

Efficient Optimization with Integrated Gradient Approximations

JOHN W. BANDLER, FELLOW, IEEE, SHAO HUA CHEN, STUDENT MEMBER, IEEE,
SHAHROKH DAIJAVAD, MEMBER, IEEE, AND KAJ MADSEN

Abstract—A flexible and effective algorithm is proposed for efficient optimization with integrated gradient approximations. It combines the techniques of perturbations, the Broyden update, and the special iterations of Powell. Perturbations are used to provide an initial approximation as well as regular corrections. The approximate gradient is updated using Broyden's formula in conjunction with the special iterations of Powell. A modification to the Broyden update is introduced to exploit possible sparsity of the Jacobian. Utilizing this algorithm, powerful gradient-based nonlinear optimization tools for circuit CAD can be employed without the effort of calculating exact derivatives. Applications of practical significance are demonstrated. The examples include robust small-signal FET modeling using the l_1 techniques and simultaneous processing of multiple circuits, worst-case design of a microwave amplifier, and minimax optimization of a five-channel manifold multiplexer. Computational efficiency is greatly improved as compared to estimating derivatives entirely by perturbations.

I. INTRODUCTION

MANY POWERFUL gradient-based algorithms have been developed in recent years for nonlinear optimization and applied to circuit CAD problems. For example, Bandler, Kellermann, and Madsen have described algorithms for linearly constrained minimax and l_1 optimization [1], [2]. However, the effort to extend their application to a wide range of practical problems has been frustrated by the requirement of exact gradients of all functions with respect to all variables. For some applications, either an explicit sensitivity expression is not available, e.g., when time-domain analysis and nonlinear circuits are involved, or the actual evaluation of such an expression is very tedious and time-consuming, e.g., for large-scale networks. Partly due to these difficulties, exact sensitivity calculations have not been implemented in many general-purpose CAD software packages, although the concept of adjoint network has been in existence for nearly

Manuscript received February 10, 1987; revised August 6, 1987. This work was supported in part by the Natural Science and Engineering Research Council of Canada under Grants A7239 and G1135, 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., 163 Watson's Lane, 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.

S. Dajavad was with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA. He is now with the IBM T. J. Watson Research Center, Yorktown Heights, NY.

K. Madsen is with the Institute for Numerical Analysis, Technical University of Denmark, Building 303, DK-2800 Lyngby, Denmark.
IEEE Log Number 8717990.

two decades and has had success in many specialized applications. The inability or inconvenience in calculating the exact derivatives has created a gap between the theoretical advances in gradient-based nonlinear optimization techniques and their actual implementation.

With only the function values available, as is the case for many CAD packages on the market, one usually resorts to the method of perturbations (finite differences) for gradients. However, this seemingly simple alternative becomes extremely inefficient when large-scale problems have to be dealt with.

In this paper, we propose a flexible and effective approach to optimization with integrated gradient approximations. It is a hybrid approach which incorporates the use of perturbations, the Broyden update [3], and the special iterations of Powell [4]. The proposed algorithm extends the previous work by Madsen [5] and Zuberek [6] in two aspects. Perturbations are integrated in a flexible manner to allow regular corrections to the approximate gradients. Therefore, a suitable compromise between accuracy and computational labor may be achieved for various applications, especially for large-scale circuit optimization. We also propose a modified Broyden update to take advantage of a possible sparse structure of the problem.

The practical usefulness of the new algorithm is demonstrated through three diverse applications. The subjects are of primary interest to microwave circuit engineers: robust small-signal modeling of FET devices, worst-case fixed tolerance design of a microwave amplifier, and large-scale optimization of manifold multiplexers. Applying an approach to robust device modeling proposed by the authors [7] which employs the l_1 optimization techniques and a novel concept of simultaneous processing of multiple circuits, we have obtained self-consistent models of a FET device using real measurement data. By integrating gradient approximations with a powerful minimax algorithm [1], we are able to optimize a five-channel noncontiguous band multiplexer efficiently and without exact derivatives. The multiplexer problem involves 75 nonlinear variables.

II. GRADIENT APPROXIMATIONS

A. Estimating the Gradient by Perturbations

The first-order derivative of $f_j(\mathbf{x})$ with respect to x_i can be estimated by

$$\frac{\partial f_j(\mathbf{x})}{\partial x_i} \approx \frac{f_j(\mathbf{x} + h\mathbf{u}_i) - f_j(\mathbf{x})}{h} \quad (1)$$

where $\mathbf{x} = [x_1 \ x_2 \ \cdots \ x_n]^T$ is the vector of variables, and \mathbf{u}_i is a column vector which has 1 in the i th position and zeros elsewhere. The accuracy of such an estimate may be improved by using a smaller h as well as by averaging the results of a two-sided approximation (using both positive and negative perturbations). This method is straightforward and reliable. However, the computational labor involved grows in proportion to the dimension of the problem.

In the new algorithm described in this section, perturbations are used to obtain an initial approximation to the gradient at the starting point of an optimization process. During the optimization, we may also incorporate a regular use of perturbations to maintain the accuracy of gradient approximations at a desirable level.

B. The Broyden Update

The Broyden update refers to a rank-one formula proposed by Broyden [3] as

$$\mathbf{G}_{k+1} = \mathbf{G}_k + \frac{\mathbf{f}(\mathbf{x}_k + \mathbf{h}_k) - \mathbf{f}(\mathbf{x}_k) - \mathbf{G}_k \mathbf{h}_k}{\mathbf{h}_k^T \mathbf{h}_k} \mathbf{h}_k^T \quad (2)$$

where \mathbf{G}_k is an approximation of the Jacobian $[\partial \mathbf{f}^T / \partial \mathbf{x}]^T$ at \mathbf{x}_k , \mathbf{h}_k is an increment vector and \mathbf{G}_{k+1} provides an updated Jacobian. The values of the function \mathbf{f} at \mathbf{x}_k and $(\mathbf{x}_k + \mathbf{h}_k)$ are assumed available. If the two points $(\mathbf{x}_k$ and $(\mathbf{x}_k + \mathbf{h}_k))$ are iterates of the optimization process, then the Broyden update requires no additional function evaluations, regardless of the dimension of the problem.

Apparently, the approximate Jacobians generated by the Broyden update are in general less accurate than those obtained from perturbations. Hence, the optimization may require more steps to reach the solution or may not reach the correct solution at all. Broyden [3] has shown that for quadratic functions the Broyden update will converge and will reduce the overall computational effort. Although such properties cannot be proved for a general nonlinear problem, the Broyden update still provides an efficient alternative for approximating derivatives.

The updated approximation \mathbf{G}_{k+1} satisfies the following equation:

$$\mathbf{f}(\mathbf{x}_k + \mathbf{h}_k) - \mathbf{f}(\mathbf{x}_k) = \mathbf{G}_{k+1} \mathbf{h}_k. \quad (3)$$

In other words, \mathbf{G}_{k+1} provides a perfect linear interpolation between the two points \mathbf{x}_k and $(\mathbf{x}_k + \mathbf{h}_k)$.

Some difficulties in the application of the Broyden update have been observed by many researchers (see, for example, [4], [5], and [6]).

