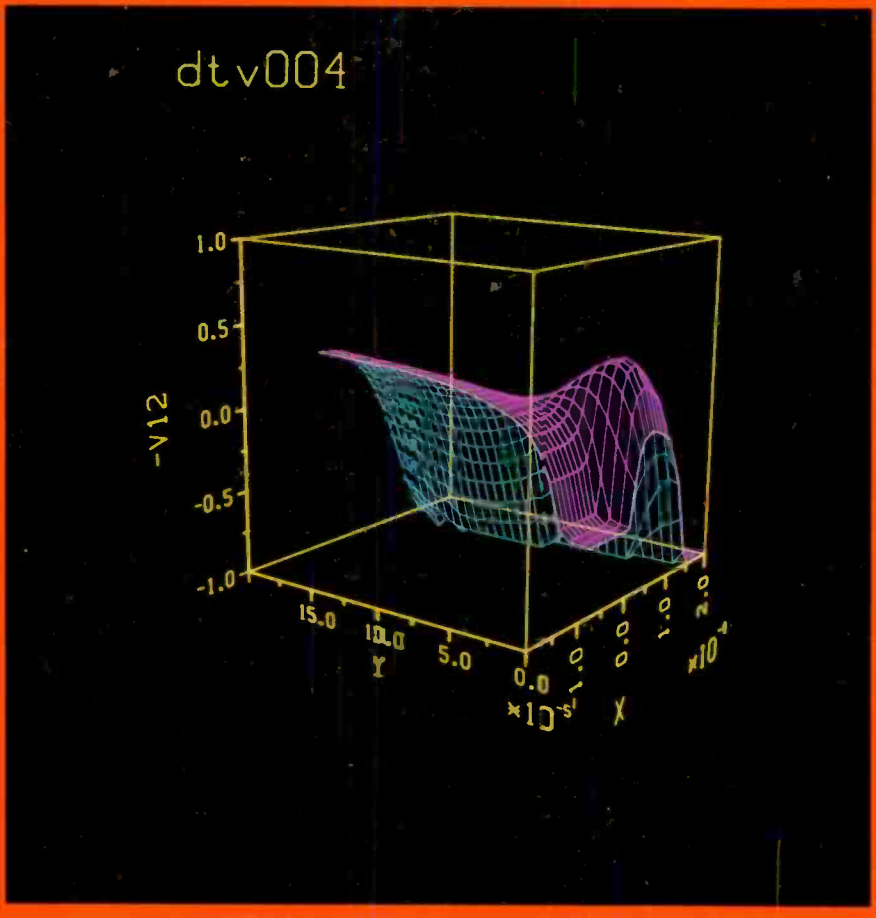
RCA Review



Simulated Plot of Potential Distribution in MOS Transistor

September 1985 Volume 46 No. 3

RCARCI46(3)279–416 (1985)

RCA Review

RCA Review (ISSN 0033-6831) is a technical journal published quarterly by RCA Laboratories in cooperation with the subsidiaries and divisions of RCA.

Contents

September 1985 Volume 46 Number 3 ISSN 0033-6831

An Introduction to Modeling and Simulation

Robert Amantea, Guest Editor
RCA Laboratories, Princeton, NJ 08540

Abstract—This special issue of *RCA Review* deals with some applications of modeling and simulation to electronic devices, circuits, and systems. In this short introduction to the issue, we give a brief discussion of types of models and some considerations in model building and model calibration and validation.

Webster's *New World Dictionary* defines 'model' as a "preliminary representation for something, serving as the plan from which the final object, often larger, is to be constructed." Clearly this definition falls short of our needs, as we are more concerned with relating physical processes to mathematical symbols than we are to constructing objects.

Physical processes are usually characterized by observations of experiments relating physical objects. Real physical processes are so complex that considerable simplification is necessary before a process can be described mathematically. Mathematics deals with the formal manipulation of symbols. Relating these symbols to the physical objects and their manipulation to the physical processes is the concept of modeling that we are seeking.

Thus, by a model we mean a mathematical model and, in particular, a model that is capable of describing quantitatively the essential details of the processes acting on the objects of interest. Furthermore, by a simulation, we mean the results of evaluating a model.

Since the papers in this issue of *RCA Review* are concerned with developing computational tools to deal predictively with real world situations, they deal mainly with simulators. All of these simulators, however, are based on models, or modifications of models, in-

tended to represent devices or circuits. It is important, therefore, that we be clear as to what is meant by the term model. In this introduction, we define the various ways in which this term is used in the literature and briefly discuss model building and model calibration and validation.

Types of Models

We will consider a model to be a set of axioms (i.e., assumptions) from which we may derive, via applied mathematics, the behavior of the system (i.e., physical processes and objects) under consideration. With this in mind, we consider specializations of this concept.

The totality of models may be subdivided in several fashions. We have all heard of qualitative and quantitative models; however, there are several classes of models that have not been adequately described. In the following paragraphs, we will discuss the concepts of qualitative, quantitative, exact, approximate, interpolative, extrapolative, compact, and macro models.

Qualitative and Quantitative Models

A qualitative model for a physical process relates the physical objects involved in the process to one another with words derived from models in the support level of knowledge. In a sense, a qualitative model is a verbal model where the words describing the system behavior are drawn from the relevant field of knowledge. For example, we say that current can flow in a semiconductor device by two mechanisms, drift and diffusion. This statement is a qualitative model for current flow in a semiconductor. The meanings of the words drift and diffusion, however, are rooted in the field of solid-state physics. Specifically, drift relates the movement of a carrier (e.g., an electron or hole) to an applied electric field (i.e., the velocity of the carrier is proportional to the electric field, where the proportionality constant, mobility, is a quantity derived in solid-state physics). Diffusion relates the movement of carriers due to random (i.e., thermal) behavior and nonuniform distribution (i.e., the velocity of the carrier is proportional to the logarithmic derivative of the density of carriers, where the proportionality constant, the diffusion constant, is a quantity derived in solid-state physics).

In contrast to the qualitative model, the quantitative model describes physical processes and objects in terms of mathematical symbols and ultimately in terms of quantitative experiments that can be performed on the system. For example, the Poisson-Shockley

equation models the behavior of a depletion region in a semiconductor. The form of this equation is based on electrostatic theory (i.e., Poisson's equation) and the characterization of carriers in a semiconductor (i.e., the density of carriers is proportional to an exponential factor of the potential) which is derived in solid-state physics. The solutions to this equation provide us with a way of characterizing the capacitance of the device with respect to a voltage applied to its terminals. Note that capacitance is something that is used in describing device electronics, and thus it is a model in a higher level of abstraction, i.e., circuit analysis.

Exact and Approximate Models

Models may also be subdivided into the two classes, exact and approximate. However, we should always keep in mind that models are always approximations. A person speaking of an exact model is referring to models that contain descriptions of all known applicable physical effects that are relevant to the object in question. More precisely, a person speaking of an exact model (or solution) really means that all known physical effects relevant to the problem under consideration have been included and that the mathematical equations describing the problem have been solved in as exact a way as possible. In the case where analytic solutions are not possible, this often means that the numerical solution has been made very accurate.

There is no general method of measuring the exactness of a model. In the case of interpolative models, there are several possible criteria based upon the relative deviation of the data from the fit, e.g., maximum error, root-mean-square (rms) error, weighted rms error, etc. These same criteria can be applied to extrapolative models; however, they say nothing about the error outside of the range in which there is data.

Interpolative and Extrapolative Models

An interpolative model (sometimes referred to as an empirical model) is generated by assuming that some functional form describes experimental data and then fitting parameters into the functional form by minimizing the error between the experimental data and simulated data over the range of the experimental data. Clearly, there is no reason why this functional form should fit the data outside the range of the experimental data.

In contrast to the interpolative model, the extrapolative model is strongly based on the physics of the device. Model parameters are fitted in a similar manner, but because of the physical basis of the model, the parameters have strong physical meaning and can be considered to hold true outside the range of experimental data. In this case the model can be used outside the experimentally determined range with considerable confidence. Therefore, an extrapolative model is very useful in predicting behavior outside the range of known behavior or in determining the behavior of devices not yet fabricated.

Compact and Macro Models

A compact model can be thought of as a model that has a small computational cost. For example, an analytic model (i.e., a closed form solution) is usually easy to compute for either a human or a computer. On the other hand, models requiring the integration of partial differential equations or even ordinary differential equations are difficult to compute. In general, compactness is a relative term relating one model to another. That is, if one model is much more easily computed than a second so that it may be imbedded in the second without an appreciable increase in the cost of computing the second, then the first model is compact with respect to the second. For example, suppose we have an electronic device that is modeled by an ordinary differential equation. And furthermore, we would like to simulate the behavior of the device in an electronic circuit by using the device model in a circuit simulation program. Clearly, if we have to numerically solve the ordinary differential equation describing the device for each iteration of the circuit analysis program, the computational cost will rise significantly. However, if we use a closed-form solution of the differential equation in the circuit analysis program, then the computational cost increases only marginally.

A macro model is similar to a compact model. Generally, this concept is used when describing the joint behavior of a collection of devices. For example, replacing a set of electronic components that form an operational amplifier by a compact model of an operational amplifier will greatly reduce the number of parameters and the time required in a circuit analysis of a system of interconnected operational amplifiers. The purpose of the macro, like that of the compact, model is to reduce computational costs.

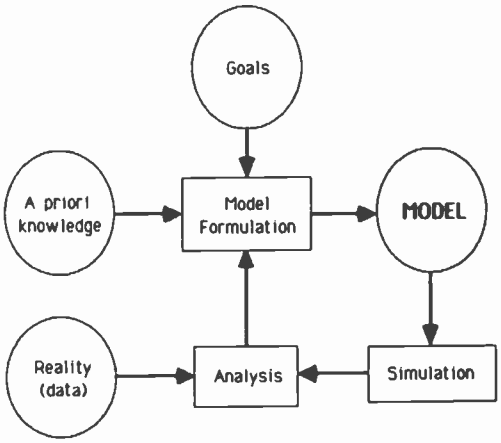


Fig. 1—Model formulation.

Analytic and Numerical Models

An analytic model is a model whose solution can be represented as a closed form mathematical expression. In fact, our goal in deriving a compact model is to obtain an analytic model. However, sometimes the relationships in the analytic model are implicit in nature and an analytic inversion does not exist. In this case the model is still compact, because analytic expressions can usually be inverted without much numerical complexity. However, the actual model becomes numerical.

A numerical model is one in which the solution to the model equations must be obtained numerically (e.g., by numerical computation). In this case the model is usually referred to as a simulator.

Model Building

Fig. 1 shows an overview of the modeling process. The inputs to the model building process are purposes, data, and a priori knowledge. The purposes or goals of the model usually determine the necessary details (i.e., the physical mechanisms) that must be included in the model. For example, if the model is to be used to introduce a student to the behavior of a device, then only the crudest model containing only dominant physical effects is necessary. Alternatively, if the model is to be used for obtaining device limitations and possible failure modes, then a much more complete model is necessary, often requiring all known applicable physical mechanisms. Analysis is the joint process of validation and calibration.

Goals

Summarizing, some of the purposes of a model are:

- Organize theoretical and empirical knowledge of deduce implications.
- Identify critical factors in the operation of a device.
- Identify possible areas for improved performance (or decreased cost).
- Identify methods for testing.
- Isolate deficiencies in devices that cause unexpected subpar performance.
- Allow changes in more input parameters than the experimental approach would allow.
- Produce an optimal design with minimum sensitivities to device parameters.
- Produce simulations to check new designs.
- Generate compact models for use in higher levels of modeling.

Constraints

Some reservations that must be kept in mind when developing models are:

- The model developed may not sufficiently describe reality and further development may be prohibitively costly.
- The range of applicability of the model may be severely limited.
- The model may not have all of the useful features expected.

A Priori Knowledge

A priori knowledge helps us to formulate the fundamental equations to be used for the model Often, these equations are very complex partial differential equations (or their equivalent in complexity) that cannot be solved in closed form. In cases like these, we use a priori knowledge to aid the simplification process to obtain equations that may be solvable. Alternatively, we may numerically solve the equations to gain new knowledge which, in turn, may help us to simplify the equations for further study. A priori knowledge can be found in basic physical principles, physical conservation laws, prior models, and qualitative reasoning.

Data

Data is what generally provides us with insight into the dominant physical mechanisms. Data may be measured or be the result of a simulation of a more exact model. In either case, by studying the data we can infer from various simplified models what are the relevant physical mechanisms that must be included in our final model to be consistent with the purposes of the modeling process.

Model calibration and validation

After we have successfully formulated a model we often find that the model contains one or more undetermined parameters. It would be most fortunate if there were no undetermined parameters, that is, all parameters could be directly related to and computed from fundamental physical quantities. However, in the process of solving the model, we are often forced to make simplifying approximations that break the direct link between the model parameters and the physical objects or quantities that they represent. In cases like these we must use model calibration and validation procedures.

Model calibration is the process of extracting model parameter values by comparing predicted data from the model to experimentally measured data. The usual practice is to fit the measured data to the model by a least-squares method (i.e., we minimize the sum of the square deviations between the measured data and the simulated data). In many cases this parameter extraction process is strongly nonlinear and numerical methods must be used.

Model validation is the process of verifying that the mathematical model truly represents the process being modeled. The most widely accepted procedure is to show that measured and simulated data agree very well. This procedure works fine for interpolative models but it tells us nothing about the validity of the model outside the region of measured data. Thus for extrapolative models, we must rely on the strong physical basis of the model for confidence. We can raise our confidence in a model by minimizing the number of undetermined parameters or by determining the parameters by some other unrelated method. Alternatively, we can use a noncompact more exact model to generate simulated data outside the range of measured data. In this case if we have strong confidence in the more exact model and, if the simulated data agree well with the compact model's predictions, we can transfer our confidence in the more exact model to the model under consideration.

Summary

The definitions presented in this introduction need not be rigidly applied in all circumstances, but they do provide a means to accurately represent the efforts of the modeler and the achievements of the model. The key issues, regardless of nomenclature, are to know your goals and constraints before embarking on a modeling project and to provide convincing validation at its conclusion. Otherwise, the probability of achieving a successful model, one that others can use with confidence, is very small.

Bibliography

1. J. A. Spriet and G. C. Vansteenkiste, *Computer-Aided Modelling and Simulation*, Academic Press (1982).
2. T. L. Saaty and J. M. Alexander, *Thinking With Models*, Pergamon Press (1981).
3. Z. A. Melzak, *Companion to Concrete Mathematics*, John Wiley and Sons, Inc. (1973).
4. Z. A. Melzak, *Mathematical Ideas, Modeling, and Applications*, John Wiley and Sons, Inc. (1975).
5. F. Y. M. Wan, "Mathematical Models and Their Formulation," in *Handbook of Applied Mathematics*, ed. by C. E. Pearson, VanNostrand Reinhold Co. (1983).

Integrated Simulation of CMOS Transistors

Robert Amantea and Beatrice Hwong
RCA Laboratories, Princeton, NJ 08540

Abstract—A library of computer programs has been developed capable of simulating the physical and electrical properties of semiconductor devices from basic processing through circuit design. Both one- and two-dimensional behaviors can be characterized. This paper describes these programs and illustrates the manner in which they have been coupled to maximize user friendliness and to achieve comprehensive simulation of circuits using a 2-μm CMOS technology.

1. Introduction

Modern VLSI design and fabrication is an extraordinarily complex effort that encompasses diverse engineering tasks such as process and device design, circuit design, photomask layout and fabrication, design verification, and chip testing. Each of these tasks is sufficiently complex in itself to require computer support.

Recently, we have acquired or developed a set of computer programs that perform VLSI process, device, and circuit simulation. Such tools give users great freedom in testing design changes computationally rather than performing tests with actual circuits. They provide a way to interpret pathological behavior of devices when experimental techniques would be too costly or time consuming. They also allow circuit simulations to be performed before the silicon chips are available, so that the entire design effort can proceed in parallel with technology development.

This paper describes the principles, procedures, and results obtained for complete two-dimensional simulation of CMOS transistors (both n and p types) from process structure through circuit behavior. The process simulation program (SUPREM[1,6]), the MOS

device simulation program (MINIMOS[2]), the MOS model parameter extraction program (MOSFIT[3]), and the circuit simulation program (R-CAP[4]) are all used with the computer-aided modeling system (Model-Graph[5]). Model-Graph pipes data between programs and generates outputs in graphic form. As an example, in this paper, we completely simulate both CMOSIII n-channel and p-channel transistors.

Fig. 1 is a block diagram showing how these software tools are integrated. The boxes labeled 'transfer', 'prepare', and 'graph' in the shaded area represent some of the functions in the Model-Graph system. The graphics data base is the repository of all the data used by the system.

SUPREM input files are written by interpreting the process specification in terms of the SUPREM process models. This program generates two sets of impurity profiles, one for the source-drain region and one for the channel region, that are subsequently transferred into the Model-Graph data base. These data together with

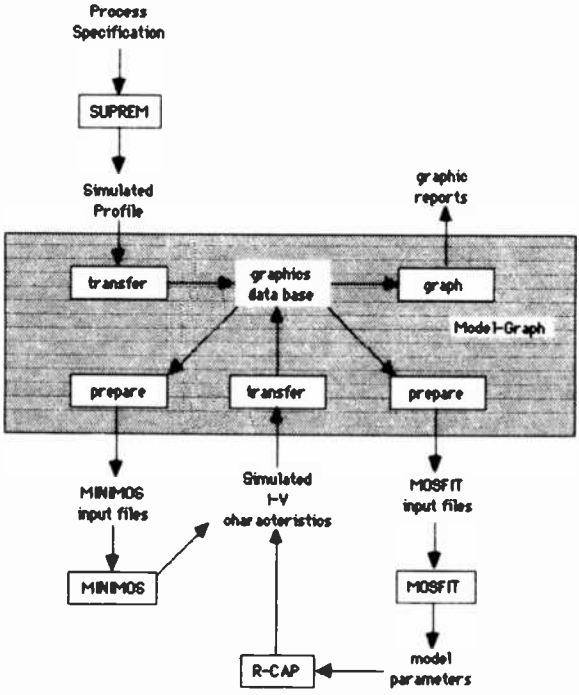


Fig. 1—Block diagram showing integration of simulation software using Model Graph.

device and bias specifications, are then used to generate the MINIMOS input files. The output of MINIMOS, *I-V* characteristics and distributions (such as potential, field, etc.), are again transferred into the Model-Graph data base. The *I-V* characteristics are then used to generate the MOSFIT input files. MOSFIT extracts compact-model parameters for R-CAP and writes a file in R-CAP syntax for insertion into an R-CAP input file for circuit simulation.

2. Simulator Background

2.1 SUPREM

SUPREM[1,6] is a process simulation program capable of simulating a large variety of semiconductor device processes. Among its capabilities are

(1) up to ten material layers;
(2) process models for oxidation, diffusion, epitaxy, and ion implantation;
(3) internal graphics and interactive graphics (via Model-Graph);
(4) a descriptive input language; and
(5) electrical models for sheet resistance, threshold voltage, and net layer charge.

SUPREM can store the simulated one-dimensional impurity profile in several different fashions, including graphical information. One of these storage modes, the 'device' file, can be read by the Model-Graph 'transfer' function to move the data into a Model-Graph graphics data base. In particular, if the MINIMOS application is selected during the transfer, the data will be stored in the form required by the Model-Graph 'prepare' function for MINIMOS, thereby integrating SUPREM and MINIMOS.

2.2 MINIMOS

MINIMOS solves the basic semiconductor device equations: Poisson's equation, the current continuity equations, and the current-transport equations under the following assumptions:

(1) homogeneity of permittivity (the oxide and semiconductor dielectrics are constants),
(2) total ionization of all impurities,
(3) no bandgap narrowing, and
(4) homogeneous temperature distribution.

The program includes a sophisticated model to calculate the potential distribution, impurity concentration, carrier concentration, and carrier mobility in two dimensions. Based upon these quantities, we can process the MINIMOS output data to generate the electric field, the potential barriers, and other quantities of interest.

MINIMOS requires the bias point or set of bias points at which the currents are to be computed (i.e., V_{GS}, V_{DS}, and V_{BS}). It also requires gate-oxide thickness, gate length, and the electrically active impurity profile over the whole device.

The impurity profile can be simulated using the approximate process simulator inside of MINIMOS or it can be specified as two profile slices (one through the source-drain region and one through the channel region). From these profiles, a two-dimensional impurity profile is generated via an elliptic rotation of the source-drain impurity profile. The numerical impurity profiles can be obtained using measured data or through simulation with SUPREM. If measured data is used, the numerical profile is smoothed by fitting the profile with several Gaussian terms and then computing the smooth approximate profile using the Gaussian terms and the parameters that characterize them.

The input files for MINIMOS can be created automatically with the Model-Graph system if the two profile slices are in a graphics data base and the bias data are specified in a Model-Graph 'VRBLS' file, the standard Model-Graph input-file format.

MINIMOS can run interactively under IBM's CMS (Conversational Monitor System) operating system, on a batch CMS machine, or on the FPS (Floating Point Systems Co.) attached array processor. The preferred (and cheapest) method is to use the attached FPS array processor. When used in this manner, the input files and output files are moved automatically from CMS to the FPS and vice-versa by invoking a command file known as an EXEC procedure. The output data can be transferred to a Model-Graph graphics data base for further processing (such as graphics).

2.3 MOSFIT

MOSFIT is a parameter extraction program for obtaining the circuit model parameters from *I-V* data for the R-CAP MOS model. The source of data can be either measured or simulated I-V characteristics. This model is based upon the following assumptions:

(1) In the linear-regime of operation, device analysis is split into two one-dimensional problems: Gauss' Law into the bulk and current continuity along the channel.

(2) The Si/SiO_2 interface is fully inverted.
(3) The effects of nonuniform channel doping are modeled by adding higher order terms to the threshold voltage model.
(4) The carrier mobility decreases with both tangential and perpendicular fields.
(5) Diffusion currents are ignored.
(6) In the saturation-regime of operation, the channel is divided into two regions; adjacent to the source where the charge-sheet approximation holds and adjacent to the drain where the current abruptly fans out. The regions are joined at a critical field.

MOSFIT uses a robust global nonlinear least-squares algorithm based on the Levenberg-Marquardt algorithm in its derivative-free form to extract model parameters from the data. The program requires three input files that can be generated automatically via the Model-Graph system. Program output consists of the set of model parameters and a computed set of I-V characteristics based upon the MOS circuit model.

2.4 MODEL-GRAPH

Model-Graph is a system of programs serving a central file termed the graphics data base. The system is implemented on CMS and makes extensive use of CMS EXEC procedures to check for necessary programs, data sets, and system requirements. Model-Graph can be thought of as the glue that sticks the various simulation programs together. It provides functions for data translation, high-level graphics, and data transformation.

3. Required Data

Complete simulation of devices requires specification of the geometry, the process, and the bias points of interest. The limited capabilities of the simulation programs restricts the range and type of data that can be specified. The following sections outline these limitations.

3.1 Geometry

The typical geometry for the simulated device is shown in Fig. 2. The gate length is specified, but the actual channel length is determined by the effective sidewall diffusion of the source and drain as modeled by an elliptical rotation of the source-drain impurity pro-

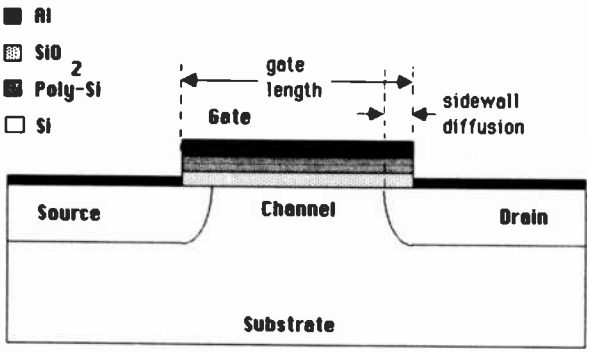


Fig. 2—CMOS device cross section showing simplified geometry imposed by limitations of MINIMOS.

file. This channel-length shrinkage can be calibrated by specifying the ellipticity of the profile rotation (e.g., 1 implies a circular junction, 0.7 is the default value).

It is important to note that the geometry of the MOS device is limited by MINIMOS. In particular, the following limitations apply:

(1) The edge of the gate is assumed to line up exactly with the edge of the source and drain openings. All gate overlap is due to the sidewall diffusion of the source and drain regions.
(2) Extended drain devices cannot be simulated because of the above restriction.
(3) The edge of the source and drain metal is assumed to line up with the edge of the source and drain openings. This results in very small series resistances, which does not reflect real devices. Consequently, the effects of series resistance must be included later.
(4) The impurity profile is assumed uniform below the deepest concentration specified in the channel region.
(5) Narrow-width effects are not simulated. The simulation behaves as if the device were infinitely wide.
(6) Multi-layer insulators cannot be simulated.

3.2 Impurity Profile

The SUPREM input file is created by mapping the process specification into the SUPREM input language. Except for selecting pro-

cess-model parameters (e.g., diffusion constants, oxide growth rates, etc.) and for selecting the numerical grid for simulation, this procedure is straightforward. The selection of the model parameters is made through a procedure called calibration, as discussed in the next section.

3.3 Bias Points

The selection of bias points affects the accuracy of the complete simulation in several ways. If the steps between the successive bias points are too big (i.e., the difference between successive gate or drain voltages), the program may not be able to accurately converge to the correct solution or the automatic regriding operation in MINIMOS may produce a grid that leads to a poor rate of convergence. If the range of bias points is not comprehensive enough, the ability of MOSFIT to extract compact-model parameters will be severely limited often leading to generating the wrong shape for the I-V characteristics (e.g., the I-V knee may be too soft).

Currently, we use 1-volt steps for the gate and drain voltage and 2-volt steps for the body voltage. We have found that this is a fair compromise between accuracy and cost. However, if it is found that the compact-model does not fit the knee region of the characteristics well, a denser grid of bias points should be selected.

4. Simulator Calibration

Before simulators can be used as design tools, the many parameters that arise in the models used by the simulators must be specified. The process of specifying these values is termed calibration. Calibration is achieved by comparing empirically derived characteristics of the processes and devices to those simulated. Where discrepancies are significant, the simulated parameters are adjusted as necessary to make the results agree.

Experiments and simulations concerning oxide growths, implantations, and diffusions were made to calibrate SUPREM. Comparisons of plots generated by the Model-Graph system with experiment, were used to adjust process–model parameters.* The fol-

* A complete set of the necessary model parameters and the appropriate model choices were provided by K. G. Amberiadis.

lowing table summarizes the necessary model choices and parameter changes required to match our processing capabilities.

Oxidation

(1) Oxidation in 100% steam requires a change from default model parameters for temperatures less than 900°C.
(2) Oxidation in oxygen with 10% steam requires the use of the dry O_2 model with minor adjustments to the model parameters.

Implantation

(1) The best fit to experimental data is achieved using the two-sided Gaussian model and the Pearson model.

Diffusion

(1) High dose (source/drain) phosphorus, short time diffusions can use the default parameters.
(2) High dose (source/drain) boron, short time diffusions require modified parameters.
(3) Low dose (well) phosphorus and boron, long time diffusions require modified parameters.

MINIMOS calibration involves the adjustment of the following parameters (default values are in parentheses):

- slow surface-state density (0)
- carrier lifetime (1 μsec.)
- mobility model parameters
- avalanche ionization parameters ($a \exp(-b/E)$)
- surface recombination velocity (100 cm/sec.)
- profile elliptic-rotation parameter (0.7)
- bulk resistor (0 Ω) $V_{BSeff} = V_{BS} - RI_B$

In all cases above except for mobility, avalanche ionization, and rotation, the parameters have little or no effect on performance unless they are outside their usually acceptable range or the simulator is used to explore parasitic behavior (e.g., leakage currents and punchthrough). The elliptic rotation parameter is adjusted so that the final channel length corresponds to that deduced from electrical measurements. The mobility model has been heavily calibrated by the MINIMOS developers and should need no further adjustment. Finally, the default avalanche ionization parameters fail to adequately describe the behavior of our devices, since calibration of these conditions has not yet been performed.

Since the performance of MINIMOS rests very heavily on the calibration of SUPREM, we have placed most of our effort in obtaining good process-model parameters.

5. Simulation Results

5.1 SUPREM

The CMOSIII process was simulated with SUPREM for four different regions:

(1) NMOS Channel
(2) NMOS Source-Drain
(3) PMOS Channel
(4) PMOS Source-Drain

We found during calibration that the diffusion-model parameters must be different depending upon dose and time, and therefore different parameters were chosen for the different regions according to the following list expressed as SUPREM statements:

For regions 1, 2, and 3,
BORON SILICON DIX.0 = 2.22E8 DIP.0 = 4.32E9.

For region 4,
BORON SILICON DIX.0 = 5.6E9 DIP.0 = 6.5E9.

All the regions used the following oxidation coefficients:
WET02 L.LIN.0 = 3.6E4 L.LIN.E = 1.64 <100>.

The simulated impurity profiles in silicon are shown in Figs. 3 and 4.

5.2 MINIMOS

MINIMOS first generates a two-dimensional impurity profile from the two supplied one-dimensional profiles by elliptically rotating the source-drain profile. The results of this rotation are shown in Fig. 5 for the n-channel device and in Fig. 6 for the p-channel device. The source-channel junction at the surface is located at $(x,y) = (0,0)$, which is at the rear of the left face of the rectangular solid. Notice the junction formed by the boron threshold-adjust implant below the gate in the p-channel device. This junction can give rise to a buried channel, which can be a source of leakage current. This behavior will be investigated in the section describing potential distributions.

Model-Graph can be used to explore the various features of this profile by rotating, scaling, or slicing the data. For example, Fig. 7 shows a profile slice along the surface ($y = 0$) of the silicon for the n-channel device. We obtain the total amount of sidewall diffusion

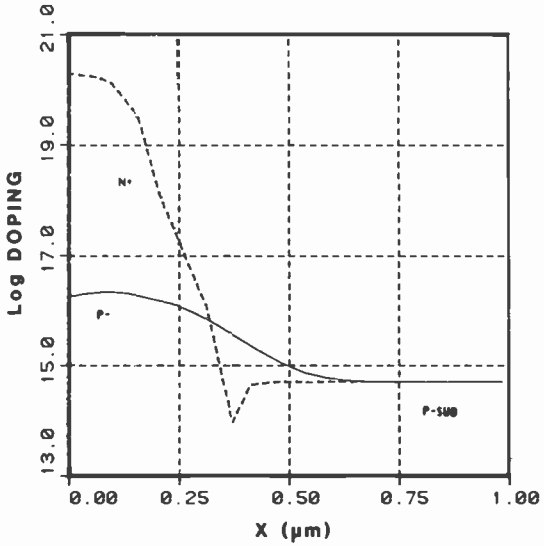


Fig. 3—CMOS n-channel impurity profiles. The solid line is for the channel region, the dashed for the source-drain region.

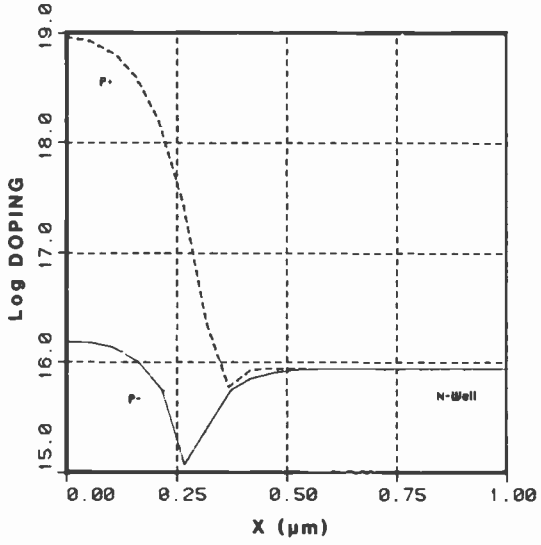


Fig. 4—CMOS p-channel impurity profiles. The solid line is for the channel region and the dashed for the source-drain region.

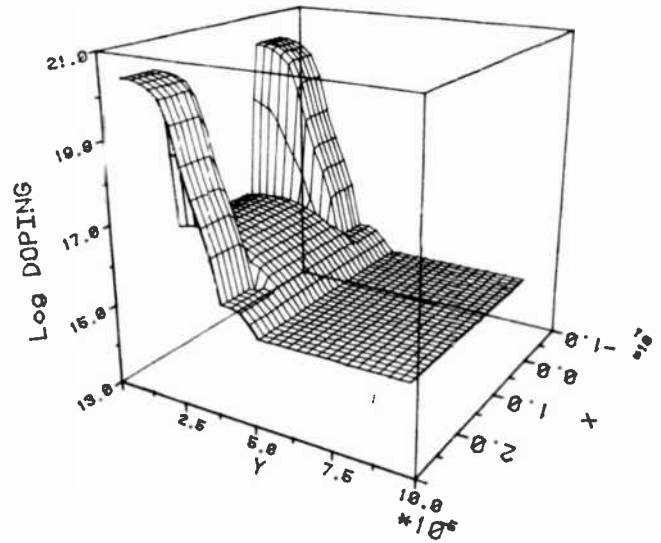


Fig. 5—CMOS n-channel impurity profile. The multiplying factor on the *y* axis is 10^{-5} cm and on the *x* axis is 10^{-4} cm.

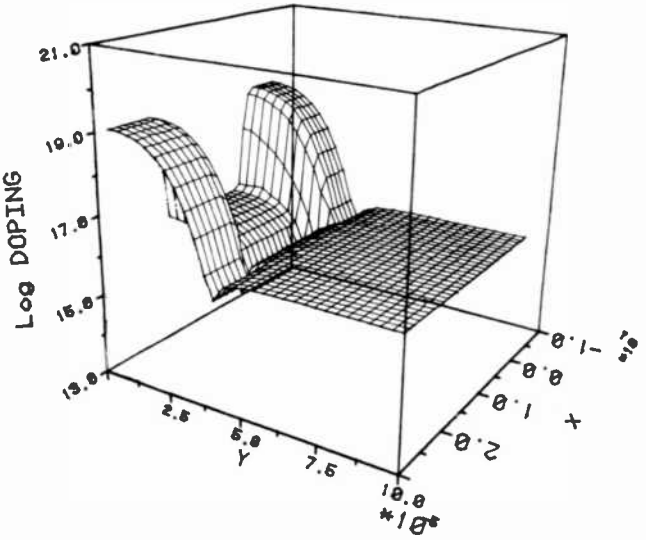


Fig. 6—CMOS p-channel impurity profile. The multiplying factor on the *y* axis is 10^{-5} cm and on the *x* axis is 10^{-4} cm.

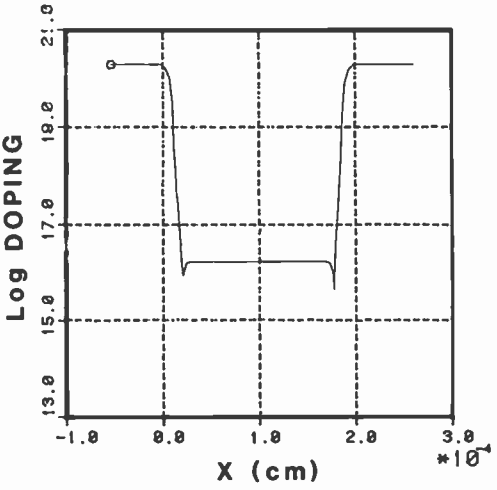


Fig. 7—CMOS n-channel impurity profile along the silicon surface.

from this plot by measuring the distance from $x = 0$ to the junction and then doubling it for $\Delta\ell$, specifically $\Delta\ell \simeq 0.44$ μm. Fig. 8 shows $\Delta\ell \simeq 0.56$ μm for the p-channel device.

The grid distribution used by MINIMOS can be obtained by plotting points instead of a solid curve. For example, Fig. 9 shows the impurity profile along the surface plotted with points. Each point

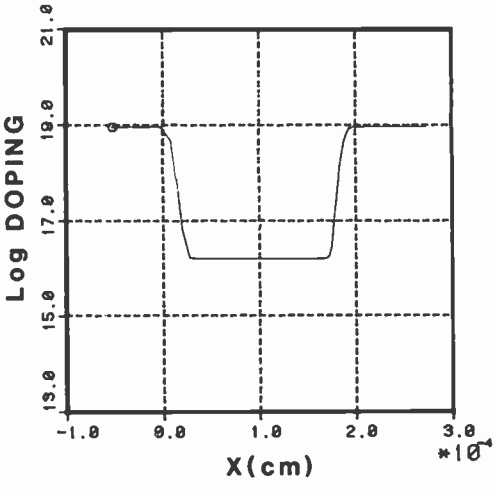


Fig. 8—CMOS p-channel impurity profile along the silicon surface.

