TWO-DIMENSIONAL DOPANT PROFILING OF SUBMICRON MOSFET'S USING NONLINEAR LEAST SQUARES INVERSE MODELING

N. Khalil, J. Faricelli, and C.L. Huang
Digital Semiconductor, ULSI Operation Group,
77 Reed Rd, Hudson, MA 01749
and
S. Selberherr
Institute for Microelectronics, Technical University of Vienna
Gußhausstraße 25-29/E360, A-1040 Vienna, Austria

We present an inverse modeling technique to determine the two-dimensional (2D) doping profile of a MOSFET from electrical measurements. In our method, the profile is formulated using two tensor product splines (TPS). This analytical representation is general, compact and flexible. It simplifies the profile determination problem to the extraction of the TPS coefficients from experimental data. We show the results of applying the new technique on data collected from a sub $0.5\mu m$ CMOS technology with various source/drain implants. We also compare the measured and simulated I–V and C–V characteristics. The results illustrate the importance of accurate 2D doping profiles for short channel device simulation and modeling.

I. INTRODUCTION

With the scaling of MOSFET's dimensions into the submicron regime, the influence of the distribution of dopants on short channel device characteristics increases dramatically. The complex multi-dimensional fields created by the doping profile become one of the most important factors in determining the electrical behavior of MOSFET's. One-dimensional (1D) profiling tools such as spreading resistance (SRP) and secondary ion mass spectrometry (SIMS) are available. However, due to the shallow vertical and lateral junctions, the proximity effects, and the interaction between dopants species, 1D profiles are less indicative of actual 2D profiles. Attempts to extend the 1D profiling tools to higher dimensions (e.g. 2D SRP, 2D SIMS) have met with limited success when applied to state-of-the-art CMOS technology (1-3). Newer techniques aimed at addressing these shortcomings are under development.

Scientists in other fields, such as in geophysics, facing a similar lack of direct experimental measurements resort to inverse modeling (4,5). Inverse techniques deal with the determination of the parameters values of a physical system from experimental measurements, along with the physical laws and theories that relate the inputs of the system to its outputs. For semiconductors, the theoretical relationship takes the form of the basic semiconductors equations, namely Poisson's equation and the current continuity equations (6). In the case of thermal equilibrium and negligible current flow, the solution of the continuity equations can be ignored. Hence, the space charge density within a device can be calculated by solving Poisson's equation:

$$\Delta^2\psi = -\frac{q}{\epsilon}(n - p + N_D^+ - N_A^-) \tag{1}$$

where: q is the elementary charge, ϵ the semiconductor permitivity, ψ the electrostatic potential, n, p the electron and hole concentrations, N_D^+, N_A^- the donor and acceptor concentrations.

The charges associated with the device terminals are calculated by integrating the space charge density over a device region $\mathcal{A}$:

$$Q_i = \int_{\mathcal{A}} (n - p + N_D^+ - N_A^-) d\mathcal{A} \tag{2}$$

or by applying Gauss's law to calculate the gate charge by evaluating the line integral of the gate electric field on a closed loop surrounding the gate:

$$Q_{gate} = \oint (\vec{E}.\hat{n})dl \tag{3}$$

The device capacitances are then approximated by numerically differentiating the terminal charges:

$$C_{ij} = \frac{\delta Q_i}{\delta V_j} \approx \frac{\Delta Q_i}{\Delta V_j} \tag{4}$$

As an inverse problem, the profile extraction consists of finding the doping profile that minimizes the weighted least squares fit criterion (SSQ) between experimental and simulated capacitance values:

$$SSQ = \sum w_i \times (C_i^{experimental} - C_i^{simulated})^2 \tag{5}$$

where the w_i's are the weights associated with each data point. This is a continuous minimization problem. The target is to determine the complete functional variation of the profile. We convert it to a discrete problem by representing the profile variation using B-splines in 1D and tensor product spline (TPS) in 2D. With a fixed sequence of breakpoints, i.e. knots, the inverse problem simplifies the profile extraction to the determination of the B-spline or TPS coefficients from the capacitance data.

