A Methodology for Statistical Parameter Extraction from DC and Small Signal Measurements

M. CONTI, P. CRIPPA, S. ORCIONI, C. TURCHETTI
Department of Electronics
University of Ancona
via Brecce Bianche, 60131 Ancona
ITALY

Abstract: - Device model accuracy not only depends on the model itself but also on the procedure used for parameter extraction and on the experimental data collected for the extraction. In analog circuits, many of the performances depend on the small signal parameters of the devices rather than on the static current (for example the gain of a differential amplifier). Hence experimental measurements of small signal parameters must be collected to increase model accuracy. In this paper an extraction procedure that uses experimental data both on drain current and on small signal parameters of MOSFET's is proposed. It is shown the correspondence between the inclusion of small signal parameters in the optimization procedure and the use of the off-diagonal terms of the covariance matrix in the Maximum Likelihood Estimation Method.

Key-Words: -.Device Modeling, Small Signal Parameters, Statistical Analysis

1 Introduction

One of the problems encountered in device modeling for analog circuit CAD is related to the prediction of small-signal parameters. As was recently pointed out in [1], MOSFET models, which are able to accurately predict drain current, may make very poor predictions of small-signal parameters. This problem becomes particularly serious in the design of voltage amplifiers where the voltage gain inversely depends on g_{ds}. Thus, in general a MOSFET model for analog design should give accurate values for small-signal quantities such as g_m, g_{mb}, g_{ds}, and capacitances.

Many models are inherently incapable of giving a reasonable fit to small signal parameters due to discontinuities or missing effects in the equations. However, in many cases this problem is not a consequence of weakness in the model. Instead, it might depend on the extraction method for the unknown parameters, which uses experimental data on drain current I_D alone, without any information on the actual behavior of small-signal parameters. In order to obtain a model adequate also for small-signal analysis, these two conditions have to be met:

i) experimental data on small-signal parameters must be collected together with DC drain current measurements;

ii) the extraction method must take into account the error between model and data for DC current and small signal parameters.

Following these assumptions some measurements of DC drain current and small signal parameters have been carried out on test devices and a model for the MOSFET drain current derived by taking into account the small signal measurements has been obtained. Section 2 shows the theory underlying the proposed methodology. Section 3 resumes the obtained experimental results. Finally Section 4 concludes the work.

2 Theory outline

Let us consider that a model for the DC drain current has been chosen and defined by a function

$$y = f(\boldsymbol{x},\boldsymbol{s}) \qquad (1)$$

where y is the drain current, $\boldsymbol{x}$ is the vector of independent variables, i. e. $[V_{gs}, V_{ds}, V_{bs}, W, L]^{\mathrm{T}}$ and $\boldsymbol{s}$ is the parameter vector, which usually represents quantities that have physical significance. Once the model (1) has been assumed, the unknown parameters $\boldsymbol{s}$ can be derived from experimental data by using estimation theory.

A complete characterization of device behavior can be achieved by gathering several sets of data

$$\boldsymbol{d}_\mu = (d_{\mu,1}, \ldots, d_{\mu,\eta}, \ldots, d_{\mu,N})^{\mathrm{T}} \qquad \mu = 1, \ldots, \mathrm{M} \qquad (2)$$

from M different dice by imposing the same value $\boldsymbol{x}_\eta$ for the independent variables, where $\eta = 1, \ldots, \mathrm{N}$. The error ε, between model and experimental data, can be considered as a realization of a random variable that takes into account for measurement errors, intra-die and inter-die technological tolerances and model inaccuracy [2]. Here we assume that the error vector ε_μ in the μ–th die have zero mean with normal distribution $N_r(0, \boldsymbol{Q})$, that is

$$p\left(\varepsilon_\mu|\boldsymbol{Q}\right)=\left[(2\pi)^N\det\boldsymbol{Q}\right]^{-\left(1/2\right)}\exp\left(\frac{1}{2}\varepsilon_\mu^T\boldsymbol{Q}^{-1}\varepsilon_\mu\right)\tag{3}$$

