Worst-Case MOSFET Parameter Extraction for a 2μm CMOS Process

K. Burke, J.A. Power*, B. Donnellan, K. Moloney and W.A. Lane**

Analog Devices B.V., Raheen Industrial Estate, Limerick, IRELAND.

*** Silvaco Data Systems, Santa Clara, CA 94054.**

**** National Microelectronics Research Centre (NMRC), Lee Maltings, Prospect Row, Cork, IRELAND.**

Abstract

This paper will describe the process by which realistic nominal and worst-case DC MOSFET model parameter sets were determined and validated for a 2 μm CMOS technology. The steps involved in this task, which will be detailed, ranged from the definition of a suitable circuit simulator model, through the collection of statistical parametric data, to the generation and verification of the worst-case model sets obtained from this data.

1. Introduction

In today's competitive IC design and manufacturing industries the importance of factors like product time-to-market, cost, yield and reliability has meant that design-for-manufacturability (DFM) [1] issues like the installation of statistical circuit design techniques are essential. At the very least some form of realistic worst-case design procedures should be possible. To achieve this, a statistical parameter extraction approach was formulated and implemented for a 2 μm CMOS technology culminating in the generation of nominal and worst-case model parameter sets. This framework, as it applies to the extraction of DC MOSFET model parameters, will be detailed in this paper.

2. Definition of Circuit Simulator Model

The selection of a suitable MOSFET model and the subsequent definition of parameter extraction schemes appropriate for the collection of large amounts of parametric data as part of in-line process monitor tests is the first stage in successful parametric yield modeling. The model equations utilized were based on a previously published model [2] with certain modifications to allow accurate predictions of device transconductance, output conductance, substrate bias effects, and subthreshold behavior. Techniques for the extraction of model parameters via optimization algorithms were implemented and employed for model validation purposes. However, this parameter extraction procedure with its large measurement requirements, its CPU demands, and the sheer amount of time involved was not suitable for inclusion into any in-line process monitor test framework. In order to achieve parameter extraction in a form appropriate for our application, direct parameter extraction strategies were utilized. Direct parameter extraction requires the measurement of a minimized data set and the generation of parameter values via analytical equation solving techniques formulated for the particular model in use. Direct parameter extraction techniques are fast, accurate, and minimize instances of parameter extraction induced correlations as well as preventing parameters attaining unrealistic or non-physical values. These problems are a characteristic feature of parameter optimization techniques where parameters which have similar influence on device characteristics can interact, or a parameter may be optimized to data measured in a region of operation in which it should have no relevance. Direct parameter extraction methodologies are however not as versatile as their optimization counterparts because knowledge of the model equations is inherently in-built into the procedures and model equation changes normally require significant re-working of these procedures.

The following are the MOSFET model equations for predicting strong-inversion currents. The implementation of subthreshold current is virtually the same as the BSIM1 [3] model. Table 1 lists and describes the relevant model parameters.

Name	Comment	Units
VFB	Flat-band voltage	V
GAMMA (γ)	Bulk threshold parameter	$V^{1/2}$
PHI (ϕ)	Surface Potential	V
LAMBDA (λ)	Channel length modulation	V^{-1}
TOX	Oxide thickness	m
UO	Surface mobility	cm^2/Vs
VMAX	Maximum carrier drift velocity	m/s
THETA (θ)	Mobility modulation	V^{-1}
THETAB (θ_B)	Body effect mobility parameter	V^{-1}
GA2 (γ_2)	Linear GAMMA parameter	-
TH2 (θ_2)	V_{DS} dependent mobility term	m/v
SIGMA (σ)	Static drain feedback	-
SIGMAB (σ_B)	Body effect static drain feedback	V^{-1}
GG	Output conductance parameter	-
N0	Zero bias gate drive coefficient	-
ND0	V_{DS} bias gate drive coefficient	V^{-1}
NB0	V_{BS} bias gate drive coefficient	V^{-1}
NC	Weak inversion fitting parameter	-
LD	Channel length reduction parameter	m
DW	Channel width reduction parameter	m
RDS (RD+RS)	Drain/source resistance parameter	ohm

