

Mismatch Analysis and Direct Yield Optimization by Spec-Wise Linearization and Feasibility-Guided Search

Frank Schenkel¹
Robert Schwencker^{1,2}

Michael Pronath¹
Helmut Graeb¹

Stephan Zizala^{1,2}
Kurt Antreich¹

¹Institute for Electronic Design Automation, Technical University of Munich, 80290 Munich, Germany
²Infineon Technologies AG, 81609 Munich, Germany

ABSTRACT

We present a new method for mismatch analysis and automatic yield optimization of analog integrated circuits with respect to global, local and operational tolerances. Effectiveness and efficiency of yield estimation and optimization are guaranteed by consideration of feasibility regions and by performance linearization at worst-case points. The proposed methods were successfully applied to two example circuits for an industrial fabrication process.

1. INTRODUCTION

In modern fabrication processes with their steadily shrinking feature size, the influence of process variations on the behavior of analog circuits cannot be neglected any more. As the local variance e.g. of a transistor's threshold voltage is inverse proportional to its area [1], the influence especially of mismatch due to local variations is getting more dominant in the future. Approaches to parametric yield optimization usually assume that the distribution of the statistical parameters like threshold voltage does not depend on the designable parameters like transistor widths and lengths. This assumption doesn't hold anymore when local process variations become important. Thus, efficient methods for yield estimation and improvement under both local and global process variations are needed for a fast and reliable design of analog circuits.

The analysis and optimization of the parametric yield of analog integrated circuits based on a Monte-Carlo analysis [2–5] is straightforward but needs a huge number of simulations if applied within an optimization loop. Moreover, the yield gradient needed for an optimization cannot be calculated, because the statistically varying parameters (e.g. oxide-thickness) and the designable parameters (e.g. transistor widths and lengths) are disjoint for the design of integrated circuits. In [6], the yield gradient is formulated using surface integrals, but a line search must be performed at every sample in order to determine the bounds of the acceptance region. In addition, gradient-based methods for direct yield optimization face the problem that yield and yield gradient are considerably different from 0 only in a small part of the whole design space. Usually, circuit performances and yield are more than weakly non-linear functions of the design parameters. This may lead to an impracticable effort for simulation or high-order modeling [4].

Methods based on geometrical approximations of the acceptance region [7, 8] are also problematic, because the designable and statistical parameters are disjoint for integrated circuits and the acceptance region defined in the space of statistical parameters depends on the designable parameters. Other methods overcome these problems by *multiple criteria optimization* (MCO) on a set of robustness objectives instead of optimizing the yield directly [10–12]. But circuit performances are often considerably correlated, which is difficult to account for in MCO. Algorithms relying on precalculated worst-case parameter sets [9] face the problem that the variances of the statistical parameters depend on the design parameters when local process variations and mismatch are to be considered. In this case, the worst-case parameter set is known to strongly depend on the design parameters and will hence change during the optimization.

In this contribution, a new approach to direct yield optimization for integrated circuits under consideration of local process variations is presented. The key concept is the strong focus on design-relevant regions in all parameter spaces, that are designable parameters, statistical parameters, and operating conditions, by means of structural constraints [13] and worst-case points [10]. This new combination significantly improves the quality of yield estimation by spec-wise linearized performance models, and therefore enables a robust and practicable technique for direct yield optimization.

In Section 2, the parametric operational yield is defined, which is the maximization goal. Based on worst-case points, Section 3 introduces a new way to analyze and detect mismatch-sensitive transistor pairs in a circuit. In Section 4 it is shown how the special statistical properties of local variations can be transformed into a more convenient statistical model that enables a Monte-Carlo based yield maximization. Section 5 describes the circuit model and the optimization algorithm that operates on this model.

In Section 6, the efficiency of the proposed method is demonstrated on two operational amplifiers. It is shown that both structural constraints and worst-case points are crucial for successful yield maximization. The proposed algorithm performs yield optimization by improving the nominal point and by reducing the variance of circuit performances simultaneously. Mismatch-relevant transistor pairs are detected and ranked in order of importance.

2. PARAMETERS, PERFORMANCES, AND YIELD

For each analog circuit, a set of *performances* $f^{(i)}$ like slew rate or phase margin is given ($i = 1, \dots, n_{\text{spec}}$). The performances of a fault-free analog circuit must satisfy a set of specifications $f^{(i)} \geq f_b^{(i)}$, e.g. the phase margin $\Phi_m \geq 60^\circ$, for all *operating parameters* θ , e.g. temperature or V_{DD} , in the operating range $\Theta = \{\theta \mid \theta^L \leq \theta \leq \theta^U\}$.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

DAC 2001, June 18–22, 2001, Las Vegas, Nevada, USA.
Copyright 2001 ACM 1-58113-297-2/01/0006 ...\$5.00.

Process fluctuations are modeled by *statistical parameters* \mathbf{s} , e.g. oxide thickness T_{ox} or threshold voltage V_{th} . Normal (Gaussian), log-normal, and uniform distributions are most commonly used. Without loss of generality, all of these distributions can be transformed into a normal (Gaussian) distribution [14,15], and therefore only this distribution is considered in the remaining part of the paper. *Design parameters* \mathbf{d} like widths and lengths of transistors are modified by the circuit designer during the sizing process.