1) If some functions are linear in some variables and if the corresponding components of \mathbf{h}_k are nonzero, then the approximations of constant derivatives are updated by nonzero values. Consider a simple example. Let $f_j = x_1^2 + 2x_3$ be a function in \mathbf{f} . Denote the variables by $\mathbf{x} = [x_1 \ x_2 \ x_3]^T$ and the gradient by $\mathbf{f}'_j(\mathbf{x}) = [g_1 \ g_2 \ g_3]^T$. Two components of the gradient, namely $g_2 = 0$ and $g_3 = 2$, are constants and can be found accurately by perturbations; g_1 is the only component that needs to be updated. Suppose that $\mathbf{x}_k = [1 \ 1 \ 1]^T$, $\mathbf{h}_k = [0.5 \ 0.5 \ 0.5]^T$, and a perfect esti-

mation of $\mathbf{f}'_j(\mathbf{x}_k)$ is available as $[2 \ 0 \ 2]^T$. The approximation to $\mathbf{f}'_j(\mathbf{x}_k + \mathbf{h}_k)$, as given by the Broyden update, would be $[2.167 \ 0.167 \ 2.167]^T$ (the true value is $[3 \ 0 \ 2]^T$).

2) Along directions orthogonal to \mathbf{h}_k the Jacobian is not updated:

$$\mathbf{G}_{k+1} \mathbf{p} = \mathbf{G}_k \mathbf{p}, \quad \text{for } \mathbf{p}^T \mathbf{h}_k = 0. \quad (4)$$

To overcome these difficulties, we implement a weighted update and the special iterations of Powell [4].

C. Weighted Broyden Update

The weighted update is to be applied to the Jacobian matrix on a row-by-row basis. The j th row vector of the approximate Jacobian, denoted by $(\mathbf{g}_j)_k$, is an approximation to $\mathbf{f}'_j(\mathbf{x}_k)$, the gradient of f_j . Suppose that the Hessian of f_j is available to us and denoted by \mathbf{H}_j ; then

$$\mathbf{f}'_j(\mathbf{x}_k + \mathbf{h}_k) \approx \mathbf{f}'_j(\mathbf{x}_k) + \mathbf{H}_j(\mathbf{x}_k) \mathbf{h}_k. \quad (5)$$

Analogously to (5), we devise an updating formula to obtain an approximation to $\mathbf{f}'_j(\mathbf{x}_k + \mathbf{h}_k)$ as

$$(\mathbf{g}_j)_{k+1} = (\mathbf{g}_j)_k + \alpha \mathbf{H}_j(\mathbf{x}_k) \mathbf{h}_k. \quad (6)$$

If we choose the coefficient α as

$$\alpha = \frac{\mathbf{f}'_j(\mathbf{x}_k + \mathbf{h}_k) - \mathbf{f}'_j(\mathbf{x}_k) - (\mathbf{g}_j)_k^T \mathbf{h}_k}{\mathbf{h}_k^T \mathbf{H}_j(\mathbf{x}_k) \mathbf{h}_k} \quad (7)$$

then the linear model as given by (3) will be preserved, namely

$$\mathbf{f}'_j(\mathbf{x}_k + \mathbf{h}_k) - \mathbf{f}'_j(\mathbf{x}_k) = (\mathbf{g}_j)_{k+1}^T \mathbf{h}_k. \quad (8)$$

In practice we are very unlikely to have access to the Hessian of any f_j . Even so, two basic facts are obvious: the Hessian of a quadratic function is constant, and if f_j is linear in x_i , then the i th row and the i th column of the Hessian contain only zeros. Hence, we propose the use of a constant diagonal matrix

$$\mathbf{W}_j = \text{diag}[w_{j1} \ \cdots \ w_{jn}], \quad w_{ji} \geq 0, \quad i = 1, \dots, n. \quad (9)$$

This leads to a weighted Broyden update as follows.

$$(\mathbf{g}_j)_{k+1} = (\mathbf{g}_j)_k + \frac{\mathbf{f}'_j(\mathbf{x}_k + \mathbf{h}_k) - \mathbf{f}'_j(\mathbf{x}_k) - (\mathbf{g}_j)_k^T \mathbf{h}_k}{\mathbf{q}_{jk}^T \mathbf{h}_k} \mathbf{q}_{jk}$$

$$\mathbf{q}_{jk} = \mathbf{W}_j \mathbf{h}_k = [w_{j1} h_{k1} \ \cdots \ w_{jn} h_{kn}]^T. \quad (10)$$

The weights w_{ji} provide a measure of the linearity of f_j . If f_j is linear in x_i , we set $w_{ji} = 0$, and if f_j is nearly linear in x_i , we assign a small value to w_{ji} . It should be clear from (10) that only the relative magnitude of the weights is important, not their absolute values.

Consider the simple example we have used in the previous section, namely $f_j = x_1^2 + 2x_3$. Since f_j is independent of x_2 and linear in x_3 , we set $w_{j2} = w_{j3} = 0$ and $w_{j1} = 1$. The approximate gradient given by (10) is $[2.5 \ 0 \ 2]^T$, compared to the result given by the Broyden update as $[2.167 \ 0.167 \ 2.167]^T$, and the true gradient $[3 \ 0 \ 2]^T$.

The assignment of weights requires some knowledge of the functional relationship of $f_j(\mathbf{x})$. Such a knowledge may come from experience or may be gained from sensitivity analyses by performing a few perturbations. For instance, for a particular circuit, it may be known that some designable parameters have little influence on the performance function over some frequency or time intervals. Using an adaptive method to find W_j might be of some theoretical interest. But it was felt to be unnecessary and too complicated to be practical at the present time.

D. Powell's Special Iterations

The Broyden update is a rank-one method. As has been shown in (4), along directions orthogonal to \mathbf{h}_k the approximate Jacobian is not updated. If some consecutive steps of optimization happen to be collinear, the updating procedure may not converge. Powell [4] suggested a method which produces strictly linearly independent directions. For this purpose, special iterations are introduced which intervene between the ordinary iterations of optimization. The increment vector of such a special iteration is not calculated to minimize the error functions; instead it serves the purpose of improving the accuracy of gradient approximations. The algorithm for computing the increment vector for a special iteration, as derived by Powell, is given in the Appendix.

III. A HYBRID APPROXIMATION ALGORITHM

Our hybrid algorithm for gradient approximations consists of an initial approximation, the Broyden update, Powell's special iterations, and regular corrections provided by perturbations.

At the starting point of optimization, the initial approximate Jacobian G_0 is usually computed by perturbations. However, G_0 may be already available; for example, it may have been stored from a previous optimization, and can be utilized to avoid unnecessary computations. This option would be useful if similar problems are being solved repetitively (e.g., the same circuit is optimized with respect to different specifications). The accuracy of G_0 is not very critical to the overall approximation. We have observed for some examples that convergence was achieved despite the erroneous estimates of G_0 .

There is little hard evidence as to how frequently the special iterations should be used. Numerical experience, ours as well as other authors', has suggested the use of a special iteration between every two ordinary ones (i.e., every third iteration is a special iteration). Also, in our implementation, a special iteration is skipped provided that the changes in the functions agree fairly well with the linear prediction by the approximate gradient. This is considered to be true if

$$\|f_j(\mathbf{x}_k + \mathbf{h}_k) - f_j(\mathbf{x}_k) - G_k \mathbf{h}_k\| < 0.1 \|f_j(\mathbf{x}_k + \mathbf{h}_k) - f_j(\mathbf{x}_k)\|. \quad (11)$$

The purpose of this provision is to avoid unnecessary computations.

