Developing Linear Error Models for Analog Devices

Gerard N. Stenbakken, *Member, IEEE*, and T. Michael Souders, *Senior Member, IEEE*

***Abstract*—Techniques are presented for developing linear error models for analog and mixed-signal devices. A simulation program developed to understand the modeling process is described, and results of simulations are presented. Methods for optimizing the size of empirical error models based on simulated error analyses are included. Once established, the models can be used in a comprehensive approach for optimizing the testing of the subject devices. Models are developed using data from a group of 13-bit A/D converters and compared with the simulation results.**

I. Introduction

TESTING of analog and mixed-signal devices is a very critical element in the design cycle of many new products. Test engineers must develop tests for these devices which can accurately sort the good devices from the not so good, with a minimum of cost and time. The process of developing an efficient test plan always involves a tradeoff of the test confidence versus the test cost. This process requires getting the maximum information possible from the tests that are performed.

Over the past several years, a comprehensive approach has been developed at NIST for maximizing the tradeoffs associated with production testing of analog and mixed-signal devices [1]–[4]. The approach is based on a simple relation between device errors measured at a small number of test points and device errors at a large number of test points that are predicted from those measurements. The simple relation is based on a linear coefficient matrix model.

One point that is emphasized in this paper is a technique for optimizing the resultant model. When an empirical model is being developed, a criterion is needed to determine the optimum number of model parameters. Simulations of error modeling show that if too few parameters are used in the model, then some of the true error structure cannot be described. On the other hand, every parameter added to the model increases the amount of measurement noise that is modeled as device error. Thus, there is an optimum number of parameters that should be used to estimate the errors of a device. This paper describes how to select a stopping criterion for empirical models that will minimize the rms difference between the predicted device error and the true device error.

Once an accurate model has been developed, algebraic operations on the model can be used to:

1) select an optimum set of test points which will minimize the test effort and maximize the test confidence,
2) estimate the parameters of the model (i.e., the error variables) from measurements made at the selected test points,
3) predict the response of the device at all candidate test points from measurements made at the selected test points, and
4) calculate the accuracy of the parameter estimates and response predictions, based on the random error in the measurement.

The purpose of this paper is to describe the procedures used to develop empirical models. Two methods are described: QR decomposition (QRD) and singular-value decomposition (SVD). A program that uses these methods to simulate a device production line is described, and results are presented. These modeling techniques are applied to data from 13-bit A/D converters and the results compared with the simulation results.

Manuscript received May 18, 1993; revised November 19, 1993.

The authors are with the National Institutes of Standards and Technology, Gaithersburg, MD 20899 USA.

IEEE Log Number 9215962.

II. QR Decomposition Model Building

Consider the problem of designing a test for a device which has error limits set at m test points. Make error measurements at all m test points and collect the errors into a vector called $\boldsymbol{y}$. Thus, $\boldsymbol{y}$ is an m-by-1 vector of measured error values. Assume that the errors are related to the process parameters $x^1, x^2, \cdots, x^p$ of the production line by a linear error model $\boldsymbol{A}$, and the measurements of the device are subject to an error of ε. Then the measurements on a device can be expressed as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \varepsilon \tag{1}$$

where $\boldsymbol{A}$ is an m-by-p matrix, $\boldsymbol{x}$ is a p-dimension vector, and ε is an m-dimension random vector with normally distributed elements with a mean of 0 and standard deviation of σ. In general the dimension of the test space m is much larger than the dimension of the parameter space p.

To build an empirical model take n devices from the production line and measure their errors. Let $\boldsymbol{Y}$ be the matrix of these n vectors, i.e.,

$$\boldsymbol{Y} = (\boldsymbol{y}^1, \boldsymbol{y}^2, \cdots, \boldsymbol{y}^n). \tag{2}$$

Use the QRD algorithm to decompose the $\boldsymbol{Y}$ matrix into two matrices. The QRD process sequentially selects or pivots the largest error vector $\boldsymbol{y}^i$ and orthogonalizes the remaining vectors to it. Denote the QRD of $\boldsymbol{Y}$ as

$$\boldsymbol{P}\boldsymbol{Y} = \boldsymbol{Q}\boldsymbol{R} \tag{3}$$

where P is the pivot matrix, Q is an orthonormal m-by-m matrix, and R is an m-by-n right triangular matrix. The pivot matrix P reorders the columns of Y (the device order) such that the diagonal of R is monotonically decreasing. This pivoting puts these columns in approximately their order of importance. The first k columns of Q (orthogonalized vectors of Y) can be used to estimate the space spanned by the linear error model A. To select models of various sizes, the vectors of Q are selected in order until the desired sizes are obtained. If Q_k denotes the first k columns of Q,

$$y^i \approx Q_k z^i + \varepsilon^i \tag{4}$$

where the elements of z do not correspond one-for-one with the x's, but the column vectors of Q_k span approximately the same space as A. The simulation results described below show how to select an optimal value for k.

