P5

Inverse Modeling of sub-100nm MOSFET with PDE-Constrained Optimization

Chen Shen and Ding Gong
Cogenda Pte Ltd, Singapore. Email: shenchen@cogenda.com

I. Introduction

The inverse modeling of MOSFET aims to extract the process and device parameters of a CMOS technology from electrical test data, such as the I-V curves. Unlike the parameter extraction for compact models, inverse modeling calculates the electrical characteristics with TCAD simulation instead of the analytical formulae of compact models (e.g. BSIM4). The parameters extracted from inverse modeling can be either the process parameters (e.g. Dose, energy, annealing time, etc.), or the device parameters (oxide thickness, peak doping concentration, char. length of Gaussian doping, etc.). Obviously, inverse modeling is an optimization problem to minimize the error between the simulated and the measured electrical characteristics.

The inverse modeling problem has been attempted by several authors with various optimization algorithms [1]–[3]. For optimization algorithms to be efficient, one needs to estimate the gradient of the object function to a reasonable accuracy. However, tradition TCAD simulation is not able to calculate derivatives, e.g. the partial derivative of MOSFET drain current w.r.t. substrate doping concentration ($\partial I_d/\partial N_{\text{sub}}$). As a result, these work either employed forward-difference to estimate the partial derivatives [1], which incurs a large number of extra simulation runs; or use optimization algorithms that does not require partial derivatives [2], [3]. Since TCAD simulations are very expensive, and these inverse modeling tasks are very time consuming, and often limited to optimizing one device dimension.

Given the high cost, one desires to exploit as much information as possible from each simulation. In this work, an inverse modeling approach that efficiently calculates the partial derivatives is outlined, and extraction result on a 65 nm CMOS technology is presented. In particular, instead of fitting to I-V curves of one transistor, this work attempts to fit to all device sizes, from the shortest gate length to the long-channel ones.

II. Quasi-2D PDE for MOSFETs

Since a large number of TCAD simulation of MOSFETs are expected, a quasi-2D set of equations are used to reduce computation time. The device is discretized to a rectangular mesh grid. The state variables consists of the electrostatic potential at each grid node $V(x_i, y_j)$, and the electron quasi-fermi-potential $V_{fn}(x_i)$ is assumed to vary only in the x-direction. 2D Poisson's equation is coupled with the 1D electron-current continuity equation along the surface of the MOSFET. The inversion charge density is integrated as $Q_i(x) = \int n(x, y)\mathrm{d}y$, and the drift-diffusion drain current by integrating the inversion charge under the charge-sheet approximation.

The drain current calculated with the quasi-2D approach was check against the full 2D TCAD simulation, and the agreement is within 1%.

III. Automatic Differentiation and PDE-Constrained Optimization

The Nested Analysis and Design approach (NAND) of PDE-constrained optimization is used [5]. The MOSFET is described by set of design parameters q, as will be detail in the next section. The state variables w consist of electrostatic potential and electron quasi-fermi potential. Given the parameters, one solves the quasi-2D device equations

$$F\left(w\left(q\right), q\right) = 0, \tag{1}$$

to obtain the state variables w, as described in the previous section. One thus has an implicitly defined function $w(q)$.

Terminal currents can then be calculated from the state variables, and we aims to minimize the difference between the simulated I-V curve and the measured data. This leads to a least square

978-1-61284-418-3/11/$26.00 © 2011 IEEE

minimization problem with the reduced objective function $\hat{f}(q) = f(w(q), q)$. For efficient optimization, one needs the derivatives $\partial w/\partial q$.

One can calculate the derivatives of state variables w.r.t. the parameters by

$$\frac{\partial w}{\partial q} = -\left(\frac{\partial F}{\partial w}\right)^{-1} \frac{\partial F}{\partial q}, \tag{2}$$

where one recognizes $\partial F/\partial w$ and $\partial F/\partial q$ are sub-blocks in the Jacobian matrix of the nonlinear PDE (1).

In practice, one solves the nonlinear equation using Netwon iterations, and after it converges, calculates the derivatives. For a problem with N_w state variables, N_q parameters, and the Jacobian matrix already L-U factored, it takes N_q back-substitutions to obtain the $N_w \times N_q$ partial derivatives in (2). Obviously, this represents a minuscule addition to the computation cost of solving the PDE.

Since one wishes to allow many device parameters, and to easily add new parameters, the PDE and its Jacobian are assembled using automatic differentiation, and implemented in the Python programming language and as part of the PyEDA package [6].

The transfer characteristics of the 65 nm MOSFET is calculated, and plotted in Fig. 1. The partial derivatives against three device parameters, are computed with the above described automatic differentiation method (AD) and forward-difference method (FD). As shown in Fig. 1, the two methods produces identical results.

