Title	An analytical model for the threshold voltage of a narrow-width MOSFET
Author(s)	Cheng, YC; Lai, PT
Citation	IEEE Transactions on Electron Devices, 1984, v. 31 n. 12, p. 1814-1823
Issued Date	1984
URL	http://hdl.handle.net/10722/205917
Rights	IEEE Transactions on Electron Devices. Copyright © IEEE.

An Analytical Model for the Threshold Voltage of a Narrow-Width MOSFET

Y. C. CHENG AND P. T. LAI

Abstract—A closed form analytical expression is derived to predict the threshold voltage of a narrow-width MOSFET. The present calculation utilizes the Fourier transform technique to analyze the voltage over the width cross section of the basic MOS device structure. No fitting parameter with experimental data is necessary because the fringe electric field is calculated directly from the relevant physical parameters to deduce the threshold voltage. The dependence of threshold voltage on channel width and substrate bias thus obtained is in reasonable agreement with experimental and numerical results. The effects of field doping and field oxide thickness on the threshold voltage are also taken into consideration. A comparison is made of the present analytical expression for threshold voltage with that, based on an adjustable weighting factor, of earlier analytical models.

I. INTRODUCTION

THE present trend of VLSI technology is to place more functional blocks and increasingly complex circuits on a single chip. One of the ways to accomplish miniaturization is to shrink the basic MOS device structure. As MOSFET's are shrunk, its length, width, oxide thickness, doping concentrations, and applied voltages must be scaled or varied according to some rules to preserve the characteristics of the corresponding large device [1]. However, a perfect scaling rule cannot be found because very small devices usually show some unexpected effects such as punchthrough, hot carriers injection, dependence of threshold voltage on the channel dimension, etc. These phenomena, which give rise to a variation in device parameter as a function of the size, are called geometry effects. Such parameter variations, will have a very important effect on circuit and layout design, and therefore must be well understood and accurately predicted.

Among the device parameters that are affected by the reduction in dimension, the most important one is the threshold voltage, which depends on both the length as well as the width of the channel. The effect of short-channel length in decreasing the threshold voltage of MOSFET has been taken into consideration in designing LSI for years [2], [3]. Later, Jeppson is one of the first to predict the increase of threshold voltage as the device width shrinks down [4]. In his model, a rectangular shape is assumed for the depletion charge, with a fitting parameter accounting for the amount of side charge (see Fig. 1). This charge gives rise to the narrow width effect if the magni-
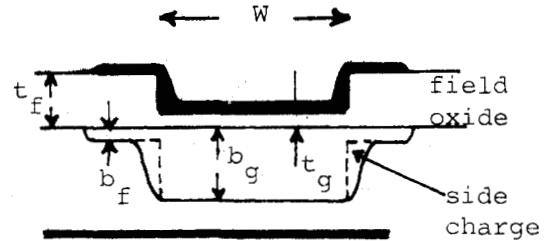
Fig. 1. The width cross section of a MOSFET showing the additional side charge.

tude of side charge becomes significant in comparison with the charge density underneath the gate electrode. Kroell and Ackermann [5] solved the two-dimensional Poisson's equation under a simple set of boundary conditions to find the threshold voltage by finite difference method. Noble and Cottrell [6] applied the finite element method to incorporate width cross-section of varying structures and doping profiles in their investigation of narrow device. Akers [7] considered different geometrical approximations to the depletion charge. Later, he and his co-workers [8] examined the effect of tapered oxide and field doping encroachment on the threshold voltage of a narrow-channel MOSFET.

At present, to account for the variation of the threshold voltage with the channel width, we have to either rely on a time-consuming numerical calculation or a simple model with an adjustable parameter which must be fitted with experimental data under different circumstances. An accurate expression describing the variation of threshold voltage with width, without the above shortcomings, does not exist. This paper aims at developing such a closed form analytical expression of the threshold voltage of a narrow-width MOSFET without invoking any adjustable parameter. The mathematical technique employed in the present investigation is similar to that used in a recent theoretical study of electrographic development [9]. The formula is derived from a Fourier transform of the potential over the width cross section focusing on the fringe electric field. The threshold voltage thus obtained is a function of the value of the gate, field oxide thickness, substrate and field doping, channel width, and backgate bias. The present model provides a simple formula for the dependence of the threshold voltage of a narrow-channel MOSFET on various physical parameters, avoiding a time-consuming numerical computation and the need for an *ad hoc* adjustable parameter. Calculated results are compared favorably with the experimental and numerical results published in [6].

II. MODEL FORMULATION

When a voltage is applied to an electrode of finite dimension, the electric field lines are mostly straight and uniform, except those near the edges. In fact, the so-called fringe field penetrates into the vicinity of the electrode. Normally, when the dimension is large such that the portion of field lines due to the fringe field is insignificant compared to all the field lines, the size effect is negligible. However, when the dimension diminishes, the fringe field becomes relatively important [10]; this gives rise to the size effect or geometrical effect in several electrostatic phenomena. A well-known example can be found in electrographic development [9], in which the resolution capability goes down as the images become finer. This is due to the fact that the fringe field due to charges in the dark area penetrates relatively more into the neighboring white area, thus losing its resolution capability of image reproduction in electrophotography.

A similar situation is encountered in the present MOS device structure. When the width of the electrode becomes really small, the fringe field becomes significant in comparison with the normal field lines underneath the gate electrode, giving rise to the effect of dependence of the threshold voltage on the width of the device. As mentioned above, a similar situation in electrophotography has been successfully analyzed, using the Fourier transform technique, by calculating the fringe field and potential of a given system. It seems natural to apply the same technique to the narrow-width device problem with, of course, modifications allowing for the specific structure and boundary conditions of the present problem.

For an MOSFET, the threshold voltage is given by

$$V_T = V_{FB} + 2\phi_B + Q_B/C_{ox} \tag{1}$$

where V_{FB} is the flat-band voltage, C_{ox} is the gate oxide capacitance, and ϕ_B is the bulk Fermi level. Q_B is the amount of depletion charge controlled by the gate electrode.

For most analytical models accounting for the geometry effects, it is always assumed implicitly that the above equation is also applicable to small device, differing only in the method adopted in the calculation of Q_B. To facilitate calculation and comparison with the experimental and numerical data, two assumptions should be made; the interface potential under the gate is uniform and the charges on the gate are distributed evenly. It can be seen later that the second assumption can be relaxed, thus eliminating the need for an adjustable parameter.

The major differences between the present and previous approaches are as follows:

1) In the previous analytical models, attention was focused on charges underneath the gate and those immediately outside the gate area (side charge in Fig. 1), whereas in the present approach we analyze the field underneath the gate as well as the fringe field, and their corresponding potentials. Both are equivalent as far as the physical picture is concerned, for all these quantities are related to each other via the Poisson's equation and appropriate differentiation.

