# EFFICIENT OPTIMIZATION WITH INTEGRATED GRADIENT APPROXIMATIONS, PART II: IMPLEMENTATION

J.W. Bandler, S.H. Chen, S. Daijavad and K. Madsen

SOS-86-13-R

December 1986

# EFFICIENT OPTIMIZATION WITH INTEGRATED GRADIENT APPROXIMATIONS

## PART II: IMPLEMENTATION

J.W. Bandler, Fellow, IEEE, S.H. Chen, Student Member, IEEE,

S. Daijavad, Member, IEEE, and K. Madsen

Abstract Various implementations of a flexible and effective approach to efficient circuit optimization with integrated gradient approximations are described. Problems of practical significance are solved utilizing advanced gradient-based optimizers without requiring analytical evaluation of derivatives or expensive numerical differentiation. These problems include robust small signal FET modelling using the techniques of nonlinear $\ell_1$ optimization with simultaneous processing of multiple circuits. Also included are worst-case tolerance design of a microwave amplifier as well as minimax optimization of a 5-channel manifold multiplexer involving 75 nonlinear variables. Drastic reduction of computational labor compared with the traditional approach is achieved in all these cases. Experiments exploiting circuit structure to further improve computational efficiency are also described. The use of approximate gradients also significantly simplifies the effort involved in programming compared with exact derivative approaches.

---

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under Grants A7239 and G1135.

J.W. Bandler and S.H. Chen are with the Simulation Optimization Systems Research Laboratory and the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada L8S 4L7.

J.W. Bandler is also with Optimization Systems Associates Inc., 163 Watson's Lane, Dundas, Ontario, Canada L9H 6L1.

S. Daijavad is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, California 94720, U.S.A.

K.Madsen is with the Institute for Numerical Analysis, Technical University of Denmark, Building 302, DK-2800 Lyngby, Denmark.

## I. INTRODUCTION

A flexible and effective approach to efficient optimization with integrated gradient approximations has been presented in our companion paper [1]. The theoretical aspects of the algorithm have been described in detail, applications have been suggested and its efficiency has been demonstrated by some well-established test problems. In the Part II of our presentation, the impact of the new approach on practical circuit CAD is illustrated by three diverse implementations. The subjects are of primary interest to microwave circuit engineers: robust small signal modelling of FET devices, worst-case tolerance design as illustrated by a microwave amplifier example and efficient optimization of manifold multiplexers.

An approach to robust device modelling has been proposed by the authors [2] which employs powerful $\ell_1$ optimization techniques and a novel concept of simultaneous processing of multiple circuits. In this paper, we apply this approach to FET modelling. Self-consistent models are achieved using real measurement data. Design of multiplexers in general using a superlinearly convergent minimax algorithm has been described by Bandler et al. [3] which well represents very large scale CAD problems in practice. In this paper, a 5-channel non-contiguous band multiplexer is taken as an example.

The optimization techniques employed in these cases are powerful gradient-based methods. Analytical evaluation of the exact derivatives could be very tedious and complicated, if at all possible, and is almost always avoided by microwave CAD software designers. The rather straightforward method of numerical differentiation, i.e., of estimating gradients by perturbations, is widely adopted. However, this seemingly simple solution becomes extremely inefficient when large-scale problems have to be dealt with. The example of FET modelling in this paper involves 408 functions and 25 variables; the one of worst-case design has 36 functions and 6 variables; the one of multiplexer design has 59 functions and 75 variables. If the gradients required by the optimizers are to be estimated entirely by perturbations the computational effort is indeed prohibitively expensive. By implementing the approach to gradient approximation proposed in Part I we are able to reduce drastically the computational labor involved in estimating the gradients. Through these examples, we show that advanced optimization tools can be utilized in a very efficient way to solve problems of practical significance even if analytical gradient evaluation is not available.

In Part I, we introduced a weighted Broyden update to improve the accuracy of gradient approximation. In this paper, we also show, through our examples, how this can be implemented, in order to exploit specific circuit structures, to the advantage of further reducing the computational effort.

## II. FET MODELLING USING $\ell_1$ OPTIMIZATION WITH APPROXIMATE GRADIENTS

### A. Introductory Remarks

The use of $\ell_1$ optimization, based on its theoretical properties, has been recommended for nonlinear data-fitting and device modelling [2],[4],[5]. A novel approach to robust modelling of microwave devices has been presented by the authors [2] which exploits the unique properties of the $\ell_1$ norm and employs the concept of simultaneous processing of multiple circuits. It has the advantage of establishing not only a good circuit model whose responses match as much as possible the measurement, but also a reliable measure of the self-consistency of the model. In the context of this paper, an example of FET modelling is given to illustrate the implementation of $\ell_1$ optimization with integrated gradient approximation.

One of the difficulties frequently encountered in practical modelling of FET devices is the non-uniqueness of the solution. A family of solutions may exist which all exhibit a reasonable match between the model responses and the measurement. As has been described in [2], we can greatly improve the chance of unique identification of the model parameters by processing simultaneously multiple circuits. In the case of FET modelling, 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 prefer the one that exhibits consistency, i.e., the one that conforms to the physical characteristics of the device. Such a self-consistent model can be achieved automatically by using $\ell_1$ optimization and choosing those model parameters that are insensitive to bias as common variables.

B. The Model and the Measurements

The circuit model for the FET that we have taken is widely used by commercial packages such as TOUCHSTONE [6] and SUPER-COMPACT [7]. It is shown in Fig. 1. 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 under the following biasing conditions were made available by R.A. Pucel [8].

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

The measurements were taken at 17 frequency points from 2 GHz to 18 GHz, 1 GHz apart.

C. Formulation of the Problem

Microwave device modelling utilizing multiple circuits has been formulated in general as an $\ell_1$ optimization problem by Bandler et al. [2]. The following are formulas (12)-(14) in [2]:

$$\underset{\mathbf{x}}{\text{minimize}} \sum_{t=1}^{n_c} \sum_{i=1}^{k_t} |f_i^t| , \tag{1}$$

where

$$f_i^t \triangleq w_i^t[F_i^c(\mathbf{x}^t) - (F_i^m)^t] \tag{2}$$

and

$$\mathbf{x} \triangleq \begin{bmatrix} \mathbf{x}^1 \\ \mathbf{x}_a^2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{x}_a^{n_c} \end{bmatrix} \tag{3}$$

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 $\mathbf{x}^t{}_a$, $t = 1, ..., n_c$, 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 $\mathbf{x}^t{}_a$ to form the vector $\mathbf{x}^t$.

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

$$\underset{\mathbf{x}}{\text{minimize}} \sum_{t=1}^{3} \sum_{i=1}^{17} \sum_{j=1}^{2} \sum_{k=1}^{2} \{|\text{Re}[f_{jk}^t(\omega_i)]| + |\text{Im}[f_{jk}^t(\omega_i)]|\}, \tag{4}$$

where

$$f_{jk}^t(\omega_i) \triangleq F_{jk}^t(\mathbf{x}^t, \omega_i) - S_{jk}^t(\omega_i) . \tag{5}$$

In (5), $F^t{}_{jk}$ and $S^t{}_{jk}$ are the scattering parameters of the model and the measurement, respectively, with superscript identifying three different biasing conditions. Having 17 frequency points with the 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 (4) are defined as

$$\mathbf{x} \triangleq \begin{bmatrix} \mathbf{x}^1 \\ \mathbf{x}_a^2 \\ \mathbf{x}_a^3 \end{bmatrix} . \tag{6}$$

The vector $\mathbf{x}^1$ actually has two parts as

$$\mathbf{x}^1 = \begin{bmatrix} \mathbf{x}^c \\ \mathbf{x}_a^1 \end{bmatrix}, \tag{7}$$

where $\mathbf{x}^c$ consists of the common variables as

$$\mathbf{x}^c = [R_g \; R_d \; L_s \; \tau]^T . \tag{8}$$

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

$$\mathbf{x}_a^t = [R_{ds}^t \; R_i^t \; R_s^t \; C_{gs}^t \; C_{dg}^t \; C_{ds}^t \; g_m^t]^T . \tag{9}$$

The total number of variables is 25.

D. Results

We have solved the problem formulated in (1) using a superlinearly convergent $\ell_1$ optimizer described by Bandler, Kellermann and Madsen [4]. The gradient required by the optimizer is provided utilizing the approximation method proposed in this paper.

We should point out that in this case analytical evaluation of the gradient is actually possible using the scheme outlined in [2]. The implementation of the theory, however, requires lengthy and complicated programming. First of all, two adjoint solutions are needed to evaluate the sensitivity expressions of the admittance matrix. From these expressions the sensitivities of the scattering parameters are derived. Since multiple circuits are processed simultaneously, a complex coding scheme is developed to associate functions arising from different circuits with the corresponding variables. The task of identifying the appropriate sensitivity expressions becomes formidable with the concept of common and independent variables. If changes have ever to be made in the circuit topology or variable designation, the amount of labor required to modify the software is such that one would rather re-write the program completely. Comparatively, the evaluation of the function values only requires much simpler effort. This, from the point of view of reducing program complexity, justifies our pursuit of gradient approximation.

Three experiments were conducted which have used different schemes of gradient approximation to solve the problem. From the same starting point they have led to basically identical results. The starting point and the solution are given in Table I. The match between the model responses and the measurements, at both the starting point and the solution, are shown in Figs. 2-7.

The first experiment corresponds to the conventional approach of straightforward numerical differentiation. At each iteration, $n+1$ model simulations are performed ($n$ being the number of variables) and the gradient is approximated by finite differences. It took 468 model simulations to reach the solution.

In the second case, the Broyden update without weights was used in conjunction with special iterations of Powell to obtain the approximate gradient as described in Part I. Regular corrections to the gradient approximation were also provided by perturbations once in every five iterations. This experiment needed only 128 model simulations for its solution.

We observed that the multiple circuit formulation introduces a natural decomposition between functions of one circuit and variables of another. For example, referring to (2) and (3), the variables in $\mathbf{x}_a^1$ belong exclusively to the first model and are therefore uncorrelated to the functions derived from the second and the third models. Obviously, the derivatives corresponding to such decoupled functions and variables are zeros. However, when we use Broyden's formula without weights, these derivatives are updated by nonzero values, thus introducing apparent errors to the approximation. We can avoid this by using the weighted Broyden update as introduced in Part I. To be more specific, we modify Broyden's formula by assigning zero weights for the decoupled functions and variables. Consequently, the corresponding derivatives are kept at zero throughout the optimization process. Applying this concept to our third experiment, we were able to reduce the use of perturbations to once in every 10 iterations and reach the solution after just 79 model simulations. This number is less than 1/5 of that of the first experiment where the more conventional method was employed. It also represents a 38% reduction in the effort of estimating gradients as compared to the second experiment.