Hence each performance $f^{(i)}$ is a function of the parameter vectors \mathbf{d} , \mathbf{s} and $\boldsymbol{\theta}$. In the space of statistical parameters, the *acceptance region* $A^{(i)}(\mathbf{d})$ of a performance $f^{(i)}$ is the set of circuits that satisfy the single specification $f^{(i)} \geq f_b^{(i)}$ in the full operating range Θ :

$$A^{(i)}(\mathbf{d}) = \{ \mathbf{s} \mid \forall_{\boldsymbol{\theta} \in \Theta} f^{(i)}(\mathbf{d}, \mathbf{s}, \boldsymbol{\theta}) \geq f_b^{(i)} \}. \quad (1)$$

There is usually a unique *worst-case operational parameter set*

$$\boldsymbol{\theta}_{\text{wc}}^{(i)} = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmin}} f^{(i)}(\mathbf{d}, \mathbf{s}, \boldsymbol{\theta}) \quad (2)$$

for each performance $f^{(i)}$. Then,

$$A^{(i)}(\mathbf{d}) = \{ \mathbf{s} \mid f^{(i)}(\mathbf{d}, \mathbf{s}, \boldsymbol{\theta}_{\text{wc}}^{(i)}) \geq f_b^{(i)} \}. \quad (3)$$

The overall acceptance region is

$$A(\mathbf{d}) = \bigcap_i A^{(i)}(\mathbf{d}). \quad (4)$$

The *parametric operational yield* Y is the percentage of produced circuits that satisfy the specification in spite of process fluctuations and for all operating conditions $\boldsymbol{\theta} \in \Theta$:

$$Y(\mathbf{d}) = \int_{A(\mathbf{d})} \text{pdf}(\mathbf{s}) \, ds, \quad (5)$$

where $\text{pdf}(\mathbf{s})$ is the probability density function of the statistical parameters.

If for example 90% of the produced circuits satisfy the specification for all $\boldsymbol{\theta} \in \Theta$, then $Y = 90\%$. If in turn all produced circuits satisfy the specification in 90% of the operating range, then $Y = 0\%$. Operating conditions are often more critical for the circuit performance than the statistical variations. Therefore, $\boldsymbol{\theta}$ must be rigorously considered as part of the specification to avoid an illusive high yield estimate.

Monte-Carlo analysis of circuits can account for this by evaluating the performance values at the respective worst-case operational parameter sets $\boldsymbol{\theta}_{\text{wc}}^{(i)}$ for each performance at N samples:

$$\tilde{Y} = \frac{1}{N} \sum_{j=1}^N \delta_j \quad (6)$$

$$\delta_j = \begin{cases} 1 & \text{if } f^{(i)}(\mathbf{d}, \mathbf{s}_j, \boldsymbol{\theta}_{\text{wc}}^{(i)}) \geq f_b^{(i)}, i = 1, \dots, n_{\text{spec}} \\ 0 & \text{else} \end{cases} \quad (7)$$

A loose upper bound for the simulation effort N^* can be given by $N^* \leq N \cdot \min(n_{\text{spec}}, 2^{\dim(\Theta)})$. Since performances may be calculated by a single simulation (like transit frequency and phase margin) and may share a common worst-case operational parameter set, N^* will usually be smaller.

3. ANALYSIS OF MISMATCH-SENSITIVE TRANSISTOR PAIRS

It is a key principle of analog circuit design to generate constant differences and ratios of currents or voltages with transistor pairs. These functional relationships are robust with respect to deviations of corresponding parameters of transistor pairs in the same direction, while being very sensitive to deviations of corresponding parameters in the opposite direction (mismatch). Finding the mismatch-sensitive transistor pairs can help identifying critical parts of a circuit to be considered thoroughly during redesign or layout.

Figure 1 shows a mismatch-sensitive circuit performance plotted over two locally varying statistical parameters $V_{\text{th}1}$ and $V_{\text{th}2}$. Pairs

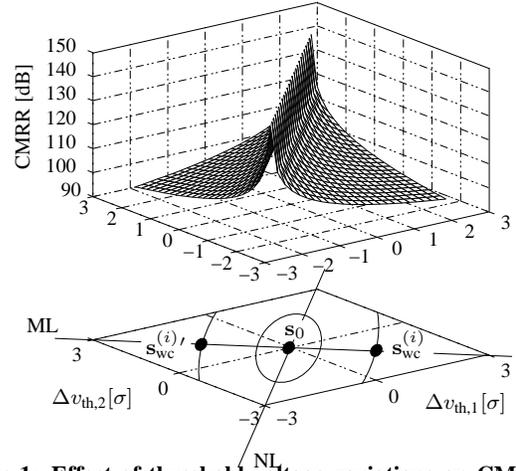


Figure 1: Effect of threshold voltage variations on CMRR of the operational amplifier (Fig. 7) before synthesis

of parameter values lying on the *neutral line* (NL) $\Delta s_1 = \Delta s_2$, have almost no influence on the performance value. In turn pairs of parameter values on the *mismatch line* (ML) $\Delta s_1 = -\Delta s_2$ result in the maximum decrease of the performance. Thus mismatch can be defined as follows:

DEFINITION 1. Performance $f^{(i)}$ is said to be *mismatch-sensitive* if there is at least one pair of statistical parameters s_k and s_l for which $f_0^{(i)} - f^{(i)}(\pm\Delta s_k, \pm\Delta s_l) \approx 0$ and $f_0^{(i)} - f^{(i)}(\pm\Delta s_k, \mp\Delta s_l) = \max$ holds ($\Delta s_k = \Delta s_l$). Then the corresponding transistors are said to be a *matching pair*.

During the mismatch analysis introduced in this section all design parameters remain constant. Since the distance factor in the matching properties of MOS transistors can be neglected [1], all locally varying parameters are uncorrelated. Hence without loss of generality the local statistical parameters which are responsible for mismatch can be assumed to be Gaussian distributed with $\mathbf{0}$ mean and the identity matrix as covariance matrix ($\mathbf{s} \sim N(\mathbf{0}, \mathbf{I})$).

