Università di Pisa

Electrostatic effect of localised charge in dual bit memory cells with discrete traps

Luca Perniola
Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni, Università di Pisa and IMEP-ENSERG, Grenoble
Sandrine Bernardini
Provence Materials and Microelec- tronics Laboratory, Polytechnique–IMT, Marseille
Giuseppe Iannaccone
Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni, Università di Pisa
Barbara De Salvo
Commissariat à l'Energie Atomique–Laboratory of Electronics, Technology, and Instrumentation, Grenoble
Gérard Ghibaudo
Institut de Microélectronique Electromagnétisme et Photonique–Centre National de la Recherche Scientifique/Institut National Polytechnique de Grenoble
Pascal Masson
Provence Materials and Microelec- tronics Laboratory, Polytechnique–IMT, Marseille
Cosimo Gerardi
Central Research and Development, STMicroelectronics, Catania

L. Perniola,S. Bernardini, G. Iannaccone,B. De-Salvo, G. Ghibaudo, P. Masson, C. Gerardi, *Electrostatic effect of localised charge in dual bit memory cells with discrete traps*, Solid-State Device Research conference, 2004. ESSDERC 2004, vol. 34, pp. 249-252, Leuven 2004.

Electrostatic effect of localised charge in dual bit memory cells with discrete traps

L. Perniola[1, 2,*], S. Bernardini[3], G. Iannaccone[1,6], B. De Salvo [4], G. Ghibaudo [2], P. Masson[3], C. Gerardi[5]

*(1)Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Pisa, Via Caruso, 56122 Pisa, Italy, *perniola@enserg.fr (2)IMEP-CNRS/INPG, Avenue de Martyrs 32, 38016 Grenoble, France (3) L2MP-Polytech – IMT Technopôle de Château Gombert, 13451 Marseille Cedex 20 France (4)CEA-LETI, Avenue de Martyrs 16, 38054 Grenoble, France (5)STMicroelectronics, Catania, Italy (6)IEIIT-CNR, Via Caruso, 56122 Pisa, Italy.*

Abstract:

In this paper the electrostatic impact of Channel Hot Electron (CHE) injection in discrete-trap memories is quantitatively addressed. The dual bit behavior of the transfer characteristic during forward and reverse read of a written cell is thoroughly analysed with the help of an analytical model. Such model allows, for the first time, to estimate the effective charged portion of the discrete storage layer, L_2, and the quantity of electrons, Q, injected in the trapping sites from the experimental parameters of the I_d-V_g characteristics, the reverse-forward threshold voltage shift ΔV_{RF}, and the total threshold voltage shift ΔV_{tot}. The viability of this model is confirmed with tests performed on nanocrystal memories, under different bias conditions. These results are confirmed with the help of a 2D drift-diffusion commercial code (ATLAS-SILVACO).

1. Introduction

Channel Hot Electron injection is widely used as a standard writing method for non-volatile discrete-trap memory products [1]-[2]. It provides the opportunity to localise the charges injected in a small region of the trapping medium, and two-bit operation is achieved through multilevel storage [3]. The basic principle on which two-bit operation resides, is common to NROM memories and nanocrystal memories [4]-[5]. It is possible to trap charges near one junction (drain or source) with a programming stress, and read them in the *reverse* mode, compared to the programming, enhancing the electrostatic effect of these charges on the conductivity of the active channel (see Fig.1).

It has been shown in the literature [1]-[3]-[6] that the threshold voltage during the *forward* read, V_{th-F}, is lower than the threshold voltage during the reverse read, V_{th-R}, when the cell is polarised in the saturation regime.

This is due to the strong two dimensional effects near the charged junction. If the injected charge, near the drain, is completely screened by the high V_{ds} applied in forward read (which induces a long pinchoff region), the I_d-V_g characteristic results very close to the characteristic of the fresh cell. In this case we have a low V_{th-F}. On the other hand, during the reverse read the high V_{ds} applied is not able to screen the effect of electrons and the conductivity of the active channel is lowered by the "bottleneck" near the low-voltage contact. In such a case we have a high V_{th-R} [2]-[6].

