Automatic BSIM3/4 Model Parameter Extraction with Penalty Function

Y. Mahotin and E. Lyumkis

Integrated System Engineering, Inc.
111 N. Market Street, Suite 710, San Jose, CA, USA, Yuri.Mahotin@ise.com

ABSTRACT

The BSIM3/4 compact models have been widely used in the microelectronic industry over the past decade. As the technology scale down achieved the 0.09µ limit, the complexity of the compact models and their parameter extraction has increased incredibly. It has been clearly recognized that the automated parameter extraction methodology can be beneficial for both semiconductor foundries and IC design houses. In this work we report the first successful automatic extraction of BSIM3/4 model parameters based on the numerical optimization of a function, which includes penalty functions. The reported algorithm is an extension of ISE simulator ISExtract [1].

Keywords: BSIM3, BSIM4, parameter extraction, penalty function.

1 INTRODUCTION

Numerical optimization algorithms, and, in particular, the Levenberg-Marquardt algorithm have been commonly used in compact model parameter extraction. This algorithm, invented almost 60 years ago [2] and improved 20 years later [3], has a few known deficiencies, such as poor convergence without a good starting point for minimization, and significant numerical difficulties to solve large dimensional optimization problems. The limited numerical capability of optimization procedures was one of the reasons why commonly used extraction procedures consist of a sequence of many extraction steps, where each of them is a local optimization, or a direct extraction step. The BSIM3/4 compact models have more than 300 parameters (if to account for binning parameters as well), and describe the complicated physics of micro- and nano-scale devices. Both the complexity of the models, and the deficiency of the optimization procedure, lead to extremely complicated extraction strategies, and even extraction experts have difficulties sometimes to achieve acceptable results.

To overcome this stalemate, new interest has recently arisen around the global optimization methodology for model parameter extraction. Watts et al. [4] have successfully applied the Genetic Algorithm for BSIM3 model parameter extraction. They concluded, in particular, that this optimization technique sufficiently reduced the engineering effort to develop a model card, and at the same time improved the model quality. Our current work also uses a global optimization approach. Our numerical algorithm is based on the quasi-newton approach, and demonstrates good convergence with a large number of optimization parameters. The main details of the approach have been reported in [5], where the method has been applied for BSIM3 model parameter extraction.

Besides numerical issues, the global optimization approach has another well-known problem: as a rule, the solution of the problem is not unique, which may lead to nonphysical values of extracted parameters. In traditional extraction methodologies this problem is often solved by direct extraction (not optimization!) of such parameters from physically appropriate sets of data. In our global optimization approach this problem is solved by the introduction of penalty functions, which keep values of extracted parameters within feasible range. Note, that because of the power of the optimization algorithm, the penalty functions did not really worsen the overall convergence of the optimization procedure.

2 OBJECTIVE FUNCTION

A well chosen objective function has a big impact on the quality of the obtained model parameters as well as on the success of the extraction process itself. The agreement between the measured and simulated data is measured by the objective function. Normally, the root mean square function (RMS) is used as the objective function for model parameter extraction:

$$F_{RMS} = \sum \left(\frac{Id_{meas} - Id_{sim}}{Id_{meas}} \right)^2 \qquad (1)$$

where the sum is over all measured data, Id_{meas} and Id_{sim} are measured and simulated (using appropriate compact model) drain currents, respectively. It is clear, that such optimization problems often might have a few, and sometimes, even an infinite number of solutions. Let us consider a simple example: the threshold voltage parameters dtvt0, dvt1, and dvt2, are responsible for the short channel effect and enter into the expression for Vth as dvt0*f(dvt1, dvt2). If, for some reason, during the optimization process, the value of dvt0 becomes equal to zero, then dvt1 and dvt2 may have arbitrary values, and will never affect the value of the objective function. Therefore, the optimization task has an infinite number of solutions in the two dimensional hyperplane, dvt1*dvt2. Due to the complexity of the BSIM3/4 models, and, hence the complexity of the objective function, the N dimensional optimization problem, in general, may have an infinite number of solutions in the K dimensional hyperspace, where K < N. In other words, the F_{RMS} function gives too much freedom to the parameters which have to be optimized.

In order to overcome the above difficulties, and to keep the values of the parameters inside the desired intervals, we use the penalty functions approach. The contribution of the penalty functions to the objective function has been used in the following form:

$$F_{PEN} = \sum \sum \begin{cases} nf \times (x_0 - x)^2 & x < x_0 \\ 0 & x \geq x_0 \end{cases} \quad (2)$$

where the first sum is over all penalties, the second sum over all data, nf is a normalization factor, x is a model parameter or an internal model variable, and x_0 is the boundary of the feasible interval of x.

