A Rigorous Exposition of the LEMMA Method for Analog and Mixed-Signal Testing

Adrian Wrixon and Michael Peter Kennedy, *Fellow, IEEE*

***Abstract*—The linear error-mechanism modeling technique is an effective tool for testing analog and mixed-signal devices which minimizes the number of measurements required to characterize the static transfer function of a circuit by determining a small number of parameters of a linear error model and then predicting the entire response error.**

This work focuses on optimizing the linear error-mechanism model algorithm (LEMMA), introducing novel refinements which are shown to improve its performance significantly. We outline the implementation of the algorithm in a tutorial manner, paying due consideration to the underlying theory where required.

***Index Terms*—Analog testing, error-mechanism, linear modeling, mixed-signal testing, test development.**

I. Introduction

THE objective of testing is to verify that a circuit meets the specifications for which it has been designed [1]. The goal of the test engineer is to ensure that this is achieved with maximum reliability and at minimum cost. This task ultimately leads to a compromise between the required accuracy and the cost of testing, where the latter can generally be specified in terms of on-line test time and the number of test stations.

The most obvious test procedure is to measure the entire response of the circuit under test (CUT). This method is called a *full-factorial* experiment and, while it is undoubtedly reliable, it is not feasible for devices with a large number of significant inputs.[1] Consider, for example, the problem of determining the integral nonlinearity[2] (INL) of a 12-bit digital-to-analog converter (DAC). An all-codes static test of this circuit involves measuring the output voltage for 2^{12} (4096) different input codes and finding the worst-case deviation from the ideal response [2].

Full-factorial testing is extremely inefficient when the number of error sources which contribute to a measured response is much smaller than the number of significant inputs. In this case, a more efficient solution may be obtained by decomposing the response into a weighted linear sum of responses due to independent error mechanisms and determining the weights associated with each error using a reduced set of test-points [3]. This is called the Linear Error-Mechanism Model Algorithm (LEMMA).

The LEMMA method is based on the assumption that the response of a circuit is determined primarily by a small number of variables and that it depends linearly on the deviation of these variables from their nominal values. Instead of *measuring* the entire response of the circuit, a limited number of measurements is made at carefully selected test-points and these are used to determine the coefficients of the underlying linear error-mechanism model. This model is used to *predict* the entire response of the CUT, and a parametric test may then be carried out on the basis of the predicted response [4], [5].

The performance of the LEMMA method is a tradeoff between the completeness of the model and the confidence of the predicted response. There are many degrees of freedom in implementing the method: how many error sources to model, which ones, and how to model them; how many test-points to select and how to choose them; and how to minimize the computation of the predicted response. In this paper, we develop near-optimal solutions to each of these questions. We divide the problem into two parts: model development and production testing. We first consider model development.

The LEMMA model may be thought of as representing the sensitivities of the measured response to the significant error-producing mechanisms. The model may thus be constructed directly from the *a priori* error signatures associated with each error source. If a good SPICE [6] model of the CUT exists, these sensitivities may be calculated directly. Other suitable *a priori* basis vectors for mixed-signal circuit include Rademacher and Walsh functions [7]. An important property of a priori basis vectors is that they are *noise-free*.

Unmodeled process variations often give rise to significant errors whose error signatures are difficult, if not impossible, to determine by simulation. In this case, the LEMMA model must be augmented with, or built entirely from, *empirical* data which are inherently noisy. In addition, each measured response is a *combination* of error signatures associated with the underlying error sources, so it is not possible to isolate the contributions of individual errors.

In an analog or mixed-signal circuit, many factors may contribute to the response error. It is not always clear how many of these are "significant." Souders and Stenbakken [4], [5] and later Huertas [1], proposed heuristic formulae for estimating the *number* of significant error sources in a LEMMA model. In Section IV of this paper, we show how

Manuscript received December 20, 1996. The first author was supported by Analog Devices, Ireland. The second author was supported by Enterprise Ireland under the Applied and Strategic Research Grants Schemes Contracts HE/1997/190 and ST/1998/110.

The authors are with the Department of Electronic and Electrical Engineering, University College Dublin, Dublin 4, Ireland (e-mail: peter.kennedy@ucd.ie).

Publisher Item Identifier S 0018-9456(99)06682-6.

[1] A *significant input* is one for which the CUT output has a specified error-limit.

[2] The integral nonlinearity of a DAC is the largest output deviation from ideality over the entire input range.

Singular Value Decomposition (SVD) [8], [9] may be used to determine, using measured data, a lower bound on the number of significant error sources.

In Section V, we demonstrate the equivalence of *a priori* and empirical models, and stress the importance of minimizing the noise-level in the model. We exploit total least-squares (TLS) to develop a rigorous noise-reduction procedure for selecting basis functions.

Section IV addresses the choice of the number of testpoints.

Souders and Stenbakken showed how QR decomposition (QRD) could be used to select an almost D-optimal set of test-points for the LEMMA method [4]; we review their procedure in Section VII.

We present a model development algorithm in detail in Section VIII and provide lower bounds in Sections IX and X for the on-line measurement and computation time of the LEMMA method in production testing.

All of the results presented in this work are based on all-codes measurements of 200 Analog Devices AD7528 8-bit current-mode DAC's. Half of the measured response error vectors have been used for model development and the other half for simulated production testing. To quantify the effects of measurement noise in Section IV, the production data has been artificially "cleaned" and a controlled amount of noise added, using the procedure described in the Appendix.