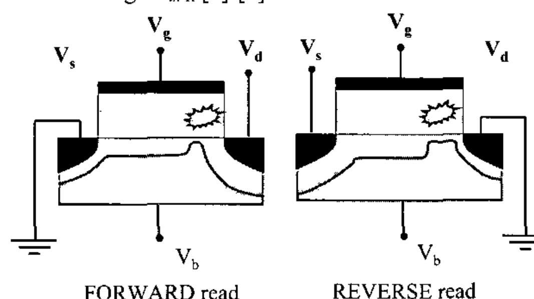

Fig.1 - The concept of forward and reverse read after Channel Hot Electron programming, considering a non zero bulk potential. In the figure the depletion width under the charged and uncharged regions is highlighted.

With a simple 1D approach (Gradual Channel Approximation) [6] it is not possible to simulate such an asymmetric behavior. In this paper, with the help of a simple quasi 2D analytical approach, we are able to quantify for the first time, both the length of the charged region, L_2, and the number of injected electrons per unit area in the trapping sites, Q, from the two quantities available from experiments, the *total threshold voltage shift*, ΔV_{th-tot}, (*i.e.*, the difference between the threshold voltage in the reverse read, V_{th-R}, and the threshold voltage of the fresh cell, $V_{th-fresh}$) and the *reverse-forward threshold voltage shift*, ΔV_{RF} (*i.e.* the difference between V_{th-R} and V_{th-F}).

The viability of the model is tested with 2D numerical simulations of a commercial TCAD code (ATLAS-SILVACO).

Main results from the theory are widely tested, under different bias conditions, on nanocrystal memories fabricated by ST-Microelectronics [4]. The influence of the bulk potential, as well as that of the drain-to-source potential, are carefully assessed in this work.

2. Quasi-2D analytical model

With a simple 1D approach it is not possible to describe properly the 2D effects near the charged region, which are the first cause of the dual bit behavior. For the first time, to our knowledge, we present an analytical approach, which starts from quasi 2D considerations [7].

0-7803-8478-4/04/$20.00 ©2004 IEEE

The analytical approach describes the behavior of the surface potential Ψ_S along the channel of a MOSFET with non-uniform threshold voltage. The distribution of trapped charge may be in general a complicate function of the spatial coordinates. We assume that the charge distribution along the channel can be described by a step function which represents an "effective" distribution. As a result, we can distinguish two regions: one close to the source with a length L_1, where storage nodes are not charged (uncharged region); and the second one close to the drain with a length L_2 which is uniformly charged by Q electrons per unit area. In Fig. 2 it can be seen how the written memory is divided in the two regions. The trapped charge raises the flat band voltage, $V_{fb\text{-}2}$, of the charged region L_2. In particular the flat band voltage difference between the charged and the uncharged region is:

$$\Delta V_{fb} = + Q \cdot q / C_2 \qquad (1)$$

Where q is the electron charge, C_2 is the top oxide capacitance per unit area.

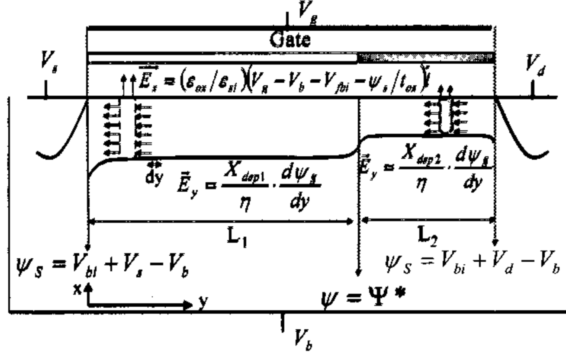

Fig. 2 - Sketch of the charged and uncharged regions. Eq.(2) and (3) come from the Gauss' law applied to the two slices of thickness dy. Boundary conditions are shown (V_{bi} is the built-in potential), considering the presence of a bulk voltage V_b. The potential reference value is the Fermi level of the bulk.

In the following model, we do not consider charges over the drain junction, but, as already noted by [6], from Fig.3 it appears that they do not influence I_d-V_g characteristics.

