Circuit-level Mismatch Modelling and Yield Optimization for CMOS Analog Circuits

Mingjing Chen and Alex Orailoglu
CSE Department, UC San Diego
La Jolla, CA 92093, USA
{mjchen,alex}@cs.ucsd.edu

Abstract

A methodology for constructing circuit-level mismatch models and performing yield optimization is presented for CMOS analog circuits. The methodology combines statistical techniques with direct investigation of circuit behavior, and achieves model simplification and computational efficiency while ensuring sufficient accuracy. The circuit-level mismatch model can be used in performance characterization and yield estimation, both important in providing information for circuit reliability analysis. The proposed yield optimization technique consists of constructing and refining a yield model over the designable parameters, and ensures fast convergence to the global optimal design. The experimental results on two representative circuits confirm the efficiency and effectiveness of the proposed method.

1. Introduction

With the fast scaling of IC fabrication technology in the past decade, device mismatch caused by process variations has become a highly challenging problem for the design and analysis of CMOS analog circuits. This trend can be attributed to two important factors. First, the structure of matched transistor pairs is widely used in most CMOS analog designs. Such a structure is robust with respect to parameter deviations of the corresponding transistor pair in the same direction, while being very sensitive to parameter deviations in opposing directions (mismatch) [14]. Second, statistically, the magnitude of mismatch increases along with the shrinkage of device feature size as process control in the manufacturing phase fails to improve at a rate comparable to scaling [12, 1]. As a result, identically designed CMOS analog circuits are exhibiting a large spread in performance metrics causing severe performance degradation and even yield loss. For example, even a few millivolts of mismatch-induced offset voltage can result in zero yield for high-precision analog-to-digital converters [13].

Mismatch-concerned design optimization is becoming increasingly important for achieving a high-precision and low-cost analog design. Two levels of such optimization techniques can be developed based on established work on transistor-level mismatch modelling [12] and characterization [1, 13]. The first level is the construction of circuit-level mismatch models which quantitatively predict the influence of mismatch on circuit performance. Such models enable fast performance characterization and yield estimation, and thus can take on the role of Monte-Carlo simulation which is no longer applicable to complex circuit design due to its high simulation cost. The second level consists of yield optimization in early design stages, which is essential for improving circuit reliability and for reducing product cost.

Due to the complex nonlinear relationship between performance distribution and process variability in CMOS analog circuits, simulation-based statistical techniques are usually desirable for both levels of the aforementioned design optimizations. Yet a typical challenge for these techniques is the *curse of dimensionality*. The accuracy generally improves with larger amounts of simulations, which, on the other hand, is bounded by the simulation cost limitation. The solution to this dilemma calls for appropriate model design and simplification techniques. Even greater challenges stem from the rapid change in fabrication technologies which have introduced new random or systematic variation-generating mechanisms, such as highly correlated random parameters or systematic spatial channel length variations. Consequently, the choice and composition of appropriate mathematical techniques for emerging technology constraints is thrown widely open to question once again at this point. Mathematical innovations such as the above, meanwhile, must be matched by thorough examinations of the dominant physical effects so as to reduce complexity.

Considerable research has been conducted for the analysis of process variability and yield-aware analog design. Methods based on line search iterations involve aggressive search for the point of higher yield along some specific direction such as yield gradient or the coordinate of the design space [14]. Such techniques face the problem of finding the optimal search direction which is quite difficult. Moreover, the greedy algorithm or simulated annealing algorithm usually employed in the search procedure may lead to a local optimal result. Some methodologies formulate yield optimization as a multi-objective optimization problem and propose techniques of generating the yield-aware pareto fronts [16, 5]. Though such techniques provide more insight on the trade-off among design criteria, it is difficult to map multiple objectives directly to yield when complex correlation mechanisms among the objectives exist. Other techniques include analyzing a diagnostic circuit to find the yield factors that can be improved [2], which requires chip fabrication and thus can not be used in early design iterations due to the attendant high fabrication cost.

1-4244-1258-7/07/$25.00 ©2007 IEEE

In this paper, a generic two-stage methodology is proposed to (1) construct circuit-level mismatch models which help predict the influence of mismatch on circuit performance and yield; and (2) perform yield optimization using a yield model constructed based on the yield estimation results of circuit-level mismatch models. The correlation information between mismatch parameters is processed without loss in the first stage using statistical techniques such as Mahalanobis Transformation [8]. Both the mismatch model and the yield model are constructed using the Response Surface Method [9]. Techniques of identifying critical model factors are developed to reduce model complexity as well as the simulation cost for model construction. The two parts of the proposed work are both compatible with SPICE models, and can be used either individually or jointly in the early design stages to guide the design iterations.