TABLE 1. MOSFET Model Parameters

94CH3380-3

In the strong-inversion region of operation

$$I_{DS} = \mathrm{UO} \cdot Fg \cdot Fm \cdot COX \cdot \frac{W}{L} \cdot Fv \cdot Fq \cdot V_{DSX} \tag{1}$$

where the low-field mobility (UO) has units $m^2/(Vs)$ and the effective dimensions are given by

$$W = Wdrawn + \mathrm{DW} \tag{2}$$

$$L = Ldrawn - (2 \cdot \mathrm{LD}) \tag{3}$$

In (1)

$$Fg = \frac{1}{1 + \theta \cdot (V_{GS} - \pm\mathrm{VFB}) + \theta_2 \cdot \frac{V_{DS}}{L} + \theta_B \cdot \gamma \cdot (\sqrt{\phi - V_{BS}} - \sqrt{\phi})} \tag{4}$$

$$Fm = 1 + \lambda \cdot (\sqrt{V_{DS} + \mathrm{GG}^2} - \mathrm{GG}) \tag{5}$$

$$Fv = \frac{1}{1 + \frac{V_{DSX} \cdot \mathrm{UO} \cdot Fg \cdot Fm}{\mathrm{VMAX} \cdot L}} \tag{6}$$

and

$$Fq = V_{GS} - VON - \left(1 + \frac{\gamma}{4 \cdot \sqrt{\phi - V_{BS}}}\right) \cdot \frac{V_{DSX}}{2} \tag{7}$$

The threshold voltage is given by the expression

$$VON = VTO + \gamma \cdot (\sqrt{\phi - V_{BS}} - \sqrt{\phi}) - \gamma_2 \cdot (\phi - V_{BS}) - f(V_{DS}) \tag{8}$$

where

$$f(V_{DS}) = \sigma \cdot V_{DS} \cdot (1 + \sigma_B \cdot \gamma \cdot (\sqrt{\phi - V_{BS}} - \sqrt{\phi})) \tag{9}$$

and

$$VTO = \pm\mathrm{VFB} + \phi + \gamma \cdot \sqrt{\phi} \tag{10}$$

The saturation voltage is

$$VSAT = \frac{\mathrm{VMAX} \cdot L}{\mathrm{UO} \cdot Fg \cdot Fm} \cdot \left(\sqrt{1 + \frac{2 \cdot (V_{GS} - VON)}{\frac{\mathrm{VMAX} \cdot L}{\mathrm{UO} \cdot Fg \cdot Fm} \cdot \left(1 + \frac{\gamma}{4\sqrt{\phi - V_{BS}}}\right)}} - 1 \right) \tag{11}$$

$$V_{DSX} = min(V_{DS}, VSAT) \tag{12}$$

In (4) and (10) the (+) sign corresponds to N-channel devices and the (-) sign corresponds to P-channel devices. These equations were found to be adequate for modeling devices from the 2 μm CMOS process used during this work.

3. Development of a Quick MOSFET Parameter Extraction Program (QMOS)

The QMOS parameter extraction program, formulated for the enhanced SPICE level 3 MOSFET model described above, employs direct parameter extraction techniques to derive model parameter values. In particular, these techniques enable the extraction of accurate model parameters in the most efficient manner possible utilizing a minimized device measurement set and no time-consuming parameter optimization procedures. Indeed the motivation behind the creation and implementation of QMOS was that it would, by its very nature, allow the extraction of MOSFET model parameter sets in far less time than any conventional parameter extraction software employing parameter optimization techniques. Thus, QMOS is suitable for implementation into a production environment where it can be utilized to extract complete MOSFET parameter sets on an on-going basis. These parameter sets can serve merely as useful process monitor information relevant to designers or, more importantly, the parametric data can be used as the basis for the development of statistical circuit design or "worst-case" design frameworks.