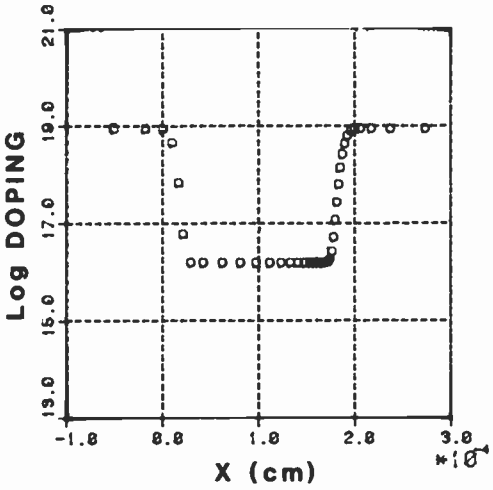


Fig. 9—CMOS p-channel impurity profile along the silicon surface plotted with points generated by MINIMOS. Each point represents the location of one vertical grid line.

represents the location of one vertical grid line. A similar plot of the profile going down into the silicon would display the locations of the horizontal grid lines. This information is sometimes necessary to check proper convergence of the MINIMOS algorithms. Dense grid lines are expected where data is changing rapidly, such as around the junctions. If this behavior is not observed, the MINIMOS results are suspect. In this particular case the grid corresponds to the bias $V_{GS} = -5$ and $V_{DS} = -5$. Note the increased grid density adjacent to the drain-channel junction; this is indicative of the high field and thus rapid change of potential in that region.

Fig. 10 shows a surface plot of -1 times the potential for the n-channel device. This is equivalent to plotting the conduction band for the structure. In this view, the source is the right front corner and the drain is the right back corner. As shown, the applied source and body voltages are 0; therefore, the potential difference shown between the source and the body (left-front) is due to the built-in potential difference. Electrons flow from the source to the drain by surmounting the potential barrier between them. In this picture the minimum potential barrier is near the surface. As the gate voltage is increased, the surface potential is pulled down, thus reducing the potential barrier between the source and the drain and increasing the drain current.

Fig. 11 is another surface plot of the potential showing the effects

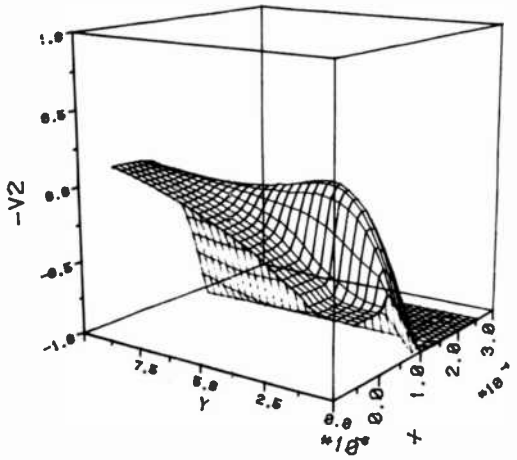


Fig. 10—CMOS n-channel conduction band at low gate voltage (volts). The multiplying factor on the *y* axis is 10^{-5} cm and on the *x* axis is 10^{-4} cm.

of the increased gate voltage. Not only is the surface barrier reduced, but the body barrier (the second minimum in the potential barrier located roughly 0.4 μm below the surface) is also reduced. In some cases this body barrier is reduced more rapidly than the

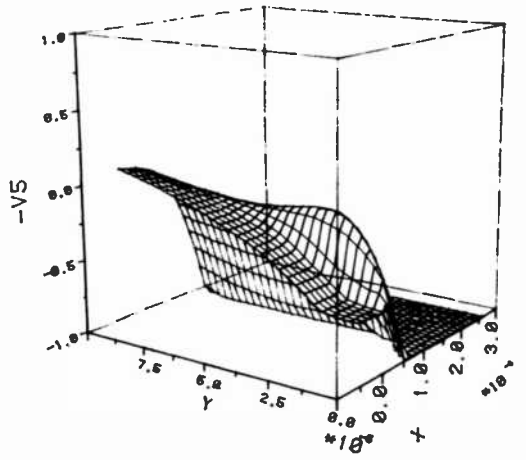


Fig. 11—CMOS n-channel conduction band at high gate voltage (volts). Note the decrease in both the surface and body potential minimum. The multiplying factor on the *y* axis is 10^{-5} cm and on the *x* axis is 10^{-4} cm.

surface barrier and the device is said to be in punch-through. This pathological case can occur when the channel region is doped very lightly, the gate length is very short, and the drain voltage is large. MINIMOS simulations can be used very effectively in exploring the sensitivity of this phenomenon to the process, in particular to the substrate doping, threshold-adjust implant, and possibly a punch-through prevention implant.

5.3 R-CAP Model Parameter Extraction With MOSFIT

MINIMOS generates the *I-V* characteristics for the device as if there were zero or very small series resistances between the device contacts and the active region due to program assumptions about the cross-sectional structure as shown in Fig. 2. In reality, there is a significant distance between the metal contact and the source-channel junction which leads to sufficient resistance to significantly alter the *I-V* characteristics. Because of this, *I-V* characteristics derived directly from MINIMOS cannot be compared with experimental data.

Fig. 12 shows the *I-V* characteristics as obtained from MINIMOS

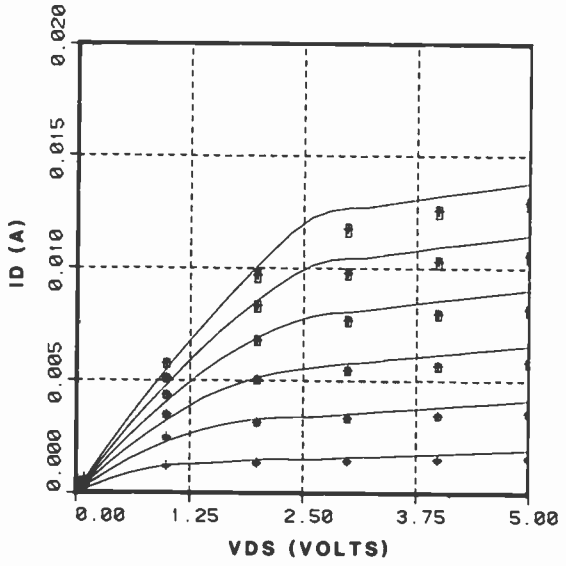


Fig. 12—CMOS n-channel simulated (curves) versus measured (boxes) *I-V* characteristics. The top of each box represents maximum value, the bottom minimum, the horizontal bar is the median, and the cross is the average.

through MOSFIT without added series resistance (curves) compared to an accumulation of measured I-V characteristics (boxes). Each box describes several different measurements. The top of the box represents the maximum value measured, the bottom represents the minimum, the horizontal bar represents the median, and the plus sign represents the average. As can be seen, the simulated data exceeds the measured data. At low currents, where the effects of resistance are very small, the discrepancy between measured and simulated data is primarily caused by differences between the actual and simulated threshold voltage. As current increases, the resistive effects reduce the measured data relative to the simulated data, so that both the slope in the linear regime (g_m) and the magnitude of current in the saturation regime are reduced.

We know approximately how much series resistance to add to the device because we know the sheet resistance in the source and drain regions from the SUPREM simulations. We also know the number of squares between the actual metal contacts and the gate edge of the source and drain regions. From these numbers we can compute a first estimate of the resistance.

To imbed the resistance values into the simulated I-V characteristics while keeping the data points on the same bias grid, we have

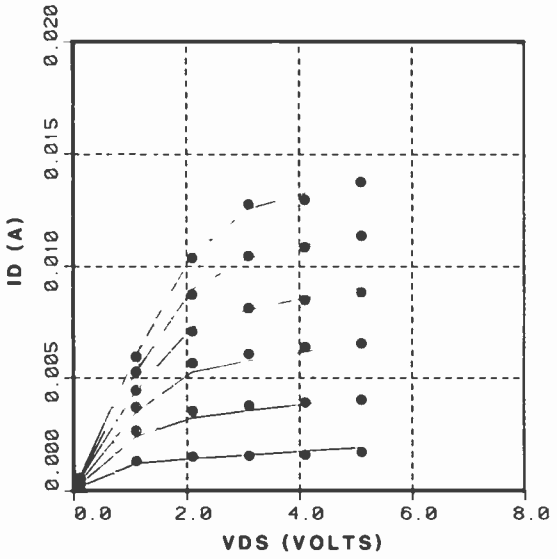


Fig. 13—CMOS n-channel I-V characteristics showing a comparison of the compact model MOSFET (curves) and MINIMOS (dots) data.

to use some intermediate model for the characteristics so that a small dc circuit analysis can be performed on the MOS transistor including the series source and drain resistances. We have elected to use the compact circuit model used in the R-CAP circuit simulation program along with the MOSFIT program to extract model parameters from the simulated *I-V* characteristics.

Fig. 13 shows a comparison between the I-V characteristics generated by MOSFIT and those by MINIMOS after a global parameter extraction has been performed. The discrepancy between the two sets of data is due to model error in the MOSFIT compact model. Even though the MOSFIT model is comprehensive, including many of the known second-order effects common to short-channel MOS transistors, it has approximately 10% maximum error in describing the actual shape of the characteristics. We hypothesize that two model errors contribute to this weakness: the approximations used in the model for the field dependence of the mobility and the use of only a single parameter to represent the effects of a varying channel impurity profile.

Resistance is added to the model by performing a circuit simulation using R-CAP and the model parameters as extracted by

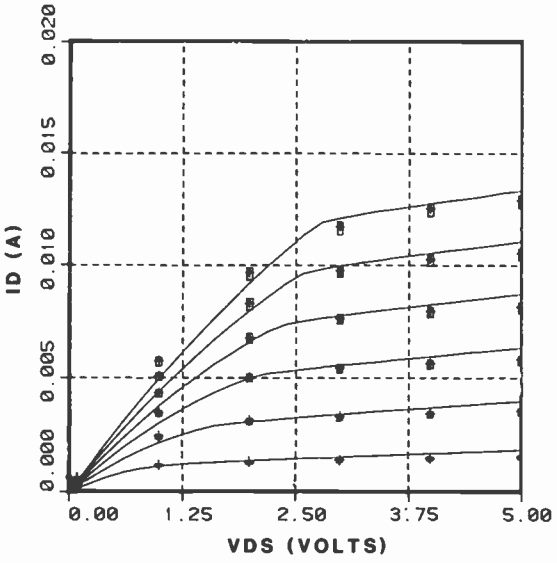


Fig. 14—CMOS n-channel *I-V* characteristics showing a comparison between the compact model with series resistance added (curves) and measured data (boxes).

MOSFIT. The resistance is added by changing the parameters describing the series resistance from a very low fixed value (used when extracting the model parameters) to the estimated resistance as determined from the sheet resistance and number of squares between the actual contact and the source-gate edge. A comparison of the *I-V* characteristics obtained in this manner and the measured data is shown in Fig. 14. A similar result is shown in Fig. 15 for the p-channel device.

6. Discussion and Conclusions

The results of this work show that excellent prediction of the *I-V* characteristics of CMOS transistors can be obtained via simulation. The simulation procedure is very involved, but the ability to use Model-Graph to perform the automatic piping of data between simulation stages makes it practical.

These procedures can be used when designing new transistors, but they are more accurate when used to diagnose imperfections and redesign existing devices. This is true principally because calibration against existing devices is more accurate. If too large a

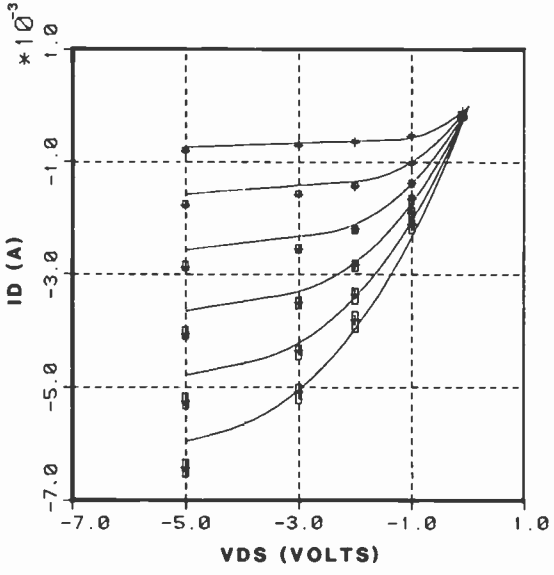


Fig. 15—CMOS p-channel *I-V* characteristics showing a comparison between the compact model with series resistance added (curves) and measured data (boxes).

change is made in the device design, however, the validity of the calibration would come into doubt and one could only guess at the accuracy of the simulation.

Finally, we note that the differences between the simulation model (the partial differential equations) and the circuit model (approximate solutions based on one-dimensional analysis) give rise to errors in the shape of the simulated *I-V* curve which can be seen when the measured and simulated characteristics are compared. The distortion of the curve due to parameter extraction using the compact model is carried forward into the final comparison (R-CAP results versus measurements). The final comparison would be significantly improved if we could add the effects of resistance during the MINIMOS simulation. That ability is expected to be available in future versions of MINIMOS.

References:

[1] *SUPREM-3 One-Dimensional Process Analysis Program,* Technology Modeling Associates, Inc. (1983).
[2] S. Selberherr et al., "MINIMOS—A Two-Dimensional MOS Transistor Analyzer," *IEEE Trans. Elec. Dev.*, **ED-27,** No. 8, Aug. 1980.
[3] C. Davis, "MOSFIT—A System for Modeling and Characterizing MOS Devices," Int. Conf. CAD, Santa Clara, Calif., Sept. 1983.
[4] C. B. Davis and M I. Payne, "R-CAP: An Integrated Circuit Simulator," *RCA Eng.*, **21**(1), p. 66, June 1975.
[5] R. Amantea and C. Davis, "MDLGRF: A System of Programs for Computer-Aided Modeling of Semiconductor Devices," Int. Conf. CAD, Santa Clara, Calif., Sept. 1983.
[6] C. P. Ho, J. D. Plummer, S. E. Hansen, and R. W. Dutton, "VLSI Process Modeling—SUPREM III," *IEEE Trans. Elec. Dev.*, **ED-30,** No. 11, p. 1439, Nov. 1983.

A SPICE II Subcircuit Representation for Power MOSFETs Using Empirical Methods

Gary M. Dolny
RCA Laboratories, Princeton, NJ 08540

Harold R. Ronan, Jr. and C. Frank Wheatley, Jr.
RCA Solid State Division, Mountaintop, PA 18707

Abstract—An accurate power MOSFET model for CAD circuit simulation is not widely available. This paper describes a power MOS *subcircuit* model that is compatible with SPICE II software and MOSFET terminal measurements. SPICE II is the circuit simulation package of choice for this work because of its universal availability, despite its inherent limitations. The limitations are circumvented through circuit means. Since the subcircuit models power MOSFET terminal behavior to be consistent with SPICE II limitations, it differs from physical models suggested in the literature.[1,2] However, it advances prior efforts[3] at simulating power MOSFET behavior, particularly in the areas of dynamic switching, third-quadrant operation, avalanche mode simulation, and diode recovery waveforms.

Introduction

Circuit simulation programs such as SPICE have been widely used as tools for the analysis and design of integrated circuits for many years. The semiconductor device models incorporated into the standard simulation programs were developed primarily for low-power devices and, hence, do not accurately simulate all modes of power-device operation. This lack of an accurate model is an obstacle to the use of circuit simulators for power electronic circuits.

Circuit simulation models for power MOSFETs are of particular interest due to the growing use of these devices as discrete compo-

nents and their potential application as the output stages of power integrated circuits. The built-in MOSFET models presently available in circuit simulation programs are not adequate to simulate a power MOS device for several reasons: (1) complex variations of capacitance with bias conditions, (2) the presence of a cascode JFET, which complicates both static and dynamic operation, and (3) the presence of a parasitic body diode, which influences third quadrant operation. At present, an accurate power MOSFET model suitable for use in CAD circuit simulation is not available.

The purpose of this work was to provide an accurate and efficient model for the power MOSFET suitable for use with commonly available circuit-simulation programs. The model described was developed specifically for the SPICE II simulation package. SPICE II was chosen because of its universal availability and because the similarity of SPICE model parameters to those used in other simulation programs simplifies the conversion of the model from one package to another.

The approach taken is to model power MOSFET behavior consistent with the inherent limitations of the SPICE II package by adding external components to the built-in SPICE II MOSFET models. In this way, a subcircuit with terminal characteristics representing those of a power MOSFET is developed. The great advantage of this approach is that the simulation package can be used in its standard form without requiring modification of internal device models that cannot be easily altered. Moreover, the general parameter extraction algorithm described here allows subcircuit model parameters to be determined from simple terminal measurements or from standard data sheets. The model is applicable to either p- or n-channel devices over a wide range of current and voltage ratings.

Recently, high-power MOSFET characteristics were simulated by adding a single lumped parasitic source resistance to a basic low-power MOSFET model.[3] This approach provides a good representation of power MOS static drain characteristics at high drain voltages and qualitative agreement at low drain voltages and under transient switching conditions. The present work advances these prior efforts to model power MOSFET behavior, particularly in the areas of dynamic switching, third-quadrant operation, avalanche mode simulation, and diode recovery waveforms.

Subcircuit Model Description

A power MOSFET unit cell is shown in cross section in Fig. 1.

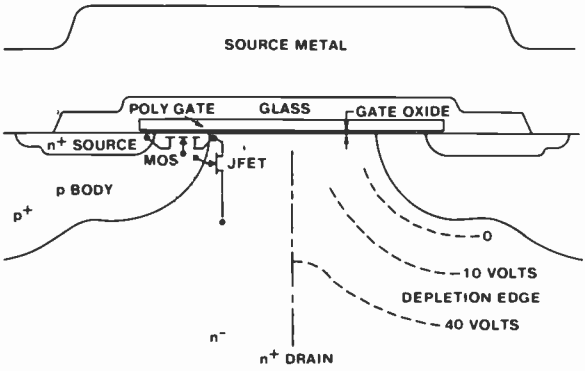


Fig. 1—Cross section of power MOSFET.

Recent work[1,2] has shown that the electrical characteristics of this structure can be understood in terms of a vertical JFET driven in cascode from a lateral low-voltage MOSFET. When the gate is positively biased with respect to the n^- bulk an accumulation layer exists in the n^- region beneath the gate. This accumulation layer acts as the drain of the lateral MOS and the source of the vertical JFET. The JFET channel is the n^- region between the two p-type body diffusions, which act as the gate of the JFET. The JFET drain is the n^- bulk, usually thought of as the power-MOSFET drain. Based on these observations the equivalent circuit of Fig. 2 has been developed.

In Fig. 2 the MOSFET and JFET are simulated by the standard SPICE II built-in model. Capacitor C_1 represents the power MOS gate-to-source capacitance. Physically, this capacitance consists of a contribution from the polysilicon gate up through the thick oxide to the source metal and a contribution from the gate down through the thin oxide to the n^+ source. This capacitance is essentially invariant with current and voltage.

C_2 represents the capacitance from the gate through the thin gate oxide to the accumulation layer. As long as the gate is positive with respect to the n^- neck region, the accumulation layer exists and C_2 is invariant. However, as the external drain voltage minus the IR drop across the n^- drift region approaches the gate potential, the accumulation layer ceases to exist and capacitance C_2 decreases abruptly.

C_3 is of significance only when no accumulation layer exists beneath the gate. Otherwise, the accumulation layer acts as an electrostatic shield.

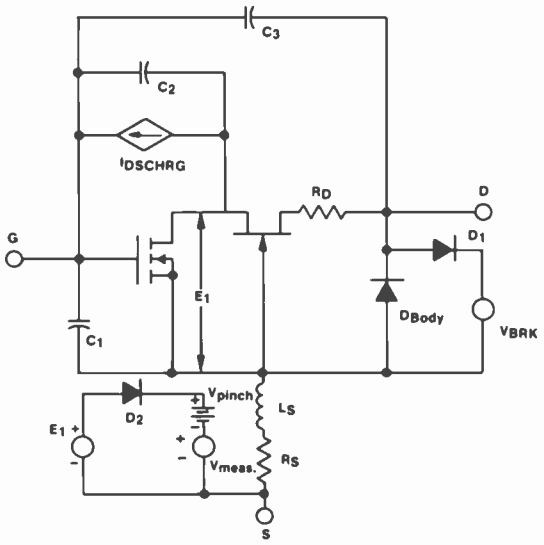


Fig. 2—Spice II subcircuit for power MOSFET simulation.

The abrupt, bias dependent changes in C_2 and C_3 cannot be easily modeled in the SPICE II package. For this reason the current controlled current source I_{DSCHRG} and the current sense subnetwork containing D2 must be employed. From Fig. 2, if the JFET source voltage E_1 is very low relative to its pinch-off voltage, the JFET is in a highly conductive state tightly coupling C_2 to the JFET drain. As the voltage E_1 approaches the pinch-off voltage of the vertical JFET (V_{PINCH}), the JFET operates in a constant current mode. This effectively decouples C_2 from the drain, thereby permitting a much faster slew rate as determined by C_3.

If E_1 exceeds the pinch-off voltage of the JFET, errors will exist in the output waveforms predicted by the model. The current controlled current source I_{DSCHRG} remedies this situation in conjunction with the subnetwork containing D2. As soon as potential E_1 exceeds the value of V_{PINCH}, current flows through D2. When this current is sensed by V_{MEAS}, the high gain current source I_{DSCHRG} is turned on rapidly discharging C_2 and effectively pinning E_1 at V_{PINCH}. The ideality factor of D_2 is set at 0.03 to assure that E_1 never exceeds V_{PINCH} by more than a few millivolts.

D_{BODY} represents the diode formed by the drain-to-body diffusion p-n junction. An external component is required for this element, since the built-in gate-to-drain diode in the SPICE II JFET model is not adequate to model third-quadrant conduction. Model param-

eters must be adjusted to assure that most of the third-quadrant current flows in D_{BODY}. This is accomplished by setting the saturation current of the gate-to-drain diode of the built-in JFET to an artificially small value such as 10^{-20} A.

Avalanche breakdown is simulated by the clamp circuit consisting of diode D_1 in series with V_{BRK}. Experience has shown that this configuration provides a better fit to measured breakdown characteristics as well as improved convergence properties compared to the SPICE II implicit diode-breakdown models.

Resistor R_D represents the series resistance of the n^- drain region. R_S represents the series source resistance made up of contributions from the n^+ source region, contact resistance, and source metal series resistance. L_S is the source inductance originating from the source metallization and bond wires.

Parameter Extraction

For the subcircuit model to be useful in simulating a power MOSFET, the subcircuit model parameters must be adjusted so that the model accurately represents the terminal characteristics of the physical device. This section outlines the procedure to determine the SPICE II model parameters required for accurate simulation of power MOSFETs.

The static *IV* characteristics of the power MOS subcircuit are determined by the low-power MOSFET of Fig. 2. This device operating in saturation is modeled in SPICE by

$$I_{DS} = \frac{K_P}{2L} W(V_{GS} - V_T)^2, \qquad [1]$$

where K_P is the process transconductance parameter and V_T is the threshold voltage. Fixing $W = L = 1$ μm in the SPICE II model, a plot of $\sqrt{I_{DS}}$ versus V_{GS} yields the parameters V_T and K_P as shown in Fig. 3. This data is also used to find the value of source resistance R_S, since this component causes the $\sqrt{2I_{DS}}$ versus V_{GS} curve to deviate from linearity at high current levels. Similarly the drain series resistance is found using the known value of R_S and plots of I_{DS} versus V_{DS} for the device operating in the linear region, as illustrated in Fig. 4.

Model parameters for the body diode are determined from Fig. 5 which plots log I_{DS} versus V_{DS} with $V_{GS} = 0.0$ for third-quadrant operation (i.e., $V_{DS} < 0$). The minority-carrier transit-time parameter T_T is chosen to provide the best fit to measured transient reverse recovery data. The junction capacitance of D_{BODY} represents

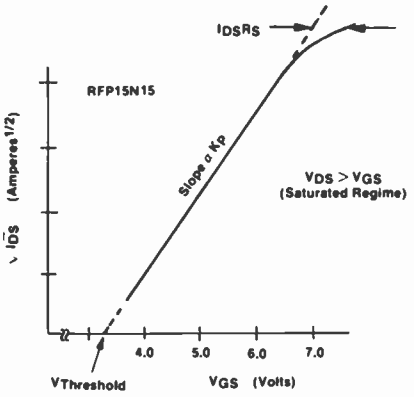


Fig. 3—Square root of drain current versus gate voltage defines $V_{threshold}$, K_p, and R_S.

power MOS output capacitance C_{OSS} which can be obtained from standard bridge measurements or data sheets.

The clamp circuit consisting of D_1 and V_{BRK} controls avalanche-breakdown simulation. V_{BRK} is set to the measured value of drain breakdown voltage, while model parameters I_S, N, and R are adjusted to provide the best fit to the measured breakdown curves.

Capacitors C_1, C_2, and C_3 along with the JFET model parameters are determined with the aid of Fig. 6. This curve plots drain voltage and gate voltage as a function of time for a power MOSFET driven with constant gate current I_G.[1] The initial slope of this curve defines

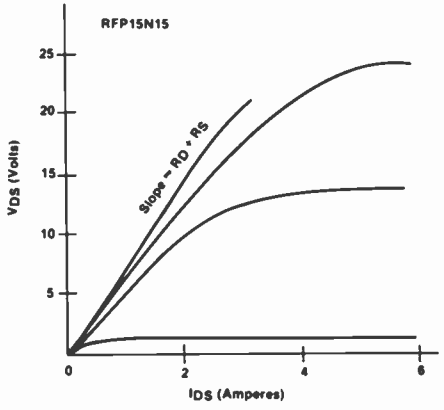


Fig. 4—Drain current versus drain voltage with constant gate voltage defines "on" resistance.

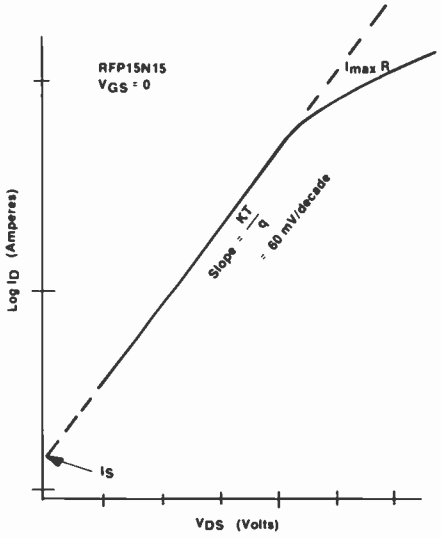


Fig. 5—Third-quadrant operation defines I_S and R of diode D_{BODY}.

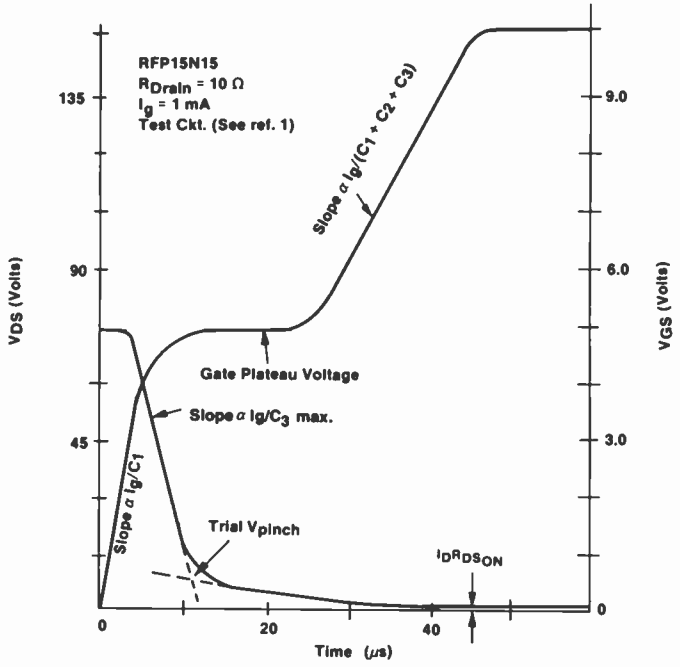


Fig. 6—Drain and gate voltages versus time determines C_1, C_2, C_3, and V_{PINCH}.

Table 1—Preferred Algorithm for Parameter Extraction

1. Determine K_P of lateral MOS
2. Determine V_{TH} of lateral MOS
3. Determine C_1
4. Determine $C_1 + C_2 + C_3$
5. Determine R_{DS}
6. Assign B of JFET = 100 × Kp of lateral MOS
7. Use trial V_{PINCH}
8. Use C_2 (min), C_3 (max) and curve-fit C's
9. Adjust V_{PINCH} to fix gate-voltage plateau

C_1, since for $V_{GS} < V_T$ the power MOSFET is in its off state and, hence, the gate-to-source capacitance charges linearly under constant-current conditions.

When threshold is reached, the power MOS turns on and its drain voltage drops toward its minimum value. Initially the JFET is operating beyond just below pinch off and the slope of the V_{DS} versus time curve is controlled by C_3. When the drain voltage falls below V_{PINCH}, the JFET conductance is high, strongly coupling C_2 to the JFET drain and greatly reducing the drain-voltage slew rate.

Table 1 gives the preferred algorithm for parameter extraction, while Table 2 summarizes the required empirical inputs.

Results

Results of typical model parameter extraction are shown in Table

Table 2—Empirical Inputs

MOSFET	Enhancement mode; $W = L = 1$ μm; K_P (Fig. 3); V_{TO} (Fig. 3); C's = 0; I_{DSO} = 1E − 12
JFET	Depletion mode; area factor = 1; $B = 100\ K_P$ (Fig. 3); $V_{TO} = V_{PINCH}$ (Fig. 6); C's = diode lifetime = R_{series} = 0; diode ideality factor = 1.0, I_{DSO} = 1E − 20
BODY DIODE	I_S from Fig. 5; Ideality Factor = 1.0; R from Fig. 5 (must be very much smaller than R_D); C (from C_{OSS}); lifetime = best fit to T_{rr}
D_1	I_S = arbitrary; C = lifetime = 0; ideality factor = best low-current fit; R = best high-current fit
D_2	I_S = 1E − 8; C = lifetime = R = 0; ideality factor = 0.03
R_S	Fig. 3
R_{DRAIN}	Fig. 4
L_S	Approx. (5L)ln(4 L/d) nH; L and d are source wire inches
V_{PINCH}	V_{TO} of JFET
V_{BRK}	Avalanche voltage
C_1	Fig. 6
C_2	Max. from Fig. 6
C_3	Min. from Fig. 6

Table 3—Input Parameters to SPICE Model

SPICE Parameter	RF15N15 Value
Lateral MOS	
V_{TO}	3.2 Volts
K_P	4.0 A/V^2
W/L	1.0
Model Level	1
Vertical JFET	
V_{TO}	−10.0 Volts
Beta	400
I_S	10^{-20} A
Body Diode	
C_{JO}	1100 pF
T_T	2×10^{-7} sec
I_S	10^{-9} A
Passive Elements	
C_1	1800 pF
C_2	40 pF
C_3	1000 pF
R_S	0.003 ohm
L_S	6×10^{-9} H
R_D	0.05 ohm
D1 V_{BRK}	165 Volts

3. These data were obtained from an RCA RFP15N15 power MOSFET rated at 15 amperes with 150 volt blocking capability.

Figs. 7 and 8 compare measured static data with calculated transfer curves and output curves. As can be seen, an excellent fit is obtained even at low values of drain voltage.

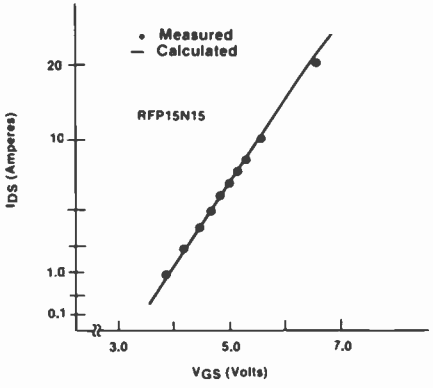


Fig. 7—Drain current (note square-root scale) versus gate voltage showing measured curve and calculated points.

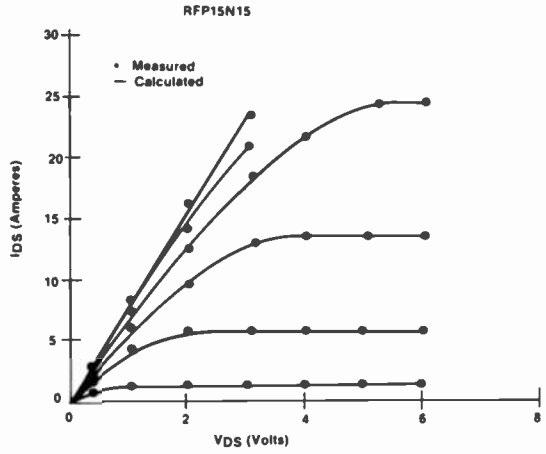


Fig. 8—Drain current versus drain voltage for constant values of gate voltage showing measured curves and calculated points.

Fig. 9 shows drain current versus drain voltage for the full first-quadrant range of operation including avalanche breakdown, and Fig. 10 compares simulated and measured results for third-quadrant operation. The excellent fit over the entire operating range of the device is evident.

Calculated switching data are compared with measured switching curves in Figs. 11 and 12. This data includes both constant gate-current drive and step gate-voltage inputs. Measured and calculated body-diode recovery curves are shown in Fig. 13.

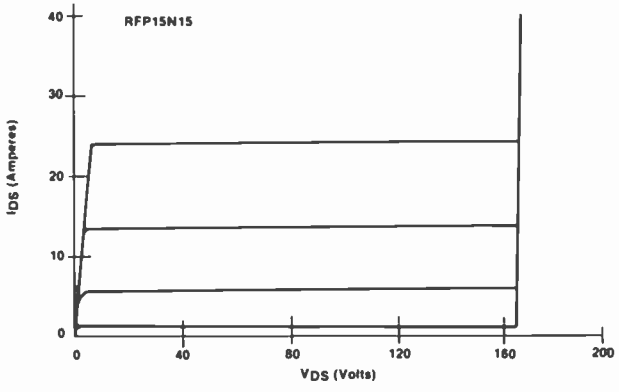


Fig. 9—First-quadrant drain current versus drain voltage, V_{GS} = constant. Note avalanche breakdown (calculated).

Conclusion

An equivalent circuit model for Power MOSFETs that is suitable for use with the SPICE CAD program has been demonstrated. The model is compatible with all versions of SPICE presently available without modification to the program's internal source code. Further, the model is easily transportable to other commonly used circuit simulators. The model addresses static and dynamic behavior over the entire operating range of the device, including avalanche breakdown. It is empirical in nature and all necessary parameters may be inferred from data sheets or simple terminal measurements. The

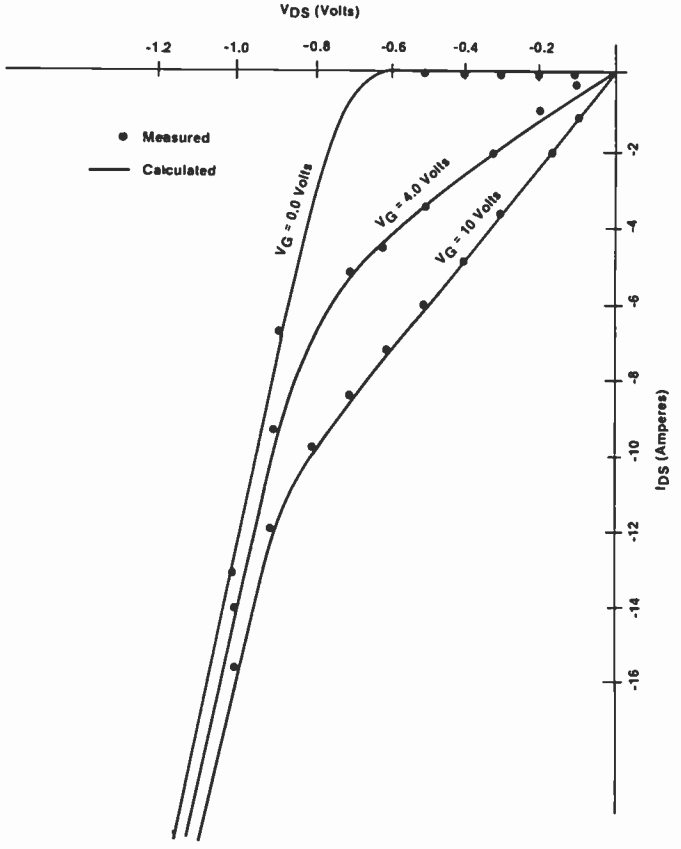


Fig. 10—Third-quadrant drain current versus drain voltage at constant gate voltage.

