A Systematic Optimization Strategy For Microwave Device Modelling

Andrew D. Patterson, Vincent F. Fusco, J. J. McKeown, and J. A. C. Stewart

Abstract—**Small signal GaAs MESFET equivalent circuit models are typically very ill-conditioned since the error function most often used is sensitive to some combinations of the model components and extremely insensitive to other combinations. Consequently the convergence of the parameter estimation routines used is slow and there is a good deal of uncertainty associated with the optimized values of the insensitive components. In this work the degree of ill-conditioning in the equivalent circuit model is formally quantified using a systematically formulated principal components sensitivity analysis procedure. Using this procedure it is possible to estimate for the first time how reliable the component values are in the optimized model. On this basis the extraction of the MESFET equivalent circuit model is compared using electrical model components and physical model parameters. In addition a new optimization strategy is presented which improves the condition number of the model so that rapid convergence and accurate models are ensured. This technique transforms the axes of the model from the equivalent circuit components which are correlated to the uncorrelated principal component axes which can be systematically scaled to eliminate ill-conditioning. Using this technique it is possible to obtain accurate estimates of the insensitive model parameters such as the parasitic resistances without resorting to direct measurement techniques.**

I. Introduction

THE SMALL-SIGNAL performance of MESFET devices at microwave frequencies is most often characterized by an equivalent circuit model, e.g. Fig. 1. This representation is convenient for incorporation into circuit simulation programs and can provide valuable insight into the operation of the device in a circuit environment. However, the equivalent circuit can only be used to validate actual device phenomenological behavior if physically representative component values can be defined for individual devices.

The component values of the equivalent circuit model can be extracted by various methods [1]–[6]. Usually the model is obtained by optimizing the component values to give the smallest sum of squares of errors between the model responses and the small-signal S-parameters measured for the device. However, typically there are several solutions that provide close fits to the measured data and the final solution obtained is dependent upon the initial values assigned to the components.

Manuscript received February 6, 1992; revised July 1, 1992. This work was supported by The Science and Engineering Research Council under Contract Numbers GR/F33605 and GR/F05237.

The authors are with the Department of Electrical and Electronic Engineering, The Queen's University of Belfast, Ashby Building, Stranmillis Road, Belfast BT 5AH, Northern Ireland.

IEEE Log Number 9205449.

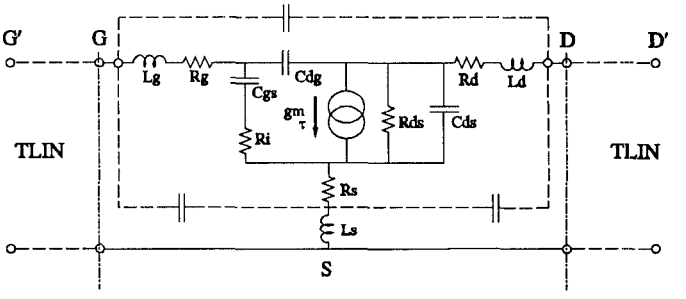


Fig. 1. MESFET device model (dashed line indicates package device elements).

There are two mechanisms by which multiple solutions originate. The first mechanism arises because the error function encompasses several distinct local minima. The optimization terminates at the closest local minimum which is not necessarily the absolute or global minimum [7], thus a sub-optimal solution may be obtained. The second mechanism arises due to the characteristic shape of the error function and the limited numerical resolution of the extraction technique. In this work we will show that in the area surrounding a minimum the error function most often used is so flat, i.e. featureless, with respect to several of the equivalent circuit elements that a large range of solutions will give equally close fits. In other words the error function suffers extremely poor sensitivity to changes in the values of some circuit elements, consequently there is uncertainty about the most physically realistic values for these elements. Fig. 2 shows the actual shape of the error function with respect to some of the components in the equivalent circuit model of Fig. 1 (using the component values given later in Table I). For example, in [6] it was found that the gate lead inductance Lg in a 10 element device model could typically vary more than 50% and still provide a satisfactory fit. In the work reported here more complex models have been used than those in [6] (from 13 up to 23 element models). This exacerbates the problem since the error function is particularly insensitive to some of the extra elements. The classical model topologies currently in use also exhibit a uniqueness problem [8]–[11]. The most significant example of this problem is the electrical similarity of the resistances Ri and Rg. This has the effect that the sum of the two components can be determined but it is difficult to determine unique individual values [11].

Uncertainty is also introduced into the equivalent circuit component values by other phenomenon: Some uncertainty is caused by small errors generated by the process of measur-

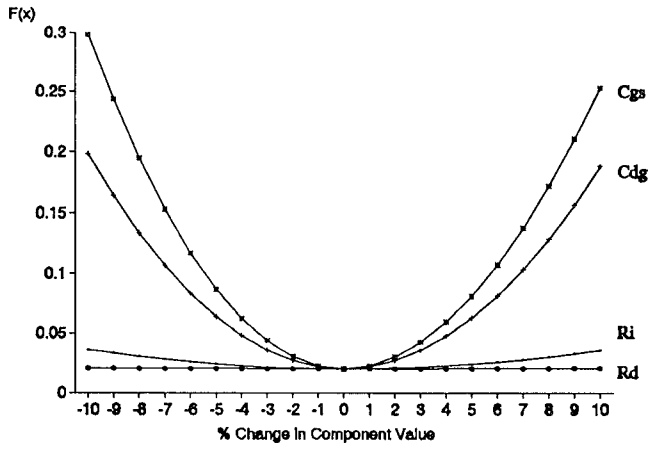


Fig. 2. Shape of error function $F(x)$ in the area of a minimum.

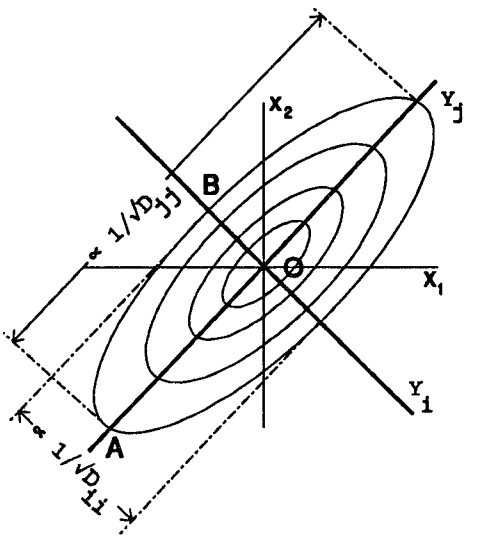


Fig. 3. Contours of error function value $F(x)$ with sensitivity directions.

ing and de-embedding the device scattering parameters [12]. Another source of uncertainty is the simplistic equivalent circuit topology used to represent actual devices. Reducing the number of components in the model topology forces the remaining components to take on less physically realistic values in order to accommodate the missing components. Also since insensitivity of the error function is responsible for uncertainty in the component values, making the model more complex by adding extra components (which are essentially less sensitive fine tuning components) increases the degree of uncertainty. The aim therefore is to create a model which is sufficiently complex to represent the important physical processes taking place in the device and yet simple enough to allow fast and accurate extraction of the component values.