This paper is organized as follows. Section 2 presents some preliminaries. The circuit-level mismatch modelling and yield optimization techniques are investigated in Sections 3 and 4, respectively. Section 5 further discusses the technical feasibility of the proposed method. Section 6 presents experimental results on two representative circuits. A set of conclusions is drawn in Section 7.

2. Preliminaries

MOS transistor mismatch is generally represented by threshold voltage mismatch ΔV_{th}, current factor mismatch $\frac{\Delta\beta}{\beta}$ and body factor mismatch $\Delta\gamma$; the first two can be easily characterized using drain current mismatch measurement on test chips [1, 10]. A widely accepted and experimentally verified model for these random variations is a normal distribution with zero mean and a variance dependent on transistor size and mutual distance [12]. The proposed methodology considers ΔV_{th} and $\frac{\Delta\beta}{\beta}$ as the major mismatch factors, as they represent the mismatch of parameters that determine the first-order behavior of transistors. Another reason is that the extra mismatch caused by $\Delta\gamma$ can be considered as a further degradation of the V_{th} matching [10].

Mismatch at different locations is usually considered independent [14]. Within each transistor pair, ΔV_{th} and $\frac{\Delta\beta}{\beta}$ have been previously considered independent because of their low correlation [12, 1]. However, recent research shows that ΔV_{th} and $\frac{\Delta\beta}{\beta}$ might be significantly correlated under some fabrication techniques, such as halo implantation which is widely used in deep-submicron devices to suppress the short channel effects [4]. Figure 1 shows a sample distribution of ΔV_{th} and $\frac{\Delta\beta}{\beta}$ with halo implantation, where the correlation coefficient reaches 0.72. This fact implies that correlation must be considered in circuit-level mismatch modelling. Principle Component Analysis (PCA) [7] has been employed to reduce the number of mismatch parameters as well as eliminate the correlation among them [11, 6]. But PCA-based methods were developed at the time of $2\mu m$ technology and fail to capture the dynamic and wide-varying nature of deep-submicron processes [3]. Since research suggests that it is common that only a few unmatched transistor pairs contribute the most to the performance deviation [14], the proposed method reduces the number of mismatch factors by considering only the mismatch sources that have statistically significant impact on circuit performance, and decorrelates mismatch factors using Mahalanobis Transformation. In this way, the error incurred by model simplification is much more controllable.

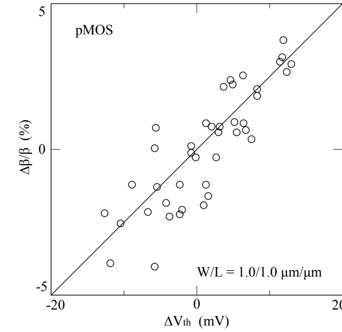

Figure 1. Correlation between ΔV_{th} and $\frac{\Delta\beta}{\beta}$ with halo implantation

As we have mentioned, mismatch is the dominant factor of yield loss in CMOS analog circuits. Theoretically, yield is defined by the following function:

$$Y(D) = \int_{R_T} \phi(D, \delta)\psi(\delta)d\delta \quad (1)$$

where D and δ are respectively the set of designable parameters and the set of random variations which are dominated by mismatch in CMOS analog circuits; $\phi(D, \delta)$ is the probability density function of δ corresponding to design D; R_T is the tolerance region of random variation; $\psi(\delta)$ is an indicator function which maps δ to 1 if $\delta \in R_T$ and to 0 otherwise. For realistic circuits, it is difficult to determine the analytical form of R_T. Thus yield is usually approximated as the result of direct sampling. Specifically, it can be computed as the number of chips whose performance values satisfy the complete set of specifications, divided by the total number of chips in the lot [14]. The yield optimization technique proposed in this paper performs yield estimation according to this realistic definition.

The proposed method can be outlined as follows. For each transistor pair possibly with potential mismatch, a small set of mismatch combinations (ΔV_{th}, $\frac{\Delta\beta}{\beta}$) are generated according to their distributions obtained by mismatch characterization. Each mismatch combination is inserted individually into the transistor pair and a performance deviation value is obtained by simulation. A metric is defined to “combine” all deviation values so that the impact can be evaluated for each transistor pair considered as a mismatch source. The mismatch-critical transistor pairs are then determined and their ΔV_{th}'s and $\frac{\Delta\beta}{\beta}$'s are considered as factors of the circuit-level mismatch model. For every performance parameter, an analytic model in terms of these factors can be constructed using the Response Surface Method after factor decorrelation. For any specific design, yield estimation can be performed using the circuit-level mismatch model given the statistical characteristics of mismatch, based on which

a yield model can be constructed in the design space. The global yield-optimal design can be achieved by refining the yield model for a small number of iterations.

3. Circuit-level Mismatch Model

The circuit-level mismatch model creates a quantitative relationship between the performance parameters and a small set of mismatch to which the performance is highly sensitive (i.e., the set of mismatch-critical transistor pairs).

