

Circuit Optimization: The State of the Art

JOHN W. BANDLER, FELLOW, IEEE, AND SHAO HUA CHEN, STUDENT MEMBER, IEEE

Invited Paper

Abstract—This paper reviews the current state of the art in circuit optimization, emphasizing techniques suitable for modern microwave CAD. It is directed at the solution of realistic design and modeling problems, addressing such concepts as physical tolerances and model uncertainties. A unified hierarchical treatment of circuit models forms the basis of the presentation. It exposes tolerance phenomena at different parameter/response levels. The concepts of design centering, tolerance assignment, and postproduction tuning in relation to yield enhancement and cost reduction suitable for integrated circuits are discussed. Suitable techniques for optimization oriented worst-case and statistical design are reviewed. A generalized I_p centering algorithm is proposed and discussed. Multicircuit optimization directed at both CAD and robust device modeling is formalized. Tuning is addressed in some detail, both at the design stage and for production alignment. State-of-the-art gradient-based nonlinear optimization methods are reviewed, with emphasis given to recent, but well-tested, advances in minimax, I_1 , and I_2 optimization. Illustrative examples as well as a comprehensive bibliography are provided.

I. INTRODUCTION

COMPUTER-AIDED circuit optimization is certainly one of the most active areas of interest. Its advances continue; hence the subject deserves regular review from time to time. The classic paper by Temes and Calahan in 1967 [102] was one of the earliest to formally advocate the use of iterative optimization in circuit design. Techniques that were popular at the time, such as one-dimensional (single-parameter) search, the Fletcher–Powell procedure and the Remez method for Chebyshev approximation, were described in detail and well illustrated by circuit examples. Pioneering papers by Lasdon, Suchman, and Warren [73], [74], [108] demonstrated optimal design of linear arrays and filters using the penalty function approach. Two papers in 1969 by Director and Rohrer [48], [49] originated the adjoint network approach to sensitivity calculations, greatly facilitating the use of powerful gradient-based optimization methods. In the same period, the work by Bandler [4], [5] systematically treated the formulation of error functions, the least p th objective, nonlinear constraints, optimization methods, and circuit sensitivity analysis.

Manuscript received May 4, 1987; revised August 20, 1987. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under Grant A7239 and in part by Optimization Systems Associates Inc.

J. W. Bandler is with the Simulation Optimization Systems Research Laboratory and the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada L8S 4L7. He is also with Optimization Systems Associates Inc., Dundas, Ontario, Canada L9H 6L1.

S. H. Chen was with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada. He is now with Optimization Systems Associates Inc., Dundas, Ontario, Canada L9H 6L1. IEEE Log Number 8717974.

Since then, advances have been made in several major directions. The development of large-scale network simulation and optimization techniques have been motivated by the requirements of the VLSI era. Approaches to realistic circuit design where design parameter tolerances and yield are taken into account have been pioneered by Elias [52] and Karafin [68] and furthered by many authors over the ensuing years. Optimization methods have evolved from simple, low-dimension-oriented algorithms into sophisticated and powerful ones. Highly effective and efficient solutions have been found for a large number of specialized applications. The surveys by Calahan [37], Charalambous [39], Bandler and Rizk [26], Hachtel and Sangiovanni-Vincentelli [63], and Brayton *et al.* [32] are especially relevant to circuit designers.

In the present paper, we concentrate on aspects that are relevant to and necessary for the continuing move to optimization of increasingly more complex microwave circuits, in particular to MMIC circuit modeling and design. Consequently, we emphasize optimization-oriented approaches to deal more explicitly with process imprecision, manufacturing tolerances, model uncertainties, measurement errors, and so on. Such realistic considerations arise from design problems in which a large volume of production is envisaged, e.g., integrated circuits. They also arise from modeling problems in which consistent and reliable results are expected despite measurement errors, structural limitations such as physically inaccessible nodes, and model approximations and simplifications. The effort to formulate and solve these problems represents one of the driving forces of theoretical study in the mathematics of circuit CAD. Another important impetus is provided by progress in computer hardware, resulting in drastic reduction in the cost of mass computation. Finally, the continuing development of gradient-based optimization techniques has provided us with powerful tools.

In this context, we review the following concepts: realistic representations of a circuit design and modeling problem, nominal (single) circuit optimization, statistical circuit design, and multicircuit modeling, as well as recent gradient-based optimization methods.

Nominal design and modeling are the conventional approaches used by microwave engineers. Here, we seek a single point in the space of variables selected for optimization which best meets a given set of performance specifications (in design) or best matches a given set of response measurements (in modeling). A suitable scalar measure

of the deviation between responses and specifications which forms the objective function to be minimized is the ubiquitous least squares measure (see, for example, Morrison [83]), the more esoteric generalized l_p objective (Charalambous [41]) or the minimax objective (Madsen *et al.* [80]). We observe here that the performance-driven (single-circuit) least squares approach that circuit design engineers have traditionally chosen has proved unsuccessful both in addressing design yield and in serious device modeling.

Recognition that an actual realization of a nominal design is subject to fluctuation or deviation led, in the past, to the so-called sensitivity minimization approach (see, for example, Schoeffler [94] and Laker *et al.* [71]). Employed by filter designers, the approach involves measures of performance sensitivity, typically first-order, that are included in the objective function.

In reality, uncertainties which deteriorate performance may be due to physical (manufacturing, operating) tolerances as well as to parasitic effects such as electromagnetic coupling between elements, dissipation, and dispersion (Bandler [6], Tromp [107]). In the design of substantially untunable circuits these phenomena lead to two important classes of problems: worst-case design and statistical design. The main objective is the reduction of cost or the maximization of production yield.

Worst-case design (Bandler *et al.* [23], [24]), in general, requires that all units meet the design specifications under all circumstances (i.e., a 100 percent yield), with or without tuning, depending on what is practical. In statistical design [1], [26], [30], [47], [97], [98], [100], [101] it is recognized that a yield of less than 100 percent is likely; therefore, with respect to an assumed probability distribution function, yield is estimated and enhanced by optimization. Typically, we either attempt to center the design with fixed assumed tolerances or we attempt to optimally assign tolerances and/or design tunable elements to reduce production cost.

What distinguishes all these problems from nominal designs or sensitivity minimization is the fact that a single design point is no longer of interest: a (tolerance) region of multiple possible outcomes is to be optimally located with respect to the acceptable (feasible, constraint) region.

Modeling, often unjustifiably treated as if it were a special case of design, is particularly affected by uncertainties and errors at many levels. Unavoidable measurement errors, limited accessibility to measurement points, approximate equivalent circuits, etc., result in nonunique and frequently inconsistent solutions. To overcome these frustrations, we advocate a properly constituted multicircuit approach (Bandler *et al.* [12]).

Our presentation is outlined as follows.

In Section II, in relation to a physical engineering system of interest, a typical hierarchy of simulation models and corresponding response and performance functions are introduced. Error functions arising from given specifications and a vector of optimization variables are defined. Performance measures such as l_p objective functions (l_p

norms and generalized l_p functions) are introduced and their properties discussed.

We devote to Section III a brief review of the relatively well-known and successful approach of nominal circuit design optimization.

In Section IV, uncertainties that exist in the physical system and at different levels of the model hierarchy are discussed and illustrated by a practical example. Different cases of multicircuit design, namely centering, tolerancing (optimal tolerance assignment), and tuning at the design stage, are identified. A multicircuit modeling approach and several possible applications are described.

Some important and representative techniques in worst-case and statistical design are reviewed in Section V. These include the nonlinear programming approach to worst-case design (Bandler *et al.* [24], Polak [89]), simplicial (Director and Hachtel [47]) and multidimensional (Bandler and Abdel-Malek [7]) approximations of the acceptable region, the gravity method (Soin and Spence [98]), and the parametric sampling method (Singhal and Pinel [97]). A generalized l_p centering algorithm is proposed as a natural extension to l_p nominal design. It provides a unified formulation of yield enhancement for both the worst case and the case where yield is less than 100 percent.

Illustrations of statistical design are given in Section VI.

The studies in the last two decades on the theoretical and algorithmic aspects of optimization techniques have produced a great number of results. In particular, gradient-based optimization methods have gained increasing popularity in recent years for their effectiveness and efficiency. The essence of gradient-based l_p optimization methods is reviewed in Section VII. Emphasis is given to the trust region Gauss-Newton and the quasi-Newton algorithms (Madsen [78], Moré [82], Dennis and Moré [46]).

The subject of gradient calculation and approximation is briefly discussed in Section VIII.

II. VARIABLES AND FUNCTIONS

In this section, we review some basic concepts of practical circuit optimization. In particular, we identify a physical system and its simulation models. We discuss a typical hierarchy of models and the associated designable parameters and response functions. We also define specifications, error functions, optimization variables and objective functions.

A. The Physical System

The physical engineering system under consideration can be a network, a device, a process, and so on, which has both a fixed structure and given element types. We manipulate the system through some adjustable parameters contained in the column vector ϕ^M . The superscript M identifies concepts related to the physical system. Geometrical dimensions such as the width of a strip and the length of a waveguide section are examples of adjustable parameters.

In the production of integrated circuits, ϕ^M may include some fundamental variables which control, say, a doping or photomasking process and, consequently, determine the geometrical and electrical parameters of a chip. External controls, such as the biasing voltages applied to an active device, are also possible candidates for ϕ^M .

The performance and characteristics of the system are described in terms of some measurable quantities. The usual frequency and transient responses are typical examples. These measured responses, or simply measurements, are denoted by $F^M(\phi^M)$.

B. The Simulation Models

In circuit optimization, some suitable models are used to simulate the physical system. Actually, models can be usefully defined at many levels. Tromp [106], [107] has considered an arbitrary number of levels (also see Bandler *et al.* [19]). Here, for simplicity, we consider a hierarchy of models consisting of four typical levels as

$$\begin{aligned} F^H &= F^H(F^L) \\ F^L &= F^L(\phi^H) \\ \phi^H &= \phi^H(\phi^L). \end{aligned} \quad (1)$$

ϕ^L is a set of low-level model parameters. It is supposed to represent, as closely as possible, the adjustable parameters in the actual system, i.e., ϕ^M . ϕ^H defines a higher-level model, typically an equivalent circuit, with respect to a fixed topology. Usually, we use an equivalent circuit for the convenience of its analysis. The relationship between ϕ^L and ϕ^H is either derived from theory or given by a set of empirical formulas.

Next on the hierarchy we define the model responses at two possible levels. The low-level external representation, denoted by F^L , can be the frequency-dependent complex scattering parameters, unterminated y -parameters, transfer function coefficients, etc. Although these quantities may or may not be directly measurable, they are very often used to represent a subsystem. The high-level responses F^H directly correspond to the actual measured responses, namely F^M , which may be, for example, frequency responses such as return loss, insertion loss, and group delay of a suitably terminated circuit.

A realistic example of a one-section transformer on stripline was originally considered by Bandler *et al.* [25]. The circuits and parameters, physical as well as model, are shown in Fig. 1. The physical parameters ϕ^M (and the low-level model ϕ^L) include strip widths, section lengths, dielectric constants, and strip and substrate thicknesses. The equivalent circuit has six parameters, considered as ϕ^H , including the effective line widths, junction parasitic inductances, and effective section length. The scattering matrix of the circuit with respect to idealized (matched) terminations is a candidate for a low-level external representation (F^L). The reflection coefficient by taking into account the actual complex terminations could be a high-level response of interest (F^H).

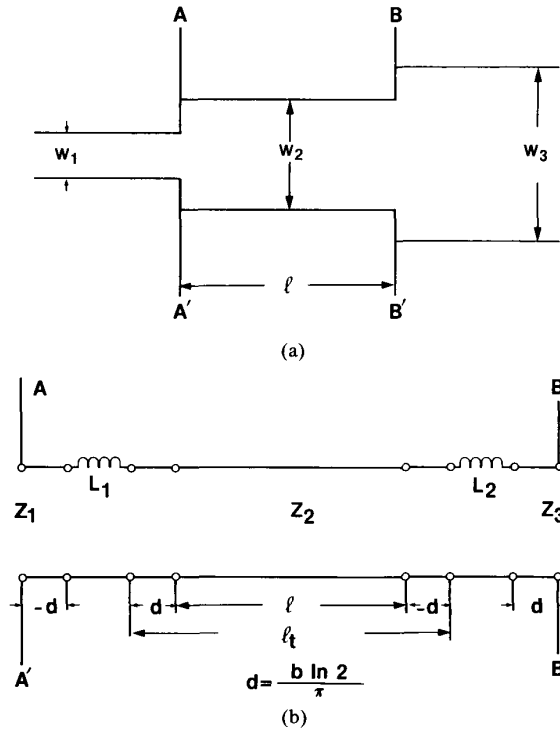


Fig. 1. A microwave stripline transformer showing (a) the physical structure and (b) the equivalent circuit model [25]. The physical parameters are

$$\phi^M = [w_1 w_2 w_3 : l \sqrt{\epsilon_{r1}} \sqrt{\epsilon_{r2}} \sqrt{\epsilon_{r3}} b_1 b_2 b_3 t_{s1} t_{s2} t_{s3}]^T$$

where w is the strip width, l the length of the middle section, ϵ_r the dielectric constant, b the substrate thickness, and t_s the strip thickness. ϕ^M is represented in the simulation model by ϕ^L . The high-level parameters of the equivalent circuit are

$$\phi^H = [D_1 D_2 D_3 L_1 L_2 l_t]^T$$

where D is the effective linewidth, L the junction parasitic inductance, and l_t the effective section length. Suitable empirical formulas that relate ϕ^L to ϕ^H can be found in [25].

For a particular case, we may choose a certain section of this hierarchy to form a design problem. We can choose either ϕ^L or ϕ^H as the designable parameters. Either F^L or F^H or a suitable combination of both may be selected as the response functions. Bearing this in mind, we simplify the notation by using ϕ for the designable parameters and F for the response functions.

C. Specifications and Error Functions

The following discussion on specifications and error functions is based on presentations by Bandler [5], and Bandler and Rizk [26], where more exhaustive illustrations can be found.