Whether perturbations should be used during optimization depends on the application. For small or mildly nonlinear problems, the Broyden update may suffice. For large-scale problems, especially in circuit applications where highly nonlinear functions are involved, the correction provided by perturbations is likely to be necessary. We have incorporated in our algorithm the use of perturbations with prescribed regularity, say, at every k th optimization iteration.

The Broyden update with or without weights, depending on whether the necessary knowledge of $f(\mathbf{x})$ is available, is employed between perturbations.

Software for gradient-based optimization typically requires a user-defined routine which accepts a set of values for \mathbf{x} as input and returns the values of $f(\mathbf{x})$ as well as the first-order derivatives. We have implemented an interface which integrates gradient approximations with optimization. Taking a set of values for \mathbf{x} from an optimizer, it calls a user-defined routine for the function values, carries out necessary operations for gradient approximations, and then returns to the optimizer the values of $f(\mathbf{x})$ as well as the approximate Jacobian. The interface is transparent to both the optimizer and the user-defined simulation routine. The optimizer is provided with the required gradients, and the user-defined routine (typically a circuit simulation module) works as if the optimizer did not require gradients.

We have integrated our gradient approximation algorithm with two recent optimization methods [1], [2], for the minimax problems as

$$\text{minimize}_x \max_j \{f_j(\mathbf{x})\} \quad (12)$$

and the l_1 problems as

$$\text{minimize}_x \sum_{j=1}^m |f_j(\mathbf{x})| \quad (13)$$

respectively. The methods described in [1] and [2] are two-stage algorithms. The second stage is to be employed near the solution to accelerate the rate of convergence, for which the accuracy of the approximate gradient may become critical. Hence, our implementation allows a more frequent use of perturbations in the second stage.

The effectiveness and efficiency of the new approach are clearly shown from the results of solving a large variety of problems. The results on some mathematical test problems can be found in [8], [9], and [10].

IV. A TWO-SECTION TRANSMISSION-LINE TRANSFORMER EXAMPLE

Consider the classical two-section 10:1 transmission-line transformer shown in Fig. 1. Originally proposed by Bandler and Macdonald [11], this problem has been widely used to test minimax algorithms. The error functions (f_j) are given by the reflection coefficient sampled at 11 frequencies normalized with respect to 1 GHz: $\{0.5, 0.6, \dots, 1.5\}$. Madsen and Schjaer-Jacobsen [12] have shown that when we take the characteristic impedances Z_1 and Z_2 as variables and keep the lengths l_1 and l_2 con-

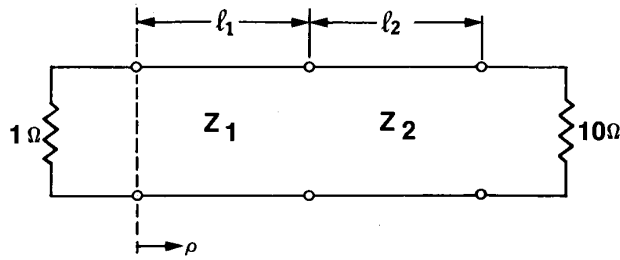


Fig. 1. Two-section, 10:1 transmission-line transformer.

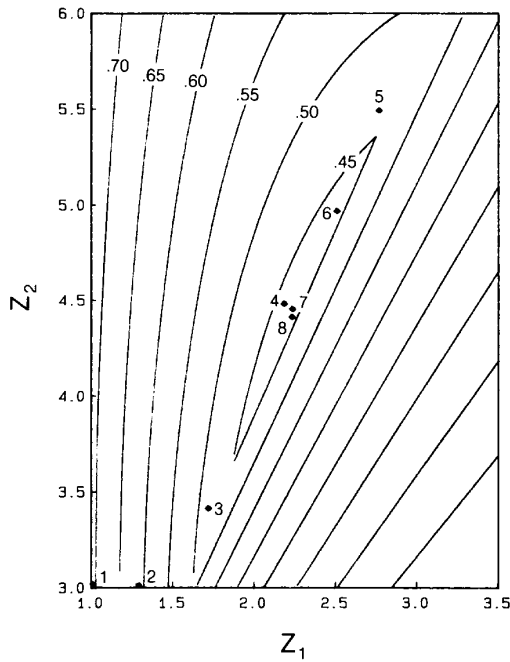


Fig. 2. Minimax contours for the two-dimensional singular minimax problem arising from optimization of the two-section transmission-line transformer. Eight iterations using exact gradients are illustrated.

stant at their optimal values (the quarter wavelength at the center frequency), the minimax problem is singular. Fig. 2 shows the minimax contours and illustrates the solution obtained using exact derivatives. If the derivatives were to be estimated by perturbations, 24 function evaluations would have to be performed. Using our gradient approximation, we obtained the solution, as shown in Fig. 3, after 18 function evaluations.

For the same transformer, we also formulate an l_1 problem. The reflection coefficient at the minimax optimum was taken as a measurement from which we attempt to identify the values of Z_1 and Z_2 . The solutions obtained with the gradients estimated entirely by perturbations and by our new algorithm are illustrated in Figs. 4 and 5, respectively.

A comparison between Figs. 2 to 5 reveals that the solutions obtained using approximate gradients require more iterations of the optimization but overall fewer function evaluations, which is expected.

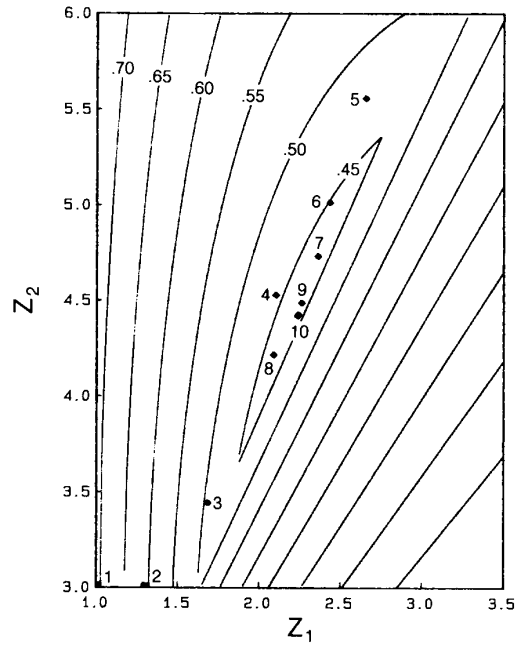


Fig. 3. Minimax optimization of the two-section transmission-line transformer. Ten iterations using approximate gradients are illustrated.

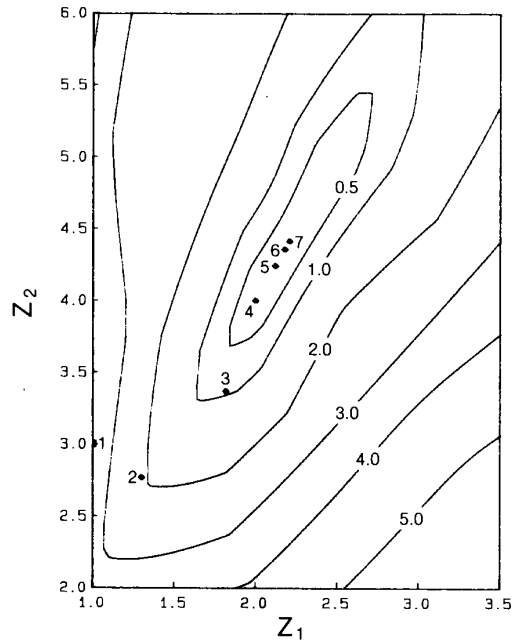


Fig. 4. l_1 contours for the problem arising from parameter identification of the two-section transmission-line transformer. Using perturbations for the gradients, the solution required 14 iterations (42 function evaluations). The first seven iterations are illustrated.

