# Statistical Integrated Circuit Design

Stephen W. Director, *Fellow, IEEE,* Peter Feldmann, *Member, IEEE,* and Kannan Krishna, *Student Member, IEEE*

***Abstract*— IC manufacturers have become increasingly interested in maximizing yield as a way to maximize profits. As a result, there has recently been renewed interest in computer aids for maximizing yield. While statistical design methods for maximizing circuit yield have been available for more than two decades, it is only recently that such methods have become practical. In this paper, we review some of these methods and illustrate the usefulness of one of these, the boundary integral method, with several examples.**

## I. Introduction

TRADITIONAL or deterministic design focuses on finding a set of circuit parameter values such that all circuit performance specifications are met. It is usually assumed that this set of values can be accurately realized when the circuit is built. Unfortunately, this assumption is not realistic for mass-produced integrated circuits where fluctuations in the manufacturing process cause deviations in the actual values of the parameters from their nominal values. The statistical variability of the circuit parameters, in turn, causes the circuit performances to show a spread of values. Thus, if an amplifier was designed for a nominal dc gain of 50 dB, we could expect to see a distribution of actual gains for a population of manufactured chips to be similar to the one shown in Fig. 1.

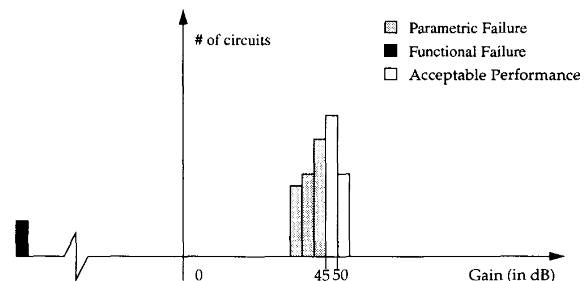


Fig. 1. Distribution of manufactured performances.

The ratio of the number of chips that meet specifications to the number of chips that are manufactured is referred to as the yield. Yield of less than 100% is due to two types of disturbances: local and global. Local disturbances, caused by spot defects in the manufacturing process, are the primary cause of catastrophic failures or functional failures which often cause a change in the basic functionality of the circuit. Global disturbances, primarily caused by phenomena that affect most devices on a chip (e.g., mask misalignment, variations in diffusion temperatures, or implant doses), are the primary cause of parametric failures. Parametric failures result when the chip is "functional" but specifications on such quantities as speed and power are not met.

Because of the close correlation between high yield and high profits, IC manufacturers strive to maximize yield. As the complexity of VLSI designs continues to increase, the need for computer aids for maximizing yield becomes more important. While statistical design methods for maximizing the circuit yield have been an active area of research for more than two decades, it is only relatively recently that such methods have been taken seriously by the IC design community. Among the reasons for this situation are that a number of practical yield maximization methods specific to integrated circuits have recently been developed. Second, it is only relatively recently that computers powerful enough to use these methods in a reasonable amount of time have appeared. Finally, the economics of chip manufacture are such that maximization of yield, rather than merely optimization of performance, has become of prime importance.

In this paper, we review several statistical design methods that have been developed to minimize the effects of IC manufacturing process disturbances on circuit performance. We also illustrate the effectiveness of the most recent of these, the boundary integral method, with several circuit design examples.

## II. Parametric Circuit Yield Maximization

In this section, we will formally define "parametric yield" and formulate the "parametric yield maximization problem" for integrated circuits.

A circuit to be designed is characterized in terms of a set of $n_\varphi$ performances, which are the components of the $n_\varphi$-dimensional vector $\varphi = (\varphi_1, \cdots, \varphi_{n_\varphi})$. For example, the performances of interest, for the circuit shown in Fig. 2, are the dc gain, the unity gain frequency, and the phase margin.

These performances are determined by a set of parameters that is denoted by the $n_p$-dimensional vector $\boldsymbol{p} = (p_1, \cdots, p_{n_p})$. Note that $\varphi = \varphi(\boldsymbol{p})$. For VLSI circuits, there are two categories of parameters comprising $\boldsymbol{p}$ that affect circuit performances: process-related parameters like the oxide thicknesses and length reduction of MOS devices, and circuit-related parameters such as device sizes and their placement. During the normal design process, nominal values are assumed for the parameters $\boldsymbol{p}_0$. During manufacture, however, inherent process disturbances cause the actual parameter values to deviate randomly from their nominal values causing actual

Manuscript received August 26, 1992; revised October 27, 1992. This work was supported in part by the Semiconductor Research Corporation.

S. W. Director and K. Krishna are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213.

P. Feldmann is with AT&T Bell Telephone Laboratories, Murray Hill, NJ 07974.

IEEE Log Number 9206123.

0018-9200/93$03.00 © 1993 IEEE

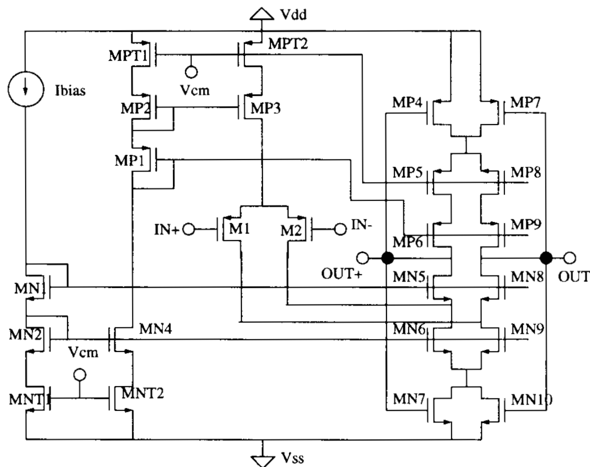


Fig. 2. Fully differential folded-cascode amplifier.

performances to vary as well. In particular, the actual layout geometries are affected by effects such as mask misalignment, lateral diffusion, and over (under) etching, while process-related parameters are primarily dependent on effects such as variations in diffusion times, temperatures, energies, doses of implants, etc.