3.1. Critical Transistor Pair Identification

As defined in [12], mismatch is the process that causes time-independent random variations in physical quantities of identically designed devices. Thus there is no need to examine all possible transistor pairs in the circuit. The proposed method groups all transistors according to their geometry dimensions. Each pair of transistors within the same group is considered as a mismatch source candidate and added into the candidate set S. Since the number of transistors with the same dimensions is usually far less than the total number of transistors in practical circuits, this strategy can reduce the size of the candidate set S to a great extent.

For any mismatch source candidate M_n in S, where $1 \leq n \leq |S|$, a small set of k mismatch combinations $(\Delta V_{th}, \frac{\Delta\beta}{\beta})^{(i)}$ $(1 \leq i \leq k)$ can be generated as follows

$$\Delta V_{th}^{(i)} = \sigma[\Delta V_{th}] \cdot x_1^{(i)} \quad (2)$$

$$\frac{\Delta\beta}{\beta}^{(i)} = \sigma[\frac{\Delta\beta}{\beta}] \cdot (r_{v\beta} \cdot x_1^{(i)} + \sqrt{1 - r_{v\beta}^2} \cdot x_2^{(i)}) \quad (3)$$

where $\sigma[\Delta V_{th}]$ and $\sigma[\frac{\Delta\beta}{\beta}]$ denote standard deviations of ΔV_{th} and $\frac{\Delta\beta}{\beta}$, and $r_{v\beta}$ denotes the correlation coefficient between them. These values can be obtained through characterization. $x_1^{(i)}$ and $x_2^{(i)}$ are sample values respectively generated from two independent normal distributions both with zero mean and a variance of 1/2. Since all the original statistical characteristics of the characterized data are preserved, this set of mismatch combinations accurately represents the practical mismatch distributions.

The k mismatch combinations are individually inserted into the corresponding transistor pair and simulations are performed k times. For any specific performance parameter P of interest, k values, $P^{(j)}[M_n]$ $(1 \leq j \leq k)$ can be obtained through simulation. Then the influence of the corresponding mismatch source candidate on performance P can be evaluated with the metric defined by equation (4)

$$\Delta P[M_n] = \frac{1}{k} \cdot \sum_{j=1}^{k} (P^{(j)}[M_n] - P_{nominal})^2 \quad (4)$$

where $P_{nominal}$ is the nominal value of performance parameter P. This metric reflects the average spread of performance P due to the mismatch of transistor pair M_n.

For each M_n in S, its corresponding $\Delta P[M_n]$ can be calculated. Then the set Z of mismatch-critical transistor pairs for performance P can be determined by examining all elements of S in descending order of ΔP and including the ones whose impact exceeds a ratio of the sum of the impact of the remaining unexamined ones. Formally all transistor pairs that are included in Z satisfy the following equation

$$\Delta P[M_l] \geq \varepsilon \cdot \sum_{M_h \in S} \Delta P[M_h], \quad \forall M_l \in Z \quad (5)$$

ε is a predefined ratio dependent on the level of targeted accuracy. Since the mismatch influence of critical transistor pairs is usually much greater than that of the non-critical ones, the differentiation result is quite insensitive to the value of ε once it exceeds some threshold value. Therefore a high value for ε is perfectly adequate in screening out from the candidate set the transistor pairs whose mismatch has dominant effects on the performance.

3.2. Model Construction

A quantitative model predicting the value of performance parameters of interest in terms of mismatch can further be constructed using the Response Surface Method (RSM) [9], which models the response parameter through a polynomial representation of several influencing factors. The basic idea of model construction with RSM consists of designating design points (the sample values of factors in the experimentation), performing experimentations at design points and calculating model parameters from the results of experimentation using regression techniques.

In the proposed methodology, the performance parameter of interest is considered as the response parameter. The mismatch parameters ΔV_{th} and $\frac{\Delta\beta}{\beta}$ of every transistor pair in set Z obtained in Section 3.1 are considered as the influencing factors since they are dominant with respect to the performance deviation of the circuit. Usually, a first order (linear) model or a second order (quadratic) model is preferred in RSM. Since the influence of mismatch on performance is often approximately symmetric to the neutral line (i.e., the set of points where the effects of the variations of two transistors cancel each other, resulting in no mismatch) [14], a quadratic model approximates practical realities sufficiently closely and accurately, while being computationally affordable. We employ it in the proposed method, therefore.