V. FET MODELING USING l_1 OPTIMIZATION WITH APPROXIMATE GRADIENTS

A. Introductory Remarks

The use of l_1 optimization, based on its theoretical properties, has been recommended for nonlinear data fitting and device modeling [1], [7], [13]. Jansen and Koster

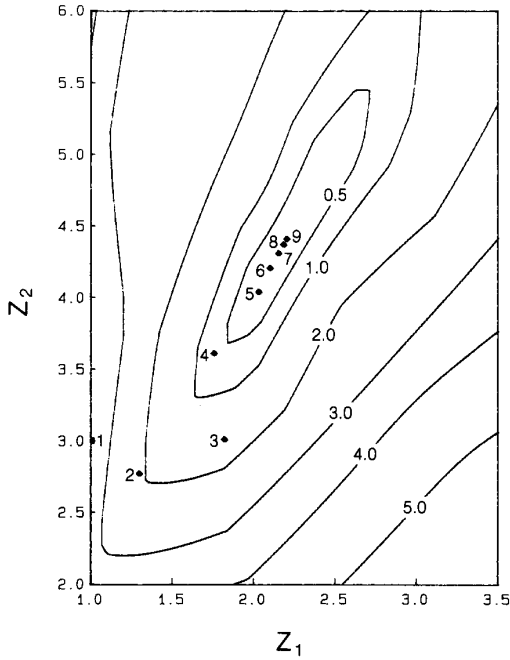


Fig. 5. Parameter identification of the two-section transmission-line transformer using l_1 optimization with approximate gradients. The solution has required 19 iterations and 27 function evaluations. The first nine iterations are illustrated.

[14] have investigated the use of generalized l_p optimization in the modeling of microwave transistors, and they concluded that values of p around unity would lead to relatively stable solutions with good convergence properties. A novel approach to robust modeling of microwave devices has been presented by the authors [7] which exploits the unique properties of the l_1 norm and employs the concept of simultaneous processing of multiple circuits. It has the advantage of establishing not only a good equivalent circuit model but also a reliable measure of the self-consistency of the model. In the context of this paper, an example of FET modeling is given to illustrate the l_1 optimization with integrated gradient approximations.

One of the concerns in practical modeling of FET devices is the uniqueness of the solution. A family of solutions may exist which all exhibit a reasonable match between the calculated and measured responses. The approach described in [7] is intended to improve the chance of unique identification of the model parameters by processing simultaneously multiple circuits. In the case of FET modeling, we create multiple circuits by taking measurements on the scattering parameters under several different biasing conditions. From the physical characteristics of the device, we know that with respect to different biasing conditions some model parameters should remain almost unchanged while the others should vary smoothly. Therefore, from a family of possible solutions we give preference to the one that exhibits the desired consistency. Such a self-consistent model can be achieved automatically by using the l_1 optimization and choosing those model

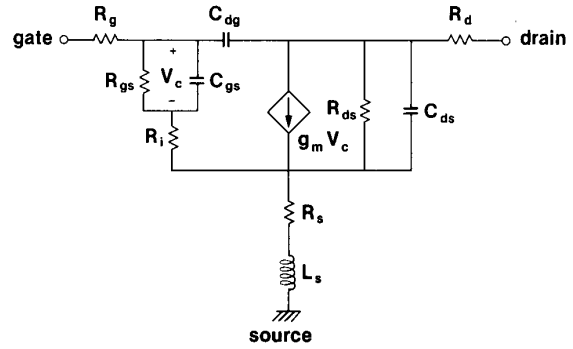


Fig. 6. The small-signal equivalent circuit model for the FET device.

parameters that are insensitive to bias as common variables.

B. The Model and the Measurements

The small-signal equivalent circuit model for the FET is shown in Fig. 6. This model is widely used by commercial programs such as TOUCHSTONE [15] and SUPER-COMPACT [16]. The model has 11 parameters that we will consider as optimization variables:

$$\{R_g, R_d, L_s, \tau, R_{ds}, R_i, R_s, C_{gs}, C_{dg}, C_{ds}, g_m\}.$$

The first four parameters are considered to be bias insensitive.

Three sets of measurements on scattering parameters of a FET device, which were taken at 17 frequency points from 2 GHz to 18 GHz, 1 GHz apart, under the following biasing conditions were made available by Pucel [17]:

- 1) $V_{ds} = 4$ V, $V_{gs} = 0.00$ V, $I_{ds} = 177$ mA
- 2) $V_{ds} = 4$ V, $V_{gs} = -1.74$ V, $I_{ds} = 92$ mA
- 3) $V_{ds} = 4$ V, $V_{gs} = -3.10$ V, $I_{ds} = 37$ mA.

C. Formulation of the Problem

Microwave device modeling utilizing multiple circuits has been formulated in general as an l_1 optimization problem by the authors [7]. The following are formulas (12) to (14) in [2]:

$$\text{minimize}_x \sum_{t=1}^{n_c} \sum_{i=1}^{k_t} |f_i^t| \quad (14)$$

where

$$f_i^t = w_i^t \left[F_i^c(x^t) - (F_i^m)^t \right] \quad (15)$$

and

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}^1 \\ \mathbf{x}^2 \\ \vdots \\ \mathbf{x}^{n_c} \end{bmatrix} \quad (16)$$

with superscript and index t identifying the t th circuit. n_c is the number of circuits and k_t is the number of functions

arising from the t th circuit. \mathbf{x}^t represents the vector of parameters of the t th circuit. Vectors \mathbf{x}_a^t , $t=1, \dots, n_c$, contain only those parameters that vary between different circuits. They do not include the common variables, i.e., those parameters that assume the same values for all circuits. For each circuit, we combine the common variables and \mathbf{x}_a^t to form the vector \mathbf{x}^t .

For the FET modeling problem under consideration, which has three sets of measurements, we specialize the formulas as follows:

$$\text{minimize}_{\mathbf{x}} \sum_{t=1}^3 \sum_{i=1}^{17} \sum_{j=1}^2 \sum_{k=1}^2 \left\{ |\text{Re}[f_{jk}^t(\omega_i)]| + |\text{Im}[f_{jk}^t(\omega_i)]| \right\} \quad (17)$$

where

$$f_{jk}^t(\omega_i) = F_{jk}^t(\mathbf{x}^t, \omega_i) - S_{jk}^t(\omega_i). \quad (18)$$

In (18), F_{jk}^t and S_{jk}^t are the calculated and measured scattering parameters, respectively, with the superscript identifying three different biasing conditions. Having 17 frequency points with real and imaginary parts of the complex S parameters being treated separately, we have a total of 408 error functions. The variables to be optimized in (17) are defined as

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}^1 \\ \mathbf{x}_a^2 \\ \mathbf{x}_a^3 \end{bmatrix}. \quad (19)$$

The vector \mathbf{x}^1 actually has two parts as $\mathbf{x}^1 = [\mathbf{x}^c \ \mathbf{x}_a^1]^T$, where \mathbf{x}^c consists of the common variables as

$$\mathbf{x}^c = [R_g \ R_d \ L_s \ \tau]^T. \quad (20)$$

These are the parameters we expect not to change with respect to different bias. The vector \mathbf{x}_a^t contains the remaining parameters of model t , namely

$$\mathbf{x}_a^t = [R_{ds}^t \ R_i^t \ R_s^t \ C_{gs}^t \ C_{dg}^t \ C_{ds}^t \ g_m^t]^T. \quad (21)$$

The total number of variables is 25.

