

Well-Tempered MOSFETs: 1D versus 2D Quantum Analysis

A. Abramo, L. Selmi
 DIEGM – University of Udine
 via delle Scienze 208, 33100 Udine (Italy)
 Z. Yu, R. W. Dutton
 Center for Integrated Systems
 Stanford University, Stanford, CA 94305 (USA)

Abstract—This paper presents the two dimensional, quantum mechanical simulation of scaled “Well-Tempered MOSFETs” featuring different effective channel lengths in the deep sub-micron range. The simulation results were obtained by means of a two-dimensional Schrödinger solver that had been previously applied to idealized MOS structures. Comparison between one- and two-dimensional approaches is presented, and the difference between the two models are highlighted.

I. INTRODUCTION

THE successful design of nano-scale MOSFETs [1] and the demands of the SIA Roadmap in terms of device features and deadlines for future technologies [2], constantly stimulate the research towards the development of predictive simulation models, to describe quantitatively device operations at such reduced channel lengths (L_{eff}).

Recently, an open-forum for the validation of different modeling tools has been proposed [3], and the paraphrastic epithet of “Well-Tempered MOSFET” (WTM) has been forged to refer to a class of well-designed, hopefully well-behaving n-channel MOSFETs, whose ultimate performance limits are being demonstrated [4], [5].

In this paper we would like to contribute to the discussion, presenting the two-dimensional (2D) quantum mechanical (QM) simulation of WTMs of different L_{eff} , obtained by using the 2D Schrödinger solver described in [6], previously applied to idealized MOS structures.

In the following, a brief summary of the physical and numerical features of the simulator will be presented, then results from the 1D and 2D models will be shown, and the difference between the two approaches will be analyzed.

II. PHYSICS AND NUMERICS SUMMARY

While the 1D [7], [8] or quasi-2D [9], [10] simulation of QM effects in MOS transistors is rather assessed and viable even for development purposes, the 2D QM simulation is still at a preliminary stage, owing to its heavy computational and memory requirements.

Since a fundamental step for the first-principles solution of the QM problem is to solve the Schrödinger equation, to

partially overcome the aforementioned limitations without resorting to specific or simplified approaches [11], [12], [13], we developed a solver for the 2D Schrödinger equation in the $\{p\}$ -representation [6], which we apply here to the simulation of nano-scale WTMs.

The simulator solves the Schrödinger equation projecting the solution on the basis of the unperturbed Hilbert space of a closed-boundary system. In other words, the following representation for the unknown eigenfunction Ψ_H relative to the H -th eigenvalue is assumed:

$$\begin{aligned} \Psi_H(x, y) &= \sum_{i,j}^{M,N} A_{ij}^{(H)} \phi_{ij} = \sum_{K=1}^{(M \times N)} A_{HK} \phi_K = \\ &= \sum_{K=1}^{(M \times N)} A_{HK} \sin(k_i x) \cdot \sin(k_j y) \end{aligned} \quad (1)$$

where M and N are the number of the discretization points in the x and y directions, A_{HK} is the amplitude of the K -th component ϕ_K of the searched solution Ψ_H relative to the H -th unknown eigenvalue ϵ_H , $k_i = \pi i/L_x$, $k_j = \pi j/L_y$ are the x and y components of the k -vector, and L_x , L_y are the dimension of the simulation domain. Substituting this expression of Ψ_H into the Schrödinger equation, the following linear system of equations is obtained:

$$\sum_{K=1}^{(M \times N)} \left[\left(V_{HK} + \frac{\hbar^2}{2m_H^*} (k_i^2 + k_j^2) \delta_{HK} \right) - \epsilon_H \delta_{HK} \right] \times A_{HK} = 0 \quad (2)$$

where $H = 1, \dots, (M \times N)$, δ_{HK} is the Kronecker delta, $(M \times N)$ is the 2D discrete grid, m_H^* is the effective electron mass in the free direction (z) for the H -th eigenvalue of the valley the electron belongs, and:

$$\begin{aligned} V_{HK} = \langle \phi_H | V | \phi_K \rangle &= \iint \phi_H V(x, y) \phi_K dx dy = \\ &= \iint F_H(x, y) \phi_K dx dy \end{aligned} \quad (3)$$

is the matrix element of the electrostatic potential in the $\{p\}$ -representation which we efficiently computed by

means of two-dimensional fast-Fourier transform (FFT) of the $(M \times N)$ functions $F_H(x, y) = \phi_H V(x, y)$, performed using the code described in [14].