We express the desirable performance of the system by a set of specifications which are usually functions of certain independent variable(s) such as frequency, time, and temperature. In practice, we have to consider a discrete set of samples of the independent variable(s) such that satisfying the specifications at these points implies satisfying them

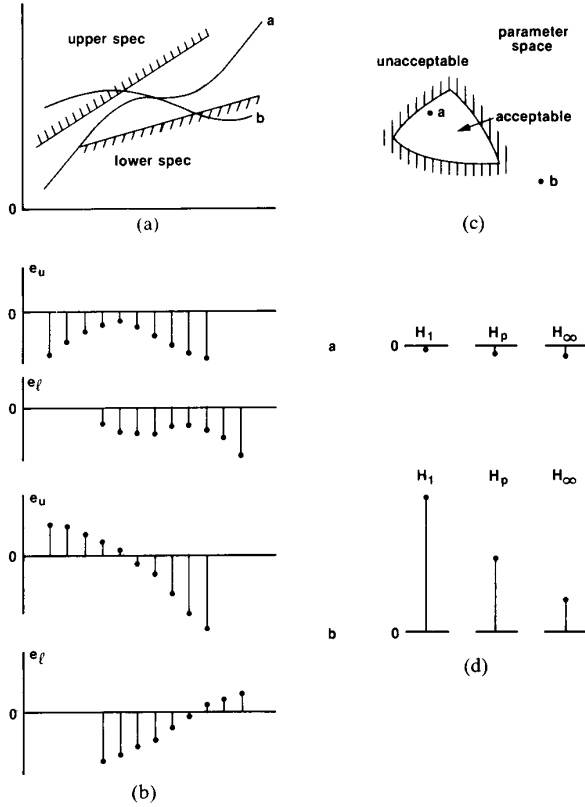


Fig. 2. Illustrations of (a) upper specifications, lower specifications, and the responses of circuits *a* and *b*, (b) error functions corresponding to circuits *a* and *b*, (c) the acceptable region, and (d) generalized l_p objective functions defined in (13).

almost everywhere. Also, we may consider simultaneously more than one kind of response. Thus, without loss of generality, we denote a set of sampled specifications and the corresponding set of calculated response functions by, respectively,

$$\begin{aligned} S_j, & \quad j=1,2,\dots,m \\ F_j(\phi), & \quad j=1,2,\dots,m. \end{aligned} \quad (2)$$

Error functions arise from the difference between the given specifications and the calculated responses. In order to formulate the error functions properly, we may wish to distinguish between having upper and lower specifications (windows) and having single specifications, as illustrated in Figs. 2(a) and 3(a). Sometimes the one-sidedness of upper and lower specifications is quite obvious, as in the case of designing a bandpass filter. On other occasions the distinction is more subtle, since a single specification may as well be interpreted as a window having zero width.

In the case of having single specifications, we define the error functions by

$$e_j(\phi) = w_j |F_j(\phi) - S_j|, \quad j=1,2,\dots,m \quad (3)$$

where w_j is a nonnegative weighting factor.

We may also have an upper specification S_{uj} and a lower specification S_{lj} . In this case we define the error

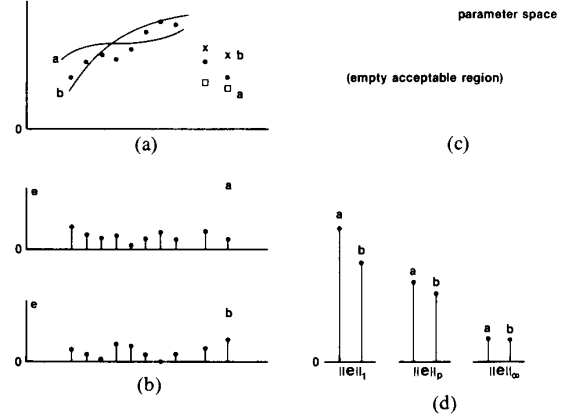


Fig. 3. Illustrations of (a) a discretized single specification and two discrete single specifications (e.g., expected parameter values to be matched), as well as the responses of circuits *a* and *b*, (b) error functions related to circuits *a* and *b*, (c) the (empty) acceptable region (i.e., a perfect match is not possible) and (d) the corresponding l_p norms.

functions as

$$\begin{aligned} e_{uj}(\phi) &= w_{uj} (F_j(\phi) - S_{uj}), & j \in J_u \\ e_{lj}(\phi) &= w_{lj} (F_j(\phi) - S_{lj}), & j \in J_l \end{aligned} \quad (4)$$

where w_{uj} and w_{lj} are nonnegative weighting factors. The index sets as defined by

$$\begin{aligned} J_u &= \{j_1, j_2, \dots, j_k\} \\ J_l &= \{j_{k+1}, j_{k+2}, \dots, j_m\} \end{aligned} \quad (5)$$

are not necessarily disjoint (i.e., we may have simultaneous specifications). In order to have a set of uniformly indexed error functions, we let

$$\begin{aligned} e_i &= e_{uj}(\phi), & j = j_i, & i = 1, 2, \dots, k \\ e_i &= -e_{lj}(\phi), & j = j_i, & i = k+1, k+2, \dots, m. \end{aligned} \quad (6)$$

The responses corresponding to the single specifications can be real or complex, whereas upper and lower specifications are applicable to real responses only. Notice that, in either case, the error functions are real. Clearly, a positive (nonpositive) error function indicates a violation (satisfaction) of the corresponding specification. Figs. 2(b) and 3(b) depict the concept of error functions.

D. Optimization Variables and Objective Functions

Mathematically, we abstract a circuit optimization problem by the following statement:

$$\underset{x}{\text{minimize}} U(x) \quad (7)$$

where x is a set of optimization variables and $U(x)$ a scalar objective function.

Optimization variables and model parameters are two separate concepts. As will be elaborated on later in this paper, x may contain a subset of ϕ which may have been normalized or transformed, it may include some statistical variables of interest, several parameters in ϕ may be tied to one variable in x , and so on.

Typically, the objective function $U(x)$ is closely related to an l_p norm or a generalized l_p function of $e(\phi)$. We shall review the definitions of such l_p functions and discuss their appropriate use in different contexts.

E. The l_p Norms

The l_p norm (Temes and Zai [103]) of e is defined as

$$\|e\|_p = \left[\sum_{j=1}^m |e_j|^p \right]^{1/p}. \quad (8)$$

It provides a scalar measure of the deviations of the model responses from the specifications. Least-squares (l_2) is perhaps the most well-known and widely used norm (Morrison [83]), which is

$$\|e\|_2 = \left[\sum_{j=1}^m |e_j|^2 \right]^{1/2}. \quad (9)$$

The l_2 objective function is differentiable and its gradient can be easily obtained from the partial derivatives of e . Partly due to this property, a large variety of l_2 optimization techniques have been developed and popularly implemented. For example, the earlier versions of the commercial CAD packages TOUCHSTONE [104] and SUPER-COMPACT [99] have provided designers solely the least-squares objective.

The parameter p has an important implication. By choosing a large (small) value for p , we in effect place more emphasis on those error functions (e_j 's) that have larger (smaller) values. By letting $p = \infty$ we have the minimax norm

$$\|e\|_\infty = \max_j |e_j| \quad (10)$$

which directs all the attention to the worst case and the other errors are in effect ignored. Minimax optimization is extensively employed in circuit design where we wish to satisfy the specifications in an optimal equal-ripple manner [3], [13], [14], [21], [40], [42], [65], [67], [80], [85].

On the other hand, the use of the l_1 norm, as defined by

$$\|e\|_1 = \sum_{j=1}^m |e_j| \quad (11)$$

implies attaching more importance to the error functions that are closer to zero. This property has led to the application of l_1 to data-fitting in the presence of gross errors [22], [29], [66], [86] and, more recently, to fault location [8], [9], [27] and robust device modeling [12].

Notice that neither $\|e\|_\infty$ nor $\|e\|_1$ is differentiable in the ordinary sense. Therefore, their minimization requires algorithms that are much more sophisticated than those for the l_2 optimization.

F. The One-Sided and Generalized l_p Functions

By using an l_p norm, we try to minimize the errors towards a zero value. In cases where we have upper and lower specifications, a negative value of e_j simply indicates that the specification is exceeded at that point which, in a

sense, is better than having $e_j = 0$. This fact leads to the one-sided l_p function defined by

$$H_p^+(e) = \left[\sum_{j \in J} |e_j|^p \right]^{1/p} \quad (12)$$

where $J = \{j | e_j \geq 0\}$. Actually, if we define $e_j^+ = \max\{e_j, 0\}$, then $H_p^+(e) = \|e^+\|_p$.

Bandler and Charalambous [10], [41] have proposed the use of a generalized l_p function defined by

$$H_p(e) = \begin{cases} H_p^+(e) & \text{if the set } J \text{ is not empty} \\ H_p^-(e) & \text{otherwise} \end{cases} \quad (13)$$

where

$$H_p^-(e) = - \left[\sum_{j=1}^m (-e_j)^{-p} \right]^{-1/p}. \quad (14)$$

In other words, when at least one of the e_j is nonnegative we use H_p^+ , and H_p^- is defined if all the error functions have become negative.

Compared to (12), the generalized l_p function has an advantage in the fact that it is meaningfully defined for the case where all the e_j are negative. This permits its minimization to proceed even after all the specifications have been met, so that the specifications may be further exceeded.

A classical example is the design of Chebyshev-type bandpass filters, where we have to minimize the generalized minimax function

$$H_\infty(e) = \max_j \{e_j\}. \quad (15)$$

The current Version 1.5 of TOUCHSTONE [105] offers the generalized l_p optimization techniques, including minimax.

G. The Acceptable Region

We use $H(e)$ as a generic notation for $\|e\|_p$, $H_p^+(e)$, and $H_p^-(e)$. The sign of $H(e(\phi))$ indicates whether or not all the specifications are satisfied by ϕ . An acceptable region is defined as

$$R_a = \{\phi | H(e(\phi)) \leq 0\} \quad (16)$$

Figs. 2(c), 2(d), 3(c), and 3(d) depict the l_p functions and the acceptable regions.

III. NOMINAL CIRCUIT OPTIMIZATION

In a nominal design, without considering tolerances (i.e., assuming that modeling and manufacturing can be done with absolute accuracy), we seek a single set of parameters, called a nominal point and denoted by ϕ^0 , which satisfies the specifications. Furthermore, if we consider the functional relationship of $\phi^H = \phi^H(\phi^L)$ to be precise, then it does not really matter at which level the design is conceived. In fact, traditionally it is often oriented to an equivalent circuit. A classical case is network synthesis

where $\phi^{H,0}$ is obtained through the use of an equivalent circuit and/or a transfer function. A low-level model $\phi^{L,0}$ is then calculated from $\phi^{H,0}$, typically with the help of an empirical formula (e.g., the number of turns of a coil is calculated for a given inductance). Finally, we try to realize $\phi^{L,0}$ by its physical counterpart $\phi^{M,0}$.

With the tool of mathematical optimization, the nominal point ϕ^0 (at a chosen level) is obtained through the minimization of $U(x)$, where the objective function is typically defined as an l_p function $H(e)$. The vector x contains all the elements of or a subset of the elements of ϕ^0 . It is a common practice to have some of the variables normalized. It is also common to have several model parameters tied to a single variable. This is true, e.g., for symmetrical circuit structures but, most importantly, it is a fact of life in integrated circuits. Indeed, such dependencies should be taken into account both in design and in modeling to reduce the dimensionality. The minimax optimization of manifold multiplexers as described by Bandler *et al.* [18], [22], [28] provides an excellent illustration of large-scale nominal design of microwave circuits.

Traditionally, the approach of nominal design has been extended to solving modeling problems. A set of measurements made on the physical system serves as single specifications. Error functions are created from the differences between the calculated responses $F(\phi^0)$ and the measured responses F^M . By minimizing an l_p norm of the error functions, we attempt to identify a set of model parameters ϕ^0 such that $F(\phi^0)$ best matches F^M . This is known as data fitting or parameter identification.

Such a casual treatment of modeling as if it were a special case of design is often unjustifiable, due to the lack of consideration to the uniqueness of the solution. In design, one satisfactory nominal point, possibly out of many feasible solutions, may suffice. In modeling, however, the uniqueness of the solution is almost always essential to the problem. Affected by uncertainties at many levels, unavoidable measurement errors and limited accessibility to measurement points, the model obtained by a nominal optimization is often nonunique and unreliable. To overcome these frustrations, a recent multicircuit approach will be described in Section IV.

IV. A MULTICIRCUIT APPROACH

The approach of nominal circuit optimization, which we have described in Section III, focuses attention on a certain kind of idealized situation. In reality, unfortunately, there are many uncertainties to be accounted for. For the physical system, without going into too many details, consider

$$\begin{aligned} F^M &= F^{M,0}(\phi^M) + \Delta F^M \\ \phi^M &= \phi^{M,0} + \Delta\phi^M \end{aligned} \quad (17)$$

where ΔF^M represents measurement errors, $\phi^{M,0}$ a nominal value for ϕ^M , and $\Delta\phi^M$ some physical (manufacturing, operating) tolerances.

For simulation purposes, we may consider a realistic representation of the hierarchy of possible models as

$$\begin{aligned} F^H &= F^{H,0}(F^L) + \Delta F^H \\ F^L &= F^{L,0}(\phi^H) + \Delta F^L \\ \phi^H &= \phi^{H,0}(\phi^L) + \Delta\phi^H \\ \phi^L &= \phi^{L,0} + \Delta\phi^L \end{aligned} \quad (18)$$

where $\phi^{L,0}$, $\phi^{H,0}$, $F^{L,0}$, and $F^{H,0}$ are nominal models applicable at different levels. $\Delta\phi^L$, $\Delta\phi^H$, ΔF^L , and ΔF^H represent uncertainties or inaccuracies associated with the respective models. $\Delta\phi^L$ corresponds to the tolerances $\Delta\phi^M$. $\Delta\phi^H$ may be due to the approximate nature of an empirical formula. Parasitic effects which are not adequately modeled in ϕ^H will contribute to ΔF^L , and finally we attribute anything else that causes a mismatch between $F^{H,0}$ and $F^{M,0}$ to ΔF^H .