The positions of design points are usually chosen symmetrically to the design center (the mismatch-free point in this specific case), so that an orthogonal design is obtained to ensure an unbiased model. In order to make such a design in accordance with the practical distributions of influencing factors, it is desirable to have all influencing factors independent of each other. As mentioned in Section 2, mismatch parameters at different locations are considered independent, but the correlation between ΔV_{th} and $\frac{\Delta\beta}{\beta}$ of the same transistor pair can not be ignored. Thus for each element in Z, its ΔV_{th} and $\frac{\Delta\beta}{\beta}$ need to be transformed to two new factors independent of each other, before they are used in model construction. Mahalanobis Transformation [8] is employed to perform this task, whose steps are concisely presented as follows. For any transistor pair, the variance-covariance matrix Σ of its ΔV_{th} and $\frac{\Delta\beta}{\beta}$ can be easily generated from characterization, as shown in equation (6).

$$\Sigma = \begin{bmatrix} \sigma^2[\Delta V_{th}] & r_{v\beta} \cdot \sigma[\Delta V_{th}] \cdot \sigma[\frac{\Delta\beta}{\beta}] \\ r_{v\beta} \cdot \sigma[\Delta V_{th}] \cdot \sigma[\frac{\Delta\beta}{\beta}] & \sigma^2[\frac{\Delta\beta}{\beta}] \end{bmatrix} \quad (6)$$

As a symmetric matrix, Σ can be written as equation (7) using Spectral (Jordan) Decomposition.

$$\Sigma = \Gamma\Lambda\Gamma^T \quad (7)$$

$\Lambda = diag(\lambda_i)$ is a diagonal matrix with its diagonal elements being the eigenvalues λ_i of Σ. Γ is an orthogonal matrix whose columns are the normalized eigenvectors of Σ. Since Σ is positive semi-definite, its inverse square root is

$$\Sigma^{-1/2} = \Gamma\Lambda^{-1/2}\Gamma^T \quad (8)$$

where $\Lambda^{-1/2} = diag(1/\sqrt{\lambda_i})$. With (8), it is easy to obtain, using equation (9), two new factors which have been proven independent of each other [8].

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \Sigma^{-1/2}(\begin{bmatrix} \Delta V_{th} \\ \frac{\Delta\beta}{\beta} \end{bmatrix} - \begin{bmatrix} \overline{\Delta V}_{th} \\ \overline{\frac{\Delta\beta}{\beta}} \end{bmatrix}) \quad (9)$$

Since the mean of mismatch parameters, $\overline{\Delta V}_{th}$ and $\overline{\frac{\Delta\beta}{\beta}}$, are equal to zero [12], (9) can be rewritten as

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \Sigma^{-1/2}(\begin{bmatrix} \Delta V_{th} \\ \frac{\Delta\beta}{\beta} \end{bmatrix}) = \Gamma\Lambda^{-1/2}\Gamma^T(\begin{bmatrix} \Delta V_{th} \\ \frac{\Delta\beta}{\beta} \end{bmatrix}) \quad (10)$$

For each mismatch-critical transistor pair M_n $(1 \leq n \leq |Z|)$, a pair of independent factors $(X_{1,n}, X_{2,n})$ can be obtained using Mahalanobis Transformation. Thus there exist a total of $2|Z|$ independent factors in the model.

A central composite design technique employing a fractional factorial design [9] can be used to choose the design points over the $2|Z|$ factors for model construction. In order to reduce the complexity of computing model parameters in later steps, all design points are coded into the $\pm1, 0$ or $\pm\kappa$, where κ is a value dependent on the number of factors [9].

The quadratic model can be written in matrix form

$$P = c_0 + \mathbf{x}^T\mathbf{c_1} + \mathbf{x}^T\mathbf{c_2}\,\mathbf{x} \quad (11)$$

where $\mathbf{x}$ are the influencing factors; P the performance parameter of interest; c_0, $\mathbf{c_1}$ and $\mathbf{c_2}$ respectively the unknown model parameters for zero order, first order and second order terms.

After SPICE simulations are performed at the design points, a set of equations can be constructed using the model expression in (11), as shown in the matrix form

$$\mathbf{P} = \mathbf{S} \cdot \mathbf{c} \quad (12)$$

where the column vector $\mathbf{P}$ consists of the set of values of the performance parameter P obtained from the simulations at the design points; $\mathbf{S}$, a matrix whose jth row is the design setting of the jth design point; the column vector $\mathbf{c}$, the unknown model parameters. Using least squares optimization, the estimate values of the model parameters are given by

$$\mathbf{c} = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\mathbf{P} \quad (13)$$

with which the analytic form of the model can be finally obtained. The procedures for model construction are summarized in Algorithm 1.