If the random $n_\xi$-dimensional vector $\xi$ denotes process disturbances, the vector $\boldsymbol{p}$ represents actual values for the parameters with nominal values $\boldsymbol{p}_0$, subjected to the disturbances $\xi$. Mathematically, the actual parameter values can be expressed as functions of their nominal values and of the disturbances: $\boldsymbol{p} = \boldsymbol{p}(\boldsymbol{p}_0, \xi)$. Note that the number of parameters is, in general, not related to the number of disturbances, i.e., a single process disturbance can simultaneously affect multiple integrated circuit parameters and a circuit parameter may be affected by multiple disturbances. The nominal values of the process-related parameters are not under the direct control of the circuit designer, who can only specify parameters such as device sizes and layout. Therefore, only a subset of the nominal parameter values $\boldsymbol{p}_0$ are truly designable in the circuit design phase. For the purpose of yield maximization, it is convenient to express circuit performances as functions of the deterministic nominal parameter values $\boldsymbol{p}_0$ and the random vector $\xi$: $\varphi = \varphi(\boldsymbol{p}_0, \xi)$. For the circuit in Fig. 2, the designable parameters are the device geometries and the bias current. Although there are 24 transistors and consequently 48 widths and lengths that are designable, only 12 of the widths and lengths are independent designables, due to matching constraints, e.g., $M1$ and $M2$ should have identical dimensions. The designable part of the $\boldsymbol{p}_0$ vector of this circuit therefore has 13 elements: 12 widths and lengths and a bias current.

The random variables that characterize the disturbances, $\xi = (\xi_1, \cdots, \xi_{n_\xi})$, may represent physically based effects at the process level. These variables, such as variations of temperatures, impurity diffusivities, etc., mimic the random variations that occur in a manufacturing line. A process and device simulator [1] can be used to predict their effect on circuit parameters. We can make reasonable approximations of the *jpdf* of $\xi$. One frequently made assumption is that the components of the random vector $\xi$, which model variations of controlling parameters at successive fabrication steps, are, at least to first order, mutually independent. A different approach to the modeling of disturbances is to use statistical techniques on data collected from the fabrication line. These techniques, such as principal factor analysis [20], model the highly dependent set of random variables as functions of a smaller set of independent random variables. While these independent factors may not necessarily be related to any physical quantities, they can be used to model the circuit parameter variations.

## III. VLSI Circuit Optimization

A manufactured circuit will be considered acceptable if all of its actual performances fall within acceptable limits, i.e., if

$$\varphi_k^L \leq \varphi_k \leq \varphi_k^U, \qquad k = 1, \cdots, n_\varphi. \tag{1}$$

In the case of our amplifier circuit in Fig. 2, acceptable performances might be expressed as

- dc gain $\geq$ 55 dB,
- unity gain frequency $\geq$ 45 MHz,
- phase margin $\geq 30°$.

The system of inequalities, (1), defines a region of acceptability, denoted by $A_\varphi$, in the $n_\varphi$-dimensional performance space, which is a cartesian space spanned by the $n_\varphi$ performances $\varphi_i, i = 1, \cdots, n_\varphi$.

We can now formally state the yield of a design as the probability that a manufactured circuit has acceptable performance, i.e., that the vector $\varphi$ belongs to the acceptability region $A_\varphi$:

$$Y = \text{Prob}\{\varphi \in A_\varphi\} = \int_{A_\varphi} f_\varphi(\boldsymbol{p}_0, \varphi)\, d\varphi.$$

where $f_\varphi(\boldsymbol{p}_0, \varphi)$ is the joint probability density function ($jpdf$) of the circuit performances that constitute the vector $\varphi$. The *jpdf* is clearly dependent on the nominal parameter values $\boldsymbol{p}_0$. Yield maximization is performed by determining a set of nominal values of the designable parameters, $\boldsymbol{p}_0$, that maximizes the probability mass of the random performances that lies within $A_\varphi$. Consequently, the yield maximization problem, in the performance space, is formulated as

$$\max_{\boldsymbol{p}_0} \left\{ Y = \int_{A_\varphi} f_\varphi(\boldsymbol{p}_0, \varphi)\, d\varphi \right\}. \tag{2}$$

It is important to observe that the *jpdf* $f_\varphi(\boldsymbol{p}_0, \varphi)$ is not explicitly available to us and, in general, no easy means is available for obtaining it. Therefore, employing (2) as a basis for maximizing yield is difficult and we are motivated to consider alternative formulations to the yield maximization problem.

We can also formulate the yield optimization problem in the space of independent statistical disturbances, viz., the disturbance space. The disturbance space is a Cartesian space in which the axes correspond to independent statistical disturbances. The region of acceptability in the disturbance space, denoted by $\boldsymbol{A}_\xi(\boldsymbol{p}_0)$, consists of all possible combinations of

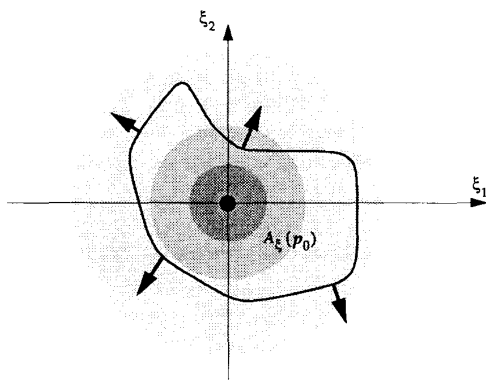


Fig. 3. Yield optimization by moving the acceptability region boundary.

disturbances that can occur in the manufacturing of a circuit, which, for specific nominal parameter values, do not result in unacceptable performance. It is given by

$$A_\xi(\boldsymbol{p}_0) = \{\xi | \varphi^L \leq \varphi(\boldsymbol{p}_0, \xi) \leq \varphi^U\}. \tag{3}$$

In the disturbance space, the acceptability region is the inverse image of the parametrized mapping $\xi \rightarrow \varphi$ which depends on the nominal parameter values $\boldsymbol{p}_0$.

Parametric yield can be expressed in the disturbance space as the probability of occurrence of a combination of acceptable disturbance values and can be formalized as

$$Y = \text{Prob}\,\{\xi \in A_\xi(\boldsymbol{p}_0)\} = \int_{A_\xi(\boldsymbol{p}_0)} f_\xi(\xi)\, d\xi. \tag{4}$$

Thus, the resulting optimization problem is

$$\max_{\boldsymbol{p}_0} \left\{ Y = \int_{A_\xi(\boldsymbol{p}_0)} f_\xi(\xi)\, d\xi \right\}. \tag{5}$$

Yield optimization is, therefore, performed by modifying the acceptability region in a way so as to increase the coverage of a fixed probability distribution. This is illustrated in Fig. 3 for a two-dimensional disturbance space.

For the case of the differential amplifier in Fig. 2, we assumed that the processing variations were adequately modeled by the variations in length and width of the n- and p-type devices, flat-band voltage variation of the n- and p-type devices, and the variation in oxide thickness. Thus, we have a seven-dimensional disturbance space.