The rest of the paper is divided as follows: The extraction procedures and related topics are presented in the next section. In Section III, the extracted 2D doping profiles of a sub-half micron CMOS process that has various source/drain implants are shown. Section III also shows a comparison of experimental I-V and C-V characteristics with simulations using the extracted doping profiles. We discuss the limitations and accuracy of our method in Section IV and present a preliminary assessment of its resolution. Finally, in the conclusion, we offer a list of open questions for future work.

II. EXTRACTION PROCEDURE

In (7,8) a technique for the determination of 2D MOSFET doping profiles from gate and source/drain capacitance measurements is presented. A brief overview of the method for a P-channel MOSFET is given as follows:

- Input Parameters Determination: Several important parameter values are obtained by independent experimental means. The oxide thickness (t_{ox}) is determined by high frequency C–V measurements. The polysilicon gate length (L_p) and polysilicon gate concentration (N_p) are extracted by matching experimental gate-to-channel capacitances (C_{gc}) and simulated results that take into account the polysilicon depletion effect (9,14,15). The S/D diode acceptor profile is determined using SIMS measurements.
- Starting 2D profile generation: The SIMS S/D acceptor profile, the S/D donor profile extracted from diode capacitances, and the channel profile extracted from deep depletion capacitance data are combined into an initial 2D profile. We use a subdiffusion factor for rotating the S/D acceptor profile to obtain the measured effective electrical channel length for the device.
- 2D extraction: Using the gate overlap and S/D diode capacitance measurements taken on a fingered polysilicon structure over active region with varying bias voltages, the TPS coefficients are adjusted to achieve a good fit between simulated and experimental values.

In the rest of this section, some important aspects of our implementation are presented.

Profile TPS Representation: As stated previously, the determination of the doping profile is converted into a discrete parameter extraction problem using the TPS representation. A TPS is a generalization of

the one-dimensional polynomial piecewise B-spline functions to a multi-dimensional space (10). It is defined by the two knot sequences t_x and t_y and the values of the coefficients c_{ij}:

$$f(x,y) = \sum_{i=1}^{nx} \sum_{j=1}^{ny} c_{ij} * B_{i,k_x,t_x}(x) * B_{j,k_y,t_y}(y) \tag{6}$$

where $B_{i,k,t}$ is the ith B-spline of order k for the knot sequence t, nx and ny are the number of knots in the X and Y direction respectively. Using a fixed number of knots and locations, a 2D function can then be written as:

$$f(x,y) = \sum_{i=1}^{nx*ny} c_i * B_{i,xy} \tag{7}$$

In view of the wide variation range of the doping concentrations, the logarithmic dependency of the acceptor and donor concentrations are represented by two TPS. Each TPS uses a different sequence of knots to accommodate the varying concentration fields. The net doping can then be written as:

$$\rho(x,y) = \mathcal{F}(\vec{\alpha}, \vec{\delta}) \tag{8}$$

where $\vec{\alpha}$ and $\vec{\delta}$ are the vectors of TPS coefficients for the acceptors and donors respectively. This analytical formulation of the profile is more compact than a straightforward mesh representation. As a result, the amount of computation required is decreased, and a better condition for the least squares problem is achieved. Moreover, this representation does not assume *a priori* knowledge of the profile functional variation. To ensure the smoothness and continuity of the profile, quadratic or cubic splines are used to represent the profile functional variation in the horizontal and vertical directions. Choosing the appropriate number of knots in each direction is important. Whereas a large number of parameters will result in a better fit to the data, this does not ensure accurate determination of the profile as the variance errors and the computation time are increased. We are presently investigating ideas similar to (11) for the determination of the number and location of knots. We typically use 4 or 5 knots in each direction and determine their location based on process information such as junction depth, spacer width, and gate length. Another feature of the TPS representation is the ease of including 1D information. For example, in generating the 2D starting profile, the coefficients of the 1D long channel profile B-spline become the TPS coefficients in the middle of the short channel device.

Capacitance Calculation: The charge integration method is used to calculate the device capacitances by numerical differentiation. This procedure is inherently prone to numerical roundoff and integration errors. We resort to the following strategies to minimize their effects:

- We use central differences for numerical differentiation:

$$C_i \approx \frac{Q_2 - Q_1}{V_2 - V_1} \tag{9}$$

  where Q_2 and Q_1 are the terminal charges at V_2 and V_1 respectively, $V_2 = V + \delta V$, and $V_1 = V - \delta V$. The size of the voltage step, δV, should be small enough to achieve a good linearized approximation of the derivative. However a small change in the voltage yields a small change in the charge. This could result in numerical roundoff errors in calculating the quotient due to subtractive cancellation. We found that a 20mV step is a reasonable compromise between the two conflicting requirements.