These concepts can be illustrated by the one-section stripline transformer example [25] which we have considered in Section II. Tolerances may be imposed on the physical parameters including the strip widths and thicknesses, the dielectric constants, the section length and substrate thicknesses (see Fig. 1). Such tolerances correspond to $\Delta\phi^M$ and are represented in the model by $\Delta\phi^L$. We may also use $\Delta\phi^H$ to represent uncertainties associated with the empirical formulas which relate the physical parameters to the equivalent circuit parameters (the effective line widths, the junction inductances, and the effective section length). Mismatches in the terminations at different frequencies may be estimated by ΔF^H (F^H being the actual reflection coefficient; see [25] for more details).

The distinction between different levels of model uncertainties can be quite subtle. As an example, consider the parasitic resistance r associated with an inductor whose inductance is L . Both L and r are functions of the number of turns of a coil (which is a physical parameter). Depending on whether or not r is modeled by the equivalent circuit (i.e., whether or not r is included in ϕ^H), the uncertainty associated with r may appear in $\Delta\phi^H$ or in ΔF^L .

When such uncertainties are present, a single nominal model often fails to represent satisfactorily the physical reality. One effective solution to the problem is to simultaneously consider multiple circuits. We discuss the consequences for design and modeling separately.

A. Multicircuit Design

Our primary concern is to improve production yield and reduce cost in the presence of tolerances $\Delta\phi^L$ and model uncertainties $\Delta\phi^H$. First of all, we represent a realistic situation by multiple circuits as

$$\phi^k = \phi^0 + s^k, \quad k = 1, 2, \dots, K \quad (19)$$

where ϕ^0 , ϕ^k , and s^k are generic notation for the nominal parameters, the k th set of parameters, and a deviate due to the uncertainties, respectively. A more elaborate definition is developed as we proceed.

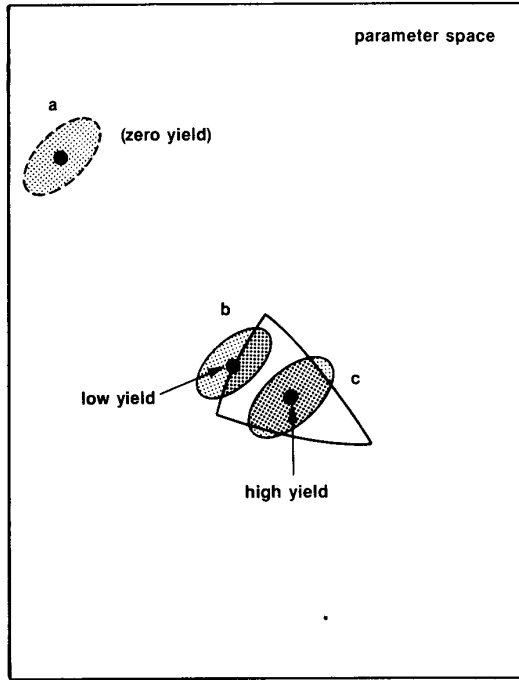


Fig. 4. Three nominal points and the related yield.

For each circuit, we define an acceptance index by

$$I_a(\phi) = \begin{cases} 1, & \text{if } H(e(\phi)) \leq 0 \\ 0, & \text{otherwise} \end{cases} \quad (20)$$

where $H(e) \leq 0$, defined in (13), indicates satisfaction of the specifications by ϕ . An estimate of the yield is given by the percentage of acceptable samples out of the total, as

$$Y \approx \left[\sum_{k=1}^K I_a(\phi^k) \right] / K. \quad (21)$$

The merit of a design can then be judged more realistically according to the yield it promises, as illustrated in Fig. 4. Now we shall have a closer look at the definition of multiple circuits.

In the Monte Carlo method the deviates s^k are constructed by generating random numbers using a physical process or arithmetical algorithms. Typically, we assume a statistical distribution for $\Delta\phi^L$, denoted by $D^L(\epsilon^L)$ where ϵ^L is a vector of tolerance variables. For example, we may consider a multidimensional uniform distribution on $[-\epsilon^L, \epsilon^L]$. Similarly, we assume a $D^H(\epsilon^H)$ for $\Delta\phi^H$. The uniform and Gaussian (normal) distributions are illustrated in Fig. 5.

At the low level, consider

$$\phi^{L,k} = \phi^{L,0} + s^{L,k}, \quad k=1,2,\dots,K^L \quad (22)$$

where $s^{L,k}$ are samples from D^L . At the higher level, we have, for each k ,

$$\phi^{H,k,i} = \phi^{H,0} + s^{H,k,i}, \quad i=1,2,\dots,K^H \quad (23)$$

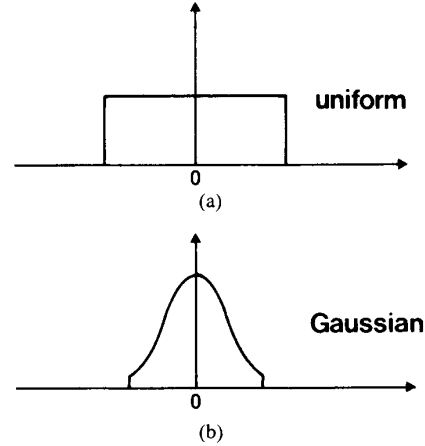


Fig. 5. Typical tolerance distributions: uniform and Gaussian (normal).

where

$$\begin{aligned} \phi^{H,0} &= \phi^{H,0}(\phi^{L,0}) \\ s^{H,k,i} &= \phi^{H,0}(\phi^{L,k}) - \phi^{H,0}(\phi^{L,0}) + \delta^{k,i} \end{aligned} \quad (24)$$

with $\delta^{k,i}$ being samples from D^H .

One might propose a distribution for $s^{H,k,i}$ which presumably encompasses the effect of distribution D^L and distribution D^H . But, while we may reasonably assume simple and independent distributions for $\Delta\phi^L$ and $\Delta\phi^H$, the compound distribution is likely to be complicated and correlated.

B. Centering, Tolerancing, and Tuning

Again, in order to simplify the notation, we use ϕ^0 for the nominal circuit and ϵ for the tolerance variables.

An important problem involves design centering with fixed tolerances, usually relative to corresponding nominal values. We call this the fixed tolerance problem (FTP). The optimization variables are elements of ϕ^0 , the elements of ϵ are constant or dependent on the variables, and the objective is to improve the yield. Incidentally, the nominal optimization problem, i.e., the traditional design problem, is sometimes referred to as the zero tolerance problem (ZTP).

Since imposing tight tolerances on the parameters will increase the cost of component fabrication or process operation, we may attempt to maximize the allowable tolerances subject to an acceptable yield. In this case both ϕ^0 and ϵ may be considered as variables. Such a problem is referred to as optimal tolerancing, optimal tolerance assignment, or the variable tolerance problem (VTP).

Tuning some components of ϕ^M after production, whether by the manufacturer or by a customer, is quite commonly used as a means of improving the yield. This process can also be simulated using the model by introducing a vector of designable tuning adjustments τ^k for each circuit, as

$$\phi^k = \phi^0 + s^k + \tau^k, \quad k=1,2,\dots,K. \quad (25)$$

We have to determine, through optimization, the value of

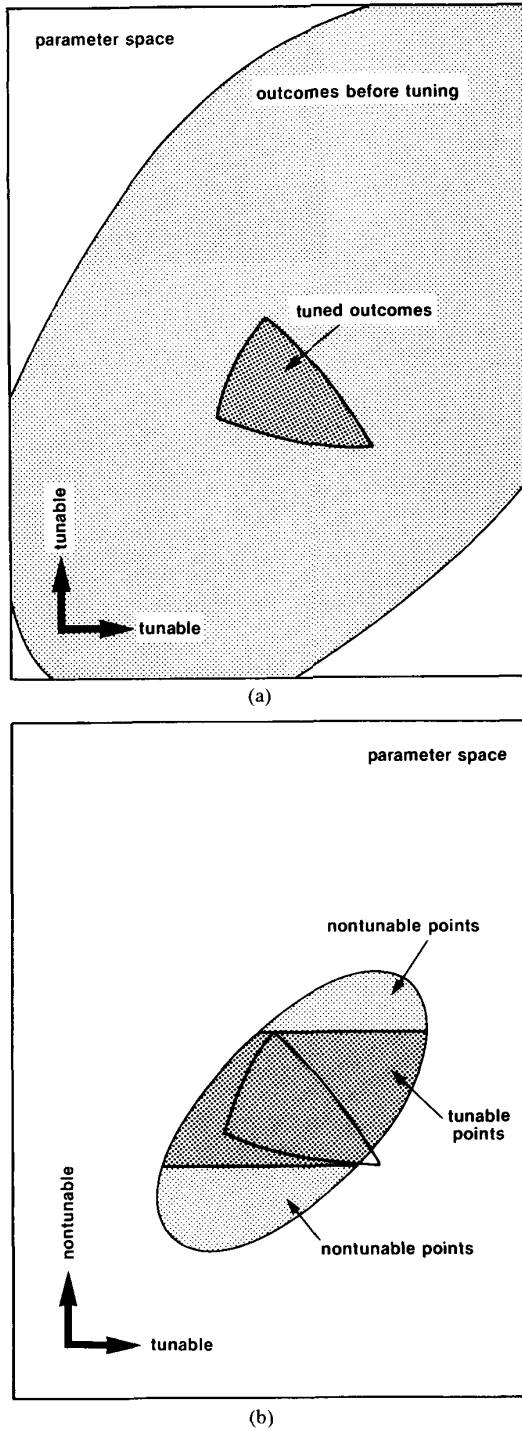


Fig. 6. Illustrations of tuning: (a) both parameters are tunable for a case in which the probability that an untuned design meets the specifications is very low and (b) only one parameter is tunable.

τ^k such that the specifications will be satisfied at ϕ^k which may otherwise be unacceptable, as depicted in Figs. 6 and 7. The introduction of tuning, on the other hand, also increases design complexity and manufacturing cost. We seek a suitable compromise by solving an optimization problem in which τ^k are treated as part of the variables.

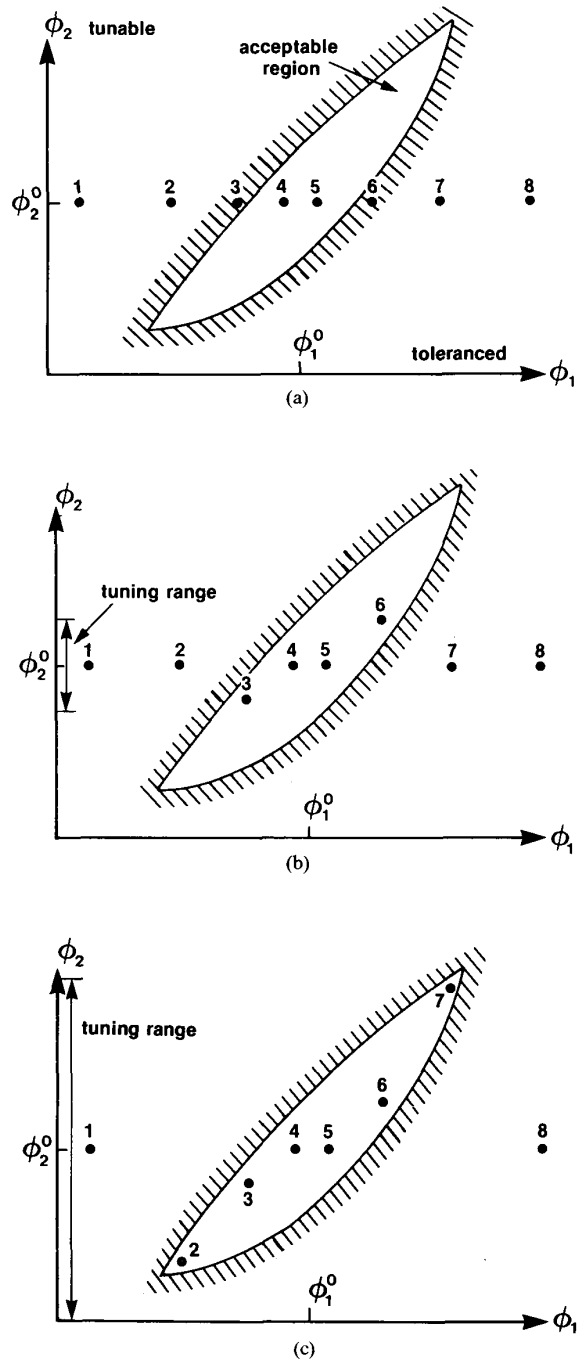


Fig. 7. An illustration of multicircuit design considering eight circuit outcomes. ϕ_1 is tolerated and ϕ_2 is tunable. (a) Without tuning the yield is 2/8 (25 percent). (b) Tuning on ϕ_2 is restricted to a small range. The improved yield is 4/8 (50 percent). (c) A 75 percent yield is achieved by allowing a large tuning range.

From nominal design, centering, optimal tolerancing, to optimal tuning, we have defined a range of problems which lead to increasingly improved yield but, on the other hand, correspond to increasing complexity. Some specific formulations are discussed in Section V. Analogously to ZTP, FTP, and VTP, we can define zero tuning, fixed tuning, and variable tuning problems [20].

C. Multicircuit Modeling

The uncertainties that affect circuit modeling can be discussed under the following categories.

- 1) Measurement errors will inevitably exist in practice, as represented by ΔF^M in (17): $F^M = F^{M,0}(\phi^M) + \Delta F^M$.
- 2) Even without measurement errors, the calculated response $F^{H,0}$ may never be able to match $F^{M,0}$ perfectly, due to, for example, the use of a model of insufficient order or inadequate complexity. Such an inherent mismatch is accounted for in (18) by $F^H = F^{H,0} + \Delta F^H$.
- 3) Even if neither ΔF^M nor ΔF^H exists so that $F^{H,0} = F^M$, we may still not be able to uniquely identify ϕ from the set of measurements that has been selected. This happens when the system of (generally nonlinear) equations $F^{H,0}(\phi) - F^M = 0$, where F^M is the data, is underdetermined. Typically, this problem occurs when, for any reason, many internal nodes are inaccessible to direct measurement. An overcomplicated equivalent circuit, including unknown parasitic elements, is frequently at the heart of this phenomenon.
- 4) The parasitic effects that are not adequately modeled by F^H contribute to the uncertainty ΔF^L . This is another source of interference with the modeling process.