Once the drain current and its partial derivatives are calculated, the SLSQP algorithm [4] is used to solve the least square optimization problem. SLSQP supports upper-/lower-bound constraints and inequality constraints. Sensible constraints must be applied to all parameters to prevent unphysical situations.

IV. Parameter Extraction of 65 nm CMOS Devices

A set of measured I-V curves of nMOS transistors with gate length ranges from 75 nm to 10 μm are used for testing the above described algorithm.

In a simplistic attempt, a total of 29 parameters to describe this process. The most critical parameters are those for the channel doping concentration.

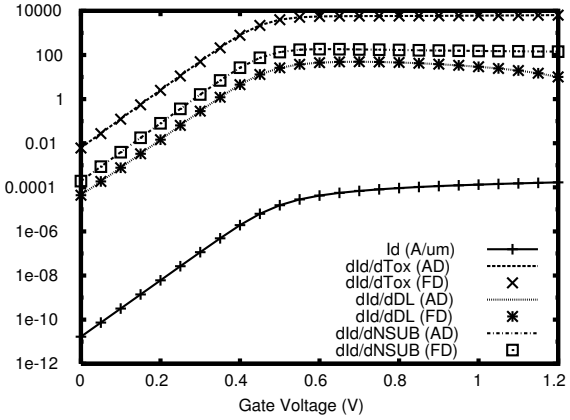


Fig. 1. Transfer characteristics of an nMOSFET transistor with $L_g = 75\text{nm}$, calculated by solving the quasi-2D device equations. Partial derivatives w.r.t. gate oxide thickness (T_{ox}), offset of metallurgical junction from gate edge (D_L) and substrate doping concentration (N_{sub}) are calculated with automatic-differentiation (lines) and forward-differences (symbols).

The long channel device is assumed to have a 1D channel doping profile $C(y, q_{\text{chn,L}})$ that consists of two Gaussian components, and the parameters $q_{\text{chn,L}}$ include peak concentration, peak position and characteristic lengths. Another double-Gaussian 1D profile $C(y, q_{\text{chn,S}})$ is assumed for an imaginary very short channel device. For the channel doping concentration at location x, y, one first interpolates the parameters $q_{\text{chn,L}}$ and $q_{\text{chn,S}}$ according to the x position (horizontal distance to the gate edge), to obtain $q_{\text{chn}}(x)$, and the doping concentration are then calculated from $C(y, q_{\text{chn}}(x))$. At the center of the channel, $q_{\text{chn}}(x)$ approaches the long channel value $q_{\text{chn,L}}$, while the opposite is true at gate edges. The interpolation function used is the complementary error function erfc(), which is smooth and, depending on the its parameters, can approximate both linear transition or exponential decay.

Other doping profile components, such as for deep source/drain and source/drain extensions are similarly described by parameters.

We optimized the parameters using measurement Id-Vg data in the sub-threshold region. After optimization, the net doping profile of the short channel transistor is shown in Fig. 2. The Id-Vg curves for both low and high drain biases are shown in Fig. 3, comparing the model data and measured data. Good agreement is observed in both cases in the sub-threshold region.

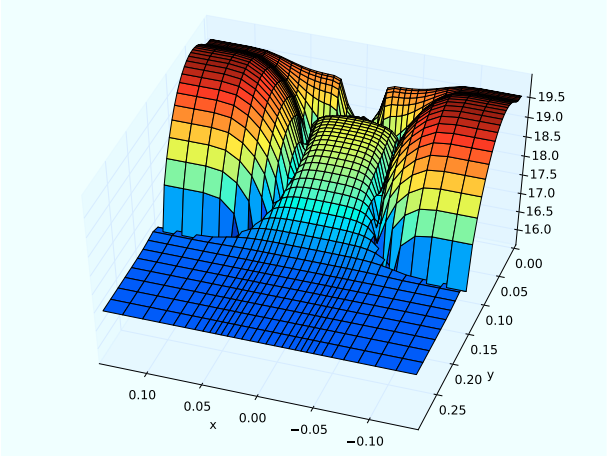
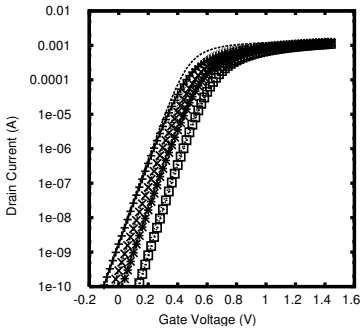
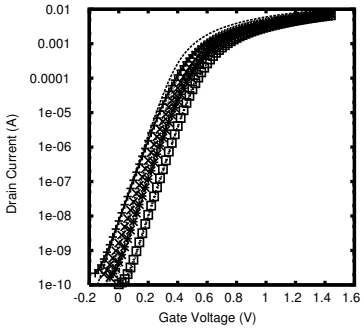


Fig. 2. Net doping concentration in the extracted 75nm nMOSFET transistor.