It must be noticed that the spectral expansion (1) allows to reconstruct Ψ_H in points of the real space which are not necessarily coincident with those of the discretization grid, thus enabling to evaluate the quantities of interest (e.g. the electron concentration) even on finer meshes than those used for the QM solution.

The linear system of equations (2) is then solved using standard libraries [15], thus obtaining the $(M \times N)$ eigenstates of the system under investigation. The accuracy of the solution against grid dimensions was verified. Consequently all simulations were performed using a 60×60 discretization grid, i.e. finding the eigenvalues and eigenvectors of a 3600×3600 matrix. The resulting numerical workout was ≈ 7 hours per simulation on a SPARC ULTRA/170 workstation, equipped with a 163 MHz UltraSPARC processor rated at 252 SPECint92 and 351 SPECfp92, and 1 GBytes RAM.

III. SIMULATION RESULTS

In the modeling framework described above, we simulated four different nano-scale MOS transistors, featuring $L_{\text{eff}} = 90, 25, 15,$ and 10 nm. The two longer devices can be properly considered WTMs, since they were obtained directly from [3], while the two shorter ones were derived shrinking the 25 nm WTM profile down to 15 nm and 10 nm, as described in [3]. Since no optimization of the two shorter structures was done, the quasi-WTM denomination may be somehow more appropriate for such devices.

We also simulated an “abrupt-junction, no-HALO” version of the two quasi-WTMs, to emphasize the 2D nature of the electrostatic potential around the junctions, as it could originate from substantial short channel effects (SCE). For this reason, we will call them non-WTMs in the following.

The Poisson-Schrödinger self-consistency was not included in the simulation scheme of this paper. Its inclusion, however, would change only the quantitative meaning of the comparison, while its qualitative significance, that is to point out the possible origin of mutual discrepancies between of the 1D and the 2D approaches, still maintains its validity.

Given the reason above, only non self-consistent solutions of the 1D and 2D Schrödinger equations were obtained in the different devices (which were biased at $V_{\text{DS}} = 0$ V, $V_{\text{GS}} > V_{\text{TH}}$), and the Fermi levels were used as a fitting parameter to obtain a quantitative agreement between the QM electron concentrations computed by the 1D and

2D models in the source of each simulated structure.

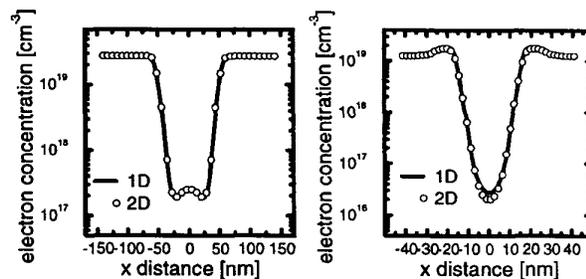


Fig. 1. QM electron concentration along the channel of the 90 nm (left) and 25 nm (right) WTMs obtained from [3]. The 1D (solid line) and 2D (symbols) solutions are compared. The oxide thickness is $t_{\text{ox}} = 4.9$ nm for the 90 nm WTM and $t_{\text{ox}} = 1.5$ nm for the 25 nm WTM.

Fig. 1 shows the 1D and 2D electron concentrations plotted along the 90 nm (left) and 25 nm (right) structures, biased above threshold, at the y coordinate of the 1D QM carrier concentration peak. The 2D curve was obtained plotting the 2D concentration along the longitudinal line at such fixed y depth, while the 1D curve was obtained taking the 1D QM value at the same y coordinate, as obtained from a transverse 1D Schrödinger simulation at the different x sections. As can be seen, no substantial difference is observed between the 1D and 2D solutions in these rather long WTM.