Algorithm 1 *Circuit − level mismatch model construction*

- Identify mismatch-critical transistor pairs
- For each mismatch-critical transistor pair, transform its ΔV_{th} and $\frac{\Delta\beta}{\beta}$ to two new independent factors using Mahalanobis Transformation
- Designate design points over all model factors using central composite design
- Perform SPICE simulations at all design points
- Compute model parameters using least squares optimization

3.3. Model-based Performance Characterization

Once a quadratic model as described in (11) has been identified, the expected values and variances of deviations of the performance parameters of interest can be directly calculated. The stationary point $\mathbf{x_0}$ of the response surface of this model (the point at which the slope of the response surface is zero when taken in all directions) can be calculated, using a differentiating method, as

$$\mathbf{x_0} = -\frac{\mathbf{c_2}^{-1}\mathbf{c_1}}{2} \quad (14)$$

If we substitute $\mathbf{u} = \mathbf{x} - \mathbf{x_0}$ for $\mathbf{x}$ in the model of (11), it can be transformed to a new form without first-order terms, as shown in (15)

$$P = p_0 + \mathbf{u}^T\mathbf{c_2}\,\mathbf{u} \quad (15)$$

The expected value and variance of the performance can be computed based on this simplified model expression. Since $\mathbf{u}$ is obtained through Mahalanobis Transformation of the multivariate normal distribution of mismatch parameters followed by a linear shifting, it obeys a multivariate standard normal distribution. Let its expectation $\mathcal{E}[\mathbf{u}] = \mu$ and variance-covariance matrix $\mathcal{D}[\mathbf{u}] = \Sigma_{\mathbf{u}}$. Then the expectation of performance P can be computed as [15]

$$\begin{aligned} \mathcal{E}[P] &= p_0 + \mathcal{E}[\mathbf{u}^T\mathbf{c_2}\,\mathbf{u}] \\ &= p_0 + tr[\mathbf{c_2}\Sigma_{\mathbf{u}}] + \mu^T\mathbf{c_2}\,\mu \end{aligned} \quad (16)$$

where $tr[]$ denotes the trace of a matrix. The variance of P can be computed as [15]

$$\begin{aligned} \sigma^2[P] &= \sigma^2[\mathbf{u}^T\mathbf{c_2}\,\mathbf{u}] \\ &= 2tr[\mathbf{c_2}\Sigma_{\mathbf{u}}]^2 + 4\mu^T\mathbf{c_2}\Sigma_{\mathbf{u}}\mathbf{c_2}\,\mu \end{aligned} \quad (17)$$

Since the elements in $\mathbf{u}$ are independent of each other, their variance-covariance matrix $\Sigma_{\mathbf{u}}$ is a diagonal matrix, which simplifies the matrix computation greatly.

4. Yield Optimization

The circuit-level mismatch model can be further used in yield optimization in early design stages, as its fast yield prediction capability provides the possibility of constructing an analytical yield model based on which the global yield-optimal point can be directly explored. For any chip with

its mismatch data characterized, its performance value can be predicted by the model. If the complete set of predicted performance specifications is satisfied, it is classified as a good chip. The yield is then computed as the fraction of good chips in the entire lot of chips of the same design. For any specific design (which corresponds to a point D in the design space $\mathcal{F}$), a corresponding yield Y can be predicted using the circuit-level mismatch models constructed at D. Hence a quadratic yield model in terms of the designable parameters can be constructed in the original design space $\mathcal{F}^0$ using RSM again. The model construction procedure is identical to that of the circuit-level mismatch model detailed in Section 3.2, except that Mahalanobis Transformation is no longer needed in this level of model construction as the designable parameters in the yield model are independent of each other. The point D^0 that corresponds to the maximum of the yield model in $\mathcal{F}^0$ can be easily computed using numerical techniques. Though this point can not be directly used as the true optimal design due to the model inaccuracy, it can be considered as a very good approximation which is sufficiently close to the true optimal design. Since RSM theory indicates that the response surface can be more accurately modelled as a quadratic function at the neighborhood of the maximum or minimum point, a better approximation point D^1 of the true optimal design can further be obtained by constructing a new and refined yield model in the small neighbouring area $\mathcal{F}^1$ of D^0 and finding its maximum point. The new search space $\mathcal{F}^1$ can be specified by taking D^0 as the center of $\mathcal{F}^1$ and shrinking the size of the original search space by a ratio α, where $0 < \alpha < 1$. In the same way, further search can be performed within a smaller neighborhood of D^1 to obtain an even better approximation point. A small number of iterations suffices for the approximation point to approach the true optimal design due to the quasi-exponential shrinking speed of the search space size. The iteration stops when the distance between two consecutive approximation points falls below some predefined threshold (e.g. the change in transistor size is less than the feature size of the fabrication process).

In order to avoid reaching a local optimal point during the optimization iteration, the shrinkage speed of the search space needs to be correctly determined for each iteration. A fixed shrinkage ratio can be easily implemented, but it does not incorporate the information about model quality and thus is not always appropriate. We propose a flexible shrinkage ratio dependent on the accuracy of the yield model of the previous iteration. The R_A^2 statistic (which is a value between 0 and 1) can be computed for the yield model in each iteration to reflect how well the model can explain the actual yield response surface. The shrinkage ratio α for the next iteration is hence defined as $\alpha = 1 - 0.5 * R_A^2$. If the model quality is very good (i.e. R_A^2 is close to 1), the shrinkage ratio can approach 1/2; otherwise a larger value is employed to ensure that the global optimal point is contained in the new search space.