Since the local minima originating by the first mechanism above are separate from each other it is possible to select the global minimum, provided an exhaustive search of local minima can be achieved [13]. In more complex models, for some components, there is a wide range of values in the area of the solution which provide equally good fits to the measured data. Primarily this occurs because the error function most often used suffers extremely poor sensitivity to several of the equivalent circuit model components. The uncertainty associated with this low component sensitivity is more difficult to reduce than the uncertainty arising from the multiple solutions produced by the local minima effects mentioned above. If we can determine starting values that lie within the area of attraction of the global minimum, then global optimization is not required. Such starting values can be determined using a physics based equivalent circuit model [14]–[16] in which the intrinsic component values are derived from the technological device parameters. Since the sensitivity characteristics of this type of model are more uniform the uncertainty in component values arising from the second mechanism above is also reduced.

The poor sensitivity of the error function to some model parameters and the uniqueness problem associated with the model topology both give rise to ill-conditioning. This in turn causes inaccurate parameter estimates and slow convergence when finite precision numerical optimization is employed. Ill-conditioning in the device model is indicated by a large condition number which is the ratio of the sensitivities of the error function to changes in the directions comprising the most and least sensitive combinations of parameters. Contours of constant error function value are shown for an arbitrary ill-conditioned least squares problem having 2 parameters x_1 and x_2, Fig. 3. For this problem the most sensitive direction is OB since the highest rate of change is along this direction, i.e. the contours are closest together along this direction. Conversely OA is the least sensitive direction. For typical MESFET model topologies such as that in Fig. 1 it will be shown that the axis of the least sensitive direction is about 100 000 times longer than the axis of the most sensitive direction. In other words the contours lie virtually parallel to the least sensitive direction. Since a relatively large move along the least sensitive direction results in a very small change in error function value the optimized solution will not be accurately defined along this direction. The least sensitive direction may coincide with a single parameter direction or it may be a combination of two or more parameter directions. For the model considered in [6] large variations in Lg had a small effect on the error function therefore Lg itself was a direction of low sensitivity. Since the effect upon the error function of an increase in Ri can be nullified by a decrease of similar magnitude in Rg the direction given by the vector $Ri - Rg$ is also a direction of low sensitivity.

Various heuristic techniques have been proposed for enhancing the convergence and accuracy of MESFET models obtained by parameter extraction [8]–[9], but these are implemented in an ad hoc fashion and do not resolve the underlying problem of ill-conditioning. This paper demonstrates how the degree of ill-conditioning in a given model can be formally quantified using a systematically formulated sensitivity analysis procedure. This analysis technique is used as a model validation tool and also to establish confidence in solutions obtained by numerical optimization. In addition, a new optimization strategy is presented which improves the condition number of the problem so that rapid convergence and accurate solutions are ensured.

II. Model Partitioning

Model partitioning is an established technique for improving the performance of optimization of ill-conditioned models with a large number of parameters. It has been shown that the convergence of the numerical routines used and the accuracy of the final solution obtained may be improved by partitioning the model parameters into groups and optimizing each group separately. If the grouping is performed so that parameters with similar sensitivity are placed together then the condition number at each optimization step may be improved significantly. For example, Kondoh [9] used an empirically derived program of optimization steps in order to provide improved uniqueness in the solutions obtained. Once the last step is completed the procedure is repeated until a sufficiently accurate solution is achieved. This method can be improved upon by using a sensitivity analysis of the error function rather than the model responses as a basis for grouping the parameters [11], [17]. The improved method discussed here is not fixed like Kondoh's method but rather it is based upon an analysis of the shape of the error function using the current model and measurement data. In addition it provides better convergence especially when bias dependence and statistical modelling are important.

Since the model parameters are correlated they can only be grouped or partitioned in an ad hoc fashion, this can lead to slow convergence; also there is no obviously intuitive way to improve convergence by scaling the parameters. However, these problems can be circumvented if more information about the shape of the error function and correlation of the model parameters can be obtained. This can be achieved using the principal components sensitivity analysis which is described next and which forms the formal underpinning for the work reported here.

III. Principal Components Sensitivity Analysis

A sensitivity analysis of the error function, $F(\underline{x})$, to variations only along the directions of the model parameter axes gives an inadequate description of the model sensitivity characteristics. This is so because the parameters are correlated and the effect of combinations of parameters may be more important than the effect of individual parameters. This is illustrated by the least squares problem in Fig. 3 having parameters x_1 and x_2. For this problem the parameters x_1 and x_2 are equally sensitive since the contours of constant $F(\underline{x})$ are equally spaced along these two directions. However, the error function is particularly insensitive to changes in the direction OA and sensitive to changes in the direction OB. The problem is therefore ill-conditioned, even though the optimizable model parameters are equally sensitive. An optimized solution will be precisely defined along the OB direction, but much less precisely defined along the OA direction. Therefore because there is some ambiguity in the value of OA there will also be some ambiguity in the optimized values of both x_1 and x_2. The most relevant sensitivity directions are therefore different from the parameter directions x_1 and x_2, they are in fact the OA and OB directions. These sensitivity directions are called the principal components. A sensitivity analysis which defines the principal components is now summarized [18]–[19].

The sum of squares function $F(\underline{x})$ can be written as

$$F(\underline{x}) = f_1^2(\underline{x}) + f_2^2(\underline{x}) + \cdots + f_m^2(\underline{x}) = \sum_{i=1}^{m} f_i^2(\underline{x}).$$

A second-order analysis is employed because a first-order analysis has no meaning since, by definition, the function is stationary at a local minimum. The second order variation in $F(\underline{x})$ can be expressed in the neighborhood of a solution, which we call $\underline{x}^*$, as the Taylor series

$$F(\underline{x}^* + \delta\underline{x}^*) \approx F(\underline{x}^*) + \delta\underline{x}^t \nabla F(\underline{x}^*) + \tfrac{1}{2}\delta\underline{x}^t \nabla^2 F(\underline{x}^*)\delta x.$$