## III. WORST-CASE TOLERANCE DESIGN OF A MICROWAVE AMPLIFIER

The general formulation of worst-case tolerance design using optimization techniques has been discussed in [9]. Consider the vector of nominal design parameters

$$\boldsymbol{\phi}^0 = [\phi_1^0 \dots \phi_n^0]^T, \tag{10}$$

the vector of associated tolerances

$$\boldsymbol{\varepsilon} = [\varepsilon_1 \dots \varepsilon_n]^T, \tag{11}$$

and a tolerance region defined by

$$R_\varepsilon = \{\boldsymbol{\phi} \mid \boldsymbol{\phi}^0 - \boldsymbol{\varepsilon} \le \boldsymbol{\phi} \le \boldsymbol{\phi}^0 + \boldsymbol{\varepsilon}\}. \tag{12}$$

We seek an optimally centered point $\boldsymbol{\phi}^{0*}$ such that the design specifications are satisfied over the tolerance region. For minimax design, we have to

$$\underset{\boldsymbol{\phi}^0}{\text{minimize}} \ \max_j \ \max_{\boldsymbol{\phi} \in R_\varepsilon} \{f_j(\boldsymbol{\phi})\}, \tag{13}$$

where $f_j$, $j = 1, \dots, m$, are error functions derived from the design specifications. In practice, we usually consider only the worst-case vertices of the tolerance region. The $2^n$ vertices of the tolerance region are defined by

$$R_v = \{\boldsymbol{\phi} \mid \phi_i = \phi_i^0 + \varepsilon_i \mu_i, \ \ \mu_i \ \ \{-1, 1\}, \ \ i = 1, \dots, n\}. \tag{14}$$

If the functions are sufficiently smooth within the tolerance region, we can use first-order changes to predict the worst-case vertices. For $f_j$, we define the initial set of worst-case vertices

$$R_v^j = \{\boldsymbol{\phi} \mid \phi_i = \phi_i^0 + \varepsilon_i \mu_i, \ \ \mu_i = \text{sign}\left(\frac{df_j}{d\phi_i}\right), \ \ i = 1, \dots, n\}, \tag{15}$$

where $(df_j/d\phi_i)$ for all i,j are first derivatives at the nominal point $\boldsymbol{\phi}^0$. Actually the initial set for each function $f_j$ contains only one point. A minimax optimization is then performed

$$\underset{\boldsymbol{\phi}^0}{\text{minimize}} \ \max_j \ \max_{\boldsymbol{\phi} \in R_v^j} \{f_j(\boldsymbol{\phi})\} \tag{16}$$

At the solution of (16), we use the derivatives at the new nominal point to predict the worst-case vertices. If any of these vertices are not included in the initial sets $R_v^j$, they will be added to those sets and a new optimization as defined by (16) will be performed. The reason for augmenting the worst-case sets instead of replacing them is to stablize the algorithm.

As an example, we consider a microwave amplifier consisting of an NEC70000 FET and five transmission lines [10] as shown in Fig.8. The FET is characterized by tabulated scattering parameters as provided by the manufacturer. The design variables are the characteristic impedance Z and the lengths $\ell_i$ of the transmission lines. Assuming a five percent tolerance for each length $\ell_i$, we seek an optimally centered design to best satisfy the specification given by

$$7.05\text{dB} \leq 20 \log|S_{21}| \leq 8.20\text{dB}, \qquad \text{at } \omega_j = 6,7,\ldots,18 \text{ GHz}.$$

Using (15), the sets $R_v^j$, $j = 1,2,\ldots,26$ (13 frequencies with both upper and lower specifications), were initialized. Working with these sets, the optimization problem (16) was solved. At its solution, 10 new worst-case vertices were detected by first-order projection. The sets $R_v^j$ were augmented by these vertices and the optimization (16) was repeated. At the second solution we found the sets of worst-case vertices were complete. The parameter values at the starting point and the final solution are given in Table II. The total number of function evaluations is 280. We have also solved the same problem with gradients being calculated entirely by perturbations, which required a total of 585 function evaluations. Fig. 9 shows the worst-case envelope at the solution.

## IV. PRACTICAL DESIGN OF A 5-CHANNEL MULTIPLEXER

### A. Introductory Remarks

Design of a five-channel 11 GHz manifold multiplexer by minimax optimization using exact gradients has been described in detail by Bandler et al. [3]. To obtain the solution reported in [3], a program with modules capable of computing multiplexer responses and their sensitivities w.r.t. variables such as filter couplings and waveguide section lengths has been utilized. The theoretical aspects of sensitivity evaluation in multiplexer structures have been discussed by Bandler et al. [11]. However, an actual implementation of the theory 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, sensitivities with respect to all possible variables were computed even though some of them were not actually used, because otherwise the coding scheme would have been unmanageable. Large amounts of memory are required to store various adjoint solutions and intermediate expressions. Utilizing our gradient approximation, it is possible to efficiently design a multiplexer without all these troubles associated with computing the exact gradient. This greatly reduces the size and complexity of the program. Clearly, evaluation of responses alone is more straightforward than simultaneous evaluation of responses and sensitivities.

The five-channel multiplexer problem is an excellent choice for application of the efficient gradient approximation technique for two reasons. First, the problem involves 75 variables and therefore numerical differentiation is indeed prohibitively expensive. To be more specific, consider the following case. Starting from the initial parameter values suggested in [3], at which the multiplexer responses are illustrated in Fig. 10, and using similar specifications as in [3], an optimization w.r.t. all 75 variables resulted in the responses of Fig. 11 after 50 itertions (45 seconds on the FPS 264 via IBM 4381), when exact sensitivities were provided by the simulator. To achieve similar results by numerical differentiation, multiplexer responses should be computed $76 \times 50 = 3800$ times. We will show that the application of the method described in this paper reduces the number of response evaluations required significantly.

A second feature of the 5-channel multiplexer problem is the fact that it is naturally suited for the use of the weighted Broyden update in gradient approximation as described in Part I. Considering Fig. 11, it is intuitively obvious that the common port return loss at lower frequencies should be almost independent of the coupling values of the two filters with the passbands at higher frequencies. Similarly, at higher frequencies the common port return loss should be independent of the coupling parameters of the three filters with passbands at lower frequencies. We will show that the use of weights in updating the Jacobian can improve the performance of the optimization.

B. Results

Before discussing the new results obtained in this paper, we describe the details of the 5-channel multiplexer problem. Some of this information was reported in [3]. The center frequencies and bandwidths are given in Table III. The waveguide manifold width is 0.75 inches. In all experiments, we start the optimization with five identical 6th order filters having the following coupling matrix [3]:

$$M = \begin{bmatrix} 0 & 0.62575 & 0 & 0 & 0 & 0 \\ 0.62575 & 0 & 0.57615 & 0 & 0 & 0 \\ 0 & 0.57615 & 0 & 0.32348 & 0 & -0.74957 \\ 0 & 0 & 0.32348 & 0 & 1.04102 & 0 \\ 0 & 0 & 0 & 1.04102 & 0 & 1.04239 \\ 0 & 0 & -0.74957 & 0 & 1.04239 & 0 \end{bmatrix},$$

where $M_{12}$, $M_{34}$ and $M_{56}$ are screw couplings and $M_{23}$, $M_{45}$ and $M_{36}$ are iris couplings. Filters are cylindrical with a diameter of 1.07 inches. The starting values for input and output filter couplings (transformer ratios) are given as $n_1^2=0.68820$ and $n_2^2=2.04417$. The filters are lossy with an estimated Q factor of 12000.

The initial spacing for the waveguide section associated with each channel is set equal to half the guide wavelength evaluated at the center frequency of the corresponding channel filter. In all experiments, filter losses (dissipation), frequency dispersion and nonideal junctions were taken into account. The models for these nonideal effects are given in [12].

In all the experiments which follow, we use the same frequency points and the same specifications as the ones utilized when the problem was solved with the exact sensitivities. The major difference is, of course, the fact that we use a simulator which only provides responses. A total of 52 frequency points is used for common port return loss with the lower specification of 20 dB. These points are almost uniformly spaced (10 MHz apart) in the pass-band of each channel with additional single frequencies at the crossover of two contiguous channels. A lower specification of 20 dB on the transition band insertion loss of some channels is also imposed. More specifically, the following frequencies are used for the insertion loss: 10935, 11210, 11215, 11440, 11442, 11712 and 11725 MHz. In Table IV, the multiplexer parameters at the optimum solution of Fig. 11 are summarized.

Several experiments were performed on the 5-channel multiplexer using the approximate gradients.

In the first experiment, perturbation was performed at the starting point with no further corrections, i.e., only the Broyden update and special iterations were used as the optimization proceeded. A solution was reached after 266 response evaluations (81 seconds on the FPS 264), of which 75 were used for the initial perturbation. The responses at the solution are illustrated in Fig. 12. Since the solution is not as good as the one achieved using the exact gradients (compare Fig. 12 with Fig. 11), it seems that the regular updating of the gradients using perturbations is necessary in this case.

In a second experiment, we performed perturbations for every 20 iterations of the optimization. After 500 response evaluations (138 seconds), of which $5 \times 75 = 375$ were used for perturbations, the responses shown in Fig. 13 were obtained. Continuing the process until 1000 response evaluations are performed (298 seconds from the starting point), the responses in Fig. 14 were obtained. The optimum parameters corresponding to this figure are summarized in Table V. Comparing Fig. 14 with Fig. 11, it is clear that with the approximate gradients we have achieved as good a result as with the exact gradients. Recall that 3800 response evaluations would be required to achieve the responses of Fig. 11 if the exact gradient calculations were simply replaced by perturbations.

