A Physical Model for Work-Function Variation in Ultra-Short Channel Metal-Gate MOSFETs

Seid Hadi Rasouli, *Student Member, IEEE*, Chuan Xu, *Student Member, IEEE*, Navab Singh, *Senior Member, IEEE*, and Kaustav Banerjee, *Senior Member, IEEE*

***Abstract*—In this letter, an accurate physical model for work-function variation (WFV) relevant to ultrashort-channel ($<$ 32 nm) MOSFETs has been formulated that considers the work function and size of the individual grains in determining the local MOS band structure. The proposed model is shown to be much more accurate than the most recently published model. Additionally, using this new model, WFV effect in a 3-D device can be captured using 2-D device simulation, resulting in a significantly lower simulation time.**

***Index Terms*—Grain orientations (GOs), metal gate, quasi-ballistic transport, work-function variation (WFV).**

I. Introduction

WORK-FUNCTION variation (WFV) has been recently identified [1], [2] and experimentally verified [3]–[5] as the main source of threshold-voltage (V_T) variation in emerging metal-gate devices, where the gate is composed of a small number of grains [1]–[3]. WFV arises due to the dependence of the work function on the grain orientations (GOs), which cannot be controlled in the IC fabrication process resulting in V_T variation. WFV-induced V_T variation increases for a lower number of grains on the gate [1]–[7]; therefore, it is crucial to accurately model the effect of WFV on the threshold voltage of metal-gate devices.

II. WFV Modeling

A. Existing WFV Models

A statistical model was proposed in [1]–[3] that assigns an average WF to the entire gate, which provides reasonable accuracy for a relatively large number of grains (> 10) in the gate and can be conveniently used in circuit simulations. However, the problem becomes complicated for ultrashort-channel devices, with only a few grains on the gate. While "atomistic" simulation methodology [6] can be employed for such devices, it is not practical for circuit analyses. Another approach (network model) has been proposed for such small devices in [7], where each grain in the gate is replaced by a transistor. However, in the network model, it is assumed that the charge distribution of any specific region in the channel is determined solely by the grain above the given region. To explore the validity of this assumption, as an example, we consider a SOI NMOS with a channel length of 22 nm, a body thickness of 12 nm, and an effective oxide thickness (EOT) of 1.5 nm (top view of the device without and with WFV is shown in Fig. 1(a) and (b), respectively). It is assumed that the gate is composed of two types of grains with WFs of 4.3 and 4.6 eV (dark and light areas, respectively). Considering the average grain size for different metals (such as 22 nm for TiN and 17 nm for MoN [1]–[3]) in sub-22-nm technology nodes, it can be assumed that there is only one grain along the channel length (as it will be explained later, the model proposed in this letter can be easily applied to the case where there are more than one grain along the channel). As shown in Fig. 1(c), based on the network model, the device is composed of five parallel branches. Fig. 1(d) shows the electron sheet density along the width of the device based on the network model and a 3-D device simulator [8]. It can be observed that the network model substantially overestimates or underestimates the charge in different parts of the channel. The main cause of this unacceptable error is that the network model only considers the effect of the WF of a metal grain on the channel electrostatic potential and charge density under such grain but neglects the effect of the WF of adjacent metal grains. The effect of the neighboring grains arises due to the WFV-induced fringing

