Poor Man's TBR: A Simple Model Reduction Scheme

Joel R. Phillips, *Member, IEEE,* and L. Miguel Silveira, *Senior Member, IEEE*

***Abstract*—This paper presents a model reduction algorithm motivated by a connection between frequency-domain projection methods and approximation of truncated balanced realizations. The method is computationally simple to implement, has near-optimal error properties, and possesses simple error estimation and order-control procedures. Usage of the method also enables straightforward exploitation of information about the particular application and setting, as well as circuit functional information, such as frequency weighting information and correlations between network port waveforms. When such specific information is available, standard truncated balanced realization algorithms generate models far from optimal according to statistical decision criteria. Examples are shown to demonstrate that the method can outperform the standard order reduction techniques by providing similar accuracy with lower order models or superior accuracy for the same size model.**

***Index Terms*—Balanced realization, input correlation, model reduction, reduced-order systems.**

I. Introduction

MODEL reduction algorithms are now standard techniques in the integrated circuits community for analysis, approximation, and simulation of models arising from interconnect and electromagnetic structure analysis. Krylov subspace projection methods such as Pade-via-Lanczos (PVL) [1] and passive reduced-order interconnect macromodeling algorithm (PRIMA) [2] have been the most widely studied over the past decade. They are very appealing due mostly to their simplicity and their overall strong performance in terms of efficiency and accuracy.

However, Krylov projection methods are known to have two drawbacks in practical application. First, there is no general agreement on how to control error in these methods. Error estimators do exist for some methods [3], but they are seldom used in practice. The drawbacks of these estimators are that they require additional computation, which can be expensive and awkward to implement, and produce error estimates only at single frequency points, which leaves open the problem of error estimation over a range of frequencies. Second, moment-based methods, such as PRIMA, are known in some cases to produce models that are "too high" in order with the obvious consequences in terms of analysis or simulation cost [4]–[6]. Multipoint rational approximations produce more compact models than moment matching, but error theory is even less well developed [7].

An alternate class of model reduction schemes is the truncated balanced realization (TBR) family [8]. These are purported to produce "nearly optimal" models and have easy to compute *a posteriori* error bounds. As the TBR methods are too expensive to directly apply to integrated circuit problems, various two-stage and iterative Krylov methods have been proposed [5], [9]–[13] that combine Krylov subspace projection and TBR. While these hybrid techniques do a fairly good job of addressing the excessive order issue, the error-bound properties are weakened. Second, they are awkward in treating nonsymmetric, particularly very unbalanced systems, when two separate projection subspaces must be combined. Third, the methods are perceived as being complicated to implement and, so, have not been widely used in practice. Implementation of the TBR techniques requires considerable machinery from control theory and multiple numerical procedures that are tricky to implement in a stable way: solution of Lyapunov equations, balancing transformations, and/or eigendecompositions of matrix products.

The first main contribution of this paper is to illustrate a direct connection between two existing algorithms: multipoint rational approximation techniques and TBR. This connection motivates a new algorithm PMTBR, whose major attraction is its simplicity. It possesses some of the advantages of both techniques: the straightforward implementation of the projection methods and the excellent compaction properties of TBR. As a side benefit, it provides further theoretical basis for the empirically observed excellent performance of multipoint projection. PMTBR appears to have promising properties with respect to order control and error estimation, which, while not as powerful as TBR's error control, appears to be an advance over multipoint projection.

The second main contribution of this paper is to illustrate how algorithmic improvements can be made through the statistical interpretation [14] of the PMTBR procedure. In this viewpoint, the TBR procedure can be seen as a special case of a more general method that obtains reduced models as a type of maximum entropy procedure, with the specifics of candidate models determined by the statistical distribution assumed for the inputs to the state-space system. TBR corresponds to the limiting case of assuming a totally noninformative prior distribution on the inputs. Utilizing distributions with more information about a particular problem structure leads to more effective algorithms, and we will demonstrate two illustrative examples.

The first example addresses one problematic feature of the TBR approach. While TBR provides strong guarantees on error, the method is a global one in the frequency domain, with no

Manuscript received March 1, 2004; revised June 1, 2004. This paper was recommended by Guest Editor G. G. E. Gielen.