The following is an example of the methodology used to extract model parameters in a "direct" fashion. In the linear region of operation at low drain-source biases the current equation can be simplified to become

$$I_{DS} = \frac{\beta \cdot V_{DS}}{\theta} \cdot \frac{V_{GS} - (VON + \frac{V_{DS}}{2})}{V_{GS} - (\pm\mathrm{VFB} - \frac{1}{\theta})} \tag{13}$$

Expression (13) can be written in a more general form [4,5] as

$$I_{DSn} = Z_n = \frac{a \cdot (X_n - b)}{(Y - c)} \quad n = 1,2,3 \tag{14}$$

where $X_n = Y_n = V_{GSn}$ and with

$$a = \frac{\beta \cdot V_{DS}}{\theta} = \frac{\mathrm{UO} \cdot COX \cdot W}{L} \cdot \frac{V_{DS}}{\theta} \tag{15}$$

$$b = VON + \frac{V_{DS}}{2} \tag{16}$$

and

$$c = \pm\mathrm{VFB} - \frac{1}{\theta} \tag{17}$$

Thus, three measurements of device current (I_{DS}'s) at three specially chosen biases (V_{GS}'s) in the linear region of device operation can be used in solving three equations in three unknowns to yield values for a, b, and c. This is repeated for three substrate biases nominally 0.0V, -/+2.5V, and -/+5V. A typical biasing arrangement is shown below.

$$\begin{aligned} V_{GS1} &= VON_{approx} \pm 0.4V \\ V_{GS2} &= V_{GS1} + \frac{V_{GMAX} - V_{GS1}}{4} \\ V_{GS3} &= V_{GSMAX} \end{aligned} \tag{18}$$

In (18) V_{GSMAX} is usually +/- 5V. Threshold voltages (VON's) are calculated from (16) and values of b at the three chosen substrate biases. Expressions (8) and (10) are then solved to calculate VFB, GAMMA, and GA2 parameters. The parameters THETA and UO, for each substrate bias, are next extracted from the previously calculated c and a values using (17) and (15) respectively. These parameters in turn enable the calculation of the THETAB parameter. Thus, for a single device, nine measurements of device current are sufficient to extract the VFB, GAMMA, GA2, UO, THETA, and THETAB parameters in a "direct" fashion, (Figure 1).

In the next stage of the direct extraction process the parameters LD and DW, in addition to a value for the drain and source parasitic resistance parameter RDS, are determined from UO and THETA values extracted from devices with different geometries [5]. A minimum of two devices of equal width and different lengths and two devices of equal length and different widths must be involved in this case.

Similarly, direct methods are utilized to extract the N0, NB0, ND0, NC, SIGMA, and SIGMAB parameters from seven current measurements in the subthreshold region of operation, (Figure 2).

A further three measurements of current in the saturation region of operation form the basis for completing the parameter extraction by determining the VMAX and LAMBDA parameters, (Figure 3).

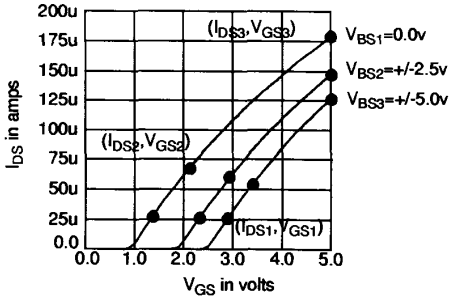

Figure 1: Plot of linear region curves showing the 9 data points used by QMOS.

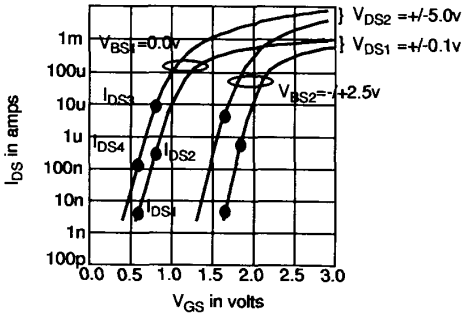

Figure 2: Plot of subthreshold curves showing the 7 data points used by QMOS.

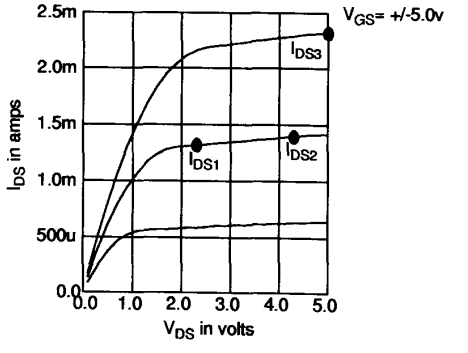

Figure 3: Plot of saturation region curves showing the 3 data points used by QMOS.