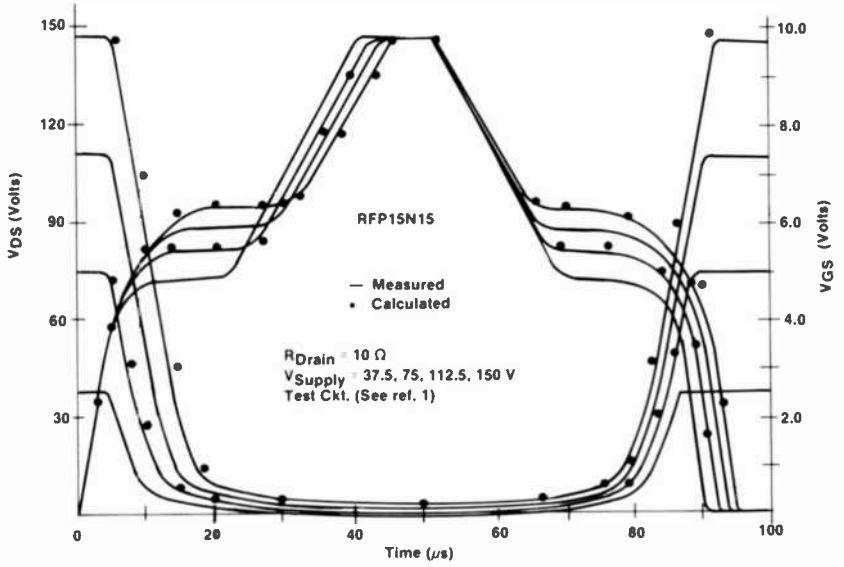


Fig. 11—Drain and gate voltages versus time for constant gate current showing measured curves and calculated points.

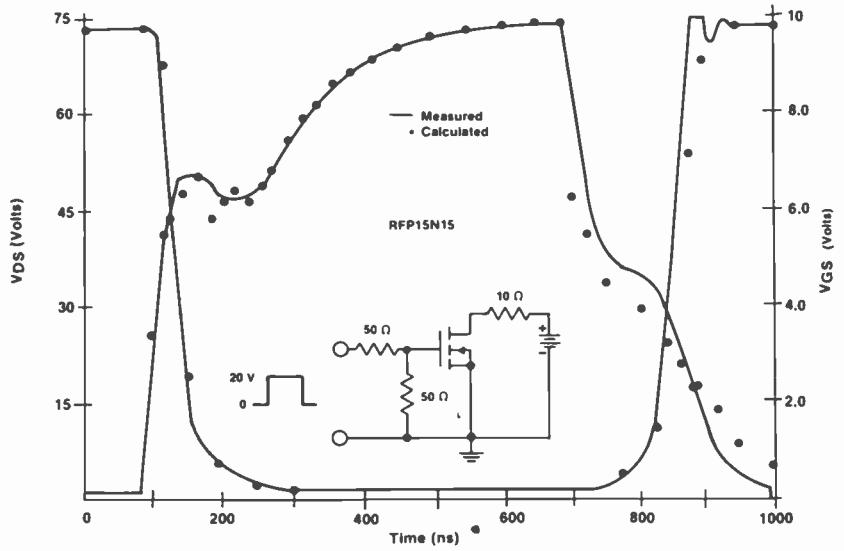


Fig. 12—Drain and gate voltages versus time for standard switching circuit showing measured curves and calculated points.

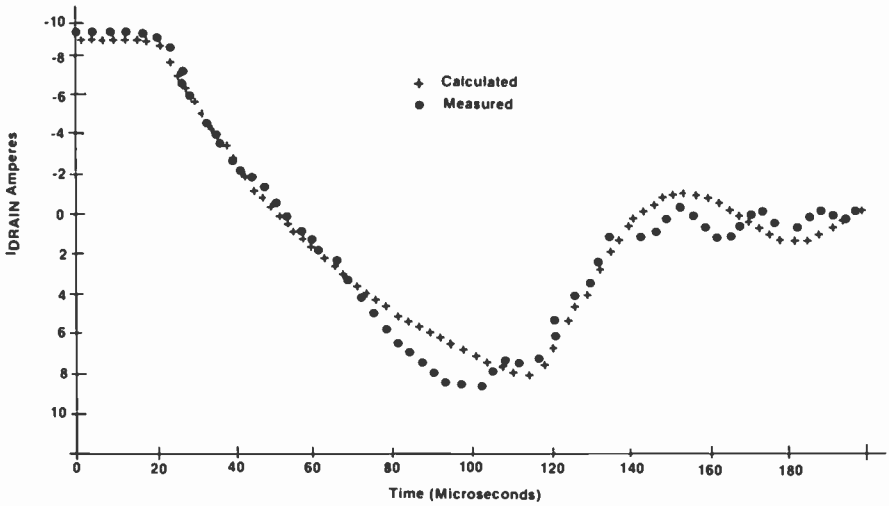


Fig. 13—Third-quadrant diode-recovery-time curve showing calculated and measured values.

model has been experimentally verified over the entire static and dynamic operating range. Excellent agreement has been obtained between measured and simulated results.

Acknowledgment

The authors would like to thank Robert Amantea of RCA Laboratories for his careful review and suggestions concerning this manuscript.

References:

[1] C. F. Wheatley, Jr. and H. R. Ronan, Jr., "Switching Waveforms of the LFET: A 5 Volt Gate-Drive Power MOSFET," *Power Electronic Specialists Conf. Record*, June 1984, p. 238.
[2] H. R. Ronan, Jr., and C. F. Wheatley, Jr., "Power MOSFET Switching Waveforms: A New Insight," *Proc. Powercon 11*, April 1984, p. C3.
[3] H. A. Nienhaus, J. C. Bowers, and P. C. Herren, Jr., "A High Power MOSFET Computer Model," Power Conversion International Conf., Jan. 1982, p. 65.

N-FET, a New Software Tool for Large-Signal GaAs FET Circuit Design

Walter R. Curtice
RCA Laboratories, Princeton, NJ 08540

Morris Ettenberg*
City College of New York, NY 10031

Abstract—A nonlinear equivalent circuit model for the GaAs FET has been used to develop an interactive computer program. The model is based on the small-signal device model and separate current measurements, including drain-gate avalanche current data. The harmonic balance technique is used to find the FET rf load-pull characteristics in an amplifier configuration under large-signal operation. Computed and experimental load-pull results show good agreement.

Introduction

GaAs FET amplifiers can be designed for good small-signal performance by use of the device scattering parameters (*S*-parameters). Extensive analytical and CAD techniques have been developed (see, for example, Yarman and Carlin[1] and Ku[2]). However, for power amplifiers, only the empirical load-pull method[3] has been useful for finding the optimum load under large-signal operation. It would be advantageous to have a nonlinear FET model to use with CAD programs to optimize efficiency, bandwidth, and other parameters before actual construction. Amplifier performance will be greatly enhanced and the amount of tuning required after construction, significantly reduced.

Since there are several general-purpose nonlinear circuit analysis

* Consultant to RCA Laboratories.

programs available, such as SPICE[4] and R-CAP[5], why not use these for designing large-signal GaAs FET amplifiers? And what about the accurate two-dimensional GaAs FET models? Can they be used? Does small-signal modeling help at all for large-signal network design? All these questions should be resolved before developing a new approach to large-signal modeling.

We will discuss why the above approaches are unsatisfactory and how a new software tool, called N-FET (nonlinear FET), is used to generate large-signal design information for GaAs power FETs.

Time-Domain FET Simulation Techniques

Nonlinear FET operation produces nonsinusoidal current and voltage waveforms at both the input and the output ports. Since the nonlinearities of the FET depend on the instantaneous voltages and currents, analysis of the FET is most accurately accomplished in the time domain. Analysis in the frequency domain by techniques such as the Volterra series is limited to weak nonlinearities because of the slow convergence of the series.

The general-purpose nonlinear circuit analysis programs that exist were designed primarily for transient (time domain) analysis of silicon integrated circuits.[4,5] By adding new models for GaAs devices, Curtice[6] and Sussman-Fort et al.[7] show that these programs can be used for studying GaAs integrated circuits. However, a more sophisticated model is required to study GaAs power FETs operated at high dc-to-rf conversion efficiency. Such a model must contain an accurate description of all the important device nonlinearities and also efficiently analyze the external microwave circuit interaction over many rf cycles. The circuit reactances lead to time constants that are large relative to the rf period. Time-domain analysis is thus inefficient and, in fact, unnecessary since the reactances are linear.

As an experiment, the nonlinear GaAs FET model in R-CAP was used to analyze an FET amplifier. An RCA GaAs Schottky-barrier junction FET was selected for study. The device has a gate length of approximately 1 μm and a gate periphery of 600 μm. It is mounted on a 1.7-mm (70 mil) carrier. The dc parameters were measured by the automated Fukui method,[8] and *S*-parameters were measured from 4 to 20 GHz. A small-signal, lumped-element circuit model was constructed for the normal operating biases by means of SUPER-COMPACT™[9] with a procedure developed previously at RCA.[10] In designing inductor/capacitor matching networks,

SUPER-COMPACT was used to match the device to 50 Ω at 12 GHz. Analysis shows that a gain of 11.1 dB should be realizable for the matched condition.

Fig. 1 shows the lumped-element model that was analyzed with R-CAP. This model consists of the small-signal model and two nonlinear elements. The first nonlinear element is the gate-source capacitance, represented as a reverse-biased diode having voltage-variable capacitance. Nonlinear gate-source conduction current can also be included (although it was not) by adjusting the diode saturation current. The second nonlinear element is the voltage-controlled current source. The current is expressed as an analytic function of gate-source and drain-source voltages, the values of which are shown in Fig. 2. This characteristic approximates the RCA device.

To simulate the model (Fig. 1) in the time domain we used R-CAP. For a 1-V amplitude of input signal at 12 GHz, we found steady-state conditions with a gain of 11.7 dB. This value agrees well with the small-signal frequency domain calculation. The input rf power is 2.35 mW, and the power delivered to the load is 35 mW. The power-added conversion efficiency is 6.6%.

We expected that the input power could be increased and new steady-state conditions achieved up to the point of high efficiency. However, for significantly larger input power, R-CAP developed a numerical instability, eliminating the possibility of physical solutions. The instability was traced to the coding of the time delay in the GaAs FET subroutine. Some solutions were found by eliminating the time delay of the voltage-controlled current source. Fig.

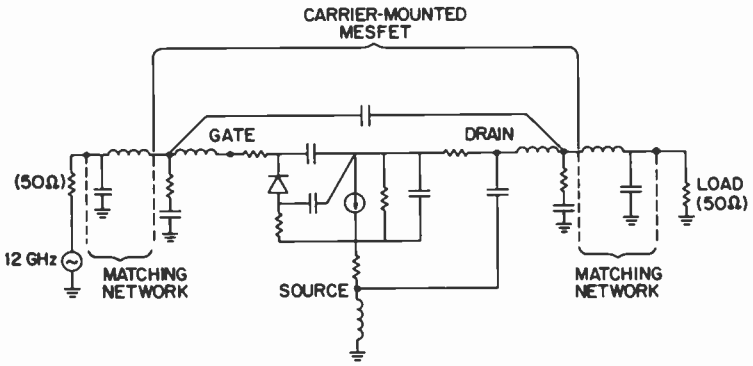


Fig. 1—Equivalent circuit for carrier-mounted GaAs FET operated as an amplifier with input- and output-matching networks.

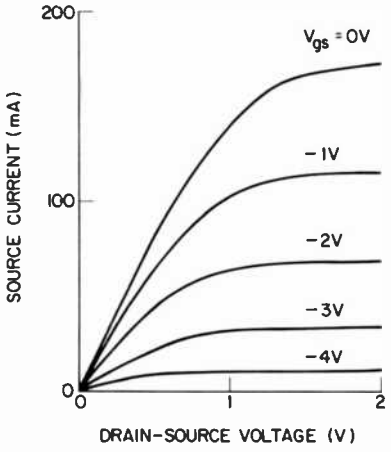


Fig. 2—Drain-source current-voltage relationships for various gate-source voltages as determined from the analytical expression that approximates the RCA device.

3 is an example of such a solution. Here 49 mW of output power is produced at a 7.8-dB gain with 9.8% efficiency. Note the nonsinusoidal current waveform. This is produced by the drain-source voltage swing to low voltages, below the current knee. Also note that although the drain-source voltage has a large ac amplitude (about equal to the bias voltage), the ac amplitude of current is low. This indicates that the device is driving too high a load impedance. Higher efficiency can be achieved by reducing the load impedance.

Solutions of the type shown in Fig. 3 were obtained after six to eight rf cycles, when steady-state operation is reached. For more complicated circuits, steady state may not occur until 20 rf periods. See, for example, Golio et al.[11] where SPICE is used to study large-signal operation of a GaAs FET amplifier. Golio et al. had to make major improvements to the J-FET model in SPICE to approximate the GaAs FET.

Although it may be possible to eliminate the numerical instability we found in R-CAP, both R-CAP and SPICE require lengthy execution (in time) to reach steady state. Ironically, it is the external linear circuit elements that cause this effect. In addition, neither the model of Golio et al. nor the model of Curtice[6] is sufficiently detailed for amplifier design. For instance, neither model includes drain-gate avalanche breakdown currents. If we start with a very detailed two-dimensional (2-D) GaAs FET model, such as described by Curtice and Yun,[12] it is possible to add external circuits

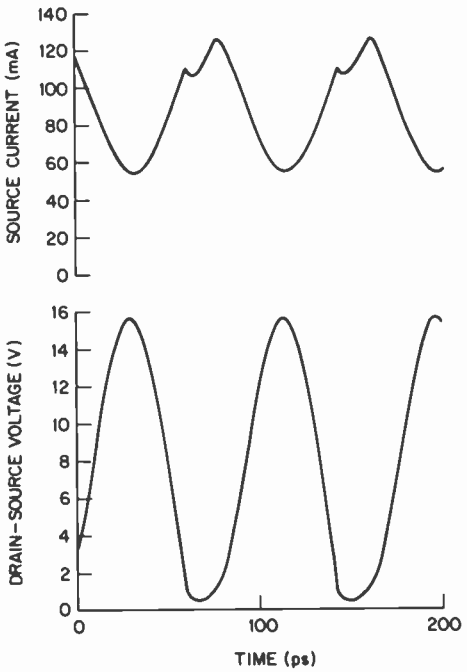


Fig. 3—Drain-source current and voltage waveforms determined from R-CAP analysis of the circuit of Fig. 1.

and develop time-domain solutions, as with R-CAP. The problem here is that the FET model is already quite CPU intensive. Adding external circuits increases the problem. Even with simple external circuits, the 2-D model must run overnight on a VAX 11/780 to reach a steady-state solution. This clearly is not useful as an interactive design tool.

N-FET, the New Nonlinear FET Program

Camacho-Penalosa,[13] Petersen et al.,[14] Gilmore and Rosenbaum,[15] and Materka and Kacprzak[16] all have utilized an analysis technique known as "harmonic balance" to find steady-state solutions for GaAs FET amplifiers under large-signal operation. We have extended the work of Peterson et al. using a more detailed FET model and have developed a useful design tool for the microwave engineer.[17]

Our nonlinear device model has evolved from the self-consistent GaAs FET small-signal model reported earlier by Curtice and

Camisa.[10] The program N-FET provides a computer-aided means to develop output circuit designs that optimize the amplifier performance (i.e., efficiency, bandwidth, etc.). Accurate prediction of large-signal load-pull performance is essential to accurately design output networks. In addition, we operate the program on Hewlett-Packard 1000 RTE minicomputers to reduce the cost of computation.

The program N-FET consists of time-domain analysis of the GaAs FET coupled with frequency-domain analysis of the input- and output-matching circuits. The nonlinear FET elements must be analyzed in the time domain to preserve their physical nature. The linear circuit's response to the FET current excitation can be analyzed in the frequency domain by standard techniques. Transformation between time and frequency domains is accomplished by use of a discrete Fourier transform. A valid physical solution is obtained when the voltage waveform at the input (or output) of the FET produces a current waveform into the device that is the negative of that into the rf circuit to within some small error. The program flow chart is shown in Fig. 4. The program uses the method of successive approximation to minimize the error between FET currents and external rf circuit currents. The input and output voltages are estimated initially by the program using either small-signal theory or a previous solution. After evaluation of the error value, Newton's method is used to produce the next trial solution. In practice, less than 100% update assures convergence.

Fig. 5 shows the equivalent circuit model assumed. This model is noticeably different from the one Curtice and Camisa used for small-signal modeling. The drain-channel capacitor is omitted to simplify node current equations. This produces some loss of accuracy. Also, two new current sources are used. The drain-gate voltage-controlled current source represents the drain-gate avalanche current that can occur at large-signal operation. The gate-source voltage-controlled current source represents gate current that occurs when the gate-source junction is forward biased. The third current source, $I_{ds}(V_{in},V_{out})$, is the large-signal form of the usual small-signal transconductance.

Table 1 summarizes the equations used to describe each of these voltage-controlled current sources. We will briefly discuss the form of each equation. Table 2 summarizes the characterization tests for a given FET.

Although a quadratic relationship is often assumed for drain current as a function of gate-source voltage, real devices often do not

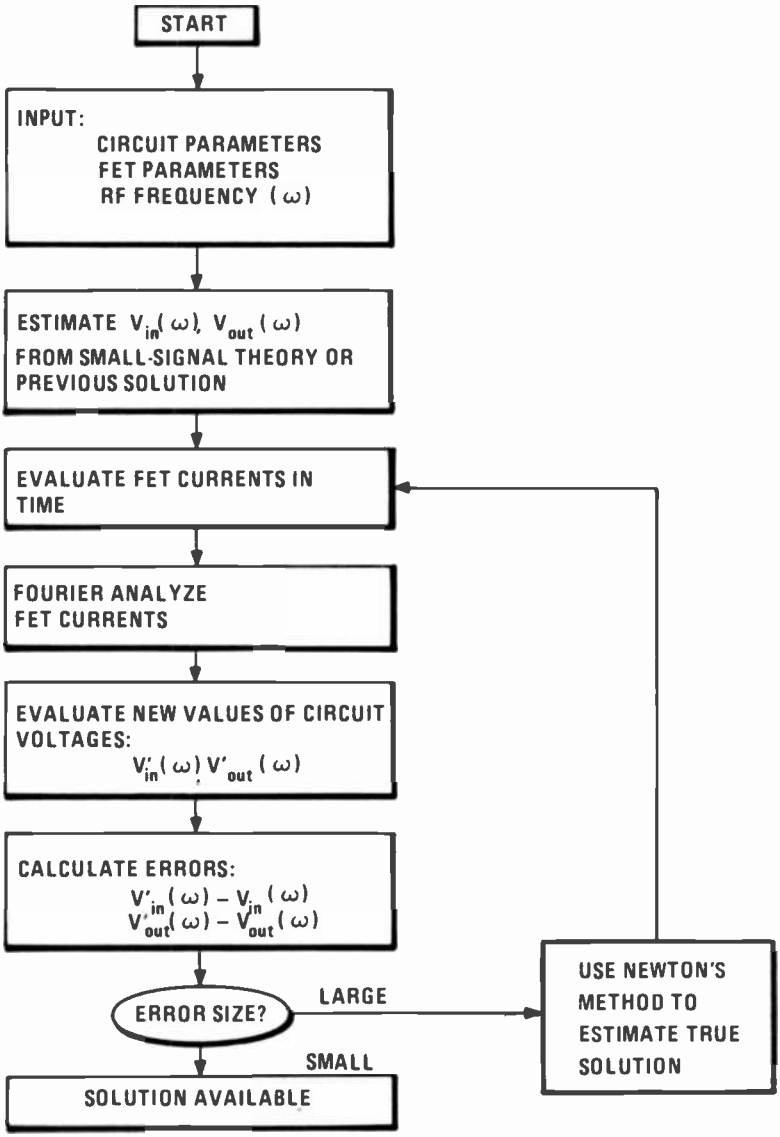


Fig. 4—Program flow chart.

exhibit such a relationship. We have found that it is more accurate to use a cubic relationship. One disadvantage of the cubic relationship is that, unlike the quadratic, a pinch-off voltage may result that makes either current or transconductance zero, but not both.

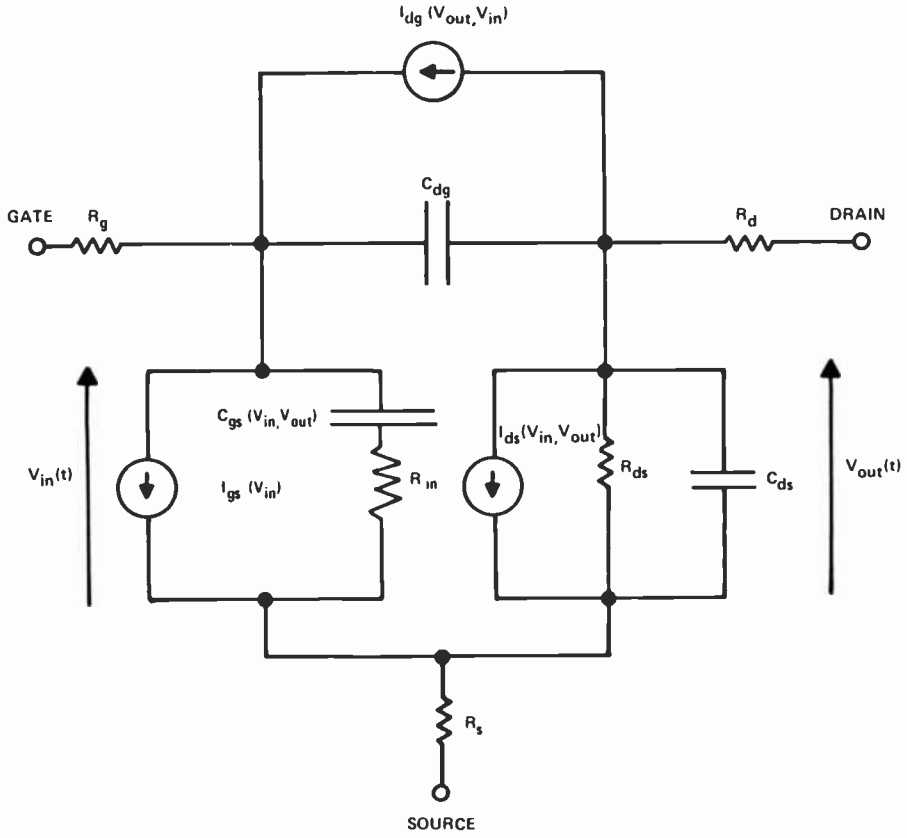


Fig. 5—Equivalent circuit model used in N-FET for the GaAs FET.

All FETs show an increase in pinch-off voltage as drain-source voltage is increased. In submicrometer-gate-length FETs this effect is quite pronounced. We have introduced coefficient β to describe this effect and have used a FORTRAN program to evaluate the coefficients A_0, A_1, A_2, A_3, and β from measured I-V data. We found the time delay, τ, at normal bias by using the small-signal model. This let us evaluate the coefficient A_5. Previous studies have shown τ to vary approximately linearly with drain-source voltage.[17]

Fig. 6 shows the current-voltage relationship calculated for a second RCA FET by means of these equations. This technique produces an accurate approximation to the measured data. Pulsed measurements of drain-gate avalanche currents are made on devices. Fig. 7 shows these data for the same device. The coefficients R_1, R_2,

Table 1—Voltage-Controlled Current Sources

(1) The Gate-Controlled Drain Current Is Expressed as:

$$I_{ds} = (A_0 + A_1 \cdot V_1 + A_2 \cdot V_1^2 + A_3 \cdot V_1^3) \cdot \tanh[\gamma \cdot V_{out}(t)]$$

where $V_1 = V_{in}(t - \tau) \cdot [1 + \beta \cdot V_{out}^{o} - \beta \cdot V_{out}(t)]$

$$\tau = A_5 \cdot V_{out}(t)$$

(2) The Drain-Gate Avalanche Current Is:

$$I_{dg} = \begin{cases} [V_{dg}(t) - V_b(t)]/R_1 & V_{dg} > V_b \\ 0 & V_{dg} < V_b \end{cases}$$

where $V_b = V_{bo} + R_2 \cdot I_{ds}$

(3) The Forward-Biased Gate Current Is:

$$I_{gs} = \begin{cases} \dfrac{V_{in}(t) - V_{bi}}{R_F} & V_{in}(t) > V_{bi} \\ 0 & V_{in}(t) < V_{bi} \end{cases}$$

where V_{bi} = built-in voltage

and V_{bo} are evaluated from these data. Note that R_2 accommodates the dependence of breakdown voltage on channel current that is usually observed.

The expression for forward-biased gate current [Eq. (3) in Table 1] confirms that there tends to be much resistance in series with the Schottky-barrier junction. These coefficients are often best evaluated from pulsed voltage data.

The values of R_g, R_d, and R_s are obtained from the automated Fukui measurements. The values of C_{dg}, C_{gs}, R_{ds}, and C_{ds} at the bias point are obtained from the small-signal model by the technique developed by Curtice and Camisa and based on *S*-parameter data. Although both C_{gs} and C_{dg} are nonlinear functions of voltage, computation including these characteristics produced only small effects upon the rf saturation characteristics. For N-FET, C_{dg} is assumed constant and C_{gs} is taken to be a weak function of $V_{in}(t)$ and $V_{out}(t)$. The particular function was derived from detailed measurements[17] of several devices over wide ranges of V_{gs} and V_{ds}.

Table 2—FET Circuit Model Characterization Tests

- Fundamental Device Parameter Determination
  (dc *I-V* and Fukui method)
  $R_s R_d R_g$
  $I_{ds}(V_{gs}, V_{ds})$
  $I_{gs}(V_{gs})$
- Linear Model and Parasitic Parameter Determination
  (Wideband *S* parameter network analysis)
  Cold FET ($V_{ds} = 0$)
  Hot FET ($V_{ds} > 0$)
- Pulsed Avalanche Measurements
  $I_{dg}(V_{gs}, V_{ds})$

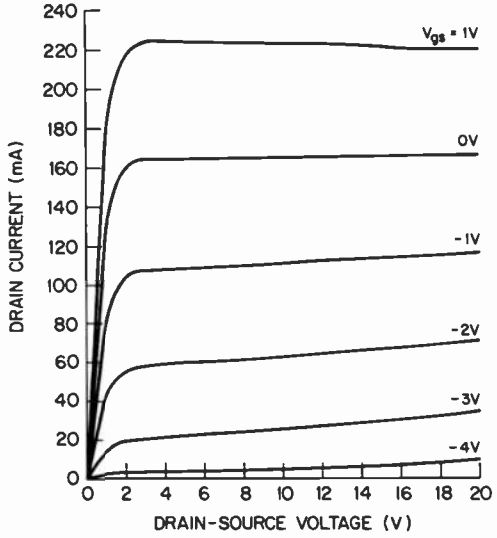


Fig. 6—Current-voltage relationships for RCA device B1512-3A as calculated by the analytical equations.

In no case did we ever find the gate-source capacitance, C_{gs}, to vary as $(V_{bi} - V_{gs})^{-1/2}$, which would occur for a simple Schottky-barrier junction. We conclude that there are velocity saturation effects in the channel as well as a significant fixed parasitic gate-source capacitance that reduce the capacitive nonlinearity.

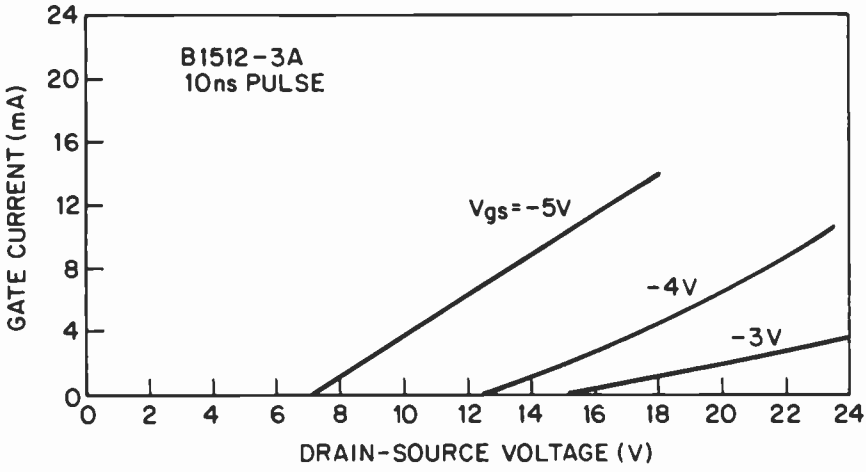


Fig. 7—Pulsed current-voltage data for RCA device B1512-3A.

The measured small-signal drain-source resistance is found to be a strong function of V_{gs} and V_{ds}. Although it is not obvious, this also occurs in N-FET. Differentation of the drain current in the model gives the following result for small-signal output rf conductance:

$$g_{ds} = \frac{1}{R_{ds}} - g_{mo}\beta\, V_{in} + \frac{\gamma I_{ds} \cdot \operatorname{sech}^2 [\gamma \cdot V_{out}(t)]}{\tanh [\gamma\, V_{out}(t)]} \qquad [1]$$

where

$$g_{mo} = (A_1 + 2A_2\, V_1 + 3A_3\, V_1^2) \cdot \tanh [\gamma \cdot V_{out}(t)] \qquad [2]$$

The three terms comprising g_{ds} may be understood to be

(1) a fixed conductance term,
(2) a substrate conductance term that causes the pinch-off voltage to change with V_{ds}, and
(3) a channel conductance term that is important only at low drain-source voltages (i.e., below current saturation).

The net result is that the output rf conductance depends on V_{gs} and V_{ds} and is consistent with actual device behavior.

Similarly, the small-signal intrinsic transconductance can be evaluated by differentation of drain current with respect to V_{in}. The result is

$$g_m = g_{mo}\{1 + \beta[V^{\circ}_{out} - V_{out}(t)]\} \qquad [3]$$

Coefficient β causes the transconductance to decrease with increasing drain-source voltage, consistent with the behavior of FETs.

In summary, the equivalent circuit model has been constructed with the simplified circuit shown in Fig. 5. The principal nonlinearities are the voltage-controlled current sources. These must be characterized for each device. Including the bias dependence of pinch-off voltage in the drain-source current was found to be important. The small-signal properties of the model proved consistent with FET behavior.

N-FET Used in Circuit Analysis

Fig. 8 shows the lumped-element equivalent circuit model of the GaAs FET amplifier. The two-port networks $[Y_{in}]$ and $[Y_{out}]$ are matching networks presumably designed for maximum power transfer at the input and output, respectively. Note that the gate and drain resistances of the FET are absorbed into these networks.

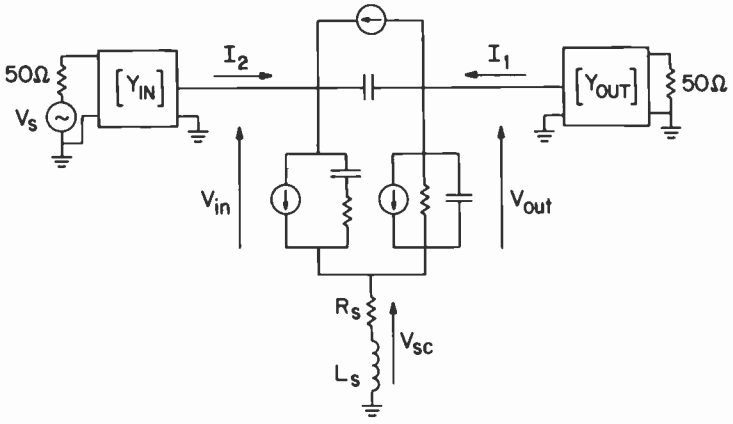


Fig. 8—Equivalent circuit model of a GaAs amplifier as used in the nonlinear analysis program.

Fig. 9 shows a typical topology for a carrier-mounted FET. To achieve conjugate matching at input and output for maximum small-signal gain, the first design of these networks is usually done by SUPER-COMPACT. Once a trial design is available, the *Y*-

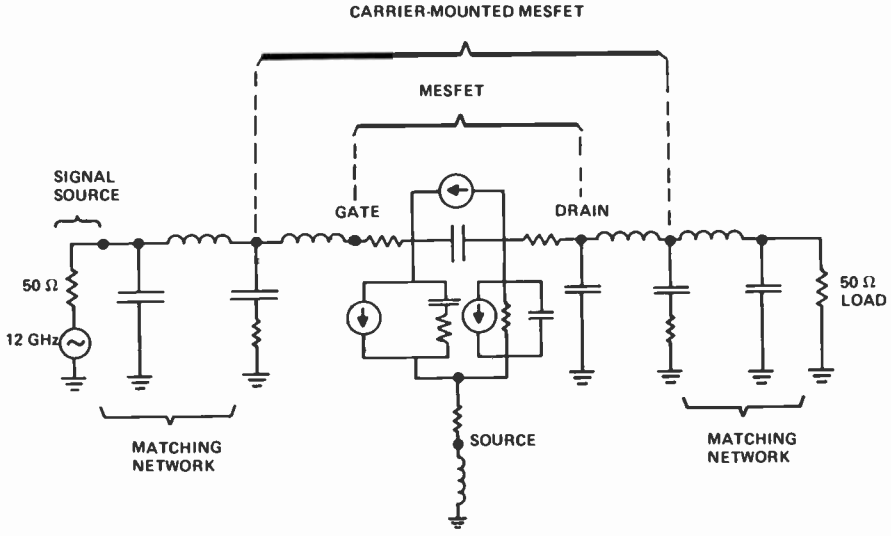


Fig. 9—Equivalent circuit model of a GaAs FET amplifier showing specific topology for the matching circuits.

transfer characteristics can be evaluated. The input voltage is

$$V_{in}(t) + V_{sc}(t) = V_{gs}^{dc} + \sum_n [a_n \sin(n\omega t) + b_n \cos(n\omega t)] \quad [4]$$

where $V_{sc}(t)$ is the voltage drop across the source resistance and inductance, assumed to be

$$V_{sc}(t) = V_{sc}^{dc} + \sum_n [c_n \sin(n\omega t) + d_n \cos(n\omega t)] \quad [5]$$

The output device voltage is

$$V_{out}(t) + V_{sc}(t) = V_{ds}^{dc} + \sum_n [f_n \sin(n\omega t) + g_n \cos(n\omega t)] \quad [6]$$

The present version of N-FET uses dc, fundamental, and second- and third-harmonic voltage components. The time-domain expressions for $V_{in}(t)$ and $V_{out}(t)$, Eqs. [4] and [6], are used to generate device currents. For example, the total drain current consists of drain-source current,

$$I_{ds}[V_{in}(t), V_{out}(t)] + C_{ds} \frac{d}{dt} V_{out}(t) + \frac{V_{out}(t) - V_{out}^{dc}}{R_{ds}} \quad [7]$$

plus drain-gate displacement current,

$$C_{dg} \frac{d}{dt} [V_{out}(t) - V_{in}(t)] \quad [8]$$

and the avalanche current, if $[V_{out}(t) - V_{in}(t)]$ is greater than the breakdown voltage, V_b.

The drain and gate currents are then analyzed by use of a discrete Fourier transform to find their frequency components. This process need not be used for linear circuit elements, such as C_{dg}, because these can be calculated from linear circuit theory.

Interactive Use of N-FET

N-FET is available on the HP1000/A900 (Hewlett-Packard Corp.). It is necessary to create a data file for the FET. This file should contain all the device coefficients described earlier and also a few others. It is necessary to specify the fundamental operating frequency, operating voltage biasing, number of voltage harmonics that need to be included, number of subdivisions of the rf cycle to use in analysis, and the harmonic impedances present at the input and output. The harmonic impedances can be evaluated from mea-

sured data or from estimations. In either case, these impedances are often difficult to evaluate accurately. N-FET is an interactive program in which the operating parameters may be changed. An algorithm is available for calculating the contours of constant power as the load is varied while the input-matching circuit and the drive power are held constant. This feature simulates the experimental load-pull techniques used to design high-power amplifiers. The algorithm is based on the facts that contours of constant power are closed curves on a Smith chart and that a contour of higher power will be enclosed within a contour of lower power. Once a load for a particular power, P_1, has been chosen, a small change is made in the amplitude of the load vector, and the power is calculated for this trial load. If the calculated power is greater than P_1, the contour trial vector added to the load vector is rotated clockwise until P_1 is found. If the trial power is less than P_1, the trial vector is rotated counterclockwise until P_1 is found. In this way the next point on the constant power contour is always found in the direction that continues the contour in the counterclockwise direction. The parameters controlling the contour generation may be changed interactively during the program run to give a finer contour with closely spaced load points or a contour that is generated more quickly and fewer points. A typical contour of 30 points will take about 10 minutes on the HP1000/A900. The contour may be plotted as the program runs, and more than one contour may be run without exiting. The pertinent data are written to a file that is saved for later retrieval. SMITHPLOT, a second program, uses the data file saved from an N-FET run and plots the contour on a Smith chart either on the terminal or on a hard-copy plotter.

