Solid-State Electronics Vol. 25, No. 2, pp. 85–90, 1982
Printed in Great Britain.

0038-1101/82/020085-06$03.00/0
© 1982 Pergamon Press Ltd.

INVESTIGATION OF PARAMETER SENSITIVITY OF SHORT CHANNEL MOSFETS

S. SELBERHERR, A. SCHÜTZ and H. PÖTZL
Institut für Allgemeine Elektrotechnik und Elektronik, Abteilung für Physikalische Elektronik, TU Wien, Gußhausstraße 27, A-1040 Wien, Austria, and Ludwig Boltzmann Institut für Festkörperphysik, Germany

(*Received* 28 *November* 1980; *in revised form* 17 *June* 1981)

Abstract—A strategy to examine the sensitivity of electrical device parameters on geometrical and technological tolerances is described. An approach is offered to determine the limit of device miniaturization for a given fabrication process and a desired operating condition. As a didactic example of practical relevance the minimum channel length for a modern silicon gate, double implant process due to threshold uncertainty is estimated. A method to calculate global sensitivity numbers for the reproducability of miniaturized devices is suggested. As an experimental determination of sensitivities is extremely difficult and expensive, numerical simulations are ideally suited for this purpose.

NOTATION

$AKEV$	implantation energy
C_{ox}	oxide capacity
Dose	implantation dose
L	channel length
N_b	bulk doping
UB	bulk bias with respect to the source
UD	drain bias with respect to the source
UT	threshold voltage with respect to the source
Rj	source/drain junction depth
Q_b	bulk charge
Q_{fs}	charge of fixed surface states
T	temperature
T_{ox}	thickness of gate oxide
W	channel width
y_c	length of depletion region below channel
ϵ_{ox}	permittivity of silicon dioxide
ϵ_{si}	permittivity of silicon
Φ_{MS}	metal-semiconductor work function difference
Φ_F	Fermi voltage

1. INTRODUCTION

VLSI is evidently connected to the miniaturization of the single transistor. Merely shrinking the physical device dimensions usually poses serious problems concerning device behaviour. Instead, all device parameters have to be scaled[1, 2] together with the device geometry according to certain rules. In general, lower voltages, heavier doping, shallower junctions and thinner oxides help to maintain applicable device characteristics as channel length is reduced. Down to about two microns channel length the device behaviour can be controlled excellently by the relevant technological steps (implantation, diffusion, oxidation, photolithography). However, as often observed in experimental investigations, this controlability is no longer ensured for devices with further reduced channel length. Reproducability tends to become worse with decreasing size, posing increasingly severe problems of tracking of the parameters of adjacent transistors, which should behave identically for certain kinds of circuits (e.g. latches).

To verify the increased process sensitivity of scaled devices, we performed an analysis of certain device parameters†) with MINIMOS, our two-dimensional MOS simulation program[3]. In this paper the sensitivity of the threshold voltage, which is usually the most important device property‡) for the designer, will be outlined for a well established short channel MOS process to determine the practical limit of miniaturization for a given technology. However, the analysis of threshold sensitivity is just an example for a strategy which is applicable to examine the sensitivity of any device property.

†PARAMETER = a variable which one can choose arbitrarily, e.g. L. W. T_{ox}.

‡PROPERTY = a physical attribute which is influenced by choice of parameters, e.g. UT, breakdown voltage, transconductance.

2. DEVICE STRUCTURE AND FABRICATION PROCESS

An n-channel silicon gate process with arsenic source/drain doping and a double channel implantation for threshold tailoring and punch-through suppression has been chosen.

Figure 1 shows the doping distribution logarithmically drawn in a quasi-three-dimensional plot for a one micron transistor. The channel implantation is performed with boron as the dopant, a dose of $3 \cdot 10^{11}\,cm^{-2}$ and an energy of 35 keV for the shallow layer and a dose of $10^{11}\,cm^{-2}$ and an energy of 160 keV for the deeper layer, respectively.

