# Hybrid Evolutionary Algorithm in MOSFET Parameter Extraction

Jarosław Arabas[1] and Sławomir Szostak[2]

[1] Warsaw University of Technology, Institute of Electronic Systems, Warsaw, Poland, email: jarabas@elka.pw.edu.pl
[2] Warsaw University of Technology, Institute of Microelectronics and Optoelectronics, Warsaw, Poland, email: s.szostak@imio.pw.edu.pl

**Abstract.** In this paper we report an ongoing research on applying evolutionary computation to the identification of technological parameters of MOS transistors (MOSFETs) using the current-voltage measurements. The identification consists in approximating the observed values of the current with the values generated by the transistor model. Values of parameters for which the smallest approximation error is observed are assumed to be the best estimations to the real values. The model is nonlinear and nondifferentiable, and the error function takes multiple local minima with respect to the parameter values. We apply a combination of an evolutionary algorithm together with the Nelder-Mead method to minimize the error function and we experimentally investigate the effectiveness of the proposed approach.

## 1 Introduction

Metal-Oxide-Semiconductor Field Effect Transistors (MOS transistors or MOSFETs) are about 90% of the total number of active elements in contemporary integrated circuits (ICs), e.g. microprocessors or memory chips. Electrophysical parameters of MOSFETs, like mobility of carriers or the gate surface, are crucial to attain good efficiency, low energy consumption or high reliability of the whole electronic system [1]. Development of the IC technology results in continuous increase of costs of technological processes, and the increasing complexity of IC layouts. This complicates the design process since more accurate models are needed to cover phenomena that could have been neglected with smaller integration scale. Such models are crucial elements of the IC design and simulation software [2]. The models are also used to extract actual parameters of semiconductor structures and to compare the extracted values to the specification. This can be used for the quality control of the technological process.

Miniaturization, new materials and technologies stimulate the development of both measurement techniques and parameter extraction methods. The most popular method of the parameter extraction is to identify the value of a single parameter at a time, and to repeat the procedure for each parameter to be extracted [8]. The method consists in analyzing measurement curves, e.g. current-voltage curves. Specific features of the curve, such as slope or the argument where the curve goes through zero, are usually used to perform such extraction. This method, though very simple, has, unfortunately,

a poor accuracy since it neglects the influence of another parameters on the curve shape. More reliable extraction results can be attained by performing *global extraction* which consists in approximating the set of measurements by the model generated values. Model nonlinearities imply that a global optimization method should be used to perform that task.

In the parameter extraction procedure one needs to choose an appropriate method of electrical characterization, then measurements have to be made, and, after that, a model can be chosen and used by an optimization method to approximate the measurement data. Several methods of the electrical characterization are in use (e.g. I-V and C-V curves, charge pumping, noise analysis — see [8]) and in this paper we concentrated on the I-V characterization, in which values of the current are measured against the voltage values. In fact, many models are designed to yield the I-V characterization, including models for conventional structures as well as for the short channel, with an alternative dielectric material, multiple gate structures, or silicon-germanium channel. In this paper we use the Pierret-Shields model [7] which can be applied for the whole range of the transistor operation, it takes a small number of parameters and is computationally inexpensive. The model neglects effects which become important for short-channel structures, so we concentrate on long channel MOSFETs only. In further research we plan to apply more sophisticated and more accurate compact models, like EKV [5].

MOSFETs are usually manufactured as elements of integrated circuits and therefore, measurements of a single transistor are usually impossible to perform. Still the approach presented here can be applied in the industrial practice to calibrate the manufacturing process. This can be achieved by interleaving the regular production of ICs with silicon wafers which contain single transistors, available for measurements, and to check if the empirical parameters are consistent with the desired values.

This paper is organized in the following way. In Section 2 we briefly characterize the operation principle of the MOSFET and the method of parameter extraction. We also define the extraction method which uses a composition of an evolutionary algorithm (EA) and a Nelder-Mead nonlinear simplex algorithm (NM). Section 3 provides results of experiments we have performed to verify the applicability of the parameter extraction method. The results indicate that the extraction error function has indeed multiple local minima, and the combination of EA and NM is a good synergy between the robustness against deep local minima (thanks to the EA) and the relative precision in localization the local minima (NM). The paper is briefly concluded in Section 4.