The linear term in $\delta\underline{x}^t$ is small in the neighborhood of a solution, and it is zero if $\underline{x}^*$ is an exact solution. The principal components sensitivity analysis is an analysis of the eigensystem of the Hessian matrix $\nabla^2 F(\underline{x}^*)$. Therefore we need to calculate the hessian matrix for the particular case of a least squares function. If we define the vector $\underline{f}$,

$$\underline{f} = [f_1 \quad f_2 \quad \cdots \quad f_m]^t$$

then we can write,

$$\begin{bmatrix} \frac{\partial F}{\partial x_1} \\ \vdots \\ \frac{\partial F}{\partial x_n} \end{bmatrix} = 2 \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_1} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_n} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix} = 2J^t f$$

where J is the matrix with components $J_{ij} = \partial f_i / \partial x_j$ called the Jacobian matrix of $F(\underline{x})$. Differentiating again:

$$\frac{\partial^2 F}{\partial x_1^2} = 2\left\{ \left(\frac{\partial f_1}{\partial x_1}\right)^2 + f_1 \frac{\partial^2 f_1}{\partial x_1^2} + \left(\frac{\partial f_2}{\partial x_1}\right)^2 + f_2 \frac{\partial^2 f_2}{\partial x_1^2} + \cdots + \left(\frac{\partial f_m}{\partial x_1}\right)^2 + f_m \frac{\partial^2 f_m}{\partial x_1^2} \right\}$$

$$\frac{\partial^2 F}{\partial x_2 \partial x_1} = 2\left\{ \frac{\partial f_1}{\partial x_2}\frac{\partial f_1}{\partial x_1} + f_1 \frac{\partial^2 f_1}{\partial x_2 \partial x_1} + \frac{\partial f_2}{\partial x_2}\frac{\partial f_2}{\partial x_1} + f_2 \frac{\partial^2 f_2}{\partial x_2 \partial x_1} + \cdots + \frac{\partial f_m}{\partial x_2}\frac{\partial f_m}{\partial x_1} + f_m \frac{\partial^2 f_m}{\partial x_2 \partial x_1} \right\}$$

and so on for the other elements of the hessian matrix. By separating into groups of first derivative terms and groups of second derivative terms, we can write the hessian matrix G as follows:

$$G = 2\left[J^t J + \sum_{i=1}^{m} f_i G_i \right]$$

where G_i is the hessian matrix of the function f_i.

TABLE I
SENSITIVITY ANALYSIS PERFORMED ON THE SMALL-SIGNAL MESFET EQUIVALENT CIRCUIT MODEL OF FIG. 1 FOR A PLESSEY F20 $4 \times 75\,\mu$m DEVICE; $V_{ds} = 5$ V, $I_{ds} = 20\% I_{dss}$

Optimizable Parameter	Final Value	Principal components directions (eigenvectors) expressed as unit vectors in terms of optimizable parameter directions												
C_{gs} (pF)	0.226	-	0.1	0.7	-0.6	-0.3	-	0.1	-	-	-	-	-	-
R_i (Ω)	8.803	-	-	0.1	-	-	-	-	0.1	-	0.4	0.4	-	0.8
R_g (Ω)	2.786	-	-	-	-	-	-	-	0.2	-0.1	0.9	-0.2	-0.2	-0.4
L_g (nH)	0.013	-	-	0.1	-0.3	0.9	-	-0.2	-	-	-	-	-	-
R_s (Ω)	1.764	-	-	-	-	-	-	-	0.4	0.6	-	0.5	0.2	-0.3
L_s (nH)	0.019	-	-	0.7	0.7	0.2	-	-	-	-	-	-	-	-
C_{ds} (pF)	0.058	0.1	0.9	-0.1	-	-	-	-	-	-	-	-	-	-
C_{dg} (pF)	0.030	0.9	-0.1	0.1	0.1	-	-	-	-	-	-	-	-	-
R_{ds} (Ω)	218.290	-	-	-	0.1	0.1	-	0.6	-	-	-	-	-	-
G_m (mS)	27.477	-	-	-	-	-	-	-	-0.4	-0.6	0.1	0.6	0.2	-0.3
L_d (nH)	0.082	-	-	-	-	0.2	-0.6	0 8	-	-	-	-	-	-
R_d (Ω)	2.534	-	-	-	-	-	-	-	-	-	0.1	-0.3	0.9	-
τ (ps)	1.191	-	-	-	-	-	-	-	0 8	-0.6	-0.2	-	-	-
Sensitivity coefficient (eigenvalue)		3E4	2E3	1E3	500	150	32	12	0.1	0.06	0.02	7E-3	2E-4	9E-6

The Gauss-Newton algorithm uses the assumption that the second term in this expression is negligible in comparison to the first, i.e. the hessian is replaced by $2J^tJ$. Since the absolute value of f_i is expected to be small in the vicinity of the minimum of $F(\underline{x})$ this is usually a good approximation for small-signal MESFET modelling. Consequently we also approximate the hessian matrix by $2J^tJ$ for the principal component sensitivity analysis, although the complete hessian can also be used. Evaluation of the eigensystem of the hessian matrix $\nabla^2 F(\underline{x}^*)$ yields a considerable amount of useful information. When the hessian matrix is replaced by $2J^tJ$ the eigenvalues are non-negative since the J^tJ is at worst positive semi-definite. The hessian can be written as follows:

$$\nabla^2 F(\underline{x}^*) = \underline{Q}\underline{D}\underline{Q}^t$$

where $\underline{Q}$ is the orthonormal matrix of eigenvectors, and $\underline{D}$ is a diagonal matrix with D_{ii} the ith eigenvalue of $\nabla^2 F(\underline{x}^*)$. Therefore, writing $\underline{y} = \underline{Q}^t \delta \underline{x}$, we have

$$F(\underline{y}) = F^* + \underline{y}^t \underline{Q}^t \underline{\nabla} F^* + \tfrac{1}{2} \underline{y}^t \underline{D} \underline{y}.$$

The contours of $F(\underline{y})$ describe a family of ellipsoids whose axis lengths are inversely proportional to $\sqrt{D_{ii}}$, and whose axes lie along the y_i axes, Fig. 3. The eigenvectors of $\nabla^2 F(\underline{x}^*)$, i.e. the columns of the matrix Q, are unit vectors which define the y_i axes. These eigenvectors i.e. the y_i axes are the principal component directions. The eigenvector or principal component corresponding to the direction of maximum eigenvalue D_{ii} is the direction of maximum sensitivity to changes in the value of $\underline{x}^*$. Conversely the principal component corresponding to the direction of minimum eigenvalue D_{jj} is the direction of minimum sensitivity to changes in the value of $\underline{x}^*$. Information concerning these two principal component directions is most useful since they describe the extremes of the model sensitivity characteristics. The number of principal components is the same as the number of model parameters. In Fig. 3, since y_i is the direction of maximum sensitivity, the position of the minimum of $F(\underline{x})$ is likely to be more accurately determined along y_i than along y_j which is the direction of minimum sensitivity. When D_{jj} is very small compared to D_{ii} then the contours become nearly parallel to the y_j axis. In this case the solution may be satisfied by a large number of solutions parallel to the y_j axis and the problem is said to be ill-conditioned.

This type of analysis has the effect of completely eliminating the correlation between the model parameters. The first principal component is the linear combination of parameters for which the variance is maximum. The second principal component also obeys this condition, subject to the overriding condition that it is uncorrelated with the first, and so on for the remaining principal components. The fact that the principal components are uncorrelated will be an important factor when the scaling of model parameters is considered later.