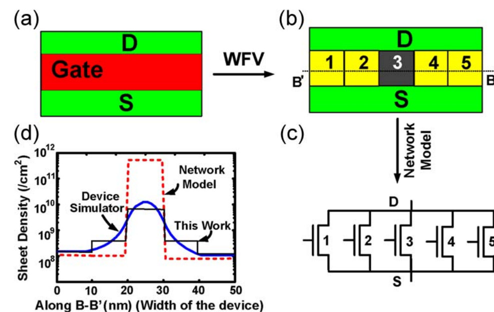

Fig. 1. Schematic of a metal-gate NMOS ($L = 22$ nm, $W = 50$ nm, and the average grain size along the width of the device is assumed to be 10 nm). (a) Without and (b) with WFV. (c) In the network model, each grain is substituted by a subtransistor. (d) Comparison between electron sheet density ($V_{\rm GS} = 0.2$ V and $V_{\rm DS} = 10$ mV) predicted by a 3-D device simulator [8] and by the network model along B-B'. WF of grain #3 is assumed to be 4.3 eV, and the WFs of the rest of the grains are assumed to be 4.6 eV. Note that, under the chosen conditions, the device operates in the linear region based on the network model, and the charge sheet density is around $10^{12}/\text{cm}^2$, while based on the proposed model, in agreement with 3-D simulation, the device is still in the subthreshold regime due to fringing electric field from or to the neighboring grains as well as the charge diffusion under the grains with different WFs.

Manuscript received July 9, 2011; revised August 14, 2011; accepted August 18, 2011. Date of publication September 28, 2011; date of current version October 26, 2011. This work was supported by UC Discovery (Intel) Award SB090042. The review of this letter was arranged by Editor L. Selmi.

S. H. Rasouli, C. Xu, and K. Banerjee are with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA (e-mail: hadi@ece.ucsb.edu; chuanxu@ece.ucsb.edu; kaustav@ece.ucsb.edu).

N. Singh is with the Institute of Microelectronics, Agency for Science, Technology and Research (A*STAR), Singapore 117685 (e-mail: navab@ime.a-star.edu.sg).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LED.2011.2166531

0741-3106/$26.00 © 2011 IEEE

electric field from or to the neighboring grains as well as the charge diffusion from the high density region (let us say the region below a grain with low WF in n-type devices) to a neighboring low density region (region below grains with high WF in n-type devices). Therefore, the charge density below a specific grain not only depends on the WF of that grain but also depends on the WFs of its neighboring grains.

B. New Physical Model

In order to consider the effect of adjacent metal grains in different parts of the channel, we assign an effective WF to each grain as explained hereinafter. Let us say that, in Fig. 1(b), we want to assign an effective WF to grain #3. The effective WF of grain #3 is given by

$$WF_{\text{eff–3}}=WF_3+\alpha_1(\Delta WF_2+\Delta WF_4)+\alpha_2(\Delta WF_1+\Delta WF_5) \quad (1)$$

where ΔWF_i is the difference between the WF of the ith grain and the WF of grain #3. ΔWF_i is a positive (negative) number if the WF of the ith grain is greater (smaller) than WF_3. α_1 and α_2 are the coefficients, which involve the WF of neighboring grains in the calculation of the charge concentration below each grain. The linear dependence of the effective WF on the WFs of the neighboring grains is explained as follows. The effective WF is extracted from the charge concentration in the channel. The charge concentration (in log scale) in the subthreshold regime is linearly dependent on the flatband voltage and, hence, on the WF of the metal. As a result, the effective WF linearly depends on the WFs of the individual grains in the gate as given by (1). It is expected that $\alpha_1 > \alpha_2$, since grains #2 and #4 are closer to grain #3 than grains #1 and #5.

Effect of Grain Size, Temperature, and EOT: The effect of the neighboring grains also depends on the average grain size. If the grain size is large (small), the neighboring grains have lower (higher) effect on the charge below a given grain. In this letter, we find these parameters through rigorous Monte Carlo simulations (needed only once) for various grain sizes from 4 to 22 nm and an EOT of 1.5 nm at a temperature of 300 K ($\alpha_1 = 0.25$ and $\alpha_2 = 0.05$ for NMOS devices considering an average grain size of 10 nm). Coefficients for other grain sizes are derived from (2) and (3), where GS is the average grain size (in nanometers) and 4 nm < GS < 22 nm

$$\alpha_1(GS) = 0.25 - 0.017(GS - 10) \quad (2)$$
$$\alpha_2(GS) = 0.05 - 0.003(GS - 10). \quad (3)$$