We have found that the disturbance space formulation is the most adequate formulation for the case of integrated circuits where different components and devices are subject to variations caused by the same random disturbances. Some other statistical design methods that assume a different formulation cannot easily accommodate this type of situation.

## IV. Statistical Design Issues

### A. Alternate Optimization Formulations

So far, we have considered only the adjustment of nominal parameter values as a means for optimizing yield. Another possible way to improve yield is to adjust component tolerances. For the case of VLSI circuits, reducing tolerances means tighter control of the process, thereby minimizing the spread in the device parameter values. This can be achieved by applying statistical design techniques at the process level and through the use of more precise, and, in most cases, more expensive manufacturing equipment. Since tightening tolerances is likely to increase the cost of production, it is desirable to identify the components, device parameters, and processing steps, for which better tolerances result in savings that offset the cost. Alternatively, one can identify parameters for which relaxing the tolerance has only minor effects on the variability of performances. This may allow the use of a less expensive processing step, or a simpler circuit topology, thus reducing manufacturing cost.

Mathematically, tolerances can be modeled as parameters, $\boldsymbol{p}_f$, that control some measure of spread, such as variances, of the statistical distributions, $f_\xi(\xi, \boldsymbol{p}_f)$, of disturbances. Let $C(\boldsymbol{p}_f)$ be a function that models the manufacturing cost associated with a given set of tolerances. The tolerancing problem can then be formulated as

$$\max_{\boldsymbol{p}_f} Y = \int_{A_\xi} f_\xi(\xi, \boldsymbol{p}_f)\, d\xi \quad \text{such that } C(\boldsymbol{p}_f) < C_{\max}. \tag{6}$$

The tolerancing problem can also be formulated as a cost minimization problem with a minimum yield constraint:

$$\min_{\boldsymbol{p}_f} C(\boldsymbol{p}_f) \quad \text{such that} \left( Y = \int_{A_\xi} f_\xi(\xi, \boldsymbol{p}_f)\, d\xi \right) > Y_{\min}.$$

Such a formulation may be more appropriate in some cases.

Sometimes a low yield results from very stringent specifications. In many cases, especially when designing a subcomponent of a larger design, specifications may be negotiable. It may be the case that one or several specifications can be relaxed and compensated for by other subcomponents at a much lower cost. The problem here is to identify specification constraints responsible for significant yield loss and to predict to what extent yield can be improved by relaxing them. Of course, such an exercise is not possible when specifications result from standards or published specification sheets.

Specifications can be modeled, mathematically, as deterministic parameters that control the acceptability region. If the function $C(\varphi^B)$ quantifies the cost, on the entire system, of modifying the specifications $\varphi^B$, which represent lower or upper bounds on circuit or device performances, the specification assignment problem becomes

$$\max_{\varphi^B} Y = \int_{A_\xi(\varphi^B, \boldsymbol{p}_0)} f_\xi(\xi)\, d\xi \quad \text{such that } C(\varphi^B) < C_{\max}. \tag{7}$$

As in the case of the tolerancing problem, this can also be formulated as a cost minimization problem with a minimum yield constraint.

Finally, while parametric yield is the most frequently used measure of statistical behavior of mass produced circuits, other measures such as quality loss or signal-to-noise ratios [10], [26] are sometimes more accurate measures of performance variability. Several statistical design methodologies [10], [17] use these as the objective for optimization.

### B. Analytical Models

The statistical design problem can be formulated as an optimization problem in which the objective function and/or constraint functions contain statistical averages, such as yield or another measure of performance variability. Optimization of a function is an iterative process that requires several evaluations of the objective and constraint functions and their gradients. Since statistical objective functions are, in general, very expensive to evaluate, it is important to choose an optimization strategy that requires a minimum number of objective function evaluations. State-of-the-art gradient-based optimization algorithms, such as [4], can significantly reduce the number of required iterations. Therefore, it is very important to be able to compute both the statistical objective function and its gradient efficiently.

In general, to determine if a circuit satisfies a given specification, a circuit simulation must be performed. Circuit simulation, especially of large circuits, is computationally expensive and has, consequently, been the bane of practical use of yield maximization strategies. Fortunately, techniques have been developed to overcome this bottleneck through the use of analytical approximations of performances. The efficiency of these techniques is determined by the cost of building sufficiently accurate approximations. These techniques require that some number of circuit simulations be performed on a sample of circuits according to some experimental plan [24], [25]. From these simulations, the analytical approximations are determined using either response surface or interpolation techniques.

Response surface models are constructed by computing the coefficients of linear or quadratic polynomials through linear regression from the simulation results at the experimental points. This requires a number of simulations that is larger than the number of coefficients to be computed, the extra degrees of freedom being used to estimate the modeling error. For quadratic polynomials, the number of coefficients increases as the square of the number of model parameters and, therefore, may quickly become prohibitively large. In order to reduce the number of coefficients that need to be computed, techniques such as parameter screening or stepwise regression [23] can be used. Some researchers have gone so far as to ignore all quadratic cross-products. This is, however, not a good compromise since the effect of this would be to neglect all interactions among parameters, information that is crucial for variability reduction [19].

A different way of generating simple analytic expressions for performances as a function of their parameters is through interpolation. In contrast to response surfaces, interpolation models pass exactly through all the experimental points, but provide no means for estimating the resulting modeling error. Interpolation methods used in statistical design methodologies range from simple linear interpolation [6], [7] to more complicated techniques such as maximally flat interpolation [15]. A method that combines the maximally flat interpolation with regression-based approximation was presented in [16]. A particularly interesting interpolation approach has been proposed in [31].

## V. Yield Maximization Methodologies for IC's

As mentioned above, a large number of circuit yield maximization methodologies have been developed for discrete circuits. Integrated circuit yield optimization is fundamentally different from discrete circuit yield optimization in that for IC's a, typically small, number of disturbances affect a large number of devices causing variations in the devices to be highly dependent on each other. Here we discuss some important methodologies that are suitable for integrated circuit yield optimization.

### A. The Method of Random Perturbations

One category of methods for IC yield optimization is based on stochastic approximation. Stochastic approximation is a general method for dealing with optimization of regression functions, $E[h(\boldsymbol{p})]$ ($E[\cdot]$ denotes expectation), where only noise corrupted observations of $h(\boldsymbol{p})$ can be obtained. Several stochastic approximation methods have been developed for both unconstrained and constrained statistical optimization problems.