Table I shows a summary of results for a typical principal components sensitivity analysis for the device topology given in Fig. 1 using measurements from a Plessey F20 process $4 \times 75\,\mu$m device [20] biased at $V_{ds} = 5$ V, $I_{ds} = 20\% I_{dss}$. The optimizable parameters and their final values are shown in the first two columns to the left. The remaining columns give the principal component directions (eigenvectors) expressed as unit vectors in terms of the model parameters. For example, the first principal component is the unit vector with components 0.9 in the direction of the gate to drain capacitance (C_{dg}) and 0.1 in the direction of the drain to source capacitance (C_{ds}), i.e. $0.9C_{dg} + 0.1C_{ds}$. The components of this principal component direction along the other model parameter directions are much smaller and have been omitted for the purpose of presentation. This first principal component is the

most sensitive direction, i.e. of all the possible directions we could choose to move along from this base point a step in the direction $0.9C_{dg} + 0.1C_{ds}$ will result in the largest change in error function value. The eigenvalue corresponding to each principal component direction is shown at the bottom of each column. A step from the base point along a given principal component direction will result in a change in error function value that is proportional to the corresponding eigenvalue. Therefore the eigenvalues can be considered to be sensitivity coefficients which indicate the sensitivity of the error function to moves along the principal component directions. The principal component directions are arranged in order from the most sensitive direction to least sensitive direction left to right. Large sensitivity coefficients indicate directions which are sensitive and small sensitivity coefficients indicate directions which are insensitive. The condition number provides a measure of the spread of sensitivities exhibited by the model and is defined as the ratio of the largest eigenvalue to the smallest eignevalue, D_{ii}/D_{jj}. In practice a value larger than 10^4 indicates ill-conditioning since truncation and rounding errors begin to have an adverse effect on the accuracy of the arithmetic involved. If the smallest eignevalue is zero then the problem is singular. For this particular model there are 13 principal component directions in total, one for each parameter in the model. The condition number is 3.3×10^9 (i.e. $3\times10^4/9\times10^{-6}$) at the solution indicating that the model is extremely ill-conditioned.

IV. Model Validation

The analysis described above can be used to attribute confidence to solutions which are obtained by standard circuit optimization techniques. From the sensitivity analysis given in Table I we can now estimate how reliable the component values in the least squares solution are. The most sensitive direction almost coincides with the direction of the gate to drain capacitance (C_{dg}) with just a small component in the direction of the drain to source capacitance (C_{ds}). Therefore C_{dg} will be precisely defined in the optimized solution since it is a sensitive component and a very small change in its value results in a large change in the error function value. The other sensitive principal component directions—say for example those with eigenvalues greater than 100, have large components in the directions of the elements C_{ds}, C_{gs}, L_s and L_g. Since the error function is sensitive to small changes in these component values, these components will also be precisely defined in the optimized solution. On the other hand components such as g_m, τ, R_i and the parasitic resistances R_s, R_d and R_g form the major components of the principal component directions of low sensitivity—those with eigenvalues less than 1. These components are less accurately defined since large changes in their values have a very small effect on the error function value. Since the sensitivity coefficients vary over such a wide range from 3×10^4 to 9×10^{-6} it is clear that the equivalent circuit element values are resolved with widely varying degrees of accuracy.

Since the eigenvalue corresponding to the least sensitive direction is very small (i.e. 9×10^{-6}) the model is practically singular. The least sensitive direction does not coincide with any one parameter direction but rather it is comprised of components in several different parameter directions (R_i, R_g, R_s and g_m). Therefore the condition number and consequently the convergence of the numerical routines used and the accuracy of the final solution will not be significantly improved by fixing any one parameter during optimization. The condition number will be improved if the least sensitive direction is fixed, a technique which will be developed more fully later.

The equivalent circuit topology in Fig. 1 can be used to model simultaneously via a physically based model the response of the device at multiple bias points by deriving values for the intrinsic equivalent circuit elements from the applied bias voltages and technological device parameters, such as the channel thickness and doping density [14]–[16]. This approach has the advantage that it provides realistic starting values for the optimization since the technological device parameters are known within small manufacturing tolerances. The optimizable parameters of this physics based model are the technological device parameters together with the extrinsic equivalent circuit components. The model is fitted simultaneously to the measured dc drain current and the device scattering parameters at each bias point. The values of the parasitic resistances R_s and R_d determine the bias voltages across the intrinsic device, therefore they affect the optimum intrinsic equivalent circuit component values [14], and so they are more sensitive in the physics based model.

Table II shows a summary of results for a typical sensitivity analysis performed on the physics based model using measurements for the same Plessey F20 process $4 \times 75\,\mu$m device biased at $V_{ds} = 5$ V, $I_{ds} = 20\%, 50\%$ and $100\%\, I_{dss}$. The physics based equivalent circuit model has a condition number of 2.35×10^6 (i.e. 9400/0.004) which is more than 1000 times smaller than the condition number of the electrical equivalent circuit model (3.3×10^9). The smallest eigenvalue of the physics based model is 0.004, much larger than the smallest eigenvalue of the electrical equivalent circuit model (0.000009) which is on the threshold of singularity. Consequently the most insensitive parameters in the physics based model are more precisely defined than the most insensitive parameters of the electrical equivalent circuit model. In other words there is less uncertainty in the equivalent circuit component values when they are obtained using the physics based equivalent circuit model rather than the electrical equivalent circuit model. With the physics based model there is still a relatively wide range of eigenvalues—from 9400 to 0.004, so the model parameters are again resolved with varying degrees of accuracy. The most sensitive principal component directions—those with eigenvalues greater than 100 are entirely comprised of the technological device parameters; the saturation velocity (V_{sat}), gate width (Z_G), doping density (N), gate length (L_G), channel thickness (W), built in potential (V_{bo}) and the space-charge layer extension coefficient (a_o). These parameters will be defined with a high degree of precision in the final solution. The model parameters with large components in the least sensitive principal component directions—those directions with eigenvalues less than 1, will be less accurately defined. The source resistance R_s forms the

TABLE II
TYPICAL SENSITIVITY ANALYSIS PERFORMED ON THE PHYSICS BASED EQUIVALENT CIRCUIT MODEL [14] FOR A PLESSEY F20 4 × 75 μm DEVICE BIASSED AT V_{ds} = 5.0 V, I_{ds} = 10%, 20% and 50% I_{dss}