## 2 Model-based extraction of MOSFET parameters

The principle of operation of a MOSFET (see Fig. 1) is modification of the resistance of semiconductor surface by means of the voltage applied to the gate electrode which is separated from the semiconductor by a very thin dielectric layer (usually silicon dioxide). If a sufficiently high positive voltage (the so called threshold voltage) is applied to the gate, enough electrons are attracted to the surface to form a conducting channel and current may flow between heavily doped n+ source and drain regions. This current depends on the channel dimensions $W$ and $L$, carrier mobility $\mu$ and the voltages applied to device electrodes. Threshold voltage depends e.g. on the dielectric constant and thickness of the gate dielectric, bulk doping (acceptors for an n-MOSFET or donors for a p-MOSFET) and workfunction of the gate electrode.

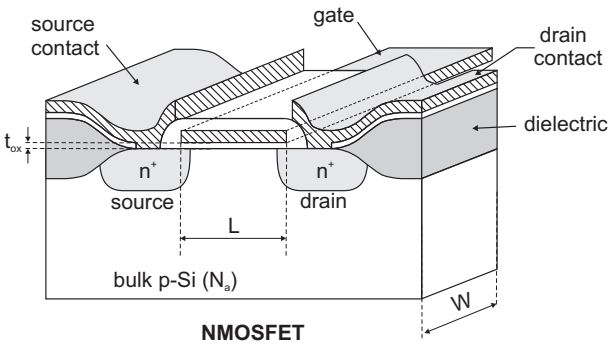


**Figure 1.** Crossection of an n-MOSFET transistor

### 2.1 Extraction as an approximation task

For the parameter extraction we use a model of the considered MOSFET. The instrument can be treated as a physical realization of a mapping $c : R^3 \to R$, which maps the set of input voltage values into the output current. The model is a mapping $m : R^k \times R^3 \to R$ that assigns a value of the output current to a combination of input voltages and the model parameters. The model parameters have usually some physical meaning, like the thickness of the oxidation layer or the channel width and length.

Extraction of the MOSFET parameters consists in reproducing of the aforementioned $k$ parameters of the model with the use of measurements of the voltage and current values. An error function is defined $e : R^k \to R$ that evaluates the discrepancy between the values of the current that have been measured and obtained from the model. A typical choice is the root mean squared error

$$e(\mathbf{p}) = \left( \frac{1}{n} \sum_{j=1}^{n} \left( m(\mathbf{p}, \mathbf{v}_j) - i_j \right)^2 \right)^{\frac{1}{2}} \tag{1}$$

where $i$ is a measurement in which a voltage-current pair $(\mathbf{v}_i, c_i)$ is observed.

The error (1) is sensitive to the choice of the measurement points, so the parameter sets $\mathbf{p}_1$ and $\mathbf{p}_2$ which resulted from minimization of the eror (1), can differ one from another. As a true set of parameters one can assume the vector $\mathbf{p}$ that will result from minimization of the error (1) when the number of measurement points $n$ goes to infinity, and the distribution of the voltage vector values $\mathbf{v}_i$ is uniform in the set of admissible values. From that perspective, $\mathbf{p}_1$ and $\mathbf{p}_2$ are estimations of the value $\mathbf{p}$. It is usually assumed that if the number of measurements is reasonably large (say, 10 measurements per each parameter to be extracted [8]) then the estimation of $\mathbf{p}$ is expected to be quite reliable.

### 2.2 Global and local optimization in the MOSFET parameter extraction

Quite usually, MOSFET models are nonlinear with respect to their parameters $\mathbf{p}$. Therefore, the error function (1) has multiple local minima and the extraction is a global

optimization task. Thus, when a minimization method is applied and returns a parameter vector $\mathbf{p}$ where the error function (1) has its local minimum, we cannot guarantee that the results is indeed a solution to the extraction task. Therefore, it is necessary to apply some global optimization technique. A reference approach is to multiply run a local optimization method, with different initial values of the optimized parameters [8]. A set of values for which the smallest error is observed is the result of the extraction task.