Load-pull contours are developed as follows: The load is adjusted for maximum output power for a given rf drive power. Then the load is changed to produce less output power (e.g., -1 dB) and a load contour for constant power is measured.

Fig. 10 shows the calculated load-pull contour and measured data for 175 mW of output power for an RCA FET that was originally tuned for 205 mW of output power. Fig. 11 shows the same comparison for an output power of 150 mW with the same drive conditions. There is good agreement regarding the output load-pull characteristics for a given output power, but disagreement in driver power (and gain) by 2.5 dB. Some of this error is attributable to losses in the input tuner used in the measurements.

Fig. 12 shows the load conditions for maximum power output at seven different output power values as computed by the nonlinear program. Measured data are also shown and are in good agreement.

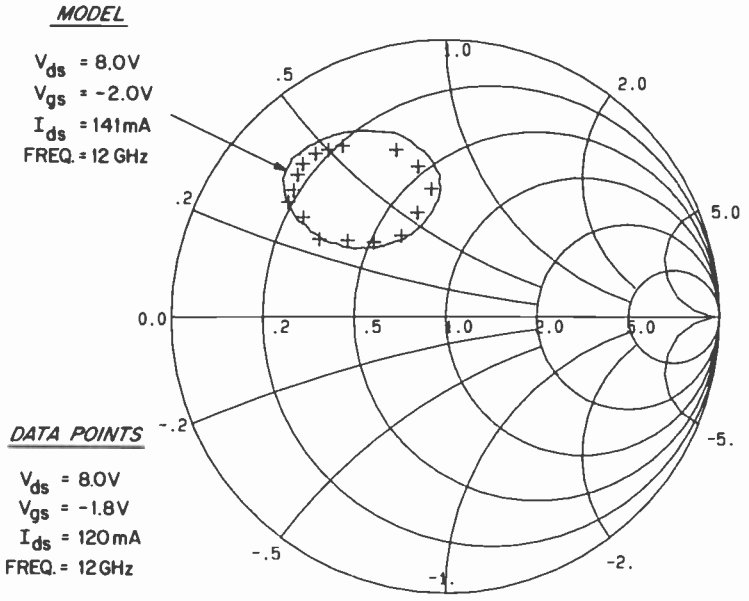


Fig. 10—Smith chart display of calculated and measured rf output loads for constant output power of 175 mW for device B1824-1C at 12 GHz.

Comparison of the Nonlinear Model with the Two-Dimensional FET Model

A direct comparison was made between the large-signal operation of the nonlinear FET model and that of an accurate 2-D time-domain model for the GaAs FET. The 2-D model[12] includes carrier-heating effects that produce velocity overshoot in GaAs. In addition, the 2-D model produces voltage waveforms that are not restricted in harmonic content. Thus, it is possible to evaluate the accuracy on the nonlinear program with regard to the harmonic content of the voltage waveforms.

We used rf input voltage amplitudes of 0.75, 1.5, and 2.5 V to make time-domain simulations for 1.5 rf cycles with the 2-D model and circuit. Each simulation required considerable computational time. The voltage waveforms were seen to be repeating, indicating that steady state was reached. The current is not calculated as accurately as the voltage, and has computational fluctuations that are not physically meaningful. Each voltage waveform was then Fourier analyzed to find its harmonic content for comparison with the nonlinear model.

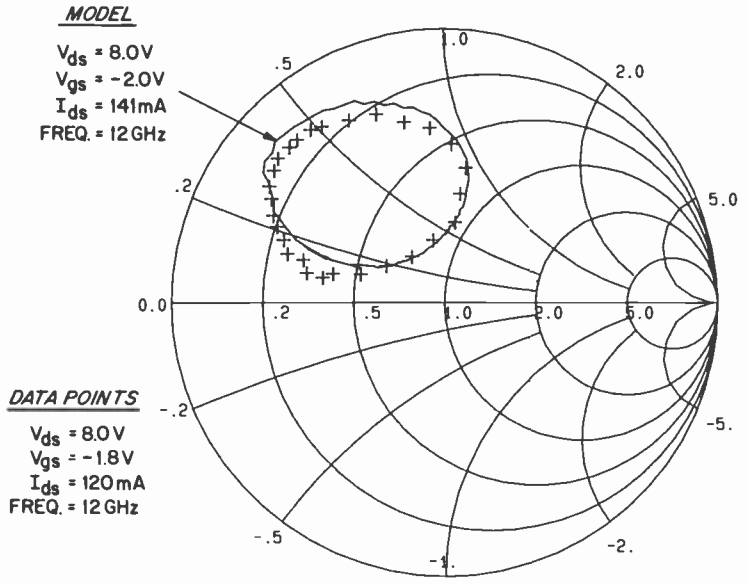


Fig. 11—Smith chart display of calculated and measured rf output loads for constant output power of 150 mW for device B1824-1C at 12 GHz.

The 2-D model was also used to calculate the steady-state drain current as a function of gate-source and drain-source voltages. The coefficients for an analytic approximation of $I_{ds}(V_{in}, V_{out})$ were then evaluated. Avalanche breakdown and gate-forward biasing were neglected.

Using these device parameters, we assumed a simple resistive rf load for the nonlinear FET amplifier program N-FET. The solid lines in Fig. 13 are values of harmonic power delivered to the load as calculated by the nonlinear program; the points are results of the 2-D simulation program.

The agreement is excellent for the fundamental output power. The agreement for second- and third-harmonic output power is good except at the lowest input power. Here the harmonic power calculation is clearly inaccurate in the 2-D program because all harmonic powers have the same value. This is a nonphysical result. As only three harmonics are used in the nonlinear program, the fourth harmonic is not evaluated.

Notice particularly that the third-harmonic output power dominates in the GaAs FET. This effect has been observed experimen-

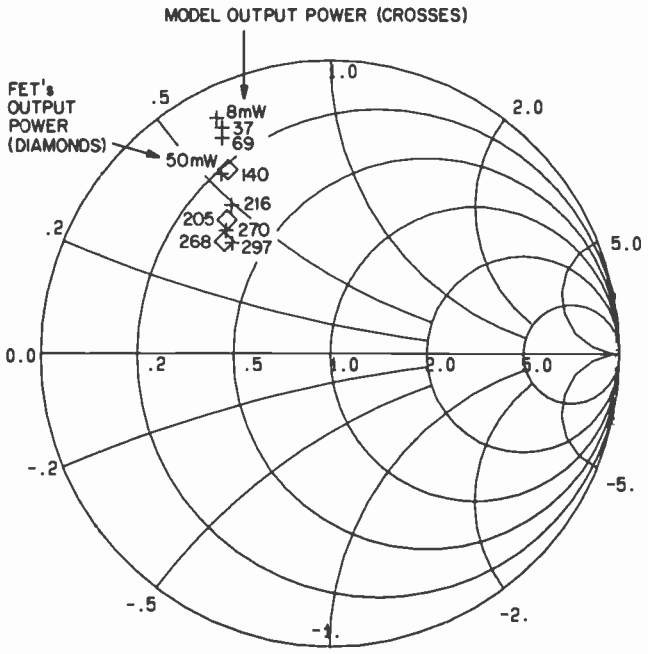


Fig. 12—Smith chart display of calculated and measured optimum rf output loads for maximum rf output power for device B1824-1C at 12 GHz.

tally by Willing et al.[18] and leads, we believe, to important third-order intermodulation distortion (IMD) in GaAs FET amplifiers. In fact, the analytical model, with some further programming effort, could be used to evaluate the IMD.

Fig. 13 clearly shows that the nonlinear program accurately predicts saturation effects that are due to the nonlinear current control characteristic; it also accurately predicts harmonic power output when harmonic impedances are known. Once the parameters of the nonlinear model were obtained it took only 1/300 of the computation time to generate the curves of Fig. 10 than it did to calculate these three cases by use of the 2-D model. The nonlinear model is clearly far more efficient.

Conclusion

We have developed a GaAs FET model suitable for efficient calculation in the large-signal region. It is useful for developing opti-

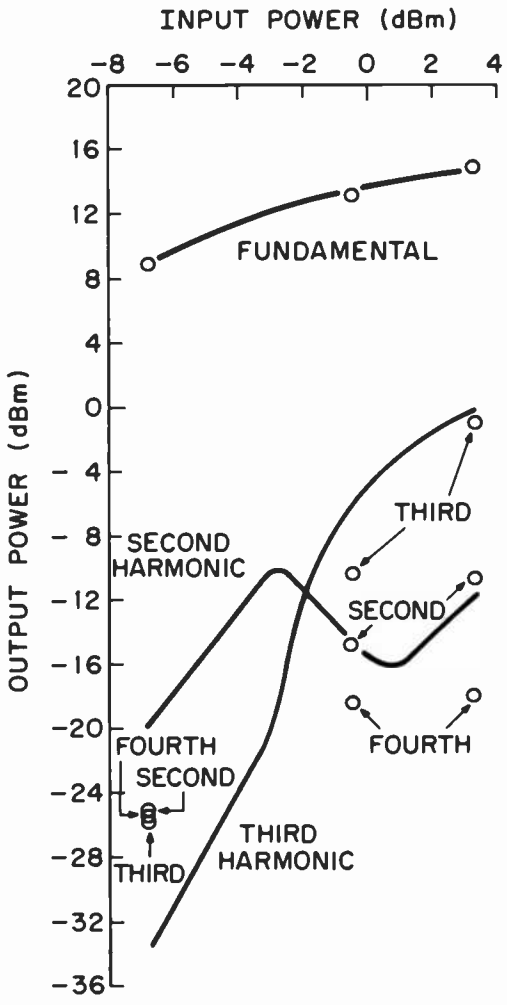


Fig. 13—Output power at fundamental and second- and third-harmonic frequencies as a function of input power. Solid lines are from N-FET, and points are from the 2-D model.

mized output network designs for high-power GaAs FET amplifiers. The program efficiency results from the use of the harmonic balance technique wherein the nonlinear FET is analyzed in the time domain and the linear circuit is analyzed in the frequency domain.

The principal nonlinearities of the FET are voltage-controlled current sources. The nonlinearity of the reactive elements does not greatly affect the large-signal solution. However, it is necessary to evaluate the characteristics of the current sources for each device

to be simulated and to develop an accurate small-signal model. The simulation can be performed with voltage waveforms containing fundamental and second-harmonic frequencies or fundamental and second- and third-harmonic frequencies. All FET current harmonics are included. Third-harmonic voltages are used only when accurate circuit impedance data are available at third-harmonic frequency.

The nonlinear FET model was coupled to a program to generate the load required for constant output power contours on a Smith chart. Excellent agreement was obtained with the measured load-pull characteristics at 12 GHz. However, the simulation predicted more gain than was measured in the experiments.

Finally, we compared N-FET with large-signal simulations made by use of an accurate two-dimensional model for the GaAs FET. This model includes carrier-heating effects that produce the phenomenon of velocity overshoot. The harmonic power contents were found to be in good agreement.

Acknowledgments

The authors are indebted to several of their colleagues at RCA Laboratories: to Dr. B. S. Perlman, for guidance in the program and for many helpful discussions; to S. M. Perlow, for making the load-pull measurements; to D. L. Rhodes, for the graphics software; and to P. D. Gardner, for device measurements.

References:

[1] Yarman, B. S., and Carlin, N. J., "A Simplified 'Real Frequency' Technique Applied to Broadband Multistage Microwave Amplifiers," *IEEE Trans. Microwave Theory Tech.* **MTT-30,** 2216–2222 (Dec. 1982).

[2] Ku, W. H., "Exact Synthesis of Interstate Matching Network for Broadband Microwave GaAs FET Amplifiers," *IEEE Int. Microwave Symp.* (San Diego, CA) *Dig.* **MTT-S,** 312–315 (1977); IEEE Cat. No. 77CH1219-5MTT.

[3] Takayama, Y., "A New Load-Pull Characterization Method for Microwave Power Transistors," *IEEE Int. Microwave Symp.* (Cherry Hill, NJ) *Dig.* **MTT-S,** 218–220 (1976); IEEE Cat. No. 76CH1087-6MTT.

[4] Nagel, L. W., "SPICE 2: A Computer Program to Simulate Semiconductor Circuits," Electronics Research Laboratory, University of California, Berkeley, Mem. No. EPL-M520, May 9, 1975.

[5] Davis, C. B., and Payne, M., "The R-CAP Program, an Integrated Circuit Simulator," *RCA Engineer* **21**(1), 66 (1975).

[6] Curtice, W. R., "A MESFET Model for Use in the Design of GaAs Integrated Circuits," *IEEE Trans. Microwave Theory Tech.* **MTT-28**(5), 448–456 (May 1980).

[7] Sussman-Fort, S. E., et al., "A Complete GaAs MESFET Model for SPICE," *IEEE Trans. Microwave Theory Tech.* **MTT-32,** 471–473 (1984).

[8] Fukui, H., "Determination of the Basic Device Parameters of a GaAs MESFET," *Bell Syst. Tech. J.* **58**(3), 771–797 (March 1979).

[9] COMSAT General Integrated Systems, Inc., SUPER-COMPACT™ *User Manual,* 1982.

[10] Curtice, W. R., and Camisa, R. L., "Self-Consistent GaAs FET Models for Amplifier Design and Device Diagnostics," *IEEE Trans. Microwave Theory Tech.* **MTT-32,** 1573–1578 (Dec. 1984).
[11] Golio, J. M., Blakey, P. A., and Grondin, R. O., "A General CAD Tool for Large-Signal GaAs MESFET Circuit Design," *IEEE Int. Microwave Symp.* (St. Louis, MO) *Dig.* **MTT-S,** 417–420 (1985); IEEE Cat. No. 85CH2163-4.
[12] Curtice, W. R., and Yun, Y. H., "A Temperature Model for the GaAs MESFET," *IEEE Trans. Electron Devices* **ED-28,** 954–962 (Aug. 1981).
[13] Camacho-Penalosa, C., "Numerical Steady-State Analysis of Nonlinear Microwave Circuits with Periodic Excitation," *IEEE Trans. Microwave Theory Tech.* **MTT-31**(9), 724–730 (Sept. 1983).
[14] Peterson, D. L., Pavio, A. M., and Kim, B., "A GaAs FET Model for Large-Signal Application," *IEEE Trans. Microwave Theory Tech.* **MTT-32,** 276–281 (1984).
[15] Gilmore, R. J., and Rosenbaum, F. J., "Modeling of Nonlinear Distortion in GaAs MESFETs," *IEEE Int. Microwave Symp.* (San Francisco, CA) *Dig.* **MTT-S,** 430–431 (1984); IEEE Cat. No. 84CH2034-7.
[16] Materka, A., and Kacprzak, T., "Computer Calculations of Large-Signal GaAs FET Amplifier Characteristics," *IEEE Trans. Microwave Theory Tech.* **MTT-33,** 129–135 (Feb. 1985).
[17] Curtice, W. R., and Ettenberg, Morris, "A Nonlinear GaAs FET Model for Use in the Design of Output Circuits for Power Amplifiers," to be published in *IEEE Trans. Microwave Theory Tech.* **MTT-33** (Dec. 1985).
[18] Willing, H. A., Rausher, C., and De Santis, P., "A Technique for Predicting Large-Signal Performance of a GaAs MESFET," *IEEE Trans. Microwave Theory Tech.* **MTT-26,** 1017–1023 (Dec. 1978).

New Algorithms for the Automated Microwave Tuner Test System

Stewart M. Perlow
RCA Laboratories, Princeton, NJ 08540

Abstract—Optimization of amplifier performance characteristics requires the termination of the active device in the proper source and load impedances. With the exception of small-signal amplifiers designed for gain, the correct terminating impedances are found by tuning. The programmable microwave tuner, along with its associated software, represents a very rapid and efficient means of obtaining the proper impedance values. This paper describes new algorithms that improve the search procedures and allow the load terminating the microwave tuner to be any arbitrary impedance of known value.

Introduction

Amplifiers can be designed to have optimum performance in one specific characteristic (such as noise figure, gain, output power or efficiency) or design tradeoffs can be made to obtain performance in more than one of these areas. These design goals can be realized only if the source and load impedances, terminating the active device, are the values required to obtain the desired performance.

S-parameter characterization of an active device supplies enough information to determine the correct source and load impedances for optimum gain performance in a small-signal amplifier. When optimizing other characteristics or in the large-signal case, the required information may be obtained experimentally by varying adjustable matching circuits or tuners until the desired performance is obtained. The technique usually employed with a manual tuner consists of attaching it to a device under test and positioning the slugs for some desired circuit performance. The tuner is then removed and its input impedance is measured on a network analyzer.

This is a very time consuming process, especially in the typical case where many data points and frequencies are required.

To overcome this difficulty, the slugs can be driven by motors controlled by a computer.[1] In this way the exact position of each slug is known by the computer. Applying this position information to a software model of the tuner results in a very precise calculation of the input impedance of the tuner. Also, the software can easily be used to change reference planes and to de-embed various circuitry, such as the test fixture.

It is extremely difficult to determine whether the impedance obtained manually truly represents the best circuit performance. If additional instrumentation capable of monitoring essential performance parameters is interfaced to the computer, software algorithms can be utilized to feed back information so that the slugs move in a prescribed pattern. An example of this would be the use of a power meter to sense output power. The measurement can be used to search for the maximum or a specific output power and to find all the output impedance points that produce the same output power (constant power contours).

The automated tuner consists of a transmission line with two moveable slugs, the associate hardware and software to move these slugs, and the instrumentation and software required for sensing the desired output parameters (e.g., power meters for a maximum power search or constant power contours).[2]

The purpose of the transmission line and moveable slugs is to provide an impedance at a specified reference plane. The software algorithms associate this impedance with the output parameter sensed by the instrumentation hardware and also govern the required logical search procedures.

Double-Slug Tuner

This section discusses the operating principles of the double-slug tuner. The tuner consists of either a uniform coaxial transmission line or a slabline with a characteristic impedance, Z_o, of 50 ohms. The line contains two identical moveable slugs that slide between the inner and outer conductors as shown in Fig. 1. The sections of line containing the slugs have a characteristic impedance, Z_s, which is lower than Z_o.

The slugs may be made of dielectric material, in which case the impedance Z_s is

$$Z_s = Z_o/e_r, \qquad [1]$$

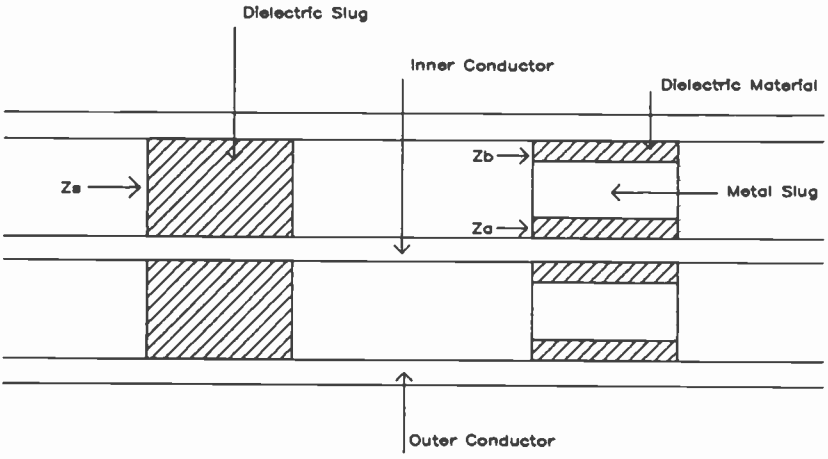


Fig. 1—Schematic of double-slug tuner. Left shows slugs made of dielectric material and the right shows slugs of a conductive material (metal) sandwiched between dielectric material.

where e_r is the relative dielectric constant of the material. The alternate approach is to use a conductive slug sandwiched between dielectric material (also shown in the Fig. 1).[3] In this case, the slug section of the line is composed of two transmission lines connected in a series configuration. The innermost line shares its center (inner) conductor with the entire tuner, while its outer conductor is the conductive (metal) slug. This conductive slug is also the inner conductor of the outermost line. The outermost line shares its outer conductor with the outer conductor of the entire line. The impedance Z_s is

$$Z_s = Z_a + Z_b, \qquad [2]$$

where Z_a and Z_b are the impedances of the innermost and outermost lines, respectively. The dielectric material used to support the conductive slug also reduces the values of Z_a and Z_b which, as will be shown, increases the impedance range of the tuner.

The entire line including the slugs is shown schematically in Fig. 2 as five cascaded transmission lines. The impedance of the two exterior lines and the center line is Z_o. The impedance of the two lines composed of the slugs is Z_s. If the last line (l5) is terminated in its characteristic impedance, $Z_1 = Z_o$, the second slug (l4) will be terminated in Z_o. Note that as long as the line is terminated in its characteristic impedance, the termination on l4 is always Z_o regardless of the length of l5.

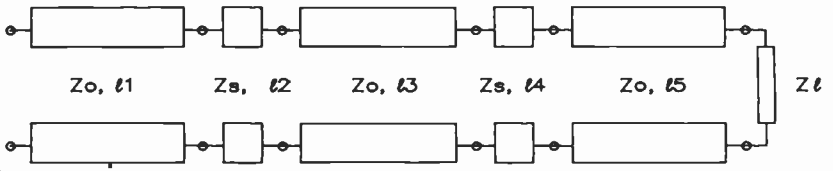


Fig. 2—Schematic of double-slug tuner transmission line. The impedance of the two exterior lines and the center line (I1, I3, I5) is Z_o and of the slugs (I2, I4) is Z_s.

The magnitude of the reflection coefficient is determined by the spacing of the slugs. At some separation, d_{min} (dimensioned in wavelengths), which is a function of the slug impedance and slug electrical length, the reflection coefficient is 0.0. In the case where the slugs are exactly one-quarter-wavelength long, this occurs when the slugs are together. Plotted on the Smith chart, the impedance would appear at the center of the chart. As the slug separation is increased, the magnitude of the reflection coefficient increases rapidly, reaching approximately half its maximum amplitude at a separation of d_{min} + 0.03 wavelengths and ninety percent at d_{min} + 0.1 wavelengths.

The magnitude reaches a maximum at a quarter wavelength beyond the separation at which the minimum occured. Beyond this separation, the magnitude begins decreasing, reaching 0.0 at a separation of d_{min} + 0.5 wavelengths.

The relative angle of the reflection coefficient is determined by the distance between the input reference plane and the position of the slug closest to that reference plane (the length of section l1). If the slug separation is held constant, the magnitude of the reflection coefficient does not change. As the two slugs are moved together, the phase angle of the reflection coefficient changes, because the distance between the slugs and the reference plane is changing.

The maximum VSWR occurs when the slug spacing is an odd multiple of a quarter wavelength and the slugs themselves are also an odd multiple of a quarter wavelength. Under these conditions the maximum VSWR is given by

$$\mathrm{VSWR}_{\max} = (Z_o/Z_s)^4. \qquad [3]$$

In general for a line with n slugs,

$$\mathrm{VSWR}_{\max} = (Z_o/Z_s)^{2n}. \qquad [4]$$

Fig. 3 shows a tuner's input impedance points plotted on a Smith chart. The curve identified as "CS = 0" represents the input imped-

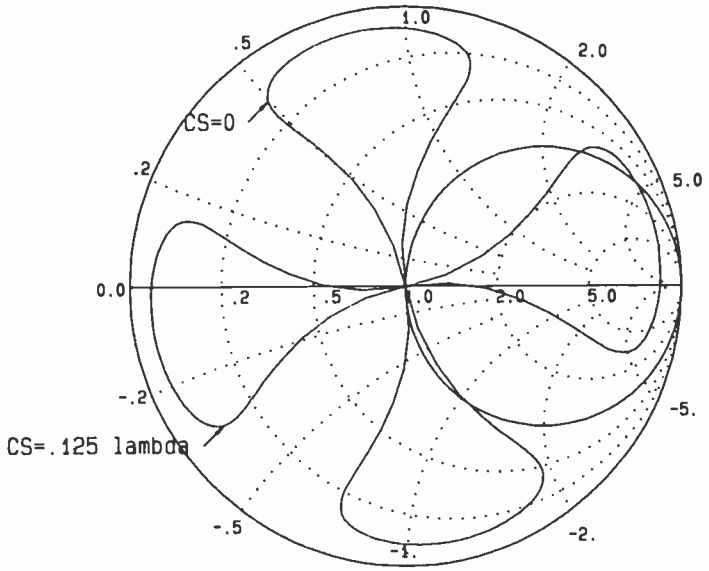


Fig. 3—Double-slug tuner input impedance points. CS = 0 represents input impedance for slugs with characteristic impedance of 25 ohms and one-quarter wavelength. CS = .125 lambda represents impedance when slugs are moved 1/8 wavelength toward input.

ance for a tuner that has slugs which have characteristics impedances of 25 ohms and are a quarter-wavelength long. The slug separation at the center of the chart is 0. As the slug separation increases, the input impedance follows the right portion of the upper curve. The slug separation at the outside edge of the cloverleaf is a quarter wavelength. The second curve, identified as "CS = .125 lambda" represents the input impedance when both slugs are moved an eighth wavelength toward the tuner's input. Note that the entire curve rotates around the Smith chart's origin.

Fig. 4 shows the effect of changing the length of the slugs. It is interesting to see that the shorter slugs produce a much narrower pattern than the longer slugs. This effect occurs because the impedance range of the tuner is limited to smaller reflection coefficient magnitudes. If the slugs were made very short but lower in impedance, so that the $VSWR_{max}$ was the same as quarter-wavelength slugs of some higher impedance (e.g., 0.0125 wavelengths and 2.61 ohms versus 0.25 wavelengths and 25 ohms, both of which have $VSWR_{max}$ of 16), the impedance plots on the Smith chart would be identical. An important point to note is that this situation is true at only one frequency. As the frequency is increased, the impedance

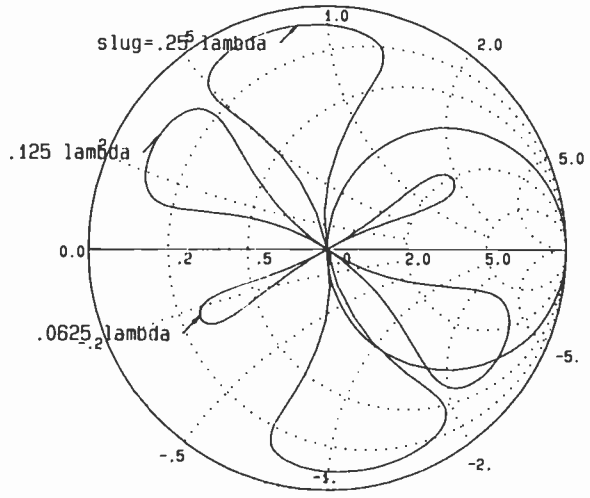


Fig. 4—Double-slug tuner input impedance points showing changes as the length of slugs is shortened.

plots will differ considerably. At twice the frequency, the quarter-wavelength slugs will be half a wavelength long and transparent, and the impedance of this tuner will be a constant 50 ohms. The very short slugs will still be short, even at twice the frequency. Therefore, the impedance diagram of this tuner (with the short slugs) will change very little, with the exception of phase angle.

The previous discussion assumes that the tuner is terminated in its characteristic impedance. This assumption eliminates the length of l5 as a parameter. If the load is not exactly equal to the characteristic impedance of the line, the impedance terminating the slug closest to the load is a function of the length of l5 rather than a constant 50 ohms. The effect on the tuner's input impedance is that movement of both slugs together does not result in a simple rotation of the impedance plot around the center of the Smith chart. The point of rotation is offset from the center of the chart and the fan shaped impedance plot distorts as it rotates. This is shown in Fig. 5 for a load reflection coefficient of .15 at an angle of 90° (VSWR = 1.35:1). The VSWR, or the magnitude of the input reflection coefficient, is no longer dependent on the slug separation alone. It is now a function of the position of the slugs, as is the phase angle of the reflection coefficient. Fig. 6 shows the locus of input reflection coefficient points as the slugs are moved together along the line for various values of slug separation. Although the contours appear to

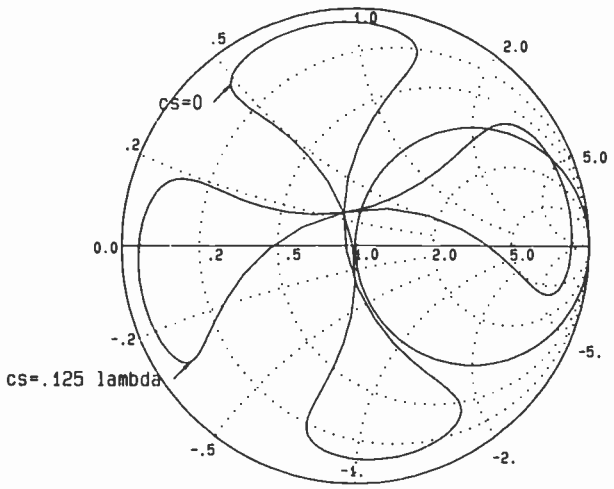


Fig. 5—Impedance plot for load reflection coefficient of .15 and angle of 90° (VSWR = 1.35:1).

be true circles they are not. Only if the load impedance is exactly the characteristic impedance of the line will they be true circles.

Impedance—Slug-Position Algorithms

To make the double-slug tuner an effective device, the impedance at the input of the tuner must be known.

The tuner model shown in Fig. 2 is used to calculate the input impedance. The parameters required for this model consist of the line and slug impedances and the electrical length of the various line sections. Certain fixed parameters are obtained using a calibration procedure: the minimum slug separation, the lengths between the input and output reference planes of the tuner and the slugs at minimum separation, the length of the entire line, the lengths of the slugs, and the impedance of the slugs. The actual position of the slugs is obtained by translating the number of steps the motors move into linear dimensions.

The slug positions are represented by two variables, the separation between the slugs, S, and the distance between the center of symmetry (the point midway between the two slugs) and an arbitrary reference point, CS. These values are shown in Fig. 7. When both slugs move together, S remains constant and CS varies. The resulting Smith chart plot is a circularly shaped constant-separa-

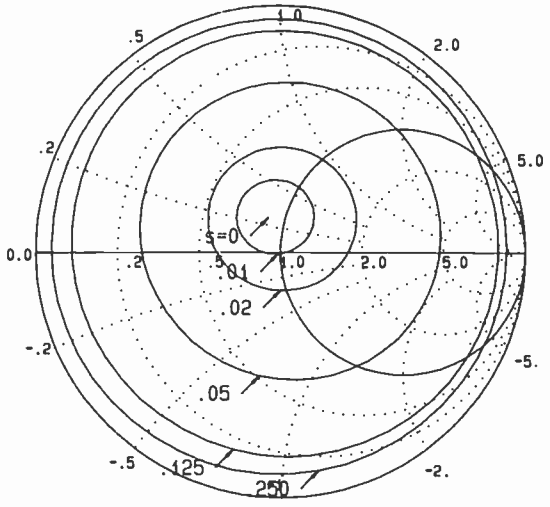


Fig. 6—Locus of input reflection coefficients as slugs are moved along transmission line for various values of slug separation.

tion contour. If one slug moves a distance d in one direction and the other moves the same distance in the opposite direction, CS remains constant but S changes. The resulting Smith chart plot is a clover leaf shaped constant CS contour. Any movement of the two slugs may be resolved into the component S and CS values.

$$S = Y - X \qquad [5]$$

$$\text{CS} = (X + Y)/2 \qquad [6]$$

where X is the distance from the reference point to the left slug and Y is the distance from the reference point to the right slug.

When the slug position is known, the calculation of the input impedance is very straight forward. It consists of either considering the tuner as a ladder-type network and calculating the input impedance or cascading the *ABCD* parameters of the five line sections with the load and solving for the input impedance. Any change in reference plane is accomplished by transforming the resultant impedance through a set of *S*-parameters for the network, representing the shift in reference plane.

The reciprocal process of obtaining the required slug positions for a desired input impedance is not as simple. One technique that has been employed consists of taking an initial educated guess at the correct slug positions. This guess is based on the simplified approximation that the magnitude of the reflection coefficient varies si-

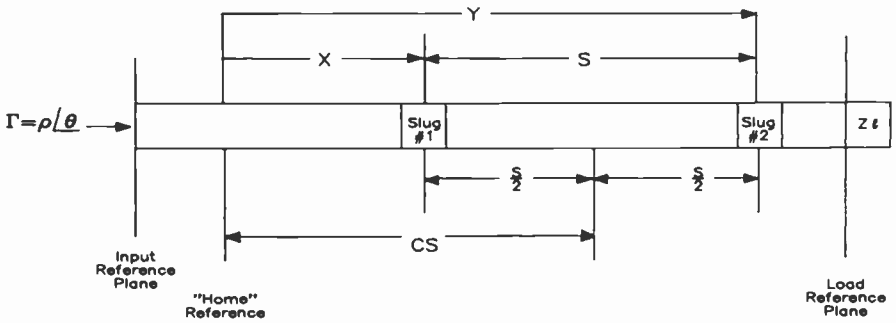


Fig. 7—Various parameters used to represent slug separation, *S*, and distance between center of symmetry (midway point between slugs) and arbitrary reference point, CS.

nusoidally with slug position.[2] The actual reflection coefficient at these slug positions is calculated using the model of the tuner. An error vector is determined as the difference between the desired input reflection coefficient and the calcualted reflection coefficient. One half of this error vector is then added to the desired reflection coefficient. The resultant reflection coefficient, or new target, is used in the simplified approximation to establish new slug positions. Using the line model again, the actual reflection coefficient for these new positions is calculated and a new error vector established. This time the new target is the resultant of the previous target and one-half the error vector. The process is repeated until the magnitude of the error vector is less than a predetermined limit.

This technique is very effective when the tuner is terminated with a load that has an impedance very close to the characteristic impedance of the line. However, it can lead to severe problems when bias tees and other components are included in the terminating circuitry. The double-slug-tuner Smith chart plot of Fig. 5 would be a typical case in which the simplified approximation would not allow the determination of the slug positions near the center of the chart. Under these circumstances a more general procedure is needed.

A technique that allows arbitrary terminating impedances to be used is depicted in Fig. 8. An extremely poor load is used as an example for purposes of illustration.

The center of rotation, the point on the Smith chart about which the contours of constant slug separation (clover leaf pattern) rotate, has the same magnitude as the load's reflection coefficient. The angle is a function of the electrical lengths of the transmission line and slugs in addition to the angle of the load's reflection coefficient. The load for the present discussion has a reflection coefficient of

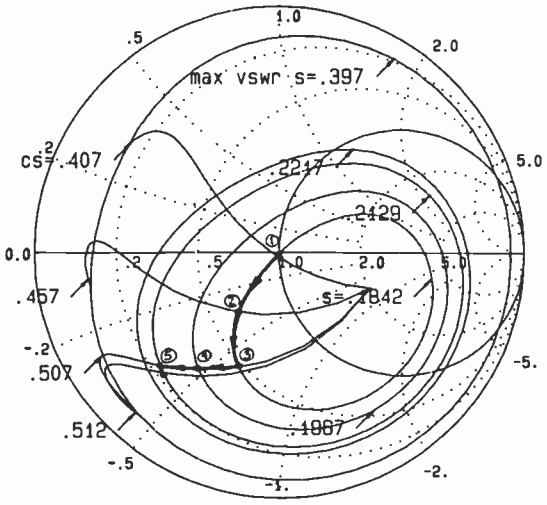


Fig. 8—Representation of technique using arbitrary terminating impedances to determine required slug positions for a desired input impedance.

$0.4\angle 90$. The slug separation at the center of rotation is 0.147 wavelengths. The desired reflection coefficient at the input to the tuner is $0.7\angle -145$. The initial starting point is completely arbitrary. In this example it was chosen at the center of the Smith chart with coordinates ($S = 0.1842$, $CS = -0.093$). The search proceeds by calculating the reflection coefficient at the input to the tuner at these coordinates (S, CS) using the tuner model. The magnitude of the error vector between the desired reflection coefficient and the present one is then determined. A small change in S is made by decrementing S by ΔS. The input reflection coefficient and error vector at (S-ΔS, CS) are calculated. If this error is smaller than the original one, the new S position is decremented by ΔS again. If not, S is incremented by ΔS and the same procedure is followed. If neither an increment or decrement of ΔS decreases the error, an increment in CS is tried. The result is to travel along a constant separation contour, which is egg shaped, or a constant center of symmetry contour, which is clover leaf shaped. The movement from Point 1 to Point 2 to Point 3 in Fig. 8 shows the travel with constant slug separation. Travel from Point 3 to Point 4 to Point 5 is along a constant CS contour. When changing S and CS no longer improves the error, the increment must be reduced. At Point 5, the value of ΔS was reduced and changed direction. This process is continued until the error is reduced to some predetermined limit or the incre-

ment becomes so small that it has no effect on the error. This latter condition could occur if the tuner is not capable of providing the requested reflection coefficient.