Optimizable Parameter	Final Value	Principal component directions (eigenvectors) expressed as unit vectors in terms of the optimizable parameter directions																
V_{bo} (V)	0.743	-0.2	-	-0.2	0.6	-0.6	-	0.1	-	-0.1	-0.1	-	-	-	-	-	-	-0.3
V_{sat} (ms^{-1})	78527	0.7	-0.1	-0.4	0.1	-	0.1	0.2	0.4	0.2	-	-	-	-	-	-	-	-
W (μm)	0.1083	0.3	0.6	0.6	-	-0.1	-	0.1	0.2	-	0.1	-0.2	-0.2	-	-	-0.2	-	-
N (m^{-3})	3.17E23	0.4	0.1	0.3	0.2	-	-	-0.1	-0.4	-	-0.3	0.3	0.3	-	-	0.4	-	-
L_G (μm)	0.5529	-0.3	-0.4	0.5	0.4	0.2	0.2	0.3	0.2	0.2	-	-0.1	0.1	-	-	0.1	-	-
Z_G (μm)	300.0	0.4	-0.6	0.3	-	-0.2	-0.1	-0.2	-0.1	-0.3	-	-	-0.2	-	-	-0.3	-	-
a_0	1.099	-	-0.1	0.2	-0.4	0.3	-0.5	0.2	0.4	0.2	-0.3	0.1	-	-	-	0.2	-	-
r_{01} (Ω)	342.64	-	-	-	-0.2	-	-	0.8	-0.5	-0.1	-	-	-	-	-	-0.1	-	-
r_{02} (Ω/V)	95.46	-	-	-	-	0.1	0.2	-	0.2	-0.3	-0.2	0.2	0.5	0.4	-	-0.5	-0.2	-0.2
r_{03} (Ω/V^2)	3.881	-	-	-	-	-0.1	-0.3	-0.1	-0.2	0.6	0.3	-0.3	0.4	0.2	-0.1	-0.2	-	-0.1
R_g (Ω)	2.786	-	-	-	-	-	-	-	-	0.3	0.2	0.7	-0.1	-0.3	-0.3	-0.2	-0.1	-0.2
L_g (nH)	0.010	-	-	-	-	-0.1	-	-	-	-0.1	-	-0.1	0.4	-0.8	-	-0.1	-	-
R_s (Ω)	1.764	-	-	-	0.1	-0.3	-	-	-	-	0.2	0.3	0.1	0.1	-	-0.1	-	0.8
L_s (nH)	0.019	-	-	-	-	-	-	-	-	-0.1	0.5	0.2	0.2	-	0.6	0.2	0.2	-0.3
C_{ds} (pF)	0.058	-	-	-	-0.3	-0.5	0.7	-	-	0.3	-	-	-	-	-	0.1	-	-
L_d (nH)	0.082	-	-	-	-0.1	-0.2	-	-	0.1	-0.3	0.4	-	0.1	0.1	-0.6	0.4	-	-
R_d (Ω)	2.534	-	-	-	-	-	-	-	-	-	-0.1	-	0.1	-	-0.3	-0.1	0.9	-
Sensitivity coefficient (eigenvalue)		9400	4600	1450	82	56	8	5	3	1	0.6	0.40	0.20	0.14	0.10	0.04	0.01	0.004

largest component of the least sensitive direction (the direction with eigenvalue 0.004), but since R_s is also a component of directions with larger eigenvalues (82 and 56) it is defined with reasonable accuracy in the final solution. On the other hand the drain resistance R_d only comprises principal component directions with eigenvalues smaller than unity.

The physics based equivalent circuit model provides a great improvement over the electrical equivalent circuit model in terms of model condition number. Nevertheless the model still suffers, to a lesser extent, slow convergence and inaccurate estimates for the most insensitive parameters as a result of ill-conditioning. However, by automatically scaling the model parameters using the information obtained by the principal component sensitivity analysis this ill-conditioning can be eliminated. In this way it is possible to obtain accurate estimates for the insensitive model parameters such as R_s, R_d and R_g without recourse to direct measurement techniques. The scaling procedure will now be discussed.

V. FULL AUTOMATIC SCALING

The analysis of the eigensystem of the Hessian matrix employed in the sensitivity analysis provides a basis for a transformation of the model space (Cgs, Ri, Cds etc.) to a new set of uncorrelated variables y_i (which are the eigenvectors or principal component directions given by the columns in Tables I and II). This transformation is equivalent to a rotation of the coordinate axes so that a unit change in y_i has the effect of altering the model parameters by an amount proportional to the ith eigenvector in $\underline{Q}$ (which is the ith column in Tables I and II). Such a change would effect the error function value by an amount which is proportional to the i th eigenvalue D_{ii} (which is the sensitivity coefficient in the ith column in Tables I and II). If the new variables y_i are scaled in inverse proportion to the square root of their corresponding eigenvalues D_{ii}, Fig. 3, then a unit change in any of the y_i variables will have equal effect on the sum of squares function for all i. A sensitivity analysis obtained for this system is shown in Table III. The error function is now locally spherical, and the model condition number is 1. Since the error function is now equally sensitive to moves in each of the principal component directions the model parameters

TABLE III
SENSITIVITY ANALYSIS AFTER FULL AUTOMATIC SCALING

Optimizable Parameter	Final Value	Principal component directions				
Vbo	0.743	1.0	0	0	...	0
Vsat	78527	0	1.0	0		0
W	0.1083	0	0	1.0		0
N	3.17E23	0	0	0	...	0
						0
rd	2.534	0	0	0		1.0
Sensitivity coefficient (eigenvalue)		1.0	1.0	1.0	...	1.0

TABLE IV
OPTIMIZATION RESULTS FOR PLESSEY F20 4 × 75 μm DEVICE

Plessey F20 4x75μm MESFET						
	Electrical Equivalent Circuit Model (one bias point)			Physics Based Equivalent Circuit Model (three bias points)		
Scaling Technique	No Scaling	Scaling to Unity	Full Automatic Scaling	No Scaling	Scaling to Unity	Full Automatic Scaling
Total Iterations	247	165	6	300*	66	14
Final RMS Error	0.00712	0.00709	0.00700	0.0455	0.0377	0.0377
Comment	starting point dependent	starting point dependent	starting point dependent	non-unique	physical solution	physical solution

* Optimization did not converge before 300 iterations

$(Cgs, Ri, Cds$ etc.) are all optimized with similar accuracy. Furthermore, the spherically shaped error function provides the fastest convergence rate possible. The transformation used is

$$\underline{x} = \underline{Q}\underline{D}^{-1/2}\underline{y},$$

where $\underline{x}$ is the new set of optimization variables, $\underline{Q}$ is the matrix of eigenvectors of the Hessian, $\underline{D}$ is the diagonal matrix of eigenvalues and $\underline{y}$ is the matrix of model parameters. Since $\underline{Q}$ and $\underline{D}$ are functions of $\underline{x}$ the transformation must be recomputed periodically during optimization to keep the condition number after scaling small as the optimization variables are adjusted. Provided the model is not ill-conditioned (after scaling) then the search direction chosen by the Gauss–Newton optimization algorithm is not dependent upon parameter scaling. Therefore the condition number must only be kept less than about 10^4, the threshold for ill-conditioning. Recomputation is needed more often if the function is locally very nonlinear (this overhead is not great for small-signal MESFET modelling) or if the starting values are far from the solution.