D. Results

To solve the problem we have formulated, the l_1 optimizer described in [2] was employed. The gradient required was provided by the approach proposed in this paper. We should point out that in this case the evaluation of exact sensitivities is actually possible using the scheme outlined in [7]. However, it involves lengthy and complicated programming. First of all, two adjoint solutions are needed to evaluate the sensitivity expressions for the admittance matrix. From these expressions the sensitivities of the S parameters are derived. Since multiple circuits are processed simultaneously, a complex coding scheme is needed to associate functions arising from different circuits with the appropriate variables. It is then very difficult to modify the software when needed. Comparatively, the calculation of the function values alone requires much simpler effort.

TABLE I
PARAMETER VALUES OF THE FET MODELS

Parameter	Starting Point	Solution		
		Case 1	Case 2	Case 3
R_g (OH)	1.0	2.6025	2.6025	2.6025
R_d (OH)	1.0	3.7630	3.7630	3.7630
R_{ds} (KOH)	0.143	0.1992	0.1638	0.1632
R_i (OH)	1.0	0.0099	0.0999	0.3891
R_s (OH)	1.0	1.0016	0.9220	0.6482
L_s (nH)	0.02	0.0039	0.0039	0.0039
C_{gs} (pF)	1.4	0.7181	0.4417	0.3454
C_{dg} (pF)	0.07	0.0306	0.0475	0.0609
C_{ds} (pF)	0.4	0.2228	0.2229	0.2151
g_m (/OH)	0.09	0.0696	0.0521	0.0410
τ (ps)	7.0	3.9558	3.9558	3.9558

Biasing Conditions

Case 1: $V_{ds}=4\text{V}$ $V_{gs}=0.00\text{V}$ $I_{ds}=177\text{mA}$

Case 2: $V_{ds}=4\text{V}$ $V_{gs}=-1.74\text{V}$ $I_{ds}=92\text{mA}$

Case 3: $V_{ds}=4\text{V}$ $V_{gs}=-3.10\text{V}$ $I_{ds}=37\text{mA}$

The starting points for the three circuits are identical.

This, from the view point of reducing software complexity, justifies the pursuit of gradient approximation.

Three experiments were conducted which have used different schemes for gradient approximation. From the starting point given in Table I, they have reached practically the same solution, which is also given in Table I. The matches between the calculated and measured responses for the first circuit, at both the starting point and the solution, are shown in Figs. 7 and 8. The match for the other two biasing conditions is similar and hence omitted.

The first experiment corresponds to the conventional approach, in which the gradients were estimated solely by perturbations. A total of 468 circuit simulations were required to reach the solution.

In the second case, the Broyden update without weights was used. Regular corrections were also provided by perturbations for every five iterations. Only 128 circuit simulations were needed for this solution.

For the third experiment, we took advantage of an inherent decomposition in the multi-circuit formulation. Notice that the responses (and error functions) of one circuit are absolutely uncorrelated to the independent parameters (\mathbf{x}_a^t) of any other circuits. Obviously, the derivatives corresponding to such decoupled functions and variables are always equal to zero. However, when we use the Broyden update without weights, these derivatives may be changed to some nonzero values, thus introducing apparent errors to the approximation. We can avoid this by using the weighted update. By assigning zero weights to decoupled functions and variables, we can keep the zero

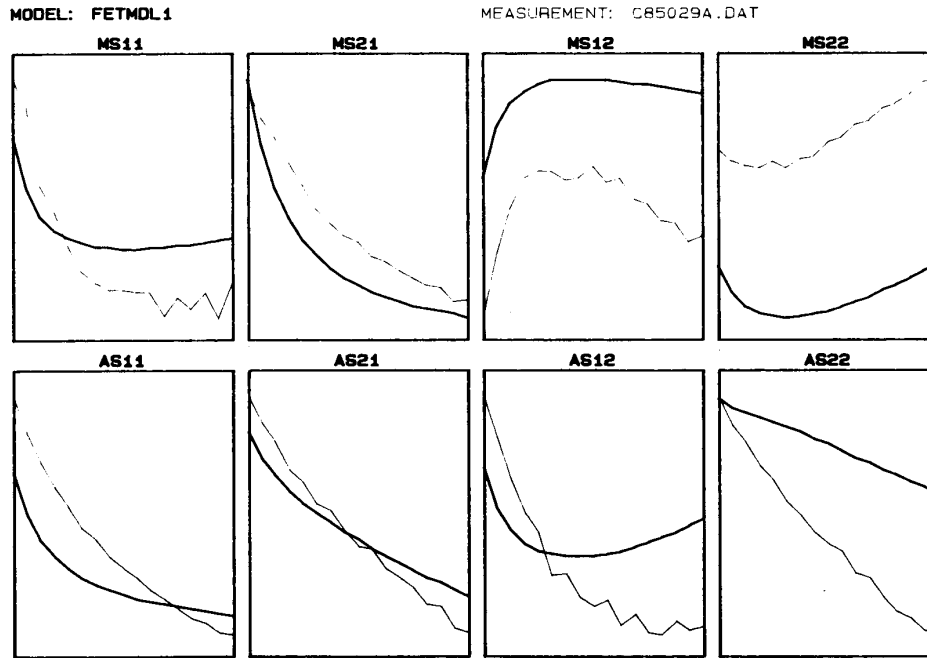


Fig. 7. The scattering parameter match between the FET model and the measurements at the starting point, for $V_{ds} = 4$ V, $V_{gs} = 0$ V, and $I_{ds} = 177$ mA.

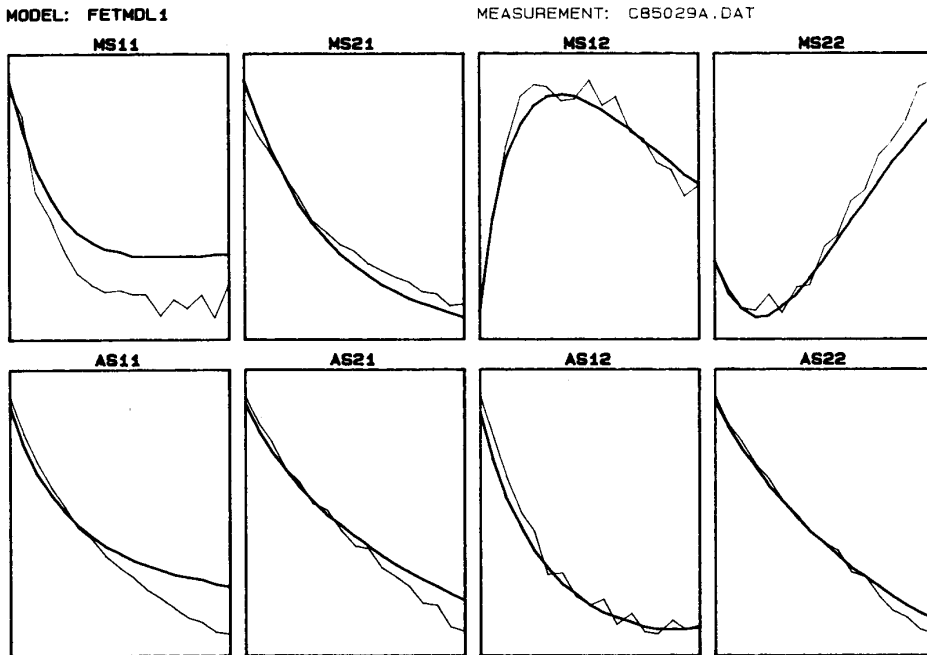


Fig. 8. The scattering parameter match between the FET model and the measurements at the solution, for $V_{ds} = 4$ V, $V_{gs} = 0$ V, and $I_{ds} = 177$ mA.