From the definition of the worst-case parameter set [10]

$$\mathbf{s}_{\text{wc}}^{(i)} = \underset{\mathbf{s}}{\operatorname{argmin}} \left\{ \mathbf{s}^T \mathbf{s} \mid f^{(i)}(\mathbf{d}, \mathbf{s}, \boldsymbol{\theta}_{\text{wc}}^{(i)}) = f_b^{(i)} \right\}, \quad (8)$$

we know that this parameter set $\mathbf{s}_{\text{wc}}^{(i)}$ represents the circuit realization for which the performance $f^{(i)}$ equals the specification $f_b^{(i)}$ and which is closest to the nominal design \mathbf{s}_0 . That means $\mathbf{s}_{\text{wc}}^{(i)}$ is also the most probable circuit realization among all manufactured circuits to reach the specification bound $f_b^{(i)}$. This problem formulation (8) implies that the worst-case parameter sets will be in the direction of maximum performance degradation. It can be shown that a large value of a component in the worst-case parameter set corresponds to a large performance sensitivity with respect to that component ($\mathbf{s}_{\text{wc}}^{(i)} = -\kappa \cdot \nabla f^{(i)}(\mathbf{s}_{\text{wc}}^{(i)})$).

Consequently, if two components of a worst-case parameter set have the same maximum absolute value and opposite signs, we can conclude that these components belong to a matching transistor pair (see Figure 1). This property can be exploited to derive a procedure for a mismatch analysis. In the following a mismatch measure guiding the procedure will be formulated.

3.1 Requirements on the Measure

1. Pairs of worst-case parameter set values lying on the mismatch line (ML) should be identified as mismatch-sensitive parameter pairs.
2. The range of the mismatch measure should be from 0 (no mismatch) to 1 (maximum mismatch).

3. A comparison between the influences of mismatch on different circuit performances should be possible.
4. The higher the robustness of a circuit performance, the lower the mismatch measure should be.

3.2 Mismatch Measure

The following proposition of a mismatch measure $m_{k,l}^{(i)}$ between the two statistical parameters s_k and s_l for a specification $f_b^{(i)}$ fulfills the requirements listed in Section 3.1:

$$m_{k,l}^{(i)} = \eta^{(i)} \cdot \frac{\max\left(\left|s_{wc,k}^{(i)}\right|, \left|s_{wc,l}^{(i)}\right|\right)}{s_{\max}^{(i)}} \cdot \Phi\left(\arctan\left(\frac{s_{wc,k}^{(i)}}{s_{wc,l}^{(i)}}\right)\right) \quad (9)$$

$$\text{with } \eta^{(i)} = \begin{cases} 1 - \frac{1}{2(-\beta_{wc}^{(i)} + 1)} & \text{for } \beta_{wc}^{(i)} = \pm\sqrt{s_{wc}^T s_{wc}} < 0 \\ \frac{1}{2(\beta_{wc}^{(i)} + 1)} & \text{else} \end{cases}$$

$$\text{and } s_{\max}^{(i)} = \max_j \left\{ \left|s_{wc,j}^{(i)}\right| \right\}, \quad j = 1, \dots, n_s,$$

where $s_{wc,j}^{(i)}$ denotes the j -th component of the worst-case point $\mathbf{s}_{wc}^{(i)}$.

The selection of pairs of parameter values lying on the mismatch-line (including an uncertainty represented by the constants Δ_1 and Δ_2) is done by the function Φ (see Figure 2).

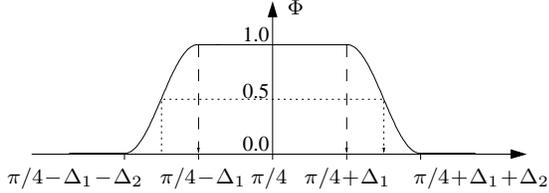


Figure 2: Angle dependent selection function Φ

The function η assures the assignment of a smaller values to more robust circuit performances (see Figure 3), while weighting higher performances with smaller robustness. η is 1/2 for $\beta_{wc}^{(i)} = 0$ and is continuously differentiable.

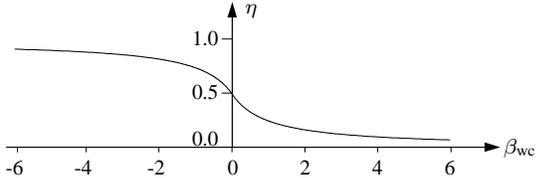


Figure 3: Robustness weighting function η

As mentioned previously, pairs of circuit parameters with a larger deviations have a stronger influence on the circuit performance to be analyzed. Through the 2nd term of eq. (9) these pairs will be weighted higher than the ones with smaller deviations. The division by $s_{\max}^{(i)}$ limits this term to maximum 1.

Since the worst-case parameter sets have to be determined anyway during the yield optimization described in Section 5, the mismatch analysis can be performed with no extra simulations.

4. YIELD OPTIMIZATION FOR CIRCUITS WITH LOCAL VARIATIONS

The task of yield optimization is to find a parameter set \mathbf{d} for which

$$Y = \int_{A(\mathbf{d})} \text{pdf}(\mathbf{s}, \mathbf{d}, \mathbf{C}(\mathbf{d})) ds \quad (10)$$

is maximized. The statistical parameters \mathbf{s} are Gaussian distributed $\mathbf{s} \sim N(\mathbf{s}_0, \mathbf{C}(\mathbf{d}))$ with a covariance matrix $\mathbf{C}(\mathbf{d})$. Algorithms for yield optimization of discrete circuits assume that $\mathbf{d} = \mathbf{s}_0$ and $A = \text{const}$, which doesn't hold for integrated circuits. Many other

approaches to yield maximization assume that $\mathbf{C} = \text{const}$. Unfortunately, this assumption doesn't hold anymore when local variations and mismatch are to be considered. Since $\sigma_{\text{Vth}}^2 \propto 1/WL$, the covariance matrix \mathbf{C} depends on the design parameters \mathbf{d} [1, 16]. With local variations, Y can be improved not only by enlarging A , but also by reducing the variance of \mathbf{s} . Depending on the initial design, both factors are necessary to increase the yield. Therefore, a modern yield optimization technique must account for both.