First we consider the case in which modeling is applied to obtain a suitable ϕ such that $F^H(\phi)$ approximates F^M . The nominal circuit approach may be able to cope with the uncertainties in 1) and 2), and comes up with a ϕ which minimizes the errors ΔF^M and ΔF^H in a certain sense. But it will not be able to overcome the problem of uniqueness. In practice, we are often unable to determine unambiguously the identifiability of a system, because all these uncertainties can be present at the same time. There will be, typically, a family of solutions which produce reasonable and similar matches between the measured and the calculated responses. We cannot, therefore, rely on any particular set of parameters.

The approach of multicircuit modeling by Bandler *et al.* [12] can be used to overcome these difficulties. Multiple circuits are created by making deliberate adjustments on the physical parameters ϕ^M . For example, we can change the biasing conditions for an active device and obtain multiple sets of measurements. By doing so, we introduce perturbations to the model which cause some parameters in ϕ to change by an unknown amount. For this approach to be successful, each physical adjustment should produce changes in only a few parameters in ϕ .

Although we do not know the changes in ϕ quantitatively, it is often possible to identify which model parameters may have been affected by the physical adjustments. Such a qualitative knowledge may be apparent from the definition of the model or it may come from practical experience. In the attempt to process multiple circuits

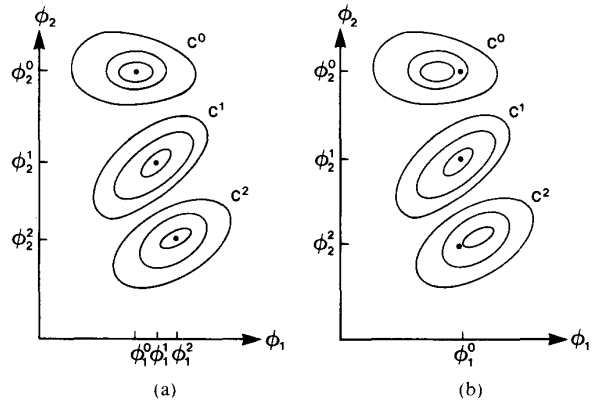


Fig. 8. An illustration of multicircuit modeling. Three circuits are created by making two physical adjustments. Assume that we know that ϕ_1 should not be affected by the physical adjustments. C^0 , C^1 , and C^2 are contours of the error functions corresponding to the three circuits. (a) By treating the three circuits separately, we obtain ϕ_1^0 , ϕ_1^1 , and ϕ_1^2 . ϕ_1^0 , ϕ_1^1 , and ϕ_1^2 turn out to have different values (which is inconsistent with our knowledge) because of uncertainties. (b) Consistent results can be obtained by defining ϕ_1 as a common variable and processing three circuits simultaneously.

simultaneously, we define those model parameters that are not supposed to change as common variables and, at the same time, allow the others to vary between different circuits. By doing so, we force the solution to exhibit the desired consistency and, therefore, improve the reliability of the result. In other words, from a family of possible solutions we select the one that conforms to the topological constraints. Bandler *et al.* have shown an example [12, Section III-A] in which ϕ can not be uniquely identified due to inaccessible nodes. The problem was effectively addressed using the multicircuit approach.

To formulate this mathematically, let

$$\phi^k = \begin{bmatrix} \phi_c^k \\ \phi_a^k \end{bmatrix} \quad (26)$$

where ϕ_c^k contains the common variables and ϕ_a^k contains the variables which are allowed to vary between the k th circuit and the reference circuit ϕ^0 . We then define the optimization variables by

$$x = [(\phi^0)^T (\phi_a^1)^T \cdots (\phi_a^K)^T]^T \quad (27)$$

and state the optimization problem as to

$$\underset{x}{\text{minimize}} U(x) = \|f\|_p \quad (28)$$

where

$$f = [e^T(\phi^0) e^T(\phi^1) \cdots e^T(\phi^K)]^T. \quad (29)$$

Although any l_p norm may be used, the unique property of l_1 discussed in detail by Bandler *et al.* [12] can be exploited to great advantage. The concept of common and independent variables is depicted in Fig. 8.

Now, suppose that we do not have a clear idea about which model parameters may have been affected by the

adjustment on ϕ^M . In this case, we let

$$x = [(\phi^0)^T (\phi^1)^T \cdots (\phi^K)^T]^T \quad (30)$$

and change the objective function to an l_p norm of

$$f = [e^T(\phi^0) \cdots e^T(\phi^K) \alpha_1(\phi^1 - \phi^0)^T \cdots \alpha_K(\phi^K - \phi^0)^T]^T \quad (31)$$

where $\alpha_1, \alpha_2, \dots, \alpha_K$ are nonnegative multipliers (weights).

Using this formulation, while minimizing the errors e , we penalize the objective function for any deviates between ϕ^k and ϕ^0 , since our only available knowledge is that only a few parameters in ϕ^k should have any significant changes. To be effective, an l_1 norm should be used. A similar principle has been successfully applied to the analog circuit fault location problem [9], [27].

A practical application to FET modeling has been described by Bandler *et al.* in [16], where multiple circuits were created by taking three sets of actual measurements under different biasing conditions.

Another important application of multicircuit modeling is to create analytical formulas which link the model ϕ to the actual physical parameters ϕ^M . Such formulas will become extremely useful in guiding an actual production alignment or tuning procedure. A sequence of adjustments on ϕ^M can be systematically made and multiple sets of measurements are taken. By nominal circuit optimization, these measurements would be processed separately to obtain a set of static models. In the presence of uncertainties, a single change in ϕ^M may seem to cause fluctuations in all the model parameters. Obviously, such results are of very little use. In contrast, multicircuit modeling is more likely to produce models that are consistent and reliable. Since the measurements are made systematically, it certainly makes sense to process them simultaneously. Actually, the variables need not be equivalent circuit model parameters. They can include coefficients of a proposed formula as well.

An example of establishing an experimental relationship between the physical and model parameters for a multicavity filter using multiple sets of actual measurements has been described by Daijavad [44].

The multicircuit approach can also be applied to model verification. This is typically related to cases where the parasitic uncertainty ΔF^L has put the validity of a model in doubt. Instead of defining common and independent variables explicitly, we use the formulation of (30) and (31). If consistent results are obtained, then our confidence in the model is strengthened. Otherwise we should probably reject the current model and consider representing the parasitics more adequately. A convincing example has been demonstrated by Bandler *et al.* [12, section V, test 2].

The commercial packages TOUCHSTONE [104], [105] and SUPER-COMPACT [99] allow a hierarchy of circuit blocks and permit the use of variable labels. Multiple circuits and common variables can be easily defined utilizing these features.

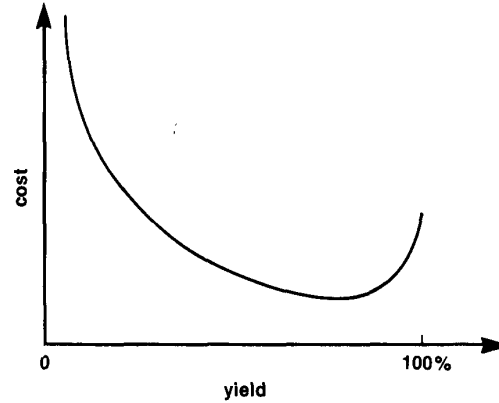


Fig. 9. A typical cost-versus-yield curve [97].

V. TECHNIQUES FOR STATISTICAL DESIGN

In Section IV we have generally discussed uncertainties at different levels, and, in particular, we have expressed our desire to maximize yield in the presence of uncertainties. Optimal tolerancing and tuning have also been identified as means to further reduce cost in the actual production.

We begin this section with a review of some existing techniques for statistical design. Some of the earliest work in this area came from Karafin [68], Pinel and Roberts [87], Butler [36], Elias [52], Bandler, Liu, and Tromp [24]. During the years, significant contributions have been made by, among others, Director and Hachtel [47] (the simplicial method), Soin and Spence [98] (the gravity method), Bandler and Abdel-Malek [1], [2], [7] (multidimensional approximation), Biernacki and Styblinski [30] (dynamic constraint approximation), Polak and Sangiovanni-Vincentelli [90] (a method using outer approximation), as well as Singhal and Pinel [97] (the parametric sampling method). Following the review, we propose a generalized l_p centering algorithm.

A commonly assumed cost versus yield curve [97] is shown in Fig. 9. Actually, hard data are difficult to obtain, and, as we shall see, rather abstract objective functions are often selected for the tolerance-yield design problem. Fig. 10 shows a design with a 100 percent yield and a second design corresponding to the minimum cost.

A. Worst-Case Design

By this approach, we attempt to achieve a 100 percent yield. Since it means that the specifications have to be satisfied for all the possible outcomes, we need to consider only the worst cases.

Bandler *et al.* [23], [24] have formulated it as a nonlinear programming problem

$$\begin{aligned} & \underset{x}{\text{minimize}} C(x) \\ & \text{subject to } e(\phi^k) \leq 0, \quad \text{for all } k \end{aligned} \quad (32)$$

where $C(x)$ is a suitable cost function and the points ϕ^k

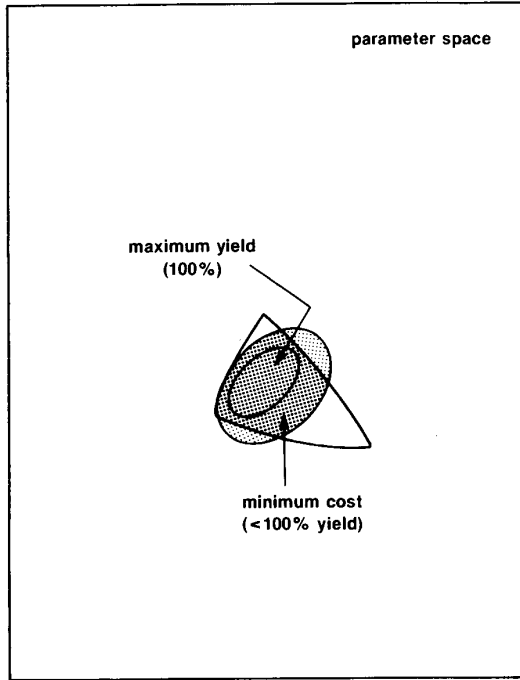


Fig. 10. A maximum yield design and a minimum cost design.

are the worst cases. For instance, we may have

$$C(x) = \sum_{i \in I_e} \frac{a_i}{\epsilon_i} + \sum_{i \in I_t} b_i t_i \quad (33)$$

where I_e and I_t are index sets identifying the tolerated and tunable parameters, respectively. ϵ_i and t_i are the tolerance and the tuning range, respectively, associated with the i th parameter. a_i and b_i are nonnegative weights. A cost function can also be defined for relative tolerances and tuning by including ϕ_i^0 into (33). A critical part of this approach is the determination of the worst cases. Vertices of the tolerance region, for example, are possible candidates for the worst cases by assuming one-dimensional convexity. The yield function does not enter (32) explicitly; instead, a 100 percent yield is implied by a feasible solution.

Bandler and Charalambous [11] have demonstrated a solution to (32) by minimax optimization. Polak and Sangiovanni-Vincentelli [90] have proposed a different but equivalent formulation which involves a nondifferentiable optimization.

A worst-case design is not always appropriate. While attempting to obtain a 100 percent yield, the worst-case approach may necessitate unrealistically tight tolerances, or demand excessive tuning. In either case, the cost may be too high. A perfect 100 percent yield may not even be realizable.

B. Methods of Approximating the Acceptable Region

Since yield is given by the percentage of model outcomes that fall into the acceptable region, we may wish to

find an approximation to that region. The acceptable region has been defined in (16) as $R_a = \{\phi | H(e(\phi)) \leq 0\}$.

Director and Hachtel [47] have devised a simplicial approximation approach. It begins by determining points ϕ^k on the boundary of R_a which is given by $\Omega_a = \{\phi | H(e(\phi)) = 0\}$. The convex hull of these points forms a polyhedron. The largest hypersphere inscribed within the polyhedron gives an approximation to R_a and is found by solving a linear programming problem. Using line searches, more points on the boundary are located and the polyhedron is expanded. The process thus provides a monotonically increasing lower bound on the yield. The center and radius of the hypersphere can be used to determine the centered nominal point and the tolerances, respectively. The application of this method is, however, severely limited by the assumption of a convex acceptable region.

Bandler and Abdel-Malek [1], [2], [7] have presented a method which approximates each $e_j(\phi)$ by a low-order multidimensional polynomial. Model simulations are performed at some ϕ^k selected around a reference point. From the values of $e_j(\phi^k)$ the coefficients of the approximating polynomial are determined by solving a linear system of equations. Appropriate linear cuts are constructed to approximate the boundary Ω_a . The yield is estimated through evaluation of the hypervolumes that lie outside R_a but inside the tolerance region. In critical regions these polynomial approximations are updated during optimization. The one-dimensional convexity assumption for this method is much less restrictive than the multidimensional convexity required by the simplicial approach. Sensitivities for the estimated yield are also available.

Recently, Biernacki and Styblinski [30] have extended the work on multidimensional polynomial approximation by considering a dynamic constraint approximation scheme. It avoids the large number of base points required for a full quadratic interpolation by selecting a maximally flat interpolation. During optimization, whenever a new base point is added, the approximation is updated. It shows improved accuracy compared with a linear model as well as reduced computational effort compared with a full quadratic model.

C. The Gravity Method

Soin and Spence [98] proposed a statistical exploration approach. Based on a Monte Carlo analysis, the centers of gravity of the failed and passed samples are determined as, respectively,

$$\begin{aligned} \phi^f &= \left[\sum_{k \in J} \phi^k \right] / K_{\text{fail}} \\ \phi^p &= \left[\sum_{k \notin J} \phi^k \right] / K_{\text{pass}} \end{aligned} \quad (34)$$

where J is the index set identifying the failed samples. K_{fail} and K_{pass} are the numbers of failed and passed samples, respectively. The nominal point ϕ^0 is then adjusted along the direction $s = \phi^p - \phi^f$ using a line search.