In practice, it may happen that the influence of some parameter $p_j$ on the value of the error function is only slightly affected by the values of the remaining parameters. In such cases, a good estimation to the $p_j$ value can be obtained separately, and it is highly recommended [8] to perform such estimation. In particular, if the model defines the way in which the parameter $p_j$ relates to the current-voltage curves (e.g., is the slope of the curve) then the estimation of $p_j$ value can be obtained by regression of the I-V curve with respect to $p_j$. Quite usually, a linear regression can be applied.

In contrast to a "perfect" regression task, in the process of MOSFET parameter extraction we may expect some dependence between variables, and the character of the dependence cannot be learnt without knowing some details of the model. In other words, one should not attempt to perform the extraction if the model is treated as a black box. For example, if a Pierret-Shields [7] model is considered, the output current is influenced by few parameters, including the channel width $W$ and length $L$, and the mobility of carriers $\mu$. These three parameters are combined together into one parameter, say $q$, given as $q = \mu(W/L)$. Due to this interdependence, it is impossible to properly reproduce the parameters $W, L$ and $\mu$, but only the parameter $q$. In some other situations, there are limitations of ranges of physically admissible values of parameters, and the limitations could be expressed in a form of inequality constraints. Unfortunately, it is sometimes impossible to give the constraints in an analytical form. Even simulation-based constraints which give only the information if the combination of parameter values is feasible, may be sometimes inaccessible (thus, a human expert opinion is needed).

Another difficulty in proper reconstruction of parameter values is the character of the influence that some parameters make on the model output. For some parameters this influence is of an exponential type, and the parameter can take its values from a range of scales. For example, in the structures we considered, the concentration of acceptor dopant $N_a$ could range from $10^{13}$ up to $10^{18}$ [cm$^{-3}$]. This makes the optimization-based regression rather hard to perform, since optimization methods typically perform their action by adding corrections to candidate solutions, and it is hard to adapt the correction step size in such broad range of scales. Therefore, a good choice is to modify the scale in which such parameters are expressed, e.g., to consider the logarithm of the concentration of acceptor dopant.

### 2.3 Evolution and MOSFET parameter extraction

Evolutionary Algorithms (EAs) are quite frequently used when the objective function has many local minima. It is well known [3] that EAs are robust against getting trapped into the neighborhood of a single local minimum, but the price for this advantage is a poor accuracy in local minimization. Therefore we decided to perform a two step procedure — in the first step, EA is run for a relatively short number of iterations, and the best of the EA-generated solutions is used to initialize the Nelder-Mead (NM) optimization algorithm [10].

The NM method is started with simplex which is the population of $k+1$ points ($k$ is the number of extracted parameters), and each of them is generated randomly in a neighborhood of the EA-generated point $\mathbf{x}_{EA}$. The neighborhood size is defined by a parameter $r$ which has the following meaning. Each point $\mathbf{y}$ which goes to the initial population of the NM method is picked up randomly such that each parameter $y_i$ is selected with a uniform distribution from the range $[\max(l_i, (1-r)x_{EA,i}), \min(u_i, (1+r)x_{EA,i})]$, where $l_i, u_i$ are the lower and upper limit for the parameter number $i$, and $x_{EA,i}$ is the value of this parameter in the solution generated by the EA.

## 3 Numerical experiments

### 3.1 Data for the experiment

In all experiments we used the Pierret-Shields model [7]. The model describes all regions of MOSFET operation (i.e. subthreshold, non-saturation and saturation) with one formula.