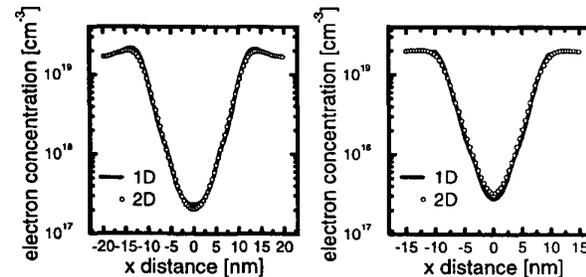


Fig. 2. QM electron concentration along the channel of the 15 nm (left) and 10 nm (right) quasi-WTMs. The MOS doping profiles were obtained as a technology shrink from the 25 nm WTM, as specified in [3]. The 1D (solid line) and 2D (symbols) solutions are compared. The oxide thickness is $t_{\text{ox}} = 1.5$ nm for both devices.

Instead looking at Fig. 2, referring to the 15 nm (left) and 10 nm (right) quasi-WTM, discrepancies between 1D and 2D solutions can be found, both around the junctions and inside the channel.

The effect is even more pronounced in Fig. 3, where the situation for the 15 nm (left) and 10 nm (right) non-WTM is depicted. This is due to the presence of SCE and of a quasi punch-through situation, resulting in a strong 2D

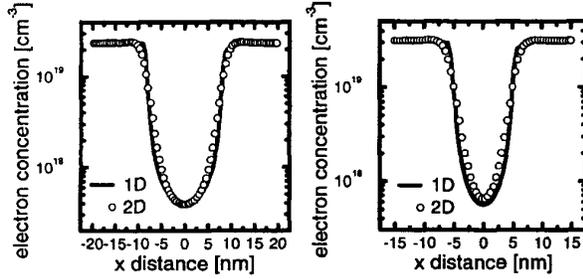


Fig. 3. QM electron concentration along the channel of the 15 nm (left) and 10 nm (right) non-WTM. The MOS structures feature constant bulk doping ($5 \times 10^{18} \text{ cm}^{-3}$), i.e. no HALO implants, and abrupt source and drain junctions. The 1D (solid line) and 2D (symbols) solutions are compared. The oxide thickness is $t_{ox} = 1.5 \text{ nm}$ for both devices.

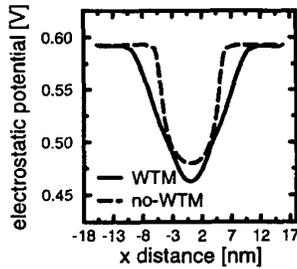


Fig. 4. The longitudinal electrostatic potential of the 10 nm quasi-WTM (solid line) and non-WTM (dashed line). The longitudinal section has been taken at the y coordinate of the 1D concentration peak.

distribution of the electrostatic potential in proximity of the junctions.

To clarify this point, Fig. 4 shows the longitudinal electrostatic potential of the 10 nm quasi-WTM and non-WTM, taken at the y coordinate of the 1D concentration peaks, while Fig. 5 shows the transverse electrostatic potential of the same devices taken at their mid-channel sections. A steeper longitudinal profile and a broader inversion layer can be observed in the non-WTM device, resulting in an enhanced 2D character of the electron concentration. This is a consequence of the absence of HALO extensions in this latter device.