Yield can be expressed as the expectation of the acceptability region indicator function

$$h_\xi(\xi, \boldsymbol{p}_0) = \begin{cases} 1, & \text{if } \xi \in A_\xi(\boldsymbol{p}_0) \\ 0, & \text{otherwise} \end{cases}$$

with respect to the *jpdf* of the statistical variables.

$$Y = \int_{A_\xi(\boldsymbol{p}_0)} f_\xi(\xi)\, d\xi = \int_\infty h_\xi(\xi, \boldsymbol{p}_0) f_\xi(\xi)\, d\xi \qquad (8)$$

Stochastic approximation methods can, therefore, be used for maximization of the yield function.

Styblinski and Ruszczynski [28] use a variant of the simplest stochastic approximation algorithm which is akin to the steepest ascent method. The main feature of this method is the fact that the Monte Carlo estimation of yield gradient is blended within the optimization algorithm. The expensive Monte Carlo analysis is not performed at every iteration to determine the yield gradient. The estimate of the gradient used in the stochastic approximation approach can be based on a very small number of samples, even one, and the direction of movement of the designable variables, at any iteration, is a convex combination of that of the previous iteration and the gradient determined during the current iteration. The rate of convergence is, however, dependent on the accuracy of the yield estimate.

Yield in expression (8) can be estimated through Monte Carlo using a sample of points in the disturbance space, produced by a random number generator that mimics the *jpdf* $f_\xi$. However, since designable parameters appear in $h_\xi$, which is

not differentiable, there is no straightforward expression for the Monte Carlo estimate of the yield gradient. The method of random perturbation is used to cope with this situation.

This method involves perturbing the original problem by adding a random component $\eta$ to the deterministic parameters $\boldsymbol{p}_0$. The variance of this random component is controlled by a parameter $\beta$. As $\beta$ is decreased to 0, the random component $\eta$ becomes a deterministic zero vector and the parameters become deterministic again. One of the possible choices for the distribution, $f_\eta(\eta, \beta)$, to be assigned to these perturbation parameters is a multivariate normal distribution with its covariance matrix scaled by $\beta$. With this random perturbation all parameters become statistical and can be decomposed into the sum of a deterministic vector (which is partially designable) and a random vector.

$$\tilde{\boldsymbol{p}} = \begin{bmatrix} \boldsymbol{p}_0 \\ 0 \end{bmatrix} + \begin{bmatrix} \eta \\ \xi \end{bmatrix} = \tilde{\boldsymbol{p}}_0 + \tilde{\xi}. \tag{9}$$

Using the indicator function $\tilde{h}$ (similar to the one introduced before), the "perturbed yield" can be expressed as

$$\tilde{Y}(\boldsymbol{p}_0, \beta) = \int_\infty \tilde{h}(\tilde{\boldsymbol{p}}_0 + \tilde{\xi}) \tilde{f}_{\tilde{\xi}}(\tilde{\xi}, \beta)\, d\tilde{\xi} \tag{10}$$

where $\tilde{f}_{\tilde{\xi}}(\tilde{\xi}, \beta) = f_\eta(\eta, \beta) f_\xi(\xi)$. By performing a simple change of variables $\zeta = \tilde{\boldsymbol{p}}_0 + \tilde{\xi}$, the designable parameters appear only in the *jpdf*, which is differentiable.

$$\tilde{Y}(\boldsymbol{p}_0, \beta) = \int_\infty \tilde{h}(\zeta) f_\zeta(\zeta - \tilde{p}_0, \beta)\, d\zeta. \tag{11}$$

In this formulation the gradient of the perturbed yield can be estimated using a Monte Carlo procedure. For the stochastic approximation algorithm, however, only one or a few samples are used at every iteration. During the course of optimization $\beta$ is also swept to 0, therefore, at convergence the "perturbed yield" becomes the original yield.

This method was used in a system for production yield optimization with respect to fundamental fabrication parameters and mask layout dimensions. A statistical process simulator (FABRICS [1]) is run in conjunction with a circuit simulator (SPICE [2]) to determine the effect of fabrication process variations on circuit performance.

The main drawback of this method is the fact that the optimization algorithm is stochastic and may require a very large number of iterations to achieve convergence. Since simulations must be performed at each iteration, stochastic approximation may turn out to be prohibitively expensive. Analytic models can be employed to reduce the cost of simulation. The approximation must be done, however, in terms of both deterministic and statistical variables over the entire space of interest, and therefore the cost of modeling can increase rapidly as the circuit size increases.

### *B. The Texas Instruments Method*

Researchers at Texas Instruments have proposed a yield optimization method tailored for digital MOS circuits. They have observed that both current and capacitance in MOS FET's are primarily sensitive to four variables [5]–[7]: width reduction, length reduction, oxide capacitance, and flat-band voltage (for both n-channel and p-channel devices for the case of CMOS). Effects of variables such as doping profile have been found to be significantly smaller. Since these parameters are determined by different steps in the manufacturing process, they can, to first order, be assumed to be independent of each other. All circuit variations are, therefore, assumed to be caused by the variations of these statistical variables. The aforementioned variables define the disturbance space.

To reduce the cost of circuit simulation, circuit performances are approximated as linear functions of the four statistical variables. A linear interpolation model for each performance is constructed from five (six for the case of CMOS circuits) circuit simulations. Performance constraints resulting from specifications applied to these interpolation models define a polytope in the disturbance space that approximates the acceptability region. Yield is defined as the integral of the disturbance *jpdf* over this approximation to the acceptability region. It can easily be estimated by Monte Carlo.

Having a means for evaluating yield, the next step is to have some means for optimizing it with respect to the designables. Two yield optimization strategies are proposed. The gradient-based approach involves computation of the yield gradient. The other is a geometric approach.

The gradient-based approach involves the computation of the yield and the yield gradient by multidimensional integration. Yield can be evaluated by quadrature techniques, by integrating the *jpdf* of the statistical variables over the approximating polytope. The gradient can be evaluated in a similar way, only that the integration is performed over the facets of the approximated acceptability region. A gradient-based algorithm using a quasi-Newton procedure is then used to optimize the circuit yield.

The geometric approach solves the problem in which the minimum distance of the mean point of the disturbance *jpdf* (the projection of the nominal point from the parameter space onto the disturbance space) from the acceptability region boundaries is maximized, i.e., the boundaries are pushed as far as possible from the center of the *jpdf* (Fig. 4). It uses an iterative approach to solve this problem in which, at each iteration, the distance functions, to the approximated acceptability region boundaries, are linearized. A linear programming problem is then formulated and solved with this linearization.