J. R. Phillips is with Cadence Berkeley Laboratories, Cadence Design Systems, San Jose, CA USA (e-mail: jrp@cadence.com).

L. M. Silveira is with IST/INESC ID–Cadence Laboratories, San Jose CA 95134 USA. He is also with the Department of Electrical and Computer Engineering, IST–Technical University of Lisbon, 1000–029 Lisboa, Portugal (e-mail: lms@inesc-id.pt).

Digital Object Identifier 10.1109/TCAD.2004.839472

control over allocation of modeling effort to different frequency bands. Thus, the near-optimal approximation properties of TBR are only near-optimal for classes of problems rarely encountered in practical circuit analysis. Various approaches to frequency weighting have been proposed that generally involve construction and reduction of a composite system by pre- and/or post-multiplying the original system by auxiliary weighting systems [15]–[17]. For narrowband applications such as in RF circuits, construction and merging of such auxiliary systems is not desirable. Projection methods, on the other hand, can be easily tuned to generate accurate approximations in any frequency bands of interest, but lack any theoretical guarantees or error bounds of any kind. As we shall see, the statistical viewpoint on the PMTBR algorithm leads directly to a *frequency-selective* variant of PMTBR.

The second example addresses the problem of reducing systems with a large number of input/output ports, also known as massively coupled systems. Such systems typically occur in substrate and package parasitic networks. Algorithms such as PRIMA [2] and PVL [1] are considered impractical for such networks. They rely on block iterations, where the size of the block equals the number of input/output ports. Therefore, each block iteration considerably increases the size of the model. For example, if a moment-matching (Krylov-subspace) algorithm is used to reduce a network with 1000 ports, and if only two (block) moments are to be matched at each port, the resulting model will have 2000 states, and the reduced system matrices will be dense. This makes simulation in the presence of nonlinear elements impractical. PMTBR is intrinsically somewhat less sensitive to the number of input ports. Much more importantly, however, in the PMTBR framework it is possible to exploit circuit functional information that results in correlations between the waveforms incident on the parasitic network ports. By exploiting this information, an *input-correlated* variant of the PMTBR procedure can be derived that enables significant further model order reduction.

This paper is structured as follows. In the next section, we review background information on model reduction algorithms. In Section III, we introduce the PMTBR algorithm and describe how it is used to construct a reduced order model. In Section IV, we describe the statistical interpretation of PMTBR and explore its implications with the use of the two examples described above. In Section V, we discuss some practical issues associated with the implementation and usage of PMTBR, namely extension to descriptor systems, order control, simultaneous approximations to both Gramians, and passivity. In Section VI, we present experimental results that demonstrate the general properties of the proposed method as well as the specific example applications discussed. Finally, in Section VII, some conclusions are drawn.

II. Model Reduction Background

A. Projection Framework

Many modern interconnect modeling technologies rely heavily on projection-based model reduction algorithms. For simplicity of exposition, consider for the moment the restricted case of linear system models

$$\frac{dx}{dt} = Ax + Bu \qquad y = Cx \tag{1}$$

with input u and output y, that are described by the matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times p}$, $C \in \mathbb{R}^{p\times n}$. These algorithms take as input a linear system of the form (1) and produce a reduced model

$$\frac{dz}{dt} = \hat{A}z + \hat{B}u, \quad y = \hat{C}z \tag{2}$$

where $\hat{A} \in \mathbb{R}^{q\times q}$, $B \in \mathbb{R}^{q\times p}$, $C \in \mathbb{R}^{p\times q}$. This is achieved by constructing matrices W and V whose columns span a useful subspace, and projecting the original equations in the column spaces of W and V as

$$\hat{A} \equiv W^T AV \qquad \hat{B} \equiv W^T B \qquad \hat{C} \equiv CV. \tag{3}$$

Most common choices are based on picking the columns of W, V to span a Krylov subspace [1], [2]. Different choices will lead to different algorithms with slightly different properties but an overall similar flavor.

