

Reliability-Aware Design Method for CMOS Circuits

Theodor Hillebrand, Nico Hellwege, Steffen Paul, Dagmar Peters-Drolshagen
 Institute of Electrodynamics and Microelectronic (ITEM.me)
 University of Bremen, Germany, +49(0)421/218-62551
 {hillebrand, hellwege, steffen.paul, peters}@me.uni-bremen.de

Abstract—In this paper a reliability-aware design method based on the g_m/I_D -methodology is presented which allows designers of integrated analog circuits to consider process as well as environmental variations and aging effects already at early design stages. Within this method the whole simulation effort is shifted to a single transistor level. With a generated stochastic Look-Up table the small signal characteristics of transistors and circuits can be predicted. Exemplarily, a reliability-aware design for common source amplifiers is shown and the predicted values are compared to those from a traditional simulation showing good data fitting and small deviations.

I. INTRODUCTION

Reliable design flows for integrated CMOS circuits must take into account all effects that influence whole circuits or several single transistors. For larger technology nodes it was sufficient to design a circuit with respect to pure electrical properties like voltages, currents or capacitance. However, the progressive downscaling of CMOS transistors necessitates a design flow that is able to consider both process variations and time-dependent degradation at early stages of the design. Other approaches will lead to an iterative simulation flow once an initial design has been found. The drawback of a design flow that depends on intensive simulations at system level is that for example aging simulations on this level are very time consuming or even unfeasible. Therefore a methodology is needed that enables designers to consider all effects on scaled down devices and does not evoke complex simulations at system level.

The two most important aging mechanisms in CMOS transistors are Bias Temperature Instability (BTI) and Hot Carrier Degradation (HCD) [1]. For both effects exist several different models that can explain some of the observed phenomena linked to these mechanisms. Nevertheless, the physics behind BTI and HCD are yet not fully understood. The main impact of those aging influences is an increase of the charge under the gate oxide of transistors causing higher threshold voltage V_{th} , higher $R_{ds,on}$ and reduced mobility $\mu_{n/p}$. Simply expressed, the transistor becomes slower.

This work is based on the g_m/I_D -methodology that was firstly introduced by [2]. The main advantage of this design methodology is its capability to consistently describe the behavior of CMOS transistors over all regimes of operation. Thus, designers are able to contrive circuits in which all transistors are operating in moderate inversion. This operating point is the best trade-off between speed and power consumption. However, in this operating region transistors are most badly affected by HCD and BTI which is shown in figure 1.

In order to ensure reliable low power circuits that are

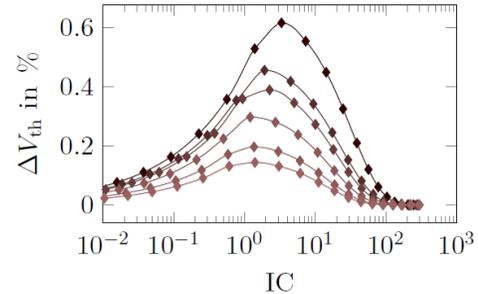


Fig. 1. Deviation of threshold voltage V_{th} due to HCI of NMOS transistors over a period of 3 Ms versus the inversion coefficient IC for different lengths L .

resistant towards process variations the well known g_m/I_D -methodology has to be enhanced [3].

The presented **Reliability-AwaRE (RARE)** design flow is based on stochastic Look-Up-Tables (LUT) including all necessary variables for the g_m/I_D -methodology. The LUT contains the distribution functions of these required parameters and their particular behavior over time for 3 Ms (≈ 10 years). The LUT entries are generated through transient simulations of 2000 P- and NMOS transistors utilizing a modified BSIM6 [4] transistor model presented in [5]. Subsequently, the extracted distributions can be used in a modified g_m/I_D -design flow and, furthermore, a direct calculation of small signal parameters can be carried out. This new approach shifts the simulation effort from the system to the single-transistor level and realizes a reliability-aware design method for CMOS circuits.