derivatives undisturbed throughout the optimization process. The application of this concept has reduced the use of perturbations and led to the solution after only 79 circuit simulations. This represents less than 1/5 of the simulations required by the first experiment as well as a 38 percent saving in computational effort as compared to the second experiment.

VI. WORST-CASE DESIGN OF A MICROWAVE AMPLIFIER

Worst-case design using optimization techniques in general has been discussed in [18]. Consider a vector of nominal designable parameters

$$\phi^0 = [\phi_1^0 \cdots \phi_n^0]^T \quad (22)$$

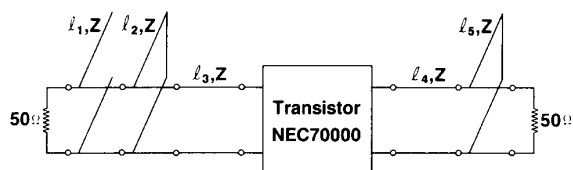


Fig. 9. A microwave amplifier.

a vector of associated tolerances

$$\boldsymbol{\varepsilon} = [\varepsilon_1 \cdots \varepsilon_n]^T \quad (23)$$

and a tolerance region defined by

$$R_\varepsilon = \{ \boldsymbol{\phi} | \boldsymbol{\phi}^0 - \boldsymbol{\varepsilon} \leq \boldsymbol{\phi} \leq \boldsymbol{\phi}^0 + \boldsymbol{\varepsilon} \}. \quad (24)$$

We seek an optimally centered design such that the specifications are satisfied over the tolerance region. It can be formulated as a minimax problem, as

$$\text{minimize}_{\boldsymbol{\phi}^0} \max_j \max_{\boldsymbol{\phi} \in R_\varepsilon} \{ f_j(\boldsymbol{\phi}) \} \quad (25)$$

where f_j , $j=1, \dots, m$, are a set of error functions derived from the design specifications. In practice, we usually consider as candidates for the worst case the vertices of the tolerance region defined by

$$R_v = \{ \boldsymbol{\phi} | \phi_i = \phi_i^0 + \varepsilon_i \mu_i, \mu_i \in \{-1, 1\}, i=1, \dots, n \}. \quad (26)$$

Consider the worst-case fixed tolerance design of a microwave amplifier. As shown in Fig. 9, the amplifier consists of a NEC70000 FET and five transmission lines [15]. The FET is characterized by tabulated scattering parameters provided by the manufacturer. The design variables are the characteristic impedance Z and the lengths l_i of the transmission lines. For each length l_i we assume a 5 percent tolerance. The design specifications are given by

$$7.05 \text{ dB} \leq 20 \log |S_{21}| \leq 8.2 \text{ dB} \quad \text{for } \omega_j = 6, 7, \dots, 18 \text{ GHz}.$$

A total of 26 error functions (f_j) arise from the upper and lower specifications at 13 frequencies.

The worst-case design was accomplished by two phases of optimization. In the first one, we predicted an initial set of worst-case vertices by first-order changes. For each f_j , a vertex $\boldsymbol{\phi}^j$ was selected by

$$\phi_i^j = \phi_i^0 + \mu_i^j \varepsilon_i, \quad \mu_i^j = \text{sign}(\partial f_j / \partial \phi_i), \quad i=1, \dots, n \quad (27)$$

where the derivatives $\partial f_j / \partial \phi_i$ were estimated at the nominal point at the start of the optimization by perturbations. Consequently, 26 vertices (one for each f_j) were considered for the minimax problem

$$\text{minimize}_{\boldsymbol{\phi}^0} \max_j \{ f_j(\boldsymbol{\phi}^j) \}. \quad (28)$$

At the solution, by using (27) with respect to the new nominal point, we found that ten of the worst-case vertices had changed (i.e., the signs of some $\partial f_j / \partial \phi_i$ had changed). The new vertices were added to the worst-case set. The corresponding old vertices were kept, rather than replaced, in order to stabilize the algorithm. We had, therefore, a

TABLE II
PARAMETER VALUES OF THE MICROWAVE AMPLIFIER

Parameter	Starting Point	Solution
l_1	52.96	69.01
l_2	148.13	152.01
l_3	26.80	18.48
l_4	24.01	5.10
l_5	46.63	36.49
Z	81.27	126.39

The starting point is a minimax nominal design

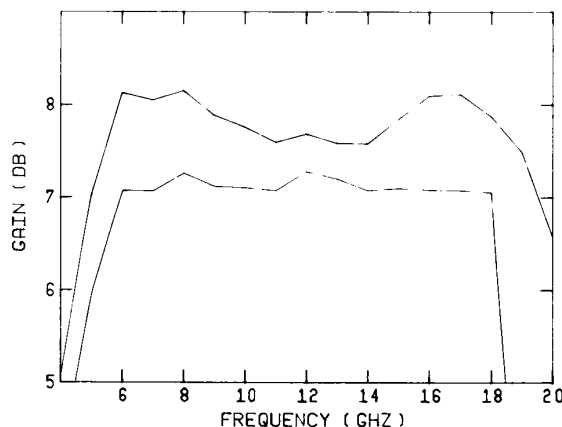


Fig. 10. Worst-case envelope for the amplifier response at the centered solution.

total of 36 worst-case vertices. A second optimization was performed and at its solution the worst-case set was found to be complete (i.e., no more sign change in (27)).

The nominal parameter values at the starting point and the final solution are given in Table II. The total number of function evaluations is 280, in contrast to the 585 required if perturbations were used throughout the optimization. Fig. 10 depicts the worst-case envelope at the solution.

VII. PRACTICAL DESIGN OF A FIVE-CHANNEL MULTIPLEXER

A. Introductory Remarks

A minimax solution of a five-channel 11-GHz noncontiguous band multiplexer was given in detail by Bandler *et al.* [1]. In order to provide the exact sensitivities required, the theory due to Bandler *et al.* [19] was implemented in a computer program which has taken months of effort to develop and test. Furthermore, because the sensitivity expressions depend highly on the circuit structure and vary from component to component, every change to the problem, such as assigning different variables, requires expert modification to the software. In fact, the sensitivi-

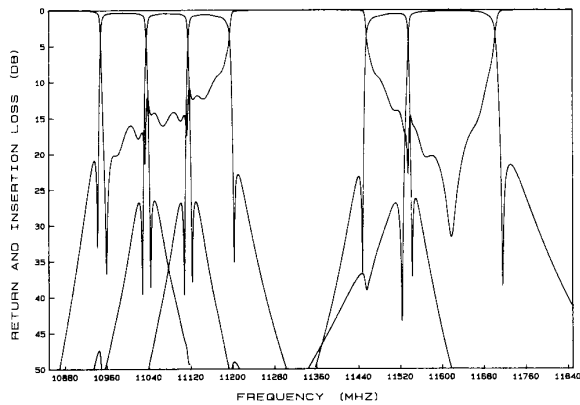


Fig. 11. Responses of the five-channel multiplexer at the starting point.

ties with respect to all possible variables were computed even though some of them have not been actually used; otherwise the coding scheme would have become unmanageable. Large amounts of computer memory were required to store various adjoint solutions and intermediate expressions. By utilizing our gradient approximation, it is possible to efficiently design a multiplexer without all these troubles associated with computing the exact sensitivities. The complexity and size of the program can therefore be considerably reduced.