B. TBR

Model reduction via balanced truncation is based on the analysis of the controllability and observability Gramians X, Y respectively. The Gramians are usually computed from the Lyapunov equations

$$AX + XA^T = -BB^T \tag{4}$$

$$A^T Y + YA = -C^T C. \tag{5}$$

Reduction is performed by projection onto the invariant subspaces associated with the dominant eigenvalues of the product of Gramians XY [8], [18]. For example, the approach of [18] corresponds to the projection procedure above with W, V the orthonormal bases arising from the Schur decomposition of the product of Gramians. One of the important features of TBR is an absolute bound on the error of approximation. If we let σ_i denote the square root of the ith largest eigenvalue of XY (XY always has real eigenvalues) then the error in the transfer function of the order q TBR approximation is bounded by $2\sum_{i=q+1}^{N} \sigma_i$ [19].

C. Multipoint Rational Approximation

An evolution of Krylov-subspace schemes are methods that construct the projection matrix V from a rational, or multipoint, Krylov subspace [7], [9], [20]. Compared to the single-point Krylov-subspace projectors, for a given model order the multipoint approximants tend to be more accurate, but are usually more expensive to construct. Given M, complex frequency points s_k, a projection matrix may be constructed whose kth column is

$$z_k = (s_k I - A)^{-1} B. \tag{6}$$

This leads to multipoint rational approximation. Multipoint projection is known to be an efficient reduction algorithm in

that the number of columns, which determines the final model size, is usually small for a given allowable approximation error, at least compared to pure moment matching approaches. Of course, there are many practical questions to ponder in an actual implementation: how many points s_k should be used, and how should the s_k be chosen? How is error determined? How is linear independence of the columns of V enforced?

Consider enforcing linear independence. An obvious strategy is to perform an SVD on the vectors z_k computed as above. A main point of this paper is that constructing projection matrices by multipoint frequency sampling, as in (6), followed by an SVD, in fact converges to the TBR algorithm. The singular values obtained from such a procedure approximate the Hankel singular values, and can thus be used for order and error control.

III. PMTBR Approach

A. Analysis in Frequency Domain

For simplicity consider the case $A = A^T$, $C = B^T$ and further assume that A is stable. This case is of more than theoretical interest as it occurs in RC circuit analysis, and the standard TBR algorithm is known to produce passive approximants [12]. It is easy to see that in this symmetrized case, both Gramians are equal and in the standard TBR procedure are obtained by solving the Lyapunov equation, (4).

The more fundamental definition of the Gramian X is obtained from the state evolution operator, also called the fundamental solution, of the differential equation $dx/dt = Ax + Bu$. The Gramian can also be computed in the time domain as

$$X = \int_0^\infty e^{At} B B^T e^{A^T t} dt. \tag{7}$$

However, noting that the Laplace transform of e^{At} is $(sI - A)^{-1}$, it follows immediately from Parseval's theorem that the Gramian X can also be computed from the expression[1]

$$X = \int_{-\infty}^{\infty} (j\omega I - A)^{-1} B B^T (j\omega I - A)^{-H} d\omega \tag{8}$$

where superscript H denotes Hermitian transpose. Consider evaluating X via applying numerical quadrature to (8). Given a quadrature scheme with nodes ω_k and weights w_k, and defining

$$z_k = (j\omega_k I - A)^{-1} B \tag{9}$$

an approximation $\hat{X}$ to X can be computed as

$$\hat{X} = \sum_k w_k z_k z_k^H. \tag{10}$$

Let Z be a matrix whose columns are z_k, and W a diagonal matrix with diagonal entries $W_{kk} = \sqrt{w_k}$. Equation (10) can be written more compactly as

$$\hat{X} = Z W^2 Z^H. \tag{11}$$

[1]Or any similar expression integrating along an appropriate contour.