2) Regarding the method of calculation, previous approaches either rely on a time-consuming numerical computer computa-

tion or assume the side charge to have a definite shape whose dimension is treated as an adjustable parameter. However, the present work attempts to compute the fringe field as accurately as possible, from all relevant basic physical parameters, without invoking any *ad hoc* adjustable parameter whose value is usually fitted with experimental data. In fact, the computed changes in the threshold voltage due to the finite size fringe field can be identified with the contribution due to the side charge, thus giving a physical picture of the previously reported adjustable parameter. The present model can predict the threshold voltage over a wide range of conditions, whereas the earlier analytical models will require a prior knowledge of the experimental threshold voltage in order that the value of the fitting parameter can be found. Of course, in the process of calculation approximation will be needed to simplify the computational complexity, and discussions will be made later to discuss the validity and the effectiveness of the approximation.

Fig. 1 illustrates the width cross section of a MOSFET with the depletion charge under the oxide layer. As the gate width W diminishes, the side charge becomes comparable to the ideal depletion charge in magnitude, thus causing an increase in threshold voltage as can be seen from (1). To study this threshold modulation quantitatively, a two-dimensional potential problem is solved for the width cross section by means of Fourier transform. The transformed quantity has no simple physical meaning, but offers a convenient means whereby we can solve the problem. However, the variable in the transformed space k can be regarded as the spatial frequency which has a simple physical interpretation. When k is zero it corresponds to an infinitely large gate electrode, i.e., the large geometry situation with no size effect. If k is nonzero, it corresponds to the finite-size situation with nonzero fringe field whose magnitude becomes proportionally larger as the spatial frequency increases. Appendix I shows that the transformed potential in the region concerned is given by

$$\bar{V}(k, z) = \frac{\sinh kz}{\psi(0, k)} \bar{V}_0 + \frac{\sinh kz}{k\epsilon_s \psi(0, k)} \int_0^{b_g} \psi(a, k)\bar{\rho}(a)\, da$$

$$- \frac{1}{k\epsilon_s} \int_0^z \bar{\rho}(a) \sinh k(z - a)\, da \tag{2}$$

where

$$\psi(a, k) = \epsilon_r \sinh kt_g \cosh k(b_g - a) + \cosh kt_g \sinh k(b_g - a) \tag{3}$$

and $\bar{V}_0$, $\bar{\rho}(a)$ are the transformed gate level voltage and transformed space-charge density, respectively.

The integrations in (2) are simplified through making $\bar{\rho}(a)$ a-independent, by assuming an ideal depletion charge boundary as shown in Fig. 2. b_g then becomes the ideal depletion depth. Notice that side charge can also be included as rectangular charge box in the diagram. Using the superposition of channel and field space-charge densities (ρ_g and ρ_f), we get
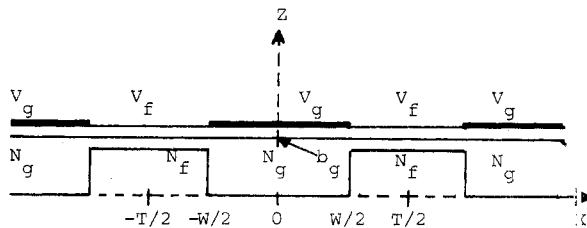
Fig. 2. Simplified structure used in the present model.

$$\bar{V}(k, z) = \frac{\sinh kz}{\psi(0, k)} \bar{V}_{0g} + \frac{\bar{\rho}_g \sinh kz}{k^2 \epsilon_s \psi(0, k)} [\epsilon_r \sinh kt_g \sinh kb_g$$

$$+ \cosh kt_g \cosh kb_g - \cosh kt_g] + \frac{\bar{\rho}_g}{k^2 \epsilon_s}$$

$$\cdot [1 - \cosh kz] + \frac{\sinh kz}{\psi(0, k)} \bar{V}_{0f} + \frac{\bar{\rho}_f \sinh kz}{k^2 \epsilon_s \psi(0, k)}$$

$$\cdot [\epsilon_r \sinh kt_g \sinh kb_f + \cosh kt_g \cosh kb_f$$

$$- \cosh kt_g] + \frac{\bar{\rho}_f}{k^2 \epsilon_s} [1 - \cosh kz] \qquad (4)$$

where V_0 is split into two components; V_{0g} in the gate region and V_{0f} in the field region.

The first term on the right-hand side of (4) is the most important term, which describes the major component of the spatial frequency dependence of the transformed voltage. When k is zero, i.e., for an infinite size, this term is at its maximum. However, when k is nonzero, i.e., for a finite-sized electrode, it decreases as k increases. The physical picture is simple: the field strength on the finite-sized electrode is smaller than that on an infinite electrode, because part of the applied voltage becomes the fringe field near the edges. Thus a larger applied voltage is needed in order to achieve the same level of inversion as in the large-geometry case, giving rise to an increase in the threshold voltage when the width of the MOS structure diminishes. Other terms in (4) involving charges, $\bar{\rho}_g$ and $\bar{\rho}_f$ are correction terms due to a combination of the space-charge effect and the finite size geometry. These are minor correction terms describing the second-order effect of the dependence of the voltage on the space charge. However, to solve for $\bar{V}(k, z)$ exactly both $\bar{\rho}_g$ and $\bar{\rho}_f$ must be known accurately; but these charges depend, in turn, on the total voltage. These unknown quantities can either be solved rigorously in a self-consistent manner or approximately by successive iterations. As we have mentioned above, these corrections are of minor effect, and to simplify calculation we have assumed that $\bar{\rho}$ is independent of the vertical distance z in obtaining (4). This also justifies the neglect of side charge as shown in Fig. 2, thus eliminating the need for an adjustable parameter.

In order to get back the physical voltage $V(x, z)$, a linear array of MOSFET's with a period T in the x-direction is used and a uniform voltage V_f is assumed between the gates as shown in Fig. 2. This assumption is made to facilitate the calculation of the correction terms, but does not affect the major fringe field component. Then $V_{0g}(x)$, $V_{0f}(x)$, $\rho_g(x)$, and $\rho_f(x)$ are all square waves and are represented by the following Fourier

series

$$V_{0g}(x) = V_g \sum_{n=0}^{\infty} A_n \cos npx; \quad \rho_g(x) = \rho_g \sum_{n=0}^{\infty} A_n \cos npx$$

$$; p = \frac{2\pi}{T}$$

$$V_{0f}(x) = V_f \sum_{n=0}^{\infty} B_n \cos npx; \quad \rho_f(x) = \rho_f \sum_{n=0}^{\infty} B_n \cos npx$$

$$(5)$$

where

$$A_0 = \frac{W}{T}; \qquad B_0 = 1 - A_0$$

$$A_n = \frac{2}{n\pi} \sin \frac{n\pi W}{T}; \qquad B_n = -A_n \quad \text{for } n \neq 0$$

and V_g is the effective gate voltage which equals applied gate voltage minus the flat-band voltage. By taking the Fourier transform of (5), substituting it into (4) and finally performing the inverse transform, the coefficient of the nth harmonic of $V(x, z)$ is found to be

$$V_n = \frac{\sinh npz}{\psi(0, np)} [(A_n + t_r B_n) V_g + (1 - t_r) B_n V_{sf}]$$

$$+ \frac{\rho_g \sinh npz}{\epsilon_s n^2 p^2 \psi(0, np)} [\epsilon_r \sinh npt_g(A_n \sinh npb_g + \rho_r B_n$$

$$\cdot \sinh npb_f) + \cosh npt_g(A_n \cosh npb_g + \rho_r B_n$$

$$\cdot \cosh npb_f) - (A_n + \rho_r B_n) \cosh npt_g] + \frac{\rho_g}{n^2 p^2 \epsilon_s}$$

$$\cdot [(A_n + \rho_r B_n) - (A_n \cosh npz + \rho_r B_n \cosh np$$

$$\cdot (z - b_g + b_f))] \qquad (6)$$

where $t_r = t_g/t_f$, $\rho_r = \rho_f/\rho_g$. Note that the interface voltage in the field region V_{sf} and field depletion width b_f can be found from Appendix II.