The total objective function is defined as the sum of (1) and (2):

$$F = F_{RMS} + F_{PEN} \quad (3)$$

3 EXAMPLES OF PENALTIES

Based on the physical and mathematical properties of the models, we introduce 49 penalties for the BSIM3 model and 66 penalties for the BSIM4 model. Because of limited space we can not describe each of these penalties, and we will narrow our discussion by considering a few examples.

Both BSIM3 and BSIM4 models impose a hard restriction on the value of the dvt1 parameter: if dvt1 < 0, the models terminate with a fatal error. This problem can be solved easily by imposing a linear constraint dvt1>0. However, it becomes more difficult to impose just such a rigid constraint if binning parameters for dvt1 have to be found as well: then dvt1 is already a function, and some restriction has to be imposed on this function. So, first of all, a smoothing function is used for dvt1 to avoid negative value, and, in addition, the following penalty function is introduced:

$$F_{PEN}(dvt1) = nf_{dvt1}\left(2 \times 10^{-3} - dvt1\right)^2 \quad (4)$$

if $dvt1 < 2 \times 10^{-3}$. Here nf_{dvt1} is the normalization factor for the dvt1 parameter.

The above example shows, how a penalty function is applied to the model parameters. Similar expressions are used for imposing penalties on internal model variables. Let us consider, for example, an effective mobility model μ_{eff}, which can be written as:

$$\mu_{eff} = \frac{\mu_0}{Denomi} \quad (5)$$

where μ_0 is the low field mobility, and *Denomi* is a function of terminal voltages and more then 100 model parameters, including binning parameters, and describes the mobility degradation due to high electric field. If we want to impose a $Denomi > A$ restriction on the value of *Denomi*, then the appropriate penalty function will have the following form:

$$F_{PEN}(Denomi) = nf_{Denomi}\left(A - Denomi\right)^2 \quad (6)$$

if $Denomi < A$, where nf_{Denomi} is the normalization factor for *Denomi*. The value of A can be chosen from numerical considerations, as it is done in the original BSIM3 model (A=0.2). We believe, that from the physical point of view, A=1 is a more appropriate value for silicon technology, and has been used in our simulations.

4 AUTOSELECTION OF BINNING PARAMETERS

It is well known that without binning the BSIM3/4 models can not provide the required accuracy to model modern technologies for various device sizes. There are no general rules or recommendations on how to choose an optimal set of binning parameters. Often the choice is based on previous experience, and is a very time consuming process as well. To automate the selection process, a procedure called "Autoselection of binning parameters" has been developed. The idea of autoselection is schematically illustrated in Fig.1. The procedure starts from the sub-optimization tasks (one-bin parameter optimization). For each parameter selected for extraction, it solves three two dimensional optimization tasks: 1) parameter plus length dependence binning parameter, 2) parameter plus width dependence binning parameter, and 3) parameter plus cross-term dependence parameter. Therefore three values of the objective function will be calculated for one parameter. After execution of all sub-optimization tasks, the 3N values of objective function are computed, where N is the number of non-binning model parameters. By selecting the minimum component of this vector, we assume, that the appropriate binning parameter is chosen. At the next step the full optimization over all non-binning parameters and the selected binning parameters is performed. The convergence criteria are checked after the full optimization procedure; if they are not satisfied, the procedure is repeated until convergence is reached.

5 PARAMETER EXTRACTION

Typically, the parameter extraction process begins by setting each parameter value to their "best guess" or "desired" value. In order to fit a set of simulated curves to measured data, a series of local optimization steps is performed. Each optimization step attempts to improve the fitting quality of some subset of measured data by adjusting a small number of model parameters dominated on this subset of data. This is the ideal case, but in reality, due to the complicated physics of sub-micron devices, an interaction of different physical effects is present for any subset of measured data and therefore, the fitting quality depends not only on the model parameters selected for local optimization, but on other model parameters, which have the initial values or values extracted before. These difficulties lead to the well-known problem of sequential extraction strategies. In contrast, good optimization algorithms can optimize many or even all model parameters, but still it is crucial to choose a good initial guess. It is well known in practical optimization, that the model parameters can be divided into two groups: the first group contains "work horse" parameters, and the second one – the "ambitious" parameters. The "work horse" parameters are very critical for the successful optimization of the objective function and are easier to estimate. The "ambitious parameters" are less important and/or are more nonlinear and hence harder to estimate. In such cases, it is very useful first to fix the "ambitious parameters" to their default values in order to obtain good initial values for the "work horse" parameters, and then rerun the optimization with all model parameters using these starting values.