VI. SUB- SPACE REDUCTION

When the condition number is excessively large, e.g. 10^9, full automatic scaling involves computations with both very large and small eigenvalues with consequent rounding and truncation errors. As a result the accuracy of the final solution and the convergence of the numerical routines used decreases. This effect is particularly pronounced as the number of model parameters increase. In order to improve the performance of the numerical optimization of such models, the model can be partitioned on the basis of the distribution of eigenvalues in Tables I and II, i.e. using the transformed model space. Since the transformed variables are uncorrelated they can be systematically divided into groups of similarly sensitive variables. We call this technique sub-space reduction because two or more sub-spaces are formed each with smaller condition numbers than the original full-space problem. For example, the equivalent circuit model giving the sensitivity analysis of Table I could be divided into two sub-spaces bounded by the sensitivity coefficients $3\times10^4-0.1$ and $0.06-9\times10^{-6}$ having condition numbers of about 3×10^5 and 7×10^3 respectively. Full automatic scaling of these sub-spaces will be performed considerably more accurately than for the full-space model with a condition number of 3×10^9. Using this technique each sub-space is optimized in turn until no further improvement can be made.

VII. RESULTS

The full automatic scaling and sub-space reduction techniques presented above have been compared with conventional optimization techniques popularly used for small-signal MESFET device modelling. The sum of squares of the errors between the measured data and the model responses was minimized using a Gauss–Newton algorithm. The modelling procedure was performed with (i) full automatic parameter scaling and sub-space reduction, (ii) no parameter scaling, and (iii) parameter scaling to unity, Tables IV and V. Scaling the parameters to unity by simply dividing each by their own value often improves convergence since it removes the effect of different units. Models were extracted for several examples of the Plessey F20 4×75 μm on-wafer MESFET and the NEC NE71083 [21] packaged MESFET. Both devices are used in I and J bands and apart from the effect of the package they provide similar performance. A packaged device was chosen because the package adds complexity to the device model making it more difficult to extract precise model parameters. S-parameters of both MESFETs were measured at bias points across their entire operating range, the packaged devices were measured in a test fixture de-embedded using a Thru-Reflect-Line calibration [22]. The on-wafer devices were measured using Line-Reflect-Match calibration [23].

The on-wafer devices were modelled using the topology given in Fig. 1. Adding transmission lines on the gate and drain, and geometric capacitances between the gate, drain and source was sufficient to simulate the package of the NE71083 device, Fig. 1. Both devices were characterized at a single bias point using the electrical equivalent circuit model, and also over multiple bias points using the physics based equivalent circuit model. The optimizable parameters for the physics based model are the extrinsic bias independent equivalent circuit model components (including package components for the NE71083) and the technological device parameters

TABLE V
OPTIMIZATION RESULTS FOR NEC NE71083 DEVICE

NEC NE71083 MESFET						
	Electrical Equivalent Circuit Model (one bias point)			Physics Based Equivalent Circuit Model (three bias points)		
Scaling Technique	No Scaling	Scaling to Unity	Full Automatic Scaling	No Scaling	Scaling to Unity	Full Automatic Scaling
Total Iterations	300*	300*	16	300*	300*	46
Final RMS Error	0.03902	0.03654	0.03648	0.06080	0.04523	0.04508
Comment	starting point dependent	starting point dependent	starting point dependent	non-unique	physical solution	physical solution

* Optimization did not converge before 300 iterations

from which the intrinsic component values are derived. The physics based equivalent circuit model preserves the inherent correlations between model components and provides reliable physically representative values for the equivalent circuit elements [16].

Tables IV and V compare typical performance of the three scaling techniques outlined above in extracting the parameters of the device models. Modelling the on-wafer devices is described first. The single bias point equivalent circuit model provided the closest match to measured scattering parameters. However, it did not provide reliable physically representative component values and the final solution achieved proved to be quite dependent upon the choice of starting values. All three scaling techniques converged to similar solutions from identical starting values, although sub-space reduction gave the lowest rms error. Scaling the parameter values to unity enabled the error to be reduced more quickly in the first few iterations than with no parameter scaling, but in the following iterations both methods converged extremely slowly. The sensitivity analysis (Table I) shows that the model is almost singular, which is the reason for the slow convergence. Initially the error reduced quickly as the sensitive components were adjusted but many iterations were subsequently used as changes made to the values of the insensitive components had little effect on the error function. Therefore, using full automatic scaling and sub-space reduction the model space was optimized in two sub-spaces giving a dramatic improvement in convergence: With sub-space reduction (each sub-space scaled once) just 6 iterations in total were required to obtain the final solution, and after 4 iterations the error was smaller than the final solutions obtained by unity parameter scaling in 165 iterations (scaled three times) and no parameter scaling in 247 iterations. The largest variance between the solutions was seen in the values of the least sensitive components, e.g. the parasitic resistances, which were resolved more accurately by sub-space reduction since the final rms error was lowest with this method.

A physical solution was obtained using the physics based equivalent circuit model. With no parameter scaling the rate of convergence was slow and the optimization was stopped without converging after 300 iterations. A solution was obtained using full automatic scaling (sub-space reduction was not used since the condition number was sufficiently small—2.35×10^6)

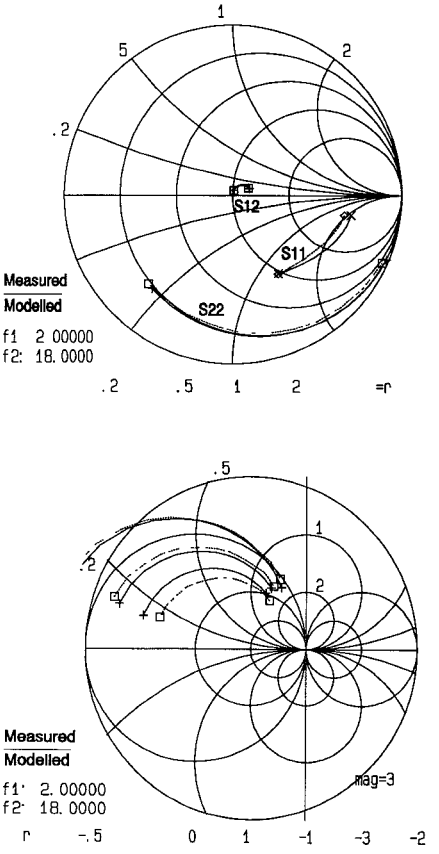


Fig. 4. S-Parameter Fits for Sample Plessey F20 $4 \times 75\,\mu$m Device: ($V_{gs} = 0.0, -0.8, -1.5$ V; $V_{ds} = 5.0$ V (Only $V_{gs} = -0.8$ V shown for S11, S12, S22).

in just 14 iterations compared to 66 when the parameters were scaled to unity, Table IV. In both solutions the final values of the technological parameters were similar but the insensitive parameters such as the parasitic resistances had more physically representative values in the solution obtained using full automatic scaling. The drain currents simulated by the physics based model—12 mA, 29 mA and 57 mA also matched the measurements closely—12 mA, 30 mA and 56 mA respectively. Sample fits obtained for the F20 process $4 \times 75\,\mu$m device using full automatic scaling are shown in Fig. 4 later in the article. S11, S22 and S12 are only shown