III. Singular- Value Decomposition Model Building

The other method that can be used to build an empirical model is singular-value decomposition (SVD). Use the SVD algorithm to decompose the Y matrix into three matrices. The SVD process selects those linear combinations of error vectors y which are best able to describe the variation in Y and orthogonalizes them. Denote the SVD of Y as

$$Y = USV^T \tag{5}$$

where U is orthogonal m by m, S is diagonal m by n with $\min(m, n)$ monotonically decreasing singular values, and V is orthogonal n by n. The first k columns of U (orthogonalized linear combinations of vectors of Y) can be used to estimate A. The SVD process puts these columns in their order of importance. To select models of various sizes, the vectors of U are selected in order until the desired size is obtained. If U_k denotes the first k columns of U, then

$$y^i \approx U_k z^i + \varepsilon^i \tag{6}$$

where again the elements of z do not correspond one-for-one with the x's, but the column vectors of U_k span approximately the same space as A.

Given an empirical model Q_k or U_k, a good set of test points can be selected using QRD on the transpose of Q_k or U_k, thus selecting rows of the empirical models or test points. Additional test points can be selected by calculating the prediction variance and sequentially selecting the test point corresponding to the maximum prediction variation [3]. The selected parts of Q_k and U_k are denoted as $\tilde{Q}_k$ and $\tilde{U}_k$ and the corresponding measurements for device i as $\tilde{y}^i$. These techniques are more fully described in [4].

IV. Model Simulation Program

A model simulation program was written to gain a better understanding of the relationships between the various factors involved in empirical modeling. This program simulates a production line that can produce devices. The data from simulated measurements on these devices can be used to study the effects of various modeling factors on how well these estimated models match the true model. Among the modeling factors that can be studied with this program are the number of model parameters, their size relative to the noise level, the total number of test points, the number of selected test points, and method of selecting test points. This program was written in a programming language designed for linear algebra called CLAM (TM).[1] Fig. 1 shows a block diagram of the model simulation program. A brief description of the program is given next, followed by a description of some of the simulations run using the program.

The true model description is a set of parameters that gives the characteristics of the true model. The simulations described below used a true model with random vectors. This model allows an arbitrary number of model vectors, an arbitrary number of test points, and an arbitrary size of each vector. The individual vectors are created from elements drawn from a uniform distribution. For real production devices this step in the simulation is equivalent to developing a processing line that builds devices. The true model describes the amount and type of process variations that result from this hypothetical production line.

The program generates two sets of random vectors which conform to the true model. This step in the program is equivalent to producing a number of devices on the production line and setting aside one group to build an empirical model and another group to validate the model. Random noise is added to the true model and validation sets to form the noisy model and validation sets. This corresponds to the noise introduced during the process of taking measurements on these devices. Measurements from real devices can also be analyzed with the simulation program. Data from real devices, shown in Fig. 1 as a predefined set, can be split into the noisy model and validation sets. In this case the true validation set (TVS) is unknown, so some later analyses cannot be performed.

The empirical model is derived from the noisy model set using either QRD or SVD analysis [5], [6]. Empirical models with a varying number of parameters (e.g., k equal to the range 1 to $\min(m, n)$ in (6) above) are used in the subsequent analyses. For each empirical model a good subset of test points is selected by minimizing the prediction variance [3] of each submodel (shown as Select Test Points in Fig. 1). These selected test points represent the small number of test points that would be measured in a real production operation.

The simulated measured errors of the corresponding test points of the noisy validation set are fitted to those of the empirical submodel. The resultant fit parameters are referred to as the Predicted Validation Parameter Set in Fig. 1. Since the number of selected test points is in general larger than the number of model parameters, a least-squares approach is needed to calculate these parameters. Thus the estimated

[1] Certain commercial software is identified in this paper in order to adequately describe the program developed. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that the software is necessarily the best available for the purpose.