A pre-analysis of circuit characteristics at the beginning of each iteration can make the yield optimization procedure more efficient. We have observed that the designable parameters that have large influence on the nominal values of performance are also critical to yield, as such parameters are decisive of the size of performance tolerance windows. Such an observation provides the possibility of simplifying the yield model as well as reducing simulation cost by only incorporating critical designable parameters in the yield model. At the beginning of each iteration, two simulations are performed for each designable parameter respectively at the lower bound and upper bound values of its corresponding search space. The difference between the nominal performance values at these two points is computed to indicate the influence of the designable parameter under consideration. Hence, for each performance, the designable parameters that are non-critical to it can be determined using the same differentiation strategy as employed in the identification of mismatch-critical transistor pairs described in Section 3.1. The designable parameters that are non-critical to all performances can be dropped from the yield model as they have little influence on yield. As a result, the dimension of the search space can be continuously decreased during the optimization iterations, which reduces the simulation cost drastically. The overall procedure for yield optimization is summarized in Algorithm 2.

Algorithm 2 *Yield optimization*

repeat
    Identify yield-critical designable parameters as yield model factors
    Designate design points for yield model construction in the search space over model factors
    Predict the yield at each design point using the circuit-level mismatch model constructed by Algorithm 1
    Construct a quadratic yield model using RSM
    Compute optimal point using numerical technique
    Compute R_A^2 and the shrinkage ratio α to specify the search space for the next iteration
until the distance between the current optimal point and the previous optimal point is less than the threshold
return current optimal point

5. Technical Feasibility

The run time complexity of the proposed methodology is dominated by SPICE simulation. The number of simulations increases in line with the number of mismatch-critical transistor pairs to be considered in the circuit-level mismatch model and that of designable parameters to be considered in the yield model. With the identification of mismatch-critical transistor pairs and yield-critical designable parameters, the simulation cost can be controlled in an affordable level, making the proposed methodology much more efficient than Monte-Carlo simulation. The reduction on the simulation count has no negative effect on the accuracy of the proposed method as the influence of dominant factors is thoroughly incorporated. Performance characterization and yield prediction can be performed efficiently because only computations over the circuit-level mismatch model are employed and no further simulations

Table 1. Circuit-level mismatch modelling for the three-stage operational amplifier

Performance	Simplified Model					Complex Model				Monte Carlo Simulation			
	Crit. Pairs	# simulations	Mean	St. Dev.	Yield	# simulations	Mean	St. Dev.	Yield	# simulations	Mean	St. Dev.	Yield
Ao	(M1,M2), (M3,M4)	30	80.5 db	73.6 db	57.1%	51	80.5 db	74.0 db	55.2%	10K	80.9 db	76.4 db	58.9%
r_o	(M1,M2), (M3,M4)		1435 Ω	407 Ω	76.8%		1459 Ω	413 Ω	78.0%		1531 Ω	435 Ω	83.7%
CMRR	(M1,M2), (M3,M4)		91.3 db	83.9 db	51.7%		91.3 db	83.7 db	52.1%		91.5 db	85.7 db	56.4%
Power	(M1,M2), (M3,M4)		2.31 mw	0.49 mw	84.7%		2.31 mw	0.49 mw	84.7%		2.31 mw	0.43 mw	82.4%
GBW	(M1,M2), (M3,M4)		4.42 MHz	1.33 MHz	69.2%		4.42 MHz	1.33 MHz	68.8%		4.44 MHz	1.23 MHz	62.0%

are needed. The proposed yield optimization technique can ensure a quick convergence to the optimal point as such a point is obtained by direct computation over an analytical model rather than by a direction-based search which may end at some locally optimal point. The proposed method can be completely embedded into an automated design tool and widely used in constraint-driven design.

6. Experimental Results

The proposed methodology has been tested by experimentation on two circuits, namely, a three-stage operational amplifier and a folded cascode amplifier. For both circuits, the experiment is designed to firstly verify the effectiveness of the circuit-level mismatch modelling technique proposed in Section 3 on the first-cut design and then that of the yield optimization technique proposed in Section 4.