Figure 4 shows plots of measured and modeled I-V curves for N-channel and P-channel 20/2 µm devices where the model parameters were extracted using the direct parameter extraction techniques described in this section. The agreement between measured and modeled data is quite reasonable even at very low gate drives, thus validating the use of the direct extraction methodology. Both the measurement of the data required, for the direct parameter extractions and the extractions themselves for five N-channel and five P-channel devices took less than two minutes on the parametric test system on which the software was installed.

Direct parameter extraction is not always an easy task. Problems can arise where the model equations are complicated and/or where it is impossible to decouple certain parameters which characterize similar effects. Care should be exercised during the formulation of a model so as to ensure that fast non-iterative forms of parameter extraction are possible. This does not always happen and the effects of individual model parameters cannot always be separated from each other. In the case of the model employed in this work the direct extraction of linear, subthreshold, and certain saturation region parameters was relatively straight forward. Difficulties were encountered in the extraction of the remaining saturation region parameters and steps had to be taken to simplify the task.

For the devices under analysis it was found, using parameter optimization, that setting the TH2 parameter to zero did not significantly affect the model's performance. Similarly it was determined that the parameter GG could be set to pre-defined values of 0.5 for N-channel devices and 1.0 for P-channel devices. This made it possible to extract the two remaining saturation region parameters, LAMBDA and VMAX using direct extraction methods which would otherwise have proved impossible without the use of unwanted iterative fitting techniques.

In addition, the parameters PHI and TOX were held at process dependent values during the extractions.

4. Generation of Model Parameter Statistics

The parameter extraction methodologies described in the previous section were employed to extract MOSFET model parameter sets over some specially fabricated wafers. Rather than build up a collection of parameter sets over a period of time from the process under investigation, it was decided to extract the required model parameters from a set of wafers which were fabricated under conditions where selected process input variables were intentionally perturbed within the extremes of their expected limits. Some of the process inputs which were varied included gate oxidation temperature, polysilicon gate length, N-well implant dose, P-well implant dose, threshold voltage adjust implant dose, N+ and P+ source/drain implant dose, and well drive-in temperature. In all, over 700 complete model parameter sets were measured, corresponding to the different process splits or perturbations.

Complete model parameter sets, including linear, subthreshold and saturation region parameters, were logged for 20/8 µm, 20/4 µm, 20/2 µm and 2/4 µm N-channel and P-channel devices on each site which was probed. Drain current (I_{DS}), transconductance (g_m) and output conductance (g_{ds}) values for a number of bias points in critical regions of device operation were also logged for these devices.

The measured parametric data was read into a statistical analysis package and merged to form one single table containing both the model parameters and the measured device characteristics for each of the devices under analysis. The parametric data was then analysed in the following manner;

(1) Parameter sets which contained "empty" parameter values i.e. parameters which were outside the limits imposed on them were deleted.

(2) Parameter sets which contained parameter values outside of their current +/- 4σ bounds were deleted.

(3) Parameter means, medians, standard deviations, maximums, minimums, skewness, ...etc. for each parameter were calculated and stored in table form.

(4) Histograms for each of the parameters were created and stored.

(5) The parameter correlation matrix was calculated and stored.

The variabilities of 15 model parameters, measured from linear, subthreshold and saturation regions of device operation, were analysed for each device of a particular polarity and geometry. The remaining, less