- Discretization is a major source of errors in the calculations. In principle one could increase the size of the grid to effectively eliminate numerical approximation errors. Whereas this is convenient and feasible for 1D problems, the computational demands in the 2D case prohibit such an approach. For that reason, we carefully refine the Poisson solver grid where the space charge density varies rapidly.

- We use the same grid at V_1 and V_2 in our solution. This reduces the integration errors since part of the errors in the calculation of the charges is canceled out by taking the difference.

- When applicable, we only integrate the charges in the device that contribute to the variation. This minimizes the loss of significant digits in the capacitance calculation. For example, in calculating the bulk charge, the ionized dopant atoms in the S/D diode region are not included in the summation because they remain constant when a small voltage perturbation is applied.

Least Squares Optimization: We solve the nonlinear least squares problem using the well established Levenberg-Marquardt algorithm with linear constraints (11,12). This is a Gauss-Newton algorithm with trust region modification that is versatile and robust. We use finite difference derivatives to approximate the Jacobian. In the following, we present some of the specific characteristics of our solution:

- Starting Guess: A good initial guess is important to avoid the trapping of the solution in local minimas and to limit the number of iterations required before convergence. This is especially critical in 2D where the computational demands are great.

- Parameter Redundancy: Poisson's equation contains a net doping term only. In the presence of PN junctions, this introduces a direct correlation between the acceptor and donor coefficients especially for knots located near the junction. This could result in a rank deficient Hessian. We bypass this problem by a two step iteration scheme in which we fix one set of coefficients while allowing the coefficients of the other type to change.

- Restriction of the Number of Parameters: We only extract the coefficients at knots when they are of the same type as the net doping at that location. Moreover, at each step, we analyze the eigenvalues of the approximate Hessian. Coefficients associated with small eigenvalues are not allowed to change.

- Parameters Bounds: To avoid deviations outside the expected range during the initial stages of the solution, we enforce linear bounds on the parameters.

III. RESULTS

We applied our method to data collected from devices fabricated using a retrograde n-well, salicided dual-gate CMOS process. All experimental I–V and C–V characteristics were obtained using an HP4145B parameter analyzer and an HP 4275A LCR meter. The measurement frequency of the HP 4275A LCR meter was set to 100 kHz. The resolution of the system is around 0.1 fF. In order to reduce the noise level in the measured results, the experimental C–V data for sub 0.5μm devices were averaged for several measurements. As a result, the actual resolution of experimental data is better than 0.1 fF.

Using deep depletion capacitance data taken on a large MOS capacitor we first extract the B-spline coefficients of the vertical channel doping. Figure 1 shows the extracted profile together with an analytically extracted profile. In contrast to the analytical results that fail to determine the doping near the surface, the inverse modeling profile is extracted up to the Si/SiO_2 interface. We note however that there is a correlation between the gate workfunction, oxide charge density Q_{ox}, and the doping near the interface.

For the S/D diode, the acceptor B-spline coefficients were determined by curve fitting the 1D SIMS profile. The SIMS profile was also used in the simulation of the reverse junction capacitance data. The B-splines coefficients of the S/D donors profiles were extracted by matching the experimental diode area capacitance. Figure 2 shows the resultant S/D net doping profile as well as the SIMS acceptor profile.

The channel and S/D profiles were then combined into an initial 2D profile. By carefully matching the experimental capacitances, we ensure that the initial profile is in the immediate proximity of the solution.

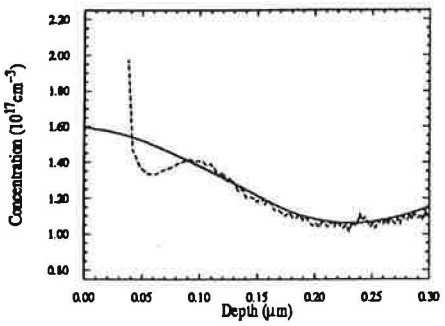

Fig. 1. MOS Capacitor Channel Doping Extracted from Deep Depletion C–V Data: Inverse Modeling (solid line) and Analytical results (dashed line).