where

$$\varepsilon_\mu=\left(\varepsilon_{\mu,1},\ldots,\varepsilon_{\mu,\eta},\ldots,\varepsilon_{\mu,N}\right)^{\mathrm{T}}$$
$$=\left(f(\boldsymbol{x}_1,\boldsymbol{s})-d_{\mu,1},\ldots,f(\boldsymbol{x}_\eta,\boldsymbol{s})-d_{\mu,\eta},\ldots,f(\boldsymbol{x}_N,\boldsymbol{s})-d_{\mu,N}\right)^{\mathrm{T}}$$
$$=\boldsymbol{f}_N(\boldsymbol{x},\boldsymbol{s})-\boldsymbol{d}_\mu$$
$$\boldsymbol{f}_N\left(\boldsymbol{x},\boldsymbol{s}\right)=\left(f\left(\boldsymbol{x}_1,\boldsymbol{s}\right),\ldots,f\left(\boldsymbol{x}_N,\boldsymbol{s}\right)\right)^T$$

and $\boldsymbol{Q}$ is the covariance matrix that is assumed to be independent from the die under consideration.

The off-diagonal terms of $\boldsymbol{Q}$ take into account the correlation between the errors in the same die (μ) for different bias conditions (η). By assuming that the errors in different dice are statistically independent each other, the joint pdf for all the experiments is given by

$$p\left(\boldsymbol{E}|\boldsymbol{Q}\right)=\left[(2\pi)^N\det\boldsymbol{Q}\right]^{-\left(M/2\right)}\exp\left(\frac{1}{2}\sum_{\mu=1}^{M}\varepsilon_\mu^T\boldsymbol{Q}^{-1}\varepsilon_\mu\right)\tag{4}$$

where $\boldsymbol{E}=\left\{\varepsilon_{\mu,\eta}\right\}$. The parameter values $\boldsymbol{s}$ of model (1) are obtained using the well-known Maximum Likelihood Estimation (MLE) method [3]. Following these statistical assumptions, likelihood can be written as

$$L(\boldsymbol{s},\boldsymbol{Q})=\left[(2\pi)^N\det(\boldsymbol{Q})\right]^{-M/2}\cdot$$
$$\cdot\exp\left\{-1/2\sum_{\mu=1}^{M}\left[\boldsymbol{d}_\mu-\boldsymbol{f}_N\left(\boldsymbol{x},\boldsymbol{s}\right)\right]^T\boldsymbol{Q}^{-1}\left[\boldsymbol{d}_\mu-\boldsymbol{f}_N\left(\boldsymbol{x},\boldsymbol{s}\right)\right]\right\}\tag{5}$$

The estimation of unknown parameters $\boldsymbol{s}$ and $\boldsymbol{Q}$ is thus obtained by maximizing $L(\boldsymbol{s},\boldsymbol{Q})$. This non-linear maximization can be solved by using the following iterative scheme [2]:

Step 1) Let us consider a trial value $\boldsymbol{s^0}$ for $\boldsymbol{s}$ and the following function is maximized

$$\max_{\boldsymbol{Q}}L(\boldsymbol{s}^0,\boldsymbol{Q})\tag{6}$$

whose solution is [3]

$$\boldsymbol{Q_{ij}^0}=\frac{1}{M}\sum_{\mu=1}^{M}\left[d_{\mu,i}-f\left(\boldsymbol{x}_i,\boldsymbol{s}^0\right)\right]\left[d_{\mu,j}-f\left(\boldsymbol{x}_j,\boldsymbol{s}^0\right)\right].\tag{7}$$

Step 2) Use $\boldsymbol{Q_{ij}^0}$ as trial value for solving the non linear minimization problem

$$\min_{\boldsymbol{s}}\sum_{\mu=1}^{M}\left[\boldsymbol{d}_\mu-\boldsymbol{f}_N(\boldsymbol{x},\boldsymbol{s})\right]^T\boldsymbol{Q}^{-1}\left[\boldsymbol{d}_\mu-\boldsymbol{f}_N(\boldsymbol{x},\boldsymbol{s})\right].\tag{8}$$

The $\boldsymbol{s}$ value so found is used as a trial value in step 1). Then the method continues until convergence is reached.

Since the number of coefficients in the matrix is high ($2N^2$), simplifying hypotheses are often introduced, such us neglecting off diagonal terms and defining a model for diagonal coefficients Q_{ii}. For example it can be assumed that the absolute error is independent on $\boldsymbol{x}_i$, resulting in $Q_{ii}=\sigma^2$.