(a) $V_d = 0.05V$

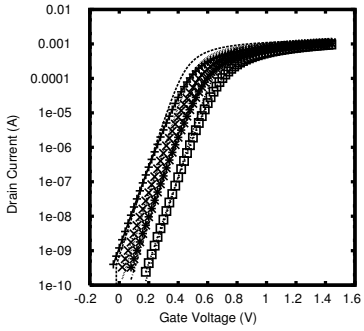


(b) $V_d = 1.2V$

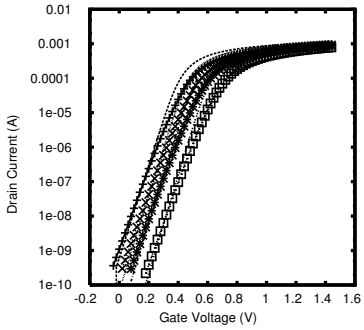
Fig. 3. Transfer characteristics of an nMOSFET transistor with $L_g = 75\text{nm}$, with substrate biases $V_b = 0, -0.3, -0.6, -1.2\text{V}$, and drain biases are a) $V_d = 0.05\text{V}$, and b) $V_d = 1.2\text{V}$. Measurement data are shown with symbols, and model data with lines.

Fig. 4 shows the comparison between model and data for other gate lengths.

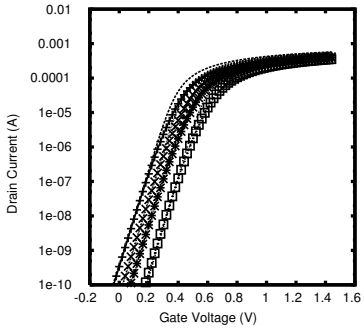
Since this study focuses on the extraction of doping profiles, no attempt was made to match the above-threshold characteristics yet, where mo-



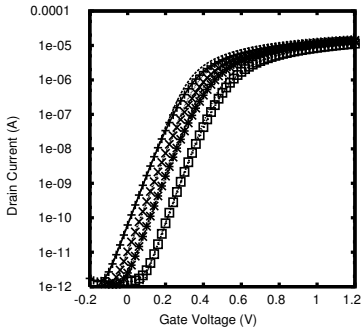
(a) $L_g = 95\text{nm}$



(b) $L_g = 135\text{nm}$



(c) $L_g = 315\text{nm}$



(d) $L_g = 10\mu\text{m}$

Fig. 4. Measured (symbol) and modeled (line) transfer characteristics of nMOSFET transistors of various gate lengths, biased at $V_d = 0.05\text{V}$.

bility and parasitic resistance parameters must be included.

It takes over 12 hours to run all the optimization steps, because 90% of the time is spent on assembling the equations in the slow Python program ($< 50\%$ in typical C++ codes), and no parallelism is exploited. With a rewrite in C++ and multithreading, one can expect to reduce the run time to well below an hour, making inverse modeling feasible as a routine job.

ACKNOWLEDGEMENT

The authors would like to thank Ms D.-M. Ji for technical assistance and Prof. G. Samudra of National University of Singapore for helpful discussions.

REFERENCES

[1] H. Hayashi, H. Matsuhashi, K. Fukuda, and K. Nishi, "Inverse modeling and its application to MOSFET channel profile extraction." *IEICE Trans. Electron.*, vol. E82-C, pp. 862–869, 1999.

[2] I. J. Djomehri and D. A. Antoniadis, "Inverse modeling of sub-100 nm MOSFETs using I-V and C-V," *IEEE Trans. Electron. Devices*, vol. 49, pp. 568–575, 2002.

[3] Y. Li and C.-K. Chen, "A simulation-based evolutionary technique for inverse doping profile problem of sub-65 nm CMOS devices," *J. Comput. Electron.*, vol. 5, pp. 365–370, 2007.

[4] D. Kraft, "Algorithm 733: TOMP-Fortran modules for optimal control calculations," *ACM Transactions on Mathematical Software*, vol. 20, pp. 262–281, 1994.

[5] S. B. Hazra, *Large-Scale PDE-Constrained Optimization in Applications*. Springer-Verlag, 2010.

[6] C. Shen, [Online]. Available: http://github.com/cogenda/pyEDA.