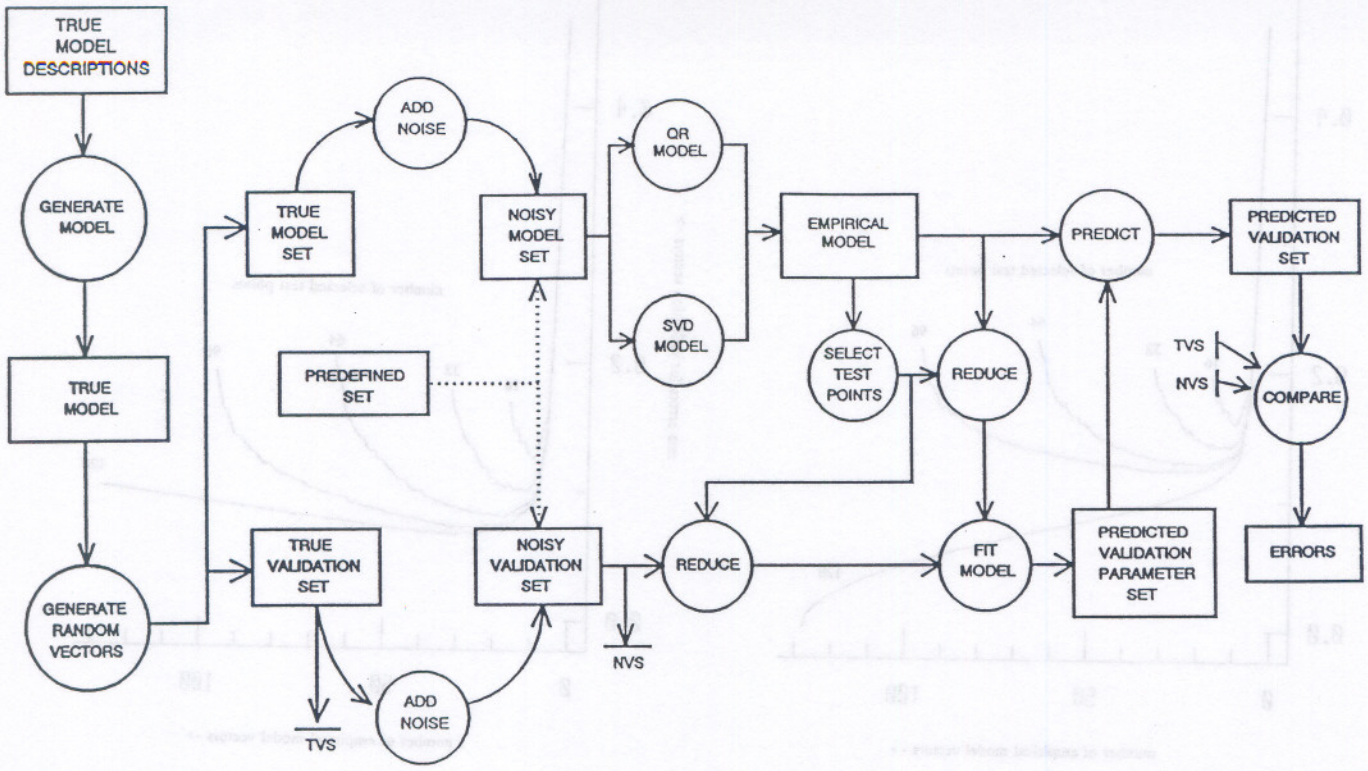

Fig. 1. Block diagram of simulation program.

parameters can be computed as

$$\hat{z}^i = (\tilde{U}_k^T \tilde{U}_k)^{-1} \tilde{U}_k^T \tilde{y}^i, \tag{7}$$

for the SVD models (use $\tilde{Q}_k$ in place of $\tilde{U}_k$ for the QRD models). Note that the estimated parameters $\hat{z}^i$ are a function of the selected test points and the number of vectors k used in the model. The parameters are multiplied by the full empirical model to generate the predicted validation set. This predicted set is compared with the true validation set (TVS in Fig. 1), if available, and the noisy validation set (NVS) to determine the adequacy of the modeling process.

V. Simulations

The simulations were used to address four questions about empirical modeling and testing. First, what is the optimum number of empirical vectors to use in the model? In the past, the number of vectors used in a model was determined by examining the r-diagonal values from the QRD [4] or the singular values from the SVD. When the r-diagonal value approached the level of the measurement noise, no additional vectors were selected. This method gave good results but was not optimal. The simulations, however, show another method which appears to be optimal. Second, how does the reduced number of test points affect the accuracy of the results and the size of the optimum empirical model? Third, what are the relative advantages of QRD versus SVD? QRD is roughly three times faster than SVD, but how do the accuracies and sizes of the model compare? Fourth, what are the effects of the noise level on these results?