In the present model, the definition for threshold voltage V_T is taken as the gate voltage at which the surface voltage directly under the middle of the gate equals to $2\phi_B - V_{BS}$, i.e.,

$$V(0, b_g) = 2\phi_B - V_{BS} \qquad (7)$$

or

$$\sum_{n=0}^{\infty} V_n \Big|_{z=b_g} = 2\phi_B - V_{BS}$$

where V_{BS} is the substrate bias. A similar situation occurred in Kroell and Ackermann's numerical model in which electric field was used instead.

Strictly speaking, the averaged surface carrier concentration over the gate width, rather than the voltage at the middle, should be used. However, in view of the observation that the surface charge is neglected in this model and the potential underneath the gate width is fairly uniform [6], and to simplify the computation, we have opted for the threshold definition as given in (7). Making use of (6) with $V_g = V_T - V_{BS}$, the

threshold voltage is given by

$$V_T = \frac{1}{\displaystyle\sum_{n=0}^{\infty} \frac{\sinh npb_g}{\psi(0, np)} (A_n + t_r B_n)} \left\{ 2\phi_B - V_{BS} - \frac{\rho_g}{\epsilon_0 p^2} \sum_{n=0}^{\infty} \right.$$

$$\cdot \frac{\sinh npt_g}{n^2 \psi(0, np)} [(A_n + \rho_r B_n) \cosh npb_g - A_n - \rho_r B_n$$

$$\cdot \cosh np(b_g - b_f)] - (1 - t_r) \left(\sum_{n=0}^{\infty} \frac{\sinh npb_g}{\psi(0, np)} B_n \right)$$

$$\left. \cdot V_{sf} \right\} + V_{BS}. \tag{8}$$

The assumptions made in Fig. 2 in treating the potential V_f between the gates to be uniform and neglecting the side charge are reasonable because they do not affect the interface potential directly below the middle of the gate significantly. As we have discussed before, the important size effect is mainly accounted for by the fringe field, and the effect due to the space charge ρ (including the side charge) only gives rise to second-order corrections. It is in these correction terms that the side charge is neglected in performing the Fourier transform. Otherwise the calculation will be very complicated. Or, an iterative procedure can also be taken for example; calculate V without the side charge, then compute the induced side charge from V, and recalculate V taking the induced side charge into consideration. In principle, the iteration can be carried several times to obtain very accurate result. However, then the computation time will be very long, losing the intended advantages over detailed numerical simulation calculation. Therefore, the approximation is made as an optimum approach such that a reasonably accurate result can be obtained with a sound physical basis, but without a time-consuming numerical computation. Discussions on the validity of the approximation will be deferred to Appendix III.

Since a linear array of MOSFET's is used instead of an individual one, the interference between the MOSFET's has to be minimized. This can be easily achieved by increasing the distance between adjacent MOSFET's. Nevertheless, accompanying with this is a mounting number of terms for the summation series required for convergence, thus lengthening the computation time. Therefore, an optimal value should be chosen for the separation of device. Calculations using (8) elucidate that V_T would remain nearly independent of the spatial period T provided that the distance between devices is about 1.4 times greater than the depletion depth b_g. This means that the fringe field extends sideway to a distance of $0.7 \ b_g$ approximately, which is quite consistent with the lateral spread of space charge commonly used in the other analytical models. For the following V_T evaluations, in order to minimize the calculation time, the spatial period is taken to be

$$T = W + 1.4b_g. \tag{9}$$

We have examined the convergence of the series in (8), by analyzing the percentage error as a function of the number of harmonics taken in the summation. Again discussions will be

deferred to Appendix III. The convergence of the series is reasonably rapid and the computational time is quite acceptable.

Now two special cases are considered. The first one is when the field oxide is much thicker than that of gate region. Then V_{sf}, t_r, and b_f are very small, and (8) is reduced to a simpler form

$$V_T = \frac{1}{\displaystyle\sum_{n=0}^{\infty} \frac{\sinh npb_g}{\psi(0, np)} A_n} \left\{ 2\phi_B - V_{BS} - \frac{\rho_g}{\epsilon_0 p^2} \sum_{n=0}^{\infty} \right.$$

$$\left. \cdot \frac{\sinh npt_g}{n^2 \psi(0, np)} [\cosh npb_g - 1] A_n \right\} + V_{BS}. \tag{10}$$

The other case is for a large device in which k tends to zero in (2). Then we have

$$\bar{V}(k, z) = \frac{z}{\epsilon_r t_g + b_g} \bar{V}_0 + \frac{\bar{\rho} z}{\epsilon_s(\epsilon_r t_g + b_g)}$$

$$\cdot \int_{t_g}^{b_g} [\epsilon_r t_g + (b_g - a)] \ da - \frac{\bar{\rho}}{\epsilon_s} \int_0^z (z - a) \ da.$$

Integrating and taking inverse transform, then for $V_0 = V_T - V_{BS}$ and $V(0, b_g) = 2\phi_B - V_{BS}$

$$V_T = 2\phi_B - \frac{\rho_g b_g t_g}{\epsilon_0} \tag{11}$$

which is the expression commonly used for the threshold voltage of a large device. This result can also be derived by putting

$$t_r = \rho_r = 1$$

and

$$b_g = b_f$$

into (8), which is equivalent to simply reducing the two-dimensional problem to an one-dimensional one.

III. THEORETICAL CURVES

Since V_T is obtained by evaluating a simple expression in the present model, it is easy to investigate its dependence on the parameters of a device. This approach should be contrasted to numerical models which require excessive computation time. (The convergence of the three summation series used in (8) is analyzed in Appendix III.)

