Solid-State Electronics Vol. 25, No. 3, pp. 177–183, 1982
Printed in Great Britain.

0038-1101/82/030177-07$03.00/0
Pergamon Press Ltd.

A TWO-DIMENSIONAL MODEL OF THE AVALANCHE EFFECT IN MOS TRANSISTORS

ALFRED SCHÜTZ, SIEGFRIED SELBERHERR and HANS W. PÖTZL
Institut für Allgemeine Elektrotechnik und Elektronik, Abteilung für Physikalische Elektronik, TU Vienna, Gusshausstrasse 27, A-1040 Wien, Austria

and

Ludwig Boltzmann Institut für Festkörperphysik

(*Received* 16 *December* 1980; *in revised form* 17 *June* 1981)

Abstract—A two-dimensional self consistent MOS transistor model accounting for the avalanche effect is described. The classical semiconductor equations—Poisson's equation and the two carrier equations—are solved with the finite difference method. The pair production rate is evaluated at any mesh point and dominates the inhomogeneity term of the carrier continuity equations in case of avalanche. Calculated and measured current–voltage characteristics are in good agreement and thus support our model. For a 3 μm device the electrical potential, the carrier densities, and the generation rates are shown in quasi three-dimensional plots from which the avalanche generation in the pinch-off region becomes apparent. Furthermore, hole storage close to the interface is seen to take place in the channel up to the vicinity of the source region. The corresponding barrier lowering leads to increased electron injection from the source and enhanced avalanche. The barrier lowering is supported by the influence of the parasitic bulk resistance.

INTRODUCTION

A model for two-dimensional MOS transistor analysis—MINIMOS—has been published recently[1]. That model, however, in order to simplify the computation neglects any hole current and generation/recombination mechanisms including avalanche. Thus only one continuity equation has to be solved which results in reduced computing time. Although the first version of MINIMOS covers a certain area of MOS transistor simulation there are applications where the modeling of the avalanche effect is essential.

It has been usual until now to treat avalanche problems by solving Poisson's equation thus obtaining a solution for the electrical potential distribution and to evaluate the ionization integral. This implies that the strongly field dependent ionization coefficients are integrated over the high field regions. As a result of this method one arrives at multiplication factors which describe the increase of currents due to avalanche. Since the carrier densities need not be calculated this method seems to be very effective in calculating breakdown voltages.

As an example of these techniques, Toyabe *et al.*[2] analysed the breakdown phenomenon in MOSFET's. With their CADDET program the carrier equation for one carrier type (electrons in *n*-channel MOSFET's and holes in *p*-channel MOSFET's) is solved. As CADDET uses the stream function technique[3] the avalanche generation rate cannot be included as inhomogeneity into the continuity equation and any influence of the increased carrier densities on the electrical potential has to be neglected.

In contrast, the present paper is based on the consistent solution of both inhomogeneous continuity equations accounting for generation and recombination terms. First, the basic equations are presented, and the ionization rates and the geometry are specified. Details of the numerical solution method are given in Appendix 1. Then a 3 μm gate length transistor is investigated. Drain and bulk currents are shown to be in good agreement with experiment. The avalanche generation rate and the electron and hole density distributions are presented and discussed in detail.

THE MODEL

As shown in [4] the following basic semiconductor equations for the steady-state have to be solved:

$$\operatorname{div} \epsilon \operatorname{grad} \psi = -q(p - n + N_D - N_A) \quad \text{(Poisson's equation)} \tag{1}$$

$$\begin{aligned} \operatorname{div} \mathbf{J}_n &= -q(G - R) \\ \operatorname{div} \mathbf{J}_p &= q(G - R) \end{aligned} \quad \text{(continuity equations)} \tag{2}$$

with the current relations

$$\begin{aligned} \mathbf{J}_n &= -q(\mu_n n \operatorname{grad} \psi - D_n \operatorname{grad} n) \\ \mathbf{J}_p &= -q(\mu_p p \operatorname{grad} \psi + D_p \operatorname{grad} p) \end{aligned} \tag{3}$$