Maximum-Power Search Algorithm

The locus of all impedance values that provide the same output power, when used as a load on an amplifier, are in general closed egg-shaped contours on the Smith chart. The maximum output power is located at a single point and the contours of constant power surround this point. One approach to finding the maximum-power point is to calculate the power gradient at the starting point of the search. The search proceeds in the direction of the max gradient by using extrapolation techniques until the maximum power is found. Mathematically this method is very powerful. However, both the gradient calculation and the extrapolation use measured data. Any measurement error in this data is greatly magnified in the calculation process and subsequently can result in aimless wandering around the Smith chart.

A direct-search technique is shown in Fig. 9. The power is measured at arbitrarly chosen Point 1. A second point, Point 2, providing the same output power is found using the method that will be described in the section on constant-power contours. Since the

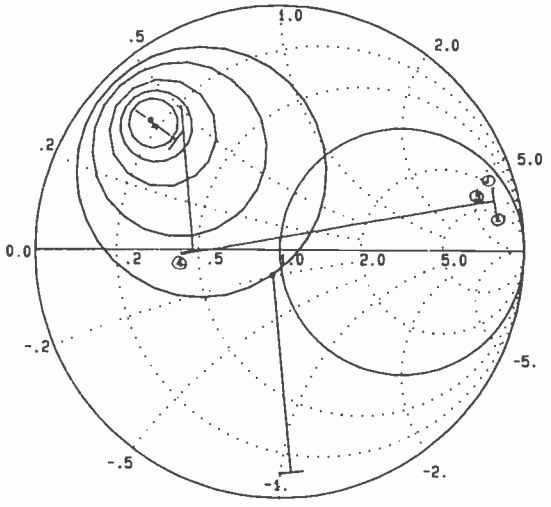


Fig. 9—Direct search technique used to determine maximum-power point and target value.

maximum-power point is located inside this segment of a constant-power contour, it is only necessary to determine which direction is the inside. To do this, reflection-coefficient points lying on the perpendicular bisector of this segment are set and the power measured. One of these points must have a lower power and the other a higher power than the two points on the contour. The proper search direction is obviously the direction of the point with the higher power, Point 3. The search is started in this direction. The next test point is determined by adding a search vector to the reflection coefficient at Point 3. The magnitude of the search vector is quite arbitrary. A typical value is 0.1, which means the entire Smith chart could be traversed with only ten measurements. The search continues by adding the search vector to each previous point, setting the tuner to the appropriate reflection coefficient, and measuring the power at each setting.

The direction of the search vector is maintained until the test point exhibits a decrease in power as shown at Point 4. At this point the adjacent point on a constant power contour is determined along with the perpendicular bisector. The preceding process is repeated again. If two adjacent points on a constant-power contour can not be obtained, it is because the search vector is too large; that is, the ends of the search vector extend beyond the contour circumference. If this occurs, the magnitude of the search vector must be reduced. The search continues until the magnitude of the search vector becomes insignificant. This typically occurs for a magnitude of 0.01.

Other Target Values

The search for any other power value can be accomplished in a manner similar to the search for maximum power. Since a target value is specified, the error between the power at any point in the search and the targeted value can be calculated. The search is initiated at an arbitrary starting point by finding two adjacent points on a constant-power contour. The powers at the two points forming the perpendicular bisector are then measured and the two corresponding error magnitudes are compared. The search proceeds in the direction of minimum error. The direction of the search vector is kept the same as long as the error magnitude continues to decrease. An increase in the error magnitude indicates that the direction of the search must be changed or that the contour containing the target value has been crossed.

To discriminate between these two possibilities a simple test is applied. If the targeted contour has been crossed, the sign of the

error will change. For example, if the initial starting point was lower in power than the target, the error will remain positive until the contour is crossed. At the search point on the other side of the contour, the power will be greater than the target and the error will become negative.

The results of this test determine the next steps in the search procedure. If the target contour has not been crossed, the direction of the search is changed by determining an adjacent point on a power contour and testing the points on the perpendicular bisector for minimum error. The search proceeds as before.

If the target contour has been crossed, the search proceeds by returning to the preceeding point (just before the contour was crossed). The magnitude of the search vector is decreased while maintaining the previous direction. The search vector's magnitude is continuously reduced in this manner until the error magnitude is reduced to zero or to some predetermined tolerance.

Constant-Power Contours

The initial reflection coefficient point, G_o, on the contour is obtained by setting the tuner manually or using a target value search. A power meter reading taken at this point is used as the reference power, P_o, for the contour. A search vector with magnitude, s_{mag}, and angle, s_{ang}, is added to the reflection coefficient of the first point. The tuner slugs are set, using the inversion algorithm, to obtain this new input reflection coefficient,

$$\Gamma = G_o + s_{mag} \angle s_{ang}. \qquad [7]$$

The output power at this point is read and the error is calculated as the difference between the power at the first point, P_o, and the new reading,

$$\text{error} = P_o - P. \qquad [8]$$

If the error is less than a predetermined tolerance, e_{tol}, the point is accepted as the second contour point.

The magnitude of the error vector and the value of e_{tol} should be selected with some care. If the values of e_{tol} or s_{mag} is made too small, the time required to plot a contour can become unnecessarily long. Making these values too large results in inaccurate contours with an insufficient number of points to obtain smooth curves. The relative values of s_{mag} and e_{tol} are also important. If e_{tol} is too large for s_{mag}, the search vector can bounce between the limits set by e_{tol}

producing a shark tooth contour. An acceptable ratio of s_{mag} to e_{tol} appears to be

$$s_{mag}/e_{tol} = 10, \quad [9]$$

while a good starting value for s_{mag} is 0.1 and for e_{tol} is 0.01. A very small contour near the maximum power point would require a reduction in s_{mag} to perhaps 0.02 and e_{tol} to 0.002.

The initial angle of the search vector, s_{ang}, is arbitrary, but a slight advantage may be obtained by making it perpendicular to the vector representing the initial reflection coefficient.

The search proceeds by rotating the search vector around the last acceptable contour point. At each test point, the tuner is set for the proper reflection coefficient, and the error is calculated from the power at that point. That the power on one side of the contour is lower than the reference power and higher on the other side means that the error has a different sign on each side. This fact can be used in setting the search angle to locate the next contour point in the following manner. Set the search angle to $-150°$, $0°$ and $150°$. At each point calculate the error. If the sign of the error changes between $0°$ and $150°$, the contour must lie between these two angles; therefore chose the next angle as 150/2 degrees. Repeat this process until the error is within the limits of e_{tol}.

The number of measurements can be reduced by using a linear extrapolation for the next angle. It is very important to realize that small measurement errors can produce very large errors in the extrapolated angle. To eliminate the possibility of the search vector spinning around in circles and going nowhere, carefull checks must be made during each extrapolation. If the search angle, s_{ang}, is assumed to be a linear function of the error, the angle at which the error should become zero is:

$$s_{ang}(3) = s_{ang}(1) - \frac{\text{error}(1)[s_{ang}(2) - s_{ang}(1)]}{\text{error}(2) - \text{error}(1)} \quad [10]$$

where (1) indicates the values at the first test point and (2) indicates the values at the second test point, etc.

The change in s_{ang}, Δs_{ang} is given by

$$\begin{aligned} \Delta s_{ang} &= s_{ang}(3) - s_{ang}(1) \\ &= -\frac{\text{error}(1)[s_{ang}(2) - s_{ang}(1)]}{\text{error}(2) - \text{error}(1)} \end{aligned} \quad [11]$$

If the magnitude of Δs_{ang} exceeds $150°$, the search can backtrack on itself. The search strategy that proved to be the most effective

is to calculate the extrapolated value first. If the magnitude of Δs_{ang} is less than 150°, s_{ang} (3) is read as the angle of the next search vector; if not, we switch to the sign-change method. This strategy combines the speed of extrapolation with the security against spinning provided by the sign-change technique.

A possible problem can occur as the contour approaches the edge of the Smith chart. If the contour requires a reflection coefficient that the tuner is not capable of providing, the search algorithm must sense that the maximum reflection coefficient has been reached. The search stops at that point and moves back to the starting point. To guarantee that the search proceeds in the opposite direction, the search angle must be made equal to the search angle connecting the initial and second contour points shifted by 180°.

The contours may be closed or they may run off the edge of the Smith chart, in which case they are open. In the case of a closed contour, the contour can be considered complete by comparing the vector representing the currently accepted point to the vector representing the original starting point. If the magnitude of the difference vector is less than the 0.05, the contour can be considered complete. The completion of an open contour occurs when the contour reaches a maximum reflection coefficient for the second time (the first time, the search returned to the initial point and went in the opposite direction).

Conclusions

The computer controlled automated tuner represents a significant improvement in amplifier design procedure, since a large number of data points can be taken in a short amount of time.

The new impedance inversion algorithm used to obtain the slug positions for the desired input reflection coefficient is fast and efficient. Since it does not make use of a simplified model, it will work for an arbitrary load terminating the tuner, even when the magnitude of the reflection coefficient of the load approaches unity.

References:

[1] J. M. Cusack, S. M. Perlow, and B. S. Perlman, "Automatic Load Contour Mapping for Microwave Power Transistors," *IEEE Trans. Microwave* Theory and Tech., **MTT-22,** No. 12, Part II, p. 1146, Dec. 1974.
[2] F. Sechi, R. Paglione, B. Perlman, and J. Brown, "A Computer-Controlled Microwave Tuner for Automated Load Pull," *RCA Rev.* **44,** No. 4. p. 566, Dec. 1983.
[3] T. Moreno, *Microwave Transmission Design Data,* Dover Publications, Inc., New York (1948).

Simulation of Microlithographic Resist Processing Using the SAMPLE Program

Dietrich Meyerhofer
RCA Laboratories, Princeton, NJ 08540

Abstract—Modeling plays an important part in the design and processing of integrated circuits. For example, it allows one to simulate a wide variety of processing conditions and determine their effects on device performance without having to perform many time-consuming experiments. The program SAMPLE was developed at the University of California at Berkeley to simulate the exposure and development of photoresist films. We have used this program to compare various processes, to try to understand experimental observations, and to determine improved techniques for processing wafers.

Applying the SAMPLE program to make realistic predictions about resist processing required modifications and extensions of the program. A model of proximity printing was added to show how diffraction effects can cause very complicated intensity patterns at the resist surface, depending on the ratio of image size to proximity gap. Then the proximity-printing algorithm was applied to the case of conformable mask printing, where the mask is in very close contact with the resist layer.

The exposure of the resist layer depends strongly on the optical properties of the material. We discuss what happens when nonbleaching dyes are added to resist films to reduce interference effects. The exposure algorithm has also been extended to the newly developed bilayer system, which consists of a contrast enhancement layer containing a bleachable dye on top of a resist film.

When the development process is modeled, cross sections of developed resist profiles are obtained. They are used to compare the dimensions of the various resist features after development with the corresponding dimensions on the object mask and to determine the fidelity of the reproduction. Another application is to compare the shape of resist profiles that can

be produced in positive and negative resists by the same kind of exposure. Finally, predictions are generated about the latitude to exposure dose that can be expected for various resists and processes and about the sensitivity of linewidth to film-thickness changes.

Introduction

Systems, processes, and procedures are often simulated by mathematical models that require numerical calculations. A model calculation is usually cheaper than an experimental run, and more variations can be explored through calculations than through experiments. A wider range of parameter values is available for investigation, and more realistic situations can be simulated. Of course, this presupposes that the model is a good one, representing the real situation closely enough to be useful. This has to be established by comparing the model with experimental observations.

In this paper we discuss the program SAMPLE that simulates the microlithographic techniques and processes required as the first step in every level of integrated circuit processing. SAMPLE (a program for the *S*imulation *A*nd *M*odeling of *P*rofiles for *L*ithography and *E*tching)[1] was developed at the University of California, Berkeley. It is in the public domain and is widely used in the industry. SAMPLE includes the following steps: the application of the resist film to the substrate, which usually is not smooth but contains varying topography from the previous steps; the illumination of the film with an optical image and its response to this exposure; and the development of the resist to form profiles on the substrate which will mask the underlying layers during the subsequent processing steps. SAMPLE also is capable of modeling electron-beam exposure, deposition of materials, and dry etching, but we will not be discussing those parts of the program in this report.

We obtained our first version of SAMPLE in January 1980 and have received other, upgraded versions since then. We have used the program for simulating many different photolithographic situations. To enable us to apply SAMPLE successfully we have made many changes in the program and have developed new algorithms for specific cases. Use of the program has proved very useful in increasing our understanding of the photolithographic variables and processes and in allowing us to make predictions about the usefulness of proposed modifications to resist systems.

As an example consider the consequences of varying the thickness of resist films. The topography of a resist film spun on a substrate with steps on it has recently been characterized in detail and modeled.[2] Sizeable variations in resist thickness are expected on either side of a step in the substrate. To make use of this information, we have determined the changes in linewidth with resist thickness for a variety of substrates and resist processes. This allows us to make suggestions for improving the linewidth control over steps. Another application has been to evaluate the promise of novel and more complicated resist systems consisting of multiple layers, such as a contrast enhancement system. We have also been interested in the effects of changing the absorption and reflection characteristics of the resist layer and the substrate.

In this paper we first present a brief description of the SAMPLE program and then list the modifications and additions we have made to it. This is followed by a discussion of applications to various practical situations.

The Berkeley "SAMPLE" Program

The SAMPLE program has been under development for many years at the University of California, Berkeley. Initially, it modeled the imaging of patterns, the resultant resist exposure, and the development of the resist. The algorithm was based on an ad hoc model of diazoquinone-based positive resists developed by Dill et al.[3] Since then, increasingly more sophisticated versions of the calculations have been made available. Etching and deposition programs, as well as an electron-beam exposure model, were added. The calculations are all limited to one dimension; the program cannot handle two-dimensional features, such as contact openings. In this report we restrict ourselves to a discussion of those parts of SAMPLE that simulate photolithographic processes.

The photolithographic model of SAMPLE consists of three separate parts: the optical image formation, the exposure of the photosensitive layer, and the development of the resist. Image formation is calculated by standard optical equations of diffraction-limited imaging. It allows for coherent, incoherent, or partially coherent illumination. The features on the object vary only in the x dimension and remain fixed in the y dimension. They may be long lines, spaces, or combinations of the two. As an example, Fig. 1 shows the intensity distribution along the x axis in the image plane of a diffraction-limited lens as a function of coherence of the illuminating radiation.

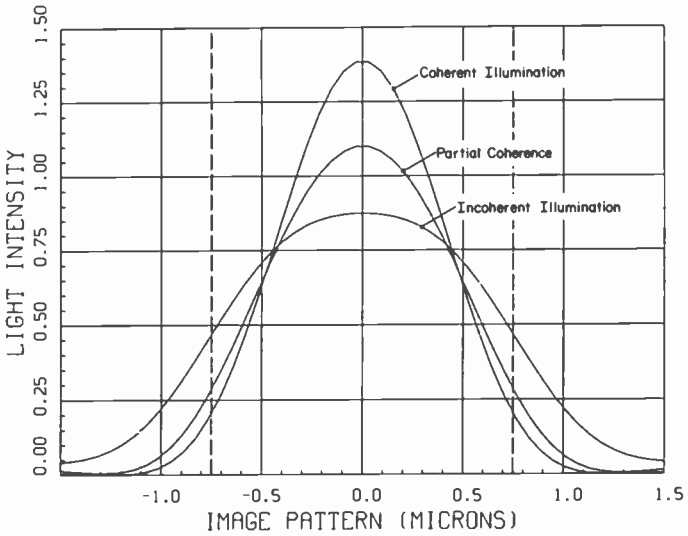


Fig. 1—Imaging properties of a diffraction-limited lens at 436 nm. Three different illumination conditions are compared, full coherence, incoherence, and partial coherence with $\sigma = 0.7$. The lens is diffraction limited with $NA = 0.28$. The object is a 1.5-μm-wide space. The dashed line indicates where the edge of a perfect image would be. The light intensity is normalized to the value of the incident uniform illumination.

A typical stepper lens is used with a numerical aperture (NA) of 0.28 and an illuminating wavelength of 436 nm. The object is a single space of 1.5-μm width in an opaque field. As expected, the image sharpens dramatically as one proceeds from incoherent to coherent illumination. At the same time oscillations can be seen in the intensity patterns at distances further away from the center than shown in Fig. 1 due to interference effects ("ringing"). The partially coherent case with fill factor $\sigma = 0.7$ is representative of the situation found in wafer steppers. It provides a good compromise between the two extremes.

SAMPLE also provides for calculating image patterns in out-of-focus conditions. Some results are shown in Fig. 2. Three images are compared, one located in the focal plane, the second displaced by one Rayleigh depth of focus ($\lambda/2[NA]^2$), and the third displaced by twice that amount. Partially coherent illumination was assumed. Fig. 2 shows that edge distortions of approximately 0.2 μm are to be expected if the focus is not controlled to within 1 Rayleigh value.

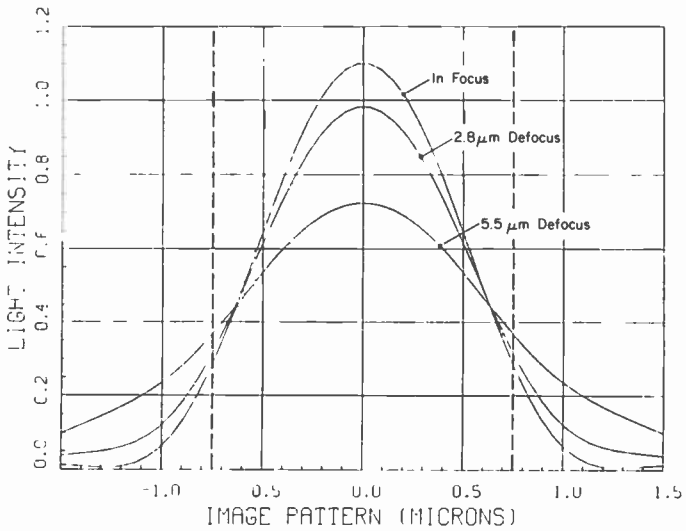


Fig. 2—The effect of focus on the quality of an image. The illumination, lens, and object are the same as in Fig. 1, with partially coherent illumination. The amount of defocus, Δf, is indicated.

Exposure

The exposure program models the photochemical reactions occurring throughout the resist film. In conventional positive resists the photoactive compound is a diazoquinone sensitizer, which absorbs the imaging light and, after chemical reaction, becomes almost completely transparent at the exposing wavelength. The light intensity within the film depends on the illumination pattern and on the amount of light reflected from the substrate and from layers deposited on it, as well as on the values of the various complex indices of refraction. As the photoactive compound is bleached, the absorption of the resist film changes locally. This, in turn, produces changes in the light distribution. The calculation is performed by dividing the exposure time into small increments. After each exposure increment the indices are changed, depending on how much bleaching has taken place, and the light intensity values are recalculated. The exposure of the resist is characterized by absorption and bleaching parameters that are determined by the diazoquinone sensitizer species in the resist. In Fig. 3 we show the result of such a calculation. The fraction of sensitizer that is not bleached is plotted against the location in the resist film, where x is in the film plane, and z at right angles to it. The projected image illuminating

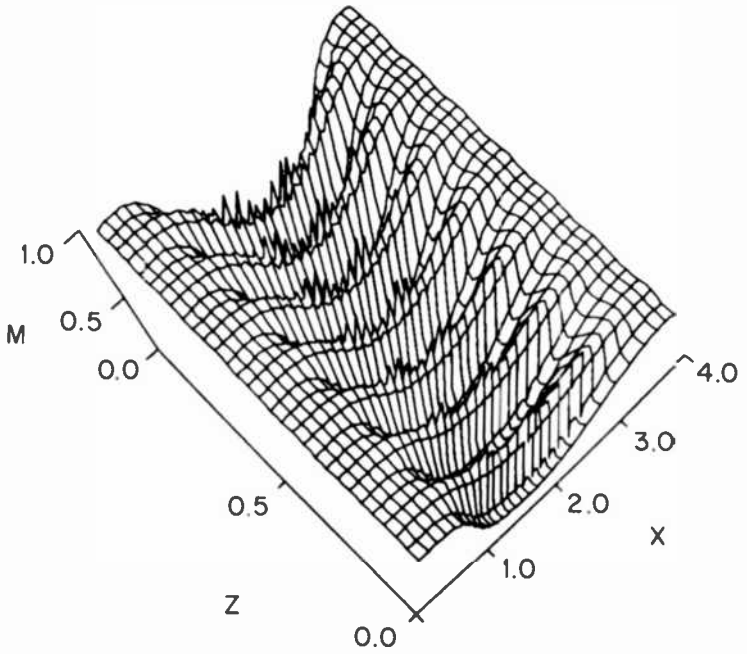


Fig. 3—Distribution of the unreacted sensitizer, *M*, in a film of AZ 1300-type resist on aluminum after exposure to the image of a 1.5-μm-wide slit. The intensity distribution of the image varies in the *x* direction and is the same as that of the in-focus curve of Fig. 2. The *z* axis is perpendicular to the resist film with the top at $z = 0$.

the resist is the one that is in focus in Fig. 2. The sinusoidal undulations in the *z* direction are due to interference between the incident and reflected light beams.

Development

The SAMPLE development program is based on an etching model in which the developer dissolves the resist off the surface, without penetrating or swelling the film. This is a realistic approximation for the common positive resists[3] consisting of diazoquinone sensitizers and novolak resins. A string model with adjustable segment lengths is used to calculate the progress of the etch front. The rate of development at each point in the resist is related to the concentration of photoactive compound remaining at that point, and each point on the interface is assumed to move at right angles to a tan-

gent at that point. This basic algorithm has been augmented by adding such refinements as surface inhibition, descumming, and multiple developments.

Parameters

In a simulation such as the one under discussion, there are a large number of parameters for which values must be determined. Values of indices of refraction can often be found in the literature. The effect of exposure on bleaching of the sensitizer can be measured on samples with large uniform areas.[3,4] The same is true of the dependence of development rates on the fraction of unexposed sensitizer[3] or on the amount of exposure.[5,6] Once the parameter values are obtained they must be inserted into SAMPLE. Since the program was originally designed to run in a batch mode on a minicomputer, the parameters were entered through simulated punched-card input. This makes the Berkeley version of the program somewhat awkward to use.

Modifications of the Model

Input-Output

After receiving our versions of SAMPLE we modified them to run on a large mainframe computer under CMS. The input-output procedures were rewritten to make them suitable for interactive use. All current values of the parameters are stored in two separate files, so that they do not have to be entered each time the program is used. They can be modified by simple interactive procedures. These changes have made the running of the program much easier and less prone to errors.

Proximity Printing

The Berkeley program allows for imaging by projection printing and, to a very limited extent, by contact printing. We have added two new imaging calculations to model other kinds of processes. First, we simulated the process of proximity printing,[7] which has been used extensively for wafer fabrication. This process is similar to contact printing, but in it the mask is suspended 1 to 40 μm above the resist layer instead of being in close contact with it. This procedure extends the life of the mask. The image, formed by "shadowing", no longer has the sharp edges of a contact print. Its quality depends on the parallelism of the illuminating light and on the

effect of diffraction at the sharp edges of the object. The same considerations also apply to the case of "soft contact printing," in which the separation of mask and image is of the order of 1 μm or less.

To calculate the diffraction pattern arising from the mask we use the Fresnel approximation to optical imaging. This approximation assumes that the incident optical rays are fairly well collimated and that the optical path differences between the various rays are relatively small. The results are valid for proximity gaps as small as about 1 μm. Typical diffraction patterns at two values of proximity gap are shown in Fig. 4. The object on the mask is a long slit of 3.0-μm width on an opaque background. The incident light is made up of parallel rays propagating at right angles to the mask. Because of the large diffraction angles, the Fresnel diffraction patterns for the small gap are very pronounced. At the large distance only a defocussing effect is observed. The more practical case of incompletely collimated light can be simulated by adding the effects of multiple incoming beams propagating in slightly different directions. Once this optical image has been determined, the exposure of the resist and the development process are simulated in the usual way.

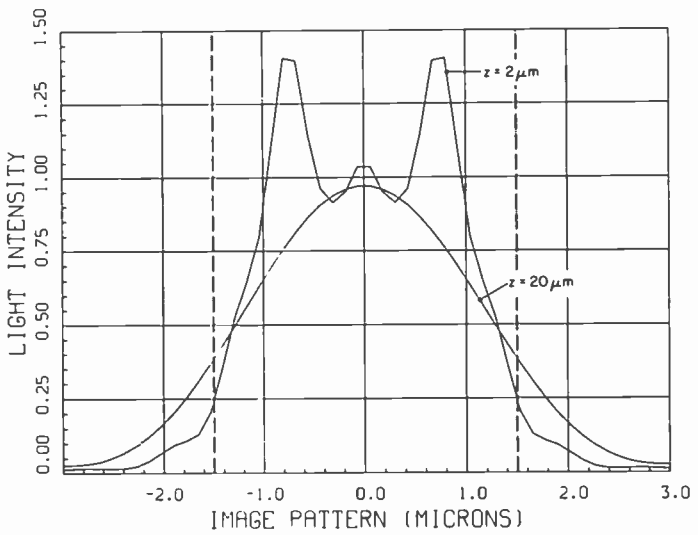


Fig. 4—Diffraction patterns produced by proximity printing. The mask consists of a 3.0-μm-wide slit and is illuminated with parallel light of 435-nm wavelength. The light distribution in the image plane is shown for two different values of the proximity gap.

We have applied this model to support experimental work on copying chrome reticles.[7] We studied the quality of the line edges that could be maintained as a measure of resolution, the consistency of the critical dimensions for features of varying size, and the dependence of the critical dimension on the proximity gap. We concluded that the latitude depends very much on the type of resist used; this agrees with experimental observations.

Conformable Mask System

The second imaging calculation that we have added is for the conformable mask system.[8] This is a two-layer resist structure, designed to improve the resolution and the shape of the resist edge profiles. The bottom layer is a film of poly(methyl methacrylate) (PMMA) that is sensitive only in the deep-UV (<250 nm). This layer is covered with a positive photoresist film, which is exposed and developed in the usual way. Since the photoresist is opaque in the deep-UV region, it acts as a mask for the flood exposure of the bottom layer. The image in the latter layer is formed by a combination of contact and diffraction imaging.

We have written a program to simulate this conformable mask system.[9] The exposure and development of the top layer is modeled in the usual way with the bottom layer acting as a fixed film on the substrate. The resultant resist structure is assumed to form a perfect mask for the far-UV exposure, i.e., where there is any resist remaining the mask is opaque. Otherwise, it is transparent. This mask is used as the object in the second step and the image is calculated by use of the proximity-printing algorithm. The bottom layer is divided into 100 equally thick horizontal sublayers. The diffracted light image is calculated for each sublayer with the proximity-printing program. The index of refraction, as well as the attenuation due to absorption in the bottom layer, must be taken into account. A difficulty arises near the top surface, because the diffraction calculation no longer provides a good approximation at distances of less than 0.5 μm. To improve the situation, we let the beam propagate without diffraction in the top region of the film and allowed it to be attenuated only according to the absorptivity of the layer. Below that region the proximity calculation is appropriate. The intensity distribution of light reflected from the substrate is calculated and added to the incoming distribution, but interference effects between the two are neglected.

These calculations were applied to study the conformable mask bilayer system of Kodak 809 resist on a 2-μm-thick sublayer of

PMMA.[9] We were particularly interested in finding out whether vertical-wall profiles could be obtained in PMMA. We also studied the effect of dyes added to the PMMA to reduce the reflection from the substrate back into the resist layer. Good qualitative agreement was obtained with experimentally formed structures.

Resist Model

The original SAMPLE program provided for only one kind of resist model, the one following Dill et al.[3] The development rate, g, is expressed as

$$g = \exp(E_1 + E_2 M + E_3 M^2),$$

where M is the fraction of sensitizer remaining unexposed and E_1, E_2, and E_3 are fitted parameters. This formulation has several disadvantages: It is inconvenient to measure the sensitizer concentration, it is not applicable to resists that do not exhibit any bleaching of photoactive sites, such as PMMA, and the equation does not approximate the data at high exposure doses well. To make the program more flexible we formulated other relationships between exposure dose and development rate. For conventional positive resists we have found the following relationship between the development rate and the exposure dose, D, to provide a good approximation to the measured values[5]:

$$g = g_0 + g_0 (D/D_e)^m$$

Here g_0 is the development rate in the absence of an exposure and D_e is the dose that doubles that rate. At high doses one must provide for a saturation at g_∞. A similar relationship is used for PMMA, but the EXPOSE algorithm is simplified because there is no bleaching and the dose distribution in the resist does not change with time. Other relationships are used for negative resists.

Contrast-Enhancement Layer

Another enhancement allows the modeling of a resist system that includes a contrast-enhancement layer (CEL).[10] The CEL is an organic film that contains a high concentration of a bleachable dye. A thin layer of it is spun on top of the resist layer. At the start of the exposure the CEL is opaque to the exposing light. Gradually the regions exposed to the highest light intensity become bleached and the underlying resist becomes exposed. If the flux required to bleach the CEL is higher than that required to expose the resist,

then the resist in the region of the higher incident intensity will be exposed long before the resist in the regions of lower incident intensity is illuminated, and sharp resist profiles tend to result.

This bilayer system has been simulated by extending the SAMPLE exposure algorithm. The substrate and the resist film are described in the usual way. The CEL is divided into 30 equal sublayers. The light intensities in each of the sublayers, as well as in the 50–100 resist sublayers and in the substrate films, are calculated. After each small exposure interval the resulting exposure dose in a given sublayer is used to calculate the amount of bleaching that has taken place in that interval. Revised values of the absorption coefficients are then determined, and the light intensity distribution is recalculated accordingly for the next exposure interval. The model assumes that both resist and CEL are bleached, but at different rates. After exposure the development process proceeds as usual.

Development Algorithm

The development algorithm of the Berkeley version describes the etching process as the movement of a string of points in the direction normal to the resist surface. This procedure exhibits difficulties when the exposure pattern varies too rapidly from point to point. Then the profile forms cusps in the regions of low development rate. As these cusps develop into sharp points, the algorithm causes the profile to generate loops. There is a provision in the program for removing these loops, but it is not always successful and the loops can become very large circles. We have avoided this problem by writing a new algorithm. When a sharp cusp forms that would later develop into a point, the progression of the string is calculated by retracting the cusp, rather than continuing to move the string forward. In addition to preventing loop formation the new algorithm generates a somewhat different wall profile, as is shown in Fig. 5. This is the case for which the difference between the two algorithms is most noticeable, since the high reflectivity of the aluminum causes very strong interference effects and large gradients in sensitizer concentration.

To determine which development algorithm is more realistic we compare the calculated profiles with the cross section of an exposed and developed resist structure, shown in Fig. 6. The micrograph shows that the interference fringes extend from the bottom to the top of the edge without a noticeable change in intensity. This is expected because of the relatively low value of absorption coefficient

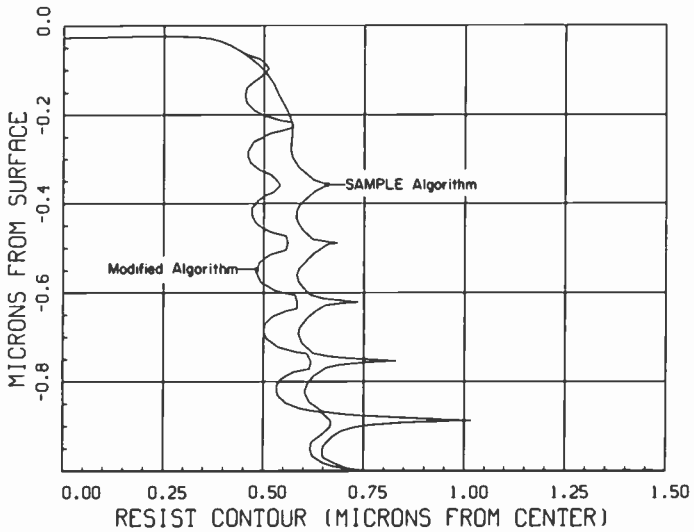


Fig. 5—Resist profiles generated in 1.0-μm AZ 1300-type resist on aluminum. The object is a 1.5-μm-wide space; the projection conditions are as in Fig. 2. Two different development algorithms are compared for calculating resist profiles. They differ significantly in the regions where the exposure dose and sensitizer concentration vary rapidly.

($0.6\ \mu m^{-1}$). Also, the undulations have a rounded rather than pointed shape. Both these observations suggest that the new algorithm simulates the profile much better than does the SAMPLE version.

Applications

Resist Profiles

The most significant results of the SAMPLE calculations are the resist profiles. They provide information on the line widths of features, on the shape of the profiles, and on the fraction of the resist remaining in unexposed areas. These characteristics define how well the resist pattern performs its task of serving as a mask for the subsequent etching, diffusion, or implant steps. The profiles that can be produced from a given exposure depend on both the kind of photoresist used and the developer that is chosen.[6]

The resist linewidth is defined as the distance between the points where the resist surface ends at the substrate. Ideally, it should always be equal to the dimensions of the corresponding object on the mask, or it should differ from it by a fixed value, regardless of

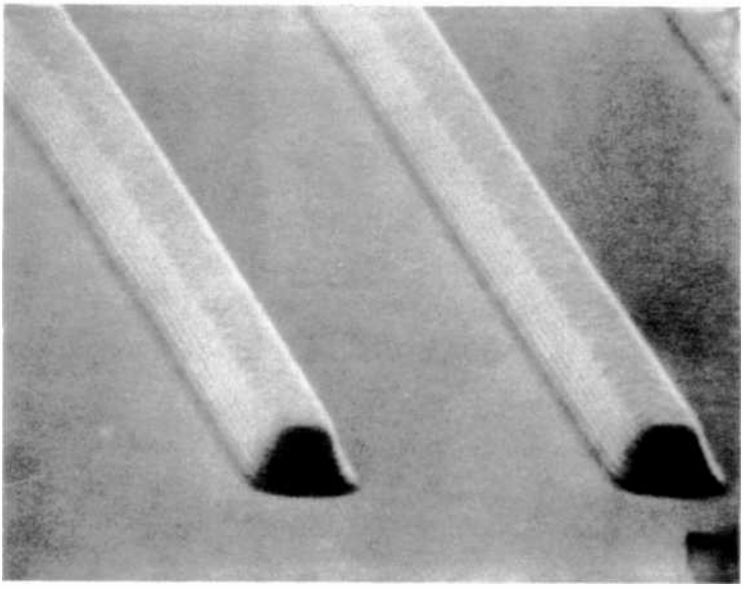


Fig. 6—Scanning electron micrograph, at 10,000 and 80° incidence, of a resist pattern on silicon. The object consisted of an array of 1.0-μm lines and 4.0-μm spaces and was projected by a TRE 800 (TRE Co.) wafer stepper at 436 nm.

the size of the object. In the latter case the desired resist dimensions can be obtained by uniformly biasing the mask.