Figure 2 shows details of the doping profile. A junction depth of 320 nm and a lateral diffusion of about 200 nm is obtained by this process. The extremely steep gradient at the junctions is typical for arsenic. The oxide thickness—the oxide is not drawn in these figures—is about 50 nm for these devices. The whole process was designed for two micron lateral dimension.

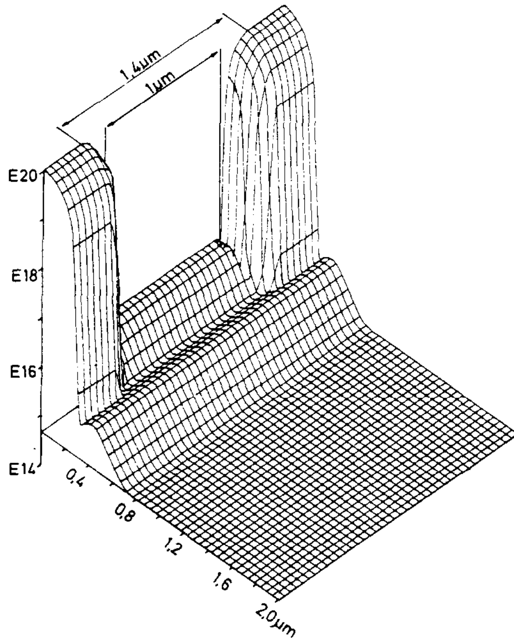

Fig. 1. Doping profile of the analyzed devices.

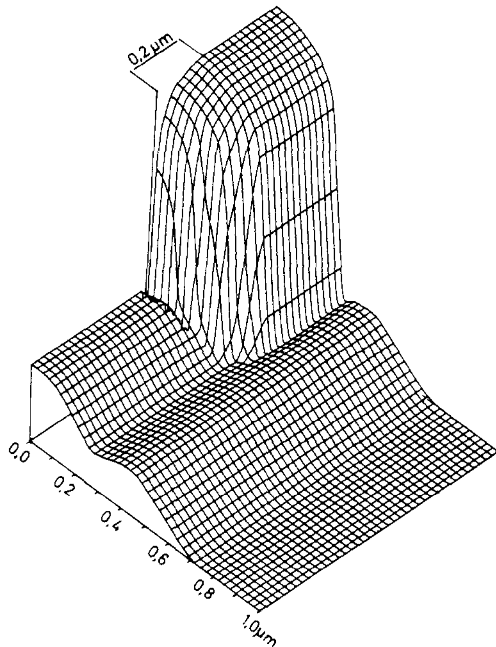

Fig. 2. Enlarged detail of Fig. 1.

3. DEFINITION OF THE THRESHOLD VOLTAGE

For an analysis of the behaviour of the threshold voltage first one has to formulate an adequate definition of the threshold voltage. The most common definitions are based on the extrapolation of an output characteristic. However, one drawback of extrapolation methods lies in their inaccuracy and in the experimental effort. Mainly owing to these mentioned reasons we define threshold voltage in the following simple and definite way: It is that applied gate voltage, at which the device sinks 0.1 microamps times the channel width per channel length. The channel length is defined as the distance between the metallurgical junctions. With this definition it is ensured that no threshold voltage shift vs channel length for long devices occurs and we can, therefore, directly obtain a quantitative measure for the influence of the short channel effects. It is probably necessary at this point to mention that drain bias and bulk bias are not explicit parameters in our definition of the threshold voltage. The dependence on those parameters has to be obtained by certain characteristics, namely: threshold voltage vs drain bias, threshold voltage vs bulk bias. Our definition is naturally arbitrary—as arbitrary as any definition—so one might have to argue about the quantitative value of the used constant (0.1 μA). For devices with a steep subthreshold characteristic, and only such devices are of practical relevance, we think that the constant we use is quite suitable. For devices with a degraded subthreshold characteristic any definition of a threshold voltage becomes meaningless.

Figure 3 shows the threshold voltage vs channel length for our devices. An operating point of 3 V drain bias and -2 V bulk bias has been chosen as a fair tradeoff for the comparison of different channel lengths. To avoid confusion all the following figures will also refer to this operating point. Figure 3 reflects the well known decrease of the threshold voltage with shrinking device length, which becomes dramatic at a length of below one micron.

