ELSEVIER

Computer Physics Communications 147 (2002) 214–217

Computer Physics Communications

www.elsevier.com/locate/cpc

Numerical simulation of quantum effects in high-k gate dielectric MOS structures using quantum mechanical models

Yiming Li [a,*], Jam-Wem Lee [b], Ting-Wei Tang [c], Tien-Sheng Chao [d], Tan-Fu Lei [b], S.M. Sze [a,b]

[a] *National Nano Device Laboratories, Hsinchu 300, Taiwan*
[b] *Institute of Electronics, National Chiao Tung University, Hsinchu 300, Taiwan*
[c] *Department of Electrical and Computer Engineering, Univ. Mass., Amherst, MA 01033, USA*
[d] *Department of Electrophysics, National Chiao Tung University, Hsinchu 300, Taiwan*

Abstract

In this paper the electrical characteristics of metal oxide semiconductor (MOS) capacitors with high-k gate dielectric are investigated with quantum mechanical models. Both the self-consistent Schrödinger–Poisson (SP) model and the density gradient (DG) model are solved simultaneously to study quantum confinement effects (QCEs) for MOS capacitors. A computationally efficient parallel eigenvalue solution algorithm and a robust monotone iterative (MI) finite volume (FV) scheme for the SP and DG models are systematically proposed and successfully implemented on a Linux cluster, respectively. With the developed simulator, we can extract the effective gate oxide thickness from capacitance voltage (C-V) measurements for TaN and Al gate NMOS capacitors with Z_rO_2 and S_iO_2 gate dielectric materials. We found that quantization effects of 5.0 nm Z_rO_2 MOS samples cannot be directly equivalent to commonly quoted effects of 1.5 nm S_iO_2 MOS samples. Achieved benchmarks are also included to demonstrate excellent performances of the proposed computational techniques.

PACS: 72.15.Rn; 73.40.Ty; 73.40.Qv; 02.70.Fj; 02.70.-c

Keywords: MOS capacitor; High-k dielectric; Quantum mechanical models; Numerical methods

1. Introduction

The operation of deep submicron MOSFETs is now entering a regime in which QCEs become noticeable and classical transport theory is no longer sufficient for accurate modeling of the operating characteristics. One of the most obvious QCEs is that the inversion layer charge density calculated using a quantum mechanical approach is smaller than that calculated classically (only solve Poisson equation) [1] for a given applied gate voltage, thus directly affecting the shift of the C-V curves. The classical method overestimates the oxide thickness of the prepared ultrathin S_iO_2 films and has received great attention in recent years [2–5]. The development of metal gate and high-k dielectric materials, such as $(B_a, S_r)T_iO_3$, T_iO_2, or Z_rO_2, is a novel alternative to fabricate devices with

* Corresponding author. Address for correspondence: Microelectronics and Information Systems Research Center, National Chiao Tung University, Hsinchu 300, Taiwan.
E-mail address: ymli@cc.nctu.edu.tw (Y. Li).

0010-4655/02/$ – see front matter
PII: S0010-4655(02)00248-5

reliable and high-quality characteristics [6–9]. However, so far there are only a few theoretically studies of QCEs for high-k materials in the literature. It is well known that the SP model is the best approach since it fully includes the QCEs, and the DG model is an approximated quantum transport approach that can be directly derived from quantum mechanics on some assumptions. In numerical simulations, to find the final self-consistent solution of the SP mode is a time-consuming task and the DG approach also introduces some artificial parameters b_n in the model, but they are still the best choices to model QCEs in semiconductor MOS devices.

In this paper, both the SP model and the DG model [2–5] are applied to study the QCEs of MOS capacitors with Z_rO_2 gate dielectric material numerically. To solve these models efficiently, we propose a parallel iterative eigenvalue solution method for the solution of the SP model [10,11] and a monotone iterative finite volume scheme [12–17] for solving the DG model. The quantum transport results are found to compare quite well with experimental data. Achieved benchmarks demonstrate excellent performances of the proposed computational techniques. We have found that quantization effects of 5.0 nm Z_rO_2 MOS samples cannot be directly equivalent to commonly quoted effects of 1.5 nm S_iO_2 MOS samples.