in which we assume the validity of Einstein's relation

$$D_{n,p} = \mu_{n,p} \frac{kT}{q}.$$

The right hand terms in the eqns (2) are given by the generation and recombination rates and are usually negligible in the non-avalanche region. However, in our model we cannot neglect it because the avalanche effect arises from electron-hole pair generation. Recombination should also not be neglected because the high level of ionization in the high field regions can lead to a drastic increase of carrier densities all over the device thus

making recombination more important. In our model, therefore, the $(G-R)$ term is due to thermal (including surface) generation/recombination, Auger recombination, and avalanche generation:

$$(G-R)=(G-R)_{\text{th}}+(G-R)_s+(G-R)_{\text{Aug}}+G_a, \tag{4}$$

with

$$(G-R)_{\text{th}}=\frac{n_i^2-p\cdot n}{\tau_n(p+p_1)+\tau_p(n+n_1)} \tag{5a}$$

$$(G-R)_s=\frac{n_i^2-p\cdot n}{(p+p_1)/s_n+(n+n_1)/s_p}\cdot\delta(y) \tag{5b}$$

$\delta(y)$ is the delta function.

$$(G-R)_{\text{Aug}}=(n_i^2-p\cdot n)(C_n n+C_p p) \tag{5c}$$

$$G_a=\frac{|\mathbf{J}_n|}{q}A_n\exp\left(-\frac{B_n|\mathbf{J}_n|}{|\mathbf{E}\cdot\mathbf{J}_n|}\right)+\frac{|\mathbf{J}_p|}{q}A_p\exp\left(-\frac{B_p|\mathbf{J}_p|}{|\mathbf{E}\cdot\mathbf{J}_p|}\right) \tag{5d}$$

with A_n (A_p) and B_n (B_p) the electron (hole) ionization coefficients.

Equations (5a) and (5b) are the usual Shockley–Read–Hall terms for the thermal bulk and surface recombination processes and eqn (5c) represents the Auger recombination term as given in [5]. In eqn (5d) we assume the validity of Chynoweth's law [6]

$$\alpha_{n,p}(\mathbf{E})=A_{n,p}\exp\left(-\frac{B_{n,p}}{|\mathbf{E}|}\right). \tag{5e}$$

In the exponential expression of eqn (5d) we do not use the absolute value of the electric field but the field component parallel to the current density. The field component perpendicular to the current flow does not cause ionization since the carriers only gain energy from the field component parallel to their motion.

Various authors have determined the ionization parameters A_n, A_p, B_n, B_p [7, 8]. The experimental method of van Overstraeten *et al.* [9] seems to be very reliable and our own investigations as well as those by other authors [10] support their results.

We now have to solve a system of three partial differential eqns (1), (2a), (2b). Essentially this is done numerically by discretisation into finite differences in a modified Gummel cycle [11] as shown in Appendix 1. In this, the rather expensive simultaneous solution is avoided and the high memory requirements [12] are relaxed. Grid spacing has to be treated very carefully to avoid too large a discretization error according to rough spacing as well as unnecessarily large computation expense because of too fine spacing. In case of avalanche this becomes eminently important as the ionization rate is very sensitive to the electric field thus an accurate calculation of the avalanche generation requires a much more accurate calculation of the electric field. To minimize computer time at justifiably small errors the finite difference mesh is adjusted to the potential-, carrier- and doping distribution and is checked and, if necessary, is modified during the iteration process. The electric field used in (5d) is derived numerically from the electric potential by simple five point discretization formulae which is consistent with the discretization of Poisson's equation.

Figure 1 shows the simulation geometry. Current flow in deep bulk as a consequence of avalanche-generated holes which are rejected from the source and drain regions causes a voltage drop across the parasitic bulk resistor. There are several options to account for this fact: (a) a truly three-dimensional analysis; (b) extension of the simulation over the entire bulk area; (c) extension of the two-dimensional simulation over the depletion region and using an (effective) bulk resistor. If one wants to avoid excessive computing time associated with (a), option (c) is to be preferred because it allows inclusion of current spread into the third dimension and, also, consumes less computing time than (b). Thus the voltage drop across the parasitic bulk resistor simulates a more positive bulk bias and, if large enough, is able to forward-bias the parasitic bipolar npn transistor (according to source, bulk and drain). This causes a larger drain current and facilitates the breakdown which then will occur at smaller drain voltages. A simple method for deriving the effective bulk resistor will be given in Appendix 2.