TABLE I
Parameters for Simulations

Number of total test points = 128
Number of true model vectors = 10
Type of true model vectors = random
Distribution of vector elements = zero mean, uniform
Size of true model vector elements
4 with maximum amplitude of ±0.50 and rms of 0.29
2 with maximum amplitude of ±0.25 and rms of 0.14
2 with maximum amplitude of ±0.10 and rms of 0.06
2 with maximum amplitude of ±0.05 and rms of 0.03
Distribution of parameters = normal
rms of parameters = 1
Standard deviation of resultant device vectors = 0.62

The simulations suggest answers to these questions, and give indications of where to look for analytical solutions. The following simulation results are from runs with true model vectors and parameters as described in Table I. In each case 130 random devices were used in the true model set, and the same number were used in the validation set.

Fig. 2 shows the average difference between the predicted error and the measured error for the validation set using models derived with QRD and an rms noise level of 0.1 at each test point. This noise level is about one sixth of the size of the true model contribution at each point. Note from Table I that this noise level is larger than the rms of the four smallest true model vector elements (0.03 and 0.06 in Table I). The horizontal axis of Fig. 2 is the number of vectors k used in the empirical model, and the vertical axis is the rms error of all test points for all devices. Results are plotted for five different

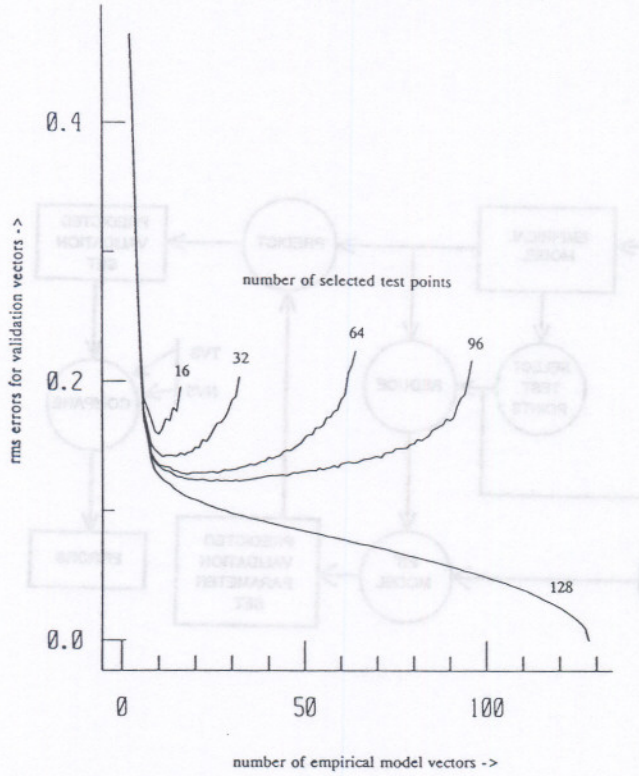

Fig. 2. Rms difference between predicted values and measured values using QRD derived models and an rms measurement noise of 0.1.

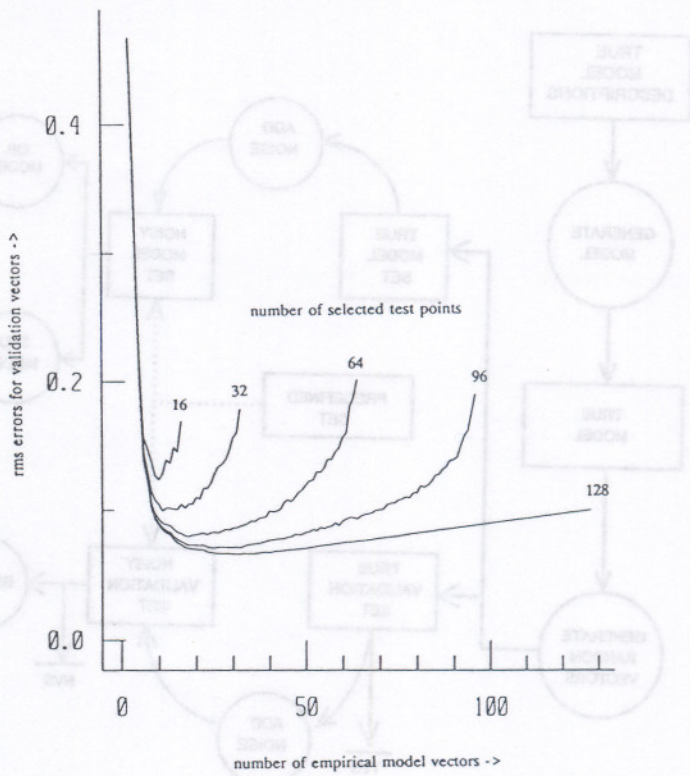

Fig. 3. Rms difference between predicted values and true values using QRD derived models and an rms measurement noise of 0.1.