A few curves are shown to illustrate the threshold modulation. Default values for substrate doping N_g and gate oxide thickness t_g are 1E15 cm^{-3} and 500 Å, respectively. Fig. 3 shows the increase in V_T as the channel width decreases. This is aggravated if the backgate bias takes a larger value. The influences of field doping N_f and field oxide thickness t_f on this narrow-width effect are also shown in Figs. 4 and 5, respectively; higher field doping and/or thicker field oxide will enhance the V_T modulation. The reason is that the ideal depletion depth in the field region thus resulted is smaller, giving rise to a larger amount of side charge in the transition region from the ideal gate depletion depth to the ideal field depletion depth.
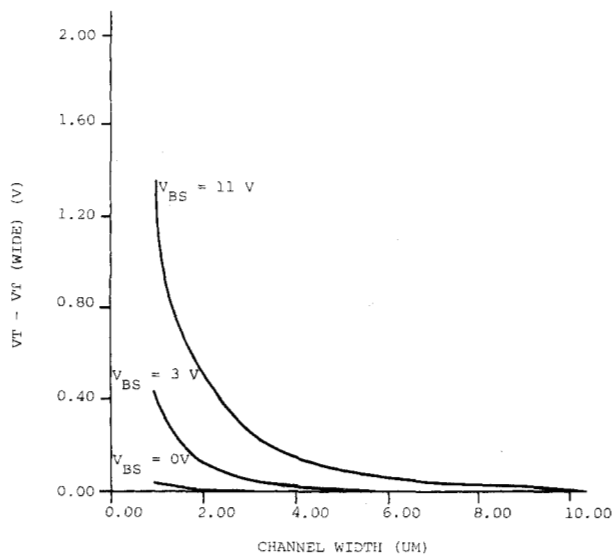
Fig. 3. Threshold voltage change versus width for various back-biases. $t_f = 85\ E - 6$ cm, $N_f = 1\ E\ 15$ cm^{-3}.
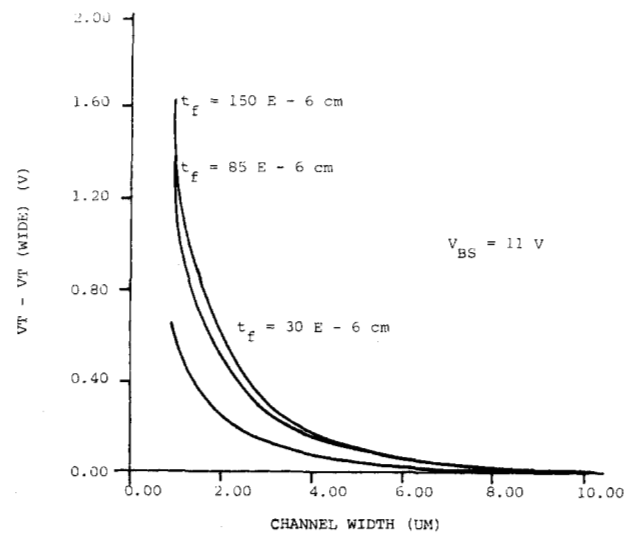

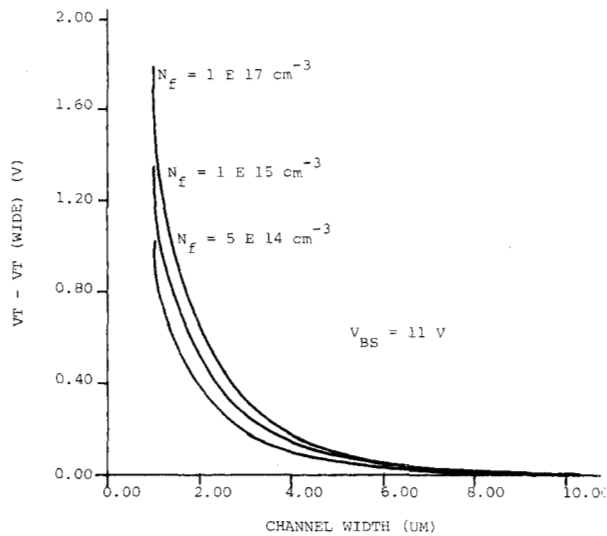Fig. 5. Threshold voltage change versus width for various field oxide thicknesses. $N_f = 1\ E\ 15$ cm^{-3}, $V_{BS} = 11$ V.

Fig. 4. Threshold voltage change versus width for various field dopings. $t_f = 85\ E - 6$ cm, $V_{BS} = 11$ V.

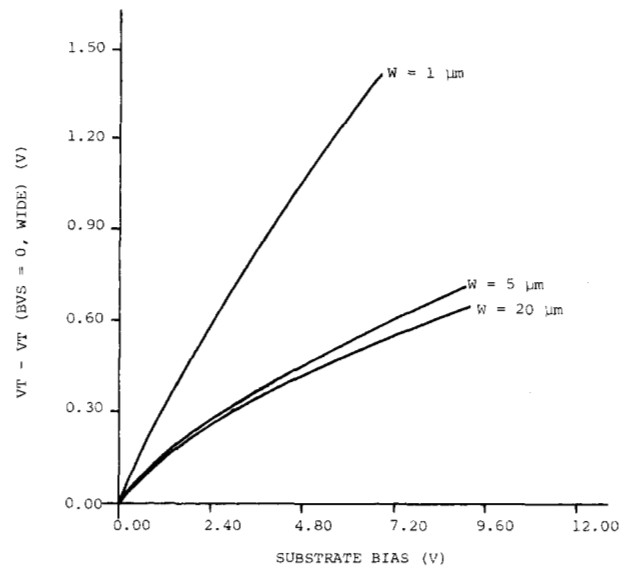Fig. 6. Threshold voltage change versus back-bias for various channel widths. $t_f = 85\ E - 6$ cm, $N_f = 1\ E\ 15$ cm^{-3}.

Fig. 6 depicts the substrate sensitivity of V_T of MOSFET. Narrower channel devices exhibit a larger increase in V_T with the backgate bias. The influences of field doping and field oxide thickness are demonstrated in Figs. 7 and 8, with large V_T-substrate bias sensitivity for higher field doping and/or thicker field oxide thickness. The explanation in the previous paragraph can also be applied here.

IV. COMPARISON WITH EXPERIMENTAL RESULTS AND EARLIER ANALYTICAL MODELS

Different definition for V_T is used for various models and experiments. Therefore, the values for V_T may be different from one another. In order to have a more meaningful comparison for the present model, only the change in V_T will be considered.

The results published in [6] are used to verify the analytical

model here. The MOSFET's used are assumed to have long channel, with parameters $N_g = 1\ E\ 15$ cm^{-3}, $t_g = 500$ Å, and $x_j = 0.3\ \mu$m unless otherwise stated.

Fig. 9 illustrates the effect of channel width on the threshold increase when V_{BS} equals to 3 and 11 V. Both numerical and experimental data are used for comparison. It can be seen that a good agreement is obtained among the three sets of data down to 2 μm range.

Fig. 10 shows the substrate sensitivity of V_T modulation for three channel widths and also illustrates the reasonable agreement between this model and the numerical model. The discrepancy appears to be larger at higher backgate bias and this may contribute to the larger amount of side charge which is neglected in the integrations of the correction terms in (2).