Optimization of eq. (5) in presence of local variations is difficult, because the integration region A and the probability measure $\text{pdf}(\mathbf{s}) ds$ both depend on \mathbf{d} . We can transform

$$\hat{\mathbf{s}} = \mathbf{G}(\mathbf{d})^{-1} \cdot (\mathbf{s} - \mathbf{s}_0) \sim \mathbf{s}(\hat{\mathbf{s}}) = \mathbf{G}(\mathbf{d}) \cdot \hat{\mathbf{s}} + \mathbf{s}_0, \quad (11)$$

where $\mathbf{G}(\mathbf{d}) \cdot \mathbf{G}(\mathbf{d})^T = \mathbf{C}(\mathbf{d})$. Then with the indicator function δ_A denoting if a sample point is in the acceptance region: $\delta_A(\mathbf{s}, \mathbf{d}) = 1$ if $\mathbf{s} \in A(\mathbf{d})$, else $\delta_A(\mathbf{s}, \mathbf{d}) = 0$,

$$\begin{aligned} Y(\mathbf{d}) &= \int_{A(\mathbf{d})} \text{pdf}(\mathbf{s}, \mathbf{d}) ds = \int_{\mathbb{R}^n} \text{pdf}(\mathbf{s}, \mathbf{d}) \delta_A(\mathbf{s}, \mathbf{d}) ds \\ &= \int_{\mathbb{R}^n} \text{pdf}(\mathbf{s}(\hat{\mathbf{s}}), \mathbf{d}) \delta_A(\mathbf{s}(\hat{\mathbf{s}}), \mathbf{d}) (\mathbf{G}(\mathbf{d}) \cdot d\hat{\mathbf{s}}) \\ &= \int_{\mathbb{R}^n} \det(\mathbf{G}(\mathbf{d})) \text{pdf}(\mathbf{s}(\hat{\mathbf{s}}), \mathbf{d}) \delta_A(\hat{\mathbf{s}}, \mathbf{d}) d\hat{\mathbf{s}} \\ &= \int_{\hat{A}(\mathbf{d})} \widehat{\text{pdf}}(\hat{\mathbf{s}}) d\hat{\mathbf{s}} = \hat{Y}(\mathbf{d}). \end{aligned} \quad (12)$$

With $\widehat{\text{pdf}}(\hat{\mathbf{s}}) = (2\pi)^{-\frac{n}{2}} \exp(-\frac{1}{2}\hat{\mathbf{s}}^T \hat{\mathbf{s}})$ and therefore $\hat{\mathbf{s}} \sim N(\mathbf{0}, \mathbf{I})$. The transformed acceptance region is then

$$\hat{A}(\mathbf{d}) = \bigcap_i \{ \hat{\mathbf{s}} \mid \hat{f}^{(i)}(\mathbf{d}, \hat{\mathbf{s}}, \boldsymbol{\theta}_{wc}^{(i)}) \geq f_b^{(i)} \} \quad (13)$$

$$\text{with } \hat{f}^{(i)}(\mathbf{d}, \hat{\mathbf{s}}, \boldsymbol{\theta}) = f^{(i)}(\mathbf{d}, \mathbf{s}(\hat{\mathbf{s}}), \boldsymbol{\theta}). \quad (14)$$

Since $Y(\mathbf{d}) = \hat{Y}(\mathbf{d})$ for every \mathbf{d} , it is sufficient for yield optimization to maximize \hat{Y} . Maximizing \hat{Y} over \hat{A} is easier than maximizing Y over A , because the covariance matrix of $\hat{\mathbf{s}}$ is constant and there is no need to calculate a derivative of the covariance matrix with regard to the design parameters. The variable variance $\mathbf{C}(\mathbf{d})$ is implicitly contained in $\hat{f}^{(i)}$.

By (12), we shift the variability of the probability measure into the integration space and transform the problem of maximizing Y over a variable probability measure into an equivalent problem with a constant measure. Therefore we can use a single approach to treat local and global variations when optimizing parametric operational yield of integrated circuits.

Note that (11) is not a simple constant norm in the statistical parameter space, but a linear transformation that depends on \mathbf{d} , i.e. it varies during the optimization. To keep the notation readable, we will use \mathbf{s} , f and A for $\hat{\mathbf{s}}$, \hat{f} and \hat{A} in the remaining part of the paper.

5. YIELD OPTIMIZATION METHODOLOGY

In this section, our algorithm for yield improvement is presented. It is based on a direct yield optimization method. In each iteration step of the optimization, a linearized performance model in \mathbf{d} and \mathbf{s} is determined and used for yield estimation and coordinate-search based yield optimization. The quality of this linearized model is crucial for the success of the optimization and will be described in the following. It consists of two parts. First the performance linearization with respect to \mathbf{d} at the current iteration point. It will be illustrated, that the constraints determine the "feasibility region" for the linearization (see Figure 4). Second the performance $f^{(i)}$ is linearized for each specification individually at the corresponding worst-case point (see Section 5.2).

The following sections are ordered according to their appearance in the optimization loop.

