Two-Dimensional Numerical Modeling of Interband Tunneling Accounting for Nonuniform Electric Field

Predrag Habaš, Andreas Lugbauer and Siegfried Selberherr
Institute for Microelectronics, Gusshausstrasse 27-29, 1040 Vienna, Austria

Abstract

A numerical calculation of the band-to-band tunneling in Si-MOSFET is presented. By analyzing the location and the amount of the generated carriers at different terminal biases, it is found that the effect is strongly dependent on the two-dimensional potential and field distribution in the critical area. Significant variations of the electric field along the tunneling path are observed. An expression for the number of the electron-hole pairs generated by internal field emission in a linearly variable field assuming direct tunneling is derived. The expression is based on the two-band $\mathbf{k}\cdot\mathbf{p}$ model and the WKBJ approximation; it reduces to the Kane expression for a constant field. The consequences of the model for the simulation of Si MOSFETs are briefly discussed.

1 Introduction

Interband tunneling has recently become increasingly important for Si MOS technology. It is a source of leakage in memory cells and MOSFETs [1], and can produce a degradation in static and transient bias conditions. In addition, it can be used to design programing memories with a high injection efficiency [2], as a monitor for hot carrier degradation, or to develop devices whose operation is based on this effect. Tunneling is a sensitive function of the electric field, which depends on several parameters such as oxide thickness t_{ox}, dopant distribution, geometry and terminal biases. Consequently, tunneling in MOSFET can only be properly modeled by a numerical approach ([3],[5],[6],[10],[11]).

We have developed and implemented in MINIMOS a two-dimensional numerical model for the band-to-band tunneling. Two approaches have been investigated: **1D** – searching for the tunneling path from the given point in the direction normal to the interface (the normal cross-section often applied in the literature [1],[3],[4],[5]). **2D** – searching from the given starting point (each point in the discretization mesh) in two dimensions for the nearest endpoint whose potential difference is larger than the characteristic tunneling band gap E_g. The tunneling path found is normal to the corresponding equipotential lines, assuming a sufficiently fine local grid. As the tunneling length l_t is less than the the curvature of the equipotential lines, the tunneling between the starting point and the endpoint is treated like a planar one-dimensional problem. To calculate the generation rate N models for direct and indirect tunneling based on the starting field E_s, average field $\overline{E} = E_g/l_t$ and field variable from E_s to E_e (endfield) are implemented and compared. The calculated generation rates

TH0424 - 2/92/0000-0135 $1.00 © 1992 IEEE

of electrons and holes, which are separated in space (Figs.1a-1d), are coupled after filtering with the continuity equations via the generation term. Note that the total charge in the device is strictly conserved. Local grid refinement is performed. Our approach relies on an accurate terminal current calculation, based on the local concentration-dependent weight functions.

2 Potential and field distribution in the critical area

The amount and the distribution of the generated carriers are investigated at different gate, drain and bulk voltages, assuming graded (P) and abrupt (As) drain junctions. Results for a graded drain junction are presented in Figs.1a-1d. Although the quantitative conclusions are different for every specific technology, some general observations can be made:

The 1D approach leads to significantly lower generation rates compared to the 2D model, because the endpoints are found to lie much deeper in the bulk (Fig.1a,1b). The tunneling current (I_D, I_B) is smaller for the 1D than for the 2D model due to: (1) The area where tunneling is possible is a bit larger in the 2D approach due to a lateral effect. Therefore, even if $N(E_s)$ is used, the current I_D will be smaller for the 1D than for the 2D approach. (2) The average t_l is much layer in the 1D model. If we assume $N(\overline{E})$ or a variable field model, I_D becomes much smaller in the 1D approach. Consequently, the 1D approach cannot be applied due to the strongly two-dimensional distribution of the potential and the field in the critical area ([6],[11]).