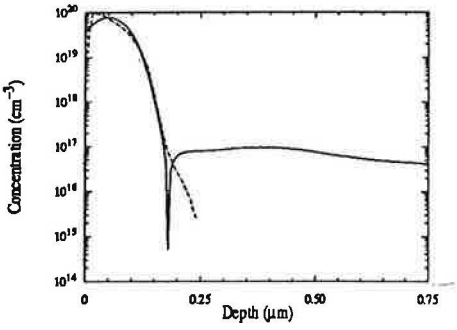

Fig. 2. Source/Drain Doping profiles: SIMS Acceptors Profile (dashed line) and Net Doping Extracted By Inverse Modeling from Area Diode Capacitance.

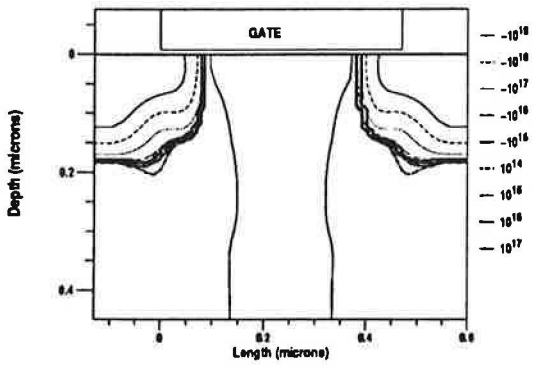

Fig. 3. P-channel Extracted 2D Net Doping for a device with L_p=0.45 μm

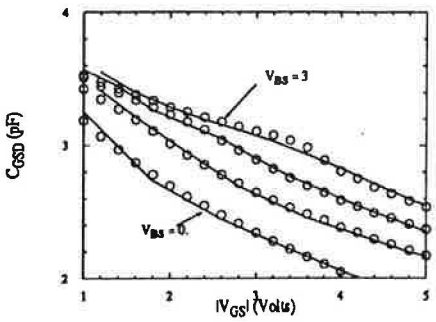

Fig. 4. Comparison of experimental (symbols) and simulated (solid lines) gate-to-source-drain capacitance (W/L = 3696/0.5 μm)

The extracted 2D doping profile for the P-channel MOSFET is shown in Fig. 3. Figure 4 illustrates the good fit achieved between the experimental data used in the extraction and simulation with the extracted 2D profile.

For validation, we compare simulated and experimental I–V and C–V characteristics using the extracted profile as input. Since the 2D profile, polysilicon gate concentration (N_p), oxide thickness (t_{ox}) and polysilicon gate length (L_p), are obtained experimentally, 2D device simulation is expected to accurately reproduce experimental I–V and C–V characteristics over a wide range of biases and devices. To ensure a good agreement we adjust the mobility parameters using long-channel device data, as well as the source/drain external resistance R_t, and the gate workfunction of the short-channel devices. Figure 5 shows measured and simulated results for the linear region currents for 3 gate lengths. Reasonable good agreement is found in all regions of bias for all three lengths. Figure 6 shows a comparison of measured and simulated results for the device with L_p = 0.45 μm. Note that a good agreement was also obtained for longer devices (not shown).

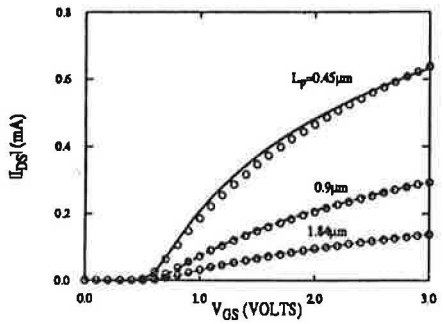

Fig. 5. Comparison of measured (symbols) and simulated (lines) I–V characteristics in the linear region ($V_{DS} = -50$ mV) for three gate lengths ($L_p = 0.45$, 0.9 and 1.84 μm) with $W = 64$ μm, $N_p = 2.7 \times 10^{19}$ cm^{-3} and $t_{ox} = 72.7$ Å.