test point sets from 16 to a full set of 128 test points. The full test point set gives the lowest difference for any model size and goes to zero for a complete model of 128 model vectors. This is to be expected since with 128 points to predict and a full set of 128 parameters the measured data can be exactly represented. Each of the reduced test point curves has a model size that gives a nontrivial minimum difference. These minima, all slightly above the noise level of 0.1, show the model size which gives the closest prediction to the full set of measured values.

The optimum model size, however, is the one which gives the minimum rms difference between the predicted values and the true values, rather than the measured values. The results given in Fig. 3 are similar to those described above but show the difference between the predicted errors and the true errors for models derived using QRD. This difference can be computed for simulated results but cannot be known for real devices. For any model size, the full set of 128 test points gives the smallest difference and for a complete model of 128 model vectors goes to 0.1 which is the measurement noise level added to the true values. All of these test point sets have a model size that gives a minimum difference, including the full test point set. Comparing these results with those of Fig. 2 shows that predicted values of the four reduced test points sets are always closer to the true values than to the measured values. The minimum difference for 32 test points is at the measurement noise level of 0.1, and the minima for 64 and 96 test points are lower than the noise level. These error levels are below the measurement noise level.

Recall that the measurement noise level is the difference between the measured values and the true values. This noise level is the error obtained when all test points are measured and no model is used. By using a model a lower error level can be obtained. Thus, by making use of the constraints on the device errors that are expressed in the models, a reduced number of test points can give results closer to the true values than measuring all test points and not using a model. At first this result seems contradictory; one can take fewer measurements and get a better result than by taking more measurements. This result shows the power of the model to give a better prediction than not using a model.

The results obtained when SVD is used to derive the empirical models are given in Figs. 4 and 5. These graphs show that the differences obtained using SVD derived models are qualitatively similar but always smaller than those from QRD derived models. For the optimal points on plots of the differences between predicted and true data, Figs. 3 and 5, the SVD models give results that are as much as 60% better than the corresponding QRD models. The SVD optimal models always have fewer vectors than the corresponding optimal QRD models.

The effects of increasing the noise level can be seen by comparing Fig. 6 with a measurement noise level of 0.3 to Fig. 5 with a measurement noise level of 0.1. Fig. 6 shows the results for model sizes of 4 to 32 model vectors. The optimal models have fewer vectors than for the corresponding simulations with a noise level of 0.1. The effects of decreasing the noise level to 0.03, shown in Fig. 7, are to increase the number of model vectors in the optimal models. This effect

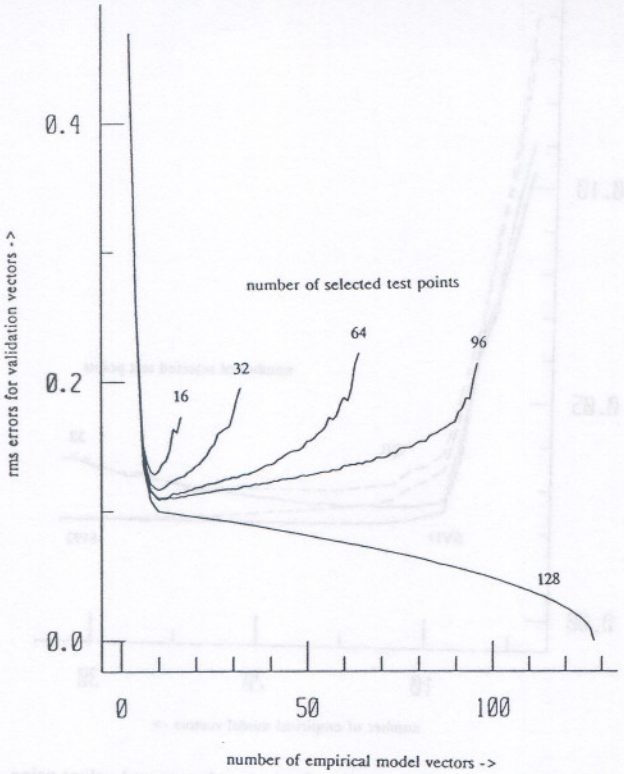

Fig. 4. Rms difference between predicted values and measured values using SVD derived models and an rms measurement noise of 0.1.