Finally, the analytical expression for the threshold voltage is compared with simple expression obtained in the earlier
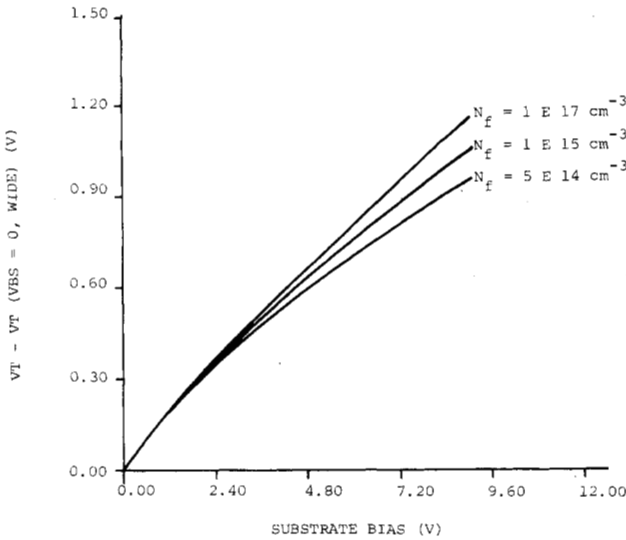
Fig. 7. Threshold voltage change versus back-bias for various field dopings. $t_f = 85\,E - 6$ cm, $W = 2\,E - 4$ cm.
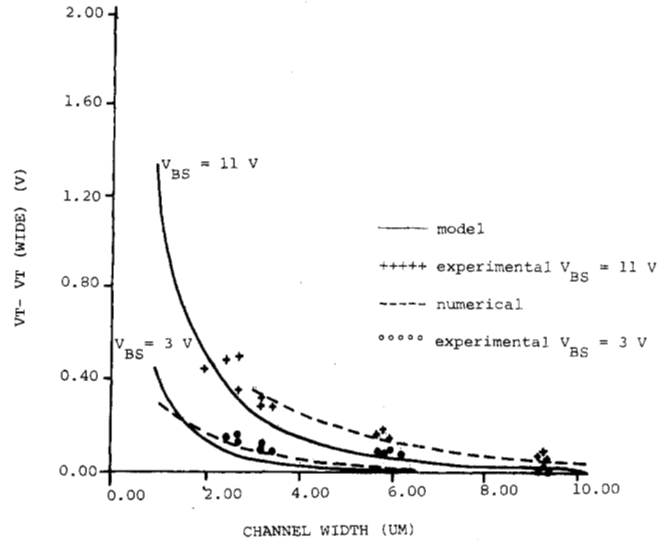

Fig. 9. Comparison of calculated, numerical, and experimental threshold voltage changes versus channel width for two back-biases. $t_f = 85\,E - 6$ cm, $N_f = 1\,E\,15$ cm^{-3}.
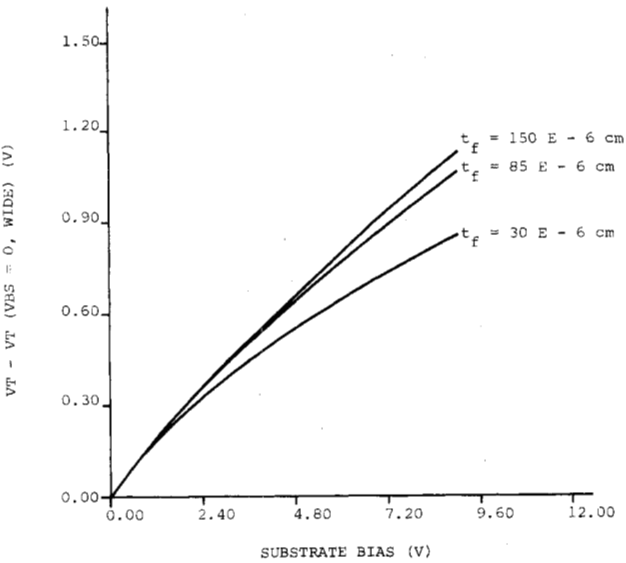

Fig. 8. Threshold voltage change versus back-bias for various field oxide thicknesses. $N_f = 1\,E\,15$ cm^{-3}, $W = 2\,E - 4$ cm.

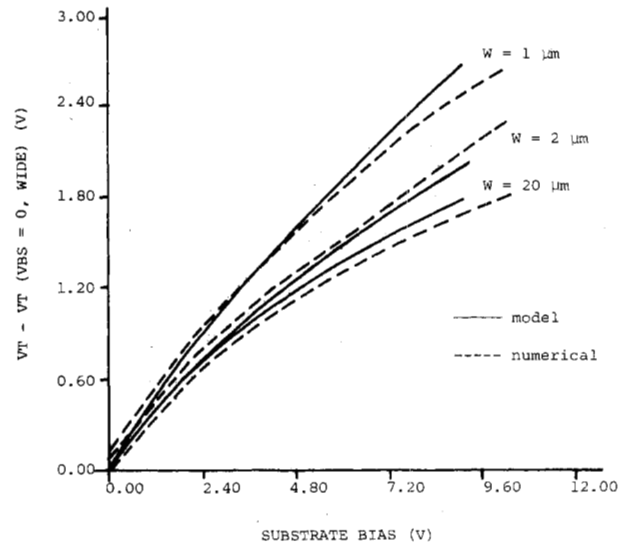Fig. 10. Comparison of calculated and numerical threshold voltage changes versus back-bias for three channel widths. $N_g = 8\,E\,15$ cm^{-3}, $t_f = 85\,E - 6$ cm, $N_f = 8\,E\,15$ cm^{-3}.

analytical models. This will shed light on the physical origin of the required adjustable parameter and the validity in treating these parametric values as constant.

After some algebraic manipulation, (8) can be transformed to a form similar to that used in other analytical models

$$V_T = V_T \text{ (large device)} - \delta b_g^2 \rho_g / C_{ox} \qquad (12)$$

where the weighting factor of the side charge is found to be

$$\delta = \left[\left(\frac{1 - D}{D} \right) \frac{1}{2\epsilon_r} + \left(\frac{N}{D} \right) \frac{1}{p^2 b_g^2} - \frac{t_g}{b_g} \right] \frac{W}{t_g} \qquad (13)$$

with

$$D = \sum_{n=0}^{\infty} \frac{\sinh np b_g}{\psi(0, np)} (A_n + t_r B_n) \qquad (14)$$

$$N = \sum_{n=0}^{\infty} \left\{ \frac{\sinh np t_g}{n^2 \, \psi(0, np)} \left[(A_n + \rho_r B_n) \cosh np b_g - A_n \right.\right.$$
$$\left. - \rho_r B_n \cosh np(b_g - b_f) \right] + (1 - t_r) \frac{\epsilon_0 p^2 \, V_{sf}}{\rho_g} \frac{\sinh np b_g}{\psi(0, np)}$$
$$\left. \cdot B_n \right\}. \qquad (15)$$