II. Theoretical Foundation of the LEMMA Method: Why It Works

The LEMMA method is based on the assumption that the components within the CUT lie sufficiently close to their nominal values that the response error may be approximated by a *first-order* Taylor series expansion, in terms of the underlying error-producing mechanisms, about the ideal response [5]. In this case, a linear model may be used to characterize the circuit. We form the linear model by assuming that the overall response error is the superposition of the response errors of each mechanism acting alone, with all other components at their nominal values.

For example, the transfer characteristic of a resistive ladder DAC is determined by the values of the resistors in the ladder. Deviation of a resistor from its nominal value produces a characteristic *error signature*. For sufficiently small errors in the resistors, the deviation of the measured response from its ideal value (the so-called *response error*) will be a linear combination of these error signatures.

In matrix form, the response error $\mathbf{e}(m\times 1)$ of the CUT may be written as a linear combination of n linearly independent column vectors $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$:

$$\mathbf{e} = \mathbf{E}\mathbf{x} \tag{1}$$

where the columns of $\mathbf{E}(m \times n)$ are the basis vectors $\mathbf{e}_i$ and $\mathbf{x}$ is some n-vector of real-valued weights.[3] Instead of *measuring* the m components of the total response error, one instead determines the vector of n weights $\mathbf{x}$ from p $(\geq n)$ measurements and *predicts* $\mathbf{e}$.

In the case of a data converter, the number of error-producing mechanisms typically scales linearly with the number of bits, while the number of significant inputs grows exponentially. Thus, $n \ll m$, and significant savings in test effort are possible.

III. LEMMA Method: Implementation Issues

The first issue to be addressed is the choice of n, the dimension of the model. This problem has been studied widely in the system identification literature (see, e.g., [10]) and in the statistics literature (see, for example, the discussion of principal component analysis in [11], [12]). We use SVD [9] to determine the rank of a matrix comprising sensitivity vectors and measured error signatures; we then follow the parsimony principle [10] in constructing $\mathbf{E}$.

The system of equations (1) is *overdetermined* in that we have m equations in n unknowns, where $m > n$. Furthermore, the presence of measurement noise in both $\mathbf{E}$ and $\mathbf{e}$ renders the equations *inconsistent*. To counter this, we measure p components of $\mathbf{e}$, where $n \leq p \ll m$, and estimate the vector of weights $\mathbf{x}$ by means of a least-squares fit $\mathbf{x_{LS}}$.

The problem of determining which points in the response error to measure is called *test-point selection*. If we let $t_1, t_2, \ldots, t_p$ denote the selected test-points, and let $\mathbf{e_R}(p\times 1)$ be the vector of associated response error measurements, then we obtain $\mathbf{x_{LS}}(n \times 1)$ by minimizing

$$\|\mathbf{E_R x_{LS}} - \mathbf{e_R}\|$$

where $\mathbf{E_R}(p \times n)$ is the reduced model formed from rows $t_1, t_2, \ldots, t_p$ of $\mathbf{E}$.[4]

Finally, we estimate the entire response error by writing

$$\mathbf{e} \approx \mathbf{E x_{LS}}.$$

An optimal selection of n, $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$, p and $\{t_1, t_2, \ldots, t_p\}$ yields an estimate which is "close" to the true response error $\mathbf{e}$ and, with appropriate guard-banding, may be used for verifying conformance to specifications. This is the principle of the LEMMA method.

In the following sections, we develop techniques for selecting the key parameters of the LEMMA method.

IV. Choice of n: How Many Basis Vectors?

When a good model for a circuit exists, the number of "significant" error-producing mechanisms may be determined analytically or by simulation. In practice, process variations and unforeseen error sources may produce a response error which deviates significantly from that predicted by an incomplete design model.

The LEMMA technique is robust in the sense that all significant errors may be accounted for in the linearized model, whether they can be interpreted as sensitivities with respect to circuit components or not. Indeed, unforeseen errors may be included in the model $\mathbf{E}$ simply by adding column vectors

[3] In the special case where an explicit model of the CUT exists, the basis vectors could be error signatures, in which case the components of $\mathbf{x}$ would be the errors in the corresponding error sources.

[4] A weighted least squares approach could also be used here.

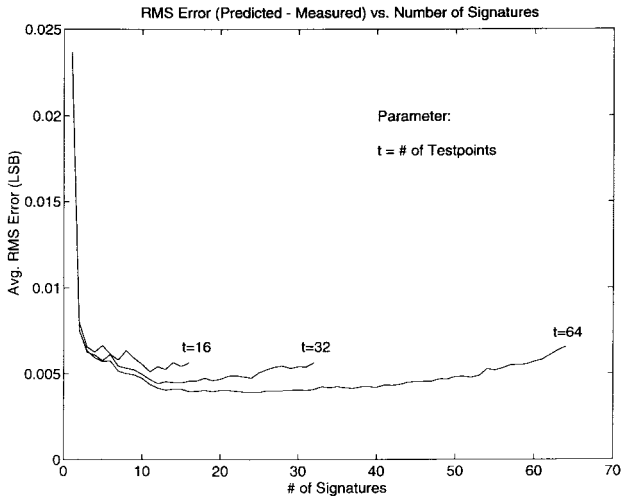

Fig. 1. RMS error between predicted values and measured values for the AD7528 with additive white Gaussian measurement noise having an rms value of 0.003 LSB, averaged over 100 data sets.