This algorithm is simple but also heuristic. It is not clear as to how the gravity centers are related to the yield in a general multidimensional problem.

D. The Parametric Sampling Method

The parametric sampling approach by Singhal and Pinel [97] has provided another promising direction. A continuous estimate of yield (as opposed to the Monte Carlo estimate, using discrete samples) is given by the following integral:

$$Y(x) = \int_{-\infty}^{+\infty} I_a(\phi) \Gamma(\phi, x) d\phi \quad (35)$$

where $I_a(\phi)$ is the acceptance index defined in (20) and $\Gamma(\phi, x)$ the parameter distribution density function which depends on the design variables x (e.g., the nominal point specifies the mean value and the tolerances control the standard deviations). Normally, in order to estimate the yield, we generate samples ϕ^k , $k=1, 2, \dots, K$, from the component density Γ , perform K circuit analyses, and then take the average of $I_a(\phi^k)$. For each new set of variables x we would have a new density function, and therefore, the sampling and circuit analyses have to be repeated.

The parametric sampling method is based on the concept of importance sampling as

$$Y(x) = \int_{-\infty}^{+\infty} I_a(\phi) \frac{\Gamma(\phi, x)}{h(\phi)} h(\phi) d\phi \quad (36)$$

where $h(\phi)$ is called the sampling density function. The samples ϕ^k are generated from $h(\phi)$ instead of $\Gamma(\phi, x)$. An estimate of the yield is made as

$$\begin{aligned} Y(x) &\approx \frac{1}{K} \sum_{k=1}^K I_a(\phi^k) \frac{\Gamma(\phi^k, x)}{h(\phi^k)} \\ &= \frac{1}{K} \sum_{k=1}^K I_a(\phi^k) W(\phi^k, x). \end{aligned} \quad (37)$$

The weights $W(\phi^k, x)$ compensate for the use of a sampling density different from the component density.

This approach has two clear advantages. First, once the indices $I_a(\phi^k)$ are calculated, no more model simulations are required when x is changed. Furthermore, if Γ is a differentiable density function, then gradients of the estimated yield are readily available. Hence, powerful optimization techniques may be employed. In practice, the algorithm starts with a large number of base points sampled from $h(\phi)$ to construct the initial databank. To maintain a sufficient accuracy, the databank needs to be updated by adding new samples during optimization.

This approach, however, cannot be applied to nondifferentiable density functions such as uniform, discrete, and truncated distributions. It can be extended to include some tunable parameters if the tuning ranges are fixed or practically unlimited. In this case the acceptance index $I_a(\phi^k)$ is defined as 1 if ϕ^k is acceptable after tuning. If ϕ^k is unacceptable before tuning, then whether it can be tuned and, if so, by how much, may have to be determined

through optimization. Variable tuning ranges (in order to minimize cost) cannot be accommodated by the parametric sampling method.

E. Generalized l_p Centering

Here, we propose a generalized l_p centering algorithm which encompasses, in a unified formulation, problems of 100 percent yield (worst-case design) and less than 100 percent yield.

First, we consider the centering problem where we have fixed tolerances and no tuning. Only the nominal point ϕ^0 is to be optimized. Define

$$f = [e^T(\phi^1) \cdots e^T(\phi^K)]^T \quad (38)$$

as the set of multicircuit error functions. We can achieve a worst-case minimax design by

$$\min_x U(x) = H_\infty(f) = \max_k \max_j \{e_j(\phi^k)\} \quad (39)$$

where the multiple circuits ϕ^k are related to ϕ^0 according to (19).

If a 100 percent yield is not attainable, we would naturally look for a solution where the specifications are met by as many points (out of K circuits) as possible. For this purpose minimax is not a proper choice, since unless and until the worst case is dealt with nothing else seems to matter. We may attempt to use a generalized l_2 or l_1 function (i.e., $H_2(f)$ or $H_1(f)$) instead of $H_\infty(f)$ in (39), hoping to reduce the emphasis given to the worst case.

In order to gain more insight into the problem, we define, for each ϕ^k , a scalar function which will indicate directly whether ϕ^k satisfies or violates the specifications and by how much. For this purpose, we choose a set of generalized l_p functions as

$$v_k(x) = H_p(e(\phi^k)), \quad k=1, 2, \dots, K. \quad (40)$$

The sign of v_k indicates the acceptability of ϕ^k while the magnitude of v_k measures, so to speak, the distance between ϕ^k and the boundary of the acceptable region. For example, with $p=\infty$ the distance is measured in the worst-case sense whereas for $p=2$ it will be closer to a Euclidean norm.

We can define a generalized l_p centering as

$$\min_x U(x) = H_p(u(x)) \quad (41)$$

where

$$u(x) = \begin{bmatrix} \alpha_1 v_1 \\ \vdots \\ \alpha_K v_K \end{bmatrix} \begin{bmatrix} \alpha_1 H_q(e(\phi^1)) \\ \vdots \\ \alpha_K H_q(e(\phi^K)) \end{bmatrix} \quad (42)$$

and $\alpha_1, \alpha_2, \dots, \alpha_K$ are a set of positive multipliers. With different p and q it leads to a variety of algorithms for yield enhancement. We discuss separately the case where a nonpositive $U(x)$ exists and the case where we always have $U(x) > 0$.

In the first case, the existence of a $U(x) \leq 0$ indicates that a 100 percent yield is attainable. We should point out

that for a given x the sign of $U(x)$ does not depend on p , q , or α_k . However, the optimal solution x at which $U(x)$ attains its minimum is dependent on p , q , and α . This means that using any values of p , q , and α we will be able to achieve a $U(x) \leq 0$ (i.e., to achieve a 100 percent yield). Furthermore, by using different p , q , and α , we influence the centering of ϕ^0 . Interestingly, the worst-case centering (39) becomes a special case by letting both $p, q = \infty$ and using unit multipliers.

Now consider the case where the optimal yield is less than 100 percent. In this case we propose the use of $p=1$ and $q=1$ in (41). Also, given a starting point x_0 , we define the set of multipliers by

$$\alpha_k = 1/|v_k(x_0)|, \quad k=1, 2, \dots, K. \quad (43)$$

Our proposition is based on the following reasoning (a more complete theoretical justification is reserved for a future paper).

Consider the l_p sum given by

$$\sum_{k \in J} [u_k(x)]^p \quad (44)$$

where $J = \{k | u_k > 0\}$. As $p \rightarrow 0$ (44) approaches the total number of unacceptable circuits which we wish to minimize. The smallest p that gives a convex approximation is 1. This leads to the generalized l_1 objective function given by

$$U(x) = \sum_{k \in J} u_k(x) = \sum_{k \in J} \alpha_k v_k(x). \quad (45)$$

With the multipliers defined by (43), the value of the objective function at the starting point, namely $U(x_0)$, is precisely the count of unacceptable circuits. Also, notice that the magnitude of v_k measures the closeness of ϕ^k to the acceptable region. A small $|v_k|$ indicates that ϕ^k is close to satisfying or violating the specifications. Therefore, we assign a large multiplier to it so that more emphasis will be given to ϕ^k during optimization. On the other hand, we de-emphasize those points that are far away from the boundary of the acceptable region because their contributions to the yield are less likely to change.

One important feature of this approach is its capability of accommodating arbitrary tolerance distributions, since they only influence the generation of ϕ^k . The numerical results we have obtained are very promising. The generalized l_p centering algorithm can also be extended to include variable tolerances and tuning.

VI. EXAMPLES OF STATISTICAL DESIGN

Example 1

The classical two-section 10:1 transmission line transformer, originally proposed by Bandler *et al.* [23] to test minimax optimizers, is a good example for illustrating graphically the basic ideas of centering and tolerancing. An upper specification on the reflection coefficient as $|\rho| \leq 0.55$ and 11 frequencies $\{0.5, 0.6, \dots, 1.5 \text{ GHz}\}$ are considered. The lengths of the transmission lines are fixed at the quarter-wavelength while the characteristic impedances Z_1 and Z_2 are to be tolerated and optimized. Fig.

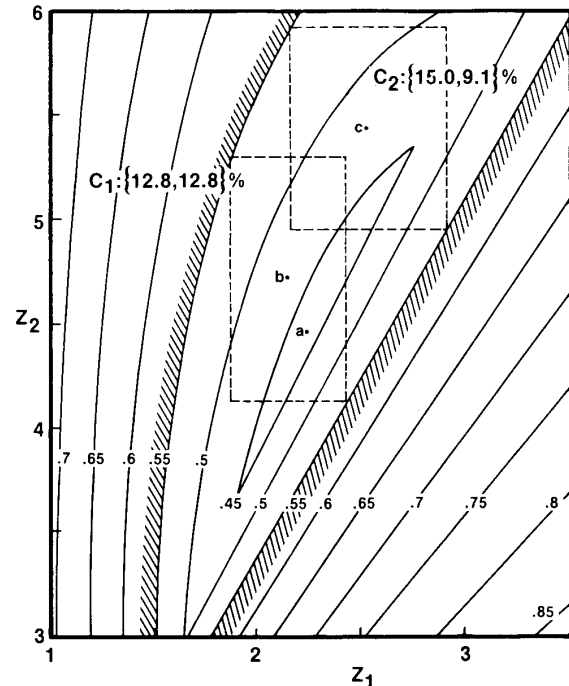


Fig. 11. Contours of $\max |\rho|$ with respect to Z_1 and Z_2 for the two-section transformer indicating the minimax nominal solution a , the centered design with relative tolerances b , and the centered design with absolute tolerances c . The values in brackets are the optimized tolerances (as percentages of the nominal values). The specification is $|\rho| \leq 0.55$.

11 shows the minimax contours, the minimax nominal solution, and the worst-case solutions [23] for

$$P0: \text{minimize } C_1 = Z_1^0/\epsilon_1 + Z_2^0/\epsilon_2 \\ \text{subject to } Y = 100 \text{ percent}$$

$$P1: \text{minimize } C_2 = 1/\epsilon_1 + 1/\epsilon_2 \text{ subject to } Y = 100 \text{ percent}$$

where ϵ_1, ϵ_2 denote tolerances on Z_1 and Z_2 (assuming independent uniform distributions), and Y is the yield. The cost functions C_1 and C_2 correspond to, respectively, relative and absolute tolerancing problems. Two problems of less than 100 percent yield have also been considered by Bandler and Abdel-Malek [7] as

$$P2: \text{minimize } C_2 \text{ subject to } Y \geq 90 \text{ percent}$$

$$P3: \text{minimize } C_2/Y.$$

The optimal tolerance regions and nominal values for $P2$ and $P3$ are shown in Fig. 12. For more details see the original paper [7].

Example 2

The statistical design of a Chebyshev low-pass filter (Singhal and Pinel [97]) is used as the second example. Fifty-one frequencies $\{0.02, 0.04, \dots, 1.0, 1.3 \text{ Hz}\}$ are considered. An upper specification of 0.32 dB on the insertion loss is defined for frequencies from 0.02 to 1.0 Hz. A lower specification of 52 dB on the insertion loss is defined at 1.3 Hz.

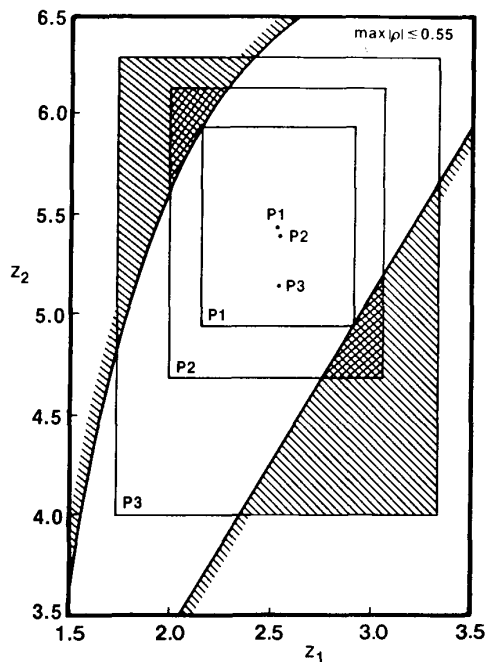


Fig. 12. The optimized tolerance regions and nominal values for the worst case design $P1$, 90 percent yield design $P2$, and minimum cost design $P3$ of the two-section transformer.

Singhal and Pinel [97] have applied the parametric sampling method to the same circuit, assuming normal distributions for the toleranced elements. But, as we have pointed out earlier in this paper, the parametric sampling method cannot be applied to nondifferentiable (such as uniform) distributions. Here, we consider a uniformly distributed 1.5 percent relative tolerance for each component. The generalized l_p centering algorithm described in Section V is used with $p=1$. The nominal solution by standard synthesis as given in [97] was used as starting point, which has a 49 percent yield (w.r.t. the tolerances specified). An 84 percent yield is achieved at the solution which involves a sequence of three design cycles with a total CPU time of 66 seconds on the VAX 8600. Some details are provided in Table I.

VII. GRADIENT-BASED OPTIMIZATION METHODS

So far we have concentrated on translating our practical concerns into mathematical expressions. Now we turn our attention to the solution methods for optimization problems.

The studies in the last two decades on the theoretical and algorithmic aspects of optimization techniques have produced a great number of results. Modern state-of-the-art methods have largely replaced the primitive trial-and-error-approach. In particular, gradient-based optimization methods have gained increasing popularity in recent years for their effectiveness and efficiency.

The majority of gradient-based methods belong to the Gauss-Newton, quasi-Newton, and conjugate gradient families. All these are iterative algorithms which, from a

TABLE I
STATISTICAL DESIGN OF A LOW-PASS FILTER USING
GENERALIZED l_1 CENTERING TECHNIQUE

Component ϕ_i	Nominal Design $\phi_i^{0,0}$	Case 1 $\phi_i^{0,1}$	Case 2 $\phi_i^{0,2}$	Case 3 $\phi_i^{0,3}$
x_1	0.2251	0.21954	0.21705	0.21530
x_2	0.2494	0.25157	0.24677	0.23838
x_3	0.2523	0.25529	0.24784	0.24120
x_4	0.2494	0.24807	0.24019	0.23687
x_5	0.2251	0.22042	0.21753	0.21335
x_6	0.2149	0.22627	0.23565	0.23093
x_7	0.3636	0.36739	0.37212	0.38225
x_8	0.3761	0.36929	0.38012	0.39023
x_9	0.3761	0.37341	0.38371	0.39378
x_{10}	0.3636	0.36732	0.37716	0.38248
x_{11}	0.2149	0.22575	0.22127	0.23129
Yield	49%	77.67%	79.67%	83.67%
Number of samples used for design		50	100	100
Starting point		$\phi^{0,0}$	$\phi^{0,1}$	$\phi^{0,2}$
Number of iterations		16	18	13
CPU time (VAX 8600)		10 sec.	30 sec.	26 sec.