Experiments one and two were both performed with all the weights $w_{ij}$, $i = 1,...,n$, $j = 1,...,m$, being equal to 1. In a third experiment we used zeros for appropriate weights in positions where it is known that a function is almost independent of a variable. For instance, the insertion loss of channels 3, 4 and 5 (channel 1 is the one with the highest center frequency) and the common port return loss over the range of frequencies in the passbands of channels 3 to 5 are almost independent of the filter couplings in channels 1 and 2. Therefore, the appropriate weights are set to zero. Repeating experiment 1, i.e., using perturbations only at the starting point, the responses of Fig. 15 were obtained after 500 response evaluations (166 seconds). Comparing Figs. 15 and 12 indicates that the use of appropriate weights has prevented the optimization from stopping prematurely due to accumulation of error in gradient approximations. The conclusion is that the use of appropriate weights effectively reduces the number of times that the time-consuming correction of gradients by perturbation is to be performed.

## V. CONCLUSIONS

Implementations of minimax and $\ell_1$ optimization with integrated gradient approximations have been presented. Examples of significant practical interest including multiple circuit modelling, fixed tolerance worst-case analysis and large-scale design have been described in detail. A summary of the computational efforts for these examples is given in Table VI. Our algorithm has demonstrated its effectiveness and flexibility in handling a large variety of problems. Compared with the commonly used method of estimating gradients by perturbations, the new approach has significantly improved the computational efficiency. We have also illustrated the weighted update which exploits decomposed structures to further reduce computation effort. The utilization of this algorithm in conjunction with existing circuit simulation packages will make it more effective and practical, taking advantage of the powerful tools of gradient-based optimization in modern computer-aided design.

## ACKNOWLEDGEMENTS

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 modelling for CAD.

## REFERENCES

[1] J.W. Bandler, S.H. Chen, S. Daijavad and K. Madsen, "Efficient optimization with integrated gradient approximations, Part I: Algorithms", submitted as a companion paper.

[2] J.W. Bandler, S.H. Chen and S. Daijavad, "Microwave device modelling using efficient $\ell_1$ optimization: a novel approach", IEEE Trans. Microwave Theory Tech., vol. MTT-34, 1986, pp. 1282-1293.

[3] 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, 1985, pp. 1519-1530.

[4] J.W. Bandler, W. Kellermann and K. Madsen, "An nonlinear $\ell_1$ optimization algorithm for design, modelling and diagnosis of networks", IEEE Trans. Circuits and Systems, vol.CAS-34, 1987.

[5] R.H. Bartels and A.R. Conn, "An approach to nonlinear $\ell_1$ data fitting", Computer Science Department, University of Waterloo, Waterloo, Canada, Report CS-81-17, 1981.

[6] TOUCHSTONE User's Manual, EEsof Inc., Westlake Village, CA 91362, Aug. 1985.

[7] SUPER-COMPACT User's Manual, Communications Consulting Corp., Patterson, NJ 07504, May 1986.

[8] R.A. Pucel, Raytheon Company, Research Division, Lexington, MA 02173, FET data, private communications, 1986.

[9] J.W. Bandler, P.C. Liu and H. Tromp, "A nonlinear programming approach to optimal design centering, tolerancing and tuning", IEEE Trans. Circuits and Systems, vol. CAS-23, 1976, pp.155-165.

[10] TOUCHSTONE User's Manual, Version 1.32, EEsof Inc., Westlake Village, CA 91362, Aug. 1985, Appendix, p.A-1.

[11] 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, 1986, pp. 93-102.

[12] S. Daijavad, "Design and modelling of microwave circuits using optimization methods", Ph.D. Thesis, McMaster University, Hamilton, Canada.

TABLE I

PARAMETER VALUES OF THE FET MODELS

| Case 1 | | Case 2 | | Case 3 | | Parameter |
| --- | --- | --- | --- | --- | --- | --- |
| starting | solution | starting | solution | starting | solution | |
| 1.0 | 2.6025 | 1.0 | 2.6025 | 1.0 | 2.6025 | $R_g$ (OH) |
| 1.0 | 3.7630 | 1.0 | 3.7630 | 1.0 | 3.7630 | $R_d$ (OH) |
| 143.0 | 199.1591 | 143.0 | 163.8249 | 143.0 | 163.1911 | $R_{ds}$ (OH) |
| 1.0 | 0.0099 | 1.0 | 0.0999 | 1.0 | 0.3891 | $R_i$ (OH) |
| 1.0 | 1.0016 | 1.0 | 0.9220 | 1.0 | 0.6482 | $R_s$ (OH) |
| 0.02 | 0.0039 | 0.02 | 0.0039 | 0.02 | 0.0039 | $L_s$ (nH) |
| 1.4 | 0.7181 | 1.4 | 0.4417 | 1.4 | 0.3454 | $C_{gs}$ (pF) |
| 0.07 | 0.0306 | 0.07 | 0.0475 | 0.07 | 0.0609 | $C_{dg}$ (pF) |
| 0.4 | 0.2228 | 0.4 | 0.2229 | 0.4 | 0.2151 | $C_{ds}$ (pF) |
| 0.09 | 0.0696 | 0.09 | 0.0521 | 0.09 | 0.0410 | $g_m$ (/OH) |
| 7.0 | 3.9558 | 7.0 | 3.9558 | 7.0 | 3.9558 | $\tau$ (ps) |

Biasing Conditions

| | | | |
| --- | --- | --- | --- |
| Case 1: | $V_{ds} = 4V$ | $V_{gs} = 0.00V$ | $I_{ds} = 177mA$ |
| Case 2: | $V_{ds} = 4V$ | $V_{gs} = -1.74V$ | $I_{ds} = 92mA$ |
| Case 3: | $V_{ds} = 4V$ | $V_{gs} = -3.10V$ | $I_{ds} = 37mA$ |

TABLE II

PARAMETER VALUES OF THE MICROWAVE AMPLIFIER

| Parameter | Starting Point | Solution |
|---|---|---|
| $\ell_1$ | 52.96 | 69.01 |
| $\ell_2$ | 148.13 | 152.01 |
| $\ell_3$ | 26.80 | 18.48 |
| $\ell_4$ | 24.01 | 5.10 |
| $\ell_5$ | 46.63 | 36.49 |
| Z | 81.27 | 126.39 |

TABLE III

MULTIPLEXER CENTER FREQUENCIES AND BANDWIDTHS

| Channel | Center Frequency | Bandwidth |
|---|---|---|
| 1 | 11618.5 | 154 |
| 2 | 11495 | 76 |
| 3 | 11155 | 76 |
| 4 | 11075 | 76 |
| 5 | 10992.5 | 81 |

TABLE IV

MULTIPLEXER PARAMETERS AFTER AN OPTIMIZATION
WITH EXACT GRADIENTS

| Parameter | Ch. 1 | Ch. 2 | Ch. 3 | Ch.4 | Ch. 5 |
|---|---|---|---|---|---|
| $M_{11}$ | −0.0417 | 0.2194 | −0.0859 | 0.0454 | 0.0459 |
| $M_{22}$ | −0.0708 | 0.0301 | −0.0596 | −0.0098 | 0.0310 |
| $M_{33}$ | −0.0209 | −0.0215 | −0.0119 | −0.0097 | 0.0069 |
| $M_{44}$ | −0.0196 | −0.0621 | 0.0158 | −0.0121 | −0.0070 |
| $M_{55}$ | 0.0414 | −0.0172 | 0.0121 | 0.0023 | 0.0141 |
| $M_{66}$ | 0.0402 | 0.0117 | 0.0339 | −0.0058 | 0.0104 |
| $M_{12}$ | 0.7598 | 0.7383 | 0.7091 | 0.6115 | 0.6592 |
| $M_{23}$ | 0.5723 | 0.6096 | 0.5845 | 0.5551 | 0.5873 |
| $M_{34}$ | 0.4239 | 0.4221 | 0.4086 | 0.3048 | 0.3644 |
| $M_{36}$ | −0.5326 | −0.6346 | −0.6021 | −0.7519 | −0.6948 |
| $M_{45}$ | 0.8971 | 1.0266 | 0.9916 | 1.0317 | 1.0468 |
| $M_{56}$ | 1.1023 | 1.1518 | 1.1715 | 1.0558 | 1.1186 |
| $n_1$ | 1.0547 | 0.9358 | 0.9343 | 0.8188 | 0.8031 |
| $n_2$ | 1.4350 | 1.4311 | 1.4286 | 1.4153 | 1.4120 |
| $\ell$ | 0.7033 | 0.6039 | 0.9219 | 0.7191 | 0.7295 |

TABLE V

MULTIPLEXER PARAMETERS AFTER AN OPTIMIZATION
WITH APPROXIMATE GRADIENTS

| Parameter | Ch. 1 | Ch. 2 | Ch. 3 | Ch.4 | Ch. 5 |
|---|---|---|---|---|---|
| $M_{11}$ | −0.0432 | 0.1801 | −0.0788 | 0.0384 | 0.0401 |
| $M_{22}$ | −0.0526 | 0.0294 | −0.0514 | −0.0067 | 0.0254 |
| $M_{33}$ | −0.0082 | −0.0178 | −0.0082 | −0.0055 | 0.0044 |
| $M_{44}$ | 0.0158 | −0.0582 | 0.0160 | −0.0064 | −0.0102 |
| $M_{55}$ | 0.0160 | −0.0206 | 0.0090 | 0.0021 | 0.0038 |
| $M_{66}$ | −0.0255 | 0.0100 | −0.0248 | −0.0037 | 0.0066 |
| $M_{12}$ | 0.7427 | 0.7077 | 0.6969 | 0.6124 | 0.6495 |
| $M_{23}$ | 0.5796 | 0.5951 | 0.5815 | 0.5567 | 0.5800 |
| $M_{34}$ | 0.3855 | 0.3876 | 0.3780 | 0.3050 | 0.3491 |
| $M_{36}$ | −0.6314 | −0.6699 | −0.6540 | −0.7520 | −0.7083 |
| $M_{45}$ | 0.9657 | 1.0338 | 1.0064 | 1.0325 | 1.0324 |
| $M_{56}$ | 1.1330 | 1.1279 | 1.1346 | 1.0553 | 1.0983 |
| $n_1$ | 0.9976 | 0.8915 | 0.8997 | 0.7993 | 0.7862 |
| $n_2$ | 1.4416 | 1.4236 | 1.4223 | 1.4140 | 1.4154 |
| $\ell$ | 0.6958 | 0.6132 | 0.9235 | 0.7228 | 0.7360 |

TABLE VI

COMPARISON OF COMPUTATIONAL EFFORT

| Examples | Number of Function Evaluations | | |
| --- | --- | --- | --- |
| | Case 1 | Case 2 | Case 3 |
| FET modelling | 468 | 128 | 79 |
| Microwave Amplifier | 585 | 280 | – |
| Multiplexer | 3800 | 1000 | 500 |

