electron concentration, and E_{ij} the quasi-bound state energies, in the ith valley and the jth subband, respectively. The total gate current density is the sum of the 2D and 3D components,

$$J_g=J_{3D}+J_{2D} \tag{9}$$

The present approach has the following features: (1) We treat the thermioinc emission, FN tunneling, and direct tunneling current components in a unified formulation with the electron supply function based on the Fermi distribution. This treatment is viable and convenient, because it avoids the separation and delimitation among calculations of these individual components and the necessary use of other approximations and fitting parameters as in many other approximate models; (2) the electrostatic potential $V(y)$ and the material parameters can be specified at each point in the multilayered stack structures, allowing evaluation of various schemes of combination having different stack materials and different layers; and (3) the possible fixed charge and interface charge distributions in the stacked layers can also accommodated in Possion's equation, as long as models for these charge distributions are available or can be set up. Therefore, the present approach is very suitable for study of multilayer stack structures of nanoscale MOSFETs. More details of the method of calculation are described in [6,7].

3. Results and discussion

3.1. Effects of various high-k materials

The calculated gate current densities of various high-k stack structures with different high-k dielectrics as a function of the gate voltage are plotted in Fig. 2, showing similar behavior and close quantities as compared with other available modeled results. For all the various high-k materials being studied, the equivalent oxide thickness (EOT) is taken to be 2.0 nm. Other device parameters are $N_A = 5.0\times10^{17}\ \mathrm{cm}^{-3}$, and $V_{FB} = 0$ V, and the barrier height and electron effective mass for various materials are listed in Table 1, which are used throughout this work.

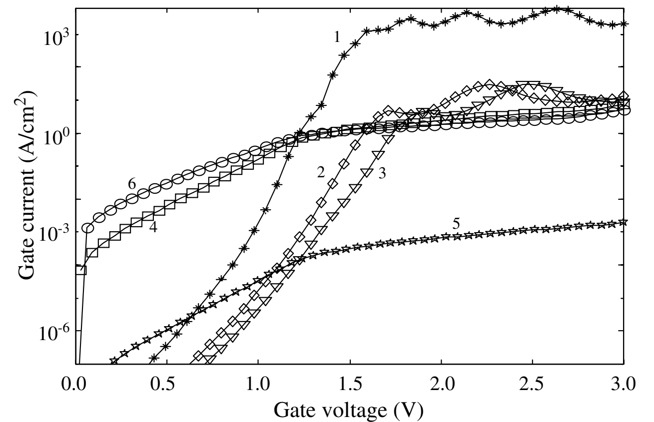

Fig. 2. Calculated gate current vs. gate voltage with different physical oxide thicknesses and the same EOT (2 nm) for different dielectric materials. Curves 1–5 correspond to Ta_2O_5, ZrO_2, HfO_2, Si_3N_4, and Al_2O_3 stack structure, respectively, each with 1 nm EOT of high-k material and 1 nm of SiO_2. Curve 6 denotes SiO_2 of 2 nm.

Table 1
Material parameters used for this work

Material	Barrier height (eV)	Dielectric constant	Effective mass ($\times m_0$)
Al_2O_3	2.80	9.00	0.30
SiO_2	3.15	3.90	0.50
Si_3N_4	2.00	7.00	0.20
Ta_2O_5	1.10	25.0	0.25
ZrO_2	1.40	25.0	0.20
HfO_2	1.50	25.0	0.10

The following observations are indicated from the calculated results: (1) All the high-*k* stacks make a reduction of gate current as compared with SiO_2 gate dielectric layer, especially at lower gate voltages applicable for nanoscale MOSFETs, due to the additional physical width in the tunneling path for the lower energy electrons that contribute most of the gate current; (2) the reduction of gate current is ineffective somewhere near and above the barrier height of the high-*k* dielectric layer, due to the dominating tunneling and thermal emission near and over the barrier of the high-*k* layer, although this is not a serious concern for the bias range of nanoscale MOS devices; (3) the resulting gate current is determined by the interplay among the barrier height and thickness, dielectric constant, the electron effective mass of the high-*k* and oxide layers, as well as the gate bias voltage; and (4) oscillations and kinks may occur in the calculated curves especially for thicker high-*k* layers. This is believed to be caused by pronounced multiple quantum reflections inherent in this type of quantum calculations in which the integration of wave functions may approach numerical instability. This type of oscillation has also been observed in other published works [10,11]. Despite the oscillations, our modeled results show good agreement with the available experimental data and results obtained with other approximate approaches, which will be verified shortly. We will not in this work discuss technological comparisons among the different dielectric materials.