B. Model Construction via SVD

To derive a model reduction procedure, consider the eigendecomposition

$$X = V_L \Sigma V_L^T. \tag{12}$$

Note that $V_L^T V_L = I$, since X is real symmetric in this special case. An obvious candidate for reduction would be to pick a projection matrix formed from the columns of V corresponding to the dominant eigenvalues of X. If the quadrature rule is accurate, $\hat{X}$ will converge to X, which by perturbation analysis of invariant subspaces [21], implies the dominant eigenspace of $\hat{X}$ converges to the dominant eigenspace of X. Now, consider the singular value decomposition of ZW.

$$ZW = V_Z S_Z U_Z \tag{13}$$

with S_Z real diagonal, V_Z and U_Z unitary matrices. Clearly

$$\hat{X} = V_Z S_Z^2 V_Z^T. \tag{14}$$

So, the dominant singular vectors in V_Z, as can be identified from the singular values in S_Z, give the eigenvectors of $\hat{X}$. Therefore, V_Z converges to the eigenspaces of X, and the Hankel singular values are obtained directly from the entries of S_Z. V_Z can then be used as the projection matriz in a model order reduction scheme.

It seems likely that the singular values of the matrix Z would have something to do with approximation error. The above illustrates that the correspondence is in fact precise; the SVD of Z reveals the same information revealed by TBR (modulo the weights W).

An obvious question is: how fast does the proposed scheme converge. In particular, how fast do the dominant singular vectors of ZW approach the dominant eigenvectors of X? As we will demonstrate, it turns out that very good models can be obtained with a fairly small number of sample points, in agreement with previous experience with multipoint approximation. For this reason, we denote our method "Poor Man's" TBR (PMTBR), since the quantities computed are cheap approximations to full TBR.

Surprisingly, as we shall shortly demonstrate, in many *practical* applications, PMTBR performs *better* than TBR in the sense of giving more accurate models for a given model size or amount of effort. This unexpected bonus demonstrates the virtues and rewards of frugality.

The PMTBR algorithm is shown as Algorithm 1 in Fig. 1. We formulate the approach to allow the sample points s_i to be arbitrary points in the right half-plane.

C. Computational Complexity

To compare the cost of computing a qth order model for a system with n states, using any of the proposed methods, we begin by quoting the complexity results for the basic operations involved. For the Krylov-subspace algorithms and PMTBR, these are either SVD or QR operations, at a cost of $O(nq^2)$, matrix solves, at a cost of $O(n^\alpha)$ (typically, $1 \le \alpha \le 1.2$ for circuits), and matrix factorizations, at a cost of $O(n^\beta)$ (typically, $1.1 \le \beta \le 1.5$ for circuits). As we saw, all of these algorithms can be cast into the projection framework, with

Algorithm 1: **PMTBR: Poor Man's TBR**

1. Do until satisfied:
2. Select a frequency point s_i.
3. Compute $z_i = [s_i I - A]^{-1} B$.
4. Form the matrix of columns
5. [real s_i]: $Z_{(i)} = [z_1, z_2, \ldots, z_i]$.
   [complex s_i]: $Z_{(i)} = [z_1, z_2, \ldots, z_i, z_i^*]$.
6. Construct the SVD of $Z_{(i)}$.
   If the error is satisfactory, set $Z = Z_{(i)}$, go to Step 7.
   Otherwise, go to Step 2.
7. Construct the projection space V from the orthogonalized column span of Z, dropping columns whose associated singular values fall below a desired tolerance.

Fig. 1. PMTBR algorithm.

appropriate projection matrices. Since such costs are similar for all the algorithms considered, we will discard the cost of such projections for purposes of comparison.

The TBR algorithm implies the computation of the Gramians X and W, from (4) and (5) and the eigenvalues of their product, at a cost of $O(n^3)$. Clearly, the cubic cost of TBR limits the use of this algorithm for small- to medium-sized problems. Krylov-subspace algorithms require one $O(nq^2)$ QR factorization, one matrix factorization, and q solves, for a total cost of $O(nq^2 + qn^\alpha + n^\beta)$.

Assuming that q frequency points are chosen in the quadrature scheme for PMTBR, examination of Algorithm 1, indicates that it involves one SVD, q solves, and q factorizations, for a total cost of $O(nq^2 + qn^\alpha + qn^\beta)$. This is the same cost as for multipoint rational approximations.[2] On the strength of the above comparison, for a given size model, it appears that Krylov-subspace algorithms provide the most efficient procedures for order reduction due to the smaller number of factorizations. However, since PMTBR and multipoint projection may not need as high an order a model, the choice depends on the relative cost of factorization, and PMTBR is likely preferred if α is close to β. In addition, when q frequency points are chosen for the PMTBR algorithm, experiences indicates that it is highly likely that the final model, after compression via the SVD will lead to a $\hat{q}$th order model, $\hat{q} < q$ (see Step 7 of Algorithm 1). If the simulation cost dominates over the model computation cost, as is often the case, then the added compactness of PMTBR makes this algorithm the most efficient alternative. For this reason, since the cost is the same, and PMTBR generally produces smaller models, there is no reason to prefer multipoint projection.