Case 1: gradient estimated entirely by perturbations

Case 2: using the new algorithm without weights

Case 3: using the new algorithm with weighted update

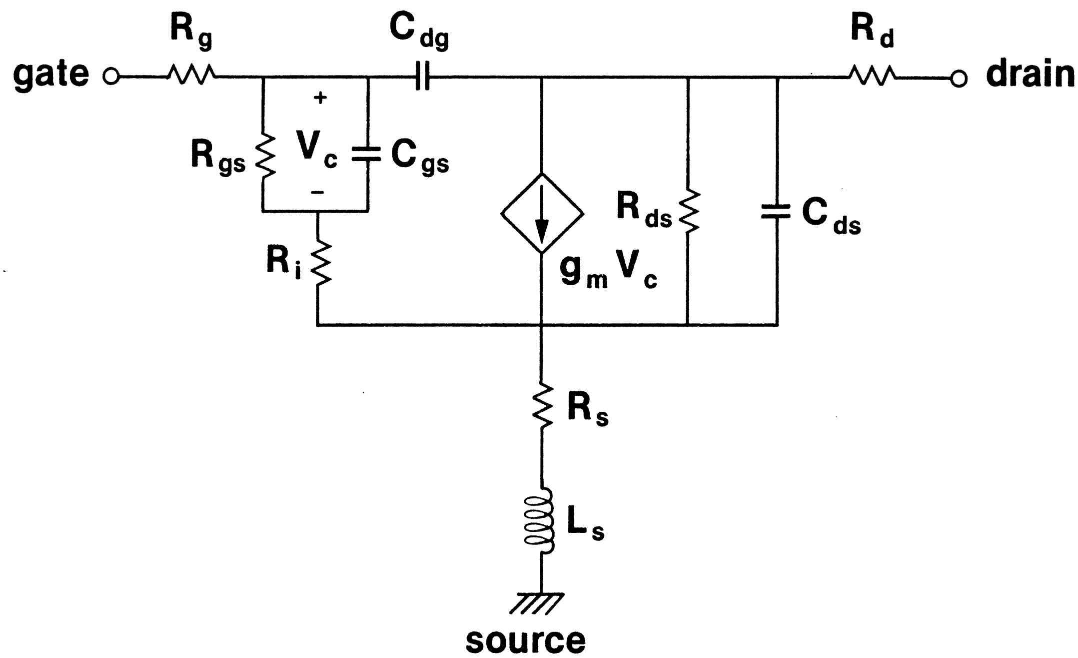


Fig. 1 The small signal equivalent circuit model for the FET.

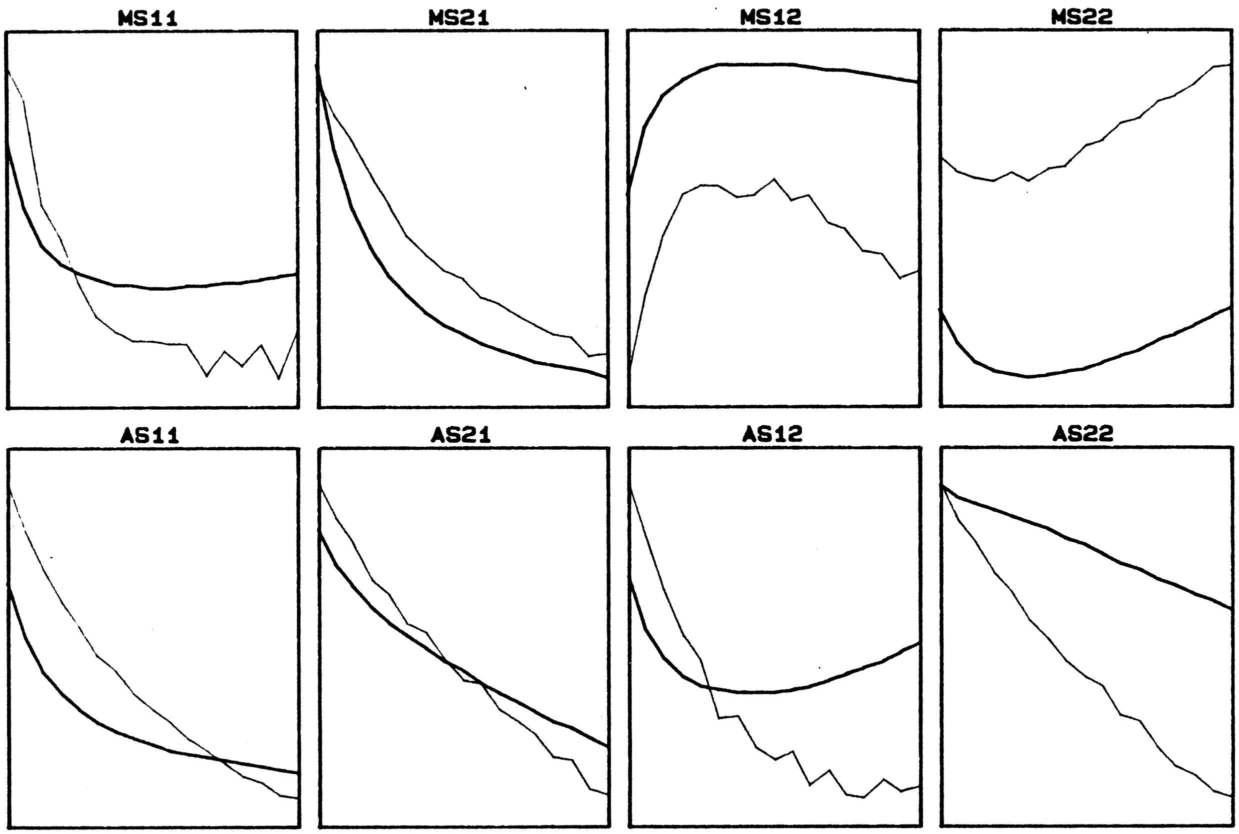


Fig. 2 The S-parameter match between the model and the measurements at the starting point. $V_{ds} = 4V$, $V_{gs} = 0V$ and $I_{ds} = 177$ mA.

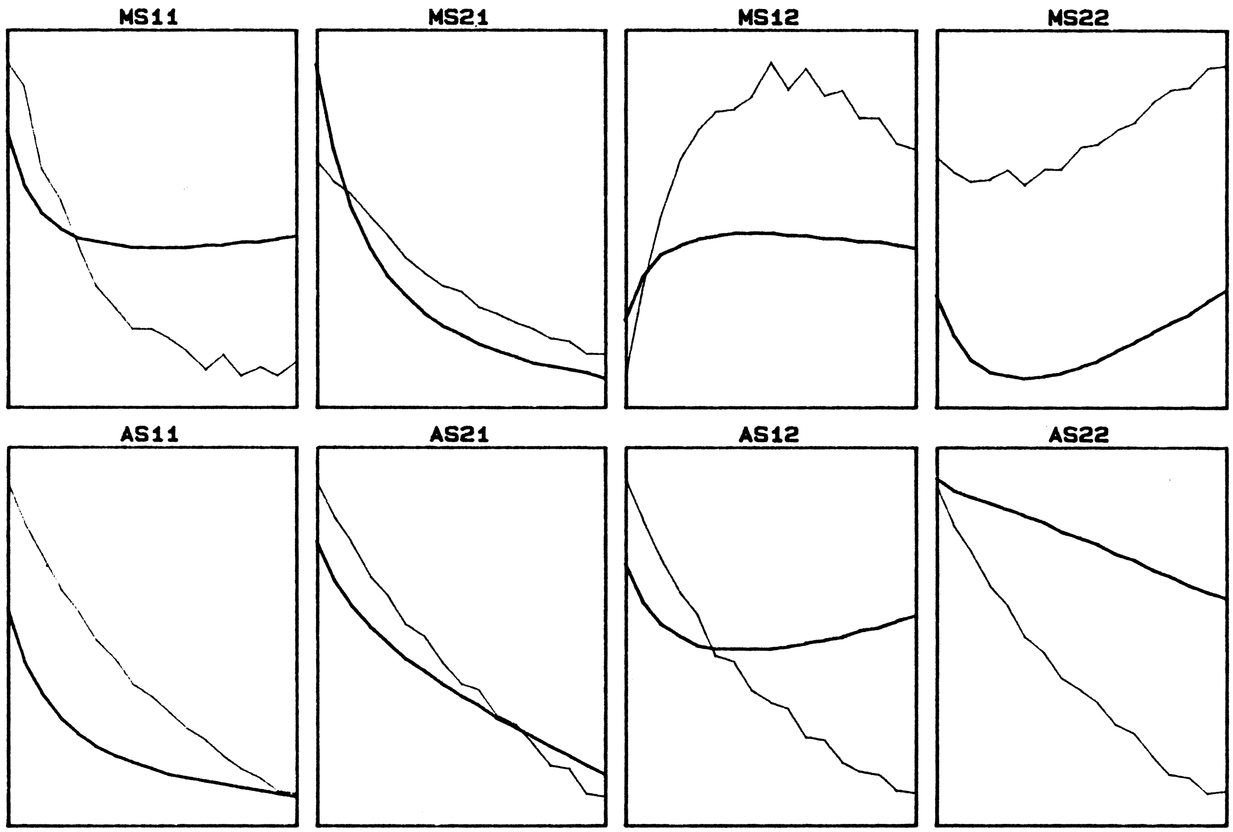


Fig. 3 The S-parameter match between the model and the measurements at the starting point. $V_{ds} = 4V$, $V_{gs} = -1.74V$ and $I_{ds} = 92$ mA.