Using this assumption (8) becomes

$$\min_{\boldsymbol{s}}\left\{\sum_{\mu=1}^{M}\frac{\left[\boldsymbol{d}_\mu-\boldsymbol{f}_N(\boldsymbol{x},\boldsymbol{s})\right]^T\left[\boldsymbol{d}_\mu-\boldsymbol{f}_N(\boldsymbol{x},\boldsymbol{s})\right]}{\sigma^2}\right\}.$$
$$=\min_{\boldsymbol{s}}\left\{\sum_{\mu=1}^{M}\sum_{\eta=1}^{N}\left[\boldsymbol{d}_{\mu,\eta}-f\left(\boldsymbol{x}_\eta,\boldsymbol{s}\right)\right]^2\right\}\tag{9}$$

Another assumption usually stated is that the relative error is independent on $\boldsymbol{x}_i$, that is $Q_{ii}=\left(\sigma y_i\right)^2$

Using this assumption (8) becomes

$$\min_{\boldsymbol{s}}\left\{\sum_{\mu=1}^{M}\sum_{\eta=1}^{N}\frac{\left[d_{\mu,\eta}-f\left(\boldsymbol{x}_\eta,\boldsymbol{s}\right)\right]^2}{d_{\mu,\eta}^2}\right\}\tag{10}$$

Although (9) and (10) are usually adopted for parameter extraction because of their simplicity, neglecting the off diagonal terms in the covariance matrix $\boldsymbol{Q}$ may induce a large error in the parameter estimation. A strong correlation between the errors of the model against experimental data for the same MOSFET but for two bias conditions close each other may occur.

In the following we shall show how the off diagonal terms can be taken into account by including small signal parameters in the extraction procedure. Let us indicate the covariance matrix $\boldsymbol{Q}$ as follows

$$\begin{matrix}a_1 & b_1 & c_1 & d_1 & \\ b_1 & a_2 & b_2 & c_1 & \\ c_1 & b_2 & a_3 & b_3 & \ldots\\ d_1 & c_1 & b_3 & a_4 & \\ & & \ldots & & \end{matrix}\tag{11}$$

and let us suppose, for example, that the data have been collected for different values of V_{GS} and ordered with increasing value of V_{GS}. We suppose that $\boldsymbol{Q}$ is a diagonal dominant matrix, that is

$$1>>\left|\frac{b_i}{a_i}\right|>>\left|\frac{b_i}{a_i}\right|^2\ \text{ and }\ \left|\frac{b_i}{a_i}\right|>>\left|\frac{c_i}{a_i}\right|>>\left|\frac{d_i}{a_i}\right|\tag{12}$$

This assumption is acceptable since it states that the correlation is stronger if the bias conditions are closer each other. With previous assumption, we obtain the following approximation for $\boldsymbol{Q}^{-1}$

$$\begin{matrix}1/a_1 & -\mathrm{b}_1/a_1a_2 & 0 & 0\\ -\mathrm{b}_1/a_1a_2 & 1/a_2 & -\mathrm{b}_2/a_2a_3 & 0\\ 0 & -\mathrm{b}_2/a_2a_3 & 1/a_3 & -\mathrm{b}_3/a_3a_4\\ 0 & 0 & -\mathrm{b}_3/a_3a_4 & 1/a_4\end{matrix}\tag{13}$$