RESULTS

In order to check our model we have measured an n-channel MOSFET with 3 μm nominal gate length. A comparison of calculated and measured drain current vs drain voltage characteristics is given in Fig. 2. To improve this presentation a different scale has been used for the $V_{GS}=1$ V curve than for the other two curves. Circles denote the simulation results with an effective bulk resistor $R_{\text{Bulk}}=3000\,\Omega$. Ignoring its influence one arrives at smaller drain currents as indicated by triangles. If impact ionization is neglected the characteristic increase in drain current cannot be observed. The effect of the bulk resistor is very pronounced close to the avalanche breakdown. In this region accounting for bulk resistance is absolutely necessary to obtain good agreement between theory and experiment. The corresponding bulk current can be seen in Fig. 3. In this plot the simulation results are drawn only for $R_{\text{Bulk}}=3000\,\Omega$ as there would be little difference to $R_{\text{Bulk}}=0\,\Omega$ due to the logarithmic scale and the bulk current would vanish if the impact ionization is neglected.

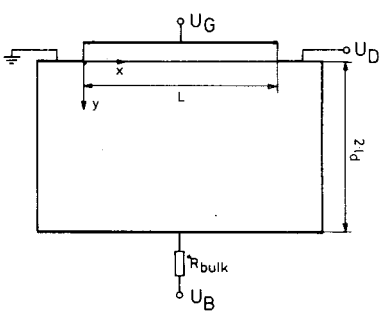

Fig. 1. The basic simulation geometry.

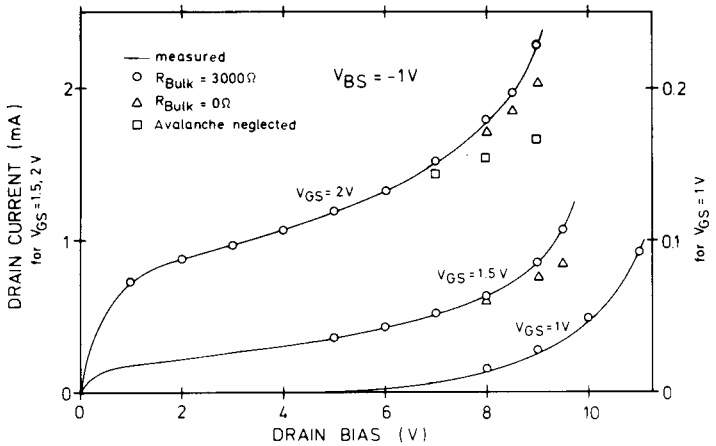

Fig. 2. Drain current vs drain voltage: Use left scale for $V_{GS} = 2$ V and $V_{GS} = 1.5$ V, whereas right scale should be used for $V_{GS} = 1$ V.

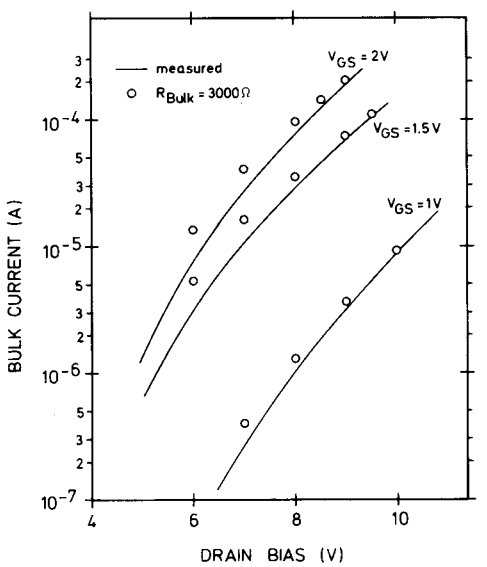

Fig. 3. Bulk current vs drain voltage.