II. THE g_m/I_D -METHODOLOGY

The main objective of the g_m/I_D -methodology is describing the operating point of MOS transistors with w/L -independent metrics. In order to achieve this goal the so called transconductance efficiency g_m/I_D and the inversion coefficient IC are used. Where

$$\frac{g_m}{I_D} = \frac{1}{I_D} \frac{\partial I_D}{\partial V_{GS}} = \frac{\partial \ln(I_D)}{\partial V_{GS}} = \frac{\partial \ln\left(\frac{I_D}{w/L}\right)}{\partial V_{GS}} \quad (1)$$

and

$$IC = \frac{I_D}{I_0 \cdot (w/L)} \quad (2)$$

with $I_0 = 2n\mu C_{ox} V_t^2$ being a technology specific current. Due to the fact that $I_D \propto w/L$ both metrics are w/L -independent. Thus, the operating point of any MOS transistor can be described by the tuple

$$\mathbb{T}_{OP} = \{IC, I_D, L\} . \quad (3)$$

By choosing an operating point, e.g. in moderate inversion, the IC is set, which leads to I_D , further L should be chosen large enough to prevent short channel effects. According to equation (2) the width can be calculated as

$$W = \frac{I_D \cdot L}{I_0 \cdot IC} . \quad (4)$$

To improve the accuracy of this method Look-up tables are used because the simulated data considers additional effects which may be included in specific transistor models and manufacturing processes. The transconductance is then derived from

$$g_m = \left(\frac{g_m}{I_D} \right)_{\text{sim}} \cdot I_D . \quad (5)$$

Utilizing this design method the operating point of every transistor in a circuit can be derived and furthermore small signal performances of the circuit can be directly analyzed.

III. GENERATION OF STOCHASTIC LUTS

Due to the fact that within the mentioned g_m/I_D -methodology every transistor is designed separately, it is sufficient for the LUT to analyze the behavior of one PMOS and one NMOS transistor. The only parameters that have to be swept, are the length L and the gate-source voltage V_{gs} .

$$\begin{aligned} L_{min} &\leq L \leq 3 \times L_{min} \\ 0 &\leq V_{gs} \leq V_{dd} = 1 \text{ V} \end{aligned} \quad (6)$$

Within this work a 65 nm CMOS technology is considered the maximum length of which is confined to $3 \times L_{min}$. For every specific sweep point a 3Ms transient simulation with 2000 transistors affected by process and temperature variation is performed. Table I shows the normal distribution functions of all considered parameters with their particular mean value μ and standard deviation σ . The values for μ and σ correspond to those which are specified by the technology.

TABLE I. PARAMETERS OF NORMAL DISTRIBUTION FUNCTIONS OF ALL CONSIDERED QUANTITIES

Parameter	Symbol	μ	σ
Temperature	T	27°C	$\sqrt{2}^\circ\text{C}$
Gate oxide thickness	t_{ox}	$t_{ox, norm}$	7 pm
Length	L	L_{spec}	2 nm
Width	W	W_{spec}	2 nm
Gate doping concentration	N_{gate}	$N_{gate, norm}$	$5 \cdot 10^{22} \text{ m}^{-3}$
Channel doping concentration	N_{dep}	$N_{dep, norm}$	$1 \cdot 10^{21} \text{ m}^{-3}$
Source / Drain doping concentration	$N_{S/D}$	$N_{S/D, norm}$	$1 \cdot 10^{23} \text{ m}^{-3}$

Figure 2 shows the test benches for NMOS and PMOS transistors.

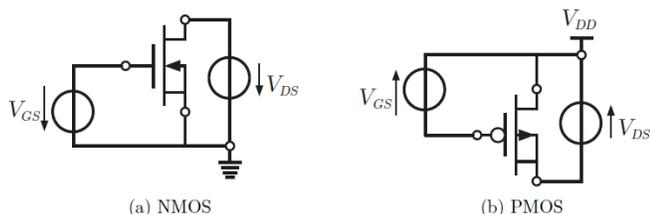


Fig. 2. Test benches for NMOS (a) and PMOS (b) transistors.

The simulation data is fitted either with a normal $f_n(\mathbf{x}, t)$ (7) or with a log-normal $f_{ln}(\mathbf{x}, t)$ (8) distribution depending on which one describes the basic population best. Two different distribution functions are required because some distribution functions are symmetrical and some possibly not.

$$f_n(\mathbf{x}, t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{\mathbf{x} - \mu}{\sigma}\right)^2\right) \quad (7)$$

$$f_{ln}(\mathbf{x}, t) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma\mathbf{x}} \exp\left(-\frac{1}{2}\left(\frac{\ln(\mathbf{x}) - \mu}{\sigma}\right)^2\right) & \mathbf{x} > 0 \\ 0 & \mathbf{x} \leq 0 \end{cases} \quad (8)$$