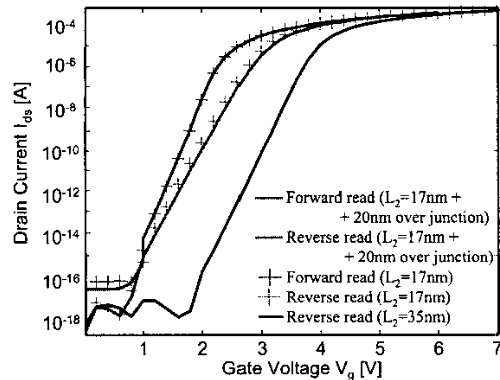

Fig.3 – ATLAS simulations to assess the influence of charge over drain junction. It is clear that a charged region over the junction of length comparable to L_2, has negligible effects on channel conductivity.

The model solves separately, in the two regions, the differential equations (2) and (3). Gauss' law is applied to a transversal slice of thickness dy along y and cross section equal to the average depletion width X_{depi} (i=1,2) (see Fig. 2), and allows us to write the flux of the electric field through the faces perpendicular to y in terms of the surface potential as $(X_{depi}/\eta) d\Psi_s/dy$, where η is a fitting parameter. During the gate voltage sweep, frequently the uncharged region goes in strong inversion, while the charged region is still pinched-off. The presence of both inversion charge Q_{inv} and depletion charge at inversion Q_{dep1}, are carefully considered in Eq.(3):

$$\varepsilon_{si} \frac{X_{dep1}}{\eta} \cdot \frac{d^2\Psi_s}{dy^2} + \varepsilon_{ox} \frac{V_g - V_b - V_{fb1} - \Psi_s}{t_{ox}} = Q_{dep1} + Q_{inv} \qquad (2)$$

$$\varepsilon_{si} \frac{X_{dep2}}{\eta} \cdot \frac{d^2\Psi_s}{dy^2} + \varepsilon_{ox} \frac{V_g - V_b - V_{fb2} - \Psi_s}{t_{ox}} = qN_{sub} X_{dep2} \qquad (3)$$

where ε_{si} (ε_{ox}) is the silicon (oxide) electric permittivity, V_g the gate voltage, t_{ox} is the equivalent oxide thickness, N_{sub} is the doping level of the bulk. If a bulk voltage, V_b, is applied, as sketched in Fig.2, the drain, source and gate potentials must be raised of V_b, setting at zero the potential of the bulk. The space charge and the inversion charge have the following expression:

$$Q_{dep} = \sqrt{2\varepsilon_{si} q N_{sub} \cdot (\Psi_{L-i} - V_b)} \qquad (4)$$

$$Q_{inv} = \frac{KT}{q} m C_{ox} \ln[1 + \frac{C_{dep}}{m C_{ox}} \exp\left(\frac{q}{mKT}(V_{gs} - V_{th-i}) - \frac{qU_{cs}}{KT}\right)] \qquad (5)$$

Where i=1,2, Ψ_{L-i} is the surface potential in the case the influence of drain and source junctions is completely neglected, U_{cs} is the quasi-Fermi potential referred to the source of the uncharged equivalent transistor.

The solution to Eq.(2) is a linear combination of hyperbolic sins:

$$\Psi_{S-1}(y) = (\Psi_{r-1} - \Psi_{L-1}) \frac{\sinh(y/\lambda_1)}{\sinh(L_1/\lambda_1)} + (\Psi_{L-1} - \Psi_{l-1}) \frac{\sinh((y - L_1)/\lambda_1)}{\sinh(L_1/\lambda_1)} \qquad (6)$$

where Ψ_{r-1} (Ψ_{l-1}) is the right-hand (left-hand) boundary condition shown in Fig.2. Ψ_{S-1} is referred to the value of the potential in the bulk. The solution to Eq.(3) is of similar form of Eq.(6).

The parameter $\lambda = \sqrt{\varepsilon_{si} t_{ox} X_{depi} / (\varepsilon_{ox} \eta)}$ describes the *short channel effects*, as it provides a clue on how strong is the influence of the junction potential on Ψ_S under the gate. The higher the value of λ, the smoother appears the Ψ_S curve along the channel, the weaker is the control effect of the gate on the channel.

From the description of the surface potential, the drain current can be calculated. As in the subthreshold region the current is a *diffusion* dominated process, the following expression can be derived [8]:

$$I_{ds} = \mu_{eff} \frac{W}{L} \sqrt{\frac{\varepsilon_{si} q N_{sub}}{2(\Psi_{S\min} + V_b)}} \left(\frac{KT}{q}\right)^2 \left(\frac{n_i}{N_{sub}}\right)^2 e^{\frac{q(\Psi_{S\min} + V_b)}{KT}} \qquad (7)$$

where μ_{eff} is the effective mobility of electrons in the channel, KT is the thermal energy, n_i is the intrinsic electron concentration in silicon and Ψ_{Smin} is the minimum of the surface potential in the channel.