The dependence of δ on the substrate bias and the channel width is shown in Figs. 11 and 12, respectively. Here we have demonstrated that the so-called weighting factor can be calculated once all the basic physical parameters, e.g., doping concentration, substrate bias, oxide thickness, etc., are known. There is therefore no necessity to treat the weighting factor as an *ad hoc* adjustable parameter. Furthermore, from Figs. 11 and 12 we see that in general, the validity of treating the
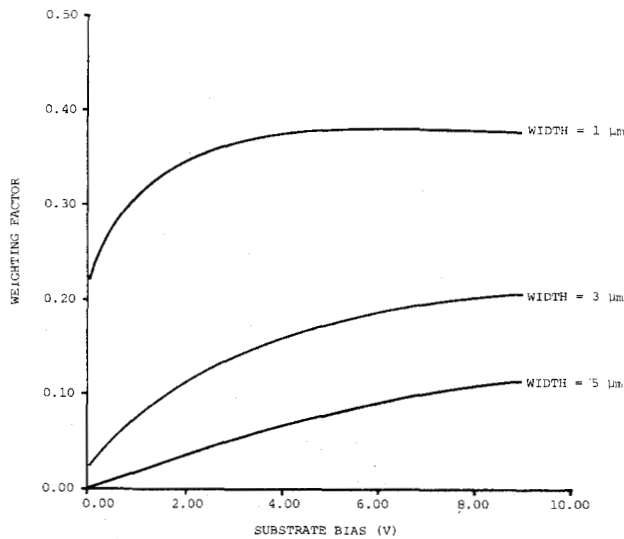
Fig. 11. Weighting factor calculated from (13) as a function of the substrate bias with the channel width taken as 1, 3, and 5 μm, respectively.
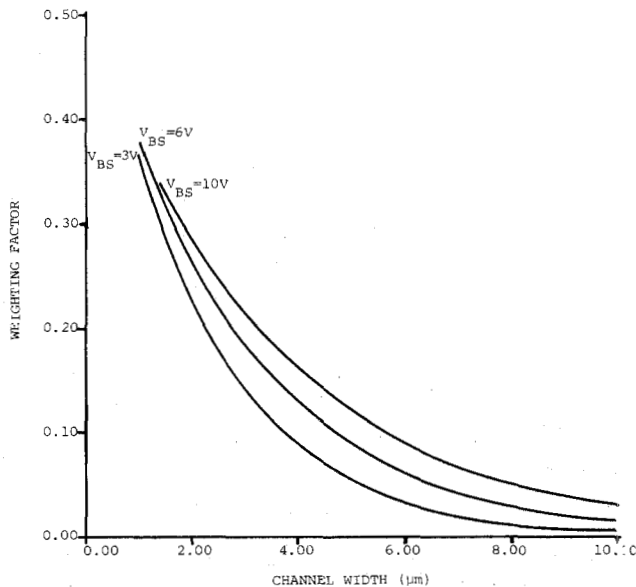
Fig. 12. Weighting factor calculated from (13) as a function of the channel width. The substrate bias is fixed at three sets of values

weighting factor as a constant is questionable. At different channel widths, the weighting factor varies slowly with the substrate bias. With a given substrate bias, the weighting factor decreases as the channel width increases; the dependence on the width is rather weak if the channel is wider than about 5 μm. These observations may explain, why in some particular cases, some earlier analytical models do seem to agree reasonably well with experimental results.

V. DISCUSSIONS AND CONCLUSION

A simple closed form analytic expression for predicting the threshold voltage of a narrow-width MOSFET is derived by analyzing the potential over its width cross section with the Fourier transform technique. Since the present approach focuses on the calculation of the fringe electric field, with no concern about the shape of the depletion charge, no fitting

parameter is required as in some existing analytical models. The present model takes into account the channel width, gate oxide thickness, substrate doping, substrate bias, field oxide thickness, and field doping. Curves are obtained based on a simple MOS structure to demonstrate the dependence of V_T on these parameters. Calculations based on a more complex or realistic structure are in progress and will be published elsewhere. For example, we have analysed the dependence of V_T on the detailed shape of the doping profile of a narrow-channel device.

A comparison of the present analytical expression for threshold voltage with that of simple expressions already published has also been made. This provides an expression of the usual weighting factor (an adjustable parameter) in earlier analytical models in terms of the basic physical quantities, such as oxide thickness, doping density, etc. The comparison sheds light on the physical origin of the fitting parameter, and gives an evaluation of the usual assumption of regarding the parameter as a constant.

Major assumptions involved in the present model are: 1) calculating the threshold voltage by identifying the potential value at the middle of the gate electrode rather than taking the averaged value over the width dimension, 2) neglecting side charge in the calculation of the correction terms of fringe electric field. While the latter assumption hardly introduces any significant error (Appendix III), the former assumption is likely to be responsible for any disagreement with the numerical results. This is especially true if the channel is extremely narrow when the potential under the gate electrode could be quite nonuniform. In that case, the averaged value of the potential should be taken, and this presents no major difficulty, but the computation time will be much longer.

In conclusion, we have formulated a simple model to compute analytically the threshold voltage of a narrow-channel MOSFET. The computer time required is negligible when compared with the usual numerical calculation, and therefore, the expression can easily be incorporated into any simulation program which requires a value for the threshold voltage. The value predicted for the threshold voltage is in reasonable agreement with experimental and numerical results. The accuracy can be improved by an averaging and/or iterative calculation, but the computation time will be longer. This presents a trade-off between accuracy and computer time, and the present model offers this flexibility.

Another advantage of the present approach is that it requires no *ad hoc* fitting parameter, and existing experimental data for different situations are, therefore, not required. Once the values for different physical quantities are given, the threshold voltage can be predicted over a wide range of operating conditions. This approach could be useful in simulation studies or as a guide prior to experimental work, but without the usual time-consuming numerical computation.

APPENDIX I

FOURIER ANALYSIS OF A TWO-DIMENSIONAL POTENTIAL PROBLEM

Fig. 13 shows a layer of oxide lying on top of a silicon substrate together with a thin layer of charge $\sigma(a)$. Boundary con-
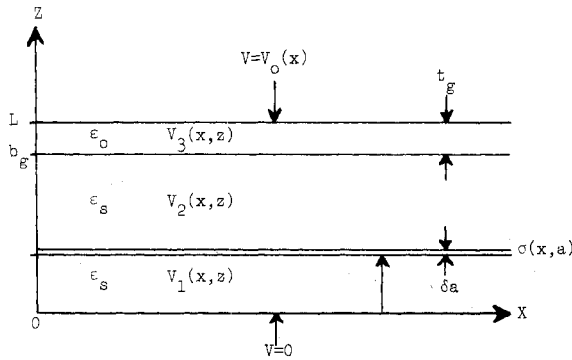
Fig. 13. An infinite strip used in the Fourier transform of a two-dimensional potential problem.

ditions give

$$V_1(0) = 0 \tag{A1}$$

$$V_3(L) = V_0 \tag{A2}$$

where $\sigma(a)$ and V_0 are x-dependent.

Continuity requires that

$$V_1(a) = V_2(a) \tag{A3}$$

$$V_2(b_g) = V_3(b_g). \tag{A4}$$

Conditions due to Gauss' law are

$$\epsilon_0 E_3(b_g) - \epsilon_s E_2(b_g) = 0 \tag{A5}$$

$$\epsilon_s E_2(a) - \epsilon_s E_1(a) = \sigma(a). \tag{A6}$$

Since there is no space charge in the three regions, the three potentials should all satisfy the Laplace's equation.

$$\frac{\partial^2}{\partial z^2} V_j + \frac{\partial^2}{\partial x^2} V_j = 0, \quad j = 1, 2, 3.$$