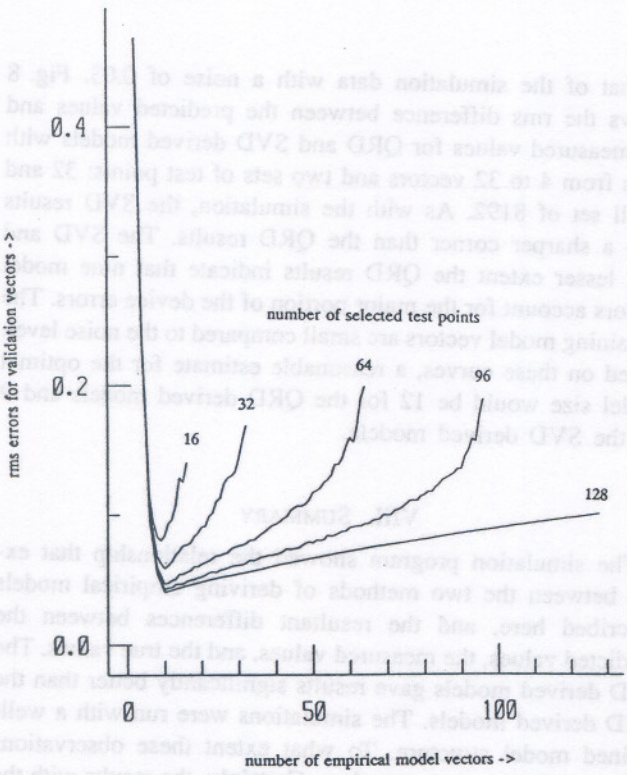

Fig. 5. Rms difference between predicted values and true values using SVD derived models and an rms measurement of 0.1.

of noise level and method of derivation are summarized in Table II. This table gives the optimum number of model vectors for the reduced set of 32 test points for both model derivation methods and three noise levels and compares them

TABLE II
OPTIMUM NUMBER OF MODEL VECTORS FOR 32 TEST POINTS

	Noise 0.03		Noise 0.1		Noise 0.3	
	QRD	SVD	QRD	SVD	QRD	SVD
Pred-Meas	13–16	10	11–17	10	10	6–8
Pred-True	16	10	11-15	10	10	7

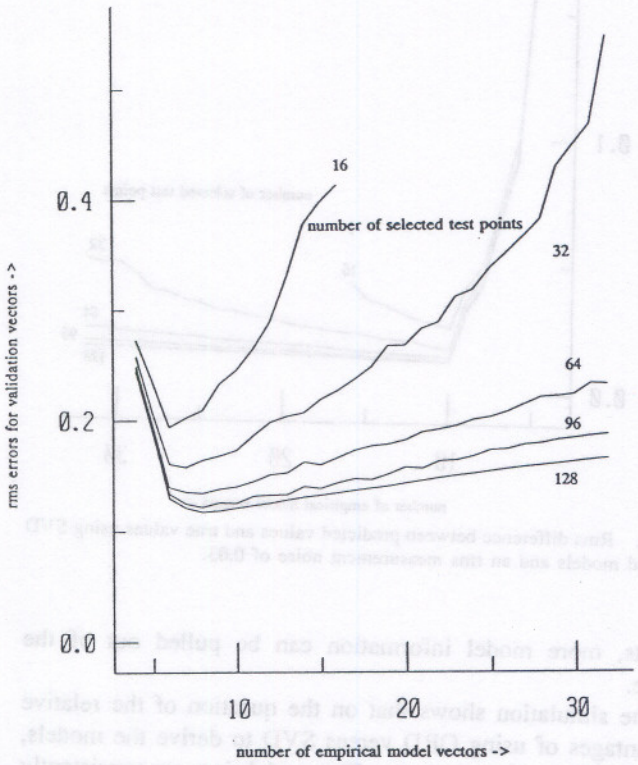

Fig. 6. Rms difference between predicted values and true values using SVD derived models and an rms measurement noise of 0.3.

with the minimum difference locations for the difference between the predicted values and the measured values. A range of model sizes is given if the other locations have differences within 1% of the minimum difference. The qualitative results for 16, 64, and 96 reduced test points are very similar to those shown in Table II.