Hence we obtain

$$\begin{aligned}
&\left[\boldsymbol{d}_\mu - \boldsymbol{f}_N(\boldsymbol{x},\boldsymbol{s})\right]^T \boldsymbol{Q}^{-1}\left[\boldsymbol{d}_\mu - \boldsymbol{f}_N(\boldsymbol{x},\boldsymbol{s})\right] \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N}\left[d_{\mu,i} - f(\boldsymbol{x}_i,\boldsymbol{s})\right]^T Q_{ij}^{-1}\left[d_{\mu,j} - f(\boldsymbol{x}_j,\boldsymbol{s})\right] \\
&= 2\sum_{i=1}^{N}\left[d_{\mu,i} - f(\boldsymbol{x}_i,\boldsymbol{s})\right]^2 \sum_{j=1}^{N} Q_{ij}^{-1} \\
&\qquad - \sum_{i=1}^{N}\sum_{j=1}^{N}\left[\left(d_{\mu,i} - d_{\mu,j}\right) - \left(f(\boldsymbol{x}_i,\boldsymbol{s}) - f(\boldsymbol{x}_j,\boldsymbol{s})\right)\right]^2 \\
&= 2\sum_{i=1}^{N}\left[d_{\mu,i} - f(\boldsymbol{x}_i,\boldsymbol{s})\right]^2\left(1/a_i - 2b_i/(a_i a_{i+1})\right) \\
&\qquad - \sum_{i=1}^{N}\left[\left(d_{\mu,i} - d_{\mu,j}\right) - \left(f(\boldsymbol{x}_i,\boldsymbol{s}) - f(\boldsymbol{x}_j,\boldsymbol{s})\right)\right]^2\left(-2b_i/(a_i a_{i+1})\right) \\
&\approx \sum_{i=1}^{N}\left[d_{\mu,i} - f(\boldsymbol{x}_i,\boldsymbol{s})\right]^2 (2/a_i) \\
&\qquad + \sum_{i=1}^{N}\left[\left(d_{\mu,i} - d_{\mu,j}\right) - \left(f(\boldsymbol{x}_i,\boldsymbol{s}) - f(\boldsymbol{x}_j,\boldsymbol{s})\right)\right]^2\left(2b_i/(a_i a_{i+1})\right)
\end{aligned} \tag{14}$$

Since data have been ordered with increasing V_{GS}, we have

$$\begin{aligned}
&\left[\boldsymbol{d}_\mu - \boldsymbol{f}_N(\boldsymbol{x},\boldsymbol{s})\right]^T \boldsymbol{Q}^{-1}\left[\boldsymbol{d}_\mu - \boldsymbol{f}_N(\boldsymbol{x},\boldsymbol{s})\right] = \\
&= \sum_{i=1}^{N}\left[I_{Di} - f(\boldsymbol{x}_i,\boldsymbol{s})\right]^2 (2/a_i) \qquad (15) \\
&+ \sum_{i=1}^{N}\left[g_m - \frac{\left(f(\boldsymbol{x}_i,\boldsymbol{s}) - f(\boldsymbol{x}_j,\boldsymbol{s})\right)}{\Delta V_{GS}}\right]^2 \left(2b_i/(a_i a_{i+1})\right)\Delta V_{GS}
\end{aligned}$$

where $g_m \cong \left(d_{\mu,i} - d_{\mu,(i-1)}\right)/\Delta V_{GS}$.

Hence to consider non vanishing off diagonal terms in the covariance matrix is equivalent to include the small signal parameters in the optimization procedure. On the contrary, neglecting the off diagonal terms in the minimization of (8) may cause model inaccuracy in the prediction of small signal parameters.

On the other hand, taking into account the off diagonal terms too, would cause an increasing in the complexity of minimization of (8). For example if 50 measurements for different bias conditions are carried out for each MOSFET, 10 MOSFET's are considered for each die and 20 dice are taken into account for the statistical characterization (i.e. M=20 and N=50x10=500), 10,000 terms have to be summed up in (8) in the case of diagonal covariance matrix, while the number of terms becomes 5,000,000 (=500x500x20) in the general case. A compromise is to accept hypothesis (12) which leads to a summation of 30,000 terms.

Since it has been shown the equivalence between considering the off-diagonal terms and adding conductances terms in the functional (8) the solution we propose is to minimize the error between drain current model and current measurements augmented by the error between conductance model and conductance measurements. That is:

$$\begin{aligned}
&\max_{s,\alpha,\beta,\gamma,\delta} L(\boldsymbol{s},\alpha,\beta,\gamma,\delta) = \\
&\max_{s,\alpha,\beta,\gamma,\delta}\left[(2\pi)^N\prod_{\eta=1}^{N}\alpha_\eta^2\right]^{-M/2}\exp\left\{\sum_{\mu=1}^{M}\sum_{\eta=1}^{N} - \frac{\left[I_{D_{\mu\eta}} - f\left(\boldsymbol{x}_{ç},\boldsymbol{s}\right)\right]^2}{2\alpha_\eta^2}\right\} \\
&+\left[(2\pi)^N\prod_{\eta=1}^{N}\beta_\eta^2\right]^{-M/2}\exp\left\{\sum_{\mu=1}^{M}\sum_{\eta=1}^{N} - \frac{\left[g_{m_{\mu\eta}} - \frac{\partial f\left(\boldsymbol{x}_\eta,\boldsymbol{s}\right)}{\partial V_{GS}}\right]^2}{2\beta_\eta^2}\right\} \\
&+\left[(2\pi)^N\prod_{\eta=1}^{N}\gamma_\eta^2\right]^{-M/2}\exp\left\{\sum_{\mu=1}^{M}\sum_{\eta=1}^{N} - \frac{\left[g_{d_{\mu\eta}} - \frac{\partial f\left(\boldsymbol{x}_\eta,\boldsymbol{s}\right)}{\partial V_{DS}}\right]^2}{2\gamma_\eta^2}\right\} \\
&+\left[(2\pi)^N\prod_{\eta=1}^{N}\delta_\eta^2\right]^{-M/2}\exp\left\{\sum_{\mu=1}^{M}\sum_{\eta=1}^{N} - \frac{\left[g_{mb_{\mu\eta}} - \frac{\partial f\left(\boldsymbol{x}_\eta,\boldsymbol{s}\right)}{\partial V_{BS}}\right]^2}{2\delta_\eta^2}\right\}
\end{aligned} \tag{16}$$