4. SENSITIVITIES

Usually in papers on short channel MOS transistors a comparison between theoretical curves and selected experimental results is given. Some of them report on statistical measurements[4], but only one paper[5], to our knowledge, deals explicitly with the sensitivity of an electrical property, namely the threshold voltage. However, with respect to the inherent dependence of most properties on the dispersion of geometry and technology, it seems to be a real necessity to analyse and present these dependences directly. Therefore, we carried out numerical investigations to extract the most important sensitivities. A two dimensional simulation

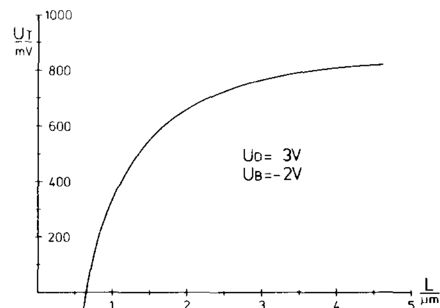

Fig. 3. Threshold voltage vs channel length.

program like MINIMOS[3] is excellently suited for numerical investigation of the sensitivity of device properties to dispersion of design and process parameters. MINIMOS solves Poisson's equation and the current continuity equation consistently over a matrix of nodes representing the cross section of a MOS transistor. First, the physical model parameters of the computer program have to be matched to those corresponding to measured characteristics, i.e. the program has to be "calibrated". This procedure has to be done with non critical transistors with relatively long channels because the measured characteristics should deviate only minimally for inaccuracies in geometry and in technology. This "calibration" procedure should certainly be done for every technology which is to be analyzed numerically as the formulae which are used in a simulation program for modelling the physical parameters (e.g. mobility) are partly heuristic. A few "constants" of those formulae have to be fitted if total agreement of simulation and measurement is desired. It is certainly absurd, and physically invalid, to change the physical model parameters when simulating transistors with just different channel lengths (for example) because all effects due to changes in the channel length are principally included in the structure of the fundamental semiconductor equations and not in their parameters.

In order to obtain a sensitivity by computer simulation, one has to vary the interesting parameter (e.g. channel length) in the vicinity of its nominal value and then to differentiate with the results (e.g. threshold voltage). This parameter variation must certainly be done within a small range because the validity of linearization which is presupposed with the whole strategy has to be ensured. On the other hand, it is necessary to have a sufficiently large range of parameter variation to avoid cancellation errors at the (numerical) differentiation.

This parameter variation within a small range cannot be performed experimentally, in general. A minute change of a process parameter which is reproducable piles up tremendous fabrication problems or inherent costs. However, with a fast modelling program the partial derivative of any electrical property with respect to any technological or geometrical parameter can be calculated easily with the outlined strategy. Thus numerical investigations are ideally suited for the performance of sensitivity analysis.

Figure 4 shows the partial derivative of the threshold voltage with regard to the channel length vs channel length for our devices; that is, the sensitivity of the threshold voltage on tolerances of the channel length. Assume a transistor with an effective channel length of one micron accurate to 10%. With this figure one can read an uncertainty of the threshold voltage of +/− 60 mV.

Figure 5 shows the sensitivity of the threshold voltage to the deviation of the oxide thickness. As one probably has not expected at first glance this sensitivity decreases for devices with short channels. This is due to the decreasing influence of the bulk charge with shrinking channel lengths. Note that this figure is qualitatively very similar to the figure showing the threshold voltage vs

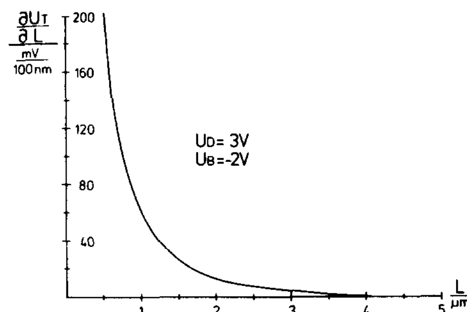

Fig. 4. Sensitivity on channel length tolerances.