5.1 Feasibility Region

Functional constraints, e.g. all transistors must be in saturation, guarantee the basic functionality and robustness of a circuit [13]. They define the *feasibility region* $\mathcal{F} = \{\mathbf{d} \mid \mathbf{c}(\mathbf{d}) \geq \mathbf{0}\}$ in the space of the design parameters \mathbf{d} . Please note the difference between \mathcal{F} and A . \mathcal{F} is defined for design parameters and DC performances of transistors and transistor pairs and can be interpreted as technology-dependent sizing rules. A is defined for arbitrary performances of a certain circuit and determines the circuit yield. Considering \mathcal{F} as described in the Sections 5.5, 5.3 and 5.4 is crucial:

1. The solution of the yield improvement has to be feasible in order to represent a technically correct circuit.
2. Most performances are only weakly nonlinear in the feasibility region, as can be seen in Fig. 4. Therefore the reduction of the design space to the feasibility region \mathcal{F} significantly improves the precision of linearized performance models for estimating the change in yield over the design parameters.
3. The constraints reduce the exploration space for the optimization algorithm and therefore improve the convergence of the algorithm.

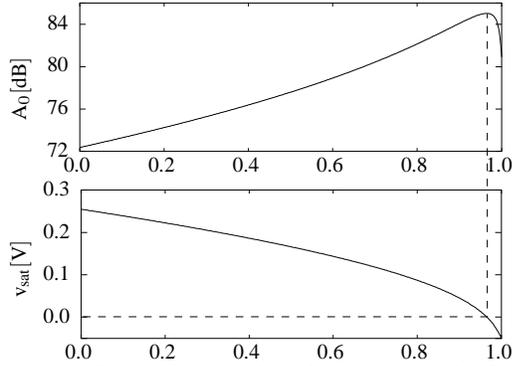


Figure 4: Performance behavior of A_0 over the feasibility region $v_{\text{sat}} \geq 0$.

The results in Section 6 show that these linearizations of the performances are sufficient for the yield optimization and that no model of higher order is needed when considering functional constraints.

During the optimization only linearized models of the constraints at a feasible point \mathbf{d}_f , i.e. a point that fulfills all constraints, are considered:

$$\bar{\mathbf{c}}(\mathbf{d}) = \mathbf{c}_0 + \nabla_{\mathbf{d}} \mathbf{c}(\mathbf{d}_f) \cdot (\mathbf{d} - \mathbf{d}_f). \quad (15)$$

This linearization is updated in each iteration step of the optimization, i.e. after each update of \mathbf{d}_f .

5.2 Specification-Wise Linearization

The convex polytope used to approximate the acceptance region A is determined by linearizations of the performances $f^{(i)}$ in their worst-case points $\mathbf{s}_{\text{wc}}^{(i)}$. Since the worst-case point $\mathbf{s}_{\text{wc}}^{(i)}$ is defined to be the parameter set with the highest probability density for which the performance value is equal to the specification bound [10], a good approximation of the acceptance region can be expected. The worst-case point is calculated for each specification separately by solving (8). The problems in finding this worst-case point in the presence of mismatch, and an algorithm to overcome these problems is presented in [12].

With the help of these worst-case points, linear models of the circuit performances are built. The performances are linearized in a feasible point \mathbf{d}_f and the worst-case points $\mathbf{s}_{\text{wc}}^{(i)}$:

$$\begin{aligned} \bar{f}^{(i)}(\mathbf{d}, \mathbf{s}) &= f_b^{(i)} + \nabla_{\mathbf{s}} f^{(i)}(\mathbf{d}_f, \mathbf{s}_{\text{wc}}^{(i)}, \boldsymbol{\theta}_{\text{wc}}^{(i)}) \cdot (\mathbf{s} - \mathbf{s}_{\text{wc}}^{(i)}) \\ &\quad + \nabla_{\mathbf{d}} f^{(i)}(\mathbf{d}_f, \mathbf{s}_{\text{wc}}^{(i)}, \boldsymbol{\theta}_{\text{wc}}^{(i)}) \cdot (\mathbf{d} - \mathbf{d}_f) \quad (16) \\ &i = 1, \dots, n_{\text{spec}} \end{aligned}$$

It has been shown that a yield estimate \bar{Y} can be obtained on this spec-wise linearized model at no extra simulation cost that in practice has an accuracy differing less than 1-2% from the results of a Monte-Carlo analysis [12].

5.3 Yield Improvement

In the yield improvement algorithm, a yield estimate \bar{Y} is maximized over the design parameters \mathbf{d} within a coordinate search loop. This yield estimate is obtained based on an evaluation of a predefined number N of Monte-Carlo samples over the linearizations of the performances, which remain unchanged during this optimization loop:

$$\bar{Y} = \frac{1}{N} \cdot \sum_{j=1}^N \delta_j \quad (17)$$

$$\delta_j = \begin{cases} 1 & \bar{f}^{(i)}(\mathbf{d}, \mathbf{s}_j, \boldsymbol{\theta}_{\text{wc}}^{(i)}) \geq f_b^{(i)}, i = 1, \dots, n_{\text{spec}}, \\ 0 & \text{else} \end{cases} \quad (18)$$

After every change of \mathbf{d} these samples are reevaluated and the yield estimation \bar{Y} is redetermined.

In order to maximize the yield within the coordinate search loop, the following optimization problem over the design parameters \mathbf{d} is solved in every iteration step and for every coordinate k :

$$\mathbf{d}^* + \mathbf{e}_k \cdot \underset{\alpha}{\text{argmax}} \{ \bar{Y}(\mathbf{d}^* + \alpha \mathbf{e}_k) \mid \bar{\mathbf{c}}(\mathbf{d}) \geq \mathbf{0} \} \longrightarrow \mathbf{d}^* \quad (19)$$

This coordinate search is performed until the yield estimate \bar{Y} obtained over the linearized models of the circuit performances cannot be further improved.