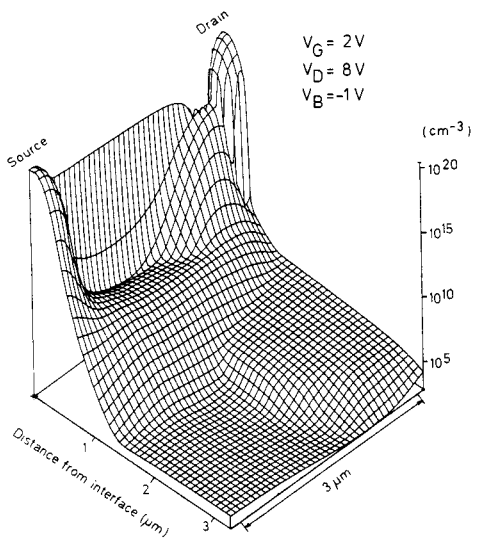

Fig. 4. Electron density.

The comparison of experimental and theoretical results is the only possible check of the numerical calculations. However, the main strength of the numerical model lies in the availability of all internal physical quantities. Thus we will now discuss the presentation of these quantities in Figs. 4–10. We chose an operating point with 2 V on gate and 8 V on drain—just 1.5 V below breakdown.

Figure 4 shows the electron density which is large in the source and drain regions due to high donor doping and also at the interface because of strong inversion. Below the inversion layer a suppressed punch channel can be seen. The current it carries is negligible compared to the inversion current.

The potential distribution is plotted in Fig. 5. The lateral electric field is small in the strongly inverted

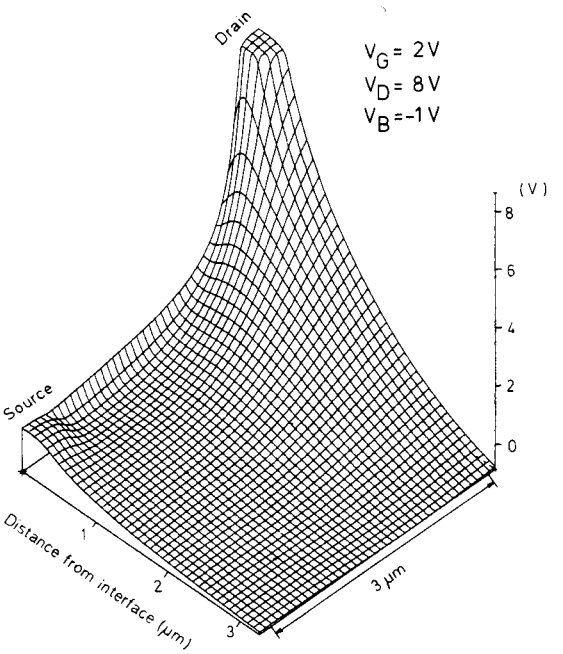

Fig. 5. Electrical potential distribution.

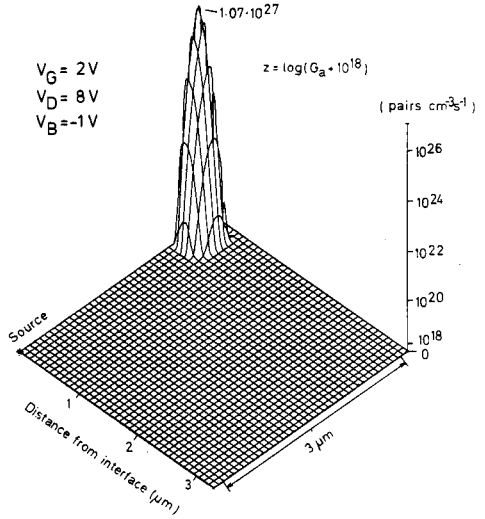

Fig. 6. Avalanche generation rate.