IV. Exploiting Statistical Information

A. Statistical Interpretation of TBR

The PMTBR approach has an interesting statistical viewpoint that is useful in motivating new algorithms. First, let us consider a statistical interpretation of the standard TBR procedure [14].

[2]Notice that we are accounting for the cost of a single SVD. However, from Algorithm 1 it appears that a new SVD is performed every time that a new sample vector z_i is added to the collection. In practice this is not done, as is discussed ahead in Section V-C, which justifies our estimation.

Consider the controllability operator $\mathcal{L}$: $[-\infty, 0]^m \rightarrow \mathbb{R}^n$, which maps the inputs of an m-input n-state linear state-space model to the state $x_0 = x(0)$ at time zero, $x_0 = \mathcal{L}u$. Suppose, for simplicity, that the system is in balanced coordinates. The controllability Gramian X as given in (4), (7) may also be written

$$X_c = \mathcal{L}\mathcal{L}^\dagger$$

(with $\mathcal{L}^\dagger$ denoting the adjoint operator). Consider interpreting $u(t)$ as a zero mean random variable with Gaussian distribution, inputs uncorrelated, and each input having autocorrelation (matrix) function $R_u(t_1, t_2) = \delta(t_1 - t_2)I$, with I the identity matrix. The time-domain inputs are unit power. x_0 is then also a Gaussian random variable with correlation matrix

$$E\left\{x_0 x_0^T\right\} = E\{\mathcal{L}uu^\dagger\mathcal{L}^\dagger\} = \mathcal{L}E\{uu^\dagger\}\mathcal{L}^\dagger = X_c.$$

Thus, in this viewpoint, the controllability Gramian can be seen as the covariance matrix associated with the state vector. Similar statements can be made for the observability Gramian, though the physical interpretation is less direct. The TBR procedure constructs the Karhunen–Loeve transformation [22] of the associated random process. The entropy of such a process is related to the variances, which are the eigenvalues of X_c.[3] The reduced models of order q created by TBR can be seen as maximum-entropy q-state models, given the assumption of uncorrelated inputs with white spectrum. In a Bayesian-like interpretation of the procedure, we would say that the TBR procedure produces a maximum entropy model associated with the prior assumption of distribution on the inputs that is uninformative. PMTBR constructs a maximum entropy model consistent with the observed column samples which are drawn according to our assumptions on the joint distribution of the inputs. This interpretation suggests a route to more efficient modeling strategies.

From an information theoretic viewpoint, the lower the entropy associated with the state x, the less effort required to represent it. In our context, the smaller the entropy of the process x, the smaller the model we can potentially obtain. A fundamental result of information theory is that conditioning reduces entropy [23]. We should expect that the stronger the degree of prior information about the input structure, the lower the entropy, and the smaller the model that can be constructed.

As an example, consider how restricting the choice of input vectors in TBR affects the size of the reduced models. A large circuit may have many possible ways to construct the input/output ports, i.e., the B/C matrices. That is, in a large circuit, there are many nodes where it is possible to *potentially* observe the output, or *potentially* attach a source. In practice, only some nodes are of interest either for drive or for observation, and we now show how this selection process affects the models produced by TBR.