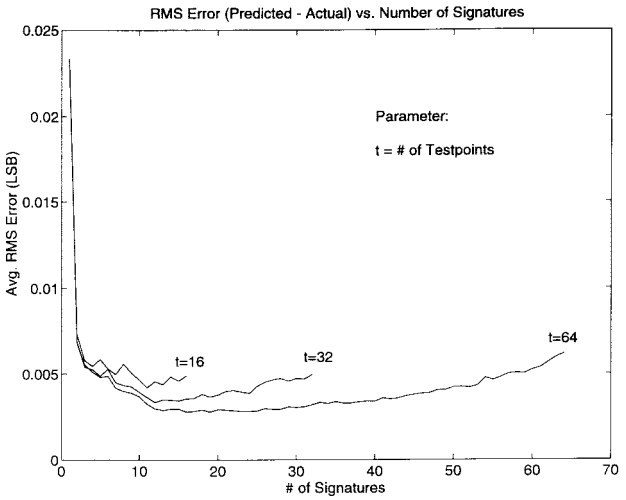

Fig. 2. RMS error (averaged over 100 data sets) between predicted values and true values for simulated data with additive measurement noise having an rms value of 0.003 LSB.

corresponding to measured response errors one at a time until the columns of $\mathbf{E}$ are no longer linearly independent.

Model development is complicated by the presence of measurement noise in the measured basis vectors. In addition to compromising the overall integrity of the model, such noise makes it difficult to determine whether or not the columns of $\mathbf{E}$ are linearly independent.

A. Noise-Free Case

Starting with a small number of sensitivity vectors or response errors as our model $\mathbf{E}$, we augment the model one basis vector at a time until the columns are deemed, using SVD [9], to be linearly dependent.

In practice this can be far from easy, and the penalty for selecting too many basis vectors is that the performance of the algorithm is compromised both in terms of on-line measurement time and accuracy (see Figs. 1 and 2).

Picture the ideal scenario with no measurement noise and exactly n error-producing mechanisms. Let $\mathbf{E_j}(m \times j)$ denote a model comprising j basis vectors. At each step we would calculate

$$\mathbf{E_j} = \mathbf{USV^T}$$

using SVD. For $j \leq n$ we would find that $\mathbf{S}$, a $j \times j$ matrix, would satisfy:

$$S_{i,i} > 0 \quad \forall i$$

For $j > n$ we would find

$$S_{i,i} > 0 \quad \forall i \leq n$$
$$S_{i,i} = 0 \quad \forall i > n$$

In the noise-free case, the optimum number of basis vectors n is simply the number of nonzero singular values of $\mathbf{E}$. When measurement noise is present, we no longer have this useful rule of thumb.

B. Real Case

Noise manifests itself by pushing each signature from the n subspace in which it should lie out into m space. Thus the diagonal elements of $\mathbf{S}$ never reach zero because the model explores every direction in $\mathbf{U}$ to some extent.

One approach to selecting n is to augment $\mathbf{E}$ with measured response errors until the last diagonal element of $\mathbf{S}$ dips below the "noise floor." Assuming measurement noise with a variance of σ^2, this amounts to selecting as n the largest number satisfying

$$S_{n,n} > \sqrt{m}\sigma.$$

Note that $\sqrt{m}\sigma$ is the rms value of the diagonal elements of $\mathbf{S}$ in the case where $\mathbf{E}$ contains only noise. In our experience, this provides a reliable guideline for estimating the *minimum* number of columns in the model [1].

C. Optimal Choice of n

In practice, more basis vectors than suggested by the above procedure may be permitted because, when forming $\mathbf{E_R}(p \times n)$, the test-points are specifically selected to minimize noise effects, implying that noise is more significant in the model as a whole than in our judiciously chosen subset. Moreover, Souders and Stenbakken [5] have correctly shown that the optimal number of basis vectors varies with the number of test-points. Hence any procedure to select an optimal number of basis vectors should take into account the number of test-points. Such a procedure does not exist as yet, and at present the only way to be entirely sure of the optimum number of basis vectors for a given number of test-points seems to lie in laboriously producing plots such as Figs. 3 and 4. These graphs show, using real data only, the ability of the LEMMA method to predict *measured* response errors, both in terms of rms prediction error and maximum prediction error.

Figs. 1 and 2 confirm the key observation by Souders and Stenbakken [5] that the optimal model size for the difference between the predicted values and the *true* values occurs at the

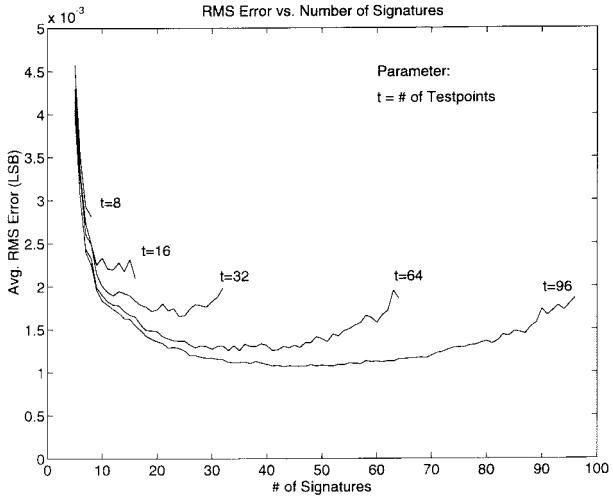

Fig. 3. RMS error (averaged over 100 data sets) between predicted values and real measured values versus number of signatures, with number of test-points as a parameter.