The five-channel multiplexer provides an excellent illustration of efficient gradient approximations for two reasons. First, it involves 75 variables and, therefore, to rely on perturbations would be prohibitively expensive. To be more specific, suppose that we use the initial parameter values and specifications suggested by Bandler *et al.* [1]. The multiplexer responses at the starting point are shown in Fig. 11. We have reached a result similar to the one reported in [1] after 50 iterations of optimization using exact derivatives. To rely on perturbations for the gradients, we would have to compute multiplexer responses 3800 times (50×76). We will show that efficient gradient approximations reduce the number of response evaluations significantly.

Also, this example is naturally suited for the use of the weighted Broyden update described earlier in this paper. From Fig. 11 it is intuitively obvious that the response functions at lower frequencies should be almost independent of the variables that are related to the filters of channels 1 and 2 (channel 1 has the highest center frequency). Similarly, the responses at higher frequencies are almost independent of the variables related to the filters of channels 3, 4 and 5. We will demonstrate the advantage of using the weighted update.

B. Results

Details of the five-channel multiplexer structure, such as the channel center frequencies, bandwidths, and coupling matrices, can be found in [1]. The channel filters are assumed lossy. Frequency dispersion and nonideal junctions are also taken into account. For all the results that

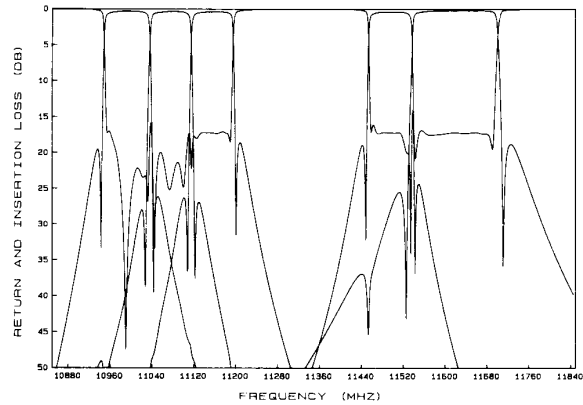


Fig. 12. Responses of the five-channel multiplexer obtained using only the Broyden update and special iterations for gradient approximations. The optimization has stopped prematurely.

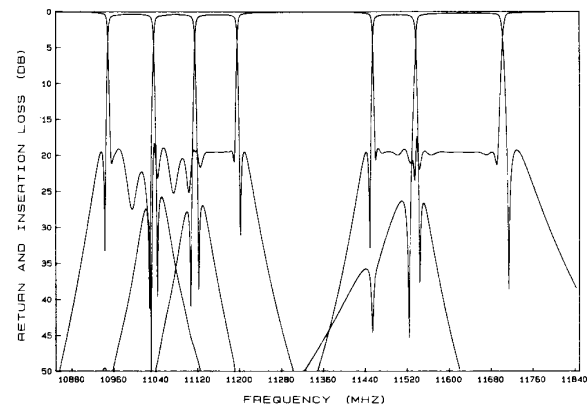


Fig. 13. Responses of the five-channel multiplexer obtained after 500 response evaluations. Regular corrections to the approximate gradient by perturbations were provided for every 20 iterations.

follow, we have used the same specifications and starting point as given in [1]. Three experiments were performed, each using a different method for gradient approximation.

In the first experiment, perturbations were used only at the starting point but not during the optimization. The approximation of gradients relied on the Broyden update in conjunction with the special iterations, which was similar to the methods of Madsen [5] and Zuberek [6]. The optimization stopped after 266 response evaluations, of which 75 were used for the initial perturbations. The responses at this solution as depicted in Fig. 12 are considerably inferior to the result reported in [1]. The optimization has stopped prematurely. This experiment has demonstrated that the Broyden update may not be sufficient for large-scale nonlinear problems.

In a second experiment, regular corrections were provided during the optimization by perturbations for every 20 iterations. After 500 response evaluations, of which 375 were used for perturbations, we obtained the responses shown in Fig. 13. Continuing the process for another 500 response evaluations, the responses shown in Fig. 14 were

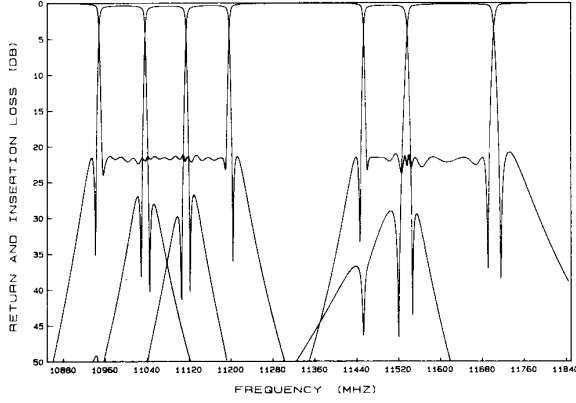


Fig. 14. Responses of the five-channel multiplexer obtained by continuing the process described in Fig. 13 for another 500 response evaluations.

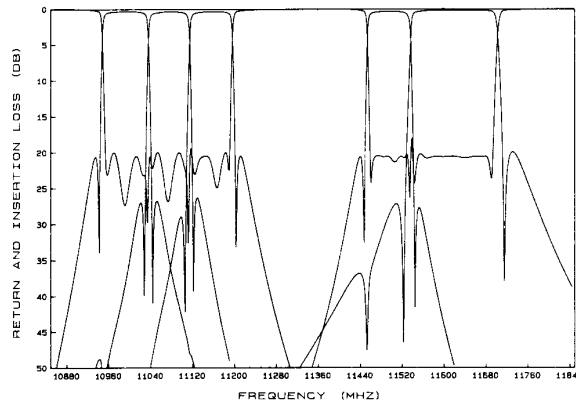


Fig. 15. Responses of the five-channel multiplexer obtained using the weighted update for gradient approximations. The use of appropriate weights has effectively prevented the optimization from stopping prematurely and reduced the use of time-consuming perturbations.

achieved, which are as good as the ones in [1]. From the starting point, a total of 1000 response evaluations were performed. Recall that 3800 response evaluations would be required if the gradient calculations were simply replaced by perturbations.

The third experiment is intended to demonstrate the weighted update proposed in this paper. To apply this updating formula, a weight w_{ji} is set to zero if we know that a function f_j is almost independent of a variable x_i . For instance, the insertion loss of channels 3, 4 and 5 and the common port return loss over the passbands of these channels are almost independent of the filter couplings in channels 1 and 2. Similarly, the responses within the frequencies of channels 1 and 2 are almost independent of the filter couplings in channels 3, 4 and 5. Therefore, we set the corresponding weights to zero.