Based on the above arguments, our extraction strategy can be summarized as following. First, we execute a sequential extraction strategy for "work horse" parameters. Second, we optimize the "work horse" parameters over all measured data. The next step is an optimization of all model parameters, "work horse" and "ambitious parameters", over all measured data. And if the fitting quality is not good enough, the autoselection of binning parameters procedure can be executed.

6 RESULTS

The new objective function (3) has been used to extract the BSIM3/4 model parameters for three different technologies. The BSIM3 model parameters were extracted for High-Voltage 0.13µ technology, using available measured data for 21 devices with different geometry size. The "model" card was extracted at the foundry using commercial extractors. From the same measured data the "new model" card was obtained using the proposed penalty function approach. A few problems have been discovered in the "model" card, one of them is illustrated in Fig. 2: the simulated Gmb becomes negative when the channel length scales down. Due to a penalty for negative Gmb, the same curves obtained using the "new model" card (Fig. 3) do not show this wrong behavior.

The penalty function approach was also successfully applied to solve similar problems in BSIM4 model parameter extraction. For a new 0.09µ technology, the measurements of 22 devices were available. Again, the "model" card and "new model" card were either provided by the foundry, or extracted using our new approach. A comparison of Gmb for the "model" and "new model" card is shown in Fig. 4: the "model" card Gmb becomes negative when Vb is below -1.5V, which may lead to bad convergence of circuit simulators. To overcome this problem, the idea from [6] was applied. In addition to measured data, we append a set of dummy points with terminal voltages up to two times higher than the normal operating voltage. For these dummy points we do not optimize the drain current, and just keep under control the penalty functions. As a result, in the "new model" card Gmb is positive within twice the range of the operating voltages. The effect of the penalty function approach on internal BSIM4 variables is shown in Fig. 5, where an effective mobility derivative over bulk voltage is plotted. By introducing a penalty function which allows only positive values of this derivative, we guarantee appropriate (physically correct) results during extraction.

To illustrate the global fitting quality of the proposed extraction techniques, Table 1 shows the RMS error calculated using formula (1) and the maximum errors, obtained for a standard 0.18µ technology with a BSIM4 model for the current range from 10^{-10} A to the maximum current. We believe that achieving such a good quality is impossible without using a method of numerical optimization with penalty functions.

7 CONCLUSION

A newly developed penalty function approach was applied to BSIM3/4 model parameter extraction. The penalty functions always keep the values of the model parameters within their physical range without manual tuning and human influence on the extraction procedure. It offers the possibility to get a high quality model card within a short time.

REFERENCES

[1] ISE TCAD Release 9.0, Volume 4b, Integrated System Engineering, 2003.
[2] Levenberg, K., "A Method for the Solution of Certain Problems in Least Squares," Quart. Appl. Math. Vol. 2, pp 164-168, 1944.
[3] Marquardt, D., "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," SIAM J. Appl. Math. Vol. 11, pp 431-441, 1963.
[4] J. Watts et al., "Extraction of Compact Model Parameters for ULSI MOSFETs Using A Genetic Algorithm," Tech. Proc. of the Second Int'l Conf. on Modeling and Simulation of Microsystems, 1999, 176-179.
[5] Y. Mahotin et al., "Parameter extraction of VLSI MOSFET mathematical models by optimization technique," All-Russia scientific and technical conference "Micro- and Nano- electronics engineering", 1998, V.2, P 3-47
[6] J. Watts, "How to Build an SOI MOSFET Compact Model without Violating the Laws of Physics," Tech. Proc. of the 2002 Int'l Conf. on Modeling and Simulation of Microsystems, 2002, 726-729.