As an example, Fig. 7 compares profiles generated by three different objects of the same dimensions: an isolated line, a space, and one of the lines of a grating of equal-sized lines and spaces. The illumination and projection conditions are the same for all three of the objects and are the same as the ones used in Fig. 2. The resist layer is assumed to be 1 μm thick and consist of AZ 1300 (American Hoechst Corp.) type of resist. For this calculation we set the index of refraction of the substrate to the same value as that of the resist film to eliminate optical reflections at the substrate interface. This is not a realistic model but it allows us to compare the various resist profiles without the complications caused by the interference between incident and reflected beams. The development step is modeled by using the parameter values of a simple inorganic alkaline developer. Exposure dose and development time are selected so that the width of the isolated space feature after development is equal to the width of the corresponding space on the object (no bias condition). It can be seen in Fig. 7 that under these conditions the

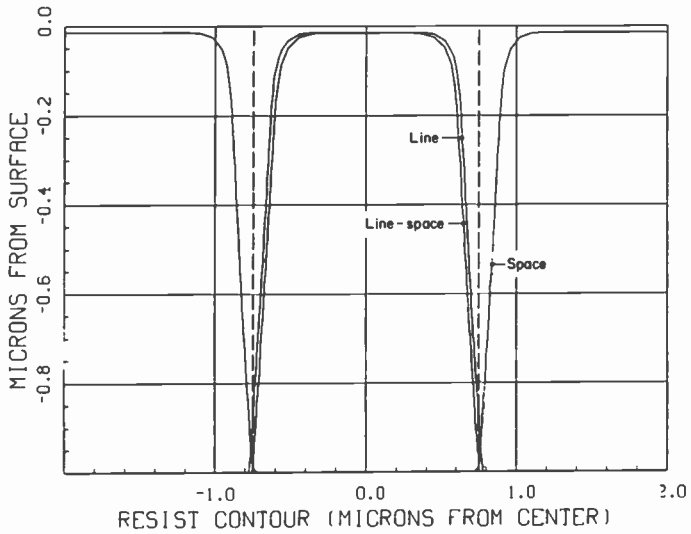


Fig. 7—Developed resist profiles of three different kinds of features, an isolated 1.5-μm-wide space, an isolated 1.5-μm line, and a similar line that is part of a grating of equal lines and spaces. The imaging conditions are those of the in-focus case of Fig. 2. The exposure dose is 25 mJ/cm^2. The resist parameters are modeled after AZ 1300-type resists with an inorganic alkaline developer and a development time of 24 s. To eliminate reflections, the substrate has the same optical properties as the resist.

developed line-space grating also shows no bias, whereas the isolated line is 0.05 μm too wide.

To investigate the influence of feature size on the linewidth the same kind of calculations are performed for other dimensions. The results are summarized in Fig. 8, which plots the deviation from the nominal dimension as a function of feature size for the three kinds of features. Features of 2 μm and larger all have about 0.03 μm of excess resist width; this can be corrected by biasing the mask 0.015 μm smaller on each edge. The remaining fluctuations in the data are caused by the discrete nature of the SAMPLE algorithms. For features smaller than 2.0 μm the deviations from the design value become more serious, reaching a range of 8% of the linewidth for 1.0 μm features.

The calculated profiles can be used to predict the latitude of a given resist system. Consider the above example of a 1.5-μm-wide isolated space in AZ 1300-type resist on a nonreflecting substrate. We varied both the exposure dose and the development time and calculated the feature size for each combination. The results are

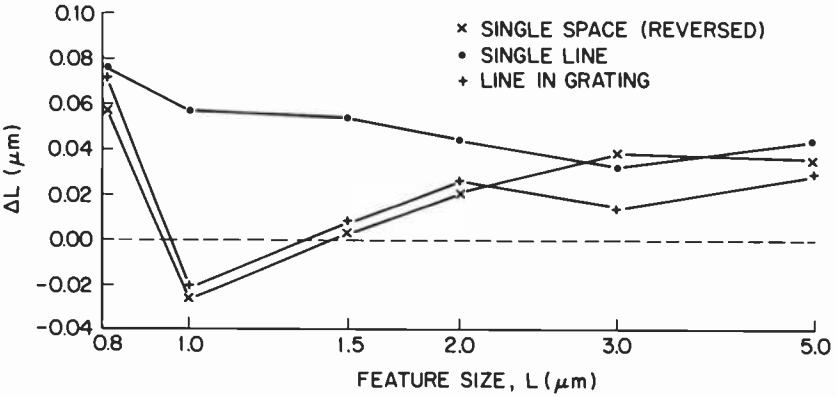


Fig. 8—Deviation of the developed linewidths from their nominal values for various features. For the case of a single space the sign of the deviation is reversed, so that for all cases $\Delta L > 0$ means that too much resist is left. The conditions are the same as the ones in Fig. 7. A slightly longer development time would place the average deviation at zero.

shown in Fig. 9. The latitude can be defined as the slope of a curve at the nominal dimension. It has a value of 7 nm/% dose, independent of the dose. This means that a 10% change in dose produces a 0.07-μm change in the width of a 1.5-μm line. If the lines are narrowed by overdevelopment, the slope gradually decreases, resulting in greater latitude.

High-Contrast Resists

The exact shape of the developed resist profiles is of great importance for the subsequent processing steps. Some of the steps, such as plasma etching, require that the resist walls be as vertical as possible for the resist to protect the underlying substrate fully. More shallow slopes are required if contact holes are to be opened. The shape of the wall profiles is determined by the contrast of the various ingredients of the lithographic process. The contrast of the imaging step is controlled by the printer and the lenses available, so that most of the contrast variation comes from the resist. Resist contrast is generally defined as the slope of a curve of resist thickness versus the logarithm of exposure dose for a fixed development process.

We have studied the influence of contrast on the shape of the resist profiles.[11] The emphasis was on high contrast, since resist manufacturers have introduced a variety of new developers that

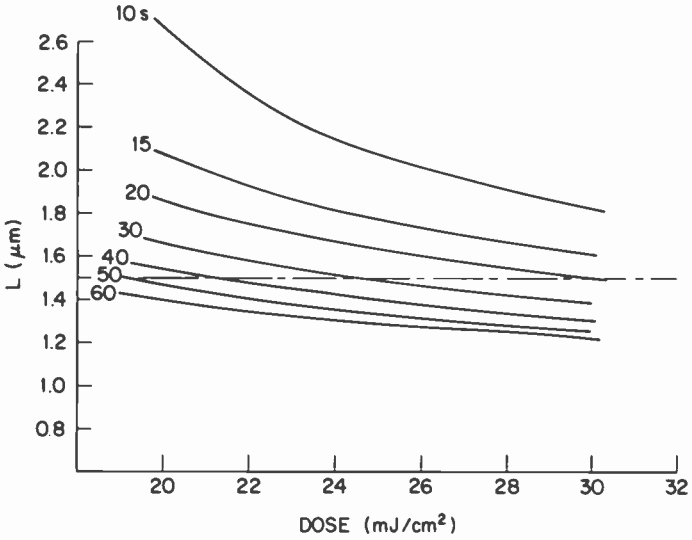


Fig. 9—Calculated width of the image of a 1.5-μm-wide space as a function of dose with development time as parameter. The resist system is the same as in Fig. 7. The slope of the curves determines the latitude to exposure variations.

appear to increase the contrast of the resist system. These developers are organic bases with surfactants added to the solution. The surfactant prevents any dissolution of the resist at low exposure intensities, while at the high intensities the effect of the surfactant disappears and the resist develops rapidly because of the aggressive nature of the developer. We modeled this situation by postulating a thin surface layer on the resist film; this layer dissolved much more slowly than the bulk, but with the same functional dependence on exposure. Because this model allows excellent fits to the measured contrast curves, the resulting parameters can be used to calculate the shapes of the profiles.[11]

Negative Resists

Negative photoresists are usually rubber-based resist systems developed in organic developers. They tend to swell during development and therefore have limited resolution. Recently, other kinds of negative resist systems, suitable for higher-resolution work, have become available.[12] They consist of sensitizer–resin combinations, can be developed in aqueous developers, and do not show significant swelling. These resists make it possible to generate profiles that are

very different from those of positive resists. In the exposed regions of the film the less exposed material at the bottom of the film etches *faster* than that at the top, the opposite of what happens in positive resists.

Matsuzawa et al.[13] have modeled such a negative resist in a manner analogous to that in which positive resists are simulated. The bleaching rate of the sensitizer was measured and the dependence of development rate on the remaining fraction of unbleached sensitizer was approximated by an exponential function. We used the same technique to simulate the developed profiles of such a resist and, in Fig. 10, compare them with those of a positive resist. A 1.5-μm-wide feature is imaged, a line in case of the positive resist, a space for the negative one. The 1.0-μm-thick resist layers are spun on silicon substrates that have a reflection coefficient of about 50%. The profiles are shown at three different development times, the middle one producing the desired linewidth, the others showing the effect of over- and underdevelopment, respectively. It can be seen that in negative resists the faster etch at the bottom allows square profiles with concave walls. Undercut structures, which are suitable for lift-off procedures, can also be generated. Such structures cannot be produced in positive resists.

Critical-Dimension Control

One of the goals of microlithographic processing is to produce resist structures with correct dimensions everywhere on the wafer. This requires extremely uniform formation, exposure, and processing of the resist layer. But it also requires that the width of a line of resist change by only a negligible amount when the line passes over a step in the substrate. However, the thickness of a spun-on resist layer changes near such a step, so that the variation of linewidth with resist thickness becomes important. We have investigated this relationship with emphasis on the case of a highly reflecting substrate where interference effects become dominant.[14] Calculations showed good agreement with experimental results.

The magnitude of the effect of resist thickness is demonstrated in Fig. 11 for a 1.5-μm-wide isolated line. The exposure and development times are adjusted so that a film of 1.0-μm thickness produces the desired dimension. For a nonreflecting substrate there is a gradual increase in the width of the line as the film thickness increases. When the substrate is highly reflecting, the slow change is overwhelmed by a large oscillation of the critical dimension. In fact, for half the thickness values, the resist is not even completely

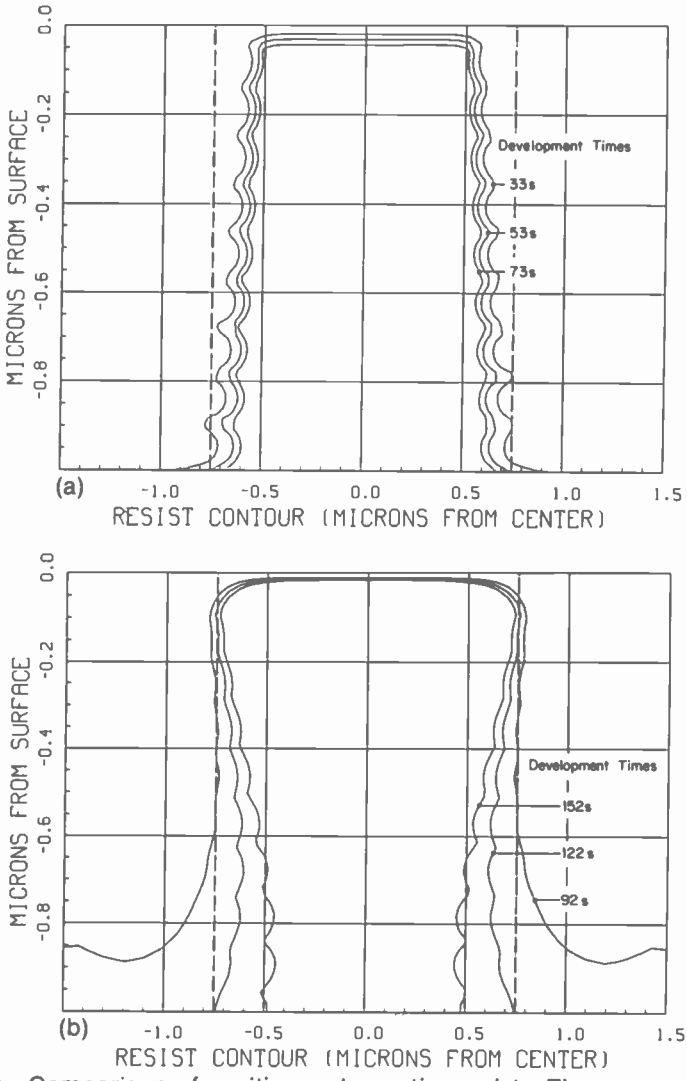


Fig. 10—Comparison of positive and negative resists. The exposure is at 365-nm wavelength, with a lens of *NA* = 0.28 and s = 0.7. The substrate is silicon. (a) 1.0-μm positive resist (AZ 1300-type), exposed with a 1.5-μm-wide line and a dose of 25 mJ/cm^2; (b) MRL[20] negative resist,[10] same thickness, exposed with a 1.5-μm-wide space and a dose of 120 mJ/cm^2. The development times are indicated on the figure.

cleared out in the exposed region. These oscillations, with a period of half a wavelength (in the resist: $\lambda = \lambda_0/n$) are due to interference between the incident and reflected waves in the resist. It not only

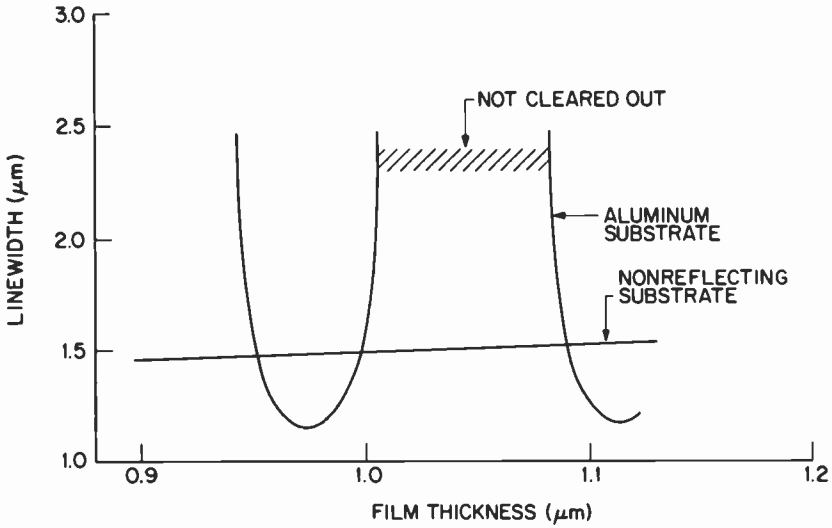


Fig. 11—Variation of the developed linewidth with resist thickness. A 1.5-μm-wide isolated line is imaged at 436 nm with a 0.28-*NA* lens onto layers of AZ 1300-type resist of varying thickness. Two cases are compared, a nonreflecting substrate and a highly reflecting (aluminum) substrate. The exposure dose is 24 mJ/cm^2 and the development times are 31.0 s for the nonreflecting substrate and 37.7 s for the aluminum substrate (95% reflectivity). The latter requires the higher dose because the interference effects cause less light to be coupled into the film at this particular thickness.

produces stratification into regions of high and low exposure, but also causes a periodic change in the amount of light that is coupled into the film. Clearly, this effect can produce great difficulties in linewidth control over steps. One of the ways the problem can be mitigated is to place an absorbing layer between the resist and the substrate. Again, SAMPLE can be used to determine how much absorption is required to provide the necessary control.

Conclusions

We have discussed the program SAMPLE and have applied it to various problems in microlithography. To make the program more generally applicable, several algorithms were modified. We also added additional models for other lithographic processes, such as proximity printing and multilayer processes. These changes have made the program very useful for predicting the results of parameter variations, for trying new structures, and for determining the

degree of process control that can be obtained. We have successfully predicted the effects of various resist parameter changes, the value of new resist structures, and the possibilities of controlling the processes.

Acknowledgments

We wish to thank F. W. Brehm for help with the initial installation of the programs; G. Lang of Laboratories RCA Ltd., Zurich, Switzerland, for detailed consultations about the operation of the system software; and D. P. Barton and J. R. Fields for assistance with generating plots and converting to batch operation. Many of the developments were discussed extensively with M. Kaplan and L. K. White, who also reviewed the manuscript, as did A. W. Levine.

References:

[1] W. G. Oldham, S. N. Nandgaonkar, A. R. Neureuther, and M. O'Toole, "A General Simulator for VLSI Lithography and Etching Processes. Part 1—Application to Projection Lithography," *IEEE Trans. Electron Devices*, vol. ED-26, no. 4, p. 717 (1979).

[2] M. M. O'Toole, E. D. Liu, M. S. Chang, "Linewidth Control in Projection Lithography Using a Multilayer Resist Process," *IEEE Trans. Electron Devices*, vol. ED-28, no. 11, p. 1405 (1981); L. K. White, "Approximating Spun-On, Thin-Film Planarization Properties on Complex Topography," *J. Electrochem. Soc.*, vol. 132, p. 168 (1985).

[3] F. H. Dill, A. R. Neureuther, J. A. Tuttle, and E. J. Walker, "Modeling Projection Printing of Positive Photoresists," *IEEE Trans. Electron Devices*, vol. ED-22, no. 7, p. 456 (1975).

[4] M. Kaplan and D. Meyerhofer, "Response of Diazoquinone Resists to Optical and Electron-Beam Exposure," *RCA Rev.*, vol. 40, no. 2, p. 166 (1979).

[5] D. Meyerhofer, "Photosolubility of Diazoquinone Resists," *IEEE Trans. Electron Devices*, vol. ED-27, no. 5, p. 921 (1980).

[6] D. Meyerhofer, "Comparison of Positive Photoresists for Use in Microlithography," in *Microcircuit Engineering 82* (Grenoble, 1982), p. 254.

[7] D. Meyerhofer and J. Mitchell, "Proximity Printing of Chrome Masks," *RCA Rev.*, vol. 43, no. 4, p. 608 (1982).

[8] B. J. Lin, "Portable Conformable Mask—a Hybrid Near-Ultraviolet and Deep-Ultraviolet Patterning Technique," *Proc. SPIE*, vol. 175, p. 114 (1979).

[9] L. K. White and D. Meyerhofer, "Image Formation in the Contact-Printed PMMA Resist Sublayers," *RCA Rev.*, vol. 45, no. 3, p. 649 (1984).

[10] B. F. Griffing and P. R. West, "Contrast Enhanced Photolithography," *IEEE Electron Devices Lett.*, vol. EDL-4, no. 1, p. 14 (1983).

[11] D. Meyerhofer, "Comparison of Positive Photoresist Systems," in *Microcircuit Engineering 83* (Academic Press, London, 1983), p. 305.

[12] T. Iwayanagi, M. Hashimoto, S. Nonogaki, S. Koibuchi, and D. Makino, "Azide-Phenolic Resin UV Resist (MRL) for Microlithography," *Polymer Eng. Sci.*, vol. 23, no. 17, p. 935 (1983).

[13] T. Matsuzawa, A. Kishimoto, T. Iwayanagi, H. Yananzawa, and H. Obayashi, "Optimization of the Composition of a New MRS-Type Negative Resist," *IEEE Trans. Electron Devices*, vol. ED-30, no. 12, p. 1780 (1983).

[14] L. K. White and D. Meyerhofer, "Positive-Resist Processing Considerations for VLSI Lithography," *RCA Rev.*, vol. 44, p. 110 (1983).

Active-Array Antenna Beam Shaping for Direct Broadcast Satellites and Other Applications

Samuel H. Colodny
RCA Astro-Electronics Division, Princeton, NJ

Roger L. Crane
RCA Laboratories, Princeton, NJ

Abstract—Current direct broadcast satellites use high-power traveling-wave-tube amplifiers and a multiple-horn feed array and reflector antenna to achieve the downlink high power required. The desirability of replacing the traveling-wave-tube amplifier with a solid-state power amplifier is well established. However, no high-efficiency, 250-W, Ku-band solid-state power amplifier now exists. The high power required can be achieved by using a large number of low-power solid-state amplifiers to drive the elements in an active-array antenna. This paper addresses the mathematical problem of determining, for each element in an active-array antenna, the amplitude and phase drive signals required to produce the desired antenna pattern on Earth.

Introduction

Satellite communication systems that transmit their signals directly to homes are called direct broadcast satellite (DBS) systems. Currently at least one such system is in operation and more are expected. Present DBS systems use conventional communication system architecture, the same as is used by most fixed-service satellite communication systems. The output components consist of a high-power traveling-wave-tube amplifier (TWTA) for each channel and an output multiplexer to combine channels at different frequencies into a single antenna port. The antenna is a multiple horn

feed array and reflector that produces a beam shaped to match closely the intended geographic coverage area as seen from geostationary orbit. Each TWTA in a typical DBS system produces about 250 W at Ku-band frequencies of approximately 12 GHz. Beams are shaped to service the United States time zones.

Recently, at lower frequencies (e.g., C-band), high-power TWTAs were replaced with solid-state power amplifiers (SSPAs). However at Ku-band there are currently no solid-state devices available with sufficient power to substitute for the TWTA.

To take early advantage of solid-state technology, an alternate communication system architecture has been developed. This architecture, shown in Fig. 1, uses a large number of relatively low power solid-state amplifiers, each feeding a small subarray of patch-radiating elements. By establishing the proper amplitude and phase of the signal for each element of the subarray, this active-array antenna (AAA) can produce the desired radiation pattern. This architecture also lends itself to implementation as a relatively low profile assembly using miniature microwave circuit techniques.

The AAA has both reliability and other performance advantages when compared with conventional satellites. For example, cathode wearout, a problem known to limit TWTA lifetime, does not exist. Since solid-state amplifiers use low-voltage power, high-voltage breakdown problems are eliminated. Further, since there are many low-power amplifiers and radiating elements, performance is insensitive to the failure of a few amplifier modules. The close proximity

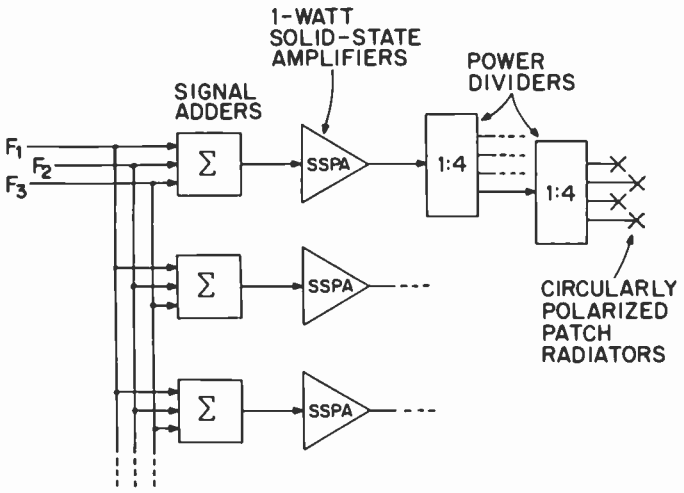


Fig. 1—DBS system using an active-array antenna.

between the output amplifier and a radiating element, and the absence of a conventional multiplexer, produce lower radio-frequency losses and, hence, improved power efficiency. Each amplifier module is associated with phase shifters that can be adjusted to maintain the proper phase relationships between elements. This permits the antenna radiation pattern to be changed in orbit to maximize the satellite's revenue-generating capability. The flexibility to produce radiation patterns not anticipated before launch is a great benefit.

The next section defines the mathematical formulation of the antenna synthesis problem and describes the method developed to determine the phase and amplitude drive signals that produce a desired radiation pattern with maximized gain. Some solutions vary only the phase of the excitation between elements (uniform drive); other solutions vary both the phase and the amplitude from element to element. Generally the latter perform better.

Mathematical Formulation of the Antenna Synthesis Problem

Classical antenna theory[1] provides the mathematical model needed to compute the strength of the signal radiated from the antenna. The model used here assumes a rectangular antenna with continuously varying amplitude and phase drive signals, $A(x,y)$ and $\psi(x,y)$, respectively. These functions, A and ψ, when appropriately specified, produce a desired radiation pattern on the Earth's surface. Let the antenna axis be defined as a line normal to the antenna surface, passing through the reference origin of the antenna system, and directed at a fixed point on Earth. This point is selected to be near the center of the desired service area. Let (R,θ,ϕ) denote the spherical coordinates of a point on a sphere of radius R, centered at the antenna origin. θ denotes the polar angle, i.e., the angle from the antenna axis, and ϕ the azimuthal angle. The relative voltage gain at a point on Earth defined by the angles θ and ϕ is approximated by

$$g(\theta,\phi,A,\Psi) = [(1 + \cos\theta)/8] \int_{-1}^{1} \int_{-1}^{1} A(x,y)\, e^{i\{\beta(x,y,\theta,\phi) - \Psi(x,y)\}}\, dxdy \quad [1]$$

where

$$\beta(x,y,\theta,\phi) = \pi(xL_x \cos\phi + yL_y \sin\phi) \sin\theta, \quad [2]$$

and L_x and L_y are the x and y dimensions, respectively, of the antenna in wavelengths. Note that the antenna coordinates have been normalized to $[-1,1]$. Relative gain is defined as gain with respect to the on-axis gain (gain with $\theta = 0$) of the same antenna with unit amplitude and constant phase drives. The relative gain in decibels is defined as

$$G(\theta,\phi,A,\psi) = 10\ log_{10}\ |g(\theta,\phi,A,\psi)|^2. \qquad [3]$$

For this formulation, $A(x,y)$ and $\psi(x,y)$ are defined to be two dimensional polynomials of the form

$$P(x,y) = \sum_{i=0}^{N} \sum_{j=0}^{i} p_{mn}\ \gamma_{m-1}(x)\ \gamma_{n-1}(y), \qquad [4]$$

where $m = j + 1$, $n = i - j + 1$, N is the polynomial degree, and $\{\gamma_i(t), i = 0,1, \ldots\}$ denotes one of two basis sets, either $1,t,t^2, \ldots$, or the Chebyshev set, $T_0(t),T_1(t),T_2(t), \ldots$. The coefficients of A and ψ are the unknowns that must be chosen to provide the desired signal in the service area.

An example for $N = 2$ and the set $1,t,t^2, \ldots$ helps clarify the ordering implied above. In this case, the p_{mn} values are coefficients multiplying the basis functions shown in the upper left triangle of the matrix

$$\begin{bmatrix} 1 & y & y^2 \\ x & xy & 0 \\ x^2 & 0 & 0 \end{bmatrix},$$

i.e., $P(x,y) = p_{11} + p_{12}y + p_{21}x + p_{13}y^2 + p_{22}xy + p_{31}x^2$.

The antenna synthesis problem can be formulated as a constrained optimization problem, that of maximizing the average gain in a given service region subject to constraints that the antenna input power remain constant and that the maximum deviation from average gain within the service area be limited to a specified tolerance. More specifically, the problem is to

$$\underset{A,\Psi}{\text{Maximize}}\ G_{avg}(A,\psi) = (1/a) \int_R \int G(\theta,\phi,A,\Psi)\ d\theta d\phi \qquad [5]$$

$$\text{subject to} \int_{-1}^{1} \int_{-1}^{1} A^2(x,y)\ dxdy = \text{constant}, \qquad [6]$$

and $\max_{\theta,\phi \in R} |G(\theta,\phi,A,\Psi) - G_{avg}(A, \Psi)| \leq \delta,$ [7]

where a is the area of the service region, R.

Numerical Approximation of the Model Problem

The sequential quadratic programming algorithm implemented in a subroutine called VMCON[2] was used to solve this nonlinear, constrained optimization problem. Several approximations to the continuous problem defined above were made in setting up the numerical problem. First, the continuous region, R, was replaced by a set of discrete points. Fig. 2 shows the set of 60 points used to represent the Eastern Time Zone. The integral over R shown in Eq. [5] was replaced by a weighted sum of values computed on the discrete set

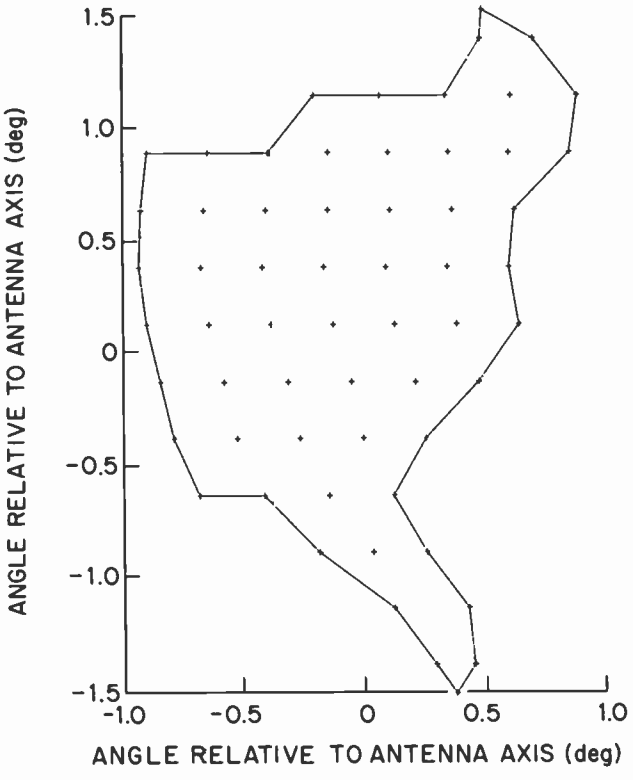


Fig. 2—60 data points defining the United States Eastern Time Zone.

just described. Likewise, the maximum deviation from average gain shown in Eq. [7] was computed by determining the maximum on the discrete set. The double integrals over the antenna surface shown in Eqs. [1] and [6] were computed by Gaussian quadrature.

Use of VMCON requires repeated evaluation of the objective function (Eq. [5]), the constraint functions (Eqs. [6] and [7]), and the derivatives of these functions with respect to the coefficients of A and ψ. Thus, the computations performed are substantial. The evaluation time was minimized by using analytic expressions for the derivatives and by removing from the inner loop the evaluation of several expressions that could be precomputed. For example, the trigonometric functions in Eq. [2] were evaluated at the Gauss sample points and stored before calling VMCON.

An initial solution estimate and a positive definite estimate of the Hessian matrix of the objective function must be provided for VMCON. Choice of the latter can be crucial to using VMCON successfully. An improperly scaled estimate of the Hessian can cause VMCON to fail. The procedure followed was to compute the Hessian at the initial point by using analytic expressions for its elements and then to compute its eigenvalues. When at least one negative eigenvalue was found, a positive constant slightly greater than the magnitude of the least eigenvalue was added to each diagonal element, whereby a positive definite initial estimate of the Hessian was obtained. After this procedure was incorporated in the program, no convergence problems were experienced.

The initial solution estimate must also be selected with some care. Multiple local solutions to the optimization problem exist, and the particular one found by VMCON depends on the initial estimate. Several strategies were used to increase the probability of finding a global optimum. Optimum solutions for N^{th} degree polynomials were used as initial estimates for degree $N + 1$ solutions; N^{th} degree solutions obtained with constant amplitude drive [$A(x,y) = 1$] were also used to start the search for solutions with both amplitude and phase variation. Solutions obtained by using one constraint tolerance, δ, were used to initiate solutions for a reduced tolerance. Sometimes the same optimum was achieved, but usually different initial estimates resulted in different solutions. There is no guarantee that the solutions presented here are globally optimum solutions of the model problem. However, since the gains achieved are comparable with those measured on an existing parabolic antenna, it is unlikely that substantially better solutions to the model problem exist.

Solutions of the Model Problem

Let N denote the degree of ψ, the phase function. The constant term in ψ does not affect the antenna pattern, and thus there are $[(N + 1)(N + 2)/2] - 1$ polynomial coefficients to be determined if only phase variations are permitted, i.e., if $A(x,y) = 1$. When both amplitude and phase drives are varied, there are $[(N + 1)(N + 2)] - 1$ unknown coefficients if it is assumed, as is true for the results shown below, that the degrees of $A(x,y)$ and $\psi(x,y)$ coincide.

All results shown below are for the Eastern Time Zone as defined by the 60 points shown in Fig. 2. Antenna dimensions L_x and L_y of 60 and 120 wavelengths, respectively, and an inequality constraint tolerance, δ, of 1.2 dB, were used. The integrals in Eqs. [1] and [6] were evaluated by using 12 Gauss sample points in each dimension. VMCON's termination parameter, TOL,[2] was set to 0.0001 for each solution. Solution accuracy and program running time depend directly on the number of Gauss samples and the termination parameter, TOL. The parameter values shown above were determined by means of computational experiments that gave sufficiently accurate solutions without excessive computing costs.

Fig. 3 shows the dependence of average gain in the service area upon the polynomial degree. Gains shown are relative to the peak gain of an equal-sized antenna driven with constant amplitude and phase. The lower trace results from phase variation only; the upper trace corresponds to both amplitude and phase variation. Also plotted are several discrete points corresponding to locally optimal solutions with gains lower than the gain for the best solution for a particular value of the polynomial degree, N. For a fixed degree, solutions with variable phase and amplitude are substantially better than constant-amplitude solutions. Also, the incremental improvement obtained by increasing the degree is becoming marginal by degree seven, particularly for solutions with variable phase and amplitude.

Figs. 4 and 5 show the antenna patterns resulting from two distinct, locally optimal, 4th degree solutions with average gains of -11.63 and -11.23 dB, respectively. Gain values in these plots have been shifted by an additive constant equal to the magnitude of the average gain. Thus, in the service area, the gains shown vary from approximately -1.2 to 1.2 dB, corresponding to the inequality constraint tolerance, δ. Examination of the plots shows that gains with magnitude greater than 1.2 dB do occur. This is, of course, possible since the solution was constrained only on a finite set of

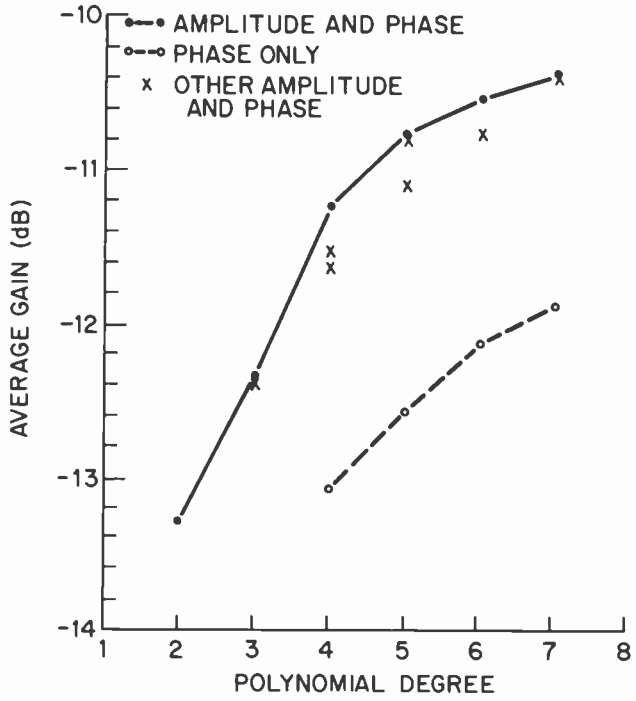


Fig. 3—Average gains in Eastern Time Zone resulting from optimal solutions for a 60λ × 120λ antenna (gains are relative to peak gains for an identical, uniformly excited antenna).

points. These two solutions are distinctly different, qualitatively and quantitatively. Yet both are plausible and appear to be valid local solutions. In each case, the −1.2-dB contour passes through several points on the boundary of the point set shown in Fig. 2.

Fig. 4, which corresponds to the poorer solution in terms of average gain, has in its interior a region of low gain, an undesirable property. Also, the pattern in Fig. 5 has at least one undesirable feature. The coarse discretization used to represent Florida has resulted in a pattern that is far from optimal for service in that area. This illustrates an important point. The antenna design problem has important aspects that have not been incorporated in the model problem. Thus, it may be wise to inspect several local solutions with comparable average gains to determine which, in fact, might be the "best" solution to the real design problem. Or, it might be wise to include more aspects of the "real" problem in the model, e.g. through the use of additional constraints. Densely populated areas could be favored by constraining the gain to be greater there. Constraints

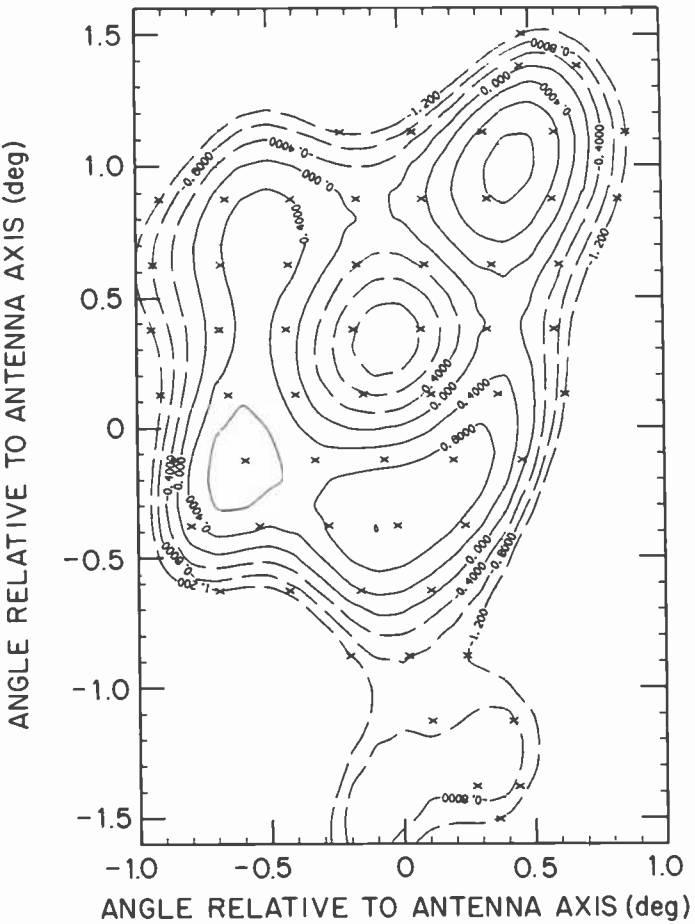


Fig. 4—Contour plot of antenna pattern for 4th-degree, −11.63-dB solution.

on design sensitivity, sidelobe constraints, and even economic constraints could be added to the model. A constrained optimization formulation of the design problem encourages the evolutionary development of progressively more realistic model problems.