While α_1 and α_2 are expected to increase with temperature (since higher temperature results in higher diffusion of carriers), our simulation results show that temperature has negligible effect on these parameters. The reason is that the difference between charge concentrations under the neighboring grains (with different WFs) reduces with temperature, which cancels out the effect of higher diffusion of carriers. α_1 and α_2 also depend on the EOT, as given by

$$\alpha_1(EOT) = \alpha_1(GS) + 0.06(EOT - 1.5) \quad (4)$$
$$\alpha_2(EOT) = \alpha_2(GS) + 0.008(EOT - 1.5) \quad (5)$$

where $\alpha_1(GS)$ and $\alpha_2(GS)$ are given by (2) and (3). EOT is the effective oxide thickness (in nanometers) (0.5 nm < EOT, 2.5 nm). For thinner (thicker) gate oxide, the effect of neighboring grains reduces (increases) since the gate has higher (lower) control on the channel.

TABLE I
ERROR IN PREDICTING THE SUBTHRESHOLD CURRENT OF AN n-TYPE MOSFET (CHANNEL WIDTH= 50 nm, $L_{\mathrm{ch}} = 22$ nm, $V_{\mathrm{GS}} = 20$ mV, $V_{\mathrm{DS}} = 20$ mV, EOT = 1.5 nm AND AVERAGE GS = 10 nm). NOTE THAT ALL DEVICES ARE IN THE SUBTHRESHOLD REGIME

WF1 (eV)	WF2 (eV)	WF3 (eV)	WF4 (eV)	WF5 (eV)	Current using 3D Simulator (nA)	Error (%) Network Model	Error (%) This Work
4.8	4.8	4.6	4.8	4.8	0.002	324	1.5
4.8	4.7	4.5	4.7	4.8	0.0758	893	1.5
4.8	4.4	4.8	4.4	4.8	0.248	4516	3.4
4.5	4.7	4.5	4.7	4.7	0.438	241	2.2
4.8	4.5	4.5	4.8	4.8	0.175	7400	2.2

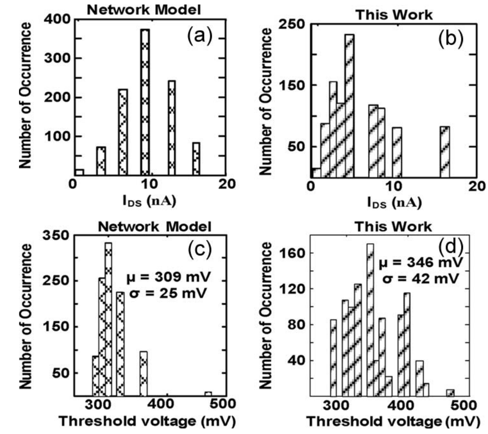

Fig. 2. MC simulation results for 1000 samples for the device shown in Fig. 1(b) with five grains along the channel width ($L = 22$ nm and $W = 50$ nm). It is assumed that there are two types of grains with WFs of 4.5 and 4.7 eV and with probabilities of 0.6 and 0.4, respectively. Vertical axes show the number of occurrences for different drain currents and threshold voltages. (a) Drain current distribution based on the network model. (b) Drain current distribution based on the model proposed in this work (in agreement with 3-D device simulator [see Table I]). (c) V_T distribution based on network model. (d) V_T distribution based on the proposed model. μ and σ are the mean value and standard deviation of V_T, respectively.

As it can be observed from Fig. 1(d), the proposed model can accurately predict the charge density inside the channel. Table I shows the error incurred in the prediction of drain current employing different models. In order to validate our model, different WFs are assigned to grains. Predicted drain currents by the network model and our model are compared with 3-D device simulation results. It should also be noted that the proposed model can accurately capture the effect of WFV on V_T variation. As an example, consider two cases of grain combinations [in the device shown in Fig. 1(b)]: Case 1) Grain #1 has low WF (4.3 eV), and other grains have high WF (4.7 eV), and Case 2) Grain #3 has low WF (4.3 eV), and other grains have high WF (4.7 eV). Previous models consider these two cases to be identical and predict the same threshold voltage for both cases (with unacceptable error compared to actual threshold voltage), while the proposed model, in agreement with 3-D device simulations, predicts two different threshold voltages for the aforementioned two grain combinations. For $V_{\mathrm{GS}} = 0.1$ V and $V_{\mathrm{DS}} = 10$ mV, the drain current in Case 2 is 22% of the drain current in Case 1. In Case 1, the effective WF of grain #1 is affected by the WFs of grains #2 and #3 (effective WF of grain #1 is 4.42 eV instead of the initial value