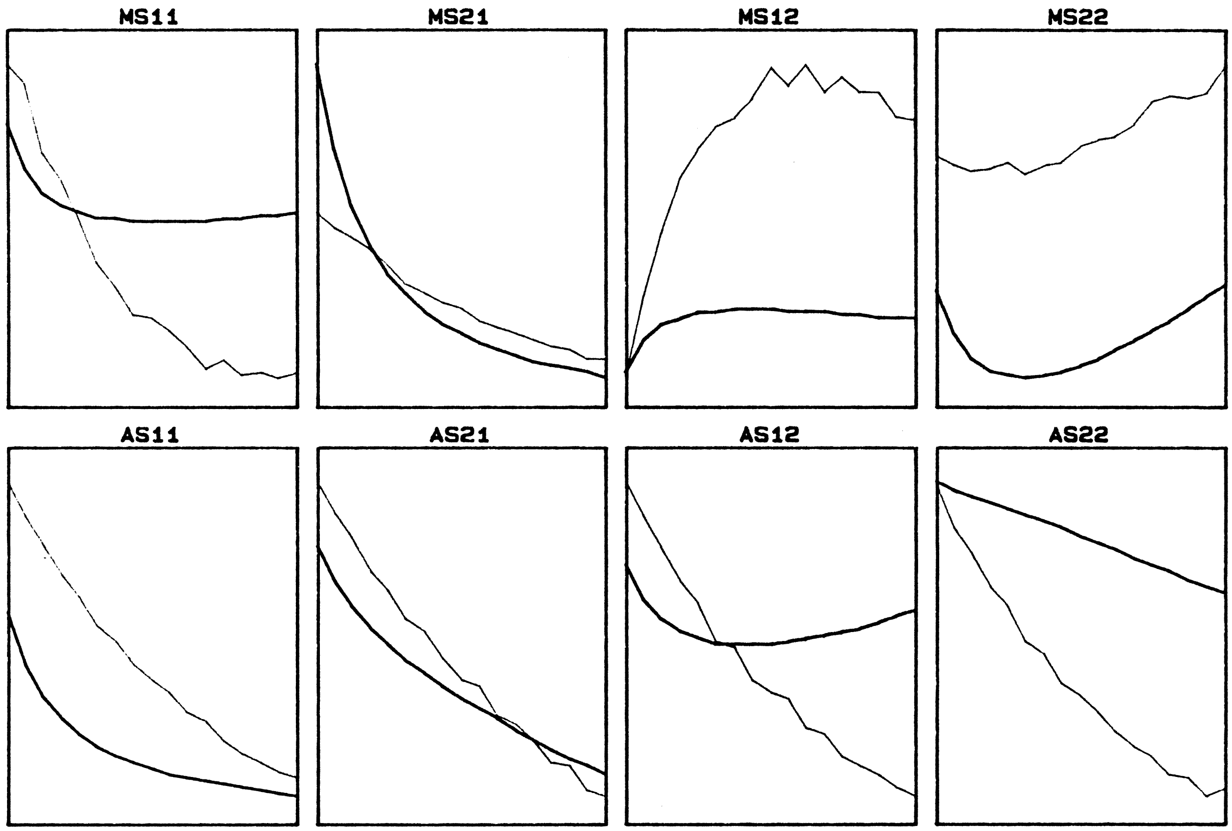


Fig. 4 The S-parameter match between the model and the measurements at the starting point. $V_{ds} = 4V$, $V_{gs} = -3.1V$ and $I_{ds} = 37$ mA.

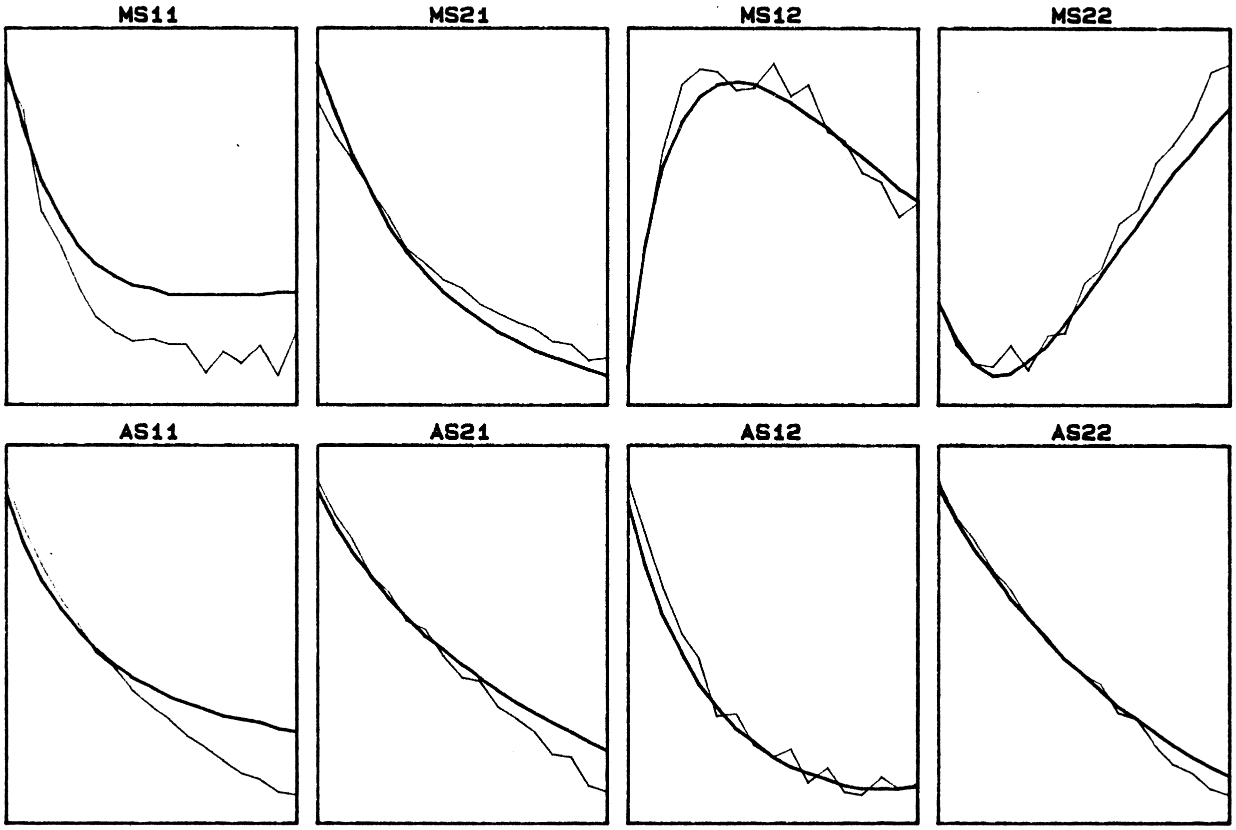


Fig. 5 The S-parameter match between the model and the measurements at the solution. $V_{ds} = 4V$, $V_{gs} = 0V$ and $I_{ds} = 177$ mA.

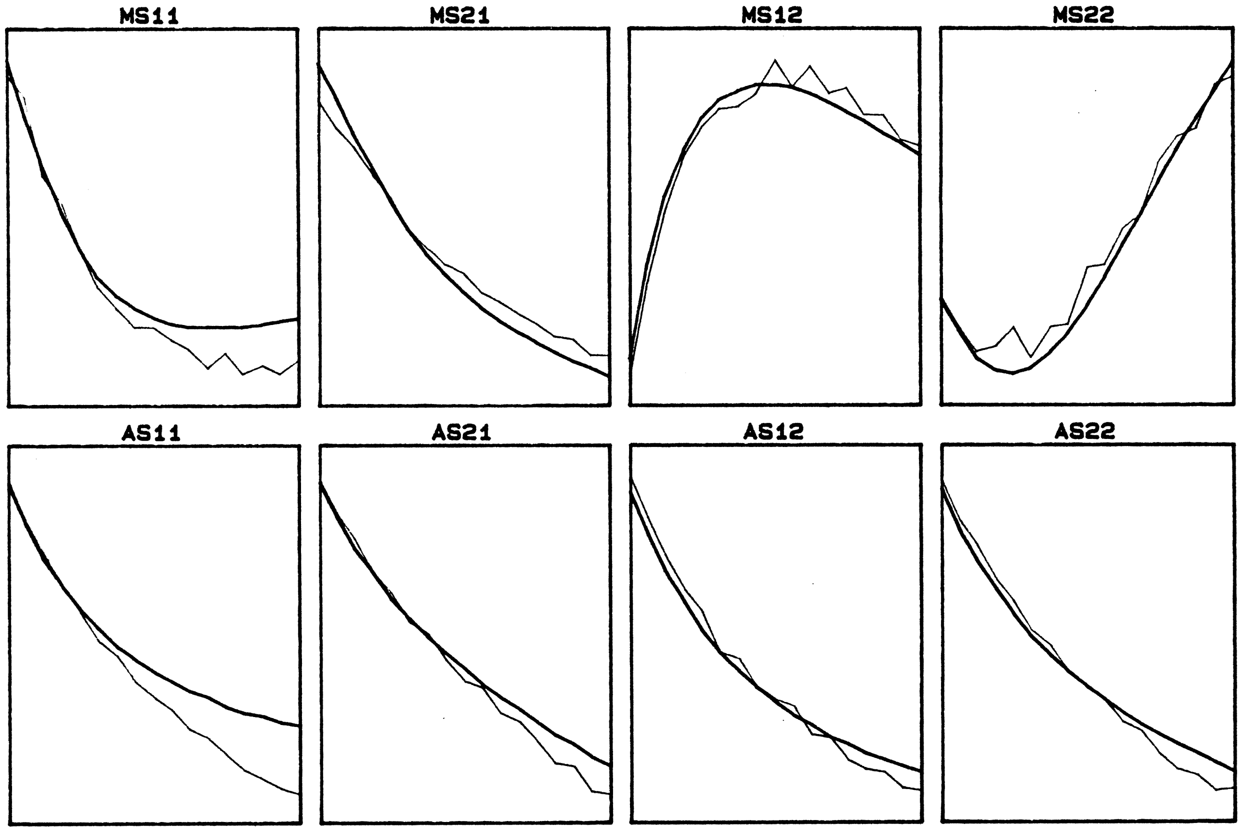


Fig. 6 The S-parameter match between the model and the measurements at the solution. $V_{ds} = 4V$, $V_{gs} = -1.74V$ and $I_{ds} = 92$ mA.