VI. SUMMARY OF SIMULATION RESULTS

The following summarizes the observations that relate to the four questions which were presented above. On the question of how to determine the optimum model size, the simulations show that, in general, the optimum size for the difference between the predicted values and the true values occurs at the same location as the optimal size for the difference between the predicted values and the measured values. For real data the true values cannot be determined; thus, one way to estimate the optimum model size is to determine the location of the minimum difference between the predicted values and the measured values using the validation set. Regarding the effects of the number of test points, as the number of test points increases the size of the optimum model increases and the differences decrease. With a larger number of test

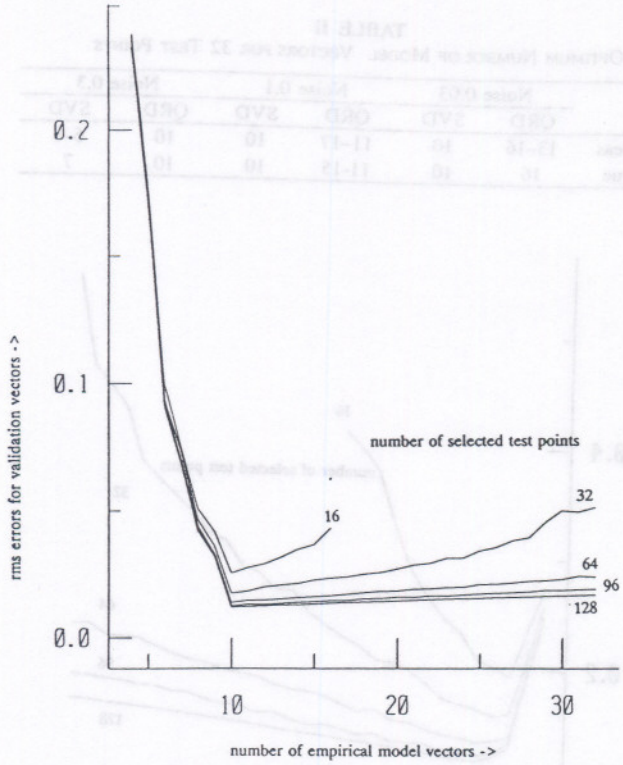

Fig. 7. Rms difference between predicted values and true values using SVD derived models and an rms measurement noise of 0.03.

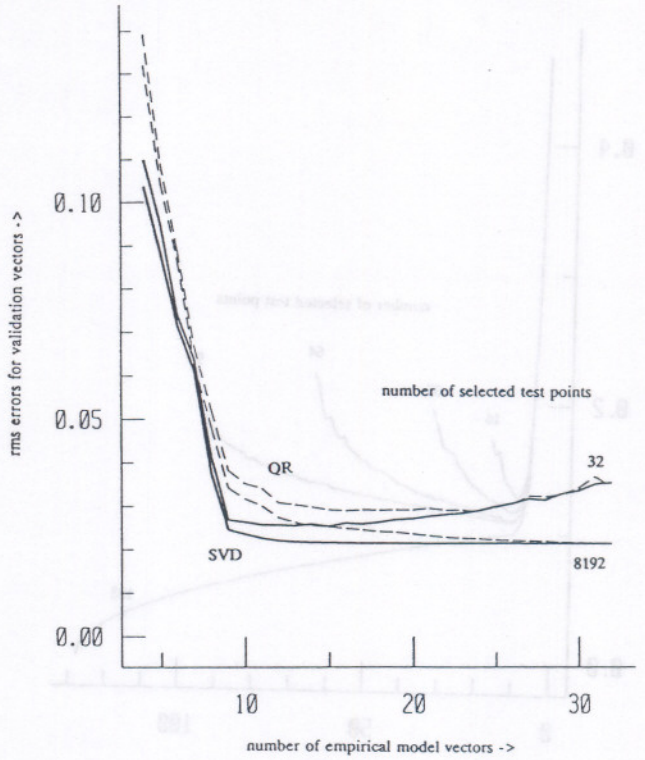

Fig. 8. Rms difference between predicted values and measured values using QRD derived models (dashed lines) and SVD derived models (solid lines) for data from 13-bit A/D converters with an rms measurement noise of about 0.021.

points, more model information can be pulled out of the noise.