\mathbf{x} denotes any fitted parameter (e.g. IC , g_m) and t is the time. These two functions were chosen because both have two fitting parameters (μ , σ) with the same meaning and this simplifies the calculus of small signal parameters. The time dependency comes in with the fitting of μ and σ , as a matter of simplicity, with a fourth order polynomial function over time which gives

$$\mu(t) = p_{1,\mu}t^4 + p_{2,\mu}t^3 + p_{3,\mu}t^2 + p_{4,\mu}t + p_{5,\mu} \quad (9)$$

and

$$\sigma(t) = p_{1,\sigma}t^4 + p_{2,\sigma}t^3 + p_{3,\sigma}t^2 + p_{4,\sigma}t + p_{5,\sigma} . \quad (10)$$

The numerical analysis is performed in MATLAB. The simulation data is reduced from 16 GB to approximately 240 MB containing the same amount of information. The data reduction can be realized by storing the coefficient of distribution and fitting functions instead of storing the whole data. The second advantage is that this simulation has to be performed only once for every technology. Moreover, the resulting time-dependent distribution functions for the parameters can subsequently be used in the g_m/I_D -methodology which is shown in the next section.

IV. CALCULATION OF SMALL SIGNAL PARAMETERS

In order to perform the calculation of small signal parameters for MOS transistors as well as circuit performances like the DC gain A_{DC} of common source amplifiers, some calculation specification has to be performed in a different manner because random variables are considered. The inversion coefficient IC is given by equation (2) where I_D , I_0 , W and L are partially independent random variables denoted with bold print. An important fact is that some parameters are not stochastically independent, thus the covariance is

$$\text{cov}(\mathbf{x}, \mathbf{y}) \neq 0 . \quad (11)$$

In common design flows the technology current I_0 is treated as a constant value for each technology. However, in this case this current is given by $I_0 = 2n\mu C_{ox} V_t^2$ which is not a constant value because each parameter is affected in a different way by the variations specified in table I. Table II shows the interdependencies of I_0 and the process/environmental variations.

TABLE II. INTERDEPENDENCIES OF I_0 AND THE PROCESS/ENVIRONMENTAL VARIATIONS

Parameter	Main Dependencies
\mathbf{n}	$T, N_{dep}, N_{gate}, t_{ox}$
μ	T
\mathbf{C}_{ox}	t_{ox}
\mathbf{V}_t	T

For the technology current follows

$$\mathbf{I}_0 = K \left(\frac{1}{\sqrt{2\pi}} \right)^N \prod_{i=1}^N \left(\frac{1}{\sigma_i x_i} \right) \cdot \dots \exp \left(-\frac{1}{2} \sum_{i=1}^N \left(\frac{z_i - \mu_i}{\sigma_i} \right)^2 \right) \cdot \exp \left(-\frac{1}{2} \left(\frac{z_T - \mu_T}{\sigma_T} \right)^2 \right) \quad (12)$$

where $z_i = \ln x_i \vee z_i = x_i$ depending on whether the variable is log-normal or normal distributed. K contains all constant factors and N is the number of independent random variables. The correlation coefficients are neglected because $r_{ij} = 0$ for \mathbf{I}_0 e.g. the change in temperature is uncorrelated to a change in doping concentration and vice versa. The resulting distribution is a normal distribution if all contributing distributions are normal distributions, and it is log-normal distributed if at least one variable follows a log-normal distribution. In order to calculate IC and g_m/I_D two random variables have to be divided and the resulting distribution must be extracted.

Therefore, let

$$\mathbf{z} = \frac{\mathbf{x}}{\mathbf{y}} \Leftrightarrow g_m/I_D = \frac{\mathbf{g}_m}{\mathbf{I}_D} \quad (13)$$

then the problem is solved for $\mathbf{x} = \mathbf{z} \cdot \mathbf{y}$. The resulting distribution is given by

$$f_z(\mathbf{z}) = \int_0^\infty \frac{\mathbf{y}}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} \exp \left[-\frac{1}{2(1-r^2)} \dots \left(\frac{(\mathbf{y}\mathbf{z})^2}{\sigma_x^2} - \frac{2r\mathbf{z}\mathbf{y}^2}{\sigma_x\sigma_y} + \frac{\mathbf{y}^2}{\sigma_y^2} + h(\mathbf{y}, \mathbf{z}) \right) \right] d\mathbf{y} \quad (14)$$

The function $h(\mathbf{y}, \mathbf{z})$ contains all terms considering the mean values of the two distributions and $r = \text{cov}(\mathbf{x}, \mathbf{y})/\sigma_x\sigma_y$. Equation (14) applies for both the calculation of IC and g_m/I_D because the change in \mathbf{I}_D is correlated to \mathbf{I}_0 and \mathbf{g}_m is correlated to \mathbf{I}_D . All calculated distributions are time-dependent because μ_i and σ_i are. Therefore the correlation coefficient r is also time-dependent and

$$\lim_{t \rightarrow \infty} r_{ij} = r_{ij,0} \quad (15)$$

holds, because all aging effects will saturate at a certain point in time and then the distribution parameters will become time-independent. This is due to the fact that every operating point has a certain maximum amount of traps that can be generated and occupied. The mentioned time dependency necessitates a repeated evaluation of these parameters for several different time steps.