3.2. Effects of high-*k* dielectric thickness

For high-*k* stack structures consisting of a single high-*k* layer and a SiO_2 layer, HfO_2 stacks are taken as an example of gate current dependence on the thickness of the high-*k* layer with the same EOT of 13 Å, as shown in Fig. 3. As expected, the physical thickness of the high-*k* layer plays a dominating role at the lower gate voltages since, for the same EOT, a higher dielectric constant translates a thicker high-*k* layer, thereby a lower gate current. The difference can be as large as 4–5 orders of magnitude between the high and low HfO_2/SiO_2 thickness ratios for the same EOT, as observed in the figure. At higher gate bias voltages slightly larger than the barrier height of the high-*k* dielectric material, however, the magnitude of gate current of the high-*k* stack structure and the SiO_2 structure becomes comparable, and that the former can be orders of magnitude larger than that of the SiO_2 structure at even higher gate voltage. In other words, at lower gate bias, the gate current is mainly affected by the dielectric constant of the high-*k* dielectric material utilized, while at higher bias, it is mainly determined by its barrier height. For ZrO_2 stack structures, similar results are obtained but will not be plotted here. In that case, for the same EOT, the gate current at lower gate voltages is almost the same as in the case of HfO_2 due to their virtually same dielectric constant, whereas at higher bias the gate current of ZrO_2 stack structures becomes slightly larger than that for the HfO_2 stacks as the former has a barrier height about 0.1 eV lower than that of HfO_2.

Other types of high-*k* stack structures using SiO_2–Ta_2O_5–SiO_2 gate dielectrics [3] and adding Al_2O_3 capping layer [13] have been proposed to alleviate the interfacial structure problems [3]. The present modeling approach is applied to study of this type of multiple layered structures, with the material and structural parameters specified accurately in the individual stacked layers. The calculated results will be compared with experiment data shortly.

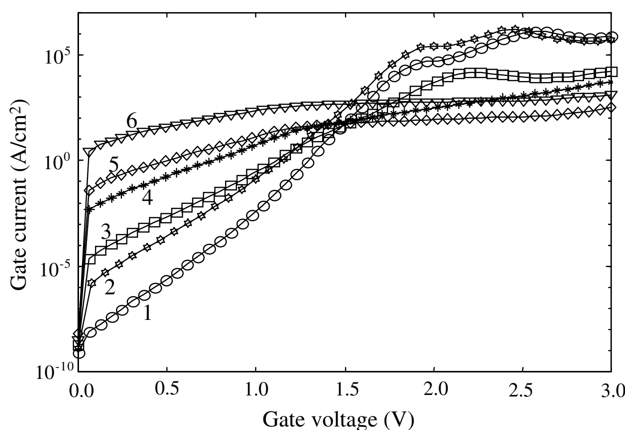

Fig. 3. Calculated gate current vs. gate voltage with different physical thickness of the high-*k* dielectric with the same EOT (1.3 nm) for HfO_2–SiO_2 structures. Curves 1–5 correspond to 84 Å of HfO_2, HfO_2 and SiO_2 = 58 and 4 Å, HfO_2 and SiO_2 = 38 and 7 Å, HfO_2 and SiO_2 = 19 and 10 Å, and HfO_2 and SiO_2 = 6 and 12 Å, respectively. Curve 6 is for SiO_2.

3.3. Effects of transition layer

In most of the previous studies of high-*k* stack structures, an abrupt transition between the high-*k* dielectric layer and interfacial layer is typically assumed. However, a gradual transition may exist in practice. Effects of interfacial layer and transition region on gate current performance have been exploited theoretically using the WKB approximation [2]. In this work, the present modeling tool is applied to a high-*k* stack structure with different transition thicknesses. Fig. 4 shows calculated gate current vs.

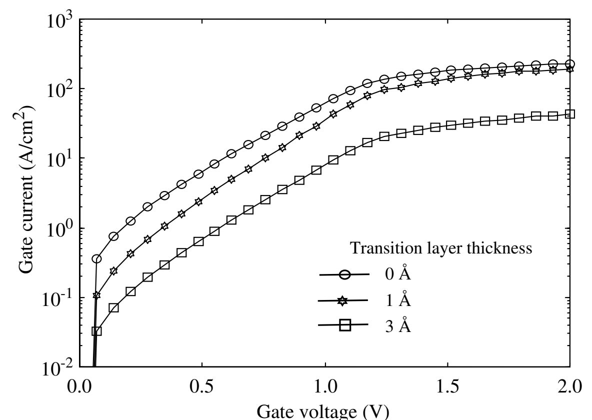

Fig. 4. Calculated gate current vs. gate voltage for high-*k* structures with transition layer between the high-*k* dielectric and silicon dioxide layers.

gate voltages of Si_3N_4-transition-SiO_2 structures with an EOT of 18 Å and an oxide thickness of 8 Å. The comparison between the calculated gate current for different transition thickness is illustrated in the figure to show the effects of transition layer. In this case, the gate current is reduced by the addition of a transition layer, and with increasing thickness of the transition layer for the same EOT. However, it is noted that the effect of gate current reduction with transition layer cannot be generalized since the magnitude of gate current in high-*k* stack structures with transition layer depends, in general, on the interplay between various factors of the structure and material parameters, such as the barrier height, electron effective masses, dielectric constant, as well as proportion of the individual layers. These effects have been discussed more extensively in [2] and will not be elaborated here.