6.1. Three-Stage Operational Amplifier

The schematic of the three-stage operational amplifier is shown in Figure 2. For each of the 5 important performance parameters (A_o, r_o, $CMRR$, $Power$, GBW), its mismatch-critical transistor pairs have been identified and a circuit-level mismatch model based on them has been constructed. For purposes of comparison, a $10K$-run Monte Carlo simulation has been performed as a baseline case for the evaluation of model accuracy, and a complex model based on all transistor pairs (without identifying mismatch-critical transistor pairs) has also been implemented to evaluate the effectiveness of the mismatch-critical pair identification technique presented in Section 3.1. Table 1 presents the experimental results for the first-cut design circuit. For each performance parameter, its mean, standard deviation and individual yield are obtained using all of the three methods. The number of simulations are also listed in Table 1.

Transistor pairs $(M1, M2)$ and $(M3, M4)$ are critical for all performance parameters at the first-cut design. The comparison between the simplified and the complex model

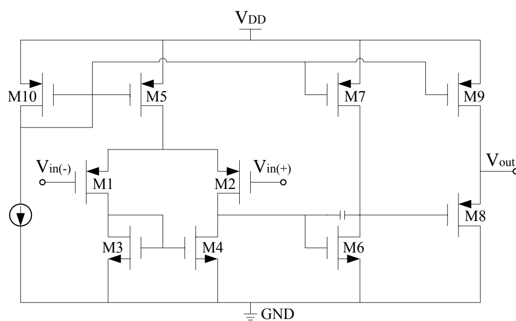

Figure 2. Three-stage operational amplifier

shows that these two models are at the same level of accuracy yet the construction of the simplified model requires fewer simulations than that of the complex one. This observation proves that the performance deviations are dominated by the mismatch-critical transistor pairs, and that the identification technique in Section 3.1 can greatly reduce the simulation cost without impairing the model accuracy. The comparison to the baseline results obtained by Monte-Carlo simulation shows that both the simplified and the complex model achieve sufficiently accurate predictions on the characteristics of performance variations and yields. However, the computational cost of the simplified model is far less than that of the one based on Monte-Carlo simulation. Considering that the number of simulations for Monte-Carlo simulation increases very quickly with the size of the circuit, the cost reduction achieved by the proposed method will be even greater for large circuits.

Table 2 presents the yield optimization results. Columns 2-4 show respectively the number of simulations, the model R_A^2 statistics and the achieved yield in each iteration. Since the yield model is quite accurate in each iteration, only two iterations are needed and a high yield is achieved. It can also be observed that the number of simulations needed in each iteration decreases very quickly due to the selection of yield-critical designable parameters.

6.2. Folded Cascode Amplifier

The schematic of the folded cascode amplifier is shown in Figure 3. The experimental setting is the same as in Section 6.1. The results for the first-cut design obtained by the three methods are shown in Table 3. Transistor pairs $(M1, M2)$, $(M3, M4)$, $(M5, M6)$ and $(M9, M10)$ are determined to be critical for all performance parameters at the first-cut design. The two modelling techniques still produce results that are at the same level of accuracy, which further confirms the effectiveness of the technique proposed in Section 3.1. Either model accurately predicts the practical performance distribution characterized by Monte-Carlo simulation. Specifically, for power consumption, the three methods reach almost identical results. Yet in terms of simulation cost, the advantage of the simplified model over the other two methods is even more evident than that in Section 6.1, as the folded cascode amplifier contains more identically designed transistor pairs.

Table 2. Yield optimization for the three-stage operational amplifier

opamp	# simulations	R^2_A	overall yield
first-cut design			49.1%
1st iteration	201	0.94	94.6%
2nd iteration	66	0.98	95.7%

Table 3. Circuit-level mismatch modelling for the folded cascode amplifier

Performance	Simplified Model					Complex Model				Monte Carlo Simulation			
	Crit. Pairs	# simulations	Mean	St. Dev.	Yield	# simulations	Mean	St. Dev.	Yield	# simulations	Mean	St. Dev.	Yield
Ao	(M1,M2), (M3,M4) (M5,M6),(M9,M10)		60.5 db	54.6 db	75.3%		60.5 db	54.5 db	75.0%		60.1 db	55.5 db	68.4%
r_o	(M1,M2), (M3,M4) (M5,M6),(M9,M10)		3.56 MΩ	1.80 MΩ	76.1%		3.57 MΩ	1.79 MΩ	76.5%		3.39 MΩ	2.00 MΩ	68.3%
CMRR	(M1,M2), (M3,M4) (M5,M6),(M9,M10)	57	97.4 db	85.4 db	74.8%	299	98.7 db	85.6 db	74.3%	10K	95.7 db	84.1 db	66.0%
Power	(M1,M2), (M3,M4) (M5,M6),(M9,M10)		302.3 μw	0.61 μw	99.9%		302.3 μw	0.61 μw	99.9%		302.3 μw	0.57 μw	99.9%
GBW	(M1,M2), (M3,M4) (M5,M6),(M9,M10)		53.6 MHz	4.34 MHz	90.5%		53.5 MHz	4.43 MHz	90.1%		53.7 MHz	3.96 MHz	91.3%

The yield optimization results for the folded cascode amplifier are presented in Table 4. It can be observed that only 3 iterations are needed and a high yield is quickly achieved.