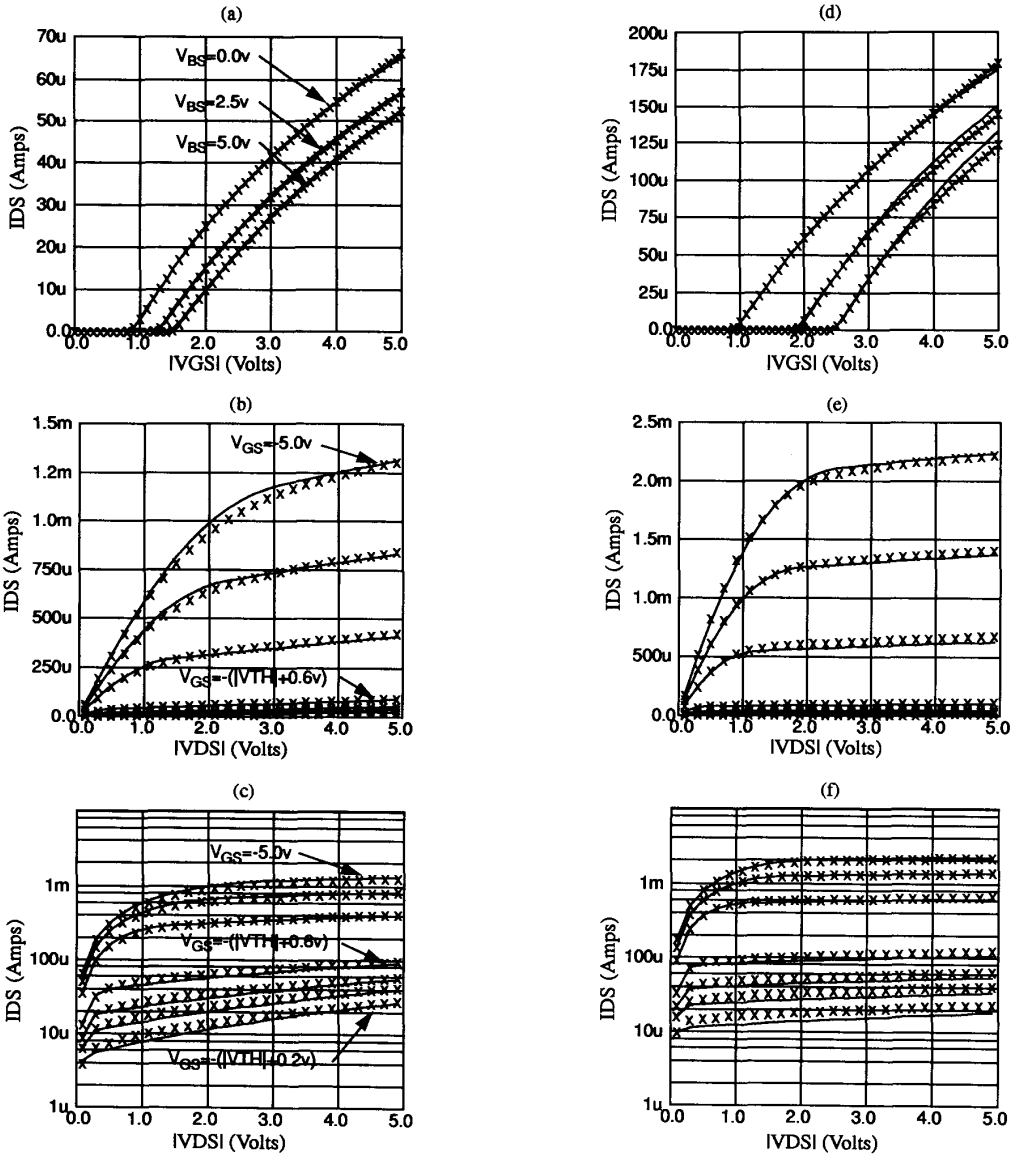

Figure 4: Measured and modeled characteristics for a 20/2μm P-channel device (a), (b) and (c) and N-channel device (d), (e) and (f). QMOS was used to extract the model parameters.

critical, model parameters were effectively excluded from the analysis by either setting them to zero or to pre-determined process dependent values. The N-channel and P-channel parameters, for the devices of each geometry, were combined to form 30-parameter CMOS sets on which the statistical analyses were performed.

5. Generation of Independent Process-Related Factors

The next stage in the construction of worst-case model parameter sets was the transformation of the correlated model parameters (P's) into a much more manageable set of independent process-related factors (X's). The aim of this task was the formulation of relationships in the form

$$P_i = \sum_{j=1}^{m} a_{ij} \cdot X_j \qquad i = 1 \text{ to } n \text{ where } m << n \tag{19}$$

In (19) n is the number of correlated parameters (in this case $n = 30$) and m is the number of uncorrelated components. Equation (19) has been normalized so that both the parameters and the factors have means of zero and unit standard deviations. A denormalized version of (19) is given below

$$P_i = \mu_{P_i} + \sum_{j=1}^{m} \left(\frac{a_{ij} \cdot (X_j - \mu_{X_j}) \cdot \sigma_{P_i}}{\sigma_{X_j}} \right) \quad i = 1 \text{ to } n \text{ where } m << n \tag{20}$$