3. Comparison between the analytical model and ATLAS results

From Eq.(2), the importance of η is apparent.
To determine the proper value of this parameter, we have used the commercial TCAD tool ATLAS from SILVACO. We have simulated a class of nanocrystal memories, fabricated from STMicroelectronics. Parameters of the memory cell are L=0.28 μm, W=0.16 μm, tunnelling oxide thickness $t_{tun\text{-}ox}$=5.5 nm, control oxide thickness t_2=8 nm. The charged nanocrystals have been considered as a uniform charged oxide region, of thickness t_{ch}=2 nm, sandwiched between the tunnelling oxide and the control gate oxide.

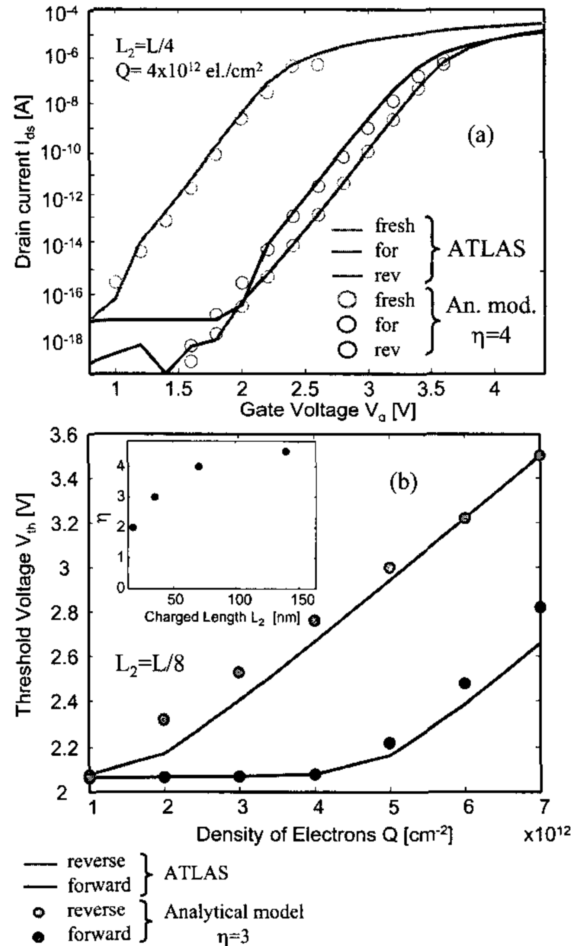

Fig. 4 (a) Comparison of the transfer characteristics from ATLAS (solid lines) and analytical model (circles) to extract the value of the fitting parameter η. In the case of L_2=$L/4$, the best value is η=4. (b) Comparison of V_{th} versus Q obtained with ATLAS (solid lines) and the analytical model (circles) shows that η is weakly dependent on the density of electrons injected in dots. In the inset of Fig. 4(b) the dependence of η on L_2 is provided.

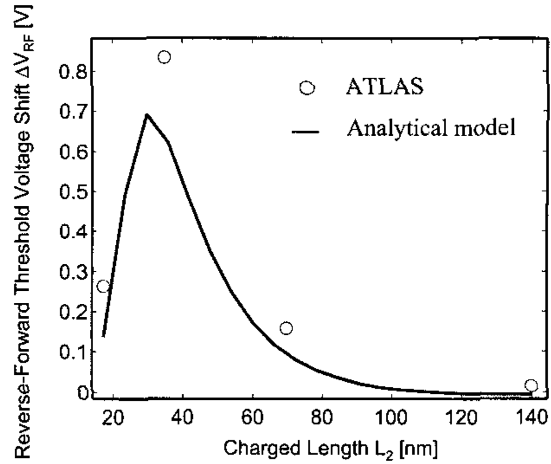

Fig.5 – Reverse-forward threshold voltage shift ΔV_{RF} as a function of the charged length L_2 obtained from ATLAS (circles) and the analytical model (solid line). Parameters are dot density 10^{12} cm^{-2}, 6 el./dot and η according to the inset of Fig. 4(b).