TABLE VI
FINAL OPTIMIZED PARAMETER VALUES FOR THE PHYSICS BASED MODEL OF THE PLESSEY F20 4 × 75 μm DEVICE

Physical Parameters	Optimized Value	Bias Independent Parameters	Optimized Value
Vbo (V)	0.743	Rg (ohms)	2.786
Vsat (ms^{-1})	78527	Lg (nH)	0.010
W (m)	0.1083	Rs (ohms)	1.764
N (m^{-3})	3.17E23	Ls (nH)	0.019
LG (m)	0.5529	Cds (pF)	0.058
a0	1.099	Ld (nH)	0.082
r01	342.64	Rd (ohms)	2.534
r02	95.46		
r03	3.881		

TABLE VII
FINAL EQUIVALENT CIRCUIT ELEMENT VALUES DERIVED FROM THE PHYSICS BASED MODEL FOR THE PLESSEY F20 4 × 75 μm DEVICE

Bias Dependent Parameter	Vgs = 0.0V Vds = 5.0V Ids = 56mA	Vgs = -0.8V Vds = 5.0V Ids = 30mA	Vgs = -1.38V Vds = 5.0V Ids = 12mA
Cgs (pF)	0.4088	0.2787	0.2261
Cdg (pF)	0.0262	0.0289	0.0300
gm (mS)	43.859	32.464	27.477
τ (pSec)	2.2817	1.5459	1.1906
Ri (ohms)	1.8433	3.4483	8.8034
Lg (nH)	0.0026	0.0036	0.0042
Rds (ohms)	342.64	274.12	218.29

at one bias point since they are very similar over the range of bias considered. The final optimized parameter values are shown in Table VI and Table VII shows the corresponding intrinsic equivalent circuit component values at the bias points measured. The gate inductance L_g consists of an intrinsic bias dependent part representing the phase delay in the gate and an extrinsic bias independent part representing the inductance of the gate contact. Comparison of the optimized technological device parameters with the device specification indicates that this solution is a good representation of the actual device physics.

Modelling the packaged device was more difficult since the equivalent circuit topology is considerably more complex than it is for on-wafer devices. The electrical equivalent circuit model uses 18 parameters and has a condition number of the order 10^{11}, and the physics based equivalent circuit model uses 23 parameters with a condition number also of the order 10^{11}. Starting values were estimated for the package components in the electrical equivalent circuit model by trial and error. But, since the technological parameters were known within small tolerances they could be fixed while the package components were optimized to provide a reasonable starting approximation for the package in the physics based equivalent circuit model. The optimization converged very slowly for both the electrical and physics based equivalent circuit models when no parameter scaling and unity parameter scaling were used, not finding a solution in less than 300 iterations. Using full automatic scaling in two subspaces solutions were found for the electrical and physics based equivalent circuit models in 16 and 46 iterations respectively. Once more the electrical equivalent circuit model gave the closest fit to the measured data but the more constrained physics based model gave more physically representative component values. The largest variance between models was seen in the values of the insensitive components, i.e. the package and parasitic components. Fig. 5 (on the next page) shows fits obtained for an NE71083 device using the physics based equivalent circuit model. By fitting this model at a larger number of bias points the sensitivity of some model parameters can be improved (thus improving the model uniqueness), this is particularly true for the package and parasitic components.

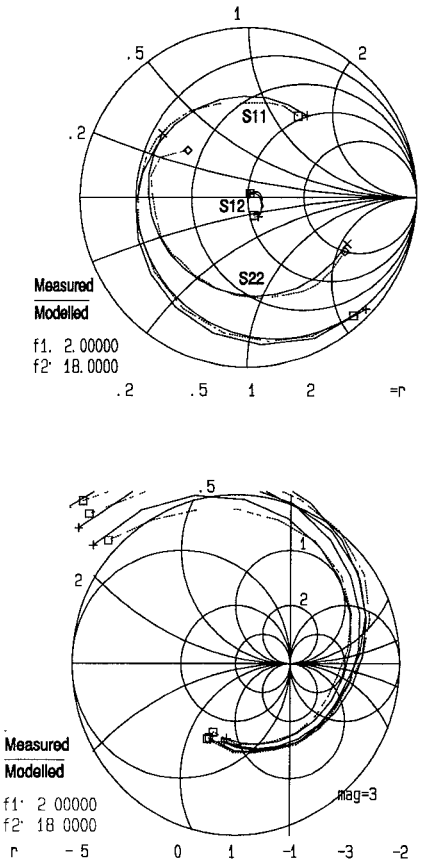


Fig. 5. S-Parameter Fits for Sample NEC NE71083 Device: ($V_{gs} = 0.0, -0.4, -0.8$ V; $V_{ds} = 5.0$ V (Only $V_{gs} = -0.4$ V shown for S11, S12, S22).

VIII. Conclusions

The degree of ill-conditioning in the small signal MESFET equivalent circuit model has been formally quantified using a systematically formulated sensitivity analysis procedure. The condition number is typically very large since the error function most often used is sensitive to changes in some combinations of model parameters but extremely insensitive to changes in other combinations. Consequently there is a good deal of uncertainty associated with the optimized values of the insensitive components, and the convergence of the optimization is slow. Using the principal components sensitivity analysis we have been able to estimate for the first time how reliable the optimized component values are in the final solution.

A new optimization technique has been presented which eliminates ill-conditioning thus ensuring rapid convergence and accurate solutions. By performing a coordinate transformation of the model variables to the principal component axes which are uncorrelated, the new principal component variables can be automatically scaled providing uniform sensitivity for all variables, i.e. a condition number of unity. For extremely ill-conditioned models numerical rounding and truncation errors reduce the effectiveness of this scaling technique. Such models can be systematically partitioned into smaller subspaces of the transformed model space on the basis of the sensitivities of the uncorrelated principal component variables. Separate optimization of sensitive and insensitive sub-spaces provides a marked improvement in convergence and a reduction in the uncertainty associated with the values of insensitive model components such as the parasitic resistances.

If the values of the intrinsic equivalent circuit components are derived from optimizable technological device parameters over multiple bias points the condition number is significantly improved. Consequently the convergence of the numerical routines used is also improved and the solution is resolved with more precision. Since the intrinsic components of the equivalent circuit model are now constrained to obey the device physics the model is unable to fit the measured data so closely. It does, however, preserve the inherent correlation between the model components and provides more physically representative values for the equivalent circuit elements than could otherwise be achieved.

Unique and physically representative small signal models have been determined for a Plessey $4 \times 75\,\mu$m MESFET and an NEC NE71083 packaged MESFET using the new optimization strategy. For both devices the convergence of the new optimization technique was significantly faster than conventional techniques. In addition insensitive parameters such as the parasitic resistances were determined more uniquely since the sensitivity of the error function to these parameters was enhanced by automatic scaling. These observations are particularly pertinent to the packaged NE71083 device since the model is more complex and more ill-conditioned than the on-wafer device model. Consequently it is very difficult to obtain a physically representative solution for the packaged device model using conventional optimization strategies.