A robust coordinate search is given preference to a gradient based algorithm, because

- the yield and the gradient of the yield can be 0 over a large part of the design space (see Figure 5). Thus gradient based algorithms have to start quite close to the optimum.
- even if the $f^{(i)}(\mathbf{d}, \mathbf{s})$ are determined through linearized performance models $\bar{f}^{(i)}(\mathbf{d}, \mathbf{s})$, the yield is strongly nonlinear and non-monotonic over \mathbf{d} (see Figure 5). This aggravates finding the maximum of the yield estimate by means of a gradient based algorithm.
- $\bar{Y}(\mathbf{d})$ is non-continuous, as it is determined through a Monte-Carlo analysis. This makes the determination of a yield gradient [6] difficult.

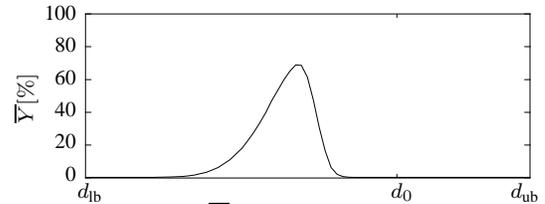


Figure 5: Yield estimate \bar{Y} over a design parameter d from its lower bound d_{lb} to its upper bound d_{ub}

To keep the computational effort low, not the whole linear model (16) is evaluated every time the yield estimate has to be recalculated over the set of Monte-Carlo samples. For a change in \mathbf{d} only $\Delta \bar{f}^{(i)} = \nabla_{\mathbf{d}} f^{(i)}(\mathbf{d}_f, \mathbf{s}_{\text{wc}}^{(i)}, \boldsymbol{\theta}_{\text{wc}}^{(i)}) \cdot (\mathbf{d} - \mathbf{d}_f)$ has to be redetermined. Hence the remaining part of equation (16) is stored for every sample $\mathbf{s}_j, j = 1, \dots, N$ as it remains constant. Moreover, since the components of the design parameter vector \mathbf{d} are changed separately one after the other during the coordinate search, only one

component of this inner product has to be calculated. Consequently the constant term $\bar{f}_{s_j}^{(i)}$ solely has to be compared with $f_b^{(i)} + \Delta f^{(i)}$:

$$\delta_j = \begin{cases} 1 & \bar{f}_{s_j}^{(i)}(\mathbf{d}_f, \mathbf{s}_j) \geq \Delta \bar{f}^{(i)} - f_b^{(i)}, i = 1, \dots, n_{\text{spec}}, \\ 0 & \text{else} \end{cases} \quad (20)$$

In the presence of mismatch, performances $f^{(i)}$ may have quadratic behavior with semidefinite Hessian matrix (see Figure 1). Then a yield estimation based on only one linearization is a poor estimate and may mislead the optimization algorithm. This problem is accounted for by introducing an additional specification and thus one more linear model for every such performance in $\mathbf{s}_{\text{wc}}^{(i) \prime}$ with

$$\mathbf{s}_{\text{wc}}^{(i) \prime} = -\mathbf{s}_{\text{wc}}^{(i)} \quad (21)$$

$$\nabla_{\mathbf{s}} f^{(i)}(\mathbf{d}_f, \mathbf{s}_{\text{wc}}^{(i) \prime}, \boldsymbol{\theta}_{\text{wc}}^{(i)}) = -\nabla_{\mathbf{s}} f^{(i)}(\mathbf{d}_f, \mathbf{s}_{\text{wc}}^{(i)}, \boldsymbol{\theta}_{\text{wc}}^{(i)}) \quad (22)$$

Only one additional simulation is needed for every specification to identify the quadratic behavior of the performance.

5.4 Line Search

Since the maximization of the yield estimate \bar{Y} has been performed using linearizations of the circuit performances and the constraints, a line search based on real circuit simulations along the line between the feasible starting point \mathbf{d}_f and the optimum \mathbf{d}^* has to be performed afterwards to assure the new iteration point $\mathbf{d}_f^{(\text{new})}$ lying in the feasibility region. Therefore, the following optimization formulation is solved with a small number of circuit simulations (e.g. 10):

$$\gamma_{\text{max}} = \underset{\gamma}{\text{argmax}} \{ \gamma | \mathbf{c}(\mathbf{d}_f + \gamma \mathbf{r}) \geq \mathbf{0} \wedge 0 \leq \gamma \leq 1 \} \quad (23)$$

with $\mathbf{r} = \mathbf{d}^* - \mathbf{d}_f$. This leads to a new point $\mathbf{d}_f^{(\text{new})} = \mathbf{d}_f + \gamma_{\text{max}} \cdot \mathbf{r}$ which serves as new feasible starting point for the next iteration.

5.5 Finding a Feasible Starting Point

In an initial step, our algorithm searches for a feasible starting point \mathbf{d}_f , i.e. a point that fulfills all functional constraints. In the case that the starting point \mathbf{d}_0 is not feasible, the closest feasible point \mathbf{d}_f in the space of the design parameters is determined.