This paper is organized as follows. Section 2 states the SP and DG numerical models and introduces computational algorithms for these models in the MOS simulation. Section 3 presents the simulation results. Section 4 draws the conclusion.

2. Physical models and solution techniques

We solve a 1D poly (metal)/Z_rO_2 (S_iO_2)/Si MOS structure along the z-direction. When electrons are confined by the surface potential, the electrical characteristics of MOS structures can be modeled by solving SP or DG models in the oxide and silicon regions without tunneling. The SP model is as follows:

$$-\frac{\hbar^2}{2}\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{m_{ni}^*}\frac{\mathrm{d}}{\mathrm{d}z}\right)\Psi_{i,j} + E\Psi_{i,j} = E_{i,j}\Psi_{i,j}, \tag{1}$$

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\varepsilon\frac{\mathrm{d}}{\mathrm{d}z}\right)\phi = \frac{q}{\varepsilon_0}\left(n - p + N_A^-\right), \tag{2}$$

$$E(z) = -q(\phi + \Delta E_c), \tag{3}$$

$$n = \frac{k_B T_L}{\pi\hbar^2}\sum_i g_i m_{di}^* \times \sum_j \ln\left(1 + \exp\left(\frac{E_F - E_{i,j}}{k_B T_L}\right)\Psi_{i,j}^2\right), \tag{4}$$

where the unknowns to be solved are $E_{i,j}$, $\Psi_{i,j}$, ϕ, and n. All variables and constants have the same meaning as in [1,3,5]. The above model to describe hole quantization phenomena is similar to Eqs. (1)–(4). In addition, a DG model is a set of coupled nonlinear equations and is given by [2,4]

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\varepsilon\frac{\mathrm{d}}{\mathrm{d}z}\right)\phi = \frac{q}{\varepsilon_0}\left(n - p + N_A^-\right), \tag{5}$$

$$2b_n\frac{\mathrm{d}^2\sqrt{n}}{\mathrm{d}z^2} = \left(\phi_F - \phi + \frac{k_B T_L}{q}F_{1/2}^{-1}\left(\frac{y^2}{N_c}\right)\right)\sqrt{n}, \tag{6}$$

where the unknowns are ϕ and n. For clarification of the analysis, we set the dielectric constant equal to 11.9, 3.9 and 20 for silicon, S_iO_2 and Z_rO_2, respectively. The choice of b_n is crucial and should be calibrated from the SP simulation results. The DG theory has been established as a possible alternative to the solution of the SP model for solving problems related to QCEs in MOS devices. Primary advantages of the DG model over the SP model are flexibility in extending to 2D and 3D, and easiness in incorporation into the conventional drift-diffusion or hydrodynamic simulator [12,15–17].

2.1. Iterative method for the SP model

The method for solving the self-consistent SP model [5] is: (i) set initial potential barrier and solve the Schrödinger equation, (ii) compute 2D electron gas with Eq. (4), (iii) solve the Poisson equation with the MI method, and (iv) update new potential Eq. (3) and back (i). The iteration will be terminated when a specified stopping criterion is reached. In our calculation experience, after about 15–20 feedback nonlinear iterations, the maximum norm error of ϕ and $\log n$ can be down to 10^{-9} monotonically. To fast solve the Schördinger equation in (i), we apply a parallel divide-conquer (DC) method for calculating the wave functions and energy levels of the large sparse band matrix.

2.2. Parallel divide-conquer algorithm

In this algorithm, the global matrix is split and scheduled equally by a host server processor and then sent to each client processor on a Linux cluster with message passing interface (MPI) library. The eigensystems are computed independently with the QR method. Only the eigenvalues will be estimated from these submatrices. Finally, the computed results are conquered by a merge method and returned to the host processor. If all exact eigenvalues are obtained from each matrix, the solved eigenvalues will be merged directly and hence the corresponding wave functions will be constructed with the inverse iteration method. If the eigenvalues are partially obtained, the multiplicity of approximated eigenvalues will be checked by a Householder transformation or by solving a secular equation for these eigenvalues. Our Linux cluster contains 16 PCs; file access and share are through a network file system and a network information system. The user datagram protocol controlled by MPI is applied to the short distance communication. In our simulation experiences, this method is stable and for a large class of matrices it is, asymptotically, faster by an order of magnitude than the conventional method.