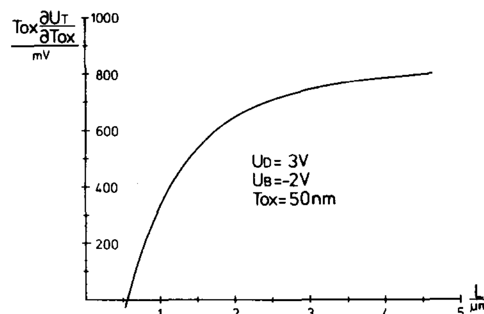

Fig. 5. Sensitivity on oxide thickness tolerances.

channel length (Fig. 3). This fact can be understood analytically by recalling the simple formula for the threshold voltage:

$$UT \doteq \Phi_{MS} + 2 \cdot \Phi_F - (Q_{fs} + Q_b)/C_{ox}$$

(without short-channel effect)

$$C_{ox} = \epsilon_{ox}/T_{ox}$$

$$\partial UT/\partial T_{ox} \doteq -(Q_{fs} + B_b)/\epsilon_{ox}$$

$$UT \doteq (\partial UT/\partial T_{ox}) \cdot T_{ox} + \text{const.}$$

With an uncertainty of 5% of the oxide thickness, one has an uncertainty of about +/−40 mV for a 5 micron device and not even half this value for a 1 micron device. However, one should not be delighted by this fact. The decrease of the sensitivity results from the decrease of the controllability of the transistor by the gate.

Figure 6 shows the sensitivity of UT on junction depth tolerances vs channel length. A one micron device with an uncertainty of 10% in the junction depth, thus has an uncertainty of about −/+40 mV of the threshold voltage. The underlying physical cause of this sensitivity is the reduction of the channel charge by the depletion regions of source and drain[6].

Figure 7 shows the sensitivity of UT on drain bias variation. A 300 mV change, that is 10% of the applied bias, results in about 30 mV change of the threshold

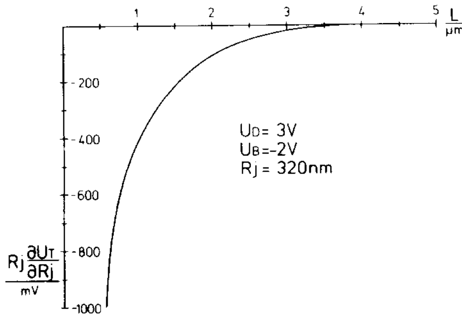

Fig. 6. Sensitivity on junction depth tolerances.

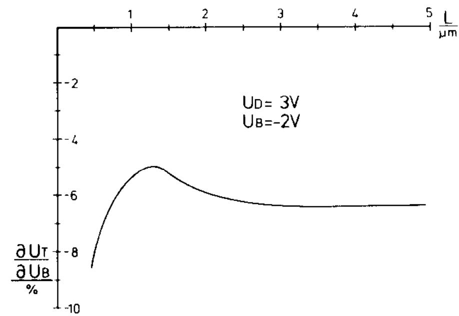

Fig. 8. Sensitivity on bulk bias variation.

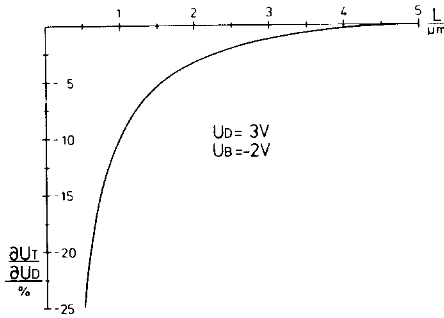

Fig. 7. Sensitivity on drain bias variation.

voltage for this operating point. Again the modulation of the depletion region of the drain is the relevant physical effect. At first glance it seems to be easy to measure this particular sensitivity in even a minimally equipped laboratory. However, in case of short channel devices just the nominal values of the process and geometry parameters are known for an individual device. The dispersion of these parameters would merely allow to extract bars by statistical measurements which again will make the analysis expensive and time consuming.