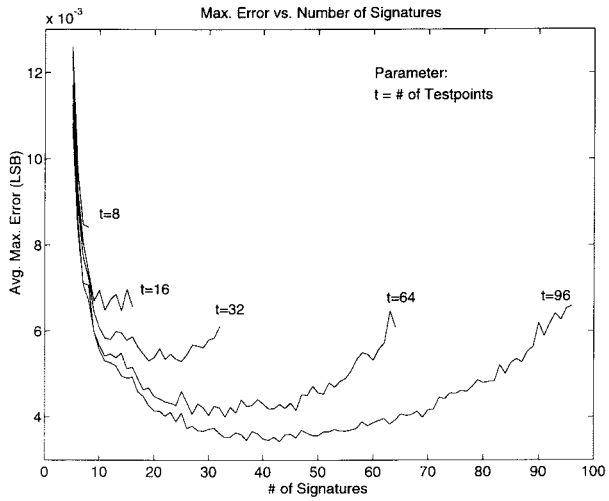

Fig. 4. Max. error (averaged over 100 data sets) between predicted values and real measured values versus number of signatures, with number of test-points as a parameter.

same location as the optimal size for the difference between the predicted values and *measured* (true plus noise) values. In a practical situation, the true values cannot be determined, but we can use this result to locate the optimal model size with confidence.

Our experimental observations lead us to conjecture that the optimal model size in terms of rms error is also optimal in terms of the maximum error (see Figs. 3 and 4).

V. Choice of $\mathbf{E}$: Which Basis Vectors?

In our discussion of the LEMMA method in Section III, we mentioned that the CUT model could be constructed from a combination of sensitivity vectors and empirical response errors. The former have the advantage of being free of measurement noise, although difficult to derive, but cannot account for unmodeled process-induced errors. The latter capture such errors but are necessarily corrupted by measurement noise.

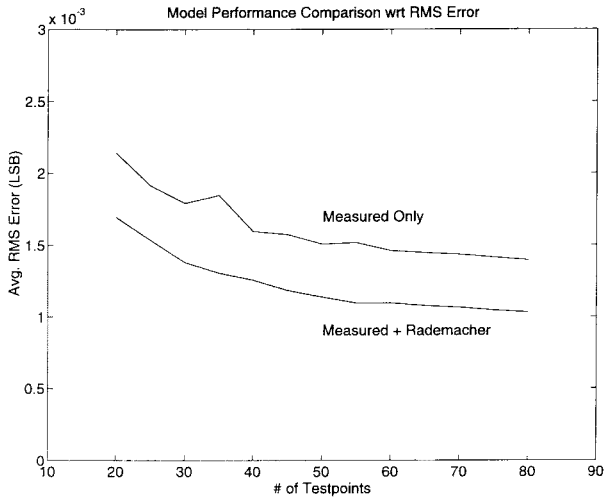

Fig. 5. Using a model of 20 basis vectors in both cases, this plot shows the performance (in terms of mean rms error) of models comprising measured vectors alone and a mixture of measured vectors and Rademacher functions.

There is no reason why a model must be entirely one or the other, however. A model with vectors of both types is equally valid, and may constitute an improvement if the overall noise in the model is reduced.

In this section we consider the choice of basis vectors for the matrix $\mathbf{E}$ and derive a procedure for selecting basis vectors from a combination of *a priori* and measured response errors.

A. Other Useful A Priori Model Vectors

In addition to simulated error signatures, a second type of *a priori* basis vector is available. Such vectors can be derived by considering particular errors which will appear due to the *functionality*, rather than the constituent parts, of a device.

An important example is the set of Rademacher functions [7] in the case of DAC's. Vectors formed from these functions describe errors that manifest themselves according to the state of switches within the DAC, and are based upon device considerations at the system level rather than the component level. The Rademacher functions can be generated by computer and are guaranteed to be noise-free, thus constituting a good choice as model vectors, when applicable.

Figs. 5 and 6 compare the performance of models with and without *a priori* Rademacher functions for the 8-bit AD7528 DAC. Since the Rademacher functions are noise-free, the model comprising Rademacher functions augmented with measured response errors clearly performs better (for the same size of reduced model) than that composed exclusively of measured response errors.

How can this observation be used to define a procedure for selecting an optimum set of n basis vectors from a choice of j measured error responses and k *a priori* vectors? We address this question in the following subsections.

B. Noise Reduction Using Total Least-Squares

A valuable noise-reduction procedure arises from examining the linear algebra method of total least-squares (TLS) [13]. This is a technique for obtaining a least squares solution of

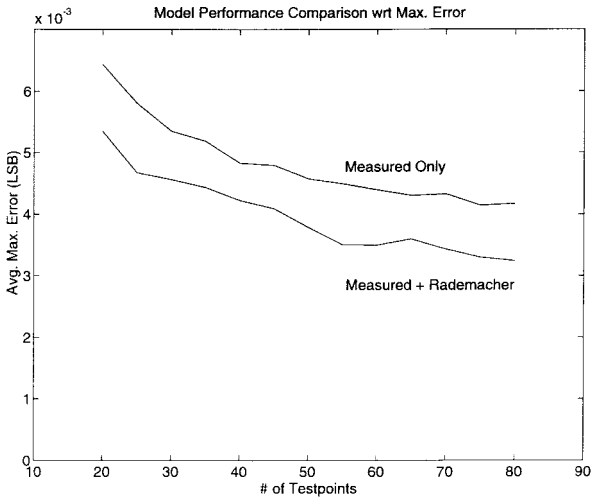

Fig. 6. Using a model of 20 basis vectors in both cases, this plot shows the performance (in terms of mean maximum error) of models comprising measured vectors only and a mixture of measured vectors and Rademacher functions.