We analyzed characterization of the NMOSFET with the following parameters: gate oxide thickness $t_{ox} = 1.210^{-7}$ [m], substrate doping $N_a = 10^{18}$ [cm$^{-3}$], mobility of carriers $\mu = 100$ [cm$^2$/Vs], channel width $W = 10^{-3}$ [cm] and channel length $L = 10^{-4}$ [cm]. The remaining model parameters were given. The characterization was performed with the use of I-V curves with $V_{sb} = 0$ [V] and the following combination of voltage values $V_{gs} = 0.25, 0.5, 0.75, \ldots, 1.75$ [V] and $V_{ds} = 0, 0.01, 0.02, \ldots, 0.3$ [V], where $V_{sb}, V_{gs}$ and $V_{ds}$ are the voltage values measured between source and bulk, gate and source, and drain and source, respectively (see Fig. 1).

We used the model to generate the I-V curves. Then we treated these curves as if they were measured and tried to reproduce the original parameters values by applying optimization techniques to minimize the approximation error. Thus we were able to verify the ability of optimization methods to perform the global extraction task.

In all numerical experiments, we rescaled the parameters under optimization such that they covered the range [0,1]. So, instead of optimizing an original parameter $x$, we performed the optimization of a parameter $y = (x-l)/(u-l)$, where $u$ and $l$ are the upper and lower limit of admissible vales for the parameter $x$. The rescaling took place after taking the logarithm of the parameter.

### 3.2 Number of local minima of the error function

First we attempted to apply NM method alone to verify if the error function indeed has many local minima, and what is the "size" of the attraction basins of these minima. Therefore we initialized the NM method using a starting simplex concentrated about a point which was each time randomly generated with the uniform distribution in a feasible area. The whole starting simplex was generated according to the method given in 2.3 using different values of the parameter $r$. With $\mathbf{s}$ denote the vector of true values of parameters which were used by the model to generate artificial "measurements", so $\mathbf{s}$ was the global minimum of the error function. The starting simplex was composed of points generated with a uniform distribution in the area $[\mathbf{s}-r\mathbf{d}, \mathbf{s}+r\mathbf{d}] \cap [\mathbf{l}, \mathbf{u}]$ where $\mathbf{l}, \mathbf{u}$ were the vectors composed of values of lower and upper limit for each parameter under extraction. In addition we tested another version of the simplex initialization which consisted in filling it with points generated independently with the uniform distribution

| $r$ | $t_{ox}$ | $N_a$ | $\mu W/L$ | log(RMS) |
|---|---|---|---|---|
| 0.1 | 1.1 | 15.57 | 0.95 | -82.04 |
| | 0.27 | 35.57 | 0.31 | 28.7 |
| 0.2 | 3.64 | 10.85 | 8.01 | -57.67 |
| | 5.12 | 30.35 | 16.95 | 53.38 |
| 0.3 | 3.47 | 28.76 | 7.13 | -49.84 |
| | 5.95 | 44.18 | 17.51 | 41.47 |
| 0.4 | 4.64 | 6.16 | 66.34 | -40.16 |
| | 9.93 | 22.17 | 126.81 | 43.08 |

**Table 1.** Results for the NM method initialized about the global maximum; for each $r$ and for each parameter we give the mean value (an upper value) and the standard deviation (a value below) of the results

in the box $[\mathbf{l}, \mathbf{u}]$. The NM method was stopped when the maximum distance between points of the simplex fell below a certain level, and the simplex was not reinitialized.

The results obtained by this procedure are given in Tab. 1. We report values of the error and the parameters which were obtained in 20 independent runs of the NM method. We provide mean value and standard deviation of the results. In case of parameters, both mean and standard deviation are expressed as percentage of the true parameter values. We report the results only for $r \leq 0.4$, since for larger values of $r$ we observed unacceptably long computation time (single run of the NM method took more than eight hours).

It can be easily observed that with the value of $r$ increasing, the NM method yields worse results. We conclude that this is the result of getting trapped into local minima of the error function. Note also that the degree of discrepancy between the real values of parameters and the values obtained from the extraction process is different for different parameters. A rule is that $t_{ox}$ is reproduced quite well and quite precisely, since the mean value of the results as well as their standard deviation do not exceed 10 times the original value. Much worse situation we observe for the ratio $\mu W/L$, and an intermediate case is for $N_a$.