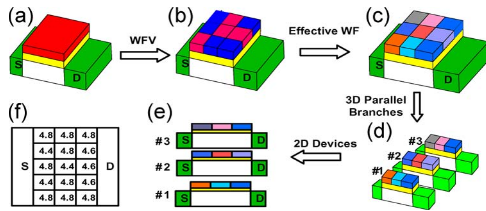

Fig. 3. Metal-gate device with two types of grains. (a) Without WFV. (b) With WFV. (c) Effective WFs of the grain are calculated using (2)–(5). Different colors [w.r.t. those in (b)] represent effective WFs of the grains. (d) After assigning effective WF to each grain, 3-D device is converted to parallel 3-D devices. (f) Top view of a hypothetical metal-gate device with three types of grains.

of 4.3 eV). In Case 2, the effective WF of grain #3 is affected by the WFs of grains #1, #2, #4, and #5 (effective WF of grain #3 is 4.54 eV instead of the nominal value of 4.3 eV). Fig. 2(a) and (b) shows the distribution of the drain current for the device shown in Fig. 1(b), where the maximum and minimum values of the drain current are found to be identical using the two methods (maximum drain current is for the case where all grains have a WF of 4.5 eV, and minimum drain current is for the case where all grains have a WF of 4.7 eV). However, the distribution of current is predicted erroneously by the network model, while our model accurately predicts the drain current for different grain combinations (as can be observed from Table I). Fig. 2(c) and (d) shows the V_T distribution based on the network model and the proposed model, respectively, where V_T distribution is significantly underestimated by the network model. It is expected that the "atomistic" simulation methodology [6] will predict slightly higher standard deviation of the threshold voltage than the one predicted by our model due to the effect of randomness in the shape of the grains. However, the effect of the randomness in the shape of the grains is expected to decrease for a lower number of grains on the gate.

It is worth noting that α_1 and α_2 can be derived for any metal-gate technology and process condition [using (2)–(5)] within the mentioned ranges of EOT and GS. From a circuit designer's perspective, knowing the size of transistors, the average GS and EOT, WFV-induced V_T distribution, and its effect on circuit performance and reliability parameters can be accurately estimated using the proposed model. For the general case, where the number of grains along the channel length is greater than one, a similar approach can be followed, as explained hereinafter.

Effective Work-Function for the General Case: For the general case, first, the gate is divided into different segments, and the WF of each grain is assigned randomly based on the probabilities of different GOs [see Fig. 3(a) and (b)]. The effective WF of each grain is determined based on the WF of the grain itself and the WF of the neighboring grains [using (1)] as well as the average grain sizes and EOT [using (2)–(5)] [see Fig. 3(c)]. Subsequently, the 3-D device can be easily converted to parallel 3-D devices [see Fig. 3(d)], such that there is no WFV along the width of the parallel 3-D devices; and therefore, they can be considered as 2-D devices [see Fig. 3(e)], resulting in a significant reduction in simulation time. For example, in a 3-D device with $L_{\mathrm{ch}} = 30$ nm, $W = 30$ nm, and $GS = 10$ nm (gate is composed of nine grains), the simulation time for each grain combination is approximately 2456 s, while employing effective WFs and 2-D-simulation results in a simulation time of less than 10 s. Note that the threshold voltage of a 2-D device (with nonuniform WF of the gate metal along the channel length) can also be calculated analytically [10]. The next step is to calculate the threshold voltage of the 3-D device from threshold voltages of the parallel 2-D devices. Assuming that the threshold voltages of 2-D devices have normal distributions, the mean value and standard deviation of the threshold voltage of the 3-D device can be analytically calculated by methods, which are employed in approximation of the sum of lognormal variables [11]. As an example, Fig. 3(f) shows the top view of a hypothetical metal-gate device with three types of grains with WFs of 4.8, 4.6, and 4.4 eV. The error resulting from using the proposed model in predicting the device current in this case is less than 7%, which is acceptable when compared to the error of the network model, which can be up to 400%.