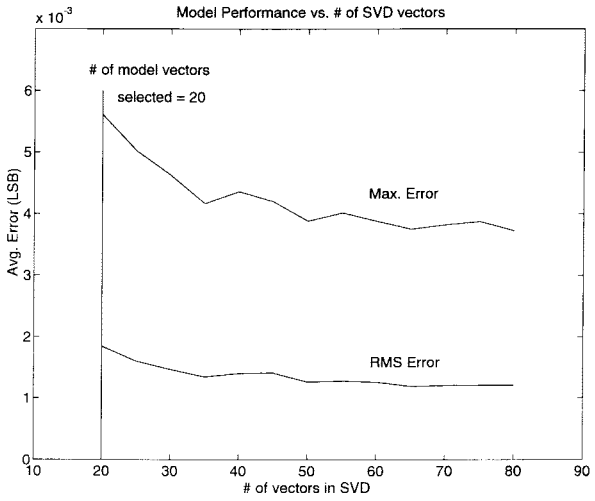

Fig. 7. The x-axis shows the number of vectors used in the SVD, after which only the most significant twenty are chosen. Performance in terms of prediction error (averaged over 100 data sets) lies along the y-axis.

the overdetermined problem $\mathbf{Ax} = \mathbf{b}$ when both $\mathbf{A}(m \times n)$ and $\mathbf{b}(m \times 1)$ are imprecisely known. It deals specifically with the case

$$\mathbf{Ax} = \mathbf{b} \quad \text{or} \quad (\mathbf{A_0} + \mathbf{\Delta A})\mathbf{x} = \mathbf{b_0} + \mathbf{\Delta b}$$

under the assumption that all elements of $[\mathbf{\Delta A}; \mathbf{\Delta b}]$ are independently and identically distributed with zero mean.

A useful application of TLS in the LEMMA method arises during model development. Suppose it is found, either theoretically or by experimentation, that a set of n signatures constitutes the optimum model, $\mathbf{E}$. Consider that another linearity plot $\mathbf{e}$ is available, so that the augmented system $[\mathbf{E}; \mathbf{e}]$ constitutes a slightly worse model than $\mathbf{E}$ alone. We know that $[\mathbf{E}; \mathbf{e}]$ should ideally consist of $n + 1$ vectors of which only n are linearly independent. However the presence of noise augments the rank of $[\mathbf{E}; \mathbf{e}]$ to $n + 1$.

Now TLS tells us that if we perform SVD as follows:

$$[\mathbf{E}; \mathbf{e}] = \mathbf{USY^T}$$

then the first n columns of $\mathbf{U}(m \times (n + 1))$ constitute a better basis for the true n dimensional subspace than $\mathbf{E}$ alone. Therefore, a model formed from these columns is an improvement on $\mathbf{E}$.

Taking this as far as possible, suppose we are forming an n-column model but there are $d > n$ candidate basis vectors, $\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_d}$, from which we must choose n. Our best option is to decompose

$$[\mathbf{e_1} \quad \mathbf{e_2} \quad \cdots \quad \mathbf{e_d}] = \mathbf{USY^T}$$

and to take the first n vectors from $\mathbf{U}$ as our model $\mathbf{E}$. These vectors define the unique subspace that minimizes the perpendicular distances to all of the candidate vectors, thereby averaging out the effects of noise.

Fig. 7 shows model performance versus the number of vectors used to build the model for the case $n = 20$. Here d ranges from 20 to 80.

C. Rigorous Procedure for Selecting Basis Vectors

The question now arises about how to combine the above facts into a robust basis vector selection procedure. Work by Souders and Stenbakken [5] has shown the importance of an orthonormal model $\mathbf{E}$, and in particular one derived by SVD instead of QRD.[5] We want to avoid placing our noise-free *a priori* vectors in the large SVD at the start of the basis selection procedure; otherwise they will be "drowned out" by noise in the measured basis vectors. We suggest the following solution.

When deriving an n-column model from a selection of $j > n$ measured signatures, $\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_j}$, and $k < n$ *a priori* vectors, $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_k}$, we decompose

$$[\mathbf{e_1} \quad \mathbf{e_2} \quad \cdots \quad \mathbf{e_j}] = \mathbf{USY^T}$$

and take the first n vectors from $\mathbf{U}$ to form $\mathbf{X}$. Then we decompose

$$[\mathbf{v_1} \quad \mathbf{v_2} \quad \cdots \quad \mathbf{v_k}; \mathbf{X}] = \mathbf{USY^T}$$

and take the first n vectors from $\mathbf{U}$ to form our model $\mathbf{E}$. The resulting model contains a low-level noise contribution from n vectors instead of a high-level noise contribution from j vectors.

Once the model $\mathbf{E}$ has been chosen, one must decide how many measurements to make and how to select the test-points. We address these questions in the following two sections.

VI. Choice of p: How Many Test-Points?

By contrast with the number of basis vectors, increasing the number of test-points for a given number of basis vectors *always* improves the accuracy of the prediction. In general, the "optimum" number of test-points is arrived at as a compromise between the permitted measurement time and the maximum

[5] Having orthonormal columns for $\mathbf{E}$ minimizes the probability that any submatrix matrix $\mathbf{E_R}$ will be singular or almost singular. In particular, an SVD-derived model best describes the directions in which errors manifest themselves and has been shown to produce better results than QR-derived models [5].

tolerable prediction error. This is a specification which will vary from application to application.

VII. Choice of $\{t_1, \ldots, t_p\}$: Which Test-Points?

The next question is how to select the test-points, i.e. how to choose the best subset $\mathbf{E_R}$ of $\mathbf{E}$. Why not, for example, simply choose the first few rows, or else use a random selection?