2.3. MI and FV methods for the DG model

The FV discretized [13] DG model leads to a system of nonlinear algebraic equations. It is directly solved by the MI method instead of the conventional Newton's iterative method [12,17]. This approach has a global convergent behavior in MOS quantum simulations. The model is simulated with the following approach:

(a) solve the Poisson equation (5) with the MI method,
(b) with new ϕ, we solve the DG equation (6) using the MI method, and
(c) update new n and back (a).

The MI method was proposed and successfully applied in 2D and 3D semiconductor device simulation by us earlier [5,12,15–17]. Similarly, the iteration will be terminated when a specified stopping criterion is reached. Our calculation experience shows that it has a fast convergence behavior. It takes about 50–60 iterative loops to meet a given convergence criterion (maximum norm error $< 10^{-9}$ on ϕ and $\log n$).

3. Results and discussion

For a 2.2 nm S_iO_2 sample, Fig. 1 shows the electron concentration distribution for classical and quantum approaches; classical theory has higher electron density (≈ 4 times) and quantum transport theory has about 0.8–1.0 nm density peak shift from the interface. For 5.0 nm Z_rO_2 sample, we also have a similar result. This is one of the reasons that the classical method results in overestimated MOS capacitances (see Fig. 2).

As shown in Fig. 2, we find that the C-V curve computed with the classical approach is away from the measured result. Compared with the SP, DG and measured curves, the classical method has about 20% overestimation at $V_G = -2$ V. Secondly, it is clearly that both the SP and DG have a good consistence in this simulation, where the b_n is taken as $\hbar^2/12qm_n^*$. The equivalent oxide thickness (T_{ox}) is an important parameter to characterize or monitor the performance

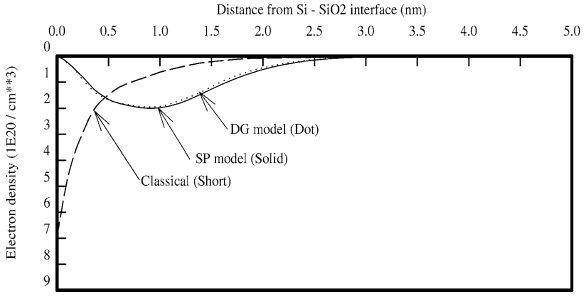

Fig. 1. Electron density for a 2.2 nm MOS at $V_G = 3$ V.

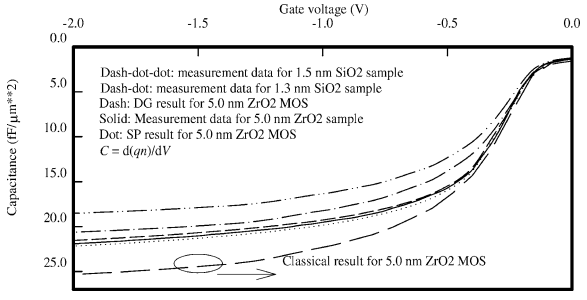

Fig. 2. Computed C-V curves for MOS capacitors.

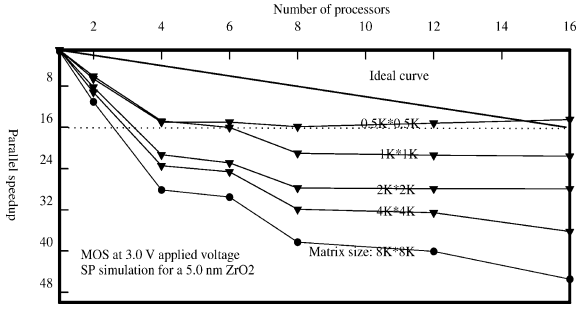

Fig. 3. Achieved speedup for parallel SP simulation.