Principal component analysis (PCA) [6], which is a branch of multivariate statistics, was used as an aid to the derivation of linear equations of the form of (19) between the model parameters and the independent factors. Using PCA techniques to analyse the correlation matrix **R** of the N-channel and P-channel model parameters for devices of any given geometry we were able to deduce that only 6 (m=6) independent factors were required to account for 80% or more of the variability of the 30 original correlated model parameters. PCA also calculates values for the coefficients relating the parameters to the factors (a_{ij}'s). These coefficients lie in the range -1 to +1, values close to +1 or -1 indicate a strong relationship between parameter P_i and factor X_j while values close to zero suggest the opposite. In the interest of making the factors slightly more interpretable a VARIMAX rotation [7] is performed on the derived factors. This ensures that the a_{ij} components are optimised so that they are as close as possible to +/- 1 or 0. By using this PCA-based technique to represent each parameter by a linear combination of uncorrelated factors we should really introduce another term (noise term) into (19) or (20) to allow for the percentage of variability of each parameter not accounted for by the 6 factors as derived. Rather than do this we chose to further adjust the a_{ij} coefficients in (19) or (20) whenever a parameter set is being determined. These coefficients are now

$$\frac{a_{ij}}{\sqrt{\frac{VarP_i}{100}}} \tag{21}$$

Here $VarP_i$ is the percentage variance of parameter P_i accounted for by the 6 components which were retained. As was explained earlier $VarP_i$ will have an average of 80% and in practice will lie in the range 60% to 100% for each of the parameters. Values closer to 60% are likely to occur for the more unimportant empirical parameters. These parameters are thus varying due to factors which are not strictly the same as those causing the variabilities of the other parameters. A large portion of the variability of these parameters may be due to noise induced by measurements or parameter extraction.

6. Interpretation of Process-Related Factors

A combination of PCA and VARIMAX techniques enabled the construction of a system of linear equations relating the 30 CMOS model parameters to 6 independent factors as in (19). In order to get an understanding of the specific process fluctuations causing the individual parameter variances it is necessary to interpret the derived factors X_1 to X_6. By doing this the construction of relationships of the form given in (21), where for example, the first component X_1 may be replaced by a variable describing oxide thickness variation (i.e. $X_1 = X_{TOX}$), are also enabled. Some of the ways by which this was achieved were;

(1) By selecting the model parameters which were most associated with a particular factor. The theoretical physical basis of these parameters may help to identify the factor under analysis.

(2) Some of the parameters were identified by relating factor scores on specific wafers to the process variable settings at which these wafers were processed.

(3) Another way by which the factors can be identified is to correlate the factor variations over the wafers utilized to some process monitor information i.e. polysilicon gate length, resistances, ...etc.

It is important to note that factor scores for any or all wafer sites can be recreated from parameter scores by performing the reverse of equation (19). This was done to all 6 selected factors to aid in their identification. This was done for each wafer site after the PCA and VARIMAX calculations had been performed.

The six independent factors X_1 to X_6, for the 2 µm length devices, have been identified as variations in oxide thickness (X_{TOX}), channel length reduction ($X_{\Delta L}$), threshold voltage adjust implant (X_{VT}), channel width reduction ($X_{\Delta W}$), p-well implant (X_{PW}) and junction depth (X_{XJ}). These independent factors are similar to the smaller set proposed in previous work [8] as been the causes of model parametric variations which would have to be considered in the statistical design of digital circuits. The first three factors each account for approximately 20% of the variance of the total model parameter set. The remaining three factors account for 6%-8% of the parameter variabilities.

7. Construction of Worst-Case Model Parameter Sets

There were 65 model parameter sets in our experiment using the 2 µm length devices (2^m +1 where m = number of factors identified and retained in the analysis). One of these parameter sets is the nominal parameter set while the remaining 64 are achieved by setting the factors to +/-N σ's. In this analysis we found that setting N to 2.5 gave very good results. In many circuit analyses we really only want to utilize 3 model parameter sets, a nominal parameter set and 2 sets enabling the prediction of the circuit upper and lower performance limits. Therefore a procedure to reduce the 64 non-nominal model parameter sets to just the required 3 parameter sets was developed.