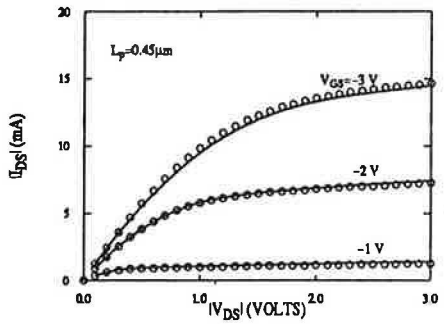

Fig. 6. Comparison of measured (symbols) and simulated (lines) I–V results in the linear and saturation regions for devices with $L_p = 0.45$ μm, $W = 64$ μm, $N_p = 2.7 \times 10^{19}$ cm^{-3}, and $t_{ox} = 72.7$Å.

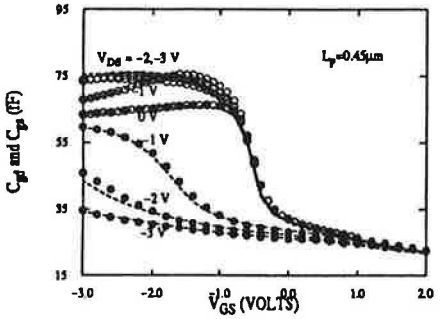

Fig. 7. Comparison of measured (symbols) C_{gd} (lower 3 curves) and C_{gs} (upper 3 curves) capacitances with simulated results (lines) as a function of V_{GS} and V_{DS} for device with $L_p = 0.45$ μm, $W = 64$ μm, $N_p = 2.7 \times 10^{19}$ cm^{-3}, and $t_{ox} = 72.7$ Å.

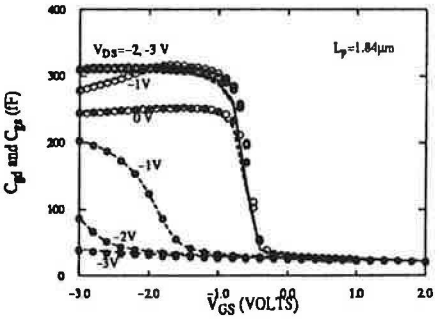

Fig. 8. Comparison of measured (symbols) C_{gd} (lower 3 curves) and C_{gs} (upper 3 curves) capacitances with simulated results (lines) as a function of V_{GS} and V_{DS} for device with $L_p = 1.84$ μm, $W = 64$ μm, $N_p = 2.7 \times 10^{19}$ cm^{-3}, and $t_{ox} = 72.7$ Å.

We also compare the gate-to-source (C_{gs}), and gate-to-drain (C_{gd}) capacitances for two lengths of P-channel MOSFET's in all regions of device operation (the accumulation, linear and saturation regions) in Fig. 7 and 8. Excellent agreement is achieved for both bias-dependent intrinsic and overlap capacitances.

Finally, we applied our method to 3 N-channel devices with varying implant conditions as shown in Table I. In Fig. 9, we show a schematic diagram of half of an N-channel MOSFET with the n- implant angle. The implant conditions for devices A and B corresponds to Large-Angle-Tilt-Implanted-Drain (LATID) implant (13), while device C received a regular Lightly-Doped Drain implant. Figure 10 shows the extracted net doping at the SiO_2/Si interface for the three devices. It is clear that the method is able to resolve the difference in the lateral diffusion length as the implant angle/energy increases.

Device Type	A	B	C
Processing conditions			
n^- S/D (ions/cm^2, keV, tilt angle)	2.4×10^{13}, 40, 38^0	3.2×10^{13}, 40, 38^0	3×10^{13}, 25, 7^0
n^+ S/D (ions/cm^2, keV)	5×10^{15}, 40	5×10^{15}, 40	5×10^{15}, 40

Table I: Processing conditions for N-channel MOSFET's

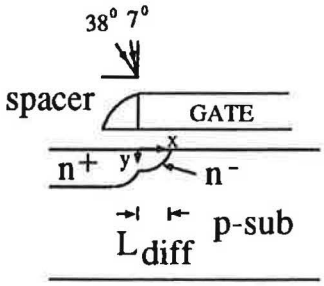

Fig. 9. A schematic diagram of half an N-channel MOSFET with an n- implant angle.