7. Conclusion

In this paper, a two-stage methodology is proposed to contribute a mismatch-concerned design optimization technique for CMOS analog circuits. The main contributions of this work include a method for constructing a circuit-level model to predict mismatch impact on circuit behavior, and a fast yield optimization technique to guide the early design iterations. The proposed methodology can be embedded into EDA tools to help performance characterization and yield-aware gate-sizing optimization. Experiments show that the circuit-level mismatch modelling technique is accurate enough while being much more efficient than Monte-Carlo simulation widely employed in practical designs, and that the proposed yield optimization technique can reach the optimal design efficiently with rather low simulation cost.

References

[1] J. Bastos, M. Steyaert, R. Roovers, P. Kinget, W. Sansen, B. Graindourze, A. Pergoot, and E. Janssens. Mismatch characterization of small size MOS transistors. *IEEE ICMTS*, pages 271–276, 1995.

[2] J. Bordelon, B. Tranchina, V. Madangarli, and M. Craig. A strategy for mixed-signal yield improvement. *IEEE Design and Test of Computers*, 19:12–21, 2002.

[3] J. Chen, M. Orshansky, C. Hu, and C. Wan. Statistical circuit characterization for deep-submicron CMOS designs. *IEEE ISSCC*, pages 90–91, 1998.

[4] J. Croon, E. Augendre, S. Decoutere, W. Sansen, and H. Maes. Influence of doping profile and halo implantation on the threshold voltage mismatch of a 0.13 μm CMOS technology. *ESSDERC*, pages 579–582, 2002.

[5] B. De Smedt and G. Gielen. HOLMES: capturing the yield-optimized design space boundaries of analog and RF integrated circuits. *IEEE DATE*, pages 256–261, 2003.

[6] E. Felt, S. Zanella, C. Guardiani, and A. Sangiovanni-Vincentelli. Hierarchical statistical characterization of mixed-signal circuits using behavioral modeling. *ACM/IEEE ICCAD*, pages 374–380, 1996.

[7] B. Flury. *Common principle components and related multivariate models*. Wiley, New York, 1988.

[8] W. Härdle and L. Simar. *Applied multivariate statistical analysis*. Springer, New York, 2003.

[9] A. Khuri and J. Cornell. *Response surface: designs and analyses*. Marcel Dekker Inc., New York, 1996.

[10] P. Kinget and M. Steyaert. *Analog VLSI integration of massive parallel signal processing systems*. Kluwer Academic Publishers, Boston, 1997.

[11] C. Michael and M. Ismail. Statistical modeling of device mismatch for analog MOS integrated circuits. *IEEE J. Solid-State Circuits*, 27:154–166, 1992.

[12] M. Pelgrom, A. Duinmaijer, and A. Welbers. Matching properties of MOS transistors. *IEEE J. Solid-State Circuits*, 24:1433–1440, 1989.

[13] M. Quarantelli, S. Saxena, N. Dragone, J. Babcock, C. Hess, S. Minehane, S. Winters, J. Chen, H. Karbasi, and C. Guardiani. Characterization and modeling of MOSFET mismatch of a deep submicron technology. *IEEE ICMTS*, pages 238–243, 2003.

[14] F. Schenkel, M. Pronath, S. Zizala, R. Schwencker, H. Graeb, and K. Antreich. Mismatch analysis and direct yield optimization by spec-wise linearization and feasibility-guided search. *ACM/IEEE DAC*, pages 858–863, 2001.

[15] G. Seber. *Linear regression analysis*. Wiley, New York, 1977.

[16] S. Tiwary, P. Tiwary, and R. Rutenbar. Generation of yield-aware pareto surfaces for hierarchical circuit design space exploration. *ACM/IEEE DAC*, pages 31–36, 2006.

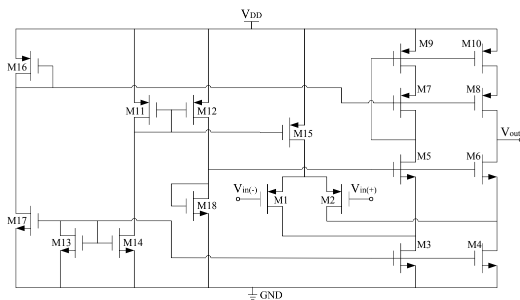

Figure 3. Folded cascode amplifier

Table 4. Yield optimization for the folded cascode amplifier

cascode	# simulations	R^2_A	overall yield
first-cut design			58.0%
1st iteration	793	0.90	89.3%
2nd iteration	244	0.87	95.7%
3rd iteration	92	0.98	98.7%