where $\alpha = \left(\alpha_1,\ldots,\alpha_\eta,\ldots,\alpha_N\right)^T$, $\beta = \left(\beta_1,\ldots,\beta_\eta,\ldots,\beta_N\right)^T$, $\gamma = \left(\gamma_1,\ldots,\gamma_\eta,\ldots,\gamma_N\right)^T$ and $\delta = \left(\delta_1,\ldots,\delta_\eta,\ldots,\delta_N\right)^T$ are the vectors of drain and conductance variances.

The minimization algorithm is the following:

Step 1) Let us consider a trial value $\boldsymbol{s}^0$ for $\boldsymbol{s}$ and the following function is maximized

$$\max_{s,\alpha,\beta,\gamma,\delta} L(\boldsymbol{s}^0,\alpha,\beta,\gamma,\delta)$$

whose solution is

$$\begin{aligned}
\alpha_\eta^0 &= \frac{1}{M}\sum_{\mu=1}^{M}\left[I_{D_{\mu,\eta}} - f(\boldsymbol{x}_\eta,\boldsymbol{s}^0)\right]^2 \\
\beta_\eta^0 &= \frac{1}{M}\sum_{\mu=1}^{M}\left[g_{m_{\mu,\eta}} - \frac{\partial f(\boldsymbol{x}_\eta,\boldsymbol{s}^0)}{\partial V_{GS}}\right]^2 \\
\gamma_\eta^0 &= \frac{1}{M}\sum_{\mu=1}^{M}\left[g_{d_{\mu,\eta}} - \frac{\partial f(\boldsymbol{x}_\eta,\boldsymbol{s}^0)}{\partial V_{DS}}\right]^2 \\
\delta_\eta^0 &= \frac{1}{M}\sum_{\mu=1}^{M}\left[g_{mb_{\mu,\eta}} - \frac{\partial f(\boldsymbol{x}_\eta,\boldsymbol{s}^0)}{\partial V_{BS}}\right]^2
\end{aligned} \tag{17}$$

Step 2) Use the values so found for $\alpha,\beta,\gamma,\delta$ to solve the nonlinear minimization of

$$\min_{s}\sum_{\mu=1}^{M}\sum_{\eta=1}^{N}\frac{\left[I_{D_{\mu\eta}}-f\left(\boldsymbol{x}_{\eta},\boldsymbol{s}\right)\right]^{2}}{\alpha_{\eta}^{2}}+\frac{1}{\beta_{\eta}^{2}}\left[g_{m_{\mu\eta}}-\frac{\partial f\left(\boldsymbol{x}_{\eta},\boldsymbol{s}\right)}{\partial V_{GS}}\right]^{2}+\frac{1}{\gamma_{\eta}^{2}}\left[g_{d_{\mu\eta}}-\frac{\partial f\left(\boldsymbol{x}_{\eta},\boldsymbol{s}\right)}{\partial V_{DS}}\right]^{2}+\frac{1}{\delta_{\eta}^{2}}\left[g_{mb_{\mu\eta}}-\frac{\partial f\left(\boldsymbol{x}_{\eta},\boldsymbol{s}\right)}{\partial V_{BS}}\right]^{2} \qquad (18)$$

The $\boldsymbol{s}$ value so found is used as a trial value in step 1). The method continues until convergence is reached.