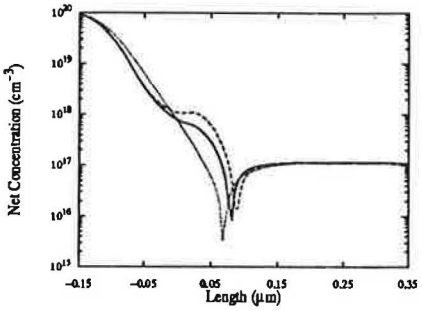

Fig. 10. Extracted one-sided net doping at the SiO_2/Si interface (device A - solid line, device B - dashed line, device C - dotted line).

IV. DISCUSSION

The following points illustrate some of our method characteristics, limitations, and associated uncertainties.

- The method capability is limited to the extraction of the electrically active net doping, not the chemical concentration of species atoms. Moreover, the required electrical measurements can only be taken after the completion of processing up to the metal layers. This limits the applicability of the method during process development.

- Uncertainties in the device geometrical structure, such as nonplanar surfaces, have a direct effect on the extracted profile. These can be resolved by independent determination of the structural information using TEM imaging. Other types of input uncertainties, e.g. errors in the S/D SIMS profile, can also influence the extraction. We plan on performing a Monte Carlo simulation analysis to estimate the accuracy of the extracted profiles subject to the inherent errors in the inputs.

- The confidence intervals of the extracted TPS coefficients were computed according to the algorithms described in (17) assuming error-free inputs. The size of the confidence interval is an estimate of the standard deviation of the profile. Figure 11 shows the extracted surface net doping for an N-channel device with a 95% confidence region. It indicates that the profile extraction is reasonably accurate in determining the doping level and the lateral junction location.

- The extent of the device region where the profile can be determined depends on the range of measurement voltage, the doping level, and the device characteristics. For example, the capability of the gate to deplete the S/D-gate overlap region of carriers under accumulation bias, without breaking down the gate oxide, limits the resolution of the method for high concentration S/D profiles.

- Modeling assumptions in solving Poisson's equations are also a source of uncertainty. For instance, there are two main approaches available in the literature to account for the anomalous threshold voltage variation in short-channel devices. The Reverse Short Channel Effect (RSCE) can be explained by either the existence of oxide charges or doping profile variation (18,19). The chosen method to model this phenomenon influence the extracted profile near the interface. Figure 12 compares the net doping along the surface in two cases: In the first one, oxide Gaussian charges were introduced in the simulation to model the RSCE. In the second, the oxide charge value was fixed to that determined from long channel device data. As seen, the two profiles are clearly different. We are presently investigating the inclusion of other types of data to eliminate this uncertainty. In particular, the use of subthreshold current data in the extraction might clarify the issue.

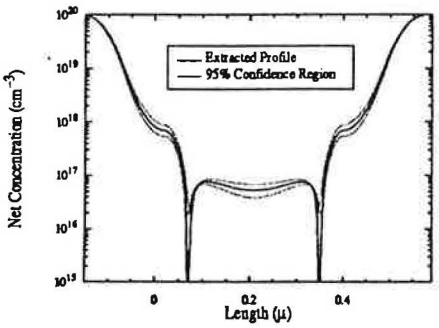

Fig. 11. Extracted N-channel net doping along the SiO_2/Si interface and the correspoding 95% confidence region.

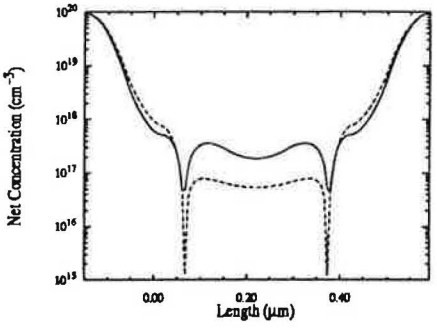

Fig. 12. N-channel net doping at the SiO_2/Si interface extracted with (solid line) and without (dashed line) Gaussian charges to model RSCE.

V. CONCLUSION

We have described a nondestructive method for MOSFET 2D profile determination. The method does not require any difficult sample preparation and uses measurements easily performed during process characterization. It fills an important gap in the available process metrology tools. The preliminary results are very encouraging, however, further development of the technique is still needed. In particular, the following areas deserve more attention:

- Development of an algorithm for selecting the appropriate number of TPS knots and their locations.
- A quantitative study of the resolution and accuracy of the method is still needed.
- A comparison between the profiles extracted using the inverse modeling technique and profiles determined by direct measurements, when possible, could serve as a cross validation check.
- Finally, the extension of the method to include other sources of electrical data, such as subthreshold current measurements, might improve the resolution beyond what is possible when relying uniquely on capacitance data.