Suppose that $\mathbf{e}$ is subject to measurement error so that our equation is actually

$$\mathbf{Ex} = \mathbf{e} + \epsilon$$

where $\mathbf{e}$ is the true response error and the components $\epsilon_1, \epsilon_2 \ldots \epsilon_m$ are normally distributed with zero mean and variance σ^2. The vector of variances of the predictions of the response error is given by

$$\sigma_{\mathbf{T}}^{\mathbf{2}} = \operatorname{diag}\{\mathbf{E}(\mathbf{E_R^T E_R})^{-1}\mathbf{E}^T\sigma^2\}. \tag{2}$$

Souders and Stenbakken [4] favor selecting the submatrix $\mathbf{E_R}$ such that the determinant of $\mathbf{E_R^T E_R}$ is maximized; this has the effect of minimizing the maximum element of $\sigma_{\mathbf{T}}^{\mathbf{2}}$. This form of optimization is called *D-optimality*.

How does one maximize $\det(\mathbf{E_R^T E_R})$? Testing the determinants of all $\binom{m}{p}$ possible subsets $\mathbf{E_R}$ of $\mathbf{E}$ is clearly too large a task to undertake, but *QR factorization* of $\mathbf{E}$ provides an efficient means of choosing a nearly optimal set of n test-points. A detailed treatment of this procedure may be found in [4].

If further test-points are desired to improve accuracy, it is a good idea to augment $\mathbf{E_R}$ with the row of $\mathbf{E}$ corresponding to the maximum prediction variance [14]. This can be repeated until the maximum prediction variance is sufficiently low.

Whether D-optimality is the best form of optimality to apply for the LEMMA method is questionable. From (2) we can form a vector α of prediction errors based upon the acceptable level of confidence. If we let $\mathbf{e_P}$ denote the predicted response error and let $\mathbf{s}$ be a vector of the specified error-limits at each significant input, then a device is deemed good if

$$\left|e_{P_i}\right| + \alpha_i < |s_i| \quad \forall i$$

If we assume that a device is equally likely to fail the above test for any i, then it is preferable to minimize the sum of the prediction variances, thereby achieving a "closest fit" to the actual response, rather than minimizing the maximum prediction variance. In this manner fewer good devices are incorrectly diagnosed as faulty. A set of test-points which minimizes the sum of the prediction variances is *A-optimal*, and the A-optimality criterion is

$$\text{minimize: } \operatorname{trace}\mathbf{E_R^T E_R}$$

The theory and performance of A-optimal test-points are examined in [15].

VIII. Algorithm I: Model Development

In this section, we summarize the model development algorithm for the LEMMA method [16]. These steps are performed off-line so we ignore the associated measurement and computation costs.

Algorithm I:

Given: Measured response error vectors $\mathbf{e_1}, \ldots, \mathbf{e_j}$ and *a priori* vectors $\mathbf{v_1}, \ldots, \mathbf{v_k}$, where all vectors are $m \times 1$. The rms measurement noise level is assumed to be σ.

Step 1) Compute the SVD, $[\mathbf{e_1} \ \cdots \ \mathbf{e_j} \ \mathbf{v_1} \ \cdots \ \mathbf{v_k}] = \mathbf{USV^T}$.

Step 2) Determine the largest n such that $S_{n,n} > \sqrt{m}\sigma$.

Step 3) Compute the SVD $[\mathbf{e_1} \ \mathbf{e_2} \ \cdots \ \mathbf{e_j}] = \mathbf{USV^T}$ and form $\mathbf{X}$ from the first n columns of $\mathbf{U}$.

Step 4) Compute the SVD $[\mathbf{v_1} \ \mathbf{v_2} \ \cdots \ \mathbf{v_k}; \mathbf{X}] = \mathbf{USY^T}$ and form the model $\mathbf{E}$ from the first n vectors of $\mathbf{U}$.

Step 5) Decompose $\mathbf{E}$ as $\mathbf{E^T P} = \mathbf{QR}$ using QRD with pivoting. Form $\mathbf{E_R}$ from the first p rows of $\mathbf{P^T E}$.

Step 6) Using (2), determine whether the maximum prediction variance is within the desired level. If not, augment $\mathbf{E_R}$ with the row corresponding to the maximum prediction variance. Repeat until the maximum prediction variance is sufficiently low.

Step 7) If the number of test-points is greater than the test time allows, try varying the number of model vectors and repeat from Step 3. If this does not work the specifications cannot be met using the LEMMA method.

Step 8) Calculate the pseudo-inverse of $\mathbf{E_R}$ using the equation $\mathbf{E_R^+} = (\mathbf{E_R^T E_R})^{-1}\mathbf{E_R^T}$.

IX. Production Testing: What Does It Cost?

Since model development is performed off-line, we have ignored the costs associated with measurement and computation. In the case of production testing, however, the time required for on-line measurement and computation will determine the throughput. Depending on the speed of the hardware, a further tradeoff may be required between measurement and computation time.

X. Algorithm II: Production Testing

In this section, we summarize the on-line step of the LEMMA method. We present the relevant steps with the associated costs in time. We assume that each measurement takes τ_{meas} s and that each floating-point operation (multiplication, addition, or comparison) takes τ_{flop} s.

Algorithm II:

Given: Model $\mathbf{E}$, test-points $t_1, \ldots, t_p$, associated submatrix pseudo-inverse $\mathbf{E_R^+}$.

Step 1) Measure the response error at the p test-points and construct the reduced error vector $\mathbf{e_R} = (e_{t_1}, e_{t_2}, \ldots, e_{t_p})^{\mathbf{T}}$. This takes $p \times \tau_{\text{meas}}$ s.