In order to select the upper and lower bound parameter sets it was necessary to code the MOSFET model equations into the statistical analysis software package procedure so that the following function could be evaluated for all 64 corner parameter sets.

$$C(P) = \sum_{k=1}^{p} \left(\frac{B_1 \cdot I_{DSk}}{I_{DSnom}} + \frac{B_2 \cdot g_{dsk}}{g_{dsnom}} + \frac{B_3 \cdot g_{mk}}{g_{mnom}} \right) \tag{22}$$

In (22) B_1, B_2, and B_3 are weighting coefficients (usually 0 or 1), and the model equations are employed to calculate the I_{DS}, g_{ds}, and g_m simulated device characteristics at any of the p biases. The nominal characteristics in the denominators in (22) correspond to those simulated with the parameters set to their nominal values (i.e. X's = 0). The parameter sets corresponding to the maximum and minimum instances of $C(P)$ are chosen to be the required two worst-case parameter sets. For the analysis of certain digital circuits B_2 and B_3 may be set to 0 and only one bias point for each device polarity (i.e. p = 2 where $V_{GS} = V_{DS}$ = +/- 5 V and V_{BS} = 0V) may have to be employed in the determination of the $C(P)$ values. Such a simplistic analysis may not be sufficient in analog applications where the device conductances and other device biasing arrangements could not be ignored. The worst-case parameter sets generated in this work were really initially only aimed at digital applications, B_2 and B_3 were set to zero and p was set to 2 (i.e. one bias point for each device polarity). The bias employed was V_{GS} = +/- 3V, V_{DS} = +/- 4V and V_{BS} = 0V. Worst-case parameter sets were extracted separately for 20/2 µm, 20/4 µm, and 20/8 µm devices.

8. Validation of Worst-Case Parameters

Device currents and conductances were measured on the wafers at the same time that the original model parameters were extracted. The measured distributions of these characteristics were compared to the worst-case limits, in an attempt to verify the accuracy of the model parameter sets which were constructed.

The biases utilized were chosen so as to span the biases at which these devices would be used in typical circuit applications. Figure 5 shows measured device current, transconductance and output conductance characteristics for a 20/2 µm N-channel and a 20/2 µm P-channel device. In this example the devices were biased in saturation at a low gate drive. The worst-case predicted limits obtained using the parameter sets generated as described in this paper are also included in these plots. There is excellent agreement between the predicted worst-case limits obtained employing the methodology described in this document and the measured device characteristic distributions. The predicted limits were found to be accurate for devices biased in the linear region of operation, devices biased in the saturation region of operation, devices with non-zero substrate biases, and devices biased at low gate drives.The predicted ranges of operation in all cases were found to be of the order of between six and seven measured standard deviations. Traditional worst-case design methodologies have a