The field distribution in the critical area close to the interface $E(x, y \approx 0)$ changes weakly as a function of U_{DB}, assuming a constant U_{GD}. The magnitude of the field is determined by U_{GD}. Deeper in the bulk ($\sim 10nm$), where the normal field component is reduced, the field changes as a function of U_{DB}, due to the lateral field. Therefore, the field E_s at the starting points, which are located close to the interface, depends strongly on U_{GD}, but weakly on U_{DB}, while the endfield E_e increases with increasing U_{DB}. These findings enable us to explain the roll-off of the tunneling current with decreasing U_{DB} at a constant U_{DG}, obtained experimentally [6], and found in our simulation (Fig.4). At $U_{DB} = 0$ equilibrium holds. When U_{DS} increases, the band-bending, necessary for the tunneling, occurs already at a low drain-bulk bias (Fig.1d). However, if we assume $N(\overline{E})$, the tunneling becomes small, although E_s is very large. By further increasing U_{DB}, the lateral field increases, the average E_e increases, the tunneling paths become skew, the average l_t is reduced and $\overline{E}$ increases (Fig.1b,1c). At the same time, E_s remains approximately constant. Note that the maximum tunneling (denoted by **B** in Fig.1) does not occur at the point of maximum field (denoted by **A**), [11]. In all three cases the magnitude of the electric field at the points **A** and **B** are: $E(A, y=0) \approx 2.15 MV/cm$, $E(B, y=0) \approx 1.87 MV/cm$. Typical field variations E_e/E_s are ~ 0.7 and ~ 0.6 in Fig.1b and Fig.1c, respectively. We found that the electric field varies typically from 2 to 5 times in common cases. The ratio E_e/E_s can be larger than 1, as is the case for tunneling towards the interface [2], and in p-channel devices. An additional effect, which increases the current by increasing U_{DB}, is a small broadening of the active tunneling area.

To demonstrate the effect of the field variation, the generation rate calculated by different direct tunneling models is shown in Fig.2. A simple one-dimensional analytical model

for the totally depleted uniformly doped region is assumed.