Let $\aleph_j$ and $\aleph_k$ denote two possible subsets of possible input vectors $B_{\aleph_j}, B_{\aleph_k}$. The conditioning property of entropy implies

$$\aleph_j \subset \aleph_k \Rightarrow H(X_j) < H(X_k)$$

[3]The Shannon entropy is essentially the sum of the log of the eigenvalues of X_c.

where $H(X_k)$ denotes entropy of the state associated with input partitioning k. It is easy to verify explicitly that addition of input vectors monotonically increases the entropy. Let

$$B_1 = B_{\aleph_j \cap \aleph_k} \qquad B_2 = B_{\aleph_k - \aleph_j \cap \aleph_k}.$$

If

$$AP_1 + P_1A^T = B_1B_1^T$$
$$AP_2 + P_2A^T = B_2B_2^T$$

then[4]

$$AP + PA^T = [B_1B_2][B_1B_2]^T \tag{15}$$

and it follows that $P = P_1 + P_2$. The claim follows from the strict concavity of the function $\log\det P$ and the symmetric positive definiteness of P_1, P_2. Thus, the fewer the inputs, or outputs, or the weaker the relation between, say, a given input and the observed state, the easier the model reduction problem. That such situations occur in practical circumstances is the basic premise of model reduction.

Given that the inputs are chosen, TBR can be interpreted as an optimal procedure for selecting a model according to the two alternate criteria.

1) For a fixed order, the expected deviation (error) in the state is minimized. This is a maximum likelihood type criteria.
2) For an allowed deviation, or variance, the minimal order model is chosen. This is a minimum description length type criteria.

However, if further information is available about the structure of the inputs, TBR will *not* produce the optimal model, according to these criteria. A more detailed analysis is necessary, and the intermediate computations (e.g., the Lyapunov equations) cannot generally be done in closed form.

We will now present two specific contexts in which more detailed knowledge about the anticipated structure of the inputs can lead to more efficient reduction procedures. PMTBR plays the role of an approximate procedure, which chooses the minimal length model, given the observed samples z_k [24].

B. Frequency-Selective TBR

Consider evaluating (8) by breaking the integral into partial sums I_k each of which is an integral over a section S_k of the imaginary axis

$$X = \sum_{k=1}^{\infty} I_k \tag{16}$$

$$I_k = \int_{S_k} (j\omega I - A)^{-1} BB^T (j\omega I - A)^{-H} w(\omega) d\omega \tag{17}$$

where $\bigcup_{i=1}^{\infty} S_i$ amounts to the whole imaginary axis.

Each I_k gives the contribution to X from the system's behavior over the interval S_k. X is the Gramian of the operator that maps input to state; its singular values give the norm of that operator. This suggests interpreting I_k as the contribution from the input u over the frequency interval S_k. The standard TBR procedure, having no *a priori* knowledge of the frequency content of the input, weights each frequency equally. However, in almost all practical problems, we have some knowledge of the actual frequency distribution of the inputs. Often, the inputs are bandlimited, or nearly so, or we might be interested only in the behavior of the system around some finite frequency interval. We propose truncating the sum in (16) to finite intervals, and using the resulting "finite-bandwidth" Gramian for model reduction. Since the resulting Gramian places more emphasis on frequencies of relevance, we expect to achieve better performance, for a given model order, on problems with finite bandwidth inputs. More generally, we may define a "frequency-weighted" Gramian as

$$X_{FW} = \int_{-\infty}^{\infty} (j\omega E - A)^{-1} BB^T (j\omega E - A)^{-H} w(\omega) d\omega \tag{18}$$

where $w(\omega)$ is the weighting function (the notational similarity with quadrature weights is deliberate). The more appropriate the weighting function to our problem at hand, the better we expect the performance of the reduction algorithm to be. Seen from this viewpoint, TBR is a generic, somewhat naive, algorithm as it presumes complete ignorance of frequency content. The weighting function in standard TBR is most appropriate for white noise inputs where nothing is really known about frequency content.

In a practical implementation, with a finite number of frequency samples, weighting can be accomplished by adjusting the weights w_k and/or location of samples ω_k. In fact, every ZW-matrix implicitly defines a frequency weighting scheme. For this reason, it is better to choose points/weights in PMTBR (perhaps adaptively) according to the expected frequency profile of the system and the inputs, than to try to achieve convergence to the TBR Gramians themselves.

This analysis provides another explanation for the empirically observed fact that multipoint projection can sometimes exhibit better relative error performance than generic TBR. Multipoint projection more highly weights points in (18) than the standard TBR weighting, resulting in better relative performance in those areas. In the search for good global error performance, TBR can over-emphasize areas of the transfer function that are large in magnitude. When such regions are of interest to the problem at hand, TBR is a nearly optimal method. However, when such regions lie out of the frequency band of interest, or lead to excessive sacrifice of relative error for absolute error, TBR may not do as well as multi-point projection. PMTBR, on the other hand, can always be tailored to the problem at hand.