The simulation shows that on the question of the relative advantages of using QRD versus SVD to derive the models, SVD does significantly better. The model sizes are consistently smaller. This means that less random error is mapped on to the model. The differences between the predicted and the true values are from 40 to 60% smaller for SVD models versus QRD models. Even though the SVD algorithm takes about three times longer than the QRD algorithm, this analysis is done off-line from the testing and so does not affect testing time. In fact, the smaller differences obtained using SVD mean that when using SVD models fewer test points are necessary to obtain the same accuracy than when using QRD models, thus speeding up the test times. Finally, the simulations show that as the noise level increases, the number of vectors in the optimum model decreases and the differences between the predicted values and the true values increase. With more noise present the QRD and SVD methods can pull less of the model information out of the empirical data.

VII. Real Data Example

This same analysis was applied to data taken on 13-bit A/D converters. Data from 78 devices were separated into 50 error vectors for the model vector set and 28 error vectors for the validation set. This data was run through the simulation program and modeled with both QRD and SVD. The measurement errors have an rms amplitude of about 0.5 least significant bits (lsb's), and the measurement noise is about 0.021 lsb's. This model-to-noise ratio is about the same as that of the simulation data with a noise of 0.03. Fig. 8 shows the rms difference between the predicted values and the measured values for QRD and SVD derived models with sizes from 4 to 32 vectors and two sets of test points: 32 and a full set of 8192. As with the simulation, the SVD results give a sharper corner than the QRD results. The SVD and to a lesser extent the QRD results indicate that nine model vectors account for the major portion of the device errors. The remaining model vectors are small compared to the noise level. Based on these curves, a reasonable estimate for the optimal model size would be 12 for the QRD derived models and 9 for the SVD derived models.

VIII. Summary

The simulation program showed the relationship that exists between the two methods of deriving empirical models described here, and the resultant differences between the predicted values, the measured values, and the true values. The SVD derived models gave results significantly better than the QRD derived models. The simulations were run with a well-defined model structure. To what extent these observations apply to other cases is not clear. Certainly, the results with the real data indicate the same kind of behavior as was observed for the simulations.

The size of the optimal model varied with the number of selected test points, the method used to derive the models, and the ratio of the model size to the noise level. In all

cases, however, the model size that gave minimum difference between the predicted values and the measured values was very near the optimal model. Thus, this appears to be a good method for locating the optimal model size. Another observed feature was that the difference between the predicted and the measured was always larger than the difference between the predicted and the true. A theoretical proof of this observation is being developed.

REFERENCES

[1] T. M. Souders and G. N. Stenbakken, "Modeling and test point selection for data converter testing," in *1985 IEEE Int. Test Conf. Proc.* New York: IEEE Press, Nov. 1985, pp. 813–817.

[2] G. N. Stenbakken and T. M. Souders, "Test point selection and testability measures via QR factorization of linear models," *IEEE Trans. Instrum. Meas.*, vol. IM-36, pp. 406–410, June 1987.

[3] T. M. Souders and G. N. Stenbakken, "A comprehensive approach for modeling and testing analog and mixed-signal devices," in *Proc. 1990 Int. Test Conf.*, IEEE Computer Society Press, Sept. 1990, pp. 169–176.

[4] G. N. Stenbakken and T. M. Souders, "Linear error modeling of analog and mixed-signal devices," in *Proc. 1991 Int. Test Conf.*, IEEE Computer Society Press, Sept. 1991, pp. 573–581.

[5] G. W. Stewart, *Introduction to Matrix Computations.* New York, NY: Academic, 1974.

[6] J. J. Dongarra, J. R. Bunch, C. B. Moler, and G. W. Stewart, *LINPACK User's Guide*, SIAM, Philadelphia, PA, 1979.

Gerard N. Stenbakken (M'71) received the Bachelor of Physics degree from the University of Minnesota in 1964, and the M.S. degree in physics from the University of Maryland in 1969, and the M.S. degree in electrical engineering from the University of Maryland in 1986.

From 1963 to 1969 he worked with Vitro Laboratories. In 1969 he joined the National Bureau of Standards (now the National Institute of Standards and Technology). There he has worked on measurement methods for semiconductor devices, instrument reliability analyses, the development of power measurement instrumentation, and methodologies for reducing the cost of testing complex electronic devices.

T. Michael Souders (M'75–SM'90) received the B.S. degree in physics from the Johns Hopkins University in 1967.

Since that time, he has been a career employee with the National Institute of Standards and Technology, Gaithersburg, MD. His primary interests in recent years include standards and test methods for data acquisition and conversion devices, and efficient testing strategies for complex systems. He has published over 30 articles on work in these fields.

Mr. Souders serves as Chairman of the IEEE Waveform Measurements and Analysis Committee.