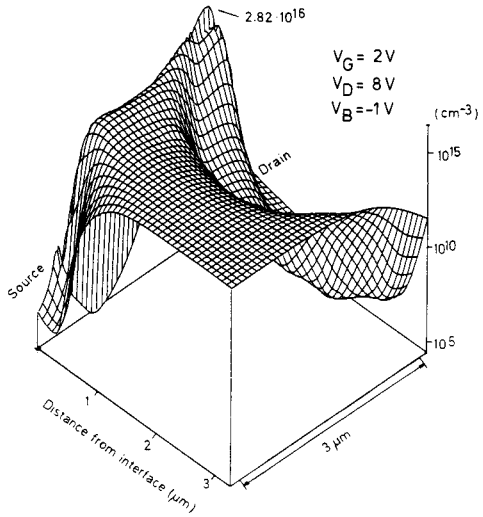

Fig. 9. Hole density with avalanche simulation.

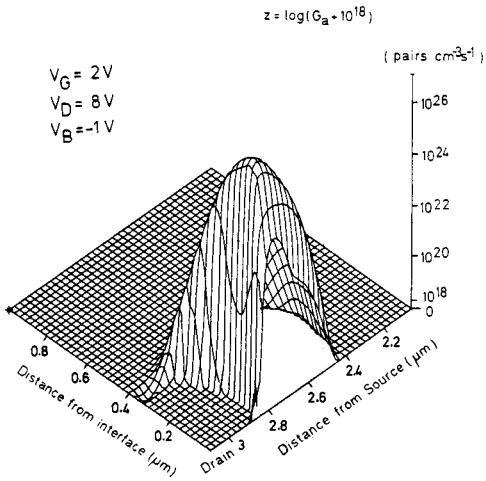

Fig. 7. Detailed plot of the avalanche generation rate.

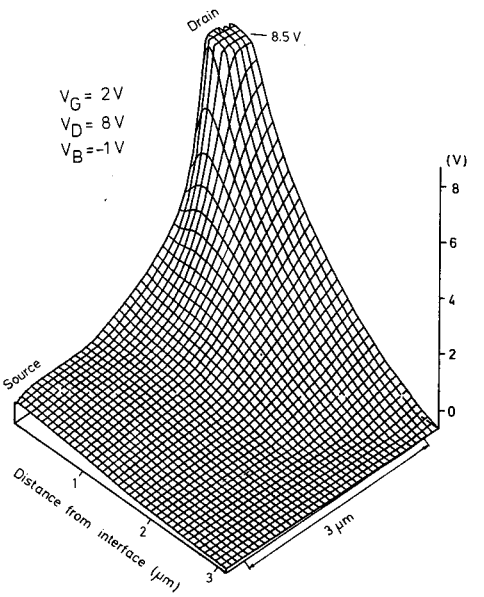

Fig. 10. Hole quasifermipotential.

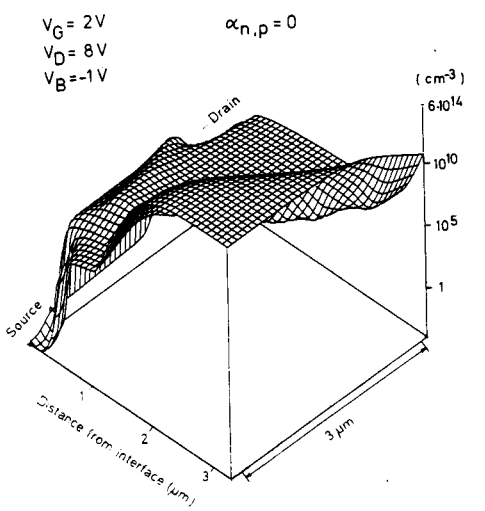

Fig. 8. Hole density without avalanche simulation.

channel, whereas it is very large in the pinch-off region. Thus we assume the avalanche to occur mainly in that region. This is corroborated by Figs. 6 and 7 which show the avalanche generation rate in pairs per s and cm^3. Figure 7 is a zoomed plot of Fig. 6 and is turned around by 180 degrees to show more details. We find the maximum of impact ionization not to occur directly at the interface but in a depth of about 0.2 μm. This is due to the fact that the electron current is pushed away from the interface in the pinch-off region. Figure 7 also shows the rapid decrease of the ionization rate in the drain region as the electric field is suppressed by the high doping level. This decrease approximately follows the curve of the *p–n* junction of the drain region.