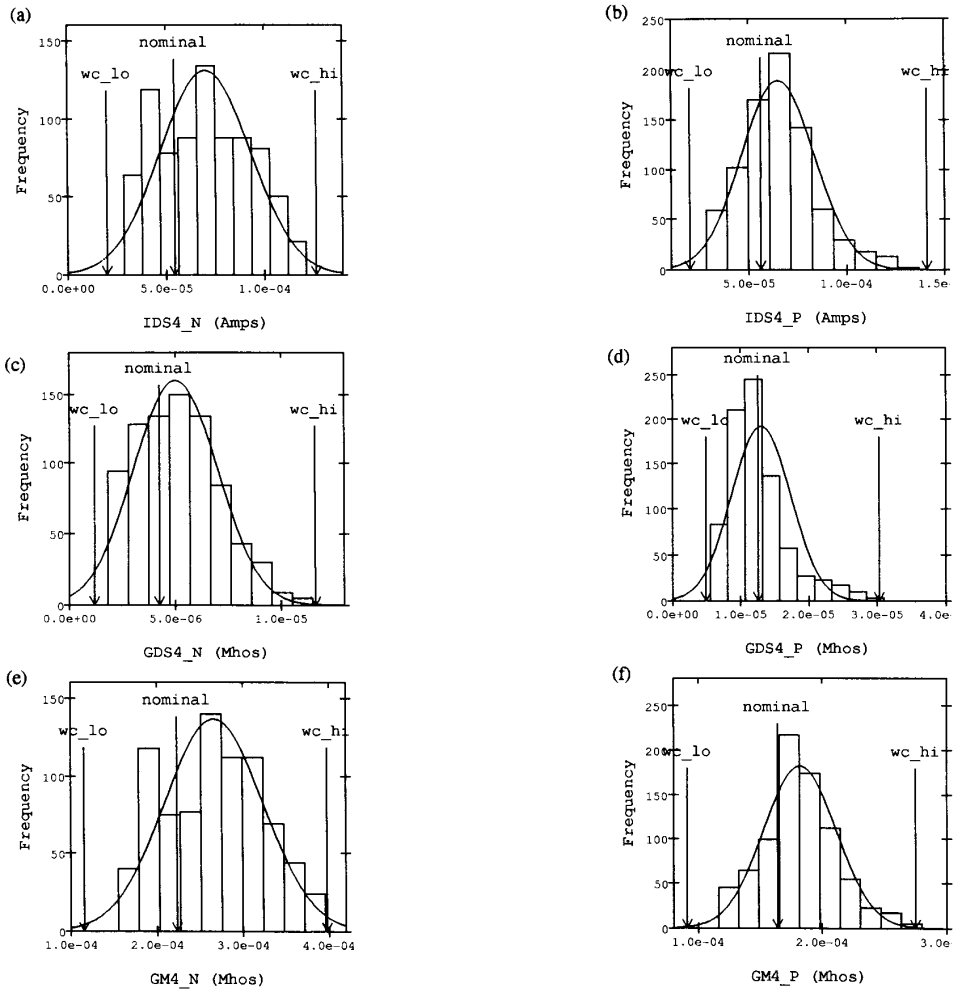

Figure 5: Plots showing the measured distributions and predicted limits for N-channel (a), (c), (e) and P-channel (b), (d), (f) currents and conductances for the 20/2μm devices with a bias of V_{BS} = 0V, $|V_{GS}|$ = 1.3V and $|V_{DS}|$ = 3.0V

tendency of predicting ranges of operation which are far more pessimistic and thus far less useful. Although the worst-case parameters were chosen essentially with digital performance in mind the predicted worst-case limits for the device output conductances and transconductances were also excellent.

9. Conclusions

This paper has described the process by which worst-case model parameter sets were generated for a 2 μm CMOS process. These model parameter sets were constructed from parametric data measured by the QMOS parameter extraction program from specially processed wafers in which key process parameters were varied between expected limits. The method by which the MOSFET model data was analysed and converted to worst-case parameter data has been described and the MOSFET model parametric data variabilities and correlations have been linked to specific sources of variability in the IC manufacturing process. Finally the worst-case parameters which were calculated have been validated utilizing measured device performance data collected over the same wafers from which the model parameter data was extracted.

Acknowledgments

The authors would like to acknowledge Mary Corby, Maureen Hallinan, Denis Doyle and Eamon Hynes of Analog Devices and Kevin McCarthy and Sean Buckley of the NMRC for their valuable contributions during the course of this work.

References

[1] L. Maliniak, Electronic Design, Vol.40, No.22, pp 39-48, October 1992.

[2] G.T. Wright, "A simple and continuous MOSFET model," *IEEE Trans. Electron Devices*, vol. ED-32, No.7, pp. 1259-1263, July 1985.

[3] B.J. Sheu, D.L. Scharfetter, P.K. Ko, and M.C. Jeng, "BSIM: Berkeley short-channel IGFET model for MOS transistors," *IEEE J. Solid-State Circuits*, vol. 22, No. 2, pp. 558-565, Aug. 1987.

[4] M.F. Hamer, "First-order parameter extraction on enhancement silicon MOS transistors," in IEE Proc., vol. 133, Pt. I, No. 2, pp. 49-54, Apr. 1986.

[5] J.A. Power and W.A. Lane, "Enhanced SPICE MOSFET model for analog applications including parameter extraction schemes," *in Proc. IEEE ICTMS'90*, vol. 3, pp. 129-134, Mar. 1990.

[6] J.E. Jackson, *A User's Guide To Principal Components*, Wiley-Interscience, 1991.

[7] R.J. Harris, *A Primer of Multivariate Statistics*, Second Edition, Academic Press, 1985.

[8] P. Yang et al., IEEE Trans. Computer-Aided Design, Vol. 5, pp 5-14, January 1986.