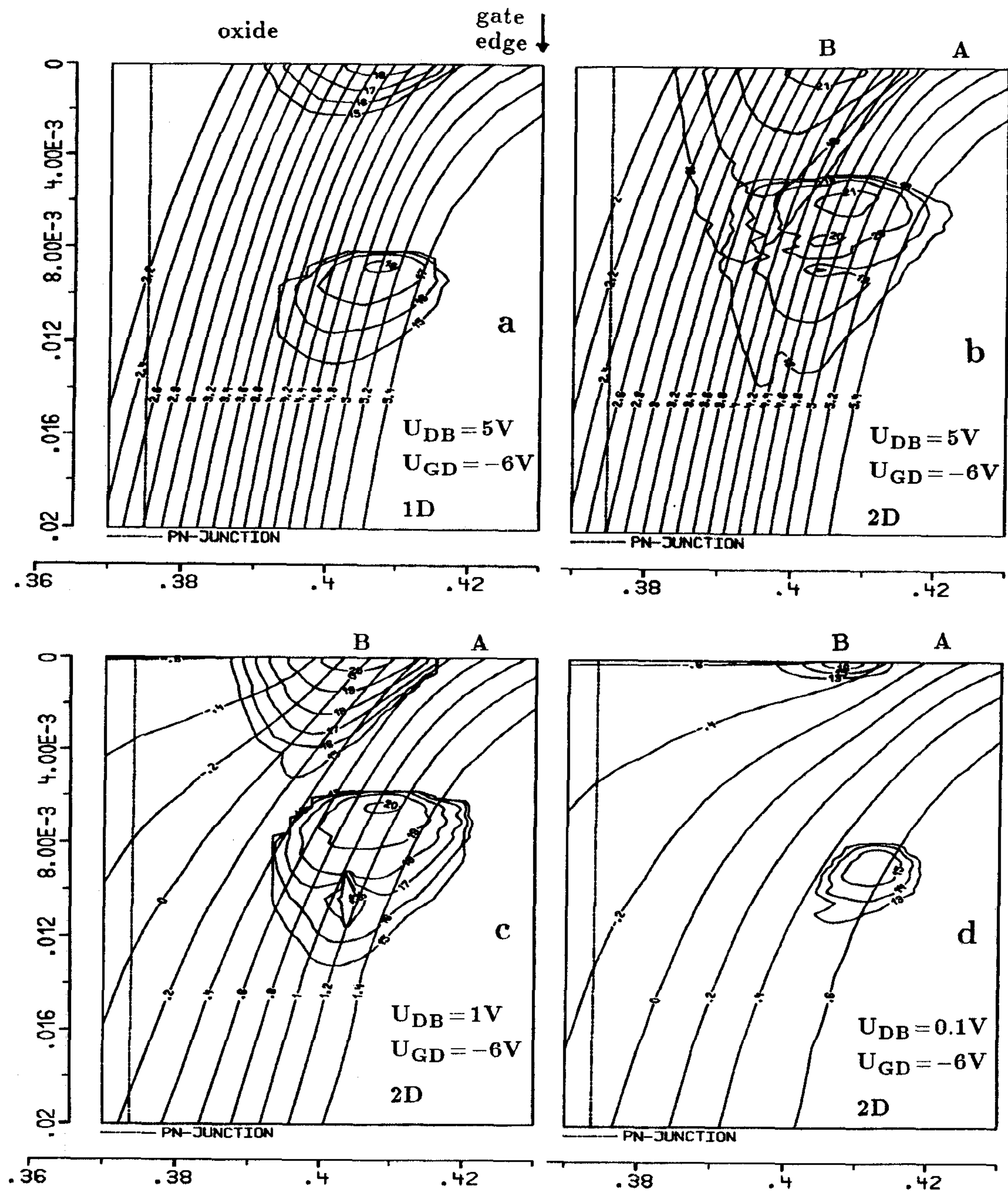

Figure 1. Distribution of the electron and hole tunneling generation rate in a MOSFET, calculated using 1D and 2D approach. Phonon-assisted tunneling model with an average field $\overline{E}$ is used. Equipotential lines are shown. Note that the tunneling paths look like to be not orthogonal to the equipotential lines due to different x and y axes scale. Device data: $L_G = 0.43\mu m$, $t_{ox} = 8nm$, $T = 300K$. Axes are given in μm.

3 Calculation of the tunneling generation rate

We consider direct tunneling processes. This is negligible in Si-devices in practice, but the results are qualitatively applicable to the dominant phonon-assisted tunneling. To model the band-to-band tunneling in MOSFET, either some result from junction-tunneling theory ([9],[13]), or for internal field emission ([7],[8],[9],[12]), have been used in the literature. Being aware of the expected effect of the vicinity of the oxide/bulk interface on the tunneling, we have applied the theory for internal field emission in an infinite medium. To derive an expression for N in a variable field, two approaches have been investigated: the derivation of the interband matrix element M_{cv} in a linearly variable field, and using the WKBJ method with an appropriate barrier for the variable field. Here we will present the second approach, where we follow the quasi-classical picture in the interpretation of an internal field emission result ([12]):

$$N = \frac{eE}{4\pi^3\hbar} \cdot \int\int P_{v\to c}(E, k_y, k_z)\; dk_y\; dk_z \tag{1}$$

where $P_{v\to c} = |\kappa_e/eE \cdot M_{cv}|^2$ is the probability of the penetration of an incident electron, with a normal impulse $\vec{k}_\perp$, from the valence to the conduction band. In this interpretation, N is the product of the band-oscillation frequency, concentration of the valence-band electrons and an average penetration probability per particle. We assumed $P_{v\to c}(\vec{k}_\perp) = \pi^2/9 \cdot T(\vec{k}_\perp)$ [13], where T is the transition probability per particle calculated by the WKBJ method. In applying the WKBJ approximation we considered two barriers: an empirical triangular barrier generalized for the case of a non-constant field, and a more realistic parabolic barrier in a linearly variable field. The parabolic barrier has been derived considering the kinetic energy ε from the two-band $\mathbf{k}\cdot\mathbf{p}$ model [7], which yields after an approximation the expression for the barrier: $\varepsilon(\mathrm{E}_g - \varepsilon)/\mathrm{E}_g$, where ε is calculated from the potential energy due to the external electric field $E(x)$, with respect to the incident energy. Assuming a field $E(x') = E_s - \beta x'$ where x' is the axis along the tunneling path, the corresponding barrier potential becomes:

$$V_b(x) = \frac{\mathrm{E}_g}{4} - \frac{(eE_o x - e\beta x^2/2)^2}{\mathrm{E}_g}, \quad \text{with} \quad E_o = \sqrt{\frac{E_s^2 + E_e^2}{2}} \quad \text{and} \quad \beta = \frac{E_s^2 - E_e^2}{2\mathrm{E}_g/e} \tag{2}$$