The amplitude and phase functions, A and ψ, are also of interest since they must be realized in hardware if an actual antenna is built. The perspective and contour plots of Figs. 6–9 show the amplitude and phase drive functions that yield the antenna pattern shown in Fig. 5. Amplitude is relative to that of an antenna with constant, unit amplitude drive. Note that the amplitude function assumes negative values. This, of course, can be implemented by

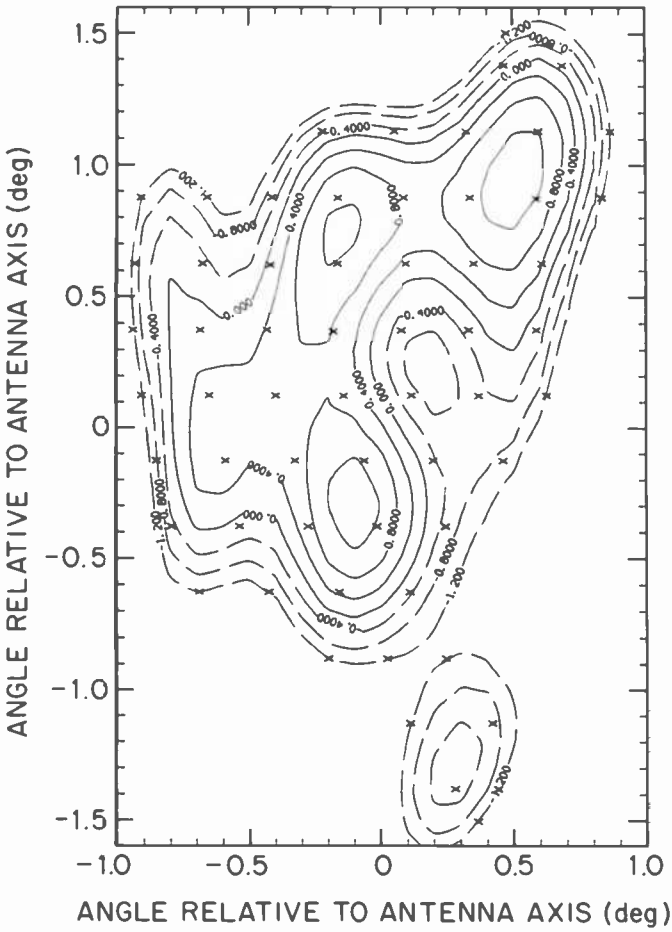


Fig. 5—Contour plot of antenna pattern for 4th-degree, −11.23-dB solution.

using the magnitude of the amplitude function and by shifting the phase 180 degrees. Phase values in radians are shown in Figs. 8 and 9.

Fig. 10 shows the antenna pattern in the service area for the best solution found, a 7th-degree phase- and amplitude-varying solution with average gain of −10.37 dB. The pattern exterior to the service area is shown in Fig. 11. Here, due to the variation of β as a function of x and y, 24 Gauss sample points in each dimension were needed to accurately evaluate the integral of Eq. [1]. Recall that no sidelobe constraints were used to obtain this solution. Since the gain has

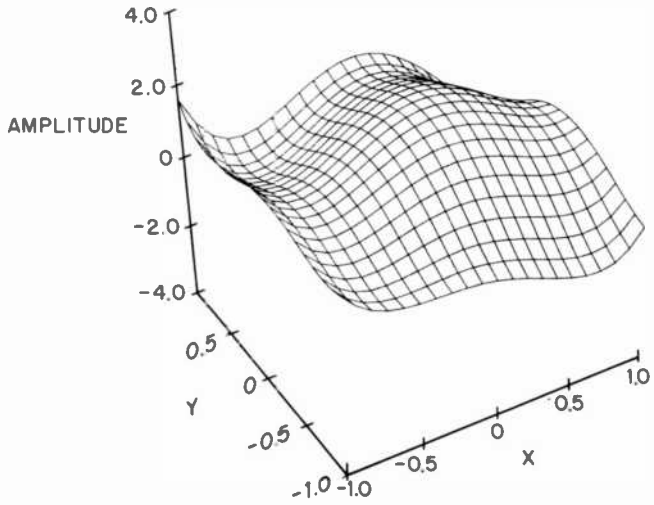


Fig. 6—Perspective plot of amplitude drive function for 4th-degree, −11.23-dB solution.

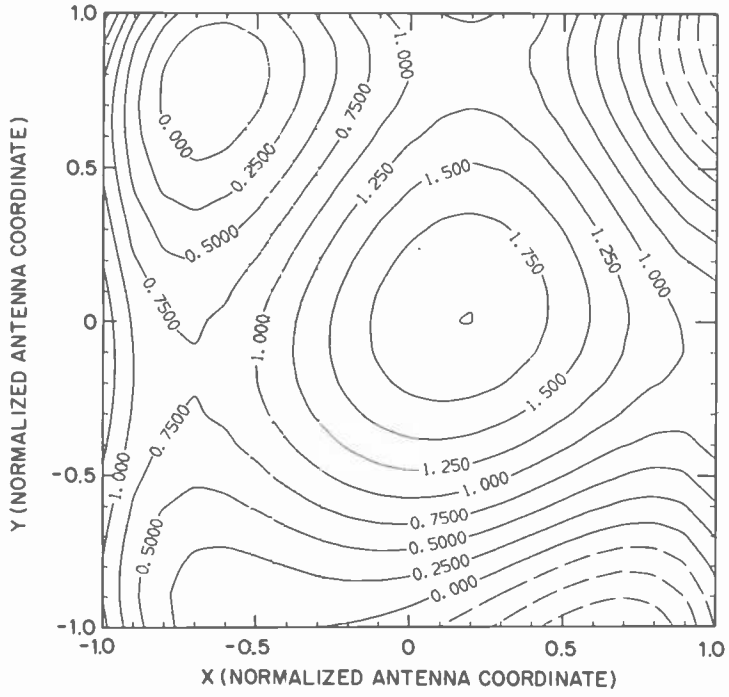


Fig. 7—Contour plot of amplitude drive function for 4th-degree, −11.23-dB solution.

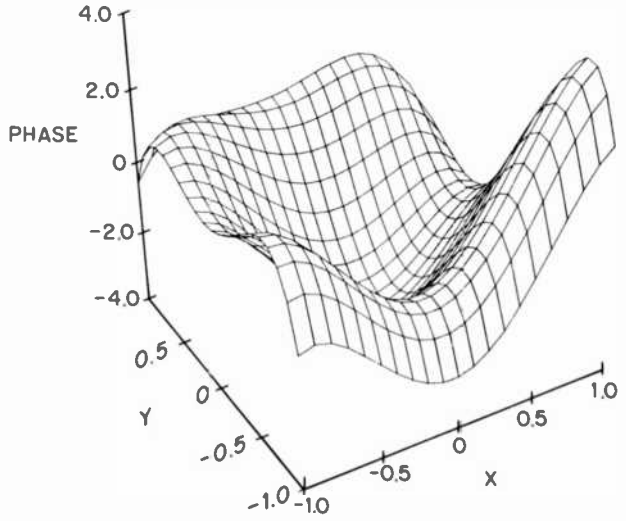


Fig. 8—Perspective plot of phase drive function for 4th-degree, −11.23-dB solution.

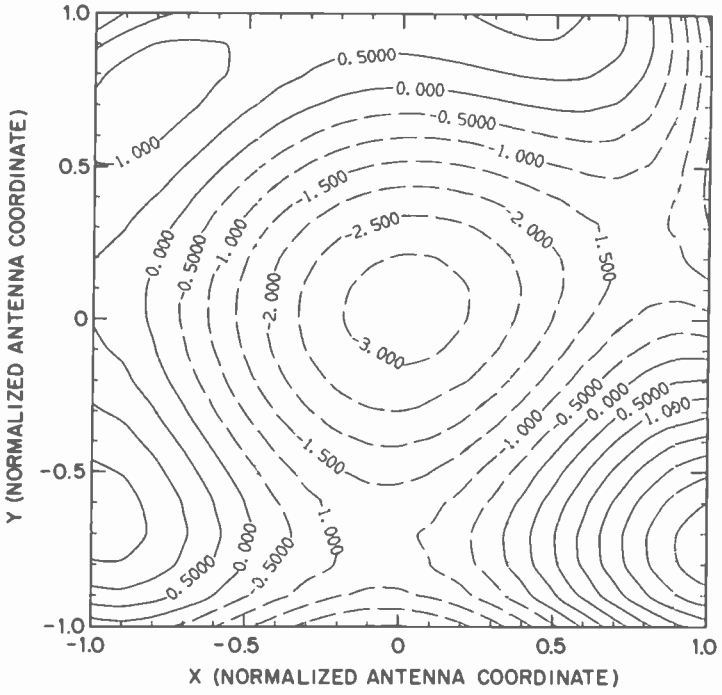


Fig. 9—Contour plot of phase drive function for 4th-degree, −11.23-dB solution.

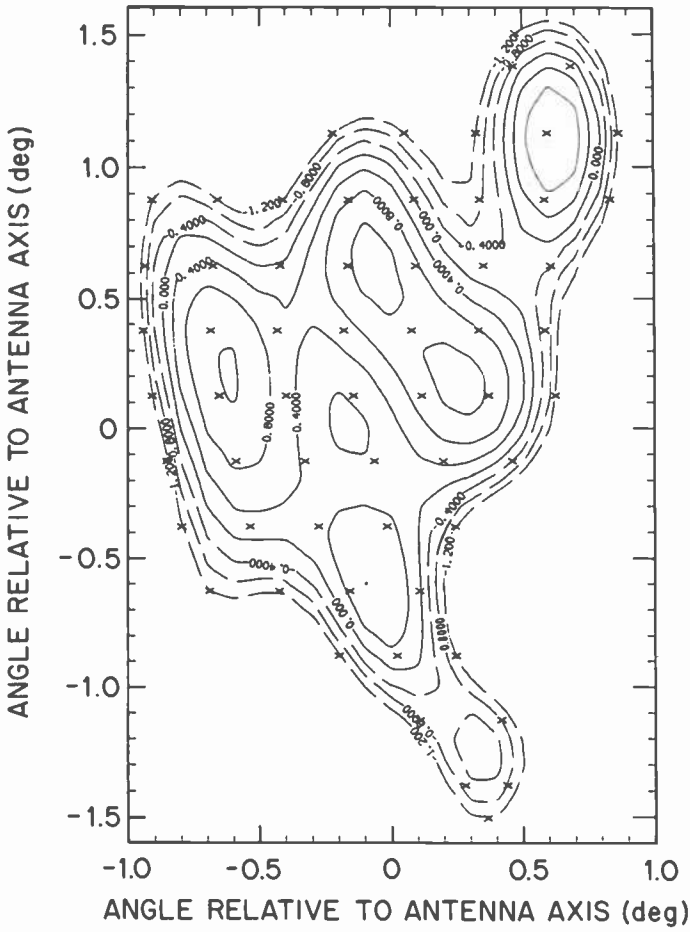


Fig. 10—Contour plot of antenna pattern for 7th-degree, −10.37-dB solution.

been maximized within the service area, one does not expect to find regions of appreciable gain outside. Nevertheless, it is necessary to check the solution. If unacceptable behavior were to be found, the model problem could be modified to include sidelobe constraints. Figs. 12–15 show the amplitude and phase functions for this 7th-degree solution in the form of perspective and contour plots. In Fig. 15, contours above two radians are one radian apart.

Conclusions

Future designs of direct broadcast and other satellites will have

solid-state power amplifiers replacing traveling-wave-tube amplifiers. Many small solid-state devices driving elements in an active-array antenna can provide the needed power. A synthesis method has been developed to determine the element amplitude and phase drive signals required to maximize the gain in a specified service area while maintaining nearly uniform coverage. This will result in a system with improved life, reliability, and efficiency that can be reconfigured while in orbit.

A broader interpretation of the results presented here shows that the shaped-beam antenna synthesis problem has been formulated and successfully solved as a constrained optimization problem. This approach is readily extendable. By including additional constraints that characterize other important aspects of "real" design problems, even more comprehensive and realistic model problems can be formulated and solved.

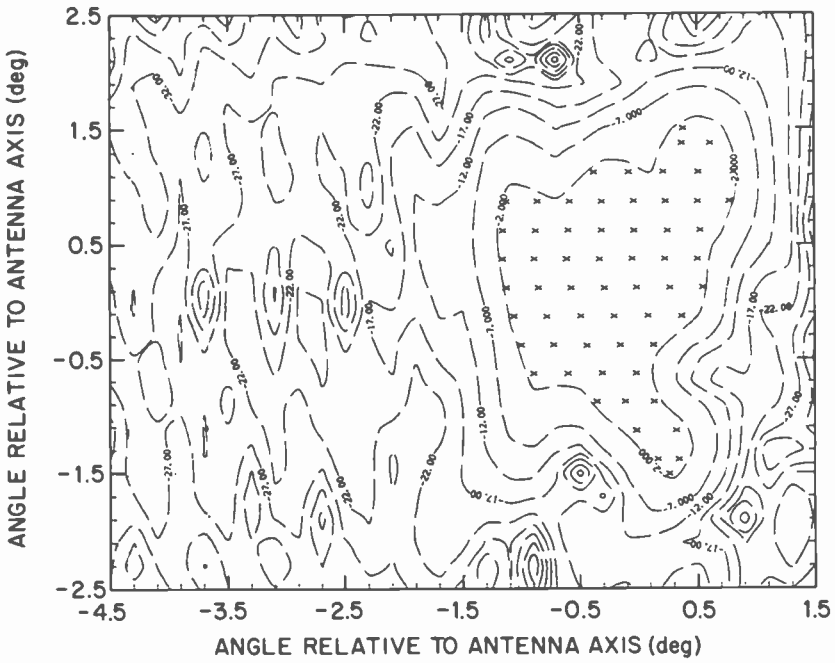


Fig. 11—Contour plot of antenna pattern exterior to service area for 7th-degree, −10.37-dB solution.

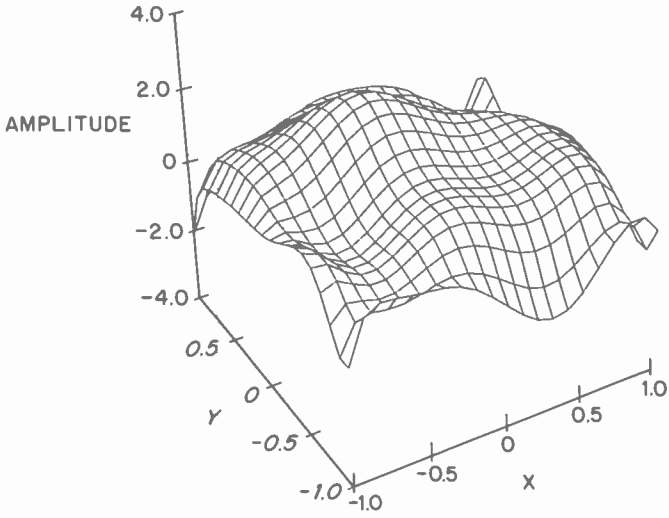


Fig. 12—Perspective plot of amplitude drive function for 7th-degree, −10.37-dB solution.

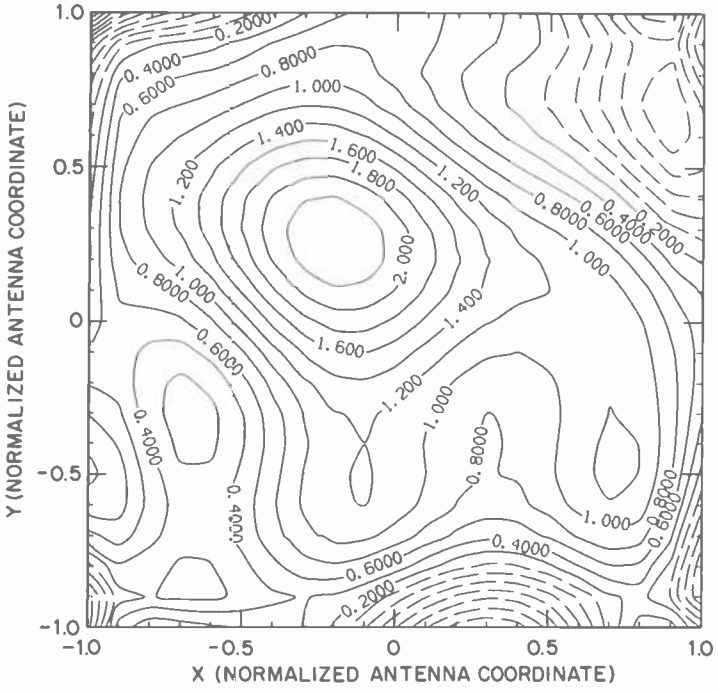


Fig. 13—Contour plot of amplitude drive function for 7th-degree, −10.37-dB solution.

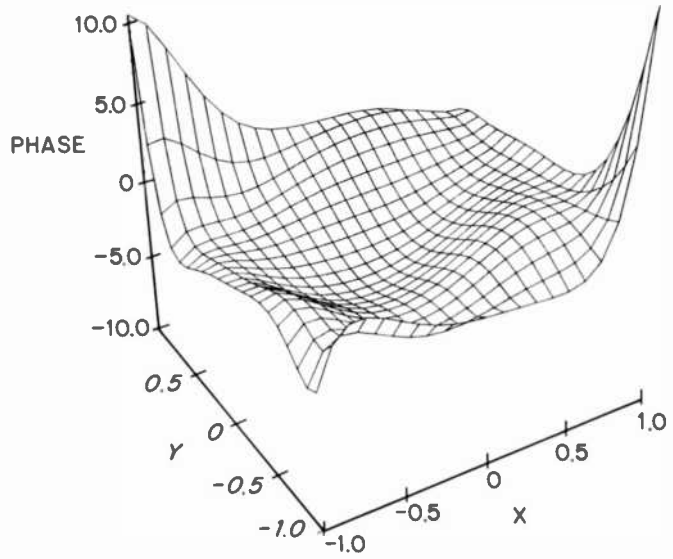


Fig. 14—Perspective plot of phase drive function for 7th-degree, −10.37-dB solution.

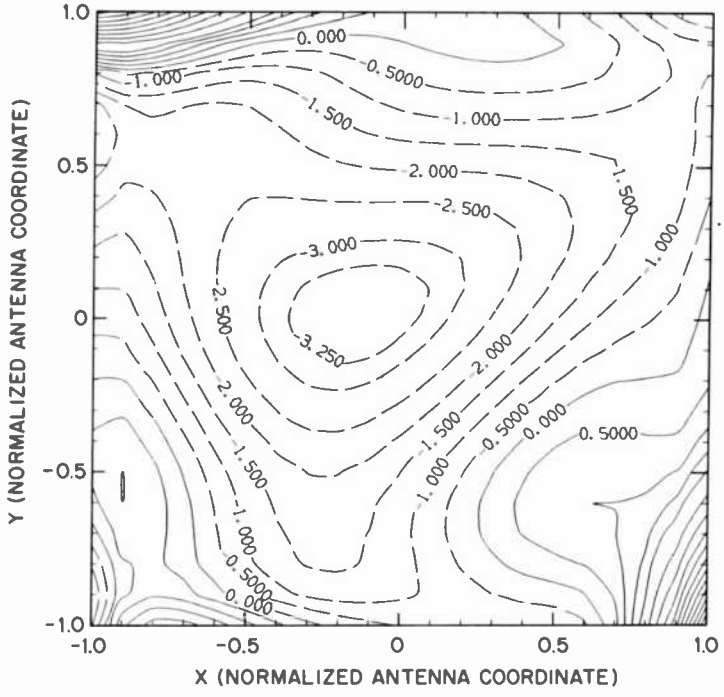


Fig. 15—Contour plot of phase drive function for 7th-degree, −10.37-dB solution.

Acknowledgments

We acknowledge with thanks the valuable contribution of R. W. Klopfenstein, who through numerous conversations helped formulate the problem and interpret the computational results. Much early progress was achieved by Toan Truong, who produced constant-amplitude, phase-varying solutions while working at RCA Astro-Electronics on the MIT Cooperative Program. The assistance of J. E. Keigler, who initiated this work, and the support of J. F. Balcewicz, D. L. Staebler, and N. D. Winarsky, as it progressed, are also gratefully acknowledged.

References:

[1] S. Silver, ed., *Microwave Antenna Theory and Design,* MIT Radiation Laboratory Series, Vol. 12, Mc-Graw Hill Book Co., New York, 1949, Chapter 6.
[2] R. L. Crane, K. E. Hillstrom, and M. Minkoff, "Solution of the General Nonlinear Programming Problem with Subroutine VMCON," ANL-80-64, Argonne National Laboratory, Argonne, IL, 1980.

A Program to Test Satellite Transponders for Spurious Signals

Benjamin R. Epstein
RCA Laboratories, Princeton, N.J.

Abstract—Software is described that has been developed to test for the emission of spurious signals from satellite transponders. The program operates over a wide frequency range and accepts spur specification levels that vary with both frequency and reference bandwidth. Spur specifications as well as spectrum analyzer settings are supplied to the program from ASCII files; this helps to eliminate user intervention, hence errors, during operation of the program. The input file structure also makes it easy to enter changes to the test specification as spur test requirements are modified.

1. Introduction

The spurious radiation testing of rf microwave systems is often an arduous, time-consuming procedure. Satellite transponders are particularly difficult to examine for spurious output because (1) often the test is over a wide frequency range, (2) carrier power levels (which can range from tens to hundreds of watts) are much higher than the spurs to be detected (often well below 0 dBm), and (3) the spur specifications may depend on both frequency and resolution bandwidth. The principal tool used to examine spurious outputs from transponders is the spectrum analyzer. Fortunately, programmable spectrum analyzers now available make the task of spur testing manageable through computer automation.

The software described in this paper has been developed to perform spur tests on transponders to be placed on board high-power direct-broadcast satellites (Satellite Television Corp.), two of which are being developed at RCA Astro-Electronics Division. The software includes a shell that provides an interface between the spur

test algorithm and transponder control programs. In addition, all specifications are supplied to the program via ASCII files, permitting the spur searching to occur under a variety of test conditions while reducing user interaction with the program during runtime.

2. Test Configuration

The spur-search software was developed for the test configuration shown in Fig. 1, where the transponder is the unit under test. Up to three Hewlett-Packard model HP-8673 synthesizers supply discrete test signals to the uplink input of the transponder. These signals range from 17.3 to 17.8 GHz and typically have a total power of less than −40 dBm. After all the signals have been combined, they are sent via coaxial cable to the transponder's input. Within the transponder, a complex manipulation of the signal occurs, including routing, conditioning, and amplification. This translates the input frequencies into a choice of output channels with frequencies ranging from 12.2 to 12.7 GHz at power levels of 50 dBm. Directional couplers route the bulk of the combined signal to air-cooled loads, while the coupled paths supply a sample of the transponder output via coaxial cable to an HP-8566B spectrum analyzer. Combined losses of the coupler and coaxial cable leading to the spectrum analyzer typically range from 40 to 50 dB.

Control of the synthesizers and spectrum analyzer is provided by an HP-1000F computer via an IEEE-488 (GPIB or HPIB) bus. Programming was written in FORTRAN 77 with Hewlett-Packard extensions. The computer also communicates with other computers in the test rack that send commands to, and receive telemetry data

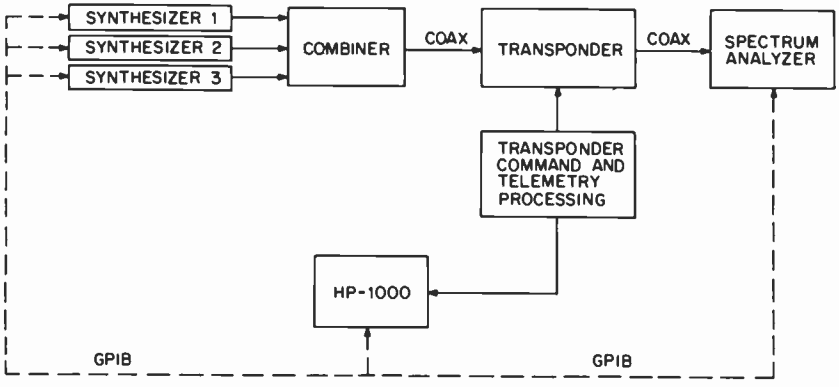


Fig. 1—Overall transponder test configuration. (See text for description.)

from, the spacecraft containing the transponder. This sets the transponder for the desired tests and monitors its status. Although the spur program was created to operate with the transponder test configuration as described, minor modifications to the program allow it to operate in a variety of situations where spur testing is required.

3. Program Shell

Fig. 2 shows a program shell in which the software for performing the spur testing is embedded. The shell sets up the test configuration according to the desired test conditions; its functions include:

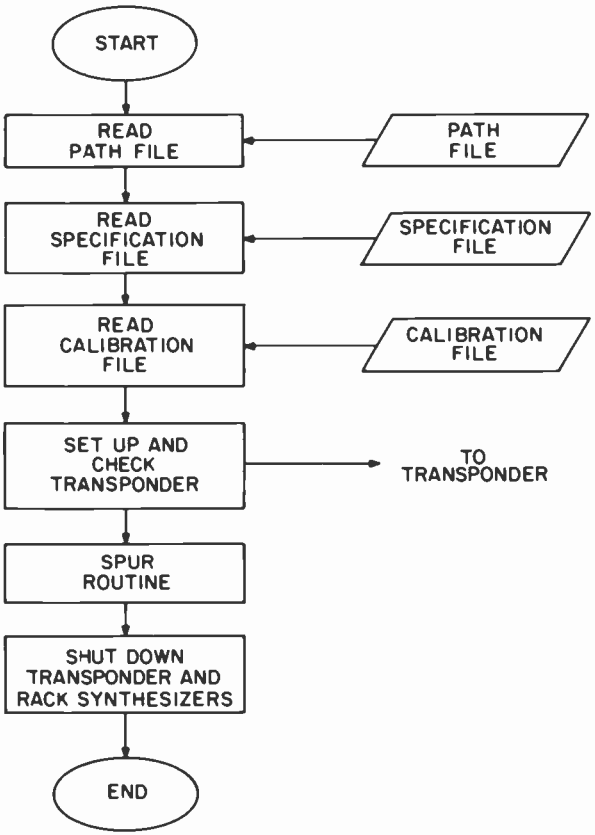


Fig. 2—Flow chart of software shell in which spur test routine is embedded. The shell performs various functions before and after the spur test itself is executed. In place of the spur test, other test routines, such as tests for transponder gain, noise, and distortion, can be embedded in the shell.

- Reading in the *specification file.* The contents of this file include the input carrier frequencies (up to three) for tests under ambient and thermal-vacuum conditions, input power levels, transponder power limits, and the name of the *path file* (which provides transponder configuration information).
- Reading in the *calibration file.* This file is generated during the execution of a separate calibration program that measures and stores all losses, as a function of frequency, between the synthesizers and transponder as well as between the transponder and spectrum analyzer.
- Initializing the transponder. This includes sending instructions to the spacecraft that, among other functions, turns on the appropriate transponder traveling-wave-tube amplifers (TWTAs) and sets signal routing and conditioning within the transponder according to the test requirements. The transponder's configuration for the test is supplied by a *path file,* which is read by the program shell.
- Updating the NAMES file. The NAMES file stores the names of the files containing output data generated by the spur program. The names of the output files are based on the convention by which the first four letters are !DAT followed by a letter that is alphabetically incremented each time the program is run. For example, if the latest output file has the name !DATF, a subsequent execution of the spur program would generate an output file named !DATG. The NAMES file also gives the date and time at which each output file name was created.
- Shutting down the synthesizers as well as the TWTAs and other electronics on board the transponder.

Note that in the above description of the shell, no functions relate specifically to spur testing. In fact, several other rf testing routines have been written that interact with this shell; these routines test for such parameters as transponder gain, group delay, distortion, and noise.

4. Spur Search Routine

This portion of the program performs the actual spur search. First an important input file will be described, followed by an explanation of the program's operation.

4.1 Description of the Band File

The spur search portion of the program first reads in information from an ASCII file called the *band file,* which is generated by the user (normally from an editor) prior to executing the program. The band file specifies what frequency intervals, i.e., bands, are to be searched for spurs. Also, for each band the file specifies the spur specification level and spectrum analyzer settings. The form of the band file is as follows:

```
Header Information (line 1)
Header Information (line 2)
BANDS:
Fstart SpStart
Fstop SpStop
Res Span
NoiseNo NumAv
BndRef
BANDS:
  .
  .
  .
BANDS:
  .
  .
  .
OF:
n
  .
  .
  .
BANDS:
  .
  .
  .
IM:
F1 F2
$
(access information placed here)
```

A typical band file is illustrated in Fig. 3. The first two lines are reserved for header information, which is entered as the user de-

```
SPUR TEST SCAN TEST FILE (Ambient conditions)
JUNE 25, 1985
BANDS:
1.0 55.0
5.0 55.0
1.0 0.15
2 15
0.004
BANDS:
5.0 55.0
10.0 55.0
1.00 0.15
2 15
0.004
OF:
2
BANDS:
10000.0 5.0
11000.0 5.0
30.0 1.5
2 10
1.0
BANDS:
11000.0 10.0
12290.0 10.0
30.0 1.5
2 10
1.0
BANDS:
12310.0 10.0
12390.0 10.0
30.0 1.5
2 10
1.0
```

Continued

```
BANDS:
12410.0 10.0
12490.0 10.0
30.0 1.5
2 10
1.0
BANDS:
12510.0 10.0
12700.0 10.0
30.0 1.5
2 10
1.0
BANDS:
12700.0 10.0
13500.0 10.0
30.0 1.5
2 10
1.0
BANDS:
13500.0 10.0
15000.0 10.0
30.0 1.5
2 10
1.0
IM:
12200 12700
$

LAST ACCESSED: 4:33 JUNE 30 1985
```

Fig. 3—A typical band specification file, named SPURF (See text for details.)

sires. The keyword BANDS indicates that the parameters which immediately follow specify the spur search bands and associated parameters. These are defined as follows:

Fstart = start frequency of band (in MHz)
Fstop = stop frequency of band (in MHz)
SpStart = specification level at Fstart (in dBm or dBc)
SpStop = specification level at Fstop (in dBm or dBc)
Res = resolution bandwidth setting on spectrum analyzer (in MHz)
Span = frequency span setting on spectrum analyzer (in MHz)
NoiseNo = no. of points within the band that are measured to determine the band's noise floor

NumAv = no. of sweeps to be averaged by spectrum analyzer
BndRef = bandwidth to which spurs are to be referenced

Note that separate specifications are given for the start and stop frequencies of each band. This allows for spur specifications that vary linearly with frequency within the band.

The keyword IM indicates the frequency range in which third-order intermodulation products (resulting from the presence of more than one carrier) are to be considered. In the band file description given above, this frequency ranges from F1 to F2 (in MHz).

The keyword OF, which stands for "offset," indicates that the first *n* bands specified describe upper and lower sidebands to be placed symmetrically about each carrier present. In this case, the parameters Fstart and Fstop represent positive frequency offsets from each carrier (for the upper sideband) and their mirror image (for the lower sideband). The use of the offset specification is explained in Section 4.3. During spur searching within the sidebands, measurements are performed relative to the carrier (i.e., in dBc); otherwise, the search is performed in absolute units (i.e., in dBm).

4.2 Operation of Spur Search Routine

To understand how the spur search operates, use the band file SPURF given in Fig. 3 to follow a typical run. Fig. 4 shows the overall flow of the spur searching algorithm. At the beginning, the user is asked to indicate which of the three carriers specified in the specification file are to be present during the test. If more than one are chosen, the third-order intermodulation-product frequencies between F1 and F2 MHz (i.e., 12,200 to 12,700 MHz, as given in SPURF) are computed and stored.

The next step is to set the transponder for the test path desired. Once all transponder diagnostics indicate that the path is properly set, the program proceeds to set the power level and frequency of each synthesizer in such a way that the sum of their power outputs feeds the transponder input at a prescribed input power level (given in the specification file). Next, the power output of the transponder at each carrier is measured by the spectrum analyzer; this is to determinate the dBc levels of possible spurs. (NOTE: If the band file SPURF did not specify offset band measurements, the carrier power measurement would be skipped.)

Once the carriers have been applied, the actual spur search procedure begins, first over the sideband offsets, then over the bands that were specified directly. Fig. 5 presents the flow of the spur

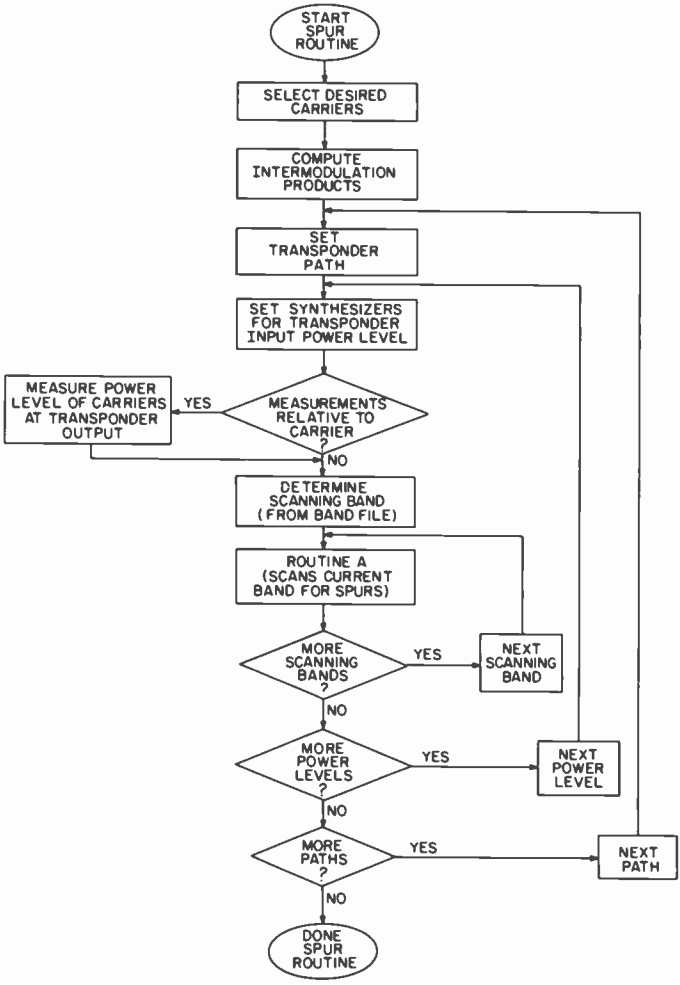


Fig. 4—Flow chart of the spur search routine.

search, as it applies to each band. Prior to the search itself, the noise floor of the band must be established. This is done by sampling the spectrum of the band at discrete frequencies within the band. The number of frequencies in which the sampling occurs is specified by the NoiseNo parameter in the band file corresponding to the band under study; the sampling occurs at equally spaced intervals within the band. The noise level is determined by use of the spectrum analyzer's built-in 1-Hz-bandwidth normalization routine. After normalization to the spectrum analyzer's resolution bandwidth setting (given in the band file), the lowest of the NoiseNo

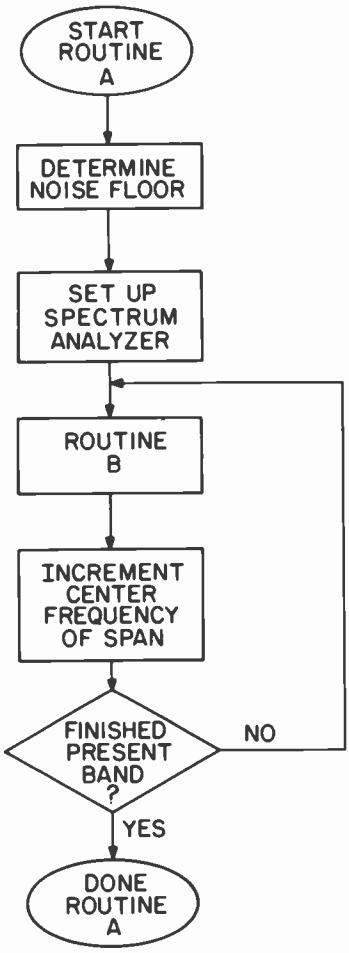


Fig. 5—Flow chart of action taken by spur search routine on each band specified in the band file.

samples is defined as the noise floor. In the case of the example band file SPURF, NoiseNo is 2 for each band; hence, the noise floor of each band is determined by a noise measurement at two frequencies, the minimum of the spectral amplitude resulting in the noise floor value.