3.4. Comparison with experiment

The present model produces very agreeable results between modeled gate current and available experimental data for various high-*k* stack structures. In Fig. 5 the results of calculated gate current vs. gate voltage and their corresponding experimental results for some high stack structures are compared. Group 1 curves in the figure plot calculated and measured gate current for a Al_2O_3–SiO_2 gate dielectric structure with an EOT of 18 Å, consisting of 25 Å of Al_2O_3 and 7 Å of SiO_2 layers. The measured data shown with the dotted line are taken from [12]. The group 2 curves correspond to a Al_2O_3–HfO_2–SiO_2 structure with an EOT of 16 Å, a 5 Å capping layer of Al_2O_3, 30 Å of HfO_2, and 9 Å SiO_2, respectively. The addition of the Al_2O_3 capping layer was proposed and demonstrated to greatly reduce the gate current to alleviate the effect of pinhole defects due to the interaction between the HfO_2 high-*k* dielectric and polysilicon gate materials in [13], where the measured data on the dotted line are taken. Both the group 1 and group 2 curves show very good agreement between the calculated and experimental results.

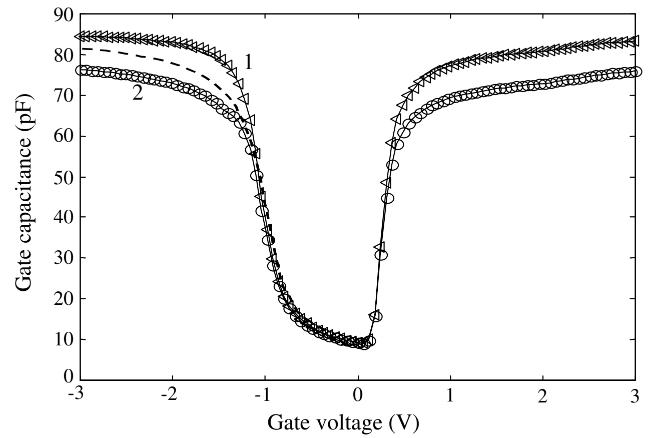

Fig. 6. Gate capacitance vs. gate voltage for a Al_2O_3–HfO_2–SiO_2 high-*k* stack structure. Curves 1 and 2 correspond to the results based on classical model and quantum model, respectively. The dotted line is based on experimental data taken from [10].

3.5. Gate capacitance of high-*k* stacks

It is well known that a high gate capacitance value is required for nanoscale MOSFET devices in order to maintain sufficient gate control over the channel conduction and device drive current. One of the features of the present modeling approach is that the gate capacitance can be evaluated simultaneously from the self-consistent solutions of the Schrödinger and Poisson equations, by integrating the space charges over the charge dipole in the device and then differentiating the space charges with respect to gate voltage, as described in [7]. Fig. 6 shows an example of calculated gate capacitance vs. gate voltage for the high-*k* Al_2O_3–HfO_2–SiO_2 stack structure calculated in the previous section, where curves 1 and 2 are results based on the classical model and quantum model, respectively.

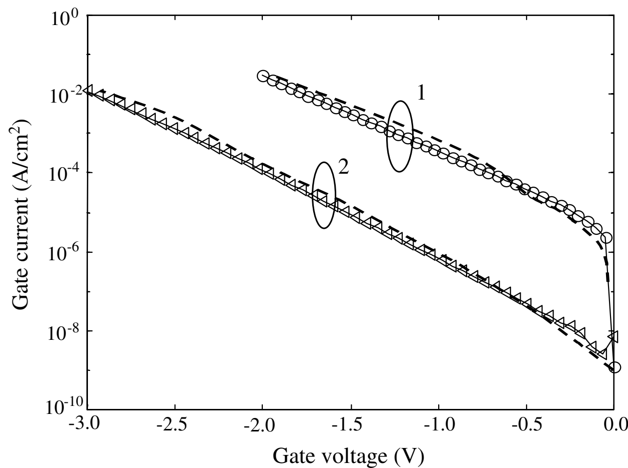

Fig. 5. Calculated gate current vs. gate voltage as compared with experimental data [9,10] for Al_2O_3–SiO_2 and Al_2O_3–HfO_2–SiO_2 high-stack structures.