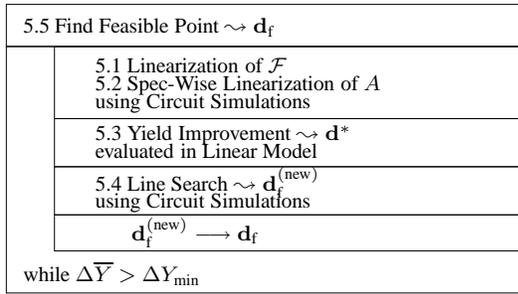


Figure 6: Structure of the yield optimization algorithm

Figure 6 summarizes the whole algorithm. It consists of an initial step to find a feasible starting point \mathbf{d}_f in the space of the design parameters \mathbf{d} , followed by three steps executed in a loop until no further improvement of the yield can be achieved: the linearization of the constraints (feasibility region) and the spec-wise linearization of the performances, a subsequent maximization of a yield estimate within a coordinate search loop over the design parameters \mathbf{d} based on those performance linearizations and a final line search assured by circuit simulations that the algorithm stays in the feasibility region. This new point \mathbf{d}_f serves as new iteration point for the next linearization of the circuit performances and constraints.

6. RESULTS

The proposed methods were applied to two example circuits for an industrial fabrication process. The folded-cascode opamp in Figure 7 was modeled with local variations. In the initial design, the total yield was 0%, mainly due to transit frequency f_t and CMRR (Table 1). The rows $f^{(i)} - f_b^{(i)}$ contain the difference between $f^{(i)}(\mathbf{d}, \mathbf{s}_0, \boldsymbol{\theta}_{\text{wc}}^{(i)})$ and the specification. The rows “bad samples” show, how many samples in the linearized model did not satisfy the respective specification. The rows \tilde{Y} show the results of a simulation-based Monte-Carlo analysis, including operational parameters according to Section 2, with a sample size of 300. This was performed at the end of the optimization to verify the results, and between the optimization steps to demonstrate the actual improvement.

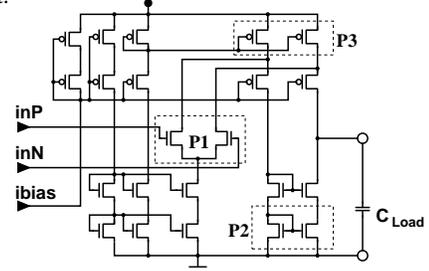


Figure 7: Folded-cascode operational amplifier

After the first iteration, a yield of 99.9% could be achieved. The second iteration still improved the robustness of the circuit. This was possible due to the high number (10,000) of Monte-Carlo samples evaluated in the linear model. After the second iteration, all 10,000 samples were inside the acceptance region A , as one can see in the rows “bad samples”.

	Performance	A_0 [dB]	f_t [MHz]	CMRR [dB]	SR_p [V/ μ s]	Power [mW]
	Specification $f_b^{(i)}$	> 40	> 40	> 80	> 35	< 3.5
Initial	$f^{(i)} - f_b^{(i)}$	10.7	-2.3	-1.9	0.18	0.54
	bad samples [%]	0.0	1000.0	980.4	272.5	0.0
	\tilde{Y}	0%				
1st Iter.	$f^{(i)} - f_b^{(i)}$	15.3	3.69	4.70	0.96	0.50
	bad samples [%]	0.0	0.0	0.9	0.2	0.0
	\tilde{Y}	99.9%				
2nd Iter.	$f^{(i)} - f_b^{(i)}$	17.7	4.15	12.8	1.63	0.51
	bad samples [%]	0.0	0.0	0.0	0.0	0.0
	\tilde{Y}	100%				

Table 1: Trace of yield optimization under consideration of functional constraints

Table 2 shows that the algorithm actually improves the yield in two ways: The distance of the mean values from the specification is increased (first column), and the variance of the performances is decreased (second column). Both factors are employed to improve f_t , CMRR.

Performance	$\Delta \mu_f / (\mu_f - f_b)$	$\Delta \sigma_f / \sigma_f$
A_0	+ 15.5%	+ 20.4%
f_t	+ 12.8%	- 11.5%
CMRR	+ 169%	- 53.4%
SR_p	+ 73.4%	+ 3.15%
Power	- 0.59%	- 1.69%

Table 2: Improvement between first and second iteration

To show the importance of using functional constraints, the same algorithm was applied once again to the same initial design, but without functional constraints. Table 3 shows, that the yield remains 0% after the first iteration. As one can see in the rows “bad

Performance		A_0 [dB]	f_t [MHz]	CMRR [dB]	SR_p [V/ μ s]	Power [mW]
Specification $f_b^{(i)}$		> 40	> 40	> 80	> 35	< 3.5
Initial	$f^{(i)} - f_b^{(i)}$	10.7	-2.3	-1.9	0.18	0.54
	bad samples [%]	0.0	1000.0	980.4	272.5	0.0
	\tilde{Y}	0%				
1st Iter.	$f^{(i)} - f_b^{(i)}$	-3.0	-5.0	-1.9	-1.0	0.6
	bad samples [%]	0.0	0.0	0.0	0.0	0.0
	\tilde{Y}	0%				

Table 3: Trace of yield optimization without consideration of functional constraints

samples”, the yield improvement algorithm still worked correctly and reduced the number of bad samples in the linearized models. However, this did not improve the true yield \tilde{Y} , as the linearized models were too inaccurate due to the missing functional constraints.

To show the importance of using worst-case points, the same algorithm was applied once again to the same initial design including functional constraints, but the linearizations were calculated at the nominal point $s = s_0$ instead of the respective worst-case points $s = s_{wc}$. Table 4 shows, that the yield remains 0% after the first iteration. Again, the number of bad samples in the linearized models declined, but the true yield did not grow, because the used linearized models were too inaccurate at the specification, especially for CMRR (cf. Fig. 1).