Taking the Fourier transform with respect to x gives

$$\left(\frac{\partial}{\partial z^2} - k^2\right) \bar{V}_j = 0; \quad \bar{V}_j = \int_{-\infty}^{\infty} V_j e^{-ikx} \, dx$$

whose solution is

$$\bar{V}_j = A_j e^{kz} + B_j e^{-kz}. \tag{A7}$$

Equation (A1) to (A6) all remain the same except that the variables are replaced by their transformed values. All the A's and B's can be found by making use of conditions (A1) to (A6) in (A7). The transformed potentials in the silicon region becomes

$$\bar{V}_1 = 2A_1 \sinh kz$$

$$\bar{V}_2 = 2A_1 \sinh kz - \frac{\bar{\sigma}(a)}{k\epsilon_s} \sinh k(z - a) \tag{A8}$$

with

$$2A_1 = \frac{\bar{V}_0 + \dfrac{\bar{\sigma}(a)}{k\epsilon_s} \psi(a, k)}{\psi(0, k)}$$

where

$$\psi(a, k) = \epsilon_r \sinh kt_g \cosh k(b_g - a) + \cosh kt_g \sinh k(b_g - a)$$

and

$$\epsilon_r = \epsilon_s/\epsilon_0.$$

Supposing there is a space-charge density $\rho(a)$ in the silicon region, $\sigma(a)$ is then equal to $\rho(a)\delta a$. With V_0 replaced by $V_0 \cdot (\delta a/b_g)$ and making use of superposition principle, the transformed potential is given by

$$\bar{V}(z) = \int_0^z \bar{V}_2 + \int_z^{b_g} \bar{V}_1$$

$$= \frac{\sinh kz}{\psi(0, k)} \bar{V}_0 + \frac{\sinh kz}{k\epsilon_s \psi(0, k)} \int_0^{b_g} \psi(a, k)\bar{\rho}(a) \, da$$

$$- \frac{1}{k\epsilon_s} \int_0^z \bar{\rho}(a) \sinh k(z - a) \, da \tag{A9}$$

through the substitution of (A8).

APPENDIX II

CALCULATION OF DEPLETION DEPTH b_f UNDER THE FIELD OXIDE

The charge per unit area in the area in the depletion region under the gate oxide is

$$\frac{\epsilon_0}{t_g} (V_g - V_{FB} - 2\phi_B + V_{BS}) = qN_g b_g \tag{B1}$$

while for that in the field region is

$$\frac{\epsilon_0}{t_f} (V_g - V_{FB} - V_{sf}) = qN_f b_f \tag{B2}$$

where V_{FB} is the flat-band voltage and V_{sf} is the interface potential in the field region, which can easily be found to have a value of $qN_f b_f^2/2\epsilon_s$. Insertion of (B1) into (B2) results in

$$b_f = -\epsilon_r t_f + \sqrt{(\epsilon_r t_f)^2 + \frac{2\epsilon_s(2\phi_B - V_{BS})}{qN_f} + \frac{2\epsilon_r N_g t_g b_g}{N_f}}. \tag{B3}$$

APPENDIX III

DISCUSSIONS ON THE CONVERGENCE OF THE SERIES AND THE APPROXIMATIONS TAKEN

In this section we discuss the convergence of infinite sums involved in (8), and also discuss the validity of neglecting side charge in the calculation of the correction terms.

In (8), there are three infinite sums involved in calculating the threshold voltage. We have examined separately the number of terms required in each summation in order to achieve a certain acceptable error. Fig. 14 illustrates two typical sets of results for the three sums. Even the worst sum converges reasonably rapidly as far as computation time is concerned. In our results, we require the relative error in each sum to be less than or equal to 5×10^{-4} and the time required to compute one V_T is reasonably short, with an average CPU time of 0.2 s in Univac 1100.
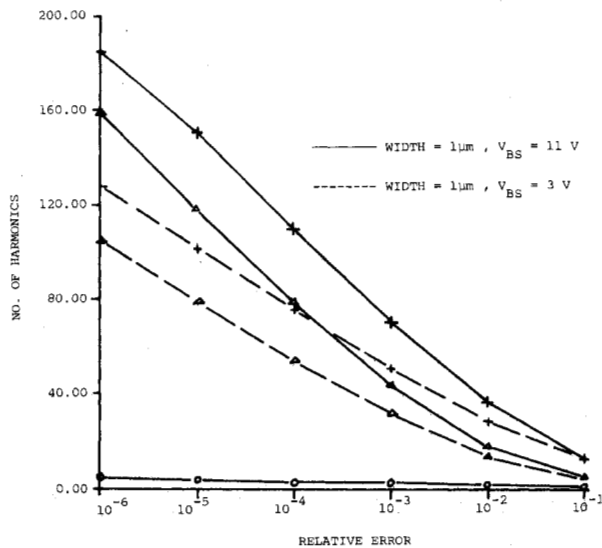
Fig. 14. Number of terms required to achieve a certain level of relative error in the infinite sum involved in calculating V_T. The three curves represent the sums in (8).
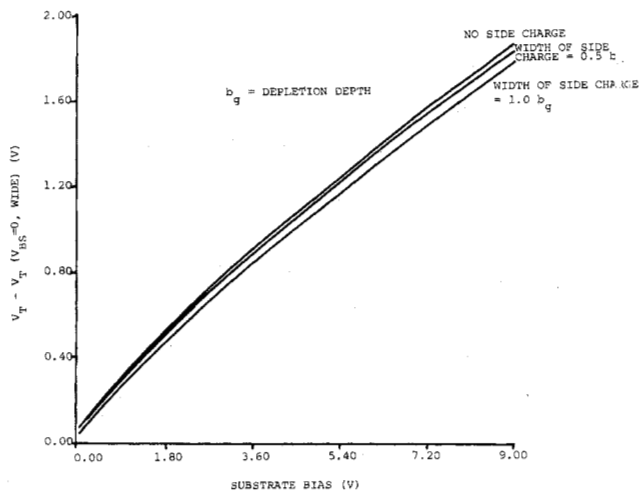

Fig. 15. Dependence of V_T on the amount of side charge. The three curves are for: 1) no side charge, 2) with side charge extending sideway to a distance equal to half a depletion depth, and 3) with side charge extending sideway to a distance equal to one depletion depth.

The approximation of neglecting side charge is discussed here. As we mentioned before, the change in V_T is attributed to the fringe field in the present work, rather than focusing on side charge as in earlier analytical model. The side charge appears explicitly in the correction terms in the present model. To show the effect is small, we have calculated V_T, assuming the existence of side charge up to about one depletion depth in the width direction beyond the electrode edge. Fig. 15 shows a summary of the calculation. The threshold voltage computed with one depletion depth of side charge differs only marginally from that based on the calculation with no side charge. In earlier section, we mentioned that the fringe field extends sideway to a distance on the order of a fraction of the depletion depth. In Fig. 15, the threshold voltage results, based on the existence of side charge extending to half a depletion width, almost coincide with those with no side charge. Therefore, we can confidently conclude that the neglect of side

charge in our calculation of correction terms in V_T will not introduce any significant error.

After all, the side charge is only the effect of the fringe field which is the actual cause of the narrow-width effect. The feedback of the side charge on the potential distribution should be small, especially near the middle of the gate provided the amount of side charge is not too large as compared to the charge directly under the gate. The steeper slopes of our curves in Figs. 9 and 10 for narrower gate width and higher backbias testify that the side charge can no longer be neglected in this operating region where it is of a considerable quantity. Fig. 15 also shows the overestimate of threshold voltage if no side charge is included.