Figure 8 shows the sensitivity of UT on bulk bias variation. A 200 mV change, that is again 10%, results in a threshold shift of about 11 mV, which is usually not dramatic. (For the practical problem, however, one has to deal with a sum of all uncertainties. Therefore, this influence may also become important.) An interesting detail of this figure is the fact that the sensitivity decreases first with shrinking channel length and at a certain length begins to increase rapidly. This behaviour is caused by a superposition of the short channel effect, which decreases this particular sensitivity, and the punch-through effect, which increases the sensitivity. For long channel devices it is fairly simple to estimate this sensitivity analytically:

$$\partial UT/\partial UB \doteq -1/C_{ox} \cdot \partial Q_b/\partial UB$$

with:

$$Q_b \doteq q \cdot (N_b \cdot y_c + \text{Dose}).$$

For the partial derivative only y_c has to be considered:

$$\partial UT/\partial UB \doteq -(T_{ox}/y_c) \cdot (\epsilon_{si}/\epsilon_{ox}).$$

With a value of about two micrometer for y_c one obtains a sensitivity of approximately -7.5% which is confirmed by the more exact two-dimensional calculations.

Figure 9 shows the influence of an implantation energy fluctuation. Qualitatively the superposition of the short channel effect and the punch-through effect is again apparent. The absolute value of this particular sensitivity is low due to the fact that the depletion region below the channel covers the whole implanted region at this operating point.

Figure 10 shows the sensitivity of UT on uncertainties of the implantation dose. The figure is rather similar to the last one as expected. An analytical estimate for the long channel transistor can be obtained in a straight forward way for this sensitivity:

$$\partial UT/\partial \text{Dose} \doteq -1/C_{ox} \cdot \partial Q_b/\partial \text{Dose}$$
$$= q/C_{ox} = 23\ \text{mV}/10^{10}\ \text{cm}^{-2}.$$

Figure 11 shows the temperature coefficient of the threshold voltage for our devices. We have, qualitatively, a similar behaviour to that already discussed, namely the superposition of short channel effect and punch through. The absolute value is around -1 mV/K. The qualitative behaviour as well as the absolute value of this sensitivity have been verified by fairly complicated experiments[7].

5. GLOBAL SENSITIVITY

The partial derivatives denote isolated sensitivities on

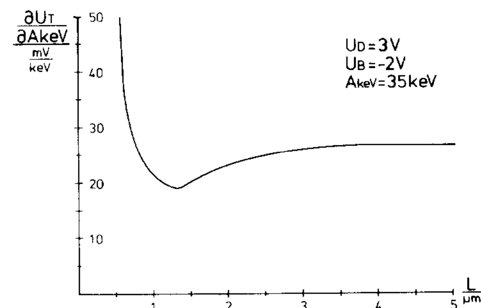

Fig. 9. Sensitivity on implantation energy tolerances.

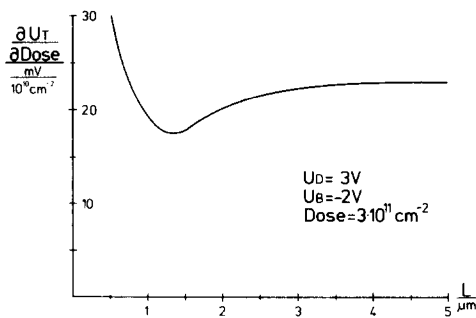

Fig. 10. Sensitivity on implantation dose tolerances.

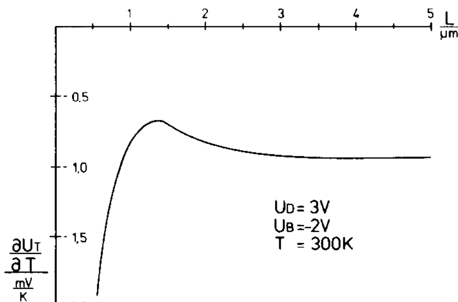

Fig. 11. Sensitivity on temperature variation.