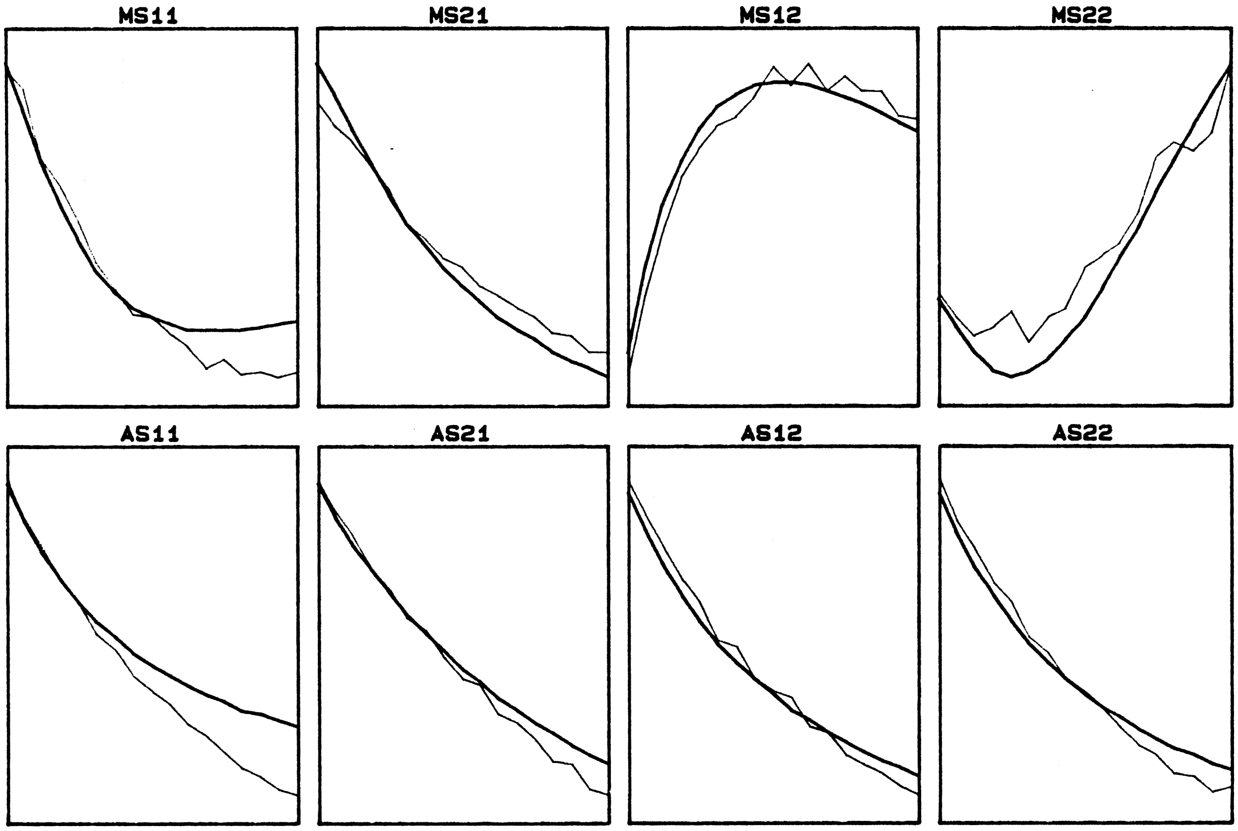


Fig. 7 The S-parameter match between the model and the measurements at the solution. $V_{ds} = 4V$, $V_{gs} = -3.1V$ and $I_{ds} = 37$ mA.

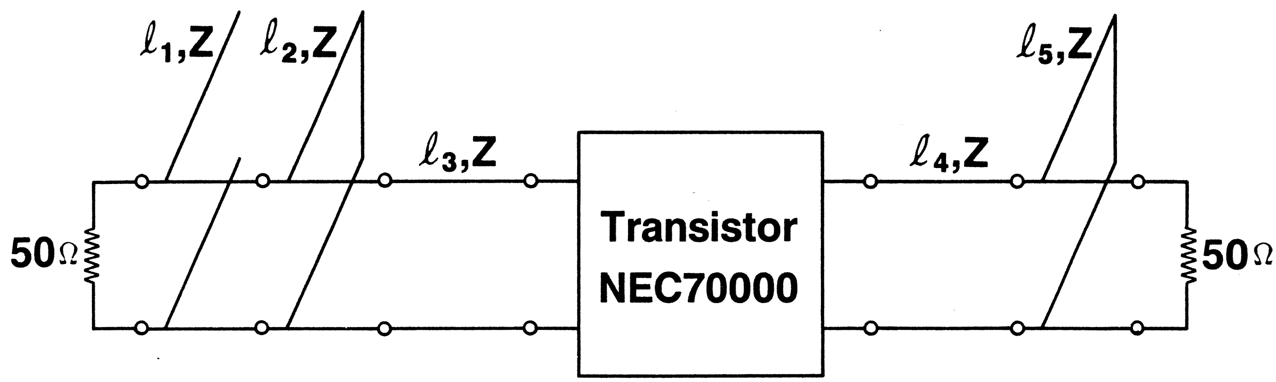


Fig. 8 A microwave amplifier [10].

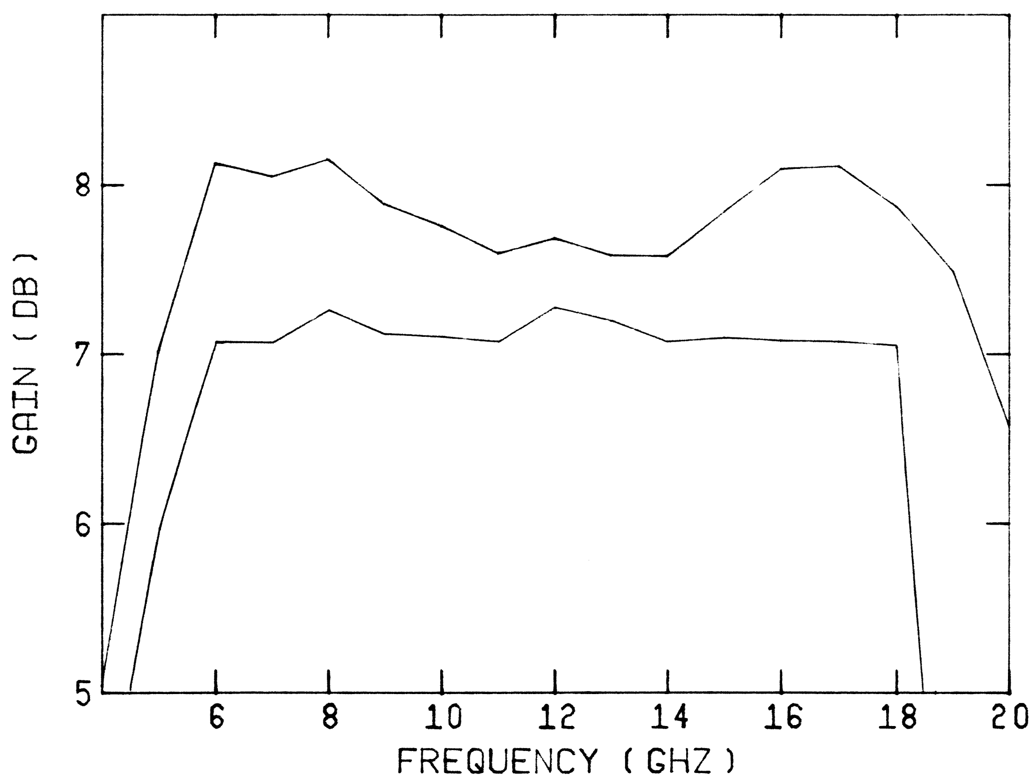


Fig. 9 The worst-case envelope for the amplifier response at the solution.

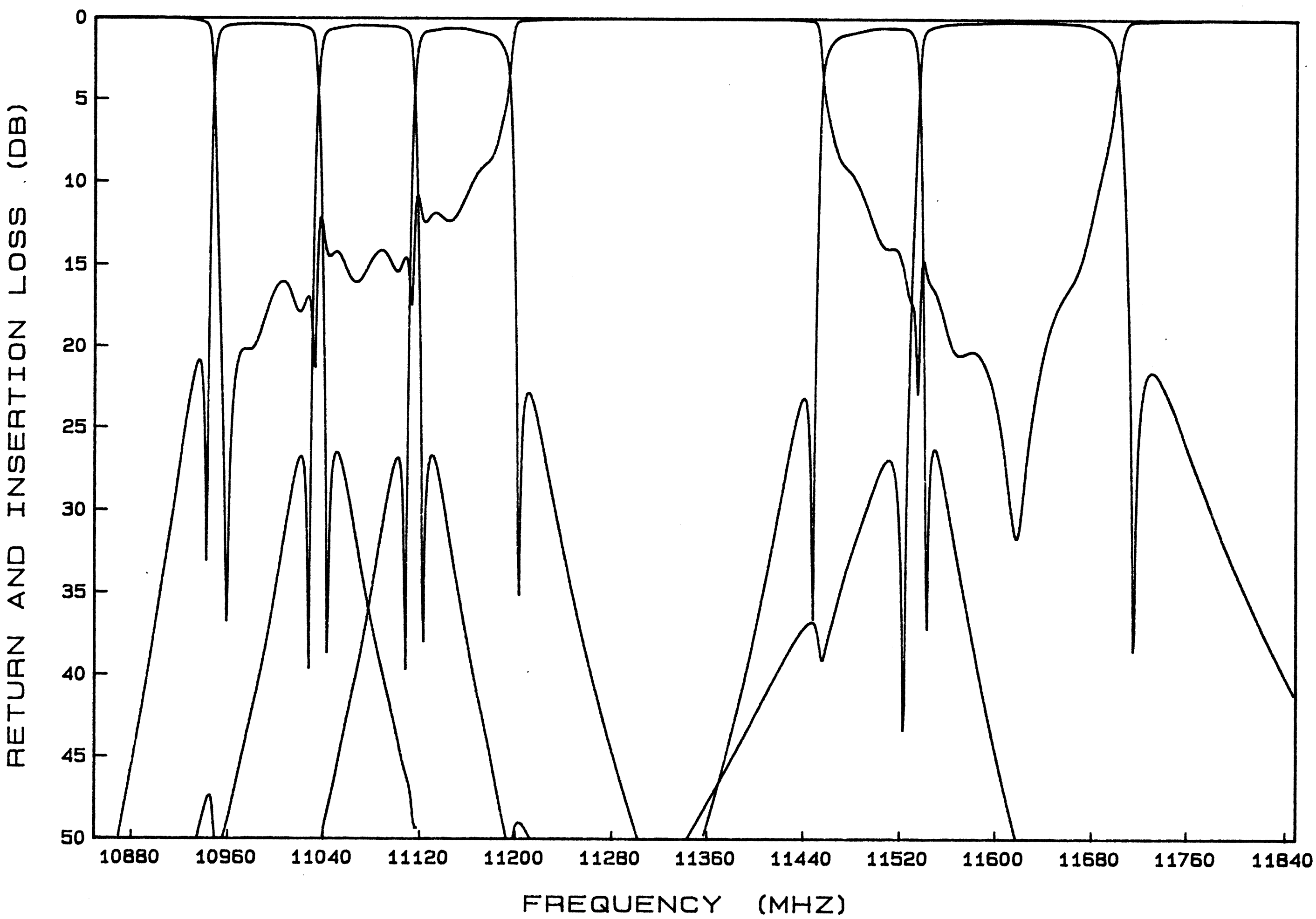


Fig. 10 Responses of the 5-channel, 11 GHz multiplexer at the starting point.