The relation between the generation rate and the carrier densities is established by the following considera-

tion: Carriers generated by ionization leave the pinch-off region with saturation velocity. This current flow must be equal to the current given by the integral over the ionization rate times the electric charge. Thus we can write, in a one-dimensional consideration,

$$I = Aqnv_s = Aq \int G_a \, dx$$

where A is the medium area of the current flow.

Assuming a medium ionization rate of $10^{27}\,cm^{-3}\,s^{-1}$ over a medium length $l = 0.2\,\mu m$ and a saturation velocity $v_s = 10^7$ cm/s we get an increase of $2 \times 10^{15}\,cm^{-3}$ in the carrier densities. This increase is unimportant for electrons as can be seen from Fig. 4. However, it will dominate over the non-avalanche hole density as we shall see in the next figures.

To study the effects of the avalanche we start from the hole density without avalanche (Fig. 8). In the undisturbed bulk region we find a hole density of $6 \times 10^{14}\,cm^{-3}$, which is the value of the substrate doping. In the depletion regions it decreases to about $10^6\,cm^{-3}$, this value being mainly due to thermal generation. In the highly doped source and drain regions the drop is even more pronounced because of the large electron density in those areas.

Figure 9 shows the hole density which results if the avalanche is included. In the pinch-off region we note a drastic increase of hole density. The accumulation at the interface is due to the electric field pulling the holes to the interface as the gate potential is more negative than the drain potential. In the adjacent channel region there is a strong transverse field component near the interface as seen from the gradient of the potential (Fig. 5). This field pulls the holes away from the interface. However, as the field quickly diminishes in transverse direction the holes are stored in a region which is not far from the interface. Looking in the direction transverse to the interface we note a maximum of hole density arising according to the above consideration. As a consequence the hole density increases markedly also close to the source in comparison to Fig. 8. This important feature is supported by a high parasitic bulk resistance because it counteracts the hole current flow into deep bulk. The increased hole density near the source will affect the electrical potential giving rise to an increased electron density and, therefore, to a larger electron current and avalanche generation. If the hole density is increased even more by ionization, the parasitic bipolar npn transistor will finally be turned on resulting in complete breakdown. As this effect includes internal feedback it can lead to negative resistance which is usually called the snap-back phenomenon. Concluding we see from Fig. 9 that holes do not flow from the pinch-off region to bulk directly but first towards the interface, from there towards source, and then into bulk. Of course this effect is less pronounced for long channel devices.

In Fig. 10 we find the distribution of the hole quasi-fermipotential (QFP). It reveals the interesting fact that the QFP in the drain region exceeds the applied drain voltage which is the value of the QFP at the drain contact. This implies the existence of a local maximum of the QFP near drain and demonstrates the effectiveness of ionization as a hole source if holes are imagined to slope down the QFP hill of this figure. The local maximum of the QFP is not seen obviously because the lower value of the QFP at the drain contact is covered by the higher peak just before it. However, the very critical reader will find the contour lines to drop at the drainmost end of this plot.

CONCLUSION

We have studied the avalanche effect in MOS transistors with the aid of a two-dimensional model. This model is shown to predict reliable values for node currents. Its main strength, however, lies in the prediction of internal quantities. Qualitatively we identified three effects which can lead to complete breakdown:

(1) infinite multiplication factor in high field regions;
(2) voltage drop due to hole flow in the substrate[2];
(3) turn-on of the bipolar transistor because of the increase of hole density near the source electrode.

Items 2 and 3 are not identical, however, they are related as stated above. Which of these processes determines the breakdown voltage will depend on device and circuit parameters. The process "voltage drop in bulk region", for example, can be made less effective by applying a more negative bulk voltage.

The program is available for anyone just for the handling costs.

Acknowledgements—This work has been supported by the "Fonds zur Förderung der wissenschaftlichen Forschung" (Project S22/11). The authors gratefully acknowledge essential help in providing MOS devices by Siemens AG Munich and want to thank the "Interuniversitäres Rechenzentrum der Technischen Universität Wien" for providing a generous amount of computer time. Further thanks are due to Prof. Dr. E. Bonek for critically reading the manuscript.