3 Results

On the basis of the theory previously outlined an accurate model for the MOS transistor has been derived and its validity has been checked with reference to the CMOS section of the 2µm CCD/CMOS technology developed at the IRST Microelectronics Laboratory for implementing application specific electro-optical sensors. Many measurements have been made on a test-strip that contains p-n channel MOS transistors of different width and length. Experimental data has been collected for DC current and small-signal parameters.

Following the optimization procedure given by (17)-(18) the terms $\alpha, \beta, \gamma, \delta$ have been estimated first by assuming a trial value $\boldsymbol{s}^0$ for the parameters and then repeating several times the steps 1 and 2 of the minimization algorithm. The obtained model is depicted in Figs. 1-11 in which different MOSFET in different bias conditions are shown. In the Figures continuous line represent the model, while symbols represent experimental data.

Since measurement errors are negligible, the dispersion on the experimental data are due to die to die technological variations in the same wafer. As you can see the agreement between model and experimental data is good for DC drain current and small signal parameters too.

4 Conclusions

Accurate analytical device models are required for the design of IC, especially for analog circuits. The accuracy is necessary not only to predict static current but small signal parameters too. To achieve this goal the parameter extraction procedure must take into account measurements on drain current and on small signal quantities such as g_m, g_{mb}, g_{ds}.

Acknowledgments

The authors are thankful to Prof. Giovanni Soncini of IRST-Istituto di ricerca scientifica e Tecnologica, Trento, Italy, for his valuable contribution to experimental characterization of MOS devices. This work was partially supported by Italian National Research Council, CNR.

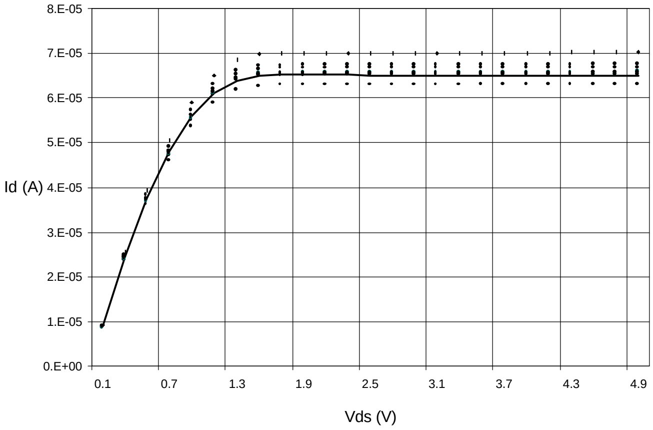

Fig. 1 I_D for a 100x100µm device with Vbs=-1.0V, Vgs=3.5V.

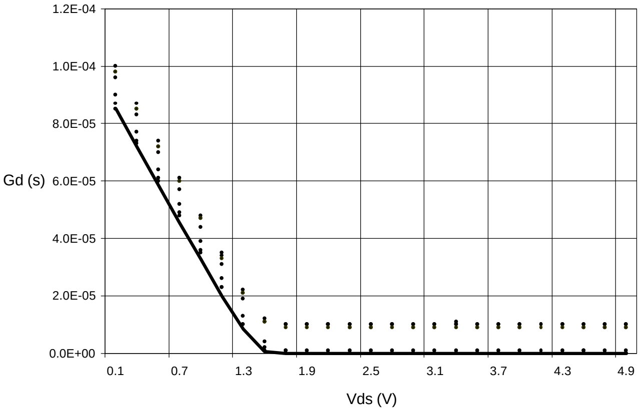

Fig. 2 G_d for a 100x100µm device with Vbs=-1.0V, Vgs=3.5V.

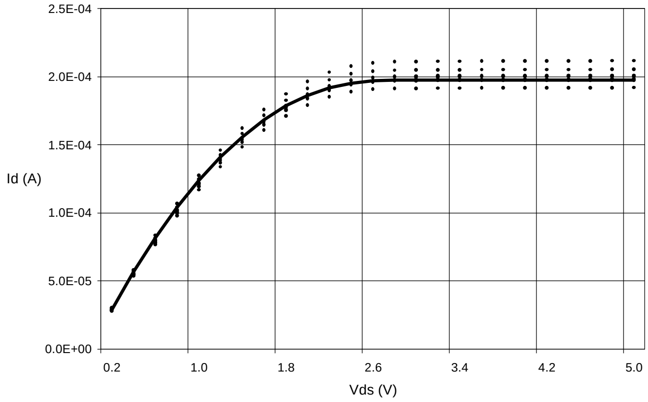

Fig. 3 I_D for a 100x100µm device with Vbs=-1.0V; Vgs=5.0V.