Next, the assumption of uniform gate-level voltage V_f in the field region is discussed. From numerical simulation, it is found that the gate-level voltage falls off very quickly to a steady value at both sides of the gate, especially when the side walls are fairly vertical. Thus the approximation of abrupt change in gate-level voltage will only introduce larger higher harmonics, but have little effect on the lower harmonics. Since the threshold voltage is calculated by summing the lower harmonics mainly, it is not much affected. In fact this smoother transition of gate-level voltage can be approximated by slightly decreasing the field oxide thickness or the field doping (This will increase V_f slightly). The resulting effect is shown in Figs. 4 and 5 and is only a small decrease in threshold voltage. This further accounts for the larger threshold voltage calculated for regions where narrow-width effect is prominent.
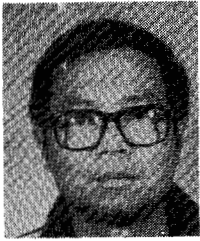
REFERENCES

[1] R. H. Dennerd, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassons, and A. LeBlanc, "Design of ion-implanted MOSFET with very small physical dimensions," *IEEE J. Solid-State Circuits*, vol. SC-9, pp. 256–268, 1974.
[2] H. S. Lee, "An analysis of the threshold voltage for short-channel IGFET's," *Solid-State Electron.*, vol. 16, pp. 1407–1417, 1973.
[3] L. D. Yau, "A simple theory to predict the threshold voltage of short-channel IGFET's," *Solid-State Electron.*, vol. 17, pp. 1059–1063, 1974.
[4] K. O. Jeppson, "Influence of the channel width on the threshold voltage modulation in MOSFET," *Electron. Lett.*, vol. 11, pp. 297–298, 1975.
[5] K. E. Kroell and G. K. Ackermann, "Threshold voltage of narrow channel field effect transistors," *Solid-State Electron.*, vol. 19, pp. 77–81, 1976.
[6] W. P. Noble and P. E. Cottrell, "Narrow-channel effects in insulated gate field-effect transistors," in *IEDM Tech. Dig.*, pp. 582–586, 1976.
[7] L. A. Akers, "Threshold voltage of a narrow-width MOSFET," *Electron. Lett.*, vol. 17, pp. 49–50, 1981.
[8] L. A. Akers, M.M.E. Beguwala, and F. Z. Custode, "A model of a narrow-width MOSFET including tapered oxide and doping encroachment," *IEEE Trans. Electron Devices*, vol. ED-28, pp. 1490–1495, 1981.
[9] Y. C. Cheng and G. C. Hartmann, "Electrographic development: An electrostatic calculation," *J. Appl. Phys.*, vol. 51, no. 5, pp. 2332–2337, May, 1980.
[10] D. N. Pattanayak, J. G. Poksheva, R. W. Downing, and L. A. Akers, "Fringing field effect in MOS devices," *IEEE Trans. Components, Hybrids, and Manufact. Technol.*, vol. CHMT-5, pp. 127–131, 1982.

*

Y. C. Cheng received the B.Sc. degree in physics and mathematics from the University of Hong Kong in 1963 and the Ph.D. degree in theoretical physics from the University of British Columbia in 1967.

From 1963 to 1978, he held various appointments at a number of universities in Canada, and at Bell-Northern and Xerox Research Lab-

oratories. During this period, he had undertaken research work on silicon devices, and had co-invented and developed the "HCl-Oxidation" technique for the production of clean oxides for integrated circuit application. In 1977 he was appointed Adjunct Professor at the University of Waterloo. He is currently a Professor in Department of Electrical Engineering and Sub-Dean of the Engineering Faculty, University of Hong Kong.
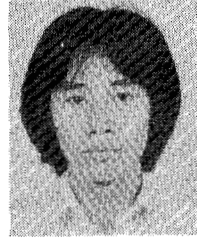
P. T. Lai was born in Hong Kong on December 19, 1958. He received the B.Sc. degree in electrical engineering from University of Hong Kong in 1981, and is now working toward the Ph.D. degree in the field of modeling of small MOS devices.

Optimum p-Channel Isolation Structure for CMOS

MICHAEL SUGINO, MEMBER, IEEE, LEX A. AKERS, SENIOR MEMBER, IEEE, AND
JENNY M. FORD, MEMBER, IEEE

Abstract—A two-dimensional numerical model of the width direction of a MOSFET is used to simulate the surface potential and the subthreshold current of p-channel devices. Fully-recessed, semi-recessed, and nonrecessed oxide isolation structures with various transition angles as well as interface charge are modeled. The nonrecessed oxide structure is superior for reducing subthreshold current, in some cases more than 20 percent. The fully-recessed oxide with a 90° transition angle provides maximum device density, a planar surface, and ease of fabrication. Experimental results indicate that for the fully-recessed oxide structure the p-channel device with interface charge will show a threshold-voltage variation of only 12 percent with widths varying from 10 to 1.5 μm, and an increase in subthreshold current of an order of magnitude compared to a wide device.

I. INTRODUCTION

WITH the feature size of VLSI devices decreasing to submicrometer dimensions, the effects of small geometries on device characteristics and performance becomes increasingly important. Such behavior as short-channel effects [1], narrow [2], [3] and inverse narrow width (also called channel or gate) effects [4]–[7], hot electron and hole generation with its subsequent substrate [8], [9] and gate current [10], and isolation oxide effects [4] must be minimized. Additional properties are also required of devices incorporated into semiconductor chips with VLSI densities.

As MOS integration densities increase, chip power dissipation approaches the thermal limits of packages. This problem can be reduced by designing with complementary structures. The related push for high-density circuits requires not only devices with short channels and narrow widths, but with small active area pitch. This requires bird's beak free non-enchroaching isolation oxides. As the width is reduced, the

current drive is reduced. Since the parasitic capacitance may not scale, and the interconnect resistance increases as devices are scaled, increasing current drive is imperative. A way to increase current drive is to reduce the threshold voltage, but since supply voltages are being scaled to reduce hot-carrier problems, it is questionable as to whether any net effect is to be expected. This emphasizes the need for maintaining mobilities, or at least a slowing down of the reduction of mobility as devices shrink [11]. With lower threshold voltages, their variation with geometries must be reduced to obtain acceptable worst case noise margins and yields. This can be accomplished by fabricating structures that tend to reduce threshold-voltage variations with device geometry. Lastly, small subthreshold currents and swing [12] are important for reduced standby power dissipation and to provide fast switching. Not all of these characteristics can be obtained simultaneously, and therefore compromises are necessary. It is proposed that at present in silicon a small geometry CMOS structure with a fully-recessed planar isolation oxide best fits these desired characteristics.

The effects of short channels on the threshold voltage has been reviewed by Akers and Sanchez [13] and by Kumar [14]. Subthreshold current in short structures was reviewed by Fitchner and Potzl [15]. Gate and substrate currents have been modeled by Tam et al. [8]. The variation of the threshold voltage in the width direction has been reviewed by Akers and Sanchez [13] and by Ji and Sah [7] for structures that produce the narrow-width effect. Only a limited amount of work has been published on the inverse narrow-width effect. Shigyo et al. [5], [6] developed a three-dimensional computer model and simulated the inverse narrow-width effect. They illustrated the effects of the transition angle of the isolation oxide on the threshold voltage in NMOS devices and showed how field implantation effects its behavior. Sugino and Akers [4] developed a two-dimensional computer simulation model of the width direction and simulated the surface potential and