Independent uniform distributions are assumed for each component with fixed tolerances

$\epsilon_i = 1.5\% \phi_i^{0,0}$. The yield is estimated based on 300 samples.

given starting point x_0 , generate a sequence of points $\{x_k\}$. The success of an algorithm depends on whether $\{x_k\}$ will converge to a point x^* and, if so, whether x^* will be a stationary point. An iterative algorithm is described largely by one of its iterations as how to obtain x_{k+1} from x_k .

We use the notation $U(x)$ for the objective function and ∇U for the gradient vector of U . When $U(x)$ is defined by an l_p function, we use f to denote the set of individual error functions so that $U = H(f)$. We also use f'_j for the first-order derivatives of f_j and G for the Jacobian matrix of f .

A. l_p Optimization and Mathematical Programming

Of the l_p family, l_1 , l_2 , and l_∞ are the most distinctive and by far the most useful members. Apart from their unique theoretical properties, it is very important from the algorithmic point of view that linear l_1 , l_2 , and l_∞ problems can be solved exactly using linear or quadratic programming techniques. Besides, all the other members of the l_p family have a continuously differentiable function and, therefore, can be treated similarly to the l_2 case.

An l_1 , l_2 , or l_∞ optimization problem can be converted into a mathematical program. The concepts of local linearization and optimality conditions are often clarified by the equivalent formulation.

For instance, the minimization of $\|f\|_1$ is equivalent to

$$\text{minimize}_{x, y} \sum_{j=1}^m y_j \quad (46)$$

TABLE II
MATHEMATICAL PROGRAMMING EQUIVALENT FORMULATIONS FOR
 l_1 , l_2 , AND l_∞ OPTIMIZATION

The original problem:	minimize $H(\mathbf{f})$ \mathbf{x}	
The equivalent problem:	minimize $V(\mathbf{x}, \mathbf{y})$ subject to the constraints as defined below \mathbf{x}, \mathbf{y}	
	$H(\mathbf{f})$	$V(\mathbf{x}, \mathbf{y})$ constraints (for $j = 1, 2, \dots, m$)
	$\mathbf{f}^T \mathbf{l}_1$	$\sum_{j=1}^m y_j$ $y_j \geq f_j, y_j \geq -f_j$
	$\mathbf{f}^T \mathbf{l}_2$	$\mathbf{y}^T \mathbf{y}$ $y_j = f_j$
	$\mathbf{f}^T \mathbf{l}_\infty$	y $y \geq f_j, y \geq -f_j$
	$H_1^+(\mathbf{f})$	$\sum_{j=1}^m y_j$ $y_j \geq f_j, y_j \geq 0$
	$H_2^+(\mathbf{f})$	$\mathbf{y}^T \mathbf{y}$ $y_j \geq f_j, y_j \geq 0$
	$H_\infty^+(\mathbf{f})$	y $y \geq f_j, y \geq 0$
	$H_\infty(\mathbf{f})$	y $y \geq f_j$

Note: A generalized ℓ_p function $H_p(\mathbf{f})$ is defined through $H_p^+(\mathbf{f})$ and $H_p^-(\mathbf{f})$. H_p^+ is a continuously differentiable function for all $p < \infty$.

subject to

$$y_j \geq f_j(\mathbf{x}), \quad y_j \geq -f_j(\mathbf{x}), \quad j = 1, 2, \dots, m.$$

Other equivalent formulations are summarized in Table II. For the convenience of presentation, we denote these mathematical programming problems by $P(\mathbf{x}, \mathbf{f})$. One important feature of $P(\mathbf{x}, \mathbf{f})$ is that it has a linear or quadratic objective function. If \mathbf{f} is a set of linear functions, then $P(\mathbf{x}, \mathbf{f})$ becomes a linear or quadratic program which can be solved using standard techniques. Equally importantly, linear constraints can be easily incorporated into the problem. Let $P(\mathbf{x}, \mathbf{f}, D)$ be the problem of $P(\mathbf{x}, \mathbf{f})$ subject to a set of linear constraints of the form

$$D: \begin{aligned} \mathbf{a}_l^T \mathbf{x} + b_l &= 0, & l = 1, 2, \dots, L_{eq} \\ \mathbf{a}_l^T \mathbf{x} + b_l &\geq 0, & l = L_{eq} + 1, \dots, L \end{aligned} \quad (47)$$

where \mathbf{a}_l and b_l are constants. If $P(\mathbf{x}, \mathbf{f})$ is a linear or quadratic program, so is $P(\mathbf{x}, \mathbf{f}, D)$. In other words, unconstrained and linearly constrained linear l_1 , l_2 , and l_∞ problems can be solved using standard linear or quadratic programming techniques.

B. Gauss-Newton Methods Using Trust Regions

For a general problem, we may, at each iteration, substitute \mathbf{f} with a linearized model $\tilde{\mathbf{f}}$ so that $P(\mathbf{x}, \tilde{\mathbf{f}})$ can be solved.

For a Gauss-Newton type method, at a given point \mathbf{x}_k , a linearization of \mathbf{f} is made as

$$\tilde{\mathbf{f}}(\mathbf{h}) = \mathbf{f}(\mathbf{x}_k) + \mathbf{G}(\mathbf{x}_k)\mathbf{h} \quad (48)$$

where \mathbf{G} is the Jacobian matrix. We then solve the linear

or quadratic program $P(\mathbf{h}, \tilde{\mathbf{f}}, D)$, where

$$D: \begin{aligned} \Lambda_k &\geq h_j, & j = 1, 2, \dots, n \\ \Lambda_k &\geq -h_j, & j = 1, 2, \dots, n. \end{aligned} \quad (49)$$

These additional constraints define a trust region in which the linearized model $\tilde{\mathbf{f}}$ is believed to be a good approximation to \mathbf{f} .

Another way to look at it is that we have applied a semilinearization (Madsen [78]) to $U(\mathbf{x}) = H(\mathbf{f})$ resulting in

$$\bar{U}(\mathbf{h}) = H(\tilde{\mathbf{f}}(\mathbf{h})). \quad (50)$$

It is important to point out that (50) is quite different from a normal linearization as $U(\mathbf{h}) \approx U(\mathbf{x}_k) + [\nabla U(\mathbf{x}_k)]^T \mathbf{h}$ which corresponds to a steepest descent method. In fact ∇U may not even exist.

Denote the solution of $P(\mathbf{h}, \tilde{\mathbf{f}}, D)$ by \mathbf{h}_k . If $\mathbf{x}_k + \mathbf{h}_k$ reduces the original objective function, we take it as the next iterate; i.e., if $U(\mathbf{x}_k + \mathbf{h}_k) < U(\mathbf{x}_k)$ then $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{h}_k$. Otherwise we let $\mathbf{x}_{k+1} = \mathbf{x}_k$. In the latter case, the trust region is apparently too large and, consequently, should be reduced. At each iteration, the local bound Λ_k in (49) is adjusted according to the goodness of the linearized model.

The above describes the essence of a class of algorithms due to Madsen, who has called it method 1. Madsen [78] has shown that the algorithm provides global convergence in which the proper use of trust regions constitutes a critical part. Such a method has been implemented as an important element in the minimax and l_1 algorithms of Hald and Madsen [65], [66]. In some other earlier work by Osborne and Watson [85], [86] the problem $P(\mathbf{h}, \tilde{\mathbf{f}})$ was solved without incorporating a trust region and the solution \mathbf{h}_k was used as the direction for a line search. For their methods no convergence can be guaranteed and $\{\mathbf{x}_k\}$ may even converge to a nonstationary point.

Normally for the least-squares objective we have to solve a quadratic program at each iteration, which can be a time-consuming process. A remarkable alternative is the Levenberg-Marquardt [76], [81] method. Given \mathbf{x}_k , it solves

$$\min_{\mathbf{h}} \mathbf{h}^T (\mathbf{G}^T \mathbf{G} + \theta_k \mathbf{1}) \mathbf{h} + 2 \mathbf{f}^T \mathbf{G} \mathbf{h} + \mathbf{f}^T \mathbf{f} \quad (51)$$

where $\mathbf{G} = \mathbf{G}(\mathbf{x}_k)$, $\mathbf{f} = \mathbf{f}(\mathbf{x}_k)$, and $\mathbf{1}$ is an identity matrix. The minimizer \mathbf{h}_k is obtained simply by solving the linear system

$$(\mathbf{G}^T \mathbf{G} + \theta_k \mathbf{1}) \mathbf{h}_k = -\mathbf{G}^T \mathbf{f} \quad (52)$$

using, for example, LU factorization. The Levenberg-Marquardt parameter θ_k is very critical for this method. First of all, it is made to guarantee the positive definiteness of (52). Furthermore, it plays, roughly speaking, an inverted role of Λ_k to control the size of a trust region. When $\theta_k \rightarrow \infty$, \mathbf{h}_k gives an infinitesimal steepest descent step. When $\theta_k = 0$, \mathbf{h}_k becomes the solution to $P(\mathbf{h}, \tilde{\mathbf{f}})$ without bounds, which is equivalent to having $\Lambda_k \rightarrow \infty$.

The concept of trust region has been discussed in a broader context by Moré in a recent survey [82].

C. Quasi-Newton Methods

Quasi-Newton methods (also known as variable metric methods) are originated in and steadily upgraded from the work of Davidon [45] and Broyden [33], [34], as well as Fletcher and Powell [55].

For a differentiable $U(x)$, a quasi-Newton step is given by

$$h_k = -\alpha_k B_k^{-1} \nabla U(x_k) \quad (53)$$

where B_k is an approximation to the Hessian of $U(x)$ and the step size controlling parameter α_k is to be determined through a line search. However, on some occasions such as in the l_1 or minimax case, the gradient ∇U may not exist, much less the Hessian.

We can gain more insight to the general case by examining the optimality conditions. Applying the Kuhn–Tucker conditions for nonlinear programming [70] to the equivalent problem $P(x, f)$, we shall find a set of optimality equations

$$R(x) = 0. \quad (54)$$

Since a local optimum x^* must satisfy these equations, we are naturally motivated to solve (54), as a means of finding the minimizer of $U(x)$. A quasi-Newton step for solving nonlinear equations (54) is given by

$$h_k = -\alpha_k J_k^{-1} R(x_k) \quad (55)$$

where J_k is an approximate Jacobian of $R(x)$. Only when $U(x)$ is differentiable will we have the optimality equations as $R(x) = \nabla U(x) = 0$ and (55) reverts to (53).

Hald and Madsen [65], [66] and Bandler *et al.* [21], [22] have described the implementation of a quasi-Newton method for the minimax and l_1 optimization in which the objective functions are not differentiable. Clarke [43] has introduced the concept of generalized gradient, with which optimality conditions can be derived for a broad range of problems.

Quasi-Newton methods, whether in (53) or (55), all require updates of certain approximate Hessians. Many formulas have been proposed over the years. The best known are the Powell symmetric Broyden (PSB) update [91], the Davidon–Fletcher–Powell (DFP) update [45], [55], and the Broyden–Fletcher–Goldfarb–Shanno (BFGS) update [35], [53], [60], [95]. The merits of these formulas and a great many other variations are often compared in terms of their preservation of positive definiteness, convergence to the true Hessian, and numerical performance (see, for instance, Fletcher [54] and Gill and Murray [59]).

Another important point to be considered is the line search. Ideally, α_k is chosen as the minimizer of U in the direction of line search so that $h_k^T \nabla U(x_k + h_k) = 0$. If exact line searches are executed, Dixon [50] has shown that theoretically all members of the Broyden family [34], [53] would have the same performance. In practice, however, exact line search is deemed too expensive and is therefore replaced by other methods. An inexact line search usually limits the evaluation of U and ∇U to only a few points.

Interpolation and extrapolation techniques (such as a quadratic or cubic fit) are then incorporated.

D. Combined Methods

The distinguishing advantage of a quasi-Newton method is that it enjoys a fast rate of convergence near a solution. However, like the Newton method for nonlinear equations, the quasi-Newton method is not always reliable from a bad starting point.

Hald and Madsen [65], [66], [78] have suggested a class of two-stage algorithms. A first-order method of the Gauss–Newton type is employed in stage 1 to provide global convergence to a neighborhood of a solution. When the solution is singular, method 1 suffers from a very slow rate of convergence and a switch is made to a quasi-Newton method (stage 2). Several switches between the two methods may take place and the switching criteria ensure the global convergence of the combined algorithm. Numerical examples of circuit applications have demonstrated a very strong performance of the approach [21], [22], [79], [80].

Powell [92] has extended the Levenberg–Marquardt method and suggested a trust-region strategy which interpolates between a steepest descent step and a Newton step. When far away from the solution, the step is biased toward the steepest descent direction to make sure that it is downhill. Once close to the solution, taking a full Newton step will provide rapid final convergence.

E. Conjugate Gradient Methods

Some extremely large-scale engineering applications involve hundreds of variables and functions. Although the rapid advances in computer technology have enabled us to solve increasingly larger problems, there may be cases in which even the storage of a Hessian matrix and the solution of an n by n linear system become unmanageable.

Conjugate gradient methods [56], [75], [88] provide an alternative for such problems. A distinct advantage of conjugate gradient methods is the minimal requirement of storage. Typically three to six vectors of length n are needed, which is substantially less than the requirement by the Gauss–Newton or quasi-Newton methods. However, proper scaling or preconditioning, near-perfect line searches and appropriate restart criteria are usually necessary to ensure convergence. In general, we have to pay the price for the reduced storage by enduring a longer computation time.