For a constant field, $\beta = 0$ holds, and $V_b(x)$ reduces to the symmetric barrier (Fig.3) which reproduces by WKBJ exactly the result of Kane for the transition probability [7], including the dependence on $\vec{k}_\perp$ (with $\pi^2/9$ added). The electrical fields at the classical turning points given by (2) are E_s and E_e, respectively. After integrating over $\vec{k}_\perp$, assuming several approximations, the number of the generated pairs in a linearly variable field follows

$$N_{vf} = \frac{1}{18\pi} \frac{e^2\sqrt{m_r}}{\hbar^2 \mathrm{E}_g^{1/2}} \frac{E_s E_o}{(1 + 3a_0^2/16)} \exp\left(-\frac{\pi}{2}\sqrt{m_r}\frac{\mathrm{E}_g^{3/2}}{\hbar e E_o}(1+3a_o^2/32)\right),\; a_o = \frac{E_s^2 - E_e^2}{E_s^2 + E_e^2} \tag{3}$$

The factor $1 + 3a_o^2/32$ are the first two terms in a series expansion. In the derivation of (3) a field variation E_s/E_e less then approximately four is assumed. This expression reduces to the Kane expression N_K for a constant field. We can conclude that the mean square

value of the field is responsible for tunneling in a linearly variable field. If the field variation E_e/E_s is less then about four, E_o becomes close to the average field along the tunneling path $\overline{E} = \mathrm{E_g}/l_t = (E_s + E_e)/2$, giving $N_{vf} \approx N(\overline{E})$. This is demonstrated using an analytically modeled depletion region (Fig.2) and by numerical simulation of a MOSFET in Fig.4. Note that we have examined the criticism related to the Kane expression presented in the resent papers [4],[8]. We found, however, that N_K is an accurate solution to the problem. These results will be published elsewhere. Based on the considerations presented, we expect that the constant-field expressions for the phonon-assisted tunneling, where an average field is used, could be a good approximation for the the corresponding tunneling in a variable field.

Acknowledgements

Our work has been considerably supported by: Siemens Corporation Munich, FRG; Sony Corporation Atsugi, Japan; and Digital Equipment Corporation Hudson, MA, USA.

References

[1] Takashi Hori, "Drain-Structure Design for Reduced Band-to-Band and Band-to-Defect Tunneling Leakage", *Proceedings Symposium on VLSI Technology (Honolulu, Hawaii),* June, 1990, pp. 69-70.

[2] I. C. Chen, C. Kaya, and J. Paterson, "Band-to-Band Tunneling Induced Substrate Hot-Electron (BBISHE) Injection: A New Programming Mechanism for Nonvolatile memory Devices", in *IEDM Tech. Dig. (Washington, D.C.),* December, 1989, pp. 263-266.

[3] W. Bergner and R. Kircher, "Modeling of Band-to-Band Tunneling Mechanisms", *Proceedings 22nd Int. Conf. on Solid-State Devices and Materials (Sendai Japan),* , August, 1990, pp. 135-137.

[4] M. Takayanagi, S. Iwabuchi, "Theory of Band-to-Band Tunneling Under Nonuniform Electric Fields for Subbreakdown Leakage Currents", *IEEE Trans. Electron Devices,* vol. 38, pp. 1425-1431, June 1991.

[5] T. Endoh, R. Shirota, M. Momodomi, F. Masuoka, "An Accurate Model of Subbreakdown Due to Band-to-Band Tunneling and Some Applications", *IEEE Trans. Electron Devices,* vol. 37, pp. 290-296, January 1990.

[6] Y. Odake, K. Kurimoto and S. Odanaka, "Three-Dimensional Numerical Modeling of the Indirect Band-to-Band Tunneling in MOSFET's", *Proceedings 22nd Int. Conf. on Solid-State Devices and Materials (Sendai Japan),* , August, 1990, pp. 131-134.

[7] E. O. Kane, "Zener tunneling in semiconductors", *Journal Phys. Chem. Solids,* vol. 12, pp. 181-188, 1959.

[8] P. K. Chakraborty, J. C. Biswas, "Theory of interband tunneling in semiconductors", *Solid-State Electronics,* vol. 28, no. 5, pp. 493-497, 1985.

[9] Ralph T. Shuey, "Theory of Tunneling Across Semiconductor Junctions", *Physical Review,* vol. 137, pp. A 1268-1277, February 1965.