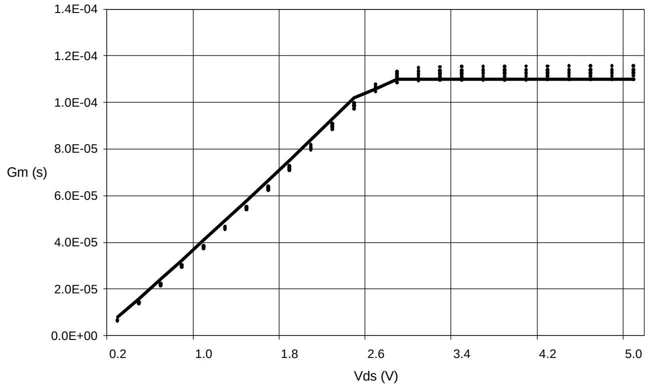

Fig. 4 G_m for a 100x100µm device with Vbs=-1.0V, Vgs=5.0V.

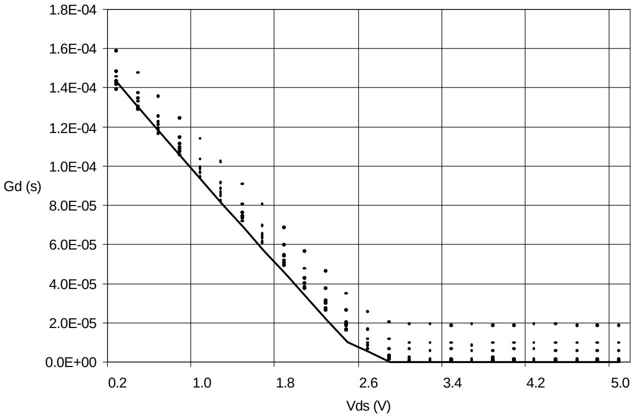

Fig. 5 G_d for a 100x100µm device with Vbs=-1.0V; Vgs=5.0V

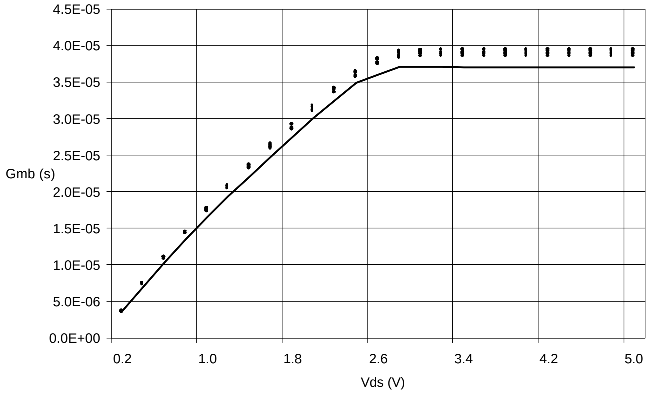

Fig. 6 G_{mb} for a 100x100µm device with Vbs=-1.0V; Vgs=5.0V

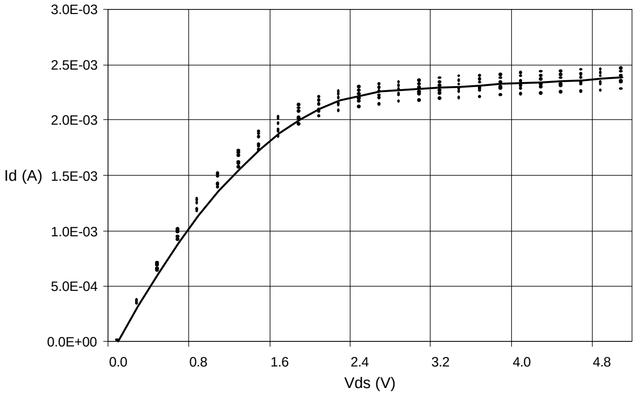

Fig. 7 I_D for a 36x4µm device with Vbs=-1.0V, Vgs=5.0V.