Performance		A_0 [dB]	f_t [MHz]	CMRR [dB]	SR_p [V/ μ s]	Power [mW]
Specification $f_b^{(i)}$		> 40	> 40	> 80	> 35	< 3.5
Initial	$f^{(i)} - f_b^{(i)}$	10.7	-2.3	-1.9	0.18	0.54
	bad samples [%]	0.0	1000.0	546.3	0.0	0.0
	\tilde{Y}	0%				
1st Iter.	$f^{(i)} - f_b^{(i)}$	19.4	5.8	-2.3	3.6	0.6
	bad samples [%]	0.0	437.8	482.1	7.7	0.0
	\tilde{Y}	0%				

Table 4: Trace of yield optimization with linearization at the nominal points $s = s_0$

Table 5 shows the three largest values of the mismatch measure defined in Section 3.2 for the folded-cascode opamp. CMRR turned out to be the only performance sensitive to mismatch. The three transistor pairs detected by this analysis are marked with P1, P2 and P3 in Figure 7.

Pair	P1	P2	P3
$m_{k,l}^{CMRR}$	0.84	0.11	0.06

Table 5: Mismatch measure for CMRR at initial point

The second example circuit was the Miller opamp of Figure 8. Only global process variations were considered for this design. The results of the yield improvement are shown in Table 6.

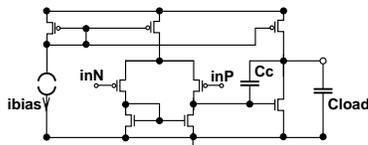


Figure 8: Miller operational amplifier

The total number of simulations and the time needed for the whole yield optimization of both examples are compiled in Table 7.

These results were obtained on a network (100 Mbit/sec) of 5 computers (500MHz Pentium III) in parallel, using the Infineon in-house simulator TITAN [17].

Performance		A_0 [dB]	f_t [MHz]	Φ_m [$^\circ$]	SR_p [V/ μ s]	Power [mW]
Specification $f_b^{(i)}$		> 80	> 1.3	> 60	> 3	< 1.3
Initial	$f^{(i)} - f_b^{(i)}$	7.4	1.6	0.8	-0.1	0.5
	bad samples [%]	3.6	0.0	166.8	636.2	0.0
	\tilde{Y}	33.7%				
1st Iter.	$f^{(i)} - f_b^{(i)}$	7.8	2.0	2.7	0.7	0.3
	bad samples [%]	2.6	0.0	0.0	0.3	0.0
	\tilde{Y}	99.3%				
2nd Iter.	$f^{(i)} - f_b^{(i)}$	7.7	1.9	3.3	0.7	0.3
	bad samples [%]	1.6	0.0	0.0	0.1	0.0
	\tilde{Y}	99.3%				

Table 6: Results for Miller operational amplifier

Circuit	# Simulations	Wall Clock Time
Folded-Cascode	689	30 min
Miller	627	8 min

Table 7: Computational effort for optimization

7. CONCLUSION

A method for mismatch analysis and automatic yield optimization of analog integrated circuits with respect to global, local and operational tolerances has been presented. This iterative yield optimization method is based upon

- spec-wise linearizations of the performances at the worst-case points in the space of the statistical parameters,
- linearizations of the feasibility region defining a “trust region” of the performance linearization with respect to the design parameters,
- a robust coordinate search.

Experimental results show that the combination of these linear approximations enable a robust and efficient yield optimization.

8. REFERENCES

- [1] M. Pelgrom, A. Duinmaier, A. Welbers, “Matching properties of MOS transistors”, *IEEE J. SC*, 1989.
- [2] M. Keramat, R. Kielbasa, “OPTOMEGA: an environment for analog circuit optimization”, *IEEE ISCAS*, 1998.
- [3] K. Antreich, R. Koblitz, “Design centering by yield prediction”, *IEEE TCAS*, 1982.
- [4] J. C. Zhang, M. A. Styblinski, *Yield and Variability Optimization of Integrated Circuits*, Kluwer, 1995.
- [5] M. Meehan, J. Purviance, *Yield and Reliability in Microwave Circuit and System Design*, Artech House, 1993.
- [6] P. Feldmann, S. Director, “Integrated circuit quality optimization using surface integrals”, *IEEE TCAD*, 1993.
- [7] H. Abdel-Malek, A. Hassan, “The ellipsoidal technique for design centering and region approximation”, *IEEE TCAD*, 1991.
- [8] S. Director, W. Maly, A. Strojwas, *VLSI Design for Manufacturing: Yield Enhancement*, Kluwer, 1990.
- [9] A. Dharchoudhury, S. M. Kang, “Worst-case analysis and optimization of VLSI circuit performances”, *IEEE TCAD*, 1995.
- [10] K. Antreich, H. Graeb, C. Wieser, “Circuit analysis and optimization driven by worst-case distances”, *IEEE TCAD*, 1994.
- [11] K. Krishna, S. W. Director, “The linearized performance penalty (LPP) method for optimization of parametric yield and its reliability”, *IEEE TCAD*, 1995.
- [12] K. Antreich, J. Eckmueller, H. Graeb, M. Pronath, F. Schenkel, R. Schwencker, S. Zizala, “WiCkED: Analog circuit synthesis incorporating mismatch”, *IEEE CICC*, 2000.
- [13] S. Zizala, J. Eckmueller, H. Graeb, “Fast calculation of analog circuits’ feasibility regions by low level functional measures”, *IEEE ICECS*, 1998.
- [14] Kevin S. Eshbaugh, “Generation of correlated parameters for statistical circuit simulation”, *IEEE TCAD*, 1992.
- [15] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, 1991.
- [16] K. Lakshmikummar, R. Hadaway, M. Copeland, “Characterization and modeling of mismatch in MOS transistors for precision analog design”, *IEEE J. SC*, 1986.
- [17] U. Feldmann, U. Wever, Q. Zheng, R. Schultz, and H. Wriedt, “Algorithms for modern circuit simulation”, *Archiv für Elektronik und Übertragungstechnik (AEÜ)*, 1992.