VIII. GRADIENT CALCULATION AND APPROXIMATION

The application of gradient-based l_p optimization methods requires the first-order derivatives of the error functions with respect to the variables.

In circuit optimization, these derivatives are usually obtained from a sensitivity analysis of the network under consideration. For linearized circuits in the frequency domain, it is often possible to calculate the exact sensitivities by the adjoint network approach [5], [31], [48].

However, we ought to recognize that an explicit and elegant sensitivity expression is not always available. For time-domain responses and nonlinear circuits, an exact formula may not exist. Even for linear circuits in the frequency domain, large-scale networks present new problems which need to be addressed.

Often, a large-scale network can be described through compounded and interconnected subnetworks. Many commercial CAD packages such as SUPER-COMPACT [99] and TOUCHSTONE [104], [105] have facilitated such a block structure. In this case, one possible approach would be to assemble the overall nodal matrix and solve the system of equations using sparse techniques (see, e.g., Duff [51], Gustavson [61], Hachtel *et al.* [62]). Another possibility is to rearrange the overall nodal matrix into a bordered block structure which is then solved using the Sherman-Morrison-Woodbury formula [63], [96]. Sometimes it is also possible to develop efficient formulas for a special structure, such as the approach of Bandler *et al.* [17] for branched cascaded networks.

In practice, perhaps the most perplexing and time-consuming part of the task is to devise an index scheme through which pieces of lower level information can be brought into the overall sensitivity expression. It may also require a large amount of memory storage for the various intermediate results. Partly due to these difficulties, methods of exact sensitivity calculations have yet to find their way into general-purpose CAD software packages, although the concept of adjoint network has been in existence for nearly two decades and has had success in many specialized applications.

In cases where either exact sensitivities do not exist or are too difficult to calculate, we can utilize gradient approximations [15], [16], [77], [109]. A recent approach to circuit optimization with integrated gradient approximations has been described by Bandler *et al.* [16]. It has been shown to be very effective and efficient in practical applications including FET modeling and multiplexer optimization.

IX. CONCLUSIONS

In this review, we have formulated realistic circuit design and modeling problems and described their solution methods. Models, variables, and functions at different levels, as well as the associated tolerances and uncertainties, have been identified. The concepts of design centering, tolerancing, and tuning have been discussed. Recent advances in statistical design, yield enhancement, and robust modeling techniques suitable for microwave CAD have been discussed in detail. State-of-the-art optimization techniques have been addressed from both the theoretical and algorithmic points of view.

We have concentrated on aspects that are felt to be immediately relevant to and necessary for modern microwave CAD. There are, of course, other related subjects that have not been treated or not adequately treated in this paper. Notable among these are special techniques for very large systems (Geoffrion [57], [58], Haines [64], Lasdon

[72]), third-generation simulation techniques (Hachtel and Sangiovanni-Vincentelli [63]), fault diagnosis (Bandler and Salama [27]), supercomputer-aided CAD (Rizzoli *et al.* [93]), the simulated annealing and combinatorial optimization methods and their application to integrated circuit layout problems [38], [69], [84], and the new automated decomposition approach to large scale optimization (Bandler and Zhang [28]).

The paper is particularly timely in that software based on techniques which we have described is being integrated by Optimization Systems Associates Inc. into SUPER-COMPACT by arrangement with Compact Software Inc.

ACKNOWLEDGMENT

The authors would like to thank Dr. K. C. Gupta, Guest Editor of this Special Issue on Computer-Aided Design, for his invitation to write this review paper. The useful comments offered by the reviewers are also appreciated. The authors must acknowledge original work done by several researchers which has been integrated into our presentation, including that of Dr. H. L. Abdel-Malek, Dr. R. M. Biernacki, Dr. C. Charalambous, Dr. S. Daijavad, Dr. W. Kellermann, Dr. P. C. Liu, Dr. K. Madsen, Dr. M. R. M. Rizk, Dr. H. Tromp, and Dr. Q. J. Zhang. M. L. Renault is thanked for her contributions, including assistance in preparing data, programs, and results. The opportunity EEsof Inc. provided to develop state-of-the-art optimizers into practical design tools, through interaction with Dr. W. H. Childs, Dr. C. H. Holmes, and Dr. D. Morton, is appreciated. Thanks are extended to Dr. R. A. Pucel of Raytheon Company, Research Division, Lexington, MA, for reviving the first author's interest and work in design centering and yield optimization. Dr. U. L. Rohde of Compact Software Inc., Paterson, NJ, is facilitating the practical implementation of advanced mathematical techniques of CAD. The stimulating environment provided by Dr. Pucel and Dr. Rohde to the first author is greatly appreciated.

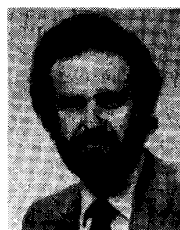
REFERENCES

- [1] H. L. Abdel-Malek and J. W. Bandler, "Yield optimization for arbitrary statistical distributions, Part I: Theory," *IEEE Trans. Circuits Syst.*, vol. CAS-27, pp. 245-253, 1980.
- [2] H. L. Abdel-Malek and J. W. Bandler, "Yield optimization for arbitrary statistical distributions, Part II: Implementation," *IEEE Trans. Circuits Syst.*, vol. CAS-27, pp. 253-262, 1980.
- [3] D. Agnew, "Improved minimax optimization for circuit design," *IEEE Trans. Circuits Syst.*, vol. CAS-28, pp. 791-803, 1981.
- [4] J. W. Bandler, "Optimization methods for computer-aided design," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-17, pp. 533-552, 1969.
- [5] J. W. Bandler, "Computer-aided circuit optimization," in *Modern Filter Theory and Design*, G. C. Temes and S. K. Mitra, Eds. New York: Wiley, 1973, pp. 211-271.
- [6] J. W. Bandler, "Engineering modelling and design subject to model uncertainties and manufacturing tolerances," in *Methodology in Systems Modelling and Simulation*, B. P. Zeigler, *et al.*, Eds. Amsterdam: North-Holland, 1979, pp. 399-421.
- [7] J. W. Bandler and H. L. Abdel-Malek, "Optimal centering, tolerancing, and yield determination via updated approximations and cuts," *IEEE Trans. Circuits Syst.*, vol. CAS-25, pp. 853-871, 1978.
- [8] J. W. Bandler, R. M. Biernacki, and A. E. Salama, "A linear programming approach to fault location in analog circuits," in

- Proc. IEEE Int. Symp. Circuits Syst.*, (Chicago, IL), 1981, pp. 256–260.
- [9] J. W. Bandler, R. M. Biernacki, A. E. Salama, and J. A. Starzyk, "Fault isolation in linear analog circuits using the L_1 norm," in *Proc. IEEE Int. Symp. Circuits Syst.*, (Rome, Italy), 1982, pp. 1140–1143.
 - [10] J. W. Bandler and C. Charalambous, "Theory of generalized least p th approximation," *IEEE Trans. Circuit Theory*, vol. CT-19, pp. 287–289, 1972.
 - [11] J. W. Bandler and C. Charalambous, "Nonlinear programming using minimax techniques," *J. Opt. Theory Appl.*, vol. 13, pp. 607–619, 1974.
 - [12] J. W. Bandler, S. H. Chen, and S. Daijavad, "Microwave device modeling using efficient l_1 optimization: A novel approach," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-34, pp. 1282–1293, 1986.
 - [13] J. W. Bandler, S. H. Chen, S. Daijavad, and W. Kellermann, "Optimal design of multicavity filters and contiguous-band multiplexers," in *Proc. 14th European Microwave Conf.* (Liege, Belgium), 1984, pp. 863–868.
 - [14] J. W. Bandler, S. H. Chen, S. Daijavad, W. Kellermann, M. Renault, and Q. J. Zhang, "Large scale minimax optimization of microwave multiplexers," in *Proc. 16th European Microwave Conf.* (Dublin, Ireland), 1986, pp. 435–440.
 - [15] J. W. Bandler, S. H. Chen, S. Daijavad, and K. Madsen, "Efficient gradient approximations for nonlinear optimization of circuits and systems," in *Proc. IEEE Int. Symp. Circuits Syst.*, (San Jose, CA), 1986, pp. 964–967.
 - [16] J. W. Bandler, S. H. Chen, S. Daijavad, and K. Madsen, "Efficient optimization with integrated gradient approximations," pp. 444–455, this issue.
 - [17] J. W. Bandler, S. Daijavad and Q. J. Zhang, "Computer aided design of branched cascaded networks," in *Proc. IEEE Int. Symp. Circuits Syst.*, (Kyoto, Japan), 1985, pp. 1579–1582.
 - [18] J. W. Bandler, S. Daijavad, and Q. J. Zhang, "Exact simulation and sensitivity analysis of multiplexing networks," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-34, pp. 93–102, 1986.
 - [19] J. W. Bandler, M. A. El-Kady, W. Kellermann, and W. M. Zuberek, "An optimization approach to the best alignment of manufactured and operating systems," in *Proc. IEEE Int. Symp. Circuits Syst.*, (Newport Beach, CA), 1983, pp. 542–545.
 - [20] J. W. Bandler and W. Kellermann, "Selected topics in optimal design centering, tolerancing and tuning," Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada, Rep. SOS-83-28, 1983.
 - [21] J. W. Bandler, W. Kellermann, and K. Madsen, "A superlinearly convergent minimax algorithm for microwave circuit design," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-33, pp. 1519–1530, 1985.
 - [22] J. W. Bandler, W. Kellermann, and K. Madsen, "A nonlinear l_1 optimization algorithm for design, modelling and diagnosis of networks," *IEEE Trans. Circuits Syst.*, vol. CAS-34, pp. 174–181, 1987.
 - [23] J. W. Bandler, P. C. Liu, and J. H. K. Chen, "Worst case network tolerance optimization," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-23, pp. 630–641, 1975.
 - [24] J. W. Bandler, P. C. Liu, and H. Tromp, "A nonlinear programming approach to optimal design centering, tolerancing and tuning," *IEEE Trans. Circuits Syst.*, vol. CAS-23, pp. 155–165, 1976.
 - [25] J. W. Bandler, P. C. Liu, and H. Tromp, "Integrated approach to microwave design," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-24, pp. 584–591, 1976.
 - [26] J. W. Bandler and M. R. M. Rizk, "Optimization of electrical circuits," *Math. Program. Study*, vol. 11, pp. 1–64, 1979.
 - [27] J. W. Bandler and A. E. Salama, "Fault diagnosis of analog circuits," *Proc. IEEE*, vol. 73, pp. 1279–1325, 1985.
 - [28] J. W. Bandler and Q. J. Zhang, "An automatic decomposition technique for device modelling and large circuit design," in *IEEE Int. Microwave Symp. Dig.*, (Las Vegas, NV), 1987, pp. 709–712.
 - [29] R. H. Bartels and A. R. Conn, "An approach to nonlinear l_1 data fitting," Computer Science Department, University of Waterloo, Waterloo, Canada, Rep. CS-81-17, 1981.
 - [30] R. M. Biernacki and M. A. Styblinski, "Statistical circuit design with a dynamic constraint approximation scheme," in *Proc. IEEE Int. Symp. Circuits Syst.*, (San Jose, CA), 1986, pp. 976–979.
 - [31] F. H. Branin, Jr., "Network sensitivity and noise analysis simplified," *IEEE Trans. Circuit Theory*, vol. CT-20, pp. 285–288, 1973.
 - [32] R. K. Brayton, G. D. Hachtel, and A. L. Sangiovanni-Vincentelli, "A survey of optimization techniques for integrated-circuit design," *Proc. IEEE*, vol. 69, pp. 1334–1362, 1981.
 - [33] C. G. Broyden, "A class of methods for solving nonlinear simultaneous equations," *Math. Comp.*, vol. 19, pp. 577–593, 1965.
 - [34] C. G. Broyden, "Quasi-Newton methods and their application to function minimization," *Math. Comp.*, vol. 21, pp. 368–381, 1967.
 - [35] C. G. Broyden, "A new double-rank minimization algorithm," *Notices Amer. Math. Soc.*, vol. 16, p. 670, 1969.
 - [36] E. M. Butler, "Realistic design using large-change sensitivities and performance contours," *IEEE Trans. Circuit Theory*, vol. CT-18, pp. 58–66, 1971.
 - [37] D. A. Calahan, *Computer-Aided Network Design*, rev. ed. New York: McGraw Hill, 1972.
 - [38] A. Casotto, F. Romeo, and A. L. Sangiovanni-Vincentelli, "A parallel simulated annealing algorithm for the placement of macro-cells," in *Proc. IEEE Int. Conf. Computer-Aided Design* (Santa Clara, CA), 1986, pp. 30–33.
 - [39] C. Charalambous, "A unified review of optimization," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-22, pp. 289–300, 1974.
 - [40] C. Charalambous, "Minimax design of recursive digital filters," *Computer Aided Design*, vol. 6, pp. 73–81, 1974.
 - [41] C. Charalambous, "Nonlinear least p th optimization and nonlinear programming," *Math. Program.*, vol. 12, pp. 195–225, 1977.
 - [42] C. Charalambous and A. R. Conn, "Optimization of microwave networks," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-23, pp. 834–838, 1975.
 - [43] F. H. Clarke, "Generalized gradients and applications," *Trans. Amer. Math. Soc.*, vol. 205, pp. 247–262, 1975.
 - [44] S. Daijavad, "Design and modelling of microwave circuits using optimization methods," Ph.D. thesis, McMaster University, Hamilton, Canada, 1986.
 - [45] W. C. Davidon, "Variable metric method for minimization," Rep. ANL-5990 Rev., Argonne National Laboratories, Argonne, IL, 1959.
 - [46] J. E. Dennis, Jr., and J. J. Moré, "Quasi-Newton methods, motivation and theory," *SIAM Rev.*, vol. 19, pp. 46–89, 1977.
 - [47] S. W. Director and G. D. Hachtel, "The simplicial approximation approach to design centering," *IEEE Trans. Circuits Syst.*, vol. CAS-24, pp. 363–372, 1977.
 - [48] S. W. Director and R. A. Rohrer, "Generalized adjoint network and network sensitivities," *IEEE Trans. Circuit Theory*, vol. CT-16, pp. 318–323, 1969.
 - [49] S. W. Director and R. A. Rohrer, "Automated network design: The frequency domain case," *IEEE Trans. Circuit Theory*, vol. CT-16, pp. 330–337, 1969.
 - [50] L. C. W. Dixon, "Quasi-Newton algorithms generate identical points," *Math. Program.*, vol. 2, pp. 383–387, 1972.
 - [51] I. S. Duff, "A survey of sparse matrix research," *Proc. IEEE*, vol. 65, pp. 500–535, 1977.
 - [52] N. J. Elias, "New statistical methods for assigning device tolerances," in *Proc. IEEE Int. Symp. Circuits Syst.*, (Newton, MA), 1975, pp. 329–332.
 - [53] R. Fletcher, "A new approach to variable metric algorithms," *Comput. J.*, vol. 13, pp. 317–322, 1970.
 - [54] R. Fletcher, "A survey of algorithms for unconstrained optimization," in *Numerical Methods for Unconstrained Optimization*, W. Murray, Ed. London: Academic Press, 1972.
 - [55] R. Fletcher and M. J. D. Powell, "A rapidly convergent descent method for minimization," *Comput. J.*, vol. 6, pp. 163–168, 1963.
 - [56] R. Fletcher and C. M. Reeves, "Function minimisation by conjugate gradients," *Comput. J.*, vol. 7, pp. 149–154, 1964.
 - [57] A. M. Geoffrion, "Elements of large-scale mathematical programming—Part I: Concepts," *Management Sci.*, vol. 16, pp. 652–675, 1970.
 - [58] A. M. Geoffrion, "Elements of large-scale mathematical programming—Part II: Synthesis of algorithms and bibliography," *Management Sci.*, vol. 16, pp. 676–691, 1970.
 - [59] P. E. Gill and W. Murray, "Quasi-Newton methods for unconstrained minimization," *J. Inst. Math. Appl.*, vol. 9, pp. 91–108, 1972.
 - [60] D. Goldfarb, "A family of variable-metric methods derived by variational means," *Math. Comp.*, vol. 24, pp. 23–26, 1970.
 - [61] F. G. Gustavson, "Some basic techniques for solving sparse systems of linear equations," in *Sparse Matrices and Their Applications*, D. J. Rose and R. A. Willoughby, Eds. New York: Plenum Press, 1971.
 - [62] G. D. Hachtel, R. K. Brayton, and F. G. Gustavson, "The sparse