As can be seen, the 2D character of the solution cannot be captured neither by a single 1D Schrödinger solution at mid channel [16] nor by a quasi-2D approach based on 1D Schrödinger solutions at different x coordinates [10]. Consequently, these results suggest the need of the accurate QM description of the charge and potential profiles around the junctions, especially for the precise prediction of SCE and charge injection into the channel, that critically depend on the actual shape of the source-to-channel

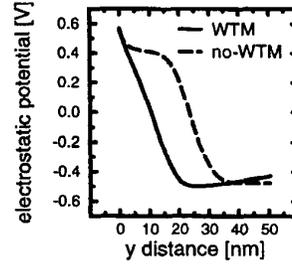


Fig. 5. The transverse electrostatic potential of the 10 nm quasi-WTM (solid line) and non-WTM (dashed line). The transverse section is taken at mid-channel.

barrier. In addition, QM carrier penetration from source to channel, adding to thermionic injection, is a further effect that 1D or quasi-2D quantum approaches will neglect, which instead may be of relevance in designing the next generation ultra-short WTM.

To be more confident on the obtained results, we also compared the 1D and 2D solvers in regions of the device where no difference was expected, such as inside the device source, where no 2D effects are normally present. As a result, the different numerical accuracy of the 1D and 2D Schrödinger solvers cannot be invoked to explain the highlighted differences between the 1D and 2D QM solutions.

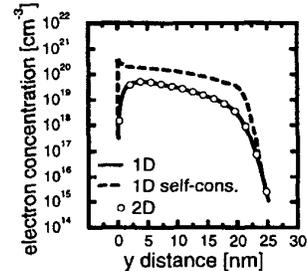


Fig. 6. The electron concentration inside the source of the 10 nm quasi-WTM, as computed by the 1D (dashed line) and 2D (symbols) Schrödinger solvers. The 1D self-consistent solution (solid line) is also shown.

In fact, a part from the relative adjustment of the Fermi levels, the two methods agree in computing the vertical electron concentration profile inside the transistor source, as can be seen in Fig. 6 for the 10 nm quasi-WTM (compare solid line and open circles). Here, the 1D self-consistent QM solution was also included, showing the charge accumulation at the Si-SiO₂ interface due to the presence of the surface hard-wall boundary condition enforced to the electron wave-functions by the high Si-SiO₂ barrier. The constraint of zero wave-function at the surface, hence zero charge, forces the potential to increase, thus inducing a very thin accumulation layer. This re-

sult suggests the need of a self-consistent iteration between Poisson and Schrödinger equation for a more quantitative prediction of the 2D effects.

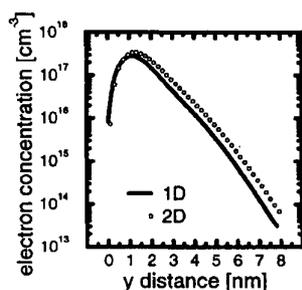


Fig. 7. The electron concentration at the mid-channel of the 10 nm quasi-WTM, as computed by the 1D (solid line) and 2D (dashed line) Schrödinger solvers.

Supported by this validation, we observe that 1D and 2D methods predict different results inside the channel. Fig. 7, in fact, shows that, besides different 1D and 2D concentration values, a different position of the concentration centroids can also be observed.

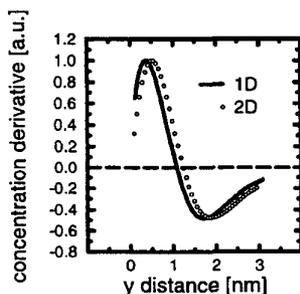


Fig. 8. Derivative of the electron concentrations of Fig. 7. For the sake of comparison, the derivatives have been normalized to unity to put in evidence the relative shift of the concentration peak.

This is clearly seen in Fig. 8, where the derivatives of the 1D and 2D electron concentrations of Fig. 7, normalized to unity for sake of comparison, are shown. A peak shift of ≈ 0.15 nm can be extracted, certainly affecting the CV characteristics of these devices with $t_{ox} = 1.5$ nm.

IV. CONCLUSIONS

In conclusion, this paper has shown that differences between 1D or quasi-2D and fully-2D QM models can be expected when simulating ultra-short WTMs. Effects that cannot be captured by 1D or quasi-2D QM models have been shown. The relevance of the issue will deserve the inclusion of self-consistency in the proposed 2D QM frame.