REFERENCES

1. S. Selberherr, A. Schütz and H. W. Pötzl, *IEEE Trans. Electron Devices* **ED-27**, 1540–1550 (1980).
2. T. Toyabe, K. Yamaguchi, S. Asai and M. S. Mock, *IEEE Trans. Electron Devices* **ED-25**, 825–832 (1978).
3. M. S. Mock, *Solid St. Electron.* **16**, 601–609 (1973).
4. W. V. van Roosbroeck, *Bell Syst. Tech. J.* **29**, 560–607 (1950).
5. D. C. d'Avanzo, *Modeling and Characterization of Short-channel Double Diffused MOS Transistors.* Stanford University Technical Report No. G-201-6, 1980.
6. A. G. Chynoweth, *Phys. Rev.* **109**, 1537–1540 (1958).
7. C. A. Lee, R. A. Logan, R. L. Batdorf, J. J. Kleimack and W. Wiegmann, *Phys. Rev.* **134**, A761–773 (1964).
8. J. L. Moll and R. van Overstraeten, *Solid-St. Electron.* **6**, 147–157 (1963).
9. R. van Overstraeten and H. de Man, *Solid-St. Electron.* **13**, 583–608 (1970).
10. V. A. K. Temple and M. S. Adler, *IEEE Trans. Electron Devices* **ED-22**, 910–916 (1975).
11. H. K. Gummel, *IEEE Trans. Electron Devices* **ED-11**, 455–465 (1964).
12. P. E. Cotrell and E. M. Buturla, *Proc. NASECODE* 1 *Conf.*, pp. 31–64, Dublin (1979).

APPENDIX 1

The modified Gummel cycle:

Gummel[11] suggested to solve Poisson's equation and both continuity equations step by step and to repeat this procedure

until convergence has been obtained. In n-channel MOS transistors without too strong an avalanche, there is only negligible hole current flow and the quasifermipotential for holes remains fairly constant, at least in regions where hole density is large. Thus, hole density is determined mainly by the electric potential. If we treat Poisson's equation as given in [1] coupling of that equation with hole continuity equation is extremely weak whereas it turns out to be very strong with the electron continuity equation, especially if the device operates in strong inversion. Thus computer time can be economized by solving more often Poisson's equation and the electron continuity equation than the hole continuity equation (Fig. 11). Only in certain cycles hole carrier equation is included into the procedure. If it turns out, after calculation of holes, that changes are large, potential and holes will be recalculated until changes in holes become small. Then the procedure starts with electron carrier equation once again unless convergence has been obtained. Thus the number of iterations is small with weak or negligible avalanche as well as with strong avalanche. The question when to change from electron to hole equation is not easy to be answered theoretically. However, it turned out to be a good compromise solving for holes whenever changes in electron densities have become small, but at least every fifth cycle.

Ionization rates are recalculated in all those cycles where the hole equation is solved. As we start the procedure without ionization, the total current will increase by a certain amount after each recalculation of the ionization rates. This increase will be approximately given by the ionization due to the excess carriers arising from updating the carrier densities. Thus the current will approximately behave like a geometric series.

Besides of economizing computer time, in case of negative resistance, this procedure offers the benefit that we only get the lowest value for the current and we have no problem with numerical instabilities. Only if drain voltage is beyond breakdown, the geometric series will not converge resulting in infinite current which is consistent with experiment.