Step 2) Calculate the coefficient vector $\mathbf{x}_{\text{LS}} = \mathbf{E_R^+ e_R}$. This takes $2 \times n \times p \times \tau_{\text{flop}}$ s.

Step 3) Calculate $\mathbf{e} = \mathbf{E}\mathbf{x}_{\mathbf{LS}}$. This takes $2 \times m \times n \times \tau_{\text{flop}}$ s.

Step 4) Check that every component of the predicted response error is within the specified error-limits. This could cost up to $2 \times m \times \tau_{\text{flop}}$ s.

A. Speed-Up of Algorithm II When Only the Largest Element of $\mathbf{e}$ *is Required*

Note that the computational cost of Algorithm II is dominated by Step 3, in which the response error is predicted by evaluating the product $\mathbf{E}\mathbf{x}_{\mathbf{LS}}$. In many cases, such as testing the INL of a DAC, the objective is to determine if the *largest* component of $\mathbf{E}\mathbf{x}_{\mathbf{LS}}$ is less (in magnitude) than some upper bound e_{MAX}.

To accelerate the computation, we exploit the Cauchy-Schwarz inequality [17] which states that

$$\mathbf{E_i} \cdot \mathbf{x_{LS}} \leq ||\mathbf{E_i}||||\mathbf{x_{LS}}|| \tag{3}$$

where $\mathbf{E_i}$ is the ith row of $\mathbf{E}$ and $||\cdot||$ denotes the Euclidean norm. This inequality permits us to eliminate from our computation all rows of $\mathbf{E}$ which satisfy $||\mathbf{E_i}|| < e_{\text{MAX}}/||\mathbf{x_{LS}}||$. This can be achieved very quickly if we presort the rows of $\mathbf{E}$ with respect to the norm during the model development phase [15].

XI. Concluding Remarks

In this work, we have developed a tutorial exposition of the LEMMA method for analog and mixed-signal testing. The technique provides an efficient way of predicting the entire response error of a circuit based on a limited number of measurements at judiciously chosen test-points. The method is particularly suited to problems where the number of significant inputs is very much greater than the number of error-producing mechanisms.

There are many degrees of freedom in implementing the method: how many error sources to model, which ones, and how to model them; how many test-points to select and how to choose them; how to minimize the computation of the predicted response. We have shown how to resolve these questions.

We have discussed the theoretical underpinnings of the LEMMA method, and provided a technique for determining the *number* of significant error sources in measured responses using SVD. By invoking TLS, we have justified a novel algorithm for constructing a low-noise model from a combination of *a priori* and measured basis vectors and we have recalled how to select the test-points by means of QR factorization.

We have provided detailed algorithms for both the model development and production testing phases of the LEMMA method. In addition, we have developed estimates of the on-line measurement and computation time and outlined a speed-up for the production testing phase when only the worst-case response is required.

Although we have resolved in this paper most of the issues involved in implementing the LEMMA method, further theoretical work is required to establish with greater rigor the relationship between the number of test-points, the choice of n, and the resulting prediction error. This is the subject of ongoing research [18], [19].

Appendix
Data Manipulation for Quantifying the Effects of Noise

All of the results presented in this work are based on all-codes measurements of 200 Analog Devices AD7528 8-bit current-mode DAC's. Half of the measured response error vectors have been used for model development and the other half for simulated production testing. To quantify the effects of measurement noise in Section IV, the production data was artificially "cleaned" and a controlled amount of noise added, as described below.

Since any measured data sets are inherently noisy to begin with, we designed the following procedure to *emulate* a system of $n = 30$ independent error mechanisms, with all measurements subjected to additive Gaussian noise with an rms value of 0.003 LSB.

1) A set of 200 measured response errors was split into two groups—set A with thirty vectors and set B with the remainder.
2) The vectors from set A were placed into a matrix $\mathbf{A}$ which was subsequently assumed to be entirely noise-free; it can be viewed as a perfect sensitivity matrix for the devices.
3) To construct a "clean" model of a desired size j, $\mathbf{E}$ was formed using j vectors from set B, which were then projected onto the subspace defined by $\mathbf{A}$. This was accomplished as follows:
$$\mathbf{E} \rightarrow \mathbf{A}(\mathbf{A}^{+}\mathbf{E})$$
where $\mathbf{A}^{+} = (\mathbf{A^T A})^{-1}\mathbf{A^T}$ denotes the pseudo-inverse of $\mathbf{A}$ [9].
4) Next, the desired level of Gaussian noise was added to the model and the required steps of the LEMMA method performed.
5) For each response error $\mathbf{e}$ tested, we calculated the "true" error $\mathbf{e_T} = \mathbf{A}(\mathbf{A}^{+}\mathbf{e})$ and the "measured" error $\mathbf{e_M} = \mathbf{e_T} + \epsilon$, where ϵ is Gaussian noise with an rms value of 0.003 LSB.
6) The LEMMA method was performed using test-points from $\mathbf{e_M}$. Letting $\mathbf{e_P}$ denote the response error as predicted by LEMMA, we gauge the performance by comparing: (a) the rms and maximum difference between the predicted and *measured* response errors $(\mathbf{e_P} - \mathbf{e_M})$, and (b) the rms and maximum difference between the predicted and *actual* response errors $(\mathbf{e_P} - \mathbf{e_T})$. These results are summarized graphically in Figs. 1 and 2.