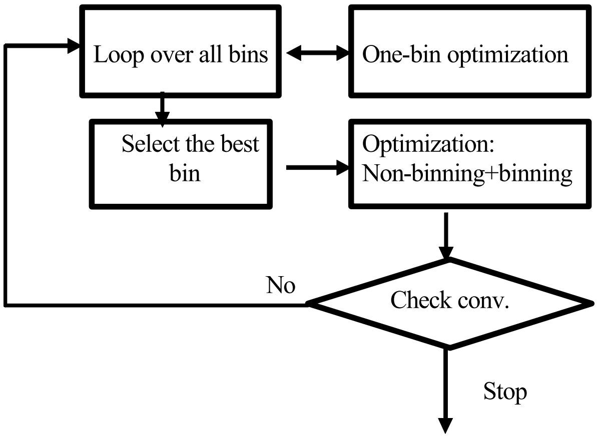

Fig. 1: Autoselection of binning parameters.

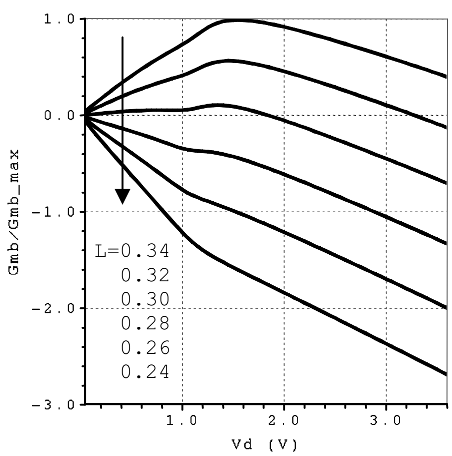

Fig. 2: The "model" card Gmb vs. Vd for different channel length.

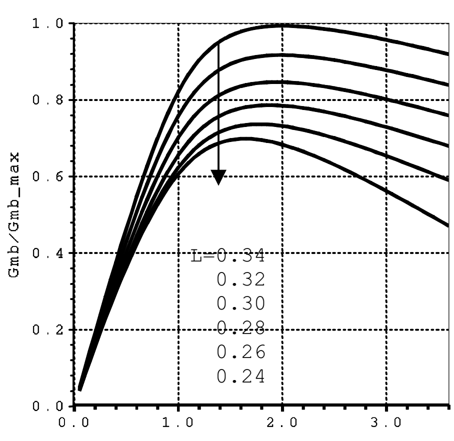

Fig. 3: The "new model" card Gmb vs. Vd for different channel length.

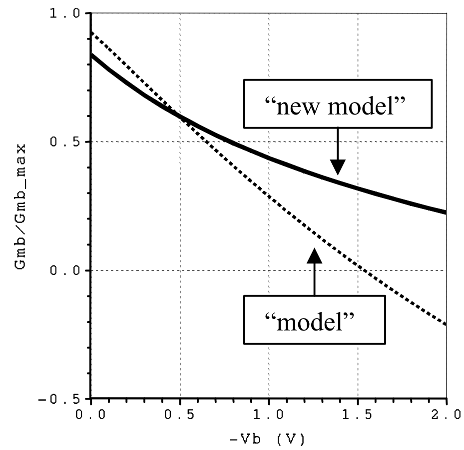

Fig. 4: Gmb vs. Vb at Vg-1.0V, Vd=1.1V

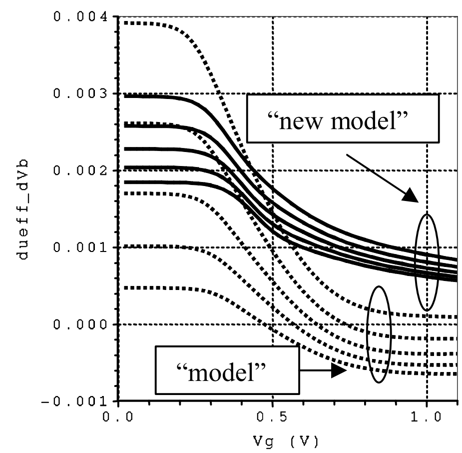

Fig. 5: $\frac{\partial \mu_{eff}}{\partial Vb}$ vs. Vg at Vd=0.05V and different Vb

Device	RMS (%)	MAX (%)
Large	1.6257	3.9397
Short1	1.9632	3.4686
Short2	0.9823	2.5600
Short3	2.4188	4.3845
Short4	2.8921	4.4679
Short5	3.2378	10.0014
Short6	1.6462	3.8046
Small	0.6841	1.3305
Narrow1	1.2388	3.5137
Narrow2	2.4042	4.9862
Narrow3	1.1989	3.2683
Narrow4	1.9566	2.8352
Narrow5	1.2715	3.5191

Table 1: RMS and maximum errors of IdVg curves at low Vd and Vb=0 for different devices.