After the noise floor is established, the band is broken up into frequency intervals, or spans, which are specified by the Span parameter (in the band file) of the band under study. For example, in SPURF the fourth band specification specifies a band ranging from 11.00 to 12.29 GHz. Because the Span parameter for this band is

1.5 MHz, the spectrum analyzer is set to this span. Consequently, the spectrum analyzer scans (12,290 − 11,000)/1.5 or 860 individual spans, each 1.5 MHz wide, with a resolution bandwidth of 30 kHz (since Res = 30).

During the observation of each span, the HP-8566 spectrum analyzer's internal peak search routine is used to find the amplitude and frequency of all peaks that are greater than 3 dB above the noise floor (see Fig. 6). The peaks result from the averaging of NumAv sweeps by the spectrum analyzer. A running summation of the amplitudes (after correction for all signal losses to the spectrum analyzer) of these peaks is made within a frequency interval specified in the band file for the band under study. Referring to SPURF, the 11.00- to 12.29-GHz band (the fourth band specification) would have these summations over a frequency interval of BndRef = 1 MHz. Whenever the summation exceeds the specification level (in the present example, 10 dBm), the user is notified at the terminal, and note is made on a printer and output disk file (the name of the file being supplied by the NAMES file, as explained in Section 3). If the frequency where the sum exceeds the specification is within 50 MHz of a frequency corresponding to an intermodulation product (the frequencies having been predicted in an earlier part of the program), the summation is considered to be an intermodulation product; otherwise it is considered to be a spur that exceeds the specification. The frequency and amplitude of the spur are then recorded at the terminal, printer, and output disk file. As spurs are detected, the program asks the user if it should continue running or if it should terminate.

After all the spurs (if any) in the span under test have been detected, the next frequency span is analyzed in the same manner. This process is continued until the entire band has been analyzed, after which the next band in the band file is analyzed. In SPURF, after the 11.00- to 12.29- GHz band has been analyzed, the 12.31- to 12.39-GHz band is the next to be analyzed. Fig. 7 shows the contents of a typical output file resulting from a run in which no spurs were detected.

4.3 Scanning Sidebands

In band file SPURF (Fig. 3), the keyword OF followed by a value of two indicates that the first two bands specified represent sidebands on either side of the carriers present. The frequency limits are positive offsets from the carriers (for the upper sidebands) and the mirror images of these offsets about the carriers (for the lower

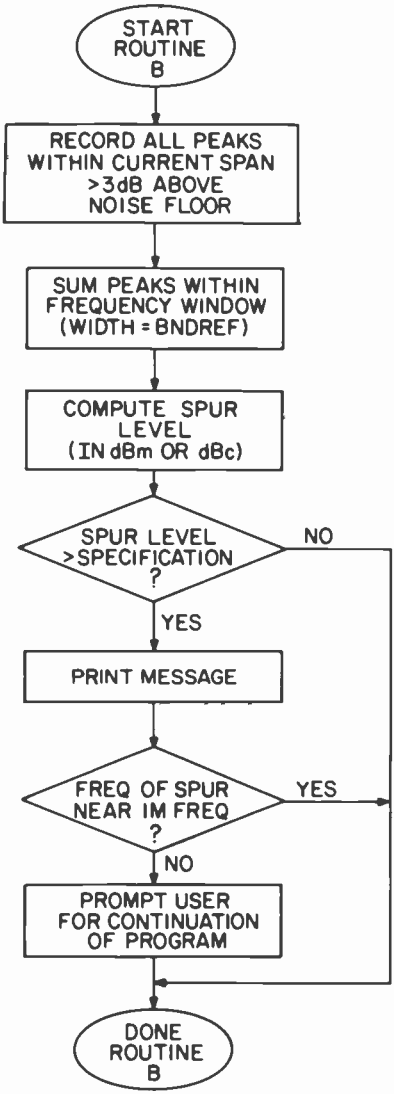


Fig. 6—Flow chart showing details of spur identification and user warning.

sidebands). Referring to SPURF, when three carriers are assumed present at the transponder's output (at 12.3, 12.4, and 12.5 GHz), the program would execute spur searches in discrete bands in the following order:

Frequency band (MHz)	Measurement type
For 12.3-GHz carrier sidebands:	
12301.0 to 12305.0	dBc
12295.0 to 12299.0	dBc
12305.0 to 12310.0	dBc
12290.0 to 12295.0	dBc
For 12.4-GHz carrier sidebands:	
12401.0 to 12405.0	dBc
12395.0 to 12399.0	dBc
12405.0 to 12410.0	dBc
12390.0 to 12395.0	dBc
For 12.5-GHz carrier sidebands:	
12501.0 to 12505.0	dBc
12495.0 to 12499.0	dBc
12505.0 to 12510.0	dBc
12490.0 to 12495.0	dBc
Normal bands to be scanned:	
10000.0 to 11000.0	dBm
11000.0 to 12290.0	dBm
12310.0 to 12390.0	dBm
12410.0 to 12490.0	dBm
12510.0 to 12700.0	dBm
12700.0 to 13500.0	dBm
13500.0 to 15000.0	dBm

Note that, for SPURF, each sideband has been broken up into two bands so that the noise floor can be evaluated more often and over smaller frequency intervals.

4.4 Program Completion

After all of the bands in the band file have been searched for spurs, the process is repeated for the remaining input power levels to the transponder (values that are supplied by the specification file). For all specified power levels, the process is repeated for each transponder path defined in the path file (see Fig. 4). On completion of all the paths, the program shell is reentered; then the transponder and the test rack's synthesizers are shut down.

```
DBS Communications Subsystem Test              3:55 pm May 30 1985
Test Phase:XPNDR AMB Test Matrix #14—Spur Search

  Path Number: 1
Recvr: 1 Channel: A DALC: 4 TWTA: 4 Output MUX: WEST Power level:
-45.30 dBm

For power level -45.30 dBm:
No. of spurs detected: 0
Spec Failures at carriers: 0

No spurs detected during test.

  Test procedure completed. 4:01 pm MAY 30 1985
```

Fig. 7—Contents of a typical disk output file resulting from a test run in which no spurs were detected.

5. Conclusion

A major advantage of the spur program is the use of input files that specify all significant test conditions. The input files can easily be changed according to changes in the spur test requirements. Once these files have been created by an engineer well acquainted with the requirements of a spur test, a user with limited experience can run the program. Note, however, that a judicious choice of spectrum-analyzer parameters must be made in setting up the band file to achieve a spur search that provides the greatest accuracy in the least amount of time. For example, if a band's resolution bandwidth setting is reduced by a factor of three, the time needed to scan that band can increase by a factor of nine. Finally, although this program was intended for use with a specific transponder, minor modifications to the program will allow it to perform spur tests on a variety of rf-microwave systems.

Acknowledgments

The author wishes to thank Dr. Jonathan L. Schepps (RCA Laboratories, Princeton), for his assistance during the development of the software described above; Mark C. Germain (RCA Astro-Electronics Div., Princeton), for his useful suggestions regarding software to perform spur tests; and Karen E. Kuczer (RCA Laboratories, Princeton), for her preparation of the flow charts.

References:

The following references are all publications of Hewlett-Packard Co., Santa Rosa, CA

[1] *HP-8566B Spectrum Analyzer Operation Manual.*
[2] *HP-8566B Spectrum Analyzer Remote Operation Manual.*
[3] Application Note AN 150, "Spectrum Analysis—Spectrum Analysis Basics."
[4] Application Note 150-8, "Spectrum Analysis—Accuracy Improvement."

Patents Issued to RCA Inventors—Second Quarter 1985

April

A. R. Balaban and S. A. Steckler Digital Television Receiver With Time-Multiplexed Analog-to-Digital Converter (4,514,760)
R. L. Berardi Apparatus for Accurately Establishing the Sealing Length of CRT Envelopes (4,507,873)
S. Bloom Mesh Lens Focus Mask for a Cathode-Ray Tube (4,514,658)
D. F. Bowman Selective Waveguide Mode Converter (4,510,469)
D. F. Bowman Dipole Radiating Element (4,513,292)
H. Chen Color Picture Tube Having Reconvergence Slots Formed in a Screen Grid Electrode of an Inline Electron Gun (4,513,222)
H. Chen Inline Electron Gun for High Resolution Color Display Tube (4,514,659)
P. Datta VideoDisc Molding Composition (4,512,916)
S. B. Deal and B. M. Cushman Method of Making a Coded Marking in a Glass Workpiece Employing a Trisilicate Coating and Product Thereof (4,514,456)
W. Denhollander Horizontal Deflection Phasing Arrangement (4,510,527)
C. E. Deyer System for Providing a Multi-Bit Input to a Computer Controlled System (4,513,394)
D. W. Fairbanks Pickup Cartridge Having Stylus Holder (4,513,410)
P. D. Filliman Switching Network With Suppressed Switching Transients (4,513,322)
P. D. Griffis and A. E. Rindal Frequency Translation Phase-Locked Loop for Separate or Intercarrier Type Television Sound Detection (4,511,924)
G. Harbeke and M. T. Duffy Optical Reflectance Method for Determining the Surface Roughness of Materials in Semiconductor Processing (4,511,800)
W. E. Harlan and J. C. Marsh, Jr. Television Receiver With Manual and Selectively Disabled Automatic Picture Control (4,511,921)
S. T. Hsu Method of Forming Self-Aligned Contact Openings (4,512,073)
A. C. Ipri and R. G. Stewart Electrically Alterable, Nonvolatile Floating Gate Memory Device (4,513,397)
E. S. Kohn Television Receiver Adaptable for Descrambler Module (4,509,210)
W. A. Lagoni and W. E. Harlan Video Signal Peaking System (4,509,080)
H. G. Lewis, Jr. Digital Television System With Truncation Error Correction (4,511,922)
E. J. Nossen Fast Correlation System (4,510,579)
M. Nowogrodzki and D. D. Mawhinney Dual Frequency Heart Rate Monitor Utilizing Doppler Radar (4,513,748)
C. L. Olson and J. F. Monahan Black Level Clamp for Television Signals (4,513,321)
S. T. Opresko Cathode-Ray Tube Having an Electron Gun Assembly With a Bimetal Cathode Eyelet Structure (4,514,660)
C. B. Patel Buzz Reduction for Intercarrier Type Television Sound Detection (4,513,323)
D. H. Pritchard and D. J. Sauer CCD Comb Filter Employing Floating Gate Subtractor Techniques (4,510,522)
S. T. Rao and F. Z. Hawrylo Bonding Pads for Semiconductor Devices (4,514,265)
A. E. Rindal Sound Signal and Impulse Noise Detector for Television Receivers (4,514,763)
J. W. Robinson and G. Kaganowicz Glow Discharge Apparatus for Use in Coating a Disc-Shaped Substrate (4,512,284)
L. R. Rockett, Jr. Pulse Duration Modulator Using Charge Coupled Devices (4,508,975)
D. J. Sauer CCD Charge Subtraction Arrangement (4,509,181)
J. H. Schier, Jr. and W. E. Brooks, Jr. Modified Deflection Yoke Coils Having Shoot-back Windings (4,511,871)
H. G. Seer, Jr. Synchronously Detected DC/DC Converter for Regulated Power Supply (4,510,564)
R. P. Stone Arc-Suppression Means for an Electron Gun Having a Split Electrode (4,514,661)

J. H. Thorn and J. J. Prusak Disc Turnover Device (4,512,455)
J. R. Tower Electro-Optical CCD Transversal Filter With Analog Reference Weights (4,514,821)
B. Vanbreemen Rear Projection Television Screen Incorporating a Prism Lens (4,512,631)
S. K. Verma and G. A. Kim Silicon Oxide Lapping Coatings (4,514,192)
L. J. Vieland Electron Beam Injection Structure for Flat Panel Display Devices (4,514, 663)
C. A. Weaver and J. W. Stephens Apparatus and Method for Automatically Controlling an Extruder for Thermoplastic Material (4,510,104)
G. J. Whitley and M. Rayl Component Lead Processing Device (4,513,493)
O. M. Woodward and M. P. Greaves, Jr. Broadband Loop Antenna With Low Wind Resistance (4,510,501)
W. C. Young Device for Indicating the Polar Coordinate Position of a Given Point Anywhere Within a Circular Area (4,513,505)

May

A. R. Balaban and S. A. Steckler Digital Television Receiver With Analog-to-Digital Converter Having Time Multiplexed Gain (4,517,586)
P. C. Basile Transceiver Having Collision Detection Capability for Contention-Formatted FDM Local Area Networks (4,519,074)
J. N. Bleacher and B. M. Cushman Method for Ablating a Coded Marking into a Glass Workpiece and Product Thereof (4,515,867)
D. B. Carlin Multi-Channel Recording/Playback Optics for Laser Diode Arrays (4,520,471)
A. T. Crowley Frequency Synthesizer Using an Arithmetic Frequency Stabilizer and Plural Phase Locked Loops (4,516,084)
L. A. DiMarco Solvent-Cast Capacitive High Density Information Disc (4,515,830)
D. F. Dion and M. J. Cantella Compression of the Dynamic Range of Video Signals (4,516,167)
R. A. Dischert and R. L. Libbey FM Video Transmission System Having Coring and Coring Compensation for Reducing Effects of Triangular Noise (4,520,396)
L. H. Ditty Method for Removing a Phosphor Layer from a Support Surface (4,517,224)
T. L. Glock and T. L. Credelle Modulator Structure and Method for Flat Panel Display Devices (4,517,489)
T. M. Gurley Worldwide Compatible Synchronizing Signal for Accurate Time Base Correction in an Analog Component Interface Standard (4,516,150)
P. E. Haferl Linearity Corrected Horizontal Deflection Circuit (4,516,058)
K. G. Hernqvist Method of Electronically Processing a CRT Mount Assembly to Reduce Arcing and Afterglow (4,515,569)
J. E. Hicks Shutdown Circuit for a Switching Regulator in a Remote Controlled Television Receiver (4,516,168)
W. Hinn and H. R. Fecht Apparatus for Accurate Adjustment of Kinescope Black Level Bias (4,518,986)
S. T. Hsu Method of Fabricating High Speed CMOS Devices (4,519,126)
K. C. Hudson Automatically Adjustable Aperture Stop for Optical Scanning System (4,516,024)
R. N. Hurst, Jr. Video Apparatus With Burst Locked Heterodyne Chroma Converter and Residual Time Base Error Correction (4,520,402)
J. Kane and J. Murr, Jr. CRT With Internal Neck Coating of Crystalline Tin Oxide for Suppressing Arcing Therein (4,518,893)
T. F. Kirschner Disc Record Player Having Player Control Apparatus (4,516,233)
D. K. Krull and S. V. Naimpally Composite Video Signal Processing Apparatus Providing Amplitude Equalization (4,516,153)
M. E. Labib and C. C. Wang Stabilizers for CED Compositions (4,517,117)
P. T. Lin Method and Apparatus for Trimming and Finishing the Outer Edge of a Molded Record (4,518,551)
T. D. Michaelis Optical Disc System Having Momentum Compensation (4,516,231)
R. D. Miller Linear Ramp Voltage Generator Circuit (4,516,036)
W. W. Murphy, 3rd and J. Hettiger Dynamic Noise Filter for an Audio Signal in a Television (4,517,602)
K. D. Peters and C. H. Anderson Modular Guided Beam Flat Display Device (31,894)

F. R. Ragland, Jr. Method for Screening Line Screen Slit Mask Color Picture Tubes (4,516,841)
G. A. Reitmeier Error Concealment System Using Dropout Severity Information for Selection of Concealment Method (4,517,600)
G. A. Reitmeier and C. R. Thompson Synchronization System that Uses All Valid Signals (4,518,996)
C. W. Reno Beam Expansion and Relay Optics for Laser Diode Array (4,520,472)
P. Schnitzler Communication System Compandor (4,518,994)
E. R. Starner Event Selector for Wide Range Probability of Occurrence (4,516,217)
E. M. Sutphin, Jr. Commutating Filter Passing Only the Fundamental Frequency and Odd Harmonics Thereof (4,518,936)
M. L. Tarng and W. S. Romito Method for Determining the Integrity of Passivant Coverage Over Rectifying Junctions in Semiconductor Devices (4,520,312)
R. J. Tarzaiski Dynamic Record Level Control for Optical Disc Recorder (4,516,235)
W. Truskalo Synchronized Switching Regulator for a Multiple Scanning Frequency Video Monitor (4,516,169)
F. Vanhekken and H. Chen Cathode-Ray Tube Having an Asymmetric Slot Formed in a Screen Grid Electrode of an Inline Electron Gun (4,520,292)
J. A. Van Raalte and G. H. Riddle VideoDisc Player With Compensation for Stylus Holder Mechanical Resonance (4,516,232)
H. H. Westerman, Jr. and J. J. Prusak Method and Apparatus for Removing Flash from a Molded Recorded Disc (4,516,301)
H. H. Westerman, Jr. Transfer Apparatus for a Record Molding Press (4,519,767)
D. H. Willis Video Signal Processor With Automatic Kinescope White Balance and Beam Current Limiter Control Systems (4,516,152)

June

M. M. Abdel-Medeid Clock Regenerator Using Two On-Off Oscillators (4,523,158)
A. Acampora Digital Signal Separation Filters (4,524,423)
E. H. Adelson and C. B. Oakley Compatible High-Definition Television System Utilizing Hadamard Basis Functions (4,523,220)
J. Armer Complementary FET Ripple Carry Binary Adder Circuit (4,523,292)
S. L. Bendell Solid-State Color TV Camera Image Size Control (4,521,804)
J. D. Bingley Cold Start Surge Current Limiting System for a Hydrazine Thruster Augmentation Heater (4,523,429)
D. Botez Semiconductor Laser With Non-Absorbing Mirror Facet (4,523,316)
D. Botez Semiconductor Laser With Reduced Absorption at a Mirror Facet (4,523,317)
R. Brown Method of Fabricating an Electronic Circuit Including an Aperture Through the Substrate Thereof (4,520,561)
G. N. Butterwick Silicon Target Support Assembly for an Image Sensing Device (4,521,713)
C. R. Carlson, E. H. Adelson, and C. H. Anderson System for Coring an Image-Representing Signal (4,523,230)
H. Chen Cathode-Ray Tube Having Asymmetric Slots Formed in a Screen Grid Electrode of an Inline Electron Gun (4,523,123)
D. Chin, E. P. Batterman, and T. V. Bolger TV Receiver Circuitry for Performing Chroma Gain, Auto-Flesh Control and the Matrixing of I and Q Signals to (R-Y), (B-Y) and (G-Y) Signals (4,523,221)
J. C. Connolly and D. Botez Semiconductor Laser Having High Manufacturing Yield (4,523,318)
R. L. Covey and M. T. Gale System for Positioning a Body (4,523,850)
F. T. D'Augustine Method for Reducing Degradation of an Optical Image in an Exposure Lighthouse (4,521,501)
N. V. Desai Record Turntable and Polymeric Coating Thereon (4,524,109)
A. G. Dingwall High Speed Sense Amplifier (4,521,703)
D. L. Elde Shaft Seal (4,521,026)
R. D. Faulkner and A. F. McDonie Photoemissive Cathode Formed on Conductive Strips (4,521,175)
A. W. Fisher Method of Making a Multi-Level Metallization Structure for Semiconductor Device (4,520,554)
W. V. Fitzgerald, Jr. Television Receiver Power Supply Regulation Responding to Beam Current Changes (4,525,739)

E. J. Freeman, J. E. Economou, J. G. Pecorari, G. John, J. B. Halter, J. E. James, R. M. Castle, and R. C. Palmer Disc Record with Tapered Groove (4,525,751)
J. E. Gilberg Transient Signal Suppression Circuit (4,525,635)
J. F. Halsey and J. M. Briggs Skylight Structure (4,520,604)
K. J. Hamalainen Automatic Tracking System for Multitrack Recorders (4,525,750)
J. J. Hanak Method of Controlling the Deposition of Hydrogenated Amorphous Silicon and Apparatus Therefor (4,525,375)
R. N. Hurst System for Synchronizing a Video Signal Having a First Frame Rate to a Second Frame Rate (4,523,227)
D. L. Jose, J. Craft, and W. S. Pike Full Wave Rectifier Circuit for Small Signals (4,523,105)
J. Kane Method of Fabricating a Silicon Lens Array (4,524,127)
F. S. Krufka System for Measuring and Identifying Line Spacing on a Curved Surface (4,525,735)
M. E. Labib, R. F., Poll, and C. C. Wang Capacitance Electronic Disc Molding Compositions (4,522,747)
M. E. Labib and E. S. Poliniak High-Density Information Disc Lubricants (4,525,402)
P. A. Levine Automatic Adjustment of the Amplitudes of Plural-Phase CCD Clocking Voltages (4,524,450)
J. J. Nigborowicz and K. L. Friedline Video Source Selector (4,521,810)
J. J. Prusak Method and Apparatus for the Manufacture of Record Stampers (4,522,659)
J. W. Stoughton and D. H. Willis Beam Current Limiting Arrangement for a Digital Television System (4,521,811)
J. C. Tallant, 2nd Automatic Bias Control System With Compensated Sense Point (4,523,233)
G. G. Tamer Synchronization Input for Television Receiver On-Screen Alphanumeric Display (4,524,387)
F. E. Vaccaro and J. Kowalik Shielded Electron Beam Guide Assembly for Flat Panel Display Devices (4,521,714)
G. J. Whitley, M. Rayl, and W. Squire Multiple Component Lead Processing Apparatus (4,520,549)
D. H. Willis Regulated Power Supply Circuit (4,525,411)
D. H. Willis and J. S. Fuhrer Digital Signal Processing Apparatus Having Digital Dither (4,524,447)
C. M. Wine Frequency Synthesis Tuning Control System for a Double-Conversion Tuner (4,521,916)
R. F. Wood, Jr., J. F. Monahan, and P. A. Levine Dark Current Measurement and Control for Cameras Having Field-Transfer CCD Imagers (4,525,743)
C. P. Wu Alloying of Aluminum Metallization (4,525,221)

AUTHORS

Robert Amantea received the BSEE degree from the City University of New York in 1965 and the MS degree in Electrophysics from the Polytechnic Institute of New York in 1969. He was a David Sarnoff Fellow at the University of California at Berkeley, which awarded him a PhD in 1977. Dr. Amantea joined RCA Laboratories, Princeton, NJ, in 1965, working in the area of power semiconductor devices. He has authored several papers on such topics as power transistors, TRAPATT devices, gated-diode physics, gate-turn-off thyristors, and computer-aided modeling of MOS transistors. He holds three U.S. patents. His more recent interests have been in the area of semiconductor device modeling, particularly, short-channel MOS transistors and buried-channel CCDs.

Dr. Amantea was a visiting lecturer at the University of California at Berkeley during the Spring 1977 quarter. He has been lecturing at LaSalle College since 1982. He is a member of IEEE, AAAI, Eta Kappa Nu, and Tau Beta Pi.

Samuel H. Colodny received the BSEE degree from Catholic University, Washington, DC, in 1942 and the MSEE degree from University of Pennsylvania in 1952. He joined RCA Astro-Electronics, Princeton, NJ in 1981 as a Principal Member of the Technical Staff and has been active in active antenna array synthesis and in satellite communication system engineering. Mr. Colodny is a Senior Member in the IEEE and has published several papers in IRE/IEEE publications. He has three patents to his credit.

Roger L. Crane received a BSEE degree from Iowa State University, graduating as the valedictorian of his class in 1956. Following three years' employment as an engineer at Sperry Rand's Univac Div., he returned to Iowa State and received the MS and the PhD degrees in Mathematics in 1961 and 1962, respectively. As a Member of the Technical Staff of RCA Laboratories, Princeton, NJ, in the summers of 1961 and 1962, Dr. Crane was engaged in scientific problem solving and research in the numerical solution of ordinary differential equations. He then taught mathematics at Iowa State for one year, returning to RCA Laboratories in 1963. He has been doing research in numerical analysis and algorithm development and in the use of digital computers for solving scientific and engineering problems. In 1978, while still an employee of RCA, he spent a year as a Resident Associate at the Argonne National Laboratory, studying mathematical optimization theory and developing optimization software.

Research and application areas in which Dr. Crane has made significant contributions include the design of circuits and logic for digital computers, algorithms for the numerical solution of ordinary differential equations and for solving nonlinear constrained optimization problems, computer system simulation and performance analysis, all-pass network syn-

thesis, optimum design of television receivers to minimize colorimetric errors, lighthouse lens design for the manufacture of color kinescopes, analysis of noise in towed underwater antennas, the design of dielectric interference filters, and the design of phased-array antennas with shaped-beam radiation patterns.

Dr. Crane has received four RCA Laboratories Outstanding Achievement Awards and, in 1973, was appointed a Fellow of the Technical Staff. He has published several papers in his fields of expertise. He is a member of the IEEE, the Society for Industrial and Applied Mathematics, the Association for Computing Machinery, the Mathematical Programming Society, and several honorary societies including Sigma Xi, Tau Beta Pi, Eta Kappa Nu, Pi Mu Epsilon, and Phi Kappa Phi.

Walter R. Curtice received the BEE, MS, and PhD degrees from Cornell University in 1958, 1960, and 1962, respectively. The subject of his Master's thesis was the ruby maser; research for his doctoral thesis was concerned with noise in linear-beam microwave devices. In 1962, he joined the Raytheon Microwave and Power Tube Division as a Senior Research and Development Engineer. He participated in microwave tube development and performed experimental and theoretical research on linear-beam and cross-field devices. In 1967, he became Visiting Assistant Professor of Electrical Engineering at the University of Michigan, and in 1969 was appointed Associate Professor. In addition to teaching courses on physical electronics and microwave measurements, he was engaged in sponsored research on microwave semiconductors with emphasis on transferred-electron devices. In 1973, Dr. Curtice joined RCA Laboratories, Princeton, NJ, as a Member of the Technical Staff, in the Microwave Technology Center. Initially, he directed the development of second-harmonic-extraction TRAPATT amplifiers for X-band operation. He later developed the two-dimensional electron-temperature model of GaAs field-effect transistors, an improved MESFET model for GaAs integrated-circuit simulation and, more recently, has been directing the nonlinear-device modeling effort. In 1984 he received an RCA Laboratories Outstanding Achievement Award for the development of advanced techniques for the computer simulation of III-V compound FETs.

Dr. Curtice has written more than 40 technical papers, has eight U.S. patents issued to him, and is a Senior member of the IEEE. He is also a member of Tau Beta Pi, Etta Kappa Nu, and Sigma Xi. He was chairman of the IEEE's Boston chapter of the Electron Devices Society from 1966 to 1967 and chairman of the Southeastern Michigan section of the combined MTT, ED, and AP chapters for 1972. For 1985–86 he is chairman of the combined MTT/ED chapters at Princeton.

Gary M. Dolny received the BS degree in Physics from Haverford College in 1977 and the MS and PhD degrees in Electrical Engineering from the University of Pittsburgh, in 1978 and 1981, respectively. From 1981 to 1984 he was a member of the faculty of the Department of Electrical Engineering, Wilkes College, where he taught courses in solid-state physics and semiconductor device fabrication. During this time he also worked with the RCA Solid State Division, Somerville, NJ, where he developed computer-aided techniques for the analysis and design of high-voltage transistor structures. In 1985, Dr. Dolny joined RCA Laboratories, Princeton, NJ, as a member

of the Technical Staff in the Advanced Silicon Technology group. His responsibilities involve research on power integrated circuits, COMFET development, and silicon process technology.

Dr. Dolny is a member of the IEEE and has published numerous papers in the areas of solid-state physics, semiconductor devices, and computer modeling and simulation.

Benjamin R. Epstein received the BS degree in Electrical Engineering (1978) from the University of Rochester and the MS and PhD degrees in Bioengineering (1980 and 1982, respectively) from the University of Pennsylvania. His PhD dissertation was titled "Dielectric studies on microemulsions." During 1982–83 Dr. Epstein served as a Postdoctoral Fellow at the Centre National de la Recherche Scientifique, Thiais, France. In 1983 Dr. Epstein joined the Microwave Technology Center of RCA Laboratories, Princeton, NJ, as a Member of the Technical Staff. He has been working on the development of computer-automated testing and manufacturing systems for microwave and millimeterwave circuits, as well as medical applications of microwaves.

Dr. Epstein is a member of the IEEE, New York Academy of Sciences, AAAS, Automated RF Techniques Group, Bioelectromagnetics Society, and Tau Beta Pi.

Morris Ettenberg received a BA degree in Mathematics in 1935 and an MS degree in Mathematics in Education in 1936, both from City College of the City University of New York. He earned a PhD in Physics from New York University in 1950. From 1942 to 1945 he worked as a radar engineer at the U.S. Navy Yard in Brooklyn. He then joined Sperry Gyroscope as an engineer and, from 1955 to 1958, headed Sperry's Department of Power TWT Development. During 1958–63 Dr. Ettenberg served as a research professor at the Polytechnic Institute of New York. He then joined the faculty of City College and, from 1968 to 1973, served as Chairman of the Electrical Engineering Department.

In addition to being a faculty member at City College, Dr. Ettenberg has been a consultant for RCA Laboratories since 1980. His research and publications are in the field of microwave tubes and semiconductor devices. He is a Life Senior member of the IEEE.

Beatrice M. Hwong received the BS degree from Tufts University in 1965, the MS degree in Electrical Engineering from Cornell University in 1967, the MSE degree in Computer Information and Control Engineering from the University of Michigan in 1974, and the PhD degree in Computer Engineering from Rutgers University in 1981. Work for the last two degrees was supported by a Danforth Fellowship for Women.

From 1967 to 1972 she worked at the AT&T Technologies Engineering Research Center, conducting integrated-circuit bonding studies, mask generation for IC bonded crossovers, and language specification for the automatic testing of printed-circuit boards. In 1973, at COMSHARE, she was engaged in applying a database language to a personnel information system. She joined RCA at

Somerville, NJ, for the summer of 1977 as an Associate Member, Technical Staff, in the Microcomputer Systems group, rejoining in 1978 as a Member, Technical Staff. There she was engaged in developing operating systems software and high-level language compilers/interpreters for the COSMAC development system. In 1984 Dr. Hwong transferred to RCA Laboratories, Princeton, NJ, in the Integrated Circuit Research Laboratory. Since then she has been developing software tools to simulate IC processing, devices, and circuit performance.

Dr. Hwong has written two technical papers. She is a member of the Association for Computing Machinery and the IEEE, and has served on the IEEE Electro Film Committee. She is also a member of Tau Beta Pi and Eta Kappa Nu.

Dietrich Meyerhofer studied Engineering Physics at Cornell University and received the PhD degree in Physics from MIT in 1957. In 1958 he joined RCA Laboratories, Princeton, NJ, as a Member of the Technical Staff, becoming a Senior Member, Technical Staff in 1983. His research has been concerned with galvanomagnetic and transport properties in semiconductors and insulators, with nonlinear optical effects and the properties of infrared lasers, and with holographic processes and materials. For two years, he was a member of an applied research group that was concerned with electronic and optic applications to printing. This involved electronic photocomposition and holographic character generation, laser exposure of various kinds of printing plates, and electronic screening of black-and-white and color images for printing. The possibility of direct etching of gravure cylinders by *Q*-switched laser was demonstrated.

In recent years Dr. Meyerhofer has been concerned with testing and characterizing resists for microlithography. In the case of the positive resist systems based on diazoquinone sensitizers in novolak resins, he studied the effect of optical exposure on the destruction of the sensitizer and the relationship between sensitizer and development rate (contrast and sensitivity). Exposures to UV radiation and to electron beams were compared. The effect of some resist parameters on resist profiles has been determined for commercial resist systems. To support the research on microlithography, modeling programs were developed that allow computer simulation of the entire process. Dr. Meyerhofer has also been providing consultation to the RCA IC mask-making facility and the optical and electron-beam lithography operations at Somerville, NJ. He has been evaluating commercial lithography equipment of all kinds.

Dr. Meyerhofer holds 10 U.S. patents and has written more than 50 technical papers. He is a member of the American Physical Society, the IEEE, and Sigma Xi.

Stewart M. Perlow received a BEE degree from the City University of New York in 1963, and an MSEE degree in 1970 from the Polytechnic Institute of New York. He joined RCA at the Advanced Communications Laboratories, New York, in 1963. For the next three years, in the Microwave Circuits group, he worked on low-frequency active-filter techniques, parametric upconverters, parametric amplifiers, and UHF solid-state power amplifiers. He was cited for achievements as an outstanding young engineer. In 1966 he became a cofounder of National Electronic Laboratories, which was purchased by Harvard Industries in 1967. When Harvard acquired

KMC Semi-conductor in 1968, Mr. Perlow became Chief Engineer of KMC. His responsibilities included directing the evaluation, specification, and testing of new semiconductor devices and the design and manufacture of ultralow-noise and broadband solid-state amplifiers and receiving systems covering the complete radio-frequency spectrum. Since joining RCA Laboratories, Princeton, NJ, as a Member of the Technical Staff in 1973, Mr. Perlow has been involved in TV tuner design, nonlinear distortion analysis of rf signal processors, and computer-aided design and measurement.

Mr. Perlow holds one U.S. patent. He has written 26 technical papers, some of which were presented at conferences, and two book chapters. He received RCA Laboratories Outstanding Achievement Awards in 1980 and 1982, respectively. He is a member of Eta Kappa Nu and a Senior member of the IEEE.

Harold R. Ronan, Jr. earned a BSME degree from the Massachusetts Institute of Technology in 1952 and an MSEE degree from the New Jersey Institute of Technology in 1960. During 1953–55, when he held the rank of 1st Lt. U.S. Army Ordnance Corps, he worked on artillery ammunition design and radar and fire control equipment. In 1959 Mr. Ronan joined RCA at Somerville, NJ, as a Member of the Technical Staff. He was promoted to Senior Member, Technical Staff, in 1968 and, in 1972, transferred to the RCA facility at Mountaintop, PA. From 1959 to 1981 he worked on a variety of projects, including the design of automatic computer-controlled test equipment for linear and digital integrated circuits and discrete devices, i.e., CCAT and SCOPE systems; machines and controls for processing equipment, such as photo exposure and crystal pulling; and automatic and semi-automatic power device test circuits, such as $I_{s/b}$SOA for bipolars, high-temperature leakage testing for thyristors, and rectifier diode-switching time equipment.

Since joining the Power Applications group at Mountaintop in 1981, he has been involved in characterizing and testing surgectors, power FETs, and COMFETs. More recently, he has also been serving on the JEDEC Committee, establishing test methods for power MOS devices.

C. Frank Wheatley, Jr., received a BSEE degree from the University of Maryland in 1951. After graduation he joined RCA at Camden, NJ, under an RCA training program. Working as an engineer at Harrison, NJ, from 1952 to 1956, he then transferred to the newly formed RCA Solid State Division, Somerville, NJ. In 1958 he was named a Member, Technical Staff, and in 1963, he was promoted to Senior Member, Technical Staff. Mr. Wheatley transferred to the RCA facility at Mountaintop, PA, in 1980. During his 34 years' work with solid-state devices, he has achieved many firsts. In 1955, he designed and reduced to practice what may be the first monolithic power integrated circuit, the integration (on a common die) of a germanium power transistor, and a "Barton" compensating diode. This was done to improve thermal coupling between the devices. As an applications engineer, he designed industry's first production-proven high-fidelity stereo amplifier, the Heathkit AA21, and codesigned industry's first production-proven all-transistor automobile radio. Mr. Wheatley published extensively on transistorized TV, with major contributions in the area of failures from transients for the horizontal deflection circuitry. The monolithic power

Darlington transistor he codesigned has found considerable industry acceptance. His work also includes a two-terminal all-electronic temperature sensor and a thermally ballasted bipolar power transistor. More recently, he and a co-worker have developed a power IC process and an operational amplifier die, which delivered 50 W of continuous sinewave power into a 10-Ω load, while maintaining all the other characteristics typical of op amps. Since 1979 his responsibilities have included the device design of power MOSFETs, COMFETs, and power bipolar transistors.

Mr. Wheatley has been issued 48 U.S. patents, among them the basic patents on the operational transconductance amplifier (OTA) and the COMFET. He has written 43 technical papers, has presented 18 papers at IEEE conferences at home and abroad, and has won two RCA awards and one IEEE Best Paper Award. He is a Senior member of the IEEE.