The main features of this method are approximation of the acceptability region boundaries by hyperplanes and yield maximization, done both directly with the yield and yield gradient and, indirectly, using a geometric approach.

The method described above makes use of a number of good ideas, such as the choice to compute yield in the disturbance space, the use of a small number of statistical variables to account for most of the performance variability, and the computation of the yield gradient as surface integrals. However, the method uses some algorithms and assumptions that restrict its application as a general yield optimization strategy. The method builds linear models for the performances in terms of the statistical variables which may not be accurate enough in other processes. Other aspects of the method, for example, using quadrature integration to determine the

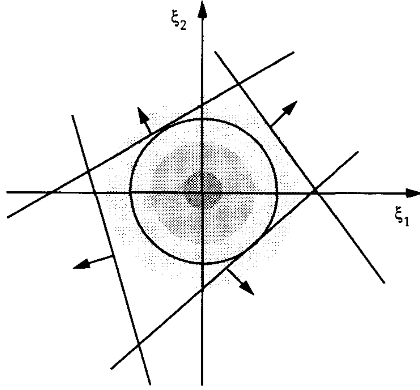


Fig. 4. Moving the acceptability region boundaries to improve parametric yield.

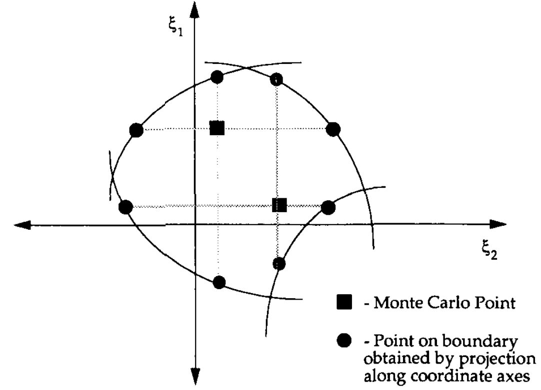


Fig. 5. Generation on points on acceptability region boundary in the boundary integral method.

yield gradient, can be applied only to acceptability regions determined by such linear models and even then only if the number of statistical variables is restricted to about four or five. Therefore, this methodology is limited in its application to the particular class of digital MOSFET circuits described by the authors.

### C. The Boundary Integral Method

The boundary integral method [8]–[11], which may be viewed as a generalization of the Texas Instruments approach, uses a formulation of yield involving the acceptability region in the space of independent statistical disturbances. However, the boundary integral method is unique in the sense that through the application of Stokes' theorem it reformulates yield itself as a surface integral on the boundary of the acceptability region in this disturbance space. The advantages of this reformulation is that yield and derivatives of yield, with respect to deterministic nominal parameters, parameters of the disturbance distribution, and performance constraints, can all be expressed as surface integrals on the acceptability region boundary, and can be computed through Monte Carlo based on the same set of sample points. This sample is obtained by generating points within the acceptability region and then projecting them onto the acceptability region boundaries, as illustrated in Fig. 5.

Suppose we generate such a sample of points on the acceptability region boundary. $S = \{\zeta^{(k)}\}_{k=1,\cdots,N_\zeta}$. According to the boundary integral method the Monte Carlo estimate of yield is

$$\hat{Y} = \frac{1}{nN} \cdot \sum_{k=1}^{N_\zeta} s_k F_{\xi,q}(\xi_q)_{\xi=\zeta^{(k)}} \tag{12}$$

where $n$ is the dimensionality of the disturbance space, $N$ is the number of original Monte Carlo points, $q$ is the projection direction used to generate point $\zeta^{(k)}, F_{\xi,q}(\xi_q)$ is the marginal probability function corresponding the $q$th component of $\zeta^{(k)}$, and $s_k$ is a sign factor.

The expression for the derivative of the yield with respect to a designable circuit parameter, say $x_i$, is

$$\frac{\partial \hat{Y}}{\partial x_i} = \frac{1}{nN} \cdot \sum_{k=1}^{N_\zeta} f_{\xi,q}(\xi_q) \cdot \left. \frac{\partial \varphi_a / \partial x_i}{|\partial \varphi_a / \partial \xi_q|} \right|_{\xi=\zeta^{(k)}} \tag{13}$$

where $\varphi_a$ is the performance that determines the boundary at point $\zeta^{(k)}$.

Similarly, the expression for the derivative of the yield with respect to a specification constraint, $\varphi_a^B, (\varphi_a - \varphi_a^B \geq 0)$ is

$$\frac{\partial \hat{Y}}{\partial \varphi_a^B} = \frac{1}{nN} \cdot \sum_{k=1}^{N_{\zeta,a}^B} f_{\xi,q}(\xi_q) \cdot \left. \frac{1}{|\partial \varphi_a / \partial \xi_q|} \right|_{\xi=\zeta_a^{(k)}} \tag{14}$$

where the terms correspond the subset of the boundary points $S_a^B = \{\zeta^{(k)}\}_{k=1,\cdots,N_{\zeta,a}^B}$ for which the constraint is active. The derivative values are the sensitivities of the yield to various constraint specifications.

In addition, expressions can be obtained for derivatives of yield with respect to parameters of the disturbance distribution. The sensitivity of the yield with respect to a parameter $p$ that may affect the distribution of more than one disturbance is given by

$$\frac{\partial Y}{\partial p} = \frac{1}{nN} \sum_{k=1}^{N_\zeta} s_k \left[ \frac{\partial}{\partial p} F_{\xi,q}(\boldsymbol{p}, \xi_q)_{\xi=\zeta^{(k)}} + \sum_{i=1, i\neq q}^{n} \left[ F_{\xi,q}(\boldsymbol{p}, \xi_q) \frac{\partial}{\partial p} \log f_{\xi,i}(\boldsymbol{p}, \xi_i) \right]_{\xi=\zeta^{(k)}} \right]. \tag{15}$$

This expression can be used to determine the effect of the variances or spreads of different process disturbances on the performance of the circuit. This expression is also used to obtain the yield gradient component contributed by disturbances that have distributions dependent on the designable parameters. Although a yield optimization procedure will only make use of the information regarding the gradient of the yield with respect to the designable parameters, the aforementioned sensitivities can provide valuable information about the circuit being designed.