### 3.3 Application of the Evolutionary Algorithm

We demonstrated that the error function indeed has many local minima, and the successful application of the NM method is possible when the initial simplex is generated in the neighborhood of the global minimum. We applied the EA to generate the starting point for the NM method.

To verify if the EA is indeed robust against local minima, we performed an experiment in a similar way as for the NM method. We initialized the EA population by filling it with points generated about the global maximum $\mathbf{s}$ for different values of $r$. In addition, we considered the case when the population was filled with points generated with the uniform distribution in the admissible set of parameter values.

We tested the EA whose settings were tuned after a series of preliminary tests. The population size was 10, elitist roulette selection was applied, no crossover was assumed, and the mutation was Gaussian with the standard deviation equal to 0.01. The method was stopped when no improvement of the best solution has been observed after 50 sub-

| $r$ | $t_{ox}$ | $N_a$ | $\mu W/L$ | log(RMS) |
|---|---|---|---|---|
| 0.1 | 1 | 2.11 | 0.84 | -37.63 |
| | 0.05 | 5.92 | 1.23 | 4.94 |
| 0.2 | 1.01 | 4.25 | 0.84 | -36.84 |
| | 0.05 | 6.67 | 0.91 | 4.89 |
| 0.3 | 1 | 2.13 | 0.72 | -36.01 |
| | 0.05 | 6.67 | 0.91 | 4.89 |
| 0.4 | 1.01 | 2.39 | 0.95 | -36.53 |
| | 0.05 | 3.39 | 1.11 | 5.57 |
| 0.5 | 1.02 | 3.82 | 1.33 | -36.98 |
| | 0.05 | 4.68 | 1.94 | 5.74 |
| 0.6 | 1.01 | 4.72 | 0.7 | -36.53 |
| | 0.06 | 6.09 | 1.75 | 5.45 |
| 0.7 | 1 | 3.28 | 0.89 | -37.08 |
| | 0.06 | 7.07 | 0.75 | 4.43 |
| 0.8 | 0.99 | 3.55 | 1.39 | -37.76 |
| | 0.04 | 4.8 | 0.77 | 4.03 |
| 0.9 | 1 | 5.48 | 1.39 | -37.48 |
| | 0.05 | 6.82 | 3.17 | 5.55 |
| unif. | 1 | 5.53 | 0.56 | -37.08 |
| | 0.04 | 8.2 | 2.62 | 6.47 |

**Table 2.** Results for the EA initialized about the global maximum and with a uniform distribution; for each $r$ and for each parameter we give the mean value (an upper value) and the standard deviation (a value below) of the results

sequent generations. The results are given in Tab. 2.

A striking feature of the EA is a relative robustness of the obtained results against the way in which the starting population was initialized. All parameter values were reproduced quite well, although the best reproduction was observed for $t_{ox}$. Note also that the error level for the EA was much worse in comparison to the approximation error that could be attained with the use of the NM method, especially when it was started nearby the global minimum.

Therefore we used a hybrid method which was the combination of the EA and NM. Initial population of the EA was started in the same way as in two previous experiments, i.e. either locally about the global minimum (according to the parameter $r$) or randomly with the uniform distribution in the admissible area. Point with the smallest error yielded by the EA (denote it by $\mathbf{s}_{EA}$) was then used to start the NM method with the value of $r = 0.1$. The best result returned by the NM was recorded as a extraction result. The results are given in Tab 3.

Note that here we again observe quite a little influence of the initialization method on the error value, which is the advantage of the EA. In addition, thanks to the use of the NM method, the error rate was reduced to the level which is only slightly worse than in the case when NM was started close to the global minimum. Note also that the standard deviation of results was dramatically reduced, and the mean values of the reproduced parameters were very close to the original values. It should be emphasized that this effect