Acknowledgment

The authors would like to thank T. O'Dwyer, New Product Development Manager, DAC Products, Analog Devices, Ireland (ADI), for bringing this problem to their attention and for his support and guidance throughout the project. Test data for the AD7528 8-bit current-mode DAC which has been used throughout this paper was also provided by ADI. The authors are also grateful to the reviewers for directing them to [10] and [11].

REFERENCES

[1] J. L. Huertas, "Test and design for testability of analog and mixed-signal integrated circuits: Theoretical basis and pragmatical approaches," in *Circuit Theory and Design 1993: Selected Topics in Circuits and Systems*, H. Dedieu, Ed. Amsterdam, The Netherlands: Elsevier, 1993, pp. 77–156.

[2] *Analog-Digital Conversion Handbook*, 3rd ed., D. H. Sheingold, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1986.

[3] T. M. Souders and G. N. Stenbakken, "Cutting the high cost of testing," *IEEE Spectrum*, pp. 48–51, Mar. 1991.

[4] 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.

[5] G. N. Stenbakken and T. M. Souders, "Developing linear models for analog devices," *IEEE Trans. Instrum. Meas.*, vol. 43, pp. 157–163, Apr. 1994.

[6] A. Vladimirescu, *The SPICE Book*. New York: Wiley, 1994.

[7] H. F. Harmuth, *The Transmission of Information by Orthogonal Functions*. Berlin, Germany: Springer-Verlag, 1972.

[8] G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," in *Linear Algebra*, Wilkinson and Reinch, Eds. Berlin, Germany: Springer-Verlag, 1971, pp. 134–151.

[9] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

[10] T. Söderström and P. Stoica, *System Identification*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[11] M. Kendall, A. Stuart, and J. K. Ord, *The Advanced Theory of Statistics*, 4th ed. London, U.K.: Griffin, 1983, vol. 3.

[12] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 2nd ed. New York: Wiley, 1984.

[13] J. Vandewalle and S. van Huffel, *The Total Least-Squares Problem*. Philadelphia, PA: Soc. Ind. Appl. Math., 1991.

[14] 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.*, pp. 169–176.

[15] A. F. Wrixon, "D-A converter test optimization," M.Eng.Sc. thesis, Univ. College, Dublin, Ireland, 1996.

[16] M. P. Kennedy, P. Grogan, J. O'Donnell, T. O'Dwyer, and A. Wrixon, "The LEMMA developer's toolbox: Semi-automated test development for analog and mixed-signal circuits," in *Proc. ISCAS'98*, Monterey, CA, May 31–June 3, 1998, vol. II, pp. 410–413.

[17] G. Strang, *Linear Algebra and Its Applications*. New York: Academic, 1976.

[18] M. P. Kennedy, D. Gillespie, P. Kiersey, L. Valsecchi, M. Vitale, and F. Ó. Cairbre, "On the use of the LEMMA method for characterising nonlinearities in resistive ladder DAC's," in *Proc. ISSC 1998*, Dublin, Ireland, June 25–26, 1998, pp. 477–484.

[19] C. Wegener, P. Grogan, T. O'Dwyer, and M. P. Kennedy, "New architecture—New trouble?," in *Proc. DEC'98*, Limerick, Ireland, Nov. 14, 1998, p. 1.

Adrian Wrixon was born in New Jersey in 1973. He received the B.E. degree in 1995 and the M.Eng.Sc. degree in 1997, both in electronic engineering, from the Department of Electronic and Electrical Engineering, University College Dublin, Dublin, Ireland, and the M.S. degree in 1998 from the University of California, Berkeley, in the field of computer-aided design.

His current research focuses on static variable ordering techniques for binary decision diagrams.

Michael Peter Kennedy (SM'95–F'98) received the B.E. degree in electronics from the National University of Ireland in 1984, and the M.S. and Ph.D. degrees from the University of California, Berkeley (UCB), in 1987 and 1991, respectively.

He was a Design Engineer with Philips Electronics, a Postdoctoral Research Engineer with the Electronics Research Laboratory, UCB, and an Invited Professor with the EPFL, Switzerland. He returned to University College Dublin (UCD) in 1992 as a College Lecturer with the Department of Electronic and Electrical Engineering, where he teaches electronic circuits, computer-aided circuit analysis, nonlinear circuits and systems, and directs the undergraduate electronics laboratory. He was appointed Senior Lecturer at UCD in 1996 and Associate Professor in 1999. He has published more than 150 articles in the area of nonlinear circuits and systems and has taught courses on nonlinear dynamics and chaos in the U.K., Switzerland, Italy, and Hungary. His research interests are in the simulation, design, analysis, synchronization, and control of nonlinear dynamical systems for applications in communications and signal processing. He is also engaged in research into algorithms for mixed-signal testing.

Dr. Kennedy received the 1991 Best Paper Award from the *International Journal of Circuit Theory and Applications* for his paper (with L. Chua) entitled "Hysteresis in electronic circuits: A circuit theorist's perspective" and the Best Paper Award at the European Conference on Cirecuit Theory and Design, 1999. He serves as a reviewer for a number of conferences, journals, and publishers, as well as national science funding agencies in Europe, Asia, and the U.S. He served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—PART I from 1993 to 1995, as Guest Editor of Special Issues on "Chaos Synchronization and Control" (1997) and "Advances in Nonlinear Electronic Circuits" (1999), and was Chair of the IEEE Technical Committee on Nonlinear Circuits and Systems from 1998 to 1999. He was elected a Fellow of the IEEE in 1998 "for contributions to the theory of neural networks and nonlinear dynamics, and for leadership in nonlinear circuits research and education."