The boundary integral method provides a way to estimate yield, yield gradients, and sensitivities that is completely decoupled from the larger statistical optimization problem. Yield, its gradient with respect to the designable parameter vector, and its derivatives with respect to performance specifications, disturbance distribution parameters, etc. can be computed without any reference to the kind of optimization that needs to be performed. One is not tied down to any one particular optimization formulation and the choice of an optimization algorithm is completely orthogonal.

The same orthogonality applies to the choice of the analytic approximation method. In [8]–[11], in order to save on the cost of performance evaluation by circuit simulation at every Monte Carlo point, the performances of the circuit are approximated by macromodels (response surfaces) that are quadratic in the disturbance variables. However, one can even choose to estimate performance employing a smaller number of boundary points obtained directly through line searches and avoid approximations altogether.

The advantages of using the surface integral formulation of yield are multiple. First, the computation involved in determining separate sets of Monte Carlo points for the yield and the yield gradient is avoided. Most importantly, though, the Monte Carlo estimate of the surface integral yield expression is consistent with the Monte Carlo estimate of the yield gradient integral, i.e., the gradient of the surface integral Monte Carlo estimate of the yield is the Monte Carlo estimate of the yield gradient. This consistency is absent if the yield is estimated as a volume integral based on a different sample. As a consequence of the consistency a gradient-based optimization using very poor estimates, even based on a few points, will still converge.

## VI. Circuit Optimization Examples with Boundary Integrals

We now illustrate the effectiveness of the boundary integral method by several examples. As a first example, we used the boundary integral method to maximize the yield of the differential amplifier example shown in Fig. 2. The specifications, disturbances, and designables were as described earlier. The circuit yield improved from 39% to 100%. Figs. 6, 7, and 8 show distributions of the three performances of interest before and after optimization.

These figures show the trade-off that has been achieved between the gain and unity gain frequency (ugf) specs, and the phase margin spec. The final spreads of all three performance measures have been reduced. However, the nominal gain and the nominal ugf have been reduced and a corresponding increase in the nominal phase margin has been obtained.

The boundary integral method affords us an efficient way to compute circuit yield and its gradient. Traditionally, yield maximization examples tend to focus on improving circuit yield using this gradient information. However, as we have

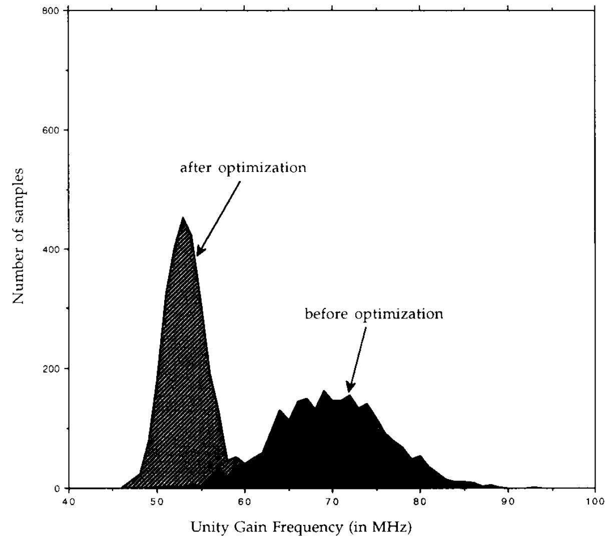


Fig. 6. Distributions of ugf (in megahertz) in a population of 2500 samples before and after optimization.

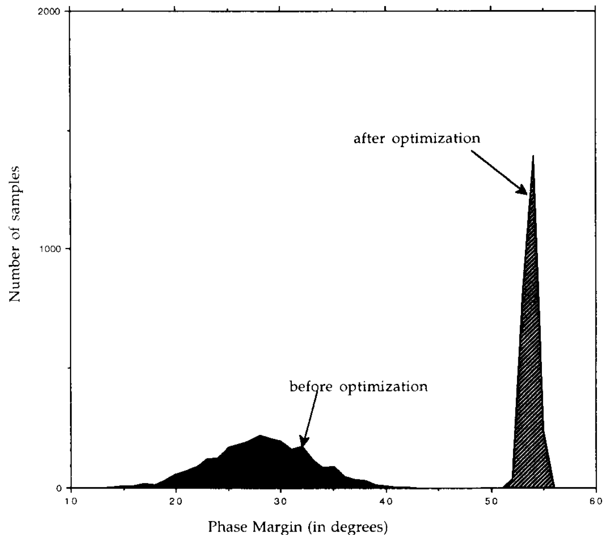


Fig. 7. Distributions of phase margin (in degrees) in a population of 2500 samples before and after optimization.

discussed in Section IV, statistical design need not restrict itself to maximizing yield. In this example we illustrate how we can increase a performance specification while maintaining a minimum yield specification. The problem formulation is of the form

$$\begin{aligned}&\max \text{ spec}\\ &\text{spec}\,\boldsymbol{p}_0\\ &\text{such that } Y(\boldsymbol{p}_0, \text{spec}) \geq Y_{\min}.\end{aligned}$$

We chose to maximize the ugf specification (spec) while

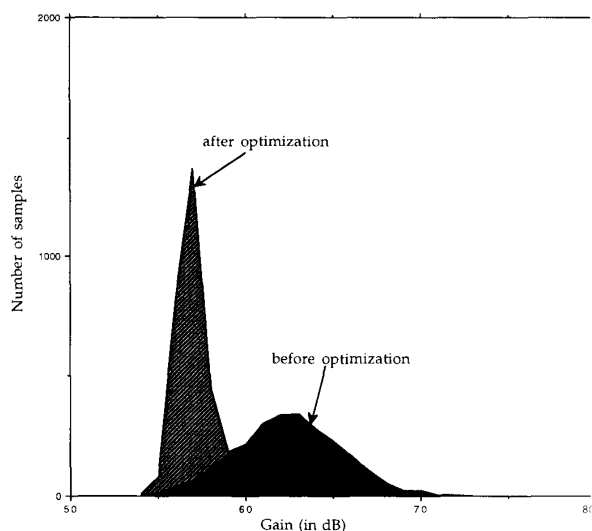


Fig. 8. Distributions of gain (in decibels) in a population of 2500 samples before and after optimization.

TABLE I
MARGINAL YIELDS AT INITIAL POINT

| Performance | Marginal Yield (in%) |
|---|---|
| Bandwidth | 88 |
| DC gain | 100 |
| Ripple | 54 |
| Roll-off | 100 |

maintaining a minimum yield ($Y_{\min}$) of 80%. The initial point had a yield of 100%. The final yield was decreased to 98% and the ugf spec was raised from 45 to 55 MHz.