- tableau approach to network analysis and design," *IEEE Trans. Circuit Theory*, vol. CT-18, pp. 101-113, 1971.
- [63] G. D. Hachtel and A. L. Sangiovanni-Vincentelli, "A survey of third-generation simulation techniques," *Proc. IEEE*, vol. 69, pp. 1264-1280, 1981.
- [64] Y. Y. Haimes, Ed., *Large Scale Systems*. Amsterdam: North Holland, 1982.
- [65] J. Hald and K. Madsen, "Combined LP and quasi-Newton methods for minimax optimization," *Math. Program.*, vol. 20, pp. 49-62, 1981.
- [66] J. Hald and K. Madsen, "Combined LP and quasi-Newton methods for nonlinear l_1 optimization," *SIAM J. Numer. Anal.*, vol. 22, pp. 68-80, 1985.
- [67] R. Hettich, "A Newton-method for nonlinear Chebyshev approximation," in *Approximation Theory*, R. Schaback and K. Scherer, Eds. (Lecture Notes in Mathematics, 556). Berlin: Springer, 1976, pp. 222-236.
- [68] B. J. Karafin, "The optimum assignment of component tolerances for electrical networks," *Bell Syst. Tech. J.*, vol. 50, pp. 1225-1242, 1971.
- [69] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671-680, 1983.
- [70] H. W. Kuhn and A. W. Tucker, "Non-linear programming," in *Proc. 2nd Symp. Math. Statistics Probability* (Berkeley, CA), 1951, pp. 481-493.
- [71] K. R. Laker, M. S. Ghausi, and J. J. Kelly, "Minimum sensitivity active (leapfrog) and passive ladder bandpass filters," *IEEE Trans. Circuits Syst.*, vol. CAS-22, pp. 670-677, 1975.
- [72] L. S. Lasdon, *Optimization Theory for Large Systems*. New York: Macmillan, 1970.
- [73] L. S. Lasdon, D. F. Suchman, and A. D. Waren, "Nonlinear programming applied to linear array design," *J. Acoust. Soc. Amer.*, vol. 40, pp. 1197-1200, 1966.
- [74] L. S. Lasdon and A. D. Waren, "Optimal design of filters with bounded, lossy elements," *IEEE Trans. Circuit Theory*, vol. CT-13, pp. 175-187, 1966.
- [75] D. Le, "A fast and robust unconstrained optimization method requiring minimum storage," *Math. Program.*, vol. 32, pp. 41-68, 1985.
- [76] K. Levenberg, "A method for the solution of certain problems in least squares," *Quart. Appl. Math.*, vol. 2, pp. 164-168, 1944.
- [77] K. Madsen, "Minimax solution of nonlinear equations without calculating derivatives," *Math. Program. Study*, vol. 3, pp. 110-126, 1975.
- [78] K. Madsen, "Minimization of non-linear approximation functions," Dr. techn. thesis, Institute of Numerical Analysis, Tech. Univ. of Denmark, DK2800 Lyngby, Denmark, 1985.
- [79] K. Madsen and H. Schjaer-Jacobsen, "New algorithms for worst case tolerance optimization," in *Proc. IEEE Int. Symp. Circuits Syst.*, (New York), 1978, pp. 681-685.
- [80] K. Madsen, H. Schjaer-Jacobsen, and J. Voldby, "Automated minimax design of networks," *IEEE Trans. Circuits Syst.*, vol. CAS-22, pp. 791-796, 1975.
- [81] D. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM J. Appl. Math.*, vol. 11, pp. 431-441, 1963.
- [82] J. J. Moré, "Recent developments in algorithms and software for trust region methods," in *Mathematical Programming, The State of the Art*. Bonn: Springer Verlag, 1982, pp. 258-287.
- [83] D. D. Morrison, "Optimization by least squares," *SIAM J. Numer. Anal.*, vol. 5, pp. 83-88, 1968.
- [84] S. Nahar, S. Sahni, and E. Shragowitz, "Simulated annealing and combinatorial optimization," in *Proc. 23rd Design Automat. Conf.*, (Las Vegas, NV), 1986, pp. 293-299.
- [85] M. R. Osborne and G. A. Watson, "An algorithm for minimax optimization in the nonlinear case," *Comput. J.*, vol. 12, pp. 63-68, 1969.
- [86] M. R. Osborne and G. A. Watson, "On an algorithm for discrete nonlinear l_1 approximation," *Comput. J.*, vol. 14, pp. 184-188, 1971.
- [87] J. F. Pinel and K. A. Roberts, "Tolerance assignment in linear networks using nonlinear programming," *IEEE Trans. Circuit Theory*, vol. CT-19, pp. 475-479, 1972.
- [88] E. Polak, *Computational Methods in Optimization: A Unified Approach*. New York: Academic Press, 1971, pp. 53-54.
- [89] E. Polak, "An implementable algorithm for the optimal design centering, tolerancing, and tuning problem," *J. Opt. Theory Appl.*, vol. 37, pp. 45-67, 1982.
- [90] E. Polak and A. L. Sangiovanni-Vincentelli, "Theoretical and computational aspects of the optimal design centering, tolerancing, and tuning problem," *IEEE Trans. Circuits Syst.*, vol. CAS-26, pp. 795-813, 1979.
- [91] M. J. D. Powell, "A new algorithm for unconstrained optimization," in *Nonlinear Programming*, J. B. Rosen, O. L. Mangasarian and K. Ritter, Eds. New York: Academic Press, 1970.
- [92] M. J. D. Powell, "A hybrid method for nonlinear equations," in *Numerical Methods for Nonlinear Algebraic Equations*, P. Rabinowitz, Ed. London: Gordon and Breach, 1970.
- [93] V. Rizzoli, M. Ferlito, and A. Neri, "Vectorized program architectures for supercomputer-aided circuit design," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-34, pp. 135-141, 1986.
- [94] J. Schoeffler, "The synthesis of minimum sensitivity networks," *IEEE Trans. Circuit Theory*, vol. CT-11, pp. 271-276, 1964.
- [95] D. F. Shanno, "Conditioning of quasi-Newton methods for function minimization," *Math. Comp.*, vol. 24, pp. 647-656, 1970.
- [96] J. Sherman and W. J. Morrison, "Adjustment of an inverse matrix corresponding to changes in the elements of a given column or row of the original matrix," *Annu. Math. Statist.*, vol. 20, p. 621, 1949.
- [97] K. Singhal and J. F. Pinel, "Statistical design centering and tolerancing using parametric sampling," *IEEE Trans. Circuits Syst.*, vol. CAS-28, pp. 692-701, 1981.
- [98] R. S. Soine and R. Spence, "Statistical exploration approach to design centering," *Proc. Inst. Elec. Eng.*, vol. 127, pt. G., pp. 260-269, 1980.
- [99] *SUPER-COMPACT User's Manual*, Compact Software Inc., Paterson, NJ 07504, May 1986.
- [100] K. S. Tahim and R. Spence, "A radial exploration approach to manufacturing yield estimation and design centering," *IEEE Trans. Circuits Syst.*, vol. CAS-26, pp. 768-774, 1979.
- [101] T. S. Tang and M. A. Styblinski, "Yield gradient estimation for non-differentiable density functions using convolution techniques and their application to yield optimization," in *Proc. IEEE Int. Symp. Circuits Syst.*, (San Jose, CA), 1986, pp. 1306-1309.
- [102] G. C. Temes and D. A. Calahan, "Computer-aided network optimization the state-of-the-art," *Proc. IEEE*, vol. 55, pp. 1832-1863, 1967.
- [103] G. C. Temes and D. Y. F. Zai, "Least p th approximation," *IEEE Trans. Circuit Theory*, vol. CT-16, pp. 235-237, 1969.
- [104] *TOUCHSTONE User's Manual*, EEsof Inc., Westlake Village, CA 91362, Aug. 1985.
- [105] *TOUCHSTONE Reference Manual*, (Version 1.5), EEsof Inc., Westlake Village, CA 91362, Mar. 1987.
- [106] H. Tromp, "The generalized tolerance problem and worst case search," in *Proc. Conf. Computer-Aided Design of Electronic and Microwave Circuits Syst.*, (Hull, England), 1977, pp. 72-77.
- [107] H. Tromp, "Generalized worst case design, with applications to microwave networks," Doctoral thesis (in Dutch), Faculty of Engineering, University of Ghent, Ghent, Belgium, 1978.
- [108] A. D. Waren, L. S. Lasdon, and D. F. Suchman, "Optimization in engineering design," *Proc. IEEE*, vol. 55, pp. 1885-1897, 1967.
- [109] W. M. Zuberek, "Numerical approximation of gradients for circuit optimization," in *Proc. 27th Midwest Symp. Circuits Syst.*, (Morgantown, WV), 1984, pp. 200-203.

✱



John W. Bandler (S'66-M'66-SM'74-F'78) was born in Jerusalem, Palestine, on November 9, 1941. He studied at Imperial College of Science and Technology, London, England, from 1960 to 1966. He received the B.Sc. (Eng.), Ph.D. and D.Sc. (Eng.) degrees from the University of London, England, in 1963, 1967, and 1976, respectively.

He joined Mullard Research Laboratories, Redhill, Surrey, England, in 1966. From 1967 to 1969 he was a Postdoctorate Fellow and Ses-

sional Lecturer at the University of Manitoba, Winnipeg, Canada. He joined McMaster University, Hamilton, Canada, in 1969, where he is currently a Professor of Electrical and Computer Engineering. He has served as Chairman of the Department of Electrical Engineering and Dean of the Faculty of Engineering. He currently directs research in the Simulation Optimization Systems Research Laboratory. Dr. Bandler is President of Optimization Systems Associates Inc., which he established in 1983. OSA currently provides consulting services and software, specializing in advanced applications of simulation, sensitivity analysis, and mathematical optimization techniques for CAE of microwave integrated circuits.

Dr. Bandler is a contributor to *Modern Filter Theory and Design* (Wiley-Interscience, 1973) and to the forthcoming *Analog Circuits: Computer-aided Analysis and Diagnosis* (Marcel Dekker). He has more than 220 publications, four of which appear in *Computer-Aided Filter Design* (IEEE Press, 1973), one in *Microwave Integrated Circuits* (Artech House, 1975), one in *Low-Noise Microwave Transistors and Amplifiers* (IEEE Press, 1981), one in *Microwave Integrated Circuits* (2nd ed., Artech House, 1985), one in *Statistical Design of Integrated Circuits* (IEEE Press, 1987), and one to be published in *Analog Fault Diagnosis* (IEEE Press). Dr. Bandler was an Associate Editor of the IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES (1969-1974). He was Guest Editor of the Special Issue of the IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES on Computer-Oriented Microwave Practices (March 1974). Dr. Bandler is a Fellow of the Royal Society of Canada and of the Institution of Electrical Engineers (Great Britain). He is a

member of the Association of Professional Engineers of the Province of Ontario (Canada).



Shao Hua Chen (S'84) was born in Swatow, Guangdong, China, on September 27, 1957. He received the B.S. degree from the South China Institute of Technology, Guangzhou, China, in 1982 and the Ph.D. degree in electrical engineering from McMaster University, Hamilton, Canada, in 1987.

From July 1982 to August 1983, he was a teaching assistant in the Department of Automation at the South China Institute of Technology.

He received a graduate scholarship from the Chinese Ministry of Education and worked in the Department of Electrical and Computer Engineering at McMaster University from 1983 to 1987. He held an Ontario Graduate Scholarship for the academic years 1985/86 and 1986/87. Currently he is working as a research engineer for Optimization Systems Associates Inc., Dundas, Ontario, Canada. His research interests include optimization methods, sensitivity analysis, device modeling, design centering, tolerancing and tuning, as well as interactive CAD software.