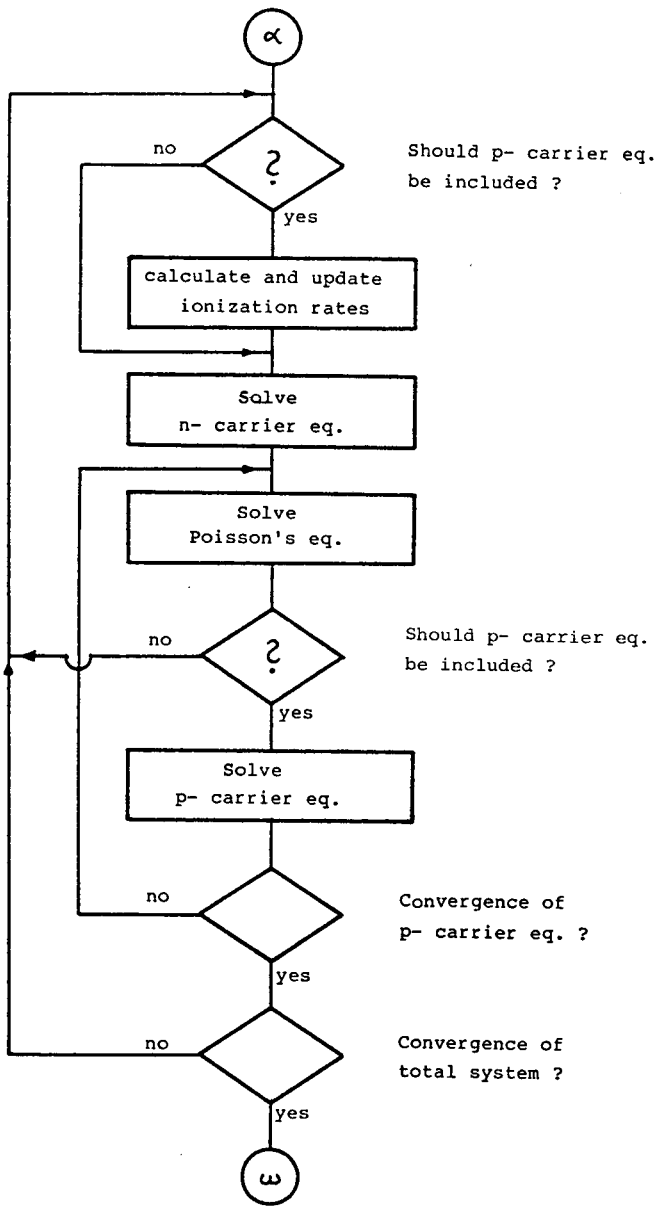

Fig. 11. The modified Gummel cycle.

APPENDIX 2

Derivation of the bulk resistor:

In this section we analyse the bulk resistor. As its value has to be entered into the input of the program, a simple expression should be found. So let us assume the current to spread by an angle of 45° into both directions perpendicular to its flow (x- and z-direction in Fig. 12). Furthermore we neglect any diffusion current. Thus we get for the electric field in deep bulk

$$\frac{d\psi}{dy} = \frac{I_B}{\kappa A} = \frac{I_B}{\kappa(L+2y)(W+2y)} \tag{A2.1}$$

with κ standing for the conductivity of the substrate and A standing for the area of the current flow. L and W are channel length and channel width, respectively. Integrating this equation along y from the end of the simulation area d_s to the bulk contact we get

$$R_{\text{Bulk}} = \frac{\int_{d_s}^{d} \frac{d\psi}{dy}\,dy}{I_B} = \frac{1}{2\kappa(W-L)}\left(\ln\left(\frac{L+2d}{L+2d_s}\right) - \ln\left(\frac{W+2d}{W+2d_s}\right)\right). \tag{A2.2}$$

For $L = W$ this equation simplifies to

$$R_{\text{Bulk}} = \frac{d-d_s}{\kappa(L+2d)(L+2d_s)}. \tag{A2.3}$$

This calculation is rather crude compared with the elaborate solution of the basic equations. However, any more accurate calculation would be very complicated and the present method is sufficient to investigate the influence of the parasistic bulk resistance at least qualitatively.

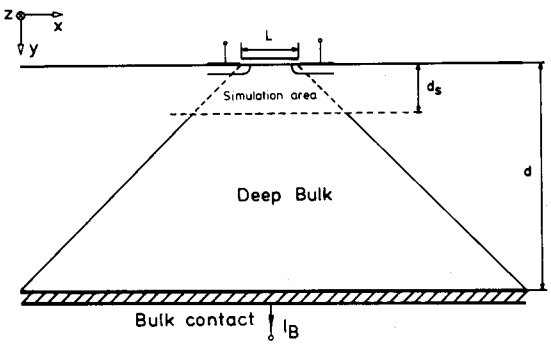

Fig. 12. Current flow in deep bulk.