As specified before, we simply assumed that the charge distribution in the oxide is a step function where the charged length can be varied and the density of charge has been chosen to fit different flat band voltages V_{fb}'s.
In Fig. 4(a) a sample comparison between ATLAS results and the analytical model for L_2=$L/4$ is shown. The parameter η to fit I_d-V_g numerical characteristics has shown a weaker dependence on the density of electrons trapped in dots, Q, than on the charged length L_2. This property is shown in Fig. 4(b), where, provided one value of L_2, it is possible to fairly fit different values of Q, with one value of η. In the inset, on the other hand, the dependence of η on L_2 is provided (strong dependence of η on L_2).

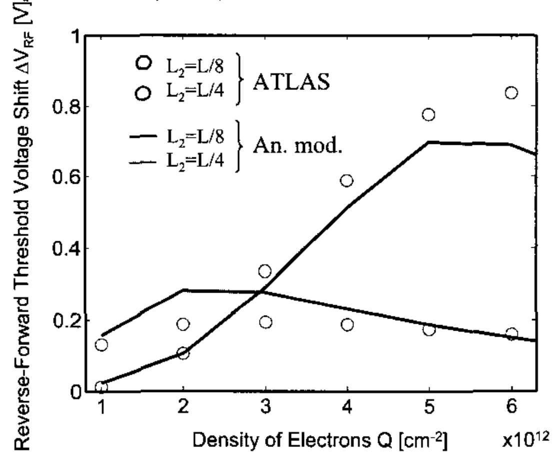

Fig.6 – Reverse-forward threshold voltage shift ΔV_{RF} as a function of the density of electrons Q obtained from ATLAS results (circles) and the analytical model (solid line). Parameters are L_2=L/8 (35 nm) η=3 and L_2=L/4 (70 nm) η=4 (according to the inset of Fig. 4(b)).

Already the plot of Fig. 4(a) shows that the forward and reverse bits are not clearly detached for L_2=$L/4$ (L_2=70 nm), and cannot be easily sensed as $V_{th\text{-}F}$ and $V_{th\text{-}R}$ are similar. From Fig. 5, it is clear that *one-bit operation* appears for L_2>70 nm ($\Delta V_{RF}\cong$0V). It is not required that

the whole region of the channel is uniformly charged to have $V_{th\text{-}F} \cong V_{th\text{-}R}$.

Another interesting property is the *charge density insensitivity* of ΔV_{RF}, once a saturation limit is achieved. In Fig. 6 such a behavior is shown for L_2=$L/8$ (L_2=35 nm) and for L_2=$L/4$ (L_2=70 nm). In the case of L_2=$L/4$, for Q>2x10^{12} el./cm^2 the window between the forward and reverse threshold voltage is almost constant; while in the case of L_2=$L/8$, this property appears for Q>5x10^{12} el./cm^2. Therefore ΔV_{RF} can be used as an *absolute* indicator of the charged region dimension.

4. Comparison between the analytical model and experiments

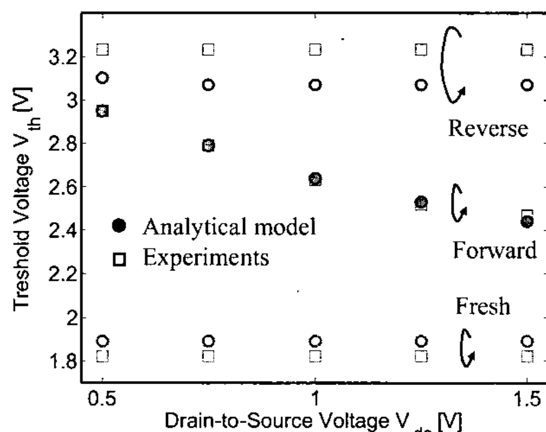

Fig. 7 – Analytical model *vs* data obtained from nanocrystal memories as a function of drain-to-source V_{ds} reading voltage.

Experiments performed on STMicroelectronics nanocrystal memory cells [4] confirm both the validity of the analytical model and suggest how to perform a reading procedure which allows to enhance the asymmetry between forward and reverse read.