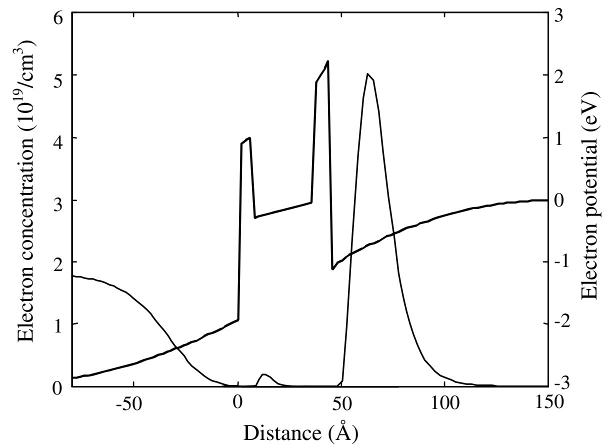

Fig. 7. An example of calculated electron charge and potential distributions obtained from self-consistent solutions to the Schrödinger and Poisson equations, from which gate capacitances are modeled.

The dotted line in Fig. 6 is based on the experimental data taken from [13], showing very good agreement. In Fig. 7, the calculated electron charge and potential distributions are plotted, from which the gate capacitances are evaluated. We note that the calculated charges in the various regions of the device are scaled for the purpose of illustration. The physical significance of the charge and potential profiles have been discussed in [7].

4. Conclusion

In this paper, a numerical model based on self-consistent solutions to the Schrödinger and Poisson equations is employed to examine effects of high-*k* gate stacks for nanoscale MOSFETs. The present model has the following features as being conceptually simple, numerically efficient and stable, and especially applicable to various high-*k* stack structures consisting of combination of different dielectric materials and thickness, as well as the inclusion of the transition and capping layers. Moreover, the device gate current and capacitance can be evaluated simultaneously which generate very good agreement with the available experimental results.

Acknowledgement

The research was partially supported by an office of Naval Research grant.

References

[1] Fan YY, Nieh RE, Lee JC, Lucovsky G, Brown GA, Register LF, et al. Voltage and temperature dependent gate capacitance and current model: Application to ZrO_2 MOSCAPs. IEEE Trans Electron Dev 2002;49:1969–78.

[2] Fan YY, Xiang Q, An J, Register LF, Banerjee SK. Impact of interfacial layer and transition region on gate current performance for high-*k* gate dielectric stack: Its tradeoff with gate capacitance. IEEE Trans Electron Dev 2006;50:433–9.

[3] Kizilyalli IC, Huang RYS, Roy PK. MOS transistors with stacked SiO_2–Ta_2O_5–SiO_2 gate dielectrics for giga-scale integration of CMOS technologies. IEEE Trans Electron Dev Lett 1998;19(11):423–5.

[4] Zhao Y, White MH. Modeling of direct tunneling current through interfacial oxide and high-*k* gate stacks. Solid-State Electron 2004;48: 1801–7.

[5] Zhang J, Yuan JS, Ma Y, Oates AS. Design optimization of stacked layer dielectrics for minimum gate leakage currents. Solid-State Electron 2000;44:2165–70.

[6] Mains RK, Sun JP, Haddad GI. Observation of intrinsic bistability in resonant tunneling diode modeling. Appl Phys Lett 1989;55:371–3.

[7] Sun JP, Mains RK, Chen WL, East JR, Haddad GI. *C–V* and *I–V* characteristics of quantum well varactors. J Appl Phys 1992;72: 2340–6.

[8] Sun JP, Haddad GI, Mazumder P, Schulman J. Resonant tunneling diodes: Models and properties. Proc IEEE 1998;86:641–61.

[9] Hou YT, Li MF, Yu HY, Kwong D-L. Modeling of tunneling currents through HfO_2 and $(HfO_2)_x(Al_2O_3)_{1-x}$ gate stacks. IEEE Electron Dev Lett 2003;24:96–8.

[10] Mudanai S, Fang YY, Quyang Q, Tasch AF, Banerjee SK. Modeling of direct tunneling current through gate dielectric stacks. IEEE Trans Electron Dev 2000;47:1851–7.

[11] Fan YY, Nieh RE, Lee JC, Lucovsky G, Brown GA, Register LF, et al. Voltage and temperature dependent gate capacitance and current model: applications to ZrO_2 n-channel MOS capacitor. IEEE Trans Electron Dev 2002;49:1969–78.

[12] Kuo CS, Hsu JF, Huang SW, Lee LS, Tsai MJ, Hwu JG. High-*k* Al_2O_3 gate dielectrics prepared by oxidation of aluminum film in nitric acid followed by high-temperature annealing. IEEE Trans Electron Dev 2004;51:854–8.

[13] Yang CW, Fang YK, Chen CH, De Wang W, Lin TY, Wang MF, et al. Dramatic reduction of gate leakage current in 161 nm HfO_2 high-*k* dielectric polysilicon gate with Al_2O_3 capping layer. Electron Lett 2002;38:1223–5.