With the calculation of these distributions the LUT contains all necessary entries for each sweep point to be used within the g_m/I_D -method. Figures 3, 4 and 5 show exemplarily the distribution of g_m , I_D as well as g_m/I_D . The distribution of

g_m/I_D shows a unique behavior over time because g_m and I_D are unequally affected by aging and this leads to a compression of the g_m/I_D distribution.

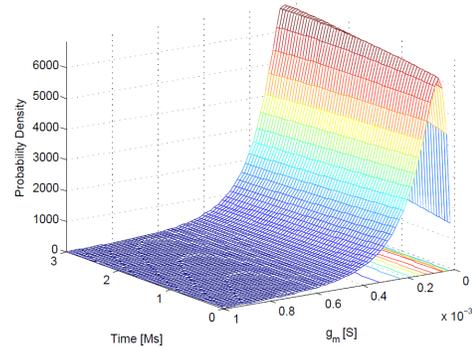


Fig. 3. Distribution function of transconductance g_m of NMOS transistors with $L = 130$ nm and $V_{gs} = 0.75$ V over 3 Ms.

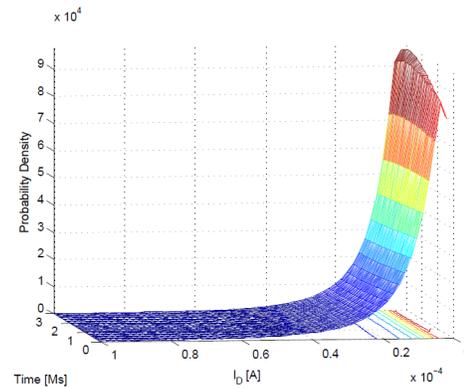


Fig. 4. Distribution function of drain current I_D of NMOS transistors with $L = 130$ nm and $V_{gs} = 0.75$ V over 3 Ms.

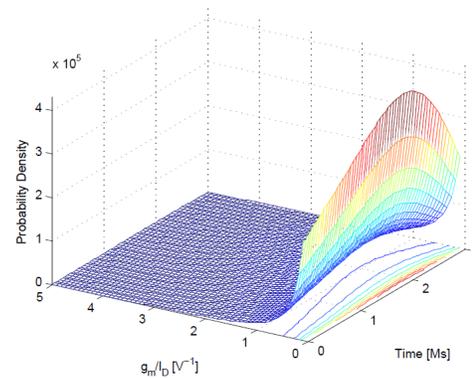


Fig. 5. Distribution function of the calculated transconductance efficiency g_m/I_D of NMOS transistors with $L = 130$ nm and $V_{gs} = 0.75$ V over 3 Ms.

V. COMMON SOURCE AMPLIFIER

In this section a common source amplifier (CS-Amp) which is shown in figure 6 is designed with the g_m/I_D -method.

Subsequently, the DC gain A_{DC} of the CS-Amp is derived with the small signal characteristics of the particular

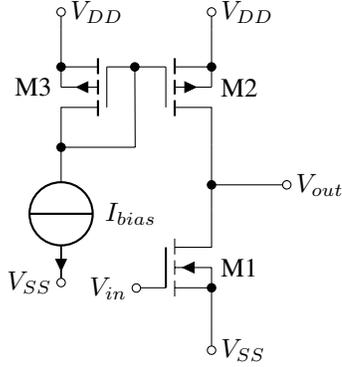


Fig. 6. Common source amplifier with active PMOS current mirror load.

MOSFETs. The calculated distribution function of A_{DC} is compared to the distribution of simulated results for the CS-Amp in CADENCE VIRTUOSO. Table III shows the chosen value for the three transistors.

TABLE III. APPROPRIATE OPERATING POINT PARAMETER VALUES FOR TRANSISTORS IN THE CS-AMP

Parameter	Value	Parameter	Value
W_1	260 nm	IC_1	5
W_2	260 nm	IC_2	50
W_3	2.6 μm	IC_3	50
L_1	130 nm		
L_2	130 nm	I_D	100 μA
L_3	130 nm	I_0	995 nA

As it can be seen, the amplifier is designed in a way that M1 operates in moderate inversion and the two transistors forming the current mirror operate in strong inversion.