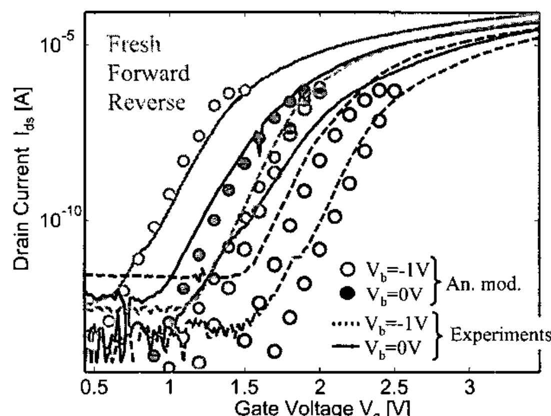

Fig. 8 – Analytical model versus experiments obtained from nanocrystal memories with V_{ds}=1.5V and different bias condition for V_b.

From Fig. 7, it is clear that raising V_{ds} enhances the electrostatic effect of trapped electrons, and decreases $V_{th\text{-}F}$ to the value of the fresh cell threshold voltage, $V_{th\text{-}fresh}$. In experiments the maximum reading V_{ds} has been put at 1.5V. In Fig. 8, the viability of this analytical model with different bulk voltages, V_b has been tested.

The most relevant conclusion of this work is presented in Fig. 9. As an example it is shown a contour plot of $\Delta V_{tot}(L_2,Q)$ and $\Delta V_{RF}(L_2,Q)$ that can be calculated from this model. From the two experimental parameters ΔV_{tot} (blu line) and ΔV_{RF} (red line) is possible to find in the contour plot the crossing point and then deduce the values of L_2 and the density of electrons Q injected in the trapping medium.

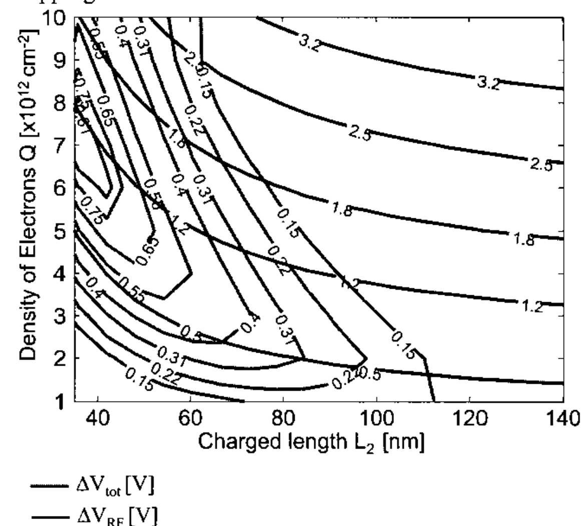

Fig. 9 – Contour plots of ΔV_{tot} and ΔV_{RF} as a function of the effective charged region L_2 and the charge density Q. Parameters are: L=0.28μm, W=0.16μm, t_{ox}=15.5nm, V_{ds}=1.5V, V_b=-1V. From the experimental data, we find the cross point between ΔV_{tot} and ΔV_{RF} values, and we deduce L_2 and Q.

5. Conclusions

A detailed model concerning Channel Hot Electron (CHE) electrostatic impact on forward/reverse reading has been provided. Both numerical simulations and experiments on nanocrystal memories, performed under different bias conditions confirm the viability of the model. At our knowledge, for the first time in literature, a contour plot is provided (Fig. 9) where, from the experimental results at hand (ΔV_{tot} and ΔV_{RF}), it is possible to assess the effective charged length, L_2 and the density of injected electrons, Q.

Support from the ADAMANT EU project (IST-2001-34234) and from the CNR through the FIRB project is gratefully acknowledged.

References

[1] B. Eitan, IEEE El. Dev. Lett., 21, 11, Nov. 2000.
[2] E. Lusky, IEEE El. Dev. Lett., 22, 11, Nov. 2001.
[3] J. Bloom, Microel. Eng., 59, 2001, 213-223.
[4]B. DeSalvo, Proc. of IEDM 2003.
[5]B. Muralidhar, Proc. of IEDM 2003.
[6] L. Larcher, IEEE Trans. El. Dev., 49, 11, Nov. 2002.
[7] Z. H. Liu, IEEE Trans. El. Dev., 40, 1, Jan. 1993.
[8] Y. Taur, "Fundamentals of Modern VLSI Devices", CUP, 1998.