Parameter	X	|ΔX|	%
L		100nm	
TOX	50nm	2.5nm	5
RJ	320nm	32nm	10
UD	3V	150mV	5
UB	-2V	100mV	5
AKEV	35keV	0.7keV	2
DOSE	$3.10^{11}cm^{-2}$	$6.10^{9}cm^{-2}$	2

Fig. 12. Desired process and operating tolerances.

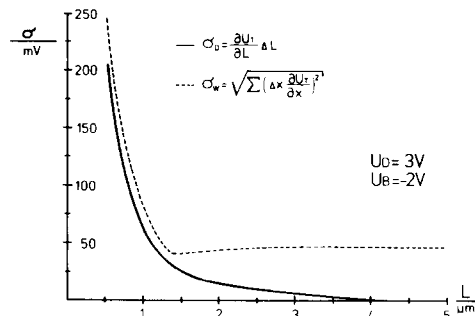

Fig. 13. Reproducability of the analyzed devices vs channel length.

a certain set of parameters. These values show which parameters are the most critical ones. However, in addition, a global sensitivity number indicating the cumulative effect of the isolated sensitivities is useful. The global sensitivity is related to a certain technology and its expected application. It should indicate the limit of channel length reduction. To obtain such a global number typical ranges of deviation of design parameters have to be specified. The table in Fig. 12 is an example for such a specification. In this example a rather small value of the absolute uncertainty of the channel length (100 nm) has been chosen. For long devices this value is unrealistic, but in consideration of a one micrometer technology 100 nm absolute uncertainty represents 10% relative dispersion, which is relatively large. The tolerances of the remaining parameters in Fig. 12, however, represent a good laboratory standard.

Figure 13 shows the global threshold voltage sensitivity based on the specifications of Fig. 12. σ_D denotes the uncertainty of the threshold voltage for identical devices on the same chip. *D* stands for device. This sensitivity is given by just the length influence, as the other parameters are commonly very homogeneous across one chip. σ_W, *W* stands for wafer, denotes the uncertainty for identical devices on wafers, which have been fabricated with different charges. Here one has to use a Euclidian norm over all deviations. Note that this value is highly constant down to a certain channel length, but then increases dramatically. The channel length at which the excellently pronounced knee is located, at 1.4 microns for our devices, thus can be interpreted as the practical limit of channel length reduction due to threshold uncertainty. Nevertheless should be noted that the data in Fig. 12 are to be understood as an example which is mainly of importance for our technology.

6. CONCLUSION

The threshold voltage sensitivity to various parameters in short channel MOS transistors of a specific technology has been evaluated with the two dimensional simulation program MINIMOS as an example of practical importance for a generally applicable method. The most critical parameters have been obtained, and furthermore, a global sensitivity number has been derived indicating the practical limit of miniaturization for the analyzed process technology.

Acknowledgements—This work has been supported by the "Fond zur Förderung der wissenschaftlichen Forschung" (Projekt Nr. S22/11). Essential help of Siemens AG, Munich, in providing MOS devices is gratefully acknowledged. Critical reading of our manuscript by Dr. J. Machek is much appreciated. Last but not least the authors wish to thank Dr. D. Schornböck and the whole staff of the computer centre for the excellent computer access.

REFERENCES

1. R. Dennard and F. Gaensslen *et al.*, *IEEE* **SC-9**, 256 (1974).
2. H. Masuda, M. Nakai and M. Kubo, *IEEE* **ED-26**, 980 (1979).
3. S. Selberherr, A. Schütz and H. Pötzl, *IEEE* **ED-27**, 1540 (1980).
4. E. Demoulin and J. Greenfield *et al.*, Process statistics of submicron MOSFET's. *IEDM Technical Digest*, 34 (1979).
5. K. Yokoyama, A. Yoshii and S. Horiguchi, *IEEE* **ED-27**, 1509 (1980).
6. R. Troutman, *IEEE* **ED-26**, 461 (1979).
7. D. Takacs, U. Schwabe and U. Bürker, *IEDM Technical Digest*, p. 569 (1980).