REFERENCES

- [1] G. Timp, J. Bude, K. Bourdelle, J. Garno, A. Ghetti, H. Gossmann, M. Green, G. Forsyth, Y. Kim, R. Kleiman, F. Klemens, A. Kornblit, C. Lochstampf, W. Mansfield, S. Moccio, T. Sorsch, D. Tennant, W. Timp, and R. Tung, "The ballistic nano-transistor," in *IEDM Tech. Dig.*, p. 55, 1999.
- [2] SIA, "International Technology Roadmap," tech. rep., Semiconductor Industry Association, 1999.
- [3] D. Antoniadis, I. Djomehri, K. Lackson, and S. Miller, "The Well-Tempered" Bulk-Si NMOSFET Device Home Page," in <http://www-mtl.mit.edu/Well/>, 1999.
- [4] F. Assad, Z. Ren, S. Datta, and M. Lundstrom, "Performance limits of Silicon MOSFET's," in *IEDM Tech. Dig.*, p. 547, 1999.
- [5] F. Assad, Z. Ren, D. Vasileska, S. Datta, and M. Lundstrom, "On the performance limits for Si MOSFET's: a theoretical study," *IEEE Trans. Electron Devices*, vol. 47, p. 232, 2000.
- [6] A. Abramo, A. Cardin, L. Selmi, and E. Sangiorgi, "Two-Dimensional Quantum Mechanical Simulation of Charge Distribution in Silicon MOSFET's," *IEEE Trans. Electron Devices*, vol. 47, p. (to appear), 2000.
- [7] S. Harelend, S. Krishnamurthy, S. Jallepalli, C. Yeap, K. Hasnat, A. T. Jr., and C. Maziar, "A computationally efficient model for inversion layer quantization effects in deep submicron n-channel MOSFET's," *IEEE Trans. Electron Devices*, vol. 43, p. 90, 1996.
- [8] S.-H. Lo, D. Buchanan, Y. Taur, and W. Wang, "Quantum-mechanical modeling of electron tunneling current from the inversion layer of ultra-thin-oxide nMOSFET's," *IEEE Electron Device Lett.*, vol. 18, p. 209, 1997.
- [9] M. Fischetti and S. Laux, "Monte Carlo study of electron transport in silicon inversion layers," *Phys. Rev. B*, vol. 48, p. 2244, 1993.
- [10] A. Spinelli, A. Benvenuti, and A. Pacelli, "Self-consistent 2-D model for quantum effects in n-MOS transistors," *IEEE Trans. Electron Devices*, vol. 45, p. 1342, 1998.
- [11] M. van Dort, P. Woerlee, and A. Walker, "A simple model for quantisation effects in heavily-doped silicon MOSFET's at inversion conditions," *Solid State Electron.*, vol. 37, p. 411, 1994.
- [12] C. Bowen, C. Fernando, G. Klimeck, A. Chatterjee, D. Blanks, R. Lake, J. Hu, J. Davis, M. Kulkarni, S. Hattangady, and I.-C. Chen, "Physical oxide thickness extraction and verification using quantum mechanical simulation," in *IEDM Tech. Dig.*, p. 869, 1997.
- [13] C. Rafferty, B. Biegel, Z. Yu, M. Ancona, J. Bude, and R. Dutton, "Multi-dimensional quantum effect simulation using a density-gradient model and script-level programming techniques," in *SIS-PAD Tech. Dig.*, (Erlangen), p. 137, Springer-Verlag, 1998.
- [14] M. Frigo and S. Johnson, "'FFTW: An Adaptive Software Architecture for the FFT,'" in *ICASSP Tech. Dig.*, vol. 35, p. 1381, 1999.
- [15] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, "LAPACK: A portable linear algebra library for high-performance computers," tech. rep., Computer Science Dept, University of Tennessee, Knoxville, 1990.
- [16] K. Krisch, J. Bude, and L. Manchanda, "Gate capacitance attenuation in MOS devices with thin gate dielectrics," *IEEE Electron Device Lett.*, vol. 17, p. 521, 1996.