Acknowledgment

Dr. D. M. Brookbanks for his helpful discussions and GEC Marconi Materials Technology Ltd for providing RF on-wafer chip devices.

References

[1] J. W. Bandler and S. H. Chen, "Circuit optimization: The state of the art," *IEEE Trans. Microwave Theory Tech.*, vol. 36, pp. 424–444, Feb. 1988.

[2] G. Dambrine, A. Cappy, F. Heliodore, and E. Playez, "A new method for determining the FET small-signal equivalent circuit," *IEEE Trans. Microwave Theory Tech.*, vol. 36, pp. 1151–1159, July 1988.

[3] J. W. Bandler, S. H. Chen, and S. Daijavad, "Microwave device modelling using efficient L_1 optimization: A novel approach," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-34, pp. 1282–1293, Dec. 1986.

[4] J. W. Bandler, S. H. Chen, S. Ye, and Q. J. Zhang, "Robust model parameter extraction using large scale optimization techniques," *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 319–322, New York, 1988.

[5] W. R. Curtice and R. L. Camisa, "Self-consistent GaAs FET models for amplifier design and device diagnostics," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-32, pp. 1573–1578, Dec. 1984.

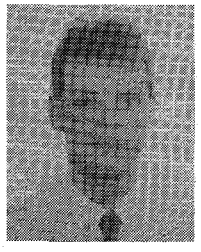
[6] R. L. Vaitkus, "Uncertainty in the values of GaAs MESFET equivalent circuit elements extracted from measured two-port S-Parameters," *IEEE Cornell Conf. on High Speed Semiconductor Dev. and Ccts.*, Cornell Univ., Ithaca, NY, Aug. 1983.

[7] G. L. Bilbro, M. B. Steer, R. J. Trew, C. R. Chang, and S. G. Skaggs, "Extraction of the parameters of equivalent circuits of microwave transistors using tree annealing," *IEEE Trans. Microwave Theory Tech.*, vol. 38, pp. 1711–1718, Nov. 1990.

[8] T. Chen and M. Kumar, "Novel GaAs FET modelling technique for MMICs," *GaAs Symp. Tech. Dig.*, pp. 49–52, Nashville, TN, Nov. 1988.

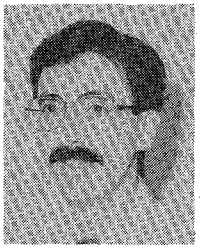
[9] H. Kondoh, "An accurate FET modelling from measured S-parameters," *IEEE MTT-S Int. Microwave Symp Dig.*, pp. 377–380, June 1986.

[10] A. D. Patterson, V. F. Fusco, J. J. McKeown, and J. A. C. Stewart, "Problems associated with small-signal MESFET equivalent circuit model parameter estimation," *14th A.R.M.M.S. Conf. Dig.,* Queen's University Belfast, Mar. 1991.
[11] ——, "Active microwave device modelling: A transfer function approach," SERC Contract Report GR/F33605.
[12] "HP8510 Specification," *HP8510B Network Analyzer System Manual,* Hewlett Packard Company, Santa Rosa, CA, 1987.
[13] P. A. Linden, Ph.D. dissertation, Department of Electrical and Electronic Engineering, The Queen's University of Belfast, Northern Ireland, 1991.
[14] P. H. Ladbrooke, *MMIC Design: GaAs FETs and HEMTs.* Norwood, MA: Artech House, 1989, pp. 91–203.
[15] S. M. Sze, *Physics of Semiconductor Devices.* New York: Wiley, 1981.
[16] J. W. Bandler, R. M. Biernacki, S. H. Chen, J. Song, S. Ye, and Q. J. Zhang, "Statistical modelling of GaAs MESFETs," *IEEE MTT-S Int. Microwave Symp. Dig.,* pp. 87–90, Boston, June 1991.
[17] A. D. Patterson, V. F. Fusco, J. J. McKeown, and J. A. C. Stewart, "Microwave device modelling using systematic optimization techniques," *IEE Colloquium on Computer Based Tools for Microwave Engineers Dig.,* Oct. 1991.
[18] J. J. McKeown, "Sensitivity analysis with respect to independent variables," in *Nonlinear Optimization - Theory and Algorithms,* L. C. W. Dixon, E. Spedicato, G. P. Szego, Birkhauser, 1980.
[19] J. J. McKeown, D. Meegan, D. Sprevak, *An Introduction to Unconstrained Optimisation.* Cambridge: Adam Hilger, ESM, 1990.
[20] *The GaAs Handbook,* 2nd ed., Towcester, Northamptonshire, UK: Plessey Three-Five Group Limited, 1989.
[21] *NEC Microwave and RF Semiconductors,* NEC Electronics (Europe) GmbH, Microwave Dept., Dusseldorf, W. Germany, 1990.
[22] G. F. Engen, and C. A. Hoer, "Thru-reflect-line; an improved technique for calibrating the dual six-port automatic network analyzer," *IEEE Trans. Microwave Theory Tech.,* vol. MTT-27, pp. 987–993, December 1979.
[23] S. Lautzenhiser, A. Davidson, and K. Jones, "Improve accuracy of on-wafer tests via L.R.M. calibration," *Microwaves and R.F.,* pp. 105–109, Jan. 1990.



Andrew Patterson was born in Northern Ireland in 1967. He received the B.Eng. degree in electrical and electronic engineering from the Queen's University of Belfast in 1988. In 1992 he completed studies for the degree of Ph.D. in microwave active active device modelling at Queen's University.

He is presently working at EEsof Inc., Westlake, CA.

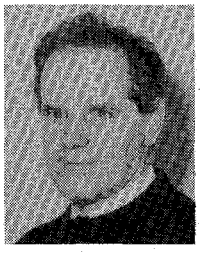


Vincent F. Fusco was educated at the Queen's University of Belfast where he obtained the degree of Ph.D.

He has worked as a research engineer on short range radar and radio telemetry systems. Currently he is a reader in Microwave Communications in the School of Electrical Engineering and Computer Science, The Queen's University of Belfast. His current research interests include nonlinear microwave circuit simulation and concurrent programming techniques for electromagnetic field problems. He has acted as consultant to a number of major companies and has published numerous research papers in these areas. He is author of the book *Microwave Circuits, Analysis and Computer Aided Design,* Prentice Hall, 1987.

Dr. Fusco is a Chartered Electrical Engineer and a Member of the Institute of Electrical Engineers.

J. J. McKeown, photograph and biography not available at the time of publication.



J. A. C. Stewart was born in Belfast, Northern Ireland in March, 1937. He received the B.Sc. degree in physics in 1959 and the Ph.D. degree in 1962, both from The Queen's University of Belfast.

After working in Shorts, Belfast as a Design Engineer and Queen's University, Kingston, Ontario, as a Post Doctoral Fellow, he joined the Department of Electrical and Electronic Engineering at The Queen's University of Belfast, where he is now Professor of Electrical Communications. His research interests include CAD and simulation of active microwave devices and their current interactions.