Next consider the five-pole current mode filter described in [32]. The circuit has a total of 175 MOS transistors. The desired specifications for this circuit are

- dc gain $\geq -1.5$ dB,
- bandwidth $\geq 35$ MHz,
- passband ripple $\leq 0.5$ dB,
- rolloff $\geq 100$ dB/decade.

Disturbances include variations in the width, length, oxide thickness, and flat-band voltage of the n- and p-type devices. We found that linear macromodels in the disturbance variables were sufficiently accurate for all four performances. The designables in this circuit were five independent transistor widths. The initial yield of the circuit was 42%.

The marginal yield of a certain circuit with respect to a particular performance is defined as the fraction of circuits that satisfies that particular performance specification. For this example, the marginal yields at the initial point are given in Table I.

It is not entirely atypical for such a situation to occur with a design. In many cases the design is such that some performance specifications are comfortably satisfied and other performances are quite close to the specification. In other cases some performances that may nominally be well within the specification bounds are extremely sensitive to process variations and cause the circuit yield to be low. This, coupled with the fact that performances are usually competing against each other (e.g., an increase in gain usually causes an decrease in bandwidth), may cause the circuit to have low parametric yield even though it may seem to have good nominal performances and good marginal yields. Since the performances are not independent of each other, the overall parametric yield is not equal to the product of the marginal yields.

TABLE II
SENSITIVITIES OF YIELDS WITH RESPECT TO DISTURBANCE VARIANCES

| Disturbance | Sensitivity of yield wrt variance |
|---|---|
| n flat-band voltage variation | 0.007466858 |
| n-device $t_{ox}$ variation | -0.993829417 |
| n-device length variation | 0.060171695 |
| n-device width variation | -0.009230657 |
| flat-band voltage variation | 0.030542683 |
| p-device $t_{ox}$ variation | -0.058411000 |
| p-device length variation | -0.059349247 |
| p-device width variation | 0.025970466 |

TABLE III
PERCENTAGE CHANGE IN DESIGNABLES

| Width # | % Change |
|---|---|
| 1 | 21.54 |
| 2 | -23.55 |
| 3 | 50.76 |
| 4 | 2.21 |
| 5 | 95.53 |

For this particular circuit, at the initial point the sensitivity of the yield to the dc gain and rolloff performance specification was 0.0. This does not come as a surprise to us after considering that these specifications are comfortably satisfied at the initial point. The normalized sensitivities of the yield to the disturbance variances at the initial point are given in Table II.

We note that variance of the oxide thickness variation of the n-type devices has the most detrimental effect on the yield. Since the variance of a disturbance is a measure of its spread, we can determine which processing step needs to be controlled most stringently to ensure good circuit yield. While such control may require costly equipment or may not be possible for a particular circuit, we might find such information useful when designing newer processing equipment.

Upon optimization, which required only 288 circuit simulations, the yield of this circuit was increased to 98%. The designable parameters changed as given in Table III.

These numbers indicate that the designable parameters may

TABLE IV
FINAL MARGINAL YIELDS

| Performance | Marginal Yield (in %) |
|---|---|
| DC gain | 100 |
| Bandwidth | 99 |
| Ripple | 99 |
| Roll-off | 100 |

undergo a significant percent change during optimization. A very small change is noticed in Width 4 whereas Width 5 has almost doubled during the course of the optimization. The marginal yields at the final point were as given in Table IV.

## VII. CONCLUSIONS

We have shown that statistical design problems can be expressed as optimization problems in which either the objective function or the constraint functions depend on expectations of random variables. Traditionally, one of the deterrents to the use of statistical design methods has been the high cost of circuit/process/device simulation. With the advent of fast computers and the development of methods that work to minimize the cost of simulation, yield maximization methods are gaining acceptance among the circuit design community. In fact, the scaling down of device sizes without a corresponding scaling down of processing variations has made such tools invaluable for circuits designed for manufacture.

Except for an attempt to analyze several methods developed to handle the case of discrete circuit yield optimization [30], no comparative performance studies exist on various IC circuit yield maximization methodologies. In the absence of objective data, we offer our, somewhat biased, conclusion.

The boundary integral method offers the most flexible formulation of the IC statistical design problem. By decoupling the computation of the yield and its gradients with respect to a variety of parameters from the actual optimization formulation, it allows the flexibility of using the objective and constraint functions that are most suited for the problem at hand.

## ACKNOWLEDGMENT

The authors would like to thank M. Anderson for helping with the examples.