| $r$ | $t_{ox}$ | $N_a$ | $\mu W/L$ | log(RMS) |
|---|---|---|---|---|
| 0.1 | 1 | 0.95 | 0.97 | -71.37 |
| | 0.05 | 0.23 | 0.16 | 9.44 |
| 0.2 | 1.01 | 1.45 | 1.14 | -69.53 |
| | 0.06 | 1.7 | 0.64 | 12.85 |
| 0.3 | 1.01 | 1.35 | 1.03 | -69.25 |
| | 0.06 | 1.7 | 0.13 | 12.85 |
| 0.4 | 1.02 | 0.98 | 1.02 | -73.53 |
| | 0.06 | 1.76 | 0.05 | 10.67 |
| 0.5 | 1.02 | 6.24 | 1.1 | -67.35 |
| | 0.05 | 0.07 | 0.47 | 3.7 |
| 0.6 | 1.01 | 0.93 | 1 | -71.48 |
| | 0.06 | 22.11 | 0.09 | 21.97 |
| 0.7 | 1.01 | 1 | 1.01 | -74.06 |
| | 0.06 | 0.23 | 0.04 | 8.06 |
| 0.8 | 1 | 1.19 | 1.03 | -71.56 |
| | 0.04 | 0.06 | 0.17 | 4.01 |
| 0.9 | 1 | 1 | 1 | -74.52 |
| | 0.05 | 0.8 | 0.04 | 8.34 |
| unif. | 1 | 1.04 | 1.02 | -71.77 |
| | 0.04 | 0.06 | 0.1 | 3.98 |

**Table 3.** Results for the EA initialized about the global maximum and with a uniform distribution; for each $r$ and for each parameter we give the mean value (an upper value) and the standard deviation (a value below) of the results

was also achieved when the population was initialized with a uniform distribution in the area of admissible parameter values, so no hints about the correct values of parameters were needed for their proper reproduction. This is a very good news for further practical use of the proposed method to extract parameters based on real measurements.

## 4 Summary and outlook

In this paper we consider an application of an evolutionary algorithm to extract electrophysical parameters of a MOS transistor. The method consists in minimizing the error of approximating the measured values of the current with the values generated by the model. We demonstrated that the task has multiple local minima and therefore it cannot be effectively solved by a local optimization method. After applying an evolutionary algorithm we recognized that the quality of results was in general independent on the initialization method, but it was fairly imprecise. A hybrid approach in which the evolutionary algorithm was used to initialize the local optimization method appeared to be a good choice which is both precise and robust.

We validated the approach on the artificial, model generated data. In further research we plan to apply the methodology for more complicated models, and in particular, for the EKV model [5]. We will also validate the results using the measurement data of real transistors manufactured in the Institute of Electron Technology in Warsaw.

**Acknowledgements**

The authors thank to Lidia Łukasiak, Warsaw University of Technology, for providing the source code of the Pierret-Shields model, and to Daniel Tomaszewski, Institute of Electron Technology, for valuable discussions on the parameter extraction task.

## Bibliography

[1] *International Technology Roadmap for Semiconductors*, Semiconductor Industry Association, 2007.

[2] N. Arora, *MOSFET Modeling for VLSI Simulation Theory and Practice*, World Scientific Publishing, 2007.

[3] J.Arabas, *Lecture Notes in Evolutionary Algorithms*, WNT, 2007 (in Polish).

[4] G. Chindalore, *et al.*, "An experimental study of the effect of quantization of the effective electrical oxide thickness in MOS electron and hole accumulation layers in heavily doped Si", *IEEE Trans. Electron Dev*, vol. 47, p. 643, 2000.

[5] C. Enz, E. A. Vittoz, *Charge-Based MOS Transistor Modeling - the EKV Model for Low-Power and RF IC Design*, Wiley, 2006.

[6] Z.Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer, 1996

[7] R.F. Pierret, J.A. Shields, "Simplified long-channel MOSFET theory", *Solid-State Electron*, vol. 26, p. 143, 1983.

[8] D. Schröder, *Semiconductor material and device characterization*, Wiley, 1990.

[9] S. Szostak, *Charge pumping in SOI structures – model and parameter extraction*, Ph.D. Thesis, Warsaw University of Technology, 2001 (in Polish).

[10] W.Press *et al.*, *Numerical Recipes in C*, Cambridge Univ. Press, 1992.