References

(1) S. H. Goodwin-Johansson, R. Subrahmanyan , C. E. Floyd and H. Z. Massoud, "Two-Dimensional Impurity Profiling with Emission Computed Tomography Techniques", *IEEE Trans. Computer-Aided Design*, CAD-8(4),1989.

(2) R. Subrahmanyan, H. Z. Massoud and R. B. Fair, "Experimental Characterization of Two-Dimensional Dopant Profiles in Silicon using Chemical Staining", *Applied Physics Letters*, 52(25), 1988.

(3) S. Kordic, E. Van Leonen, D. Dijkkamp, A. Hoeven and H. Moraal, "Scanning Tunneling Microscopy on Cleaved Silicon PN Junctions", *IEDM Tech. Digest*, 1989, pp.277-280.

(4) A. Tarantola, *Inverse Problem Theory*, Elsevier, Amsterdam, 1987.

(5) G.J.L. Ouwerling, "Nondestructive One- and Two-Dimensional Doping Profiling by Inverse Methods", Ph.D. Dissertation, 1989, The Delft University of Technology, Netherlands.

(6) S. Selberherr, *Analysis and Simulation of Semiconductor Devices*, Springer-Verlag, Wien-New York, 1984.

(7) N. Khalil, J. Faricelli, D. Bell and S. Selberherr, "The Extraction of Two-Dimensional MOS Transistor Doping via Inverse Modeling", *IEEE Electron Device Letters*, EDL-16, No. 1, p. 17, Jan. 1995.

(8) N. Khalil, C.L. Huang, J. Faricelli, and Siegfried Selberherr, "Measurements and Simulations of Short-Channel MOSFET I-V and C-V characteristics", In preparation for publication in *IEEE Trans. Electron Devices.*

(9) P. Habaŝ and J. Faricelli, "Investigation of the physical modeling of the gate-depletion effect", *IEEE Trans. Electron Devices*, ED-39, p. 1496, 1992.

(10) Carl De Boor, *A Practical Guide to Splines* , Springer-Verlag New York Inc., 1978.

(11) A.T. Watson, P.C. Richmond, P.D. Krieg, and T.M. Tao, "A Regression-Based Method for Estimating Relative Permeabilities from Displacement Experiments", *SPE Reservoir Engineering*, p. 953, Aug. 1988.

(12) K. Levenberg, "A Method for the Solution of Certain Nonlinear Problems in Least Squares", *Quarterly of Applied Mathematics*, Vol.2, pp. 164-168,1944.

(13) D. W. Marquardt, " An Algorithm for Least Squares Estimation of Nonlinear Parameters", *Journal of the Society of Industrial Math and Applied Mathematics (SIAM)*, Vol.11, pp. 431-441, 1963.

(14) SEDAN III - *A Generalized Electronic Material Device Analysis Program*, Stanford University, July 1985.

(15) S. Selberherr, A. Schütz and H. W. Pötzl, "MINIMOS — A Two-Dimensional MOS Transistor Analyzer", *IEEE Trans. on Electron Devices*, ED-27 (8), August 1980.

(16) T. Hori, "$1/4-\mu m$ LATID (Large-Tilt-Angle Implanted Drain) Technology for 3.3-V Operation", *IEDM Techn. Digest*, 1989, pp.777-780.

(17) M. Sharma and N.D. Arora, "OPTIMA: A Nonlinear Model Parameter Extraction Program with Statistical Confidence Region Algorithms," *IEEE Trans. Computer-Aided Design*, CAD-12, pp. 982-986, (1993).

(18) H. Jacobs, A.V. Schwerin, D. Scharfetter, and F. Lau, "Mosfet Reverse Short Channel Effect Due to Silicon Interstitial Capture in Gate Oxide", in *IEDM Tech. Digest*, 1993, pp. 307-310.

(19) C.S. Rafferty, H.H. Vuong, S.A. Eshraghi, M.D. Giles, M.R. Pinto, and S.J. Hillenius, "Explanation of Reverse Short Channel Effect by Defect Gradient", in *IEDM Tech. Digest*, 1993, pp.311-314.