## REFERENCES

[1] S. R. Nassif, A. J. Strojwas, and S. W. Director, "FABRICS-II: A statistical based IC fabrication process simulator," *IEEE Trans. Computer-Aided Design*, vol. CAD-3, Jan. 1984.
[2] L. Nagel, "SPICE2: A computer program to simulate semiconductor circuits," Univ. of California, Berkeley, ERL Memo. ERL-M520, May 1975.
[3] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods*. London: Metheun, 1964.
[4] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright, "User's guide for NPSOL (version 4.0)," Systems Optimization Lab., Stanford Univ., Stanford, CA, Tech. Rep. SOL 86-2, Jan. 1986.
[5] P. Cox, P. Yang, S. S. Mahant-Shetti, and P. Chatterjee, "Statistical modeling for efficient parametric yield estimation of MOS VLSI circuits," *IEEE Trans. Electron Devices*, vol. ED-32, pp. 471–478, Feb. 1985.
[6] P. Yang, D. E. Hocevar, P. F. Cox, C. Machala, and P. K. Chatterjee, "An integrated and efficient approach for MOS VLSI statistical circuit design," *IEEE Trans. Computer-Aided Design*, vol. CAD-5, pp. 5–14, Jan. 1986.
[7] D. E. Hocevar, P. F. Cox, and P. Yang, "Parametric yield optimization for MOS circuit blocks," *IEEE Trans. Computer-Aided Design*, vol. 7, pp. 645–658, June 1988.
[8] P. Feldmann and S. W. Director, "Accurate and efficient evaluation of circuit yield and yield gradients," in *Proc. IEEE Int. Conf. Computer-Aided Design*, Nov. 1990, pp. 120–123.
[9] P. Feldmann, "Statistical integrated circuit design," Ph.D. dissertation, Carnegie Mellon Univ., Pittsburgh, PA, Jan. 1991.
[10] P. Feldmann and S. W. Director, "Improved methods for IC yield and quality optimization using surface integrals," in *Proc. IEEE Int. Conf. Computer-Aided Design*, Nov. 1991.
[11] P. Feldmann and S. W. Director, "Integrated circuit quality optimization using surface integrals," submitted to *IEEE Trans. Computer-Aided Design*.
[12] P. Feldmann and S. W. Director, "A Macromodeling approach for increasing efficiency of IC yield optimization," presented at the Int. Symp. Circuits and Syst., 1991.
[13] R. Y. Rubinstein, *Simulation and the Monte Carlo Method*. New York: Wiley, 1981.
[14] M. A. Styblinski and L. J. Opalski, "Algorithms and software tools of IC yield optimization based on fundamental fabrication parameters," *IEEE Trans. Computer-Aided Design*, vol. CAD-5, pp. 79–89, Jan. 1986.
[15] R. Biernacki and M. Styblinski, "Efficient performance function interpolation scheme and its application to statistical circuit design," *Int. J. Circuit Theory Application*, vol. 19, pp. 403–422, 1991.
[16] M. A. Styblinski and S. A. Aftab, "Efficient circuit performance modeling using a combination of interpolation and self organizing approximation techniques," in *1990 IEEE Int. Symp. Circuits Syst. Proc.*, 1990.
[17] L. J. Opalski and M. A. Styblinski, "Generalization of yield optimization problem: The maximum income approach," *IEEE Trans. Computer-Aided Design*, vol. CAD-5, pp. 346–360, Apr. 1986.
[18] M. A. Styblinski and J. C. Zhang, "Orthogonal array approach to gradient based yield optimization," in *1990 IEEE Int. Symp. Circuits Syst. Proc.*, 1990.
[19] M. A. Styblinski and J. C. Zhang, "Circuit performance variability reduction: Principles, problems and practical solutions," presented at the IEEE Int. Conf. Computer-Aided Design, Nov. 1991.
[20] N. R. Draper and H. Smith, *Applied Regression Analysis* (Wiley Series in Probability and Mathematical Statistics). New York: Wiley, 1980.
[21] S. W. Director and R. A. Rohrer, "The generalized adjoint network and network sensitivities," *IEEE Trans. Circuit Theory*, vol. CT-16, pp. 318–323, Aug. 1969.
[22] P. Feldmann, T. V. Nguyen, S. W. Director, and R. A. Rohrer, "Sensitivity computation in piecewise approximate circuit simulation," *IEEE Trans. Computer-Aided Design*, vol. 10, Feb. 1991.
[23] *SAS User's Guide: Statistics*, Version 5 ed., SAS Institute, Cary, NC, 1985.
[24] J. Sacks, S. B. Schiller, and W. J. Welch, "Designs for computer experiments," *Technometrics*, vol. 31, no. 1, pp. 41–47, Feb. 1989.
[25] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, no. 2, pp. 239–245, May 1979.
[26] M. S. Phadke, *Quality Engineering Using Robust Design*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[27] M. J. M. Pelgrom, AAD C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1440, Oct. 1989.
[28] M. A. Styblinski and A. Ruszczynski, "Stochastic approximation approach to statistical circuit design," *Electron. Lett.*, vol. 19, no. 8, pp. 300–302, 1983.
[29] A. Strojwas, Ed., *Selected Papers on Statistical Design of Integrated Circuits*. New York: IEEE Press, 1987.
[30] E. Wehrhahn and R. Spence, "The performance of some design centering methods," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 1984, pp. 1424–1438.
[31] W. J. Welch, T.-K. Yu, S.-M. Kang, and J. Sacks, "Computer experiments for quality control by parameter design," *J. Quality Technology*, vol. 22, pp. 15–22, 1990.
[32] S.-S. Lee *et al.*, "A 40 MHz CMOS continuous-time current-mode filter," in *Proc. IEEE 1992 CICC*, pp. 24.5.1–24.5.4.

**Stephen W. Director** (S'65–M'69–SM'75–F'78) received the B.S. degree from the State University of New York at Stony Brook in 1965 and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1967 and 1968, respectively.

From 1968 until 1977 he was with the Department of Electrical Engineering at the University of Florida, Gainesville. From Sept. 1974 to Aug. 1975 he was a Visiting Scientist in the Mathematical Sciences Department at IBM's T. J. Watson Research Center, Yorktown Heights, NY. He joined Carnegie Mellon University, Pittsburgh, PA, in 1977 and served as Head of the Department of Electrical and Computer Engineering from 1982 to 1991. In 1982 he founded the SRC-CMU Research Center for Computer-Aided Design and served as its Director from 1982 to 1989. He is now Dean of the College of Engineering and U. A. and Helen Whitaker University Professor of Electrical and Computer Engineering.

Dr. Director has served as President of the IEEE Circuits and Systems Society, as Chairman of the CAS Technical Committee on Computer-Aided Network Design (CANDE), and as Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS. He has authored or coauthored four texts. In 1970 and 1985 he received Best Paper Awards from the IEEE Circuits and Systems Society; in 1976 he received the Frederick Emmons Terman Award from the American Society of Engineering Education; and in 1978 he received the W. R. G. Baker Prize Paper Award from the IEEE. In 1984 he received an IEEE Centennial Medal and was named a Distinguished Alumnus for the State University of New York, Stony Brook. He was elected to membership in the National Academy of Engineering in 1989. In 1992 he received the Society Award from the Circuits and Systems Society and a Best Paper Award from the ACM/IEEE Design Automation Conference.

**Peter Feldmann** (S'88–M'91) was born in Timisoara, Romania, in 1958. He received the B.Sc. degree in computer engineering in 1983 and the M.Sc. degree in electrical engineering in 1987, both from the Technion in Haifa, Israel, and the Ph.D. degree in 1991 from Carnegie Mellon in Pittsburgh, PA.

From 1985 through 1987 he designed digital signal processors at Zoran Microelectronics in Haifa. Currently, he is on the technical staff at Bell Labs in Murray Hill, NJ. His research interests include CAD for VLSI circuits, more specifically circuit simulation, optimization, and design for manufacturability.

**Kannan Krishna** (S'89) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, New Delhi, India, in 1990 and the M.S. degree from the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, in 1991, where he is currently working toward the Ph.D. degree. His research interests are in the areas of statistical design, device modeling, and circuit simulation.

Currently he is holding a temporary position as Member of Technical Staff at AT&T Bell Laboratories, Murray Hill, NJ.