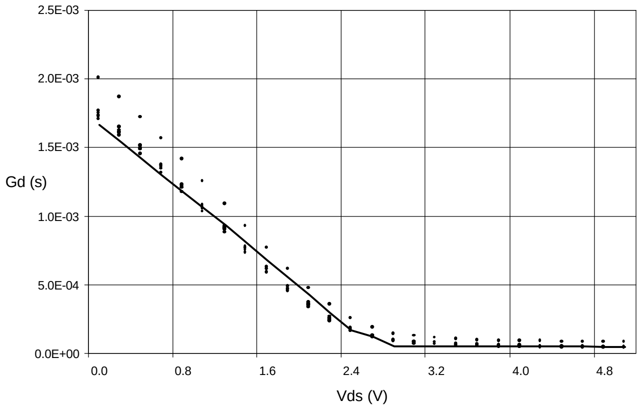

Fig. 8 G_d for a 36x4µm device with Vbs=-1.0V, Vgs=5.0V.

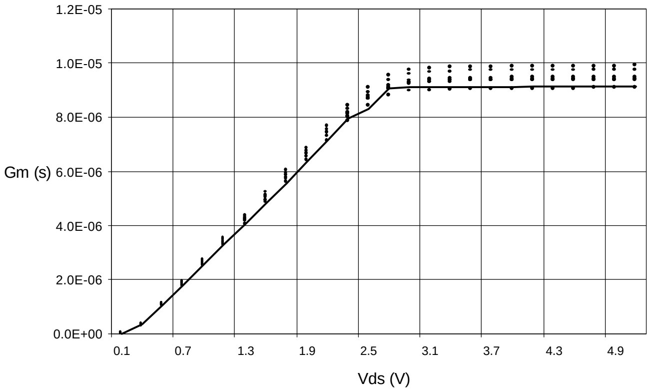

Fig. 9 G_m for a 4x36µm device with Vbs=-1.0V, Vgs=5.0V.

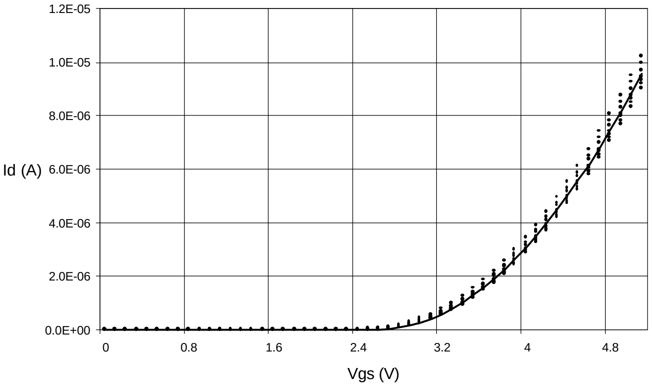

Fig. 10 I_D for a 4x36µm device with Vbs=-1.0V, Vds=5.0V.

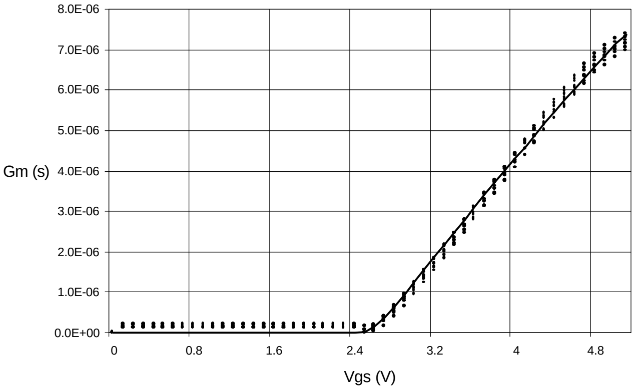

Fig. 11 G_m for a 4x36µm device with Vbs=-1.0V, Vds=5.0V.

References:

[1] Y. P. Tsividis, ‘MOSFET modeling for analog circuit CAD: Problems and prospect’, *IEEE J. Solid-State Circuits*, 1994, **29** (3) pp. 210-216

[2] M. Conti, S. Orcioni, G. Soncini, C. Turchetti, and N. Zorzi, ‘Analytical device modeling for MOS Analog ICs based on Regularization and Bayesian Estimation’, *IEEE Transaction on Computer Aided Design of Integrated Circuits and Systems*, 1996, **15** (11) pp. 1309-1323

[3] Y. Bard, ‘Nonlinear parameter estimation’, Academic Press, New York, 1974.