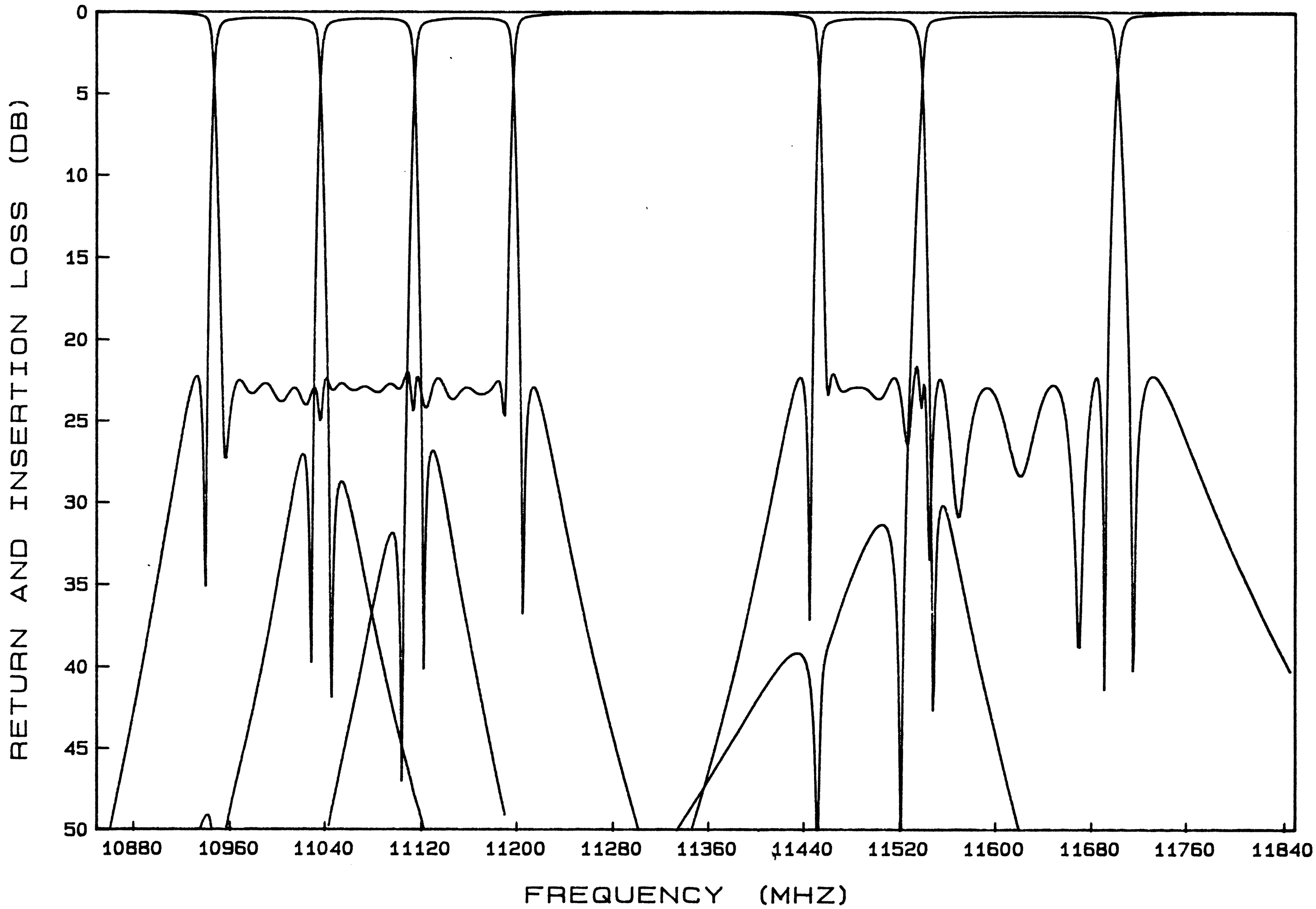


Fig. 11 Responses of the 5-channel multiplexer obtained after 50 optimization iterations using exact derivatives. It would require 3800 multiplexer response evaluations to estimate the derivatives entirely by perturbations.

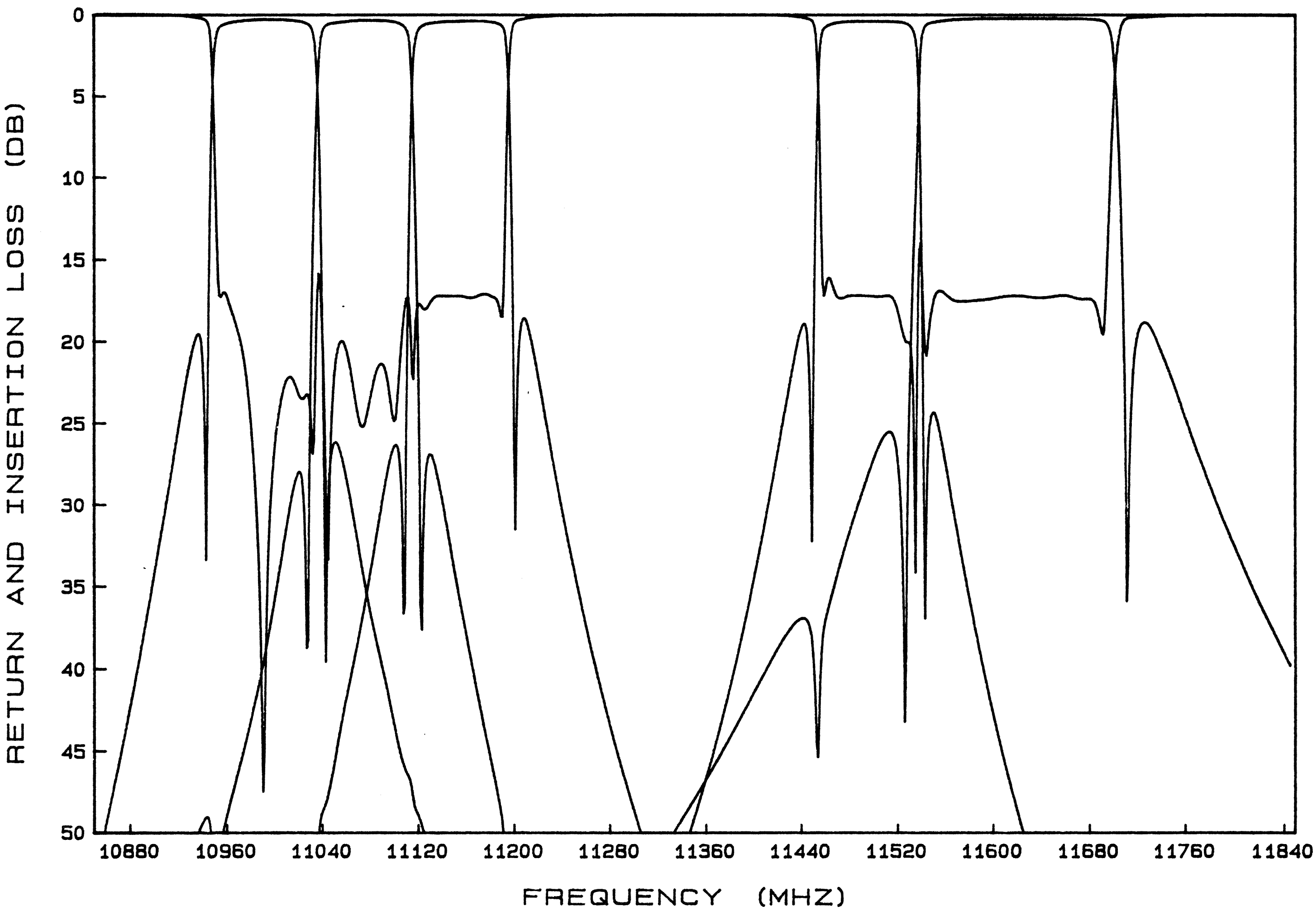


Fig. 12 Responses of the 5-channel multiplexer obtained using a simple scheme of gradient approximations. Only the Broyden update and special iterations were used as the optimization proceeded. Such a simple scheme failed to give the best possible solution.