Utilizing the weighted update, we optimize the multiplexer without any regular correction by perturbations. After 500 response evaluations we obtained the responses shown in Fig. 15. By comparing this result with experiment 1 we can clearly see that the use of appropriate weights has prevented the optimization from stopping prematurely. We

can also conclude from a comparison between experiments 2 and 3 (also, between Figs. 13 and 15) that the application of the weighted update has effectively reduced the use of time-consuming perturbations.

VIII. CONCLUSIONS

A new algorithm for gradient approximations has been presented. Integration of this algorithm with powerful gradient-based optimization techniques has been described and illustrated by the minimax and l_1 implementations. The effectiveness and efficiency of the proposed approach have been demonstrated through diverse examples of practical significance, including FET modeling, worst-case centering, and multiplexer design. A weighted update has also been proposed which exploits possible sparsity and decoupled structures to further reduce the computations involved in estimating gradients. The new approach is very useful when analytical evaluation of partial derivatives is unavailable or tedious. The prospect of integrating our method with existing CAD packages and thus bringing the full power of advanced optimization techniques into practical microwave applications is especially promising.

APPENDIX

FORMULAS FOR POWELL'S SPECIAL ITERATIONS

The formulas for computing the increment vector for a special iteration, as derived by Powell [4], are as follows.

An n by n (n being the dimension of x) orthogonal matrix D_k is constructed at each iteration. Denote the rows of D_k by d_i^T , $i=1,2,\dots,n$. At a special iteration, the increment vector is set to a multiple of the first row vector of D_k , as

$$h_k = \Lambda_k d_1, \quad (\text{A1})$$

where Λ_k is a parameter controlling the step size of h_k . Usually it is set to the step size of the latest ordinary iteration.

At the starting point D_1 is set to an identity matrix. At the k th iteration D_k is revised to produce D_{k+1} . We use y_i^T for the rows of D_{k+1} . For a special iteration, we simply let

$$y_i = d_{i+1}, \quad i=1,2,\dots,n-1 \\ y_n = d_1. \quad (\text{A2})$$

For an ordinary iteration, the following steps take place:

Step 1: Compute $\sigma_i = d_i^T h_k$, $i=1,2,\dots,n$.

Step 2: Find t which is the greatest integer such that $\sigma_t \neq 0$.

Step 3: Let $\alpha_t = 0$ and $z_t = 0$. For $i=t-1, t-2,\dots,1$, compute

$$z_i = z_{i+1} + \sigma_{i+1} d_{i+1} \\ \alpha_i = \alpha_{i+1} + \sigma_{i+1}^2 \\ y_i = (\alpha_i d_i - \sigma_i z_i) / [\alpha_i (\alpha_i + \sigma_i^2)]^{1/2}. \quad (\text{A3})$$

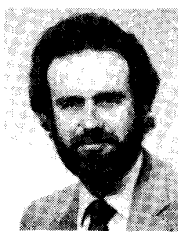
Step 4: Let $y_i = d_{i+1}$, $i=t, t+1,\dots,n-1$. Let $y_n = h_k / (h_k^T h_k)^{1/2}$.

ACKNOWLEDGMENT

Thanks are extended to Dr. R. A. Pucel of Raytheon Company, Research Division, Lexington, MA, for his permission to use the FET data quoted in this paper. The first author must also acknowledge helpful discussions with Dr. Pucel on the subject of device modeling for CAD. The useful comments offered by the reviewers are also appreciated, including bringing reference [14] to the authors' attention.

REFERENCES

- [1] 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.
- [2] 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.
- [3] C. G. Broyden, "A class of methods for solving nonlinear simultaneous equations," *Mathematics of Computation*, vol. 19, pp. 577-593, 1965.
- [4] M. J. D. Powell, "A Fortran subroutine for unconstrained minimization, requiring first derivatives of the objective functions," Atomic Energy Research Establishment, Harwell, Berkshire, England, Rep. AERE-R. 6469, pp. 20-27, 1970.
- [5] K. Madsen, "Minimax solution of nonlinear equations without calculating derivatives," *Mathematical Programming Study*, vol. 3, pp. 110-126, 1975.
- [6] W. M. Zuberek, "Numerical approximation of gradients for circuit optimization," *Proc. 27th Midwest Symp. Circuits Syst.* (Morgantown, WV), 1984, pp. 200-203.
- [7] 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.
- [8] 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.
- [9] J. W. Bandler, S. H. Chen, S. Daijavad, and K. Madsen, "Efficient optimization with integrated gradient approximations, Part I: Algorithms," Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada, Rep. SOS-86-12, 1986.
- [10] J. W. Bandler, S. H. Chen, S. Daijavad, and K. Madsen, "Efficient optimization with integrated gradient approximations, Part II: Implementation," Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada, Rep. SOS-86-13, 1986.
- [11] J. W. Bandler and P. A. Macdonald, "Cascaded noncommensurate transmission-line networks as optimization problems," *IEEE Trans. Circuit Theory*, vol. CT-16, pp. 391-394, 1969.
- [12] K. Madsen and H. Schjaer-Jacobsen, "Singularities in minimax optimization of networks," *IEEE Trans. Circuits Syst.*, vol. CAS-23, pp. 456-460, 1976.
- [13] 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.
- [14] R. H. Jansen and N. Koster, "Generalized least p th optimization in the modeling of microwave transistors," *Arch. Elek. Übertragung.*, vol. 31, pp. 475-477, 1977.
- [15] *TOUCHSTONE User's Manual*, EEsof Inc., Westlake Village, CA 91362, Aug. 1985.
- [16] *SUPER-COMPACT User's Manual*, Compact Software Inc., Paterson, NJ 07504, May 1986.
- [17] R. A. Pucel, Raytheon Company, Research Division, Lexington, MA 02173, FET data, private communications, 1986.
- [18] 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.
- [19] 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.



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 Sessional 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.

✱

Shahrokh Daijavad (S'82-M'87) was born in Isfahan, Iran, on July 7, 1960. He received the B.Eng. (summa cum laude) and Ph.D. degrees in electrical engineering from McMaster University, Hamilton, Canada, in



1983 and 1986, respectively.

From 1983 to 1986, he worked as a teaching assistant in the Department of Electrical and Computer Engineering, McMaster University, while holding an Ontario Graduate Scholarship from 1983 to 1985 and a Natural Sciences and Engineering Research Council of Canada Scholarship from 1985 to 1986. He was awarded an NSERC Postdoctoral Fellowship in 1986 and joined the University of California, Berkeley, working on optimization aspects of layout algorithms.

Currently, he is a research staff member in the VLSI packaging systems group at the IBM T. J. Watson Research Center, Yorktown Heights, NY. His research interests are in optimization methods and their application to VLSI design and microwave device modeling.



Kaj Madsen was born in Denmark in 1943. He received the Cand. Scient. degree in mathematics from the University of Aarhus in 1968.

He has been a Lecturer in numerical analysis and computer science since 1968, apart from his time at the AERE Harwell, Didcot, England. Most of the time he has been with the Institute for Numerical Analysis at the Technical University of Denmark, but during the years 1981–1983 he was at the Computer Science Department at Copenhagen University. In 1986 he obtained the

Dr. Techn. degree at the Technical University of Denmark. His fields of interest in teaching and research are programming languages, optimization, and interval analysis.