and reliability of the devices and it equals $\varepsilon_{S_iO_2} * (\text{Area}/C_{\text{meas}})$, where $\varepsilon_{S_iO_2}$ is the dielectric constant of the silicon dioxide, Area is the area of the MOS capacitor. The C_{meas} is the measured capacitance of the capacitor and it is strongly dependent on the QCEs. In the experiment, the T_{ox} without quantum calibration of the Z_rO_2 sample is 1.5 nm (C_{meas} is measured at $V_G = -3$ V), however, in Fig. 2, we find that the simulated capacitance of 1.5 nm silicon dioxide (dash-dot-dot line) is much lower than the measured capacitance (solid line). This indicates that the QCEs should cause $\approx 12\%$ error of thickness. In other words, the Z_rO_2 film is more similar to a 1.3 nm oxide (dash-dot line) than to a 1.5 nm oxide. We have found the difference of QCEs between 5.0 nm Z_rO_2 and 1.5 nm S_iO_2 samples is not exactly the same. That is, the peak of the carrier density of the 5.0 nm Z_rO_2 is farther away from the interface of silicon than that of the 1.5 nm S_iO_2. This observation shows that the 5.0 nm Z_rO_2 cannot be exactly equal to S_iO_2 with any thickness. Thus, exact film thickness and dielectric constant are necessary to simulate MOS with high k films. Fig. 3 shows the excellent speedup performance of the proposed parallel solution algorithm for the SP simulation.

4. Conclusion

In conclusion, the electrical characteristic of MOS capacitors with high-k gate dielectric material Z_rO_2 has been investigated with the SP and DG quantum mechanical models. Computationally efficient simulation algorithms for these models have been presented and successfully implemented on a Linux cluster. The developed simulator has good computing performance and has been applied to extract the effective gate oxide thickness from C-V measurements for MOS capacitors with Z_rO_2 and S_iO_2 gate dielectric materials. We found that the quantization effects of 5.0 nm Z_rO_2 MOS samples cannot be directly equivalent to commonly quoted effects of 1.5 nm S_iO_2 MOS samples.

Acknowledgements

This work was supported in part by the National Science Council of Taiwan under Contract No NSC 90-2112-M-317-001.

References

[1] S.M. Sze, Physics of Semiconductor Devices, 2nd edn., Wiley-Interscience, New York, 1981.
[2] M.G. Ancona et al., Phys. Rev. B 42 (1990) 1222–1233.
[3] S.-H. Lo et al., IBM J. Res. Dev. 43 (1999) 327–337.
[4] M.G. Ancona, IEEE Trans. ED 47 (2000) 1449–1456.
[5] Y. Li et al., in: Proc. 3rd WSES/IEEE Inter. Conf. Math. Comput. Phys., 2001, pp. 4511–4516.
[6] C.A. Richter et al., IEEE EDL 22 (2001) 35–37.
[7] K. Matsuo et al., VLSI Tech. Dig. (2000) 70–71.
[8] I. Asano et al., IEDM Tech. Dig. (1998) 755–758.
[9] W. Liu et al., IEEE Trans. ED 46 (2000) 1070–1072.
[10] E.R. Jessup et al., SIAM J. Mat. Anal. Appl. 15 (1994) 1201–1226.
[11] F. Tisseur et al., SIAM J. Sci. Comp. 20 (1999) 2223–2236.
[12] Y. Li et al., in: Proc. IEEE Int. Symp. VLSI-TSA, 1999, pp. 27–30.
[13] R.S. Varga, Matrix Iterative Analysis, Springer, New York, 2000.
[14] T. Gallouet et al., SIAM J. Num. Anal. 37 (2000) 1935–1972.
[15] Y. Li et al., Comput. Phys. Comm. 142 (2001) 285–289.
[16] Y. Li et al., in: Tech. Proc. Int. Conf. Modeling and Simulation of Microsystem, 2001, pp. 562–566.
[17] Y. Li, WSEAS Trans. Systems 1 (2002) 68–73.