The DC gain can be approximated by

$$A_{DC} = \frac{\left(\frac{g_{m,1}}{I_D}\right)_{sim}}{\left(\frac{g_{ds,1}}{I_D}\right)_{sim} + \left(\frac{g_{ds,2}}{I_D}\right)_{sim}} = \frac{g_{m,1}}{g_{ds,1} + g_{ds,2}}. \quad (16)$$

The calculation of A_{DC} follows the same scheme like that for IC and g_m/I_D . All conductances are time-dependent, therefore the expression for A_{DC} already includes the impact of aging and process/environmental variations. Moreover, it is possible to evaluate which is the most crucial parameter for both effects at this very early design stage. Table IV shows a comparison of predicted fresh/aged A_{DC} and those values simulated in CADENCE. The initial amplification as well as the impact due to aging is overestimated because the prediction neglects the load resistance.

TABLE IV. COMPARISON OF MEAN PREDICTED AND SIMULATED FRESH/AGED A_{DC} .

Fresh	Value	Aged	Value
$A_{DC,pre}$	93.1 dB	$A_{DC,pre}$	13.7 dB
$A_{DC,sim}$	82.3 dB	$A_{DC,sim}$	19.12 dB

Figures 7 and 8 show the distribution for the simulated $g_{m,1}$ as well as the relative error between the simulated and calculated distribution. Around the mean value $\mu_{g_{m,1}} = 1$ mS the relative error is less than 3%. For lower and higher values the error increases, but in order to fit this region another distribution from the LUT has to be utilized. A_{DC} cannot be

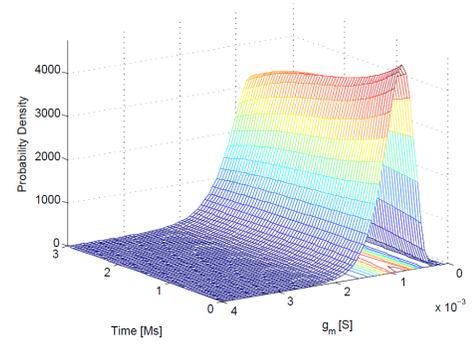


Fig. 7. Distribution function of simulated $g_{m,1}$ of a common source amplifier over 3 Ms.

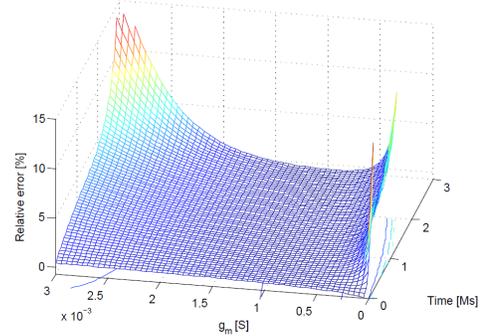


Fig. 8. Relative error between the calculated and the simulated distribution function of $g_{m,1}$.

plotted because it is a 5-dimensional function. Therefore, only the error for $g_{m,1}$ is shown, but the shown deviation applies also for $g_{ds,1}$ and $g_{ds,2}$.

Additionally, the shown calculation can be carried out for every small signal circuit performance.

VI. CONCLUSION

In this paper a design method is presented that allows designers to consider process as well as environmental variations and aging effect at early design stages. Moreover, the method shifts the simulation effort to a single transistor level by generating a stochastic Look-Up Table. This leads to reduced simulation time and amount of data that must be stored. With this data good predictions for small signal parameters of MOSFET and circuit performances can be derived. The method is evaluated with the design of a common source amplifier showing only small errors in the prediction of the initial distribution as well as for the degraded distribution.

REFERENCES

- [1] Maricaeu, Elie, and Georges Gielen. Analog IC Reliability in Nanometer CMOS. Springer Science & Business Media, 2013.
- [2] Enz, Christian C., and Eric A. Vittoz. Charge-based MOS transistor modeling. John Wiley & Sons, 2006.
- [3] Hellwege, Nico, et al. Using operating point-dependent degradation and g m/I D method for aging-aware design. *Integrated Reliability Workshop Final Report (IRW), 2013 IEEE International. IEEE*, 2013. S. 113-116.
- [4] Agarwal, et al. BSIM6.1.0 MOSFET Compact Model. University of California, 2014.
- [5] Hillebrand, et al. Stochastic analysis of degradation and variations in CMOS Transistors, ZuE 2015.