III. Summary

A new and accurate model has been proposed that captures the essential physics behind the WFV-induced threshold voltage variation in ultrashort-channel metal-gate devices. The new model considers the effect of the neighboring grains as well as the size of the grains and the effective gate-dielectric thickness. By assigning an effective WF to each grain, the effect of the adjacent metal grains is accurately considered. Moreover, by employing the proposed model, a 3-D device can be simulated using a 2-D device simulator, resulting in significant reduction in simulation time.

References

[1] H. Dadgour, V. De, and K. Banerjee, "Statistical modeling of metal-gate work-function variability in emerging device technologies and implications for circuit design," in *Proc. ICCAD*, 2008, pp. 270–277.
[2] H. Dadgour, K. Endo, V. De, and K. Banerjee, "Modeling and analysis of grain-orientation effects in emerging metal-gate devices and implications for SRAM reliability," in *IEDM Tech. Dig.*, 2008, pp. 705–708.
[3] H. Dadgour, K. Endo, V. De, and K. Banerjee, "Grain-orientation induced work function variation in nanoscale metal-gate transistors—Part I: Modeling, analysis, and experimental validation," *IEEE Trans. Electron Devices*, vol. 57, no. 10, pp. 2504–2514, Oct. 2010.
[4] T. Matsukawa, S. O'uchi, Y. Ishikawa, H. Yamauchi, Y. Liu, J. Tsukada, K. Sakamoto, and M. Masahara, "Comprehensive analysis of variability sources of FinFET characteristics," in *VLSI Symp. Tech. Dig.*, 2009, pp. 118–119.
[5] K. Ohmori, T. Matsuki, D. Ishikawa, T. Morooka, T. Aminaka, Y. Sugita, T. Chikyow, K. Shiraishi, Y. Nara, and K. Yamada, "Impact of additional factors in threshold voltage variability of metal/high-k gate stacks and its reduction by controlling crystalline structure and grain size in the metal gates," in *IEDM Tech. Dig.*, 2008, pp. 409–412.
[6] A. R. Brown, N. M. Idris, J. R. Walting, and A. Asenov, "Impact of the metal gate granularity on threshold voltage variability: A full scale three-dimensional statistical simulation study," *IEEE Electron Device Lett.*, vol. 31, no. 11, pp. 1199–1201, Nov. 2010.
[7] X. Zhang, L. Jing, M. Grubbs, M. Deal, B. Magyari-Kope, B. M. Clemens, and Y. Nishi, "Physical model of the impact of metal grain work function variability on emerging dual metal gate MOSFETs and its implication for SRAM reliability," in *IEDM Tech. Dig.*, 2009, pp. 57–60.
[8] *ATLAS User's Manual*, 2008.
[9] O. Weber, O. Faynot, F. Andrieu, C. Buj-Dufournet, F. Allain, P. Scheiblin, J. Foucher, N. Daval, D. Lafond, L. Tosti, L. Brevard, O. Rozeau, C. Fenouillet-Beranger, M. Marin, F. Boeuf, D. Delprat, K. Bourdelle, B.-Y. Nguyen, and S. Deleonibus, "High immunity to threshold voltage variability in undoped ultra-thin FDSOI MOSFETs and its physical understanding," in *IEDM Tech. Dig.*, 2008, pp. 245–248.
[10] M. J. Kumar, A. Orouji, and H. Dhakad, "New dual-material SG nanoscale MOSFET: analytical threshold-voltage model," *IEEE Trans. Electron Devices*, vol. 53, no. 4, pp. 920–923, Apr. 2006.
[11] S. H. Rasouli, H. F. Dadgour, K. Endo, H. Koike, and K. Banerjee, "Design optimization of FinFET domino logic considering the width quantization property," *IEEE Trans. Electron Devices*, vol. 57, no. 11, pp. 2934–2943, Nov. 2010.