[10] M. Orlowski, S. W. Sun, P. Blakey and R. Subrahmanyan, "The Combined Effects of Band-to-Band Tunneling and Impact Ionization in the Off Regime of an LDD MOSFET", *IEEE Electron Device Letters,* vol. 11, pp. 593-595, December 1990.

[11] I. Nedev, A. Asenov and E. Stefanov, "Experimental study and modeling of band-to-band tunneling leakage current in thin-oxide MOSFETs", *Solid-State Electronics,* vol. 34, no. 12, pp. 1401-1408, 1991.

[12] P. N. Argyres, "Theory of Tunneling and its Dependence on Longitudinal Magnetic Field", *Physical Review,* vol. 126, pp. 1386-1393, May 1962.

[13] E. O. Kane, "Theory of Tunneling", *Journal of Applied Physics,* vol. 32, pp. 83-91, January 1961.

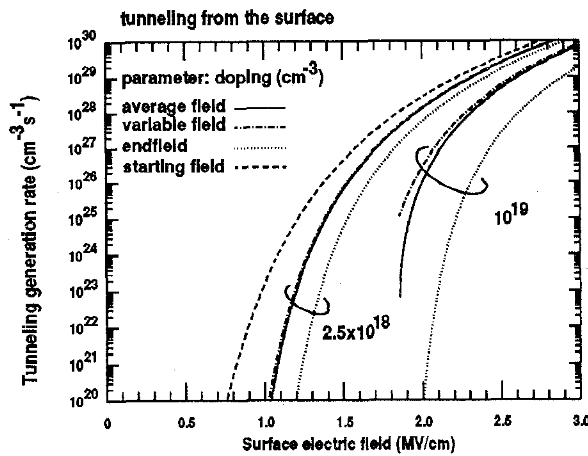

Figure 2. Tunneling rate calculated with $N_K((E_s + E_e)/2)$, $N_{vf}(E_s, E_e)$, $N_K(E_e)$ and $N_K(E_s)$. N_k is the Kane expression and N_{vf} is given with (3). An one-dimensional analytical model for the totally depleted uniformly doped bulk is used.

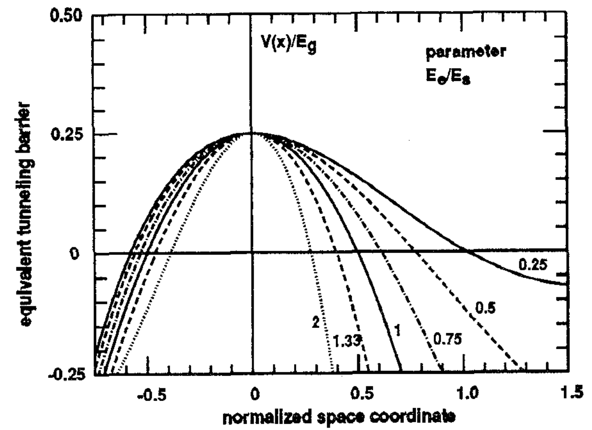

Figure 3. Normalized potential barrier in the WKBJ approach for a linearly variable field: expression (2).

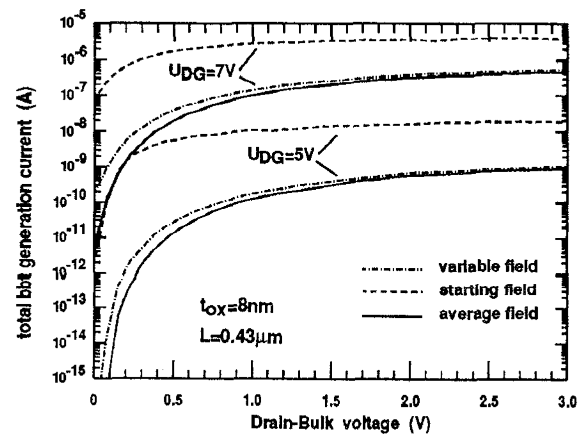

Figure 4. MINIMOS numerical simulation: 2D approach; direct tunneling with indirect Si band gap assumed. SUPREM profile is used; junctions are abrupt (As). The starting-field curve is flat, while the variable field curve shows a typical roll-off observed experimentally by reducing drain bias.