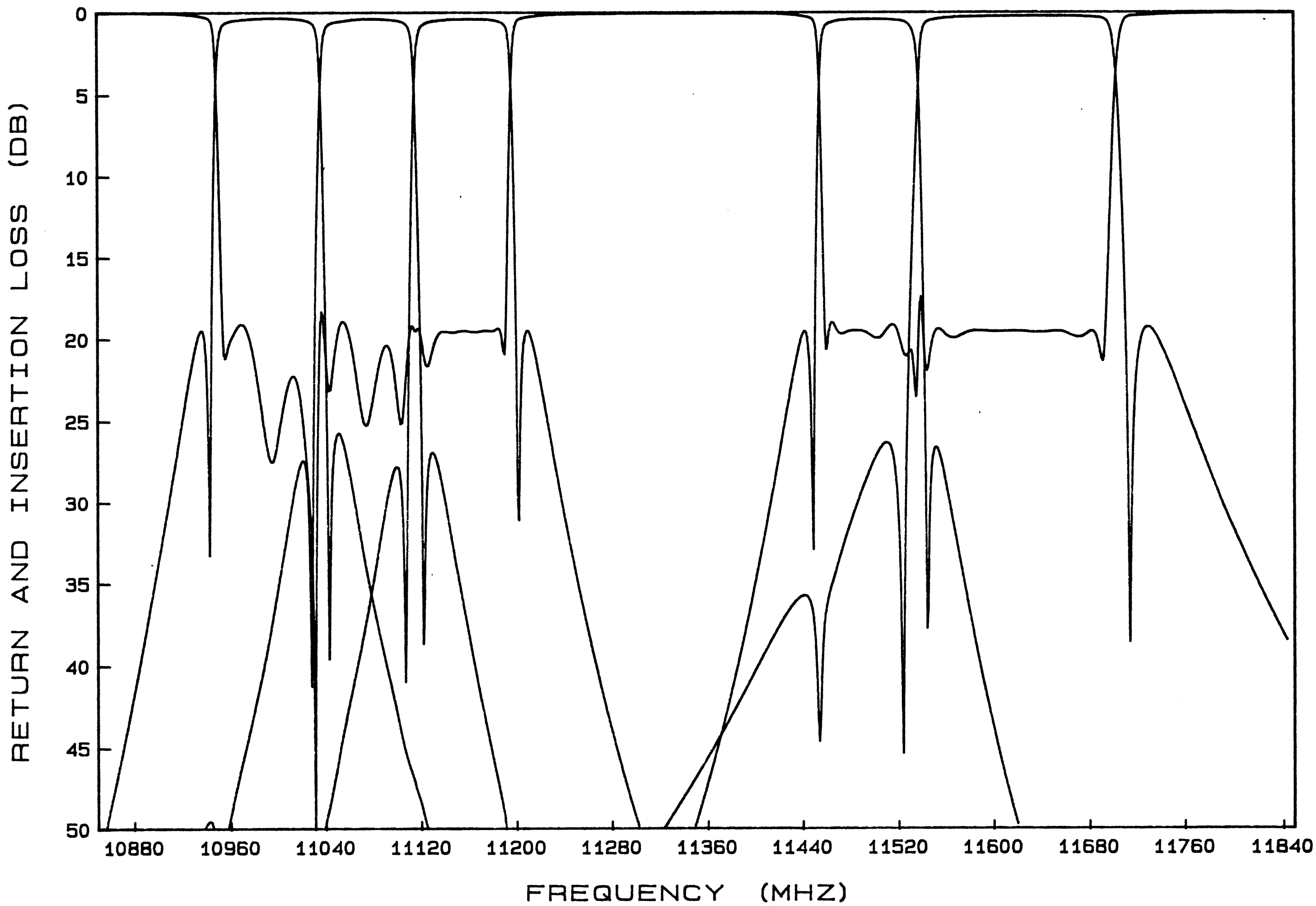


Fig. 13 Responses of the 5-channel multiplexer obtained after 500 response evaluations. Approximate gradients were used with regular corrections provided by perturbations for every 20 iterations.

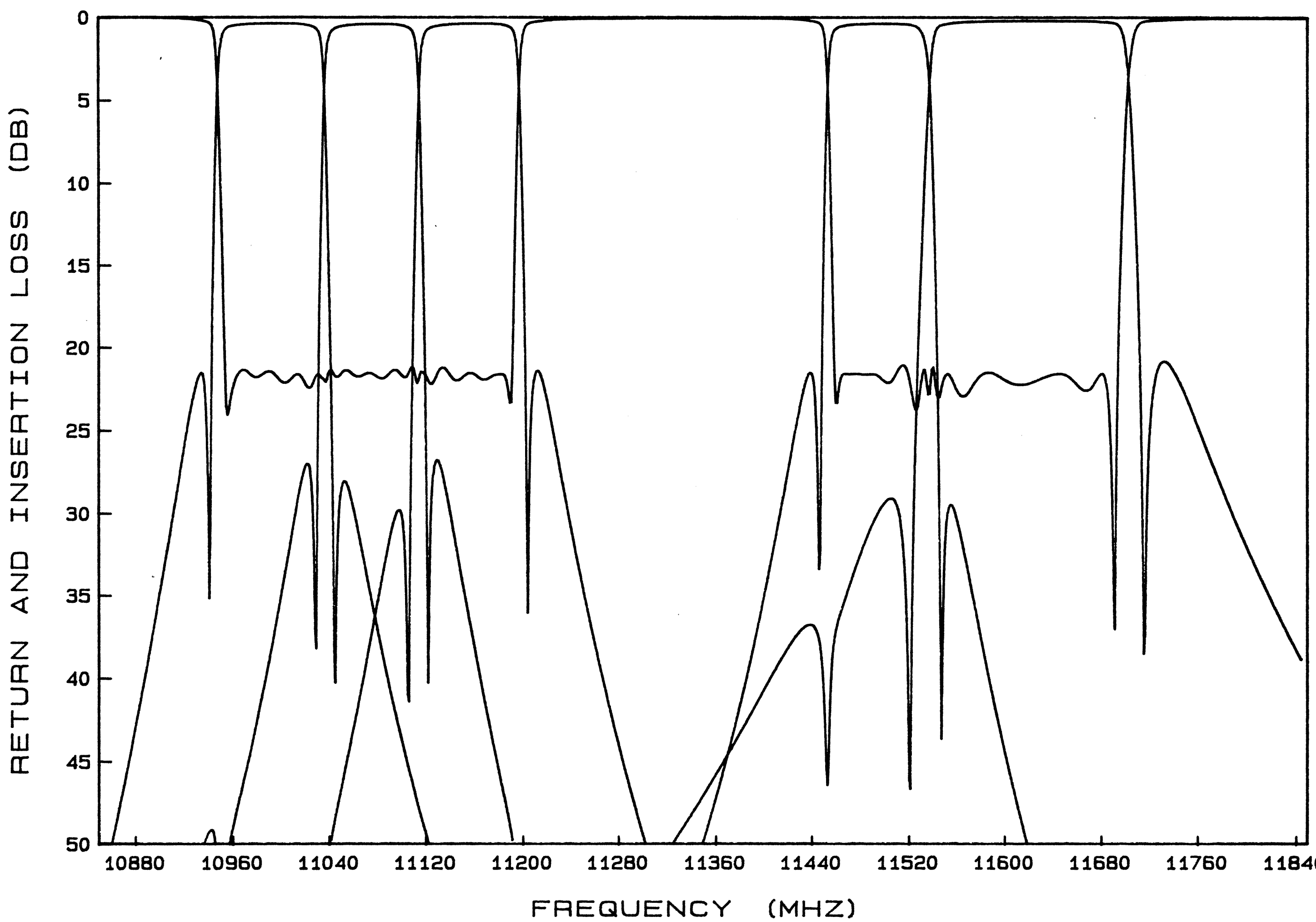


Fig. 14 Responses of the 5-channel multiplexer obtained by continuing the process described in Fig. 13 for another 500 response evaluations.

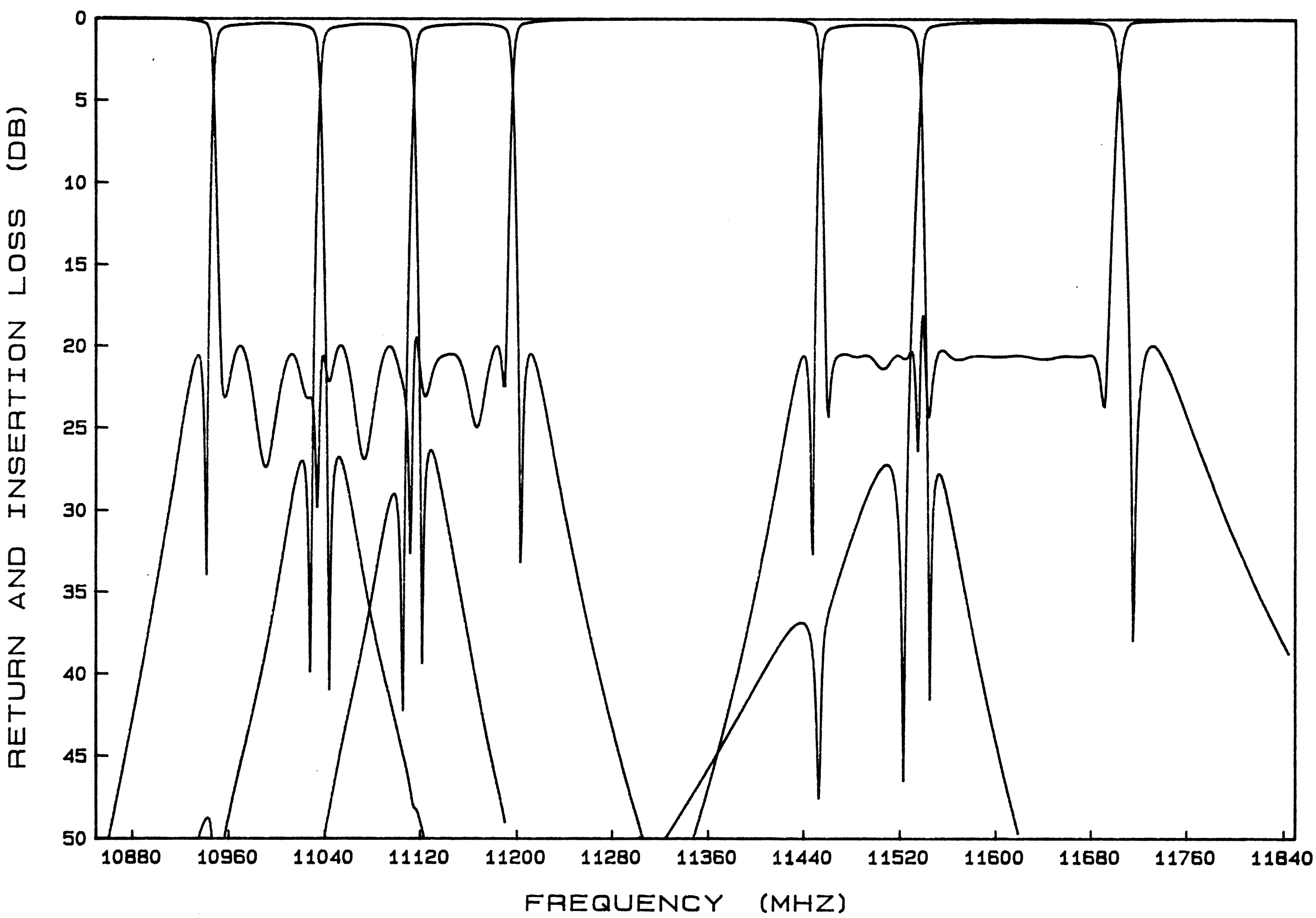


Fig. 15 Responses of the 5-channel multiplexer obtained utilizing the weighted Broyden update. It required only 500 response evaluations. The use of appropriate weights has effectively prevented the optimization from stopping prematurely (Fig. 12) and avoided the time-consuming correction by perturbations (Figs. 13 and 14).