The frequency-selective TBR procedure is shown as Algorithm 2 in Fig. 2. The similarity with Algorithm 1 should be fairly obvious, the main distinction being in the point-selection algorithm.

C. Input-Correlated TBR

In the previous section, we showed how exploiting knowledge of the frequency profiles of the inputs led to a more efficient procedure. We implicitly assumed no knowledge of the relation *between* the inputs. In this section, we discuss the input-correlated TBR procedure [25], describe how it fits into the statistical

[4]To see that this is true, observe that the right-hand side $[B_1\,B_2][B_1\,B_2]^T = B_1B_1^T + B_2B_2^T$ is just matrix multiplication written as outer products, then note that the left-hand side of (15) is linear in P.

Algorithm 2: **Frequency Selective TBR**
/* assuming M bands of interest */
1. Define the range corresponding to the frequency bands of interest, $S = \bigcup_{k=1}^{M} S_k$
2. Do until satisfied:
3. Select a frequency point s_i **within** S.
4. Compute $z_i = [\, s_i I - A]^{-1} B$.
5. Form the matrix of columns $Z_{(i)}^{fs} = [\, z_1, z_2, \ldots, z_i]$.
6. Construct the SVD of $Z_{(i)}^{fs}$.
   If the error is satisfactory, set $Z = Z_{(i)}^{fs}$, go to Step 7.
   Otherwise, go to Step 3.
7. Construct the projection space V from the orthogonalized column span of Z, dropping columns whose associated singular values fall below a desired tolerance.

Fig. 2. Frequency-selective TBR algorithm.

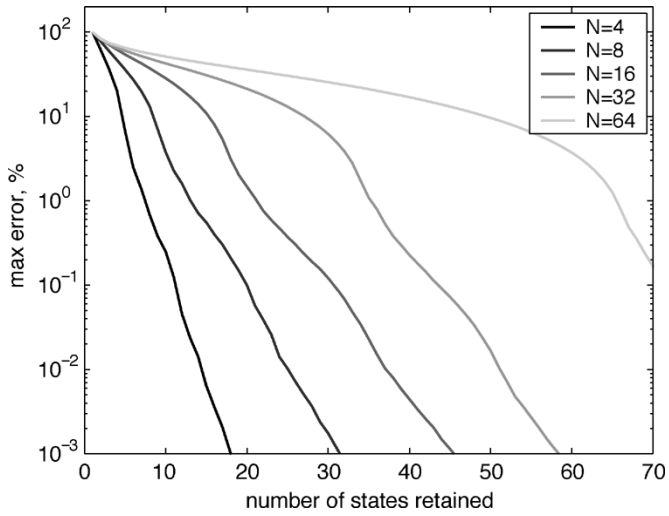

Fig. 3. TBR error bounds for 12 × 12 *RC* mesh as function of number of inputs.

context, and show how it leads to more efficient reduction procedures.

Consider applying model reduction algorithms to networks with a large numbers of input/output ports, that is, for networks with many columns in the matrices defining the inputs. Such systems occur often in practice, for instance when dealing with massively coupled parasitic networks as occur in substrate analysis, package modeling, and on large digital interconnect networks.

To motivate our algorithm, let us examine the behavior of the Hankel singular values for a simple system, such as an RC mesh, as the number of ports (i.e., columns in the B-matrix) varies. Fig. 3 shows the TBR error bound obtained from the Hankel singular values as a function of the number of inputs. We can conclude that the order needed for good accuracy grows with the number of inputs. This is contrary to the common expectation that a few poles are sufficient for *RC* systems, but in accord with the statistical analysis above. For systems with many inputs, many states may be needed because of the high dimension of the controllable space. Even in this simple RC circuit case, for the 64-input case, low accuracy (20% error bound) requires at least 40 states in the reduced model.

Other available procedures are likewise impractical. The moment-matching (Krylov-subspace) family of algorithms, such as PRIMA [2] and PVL [1], lead to models whose size is the number of ports multiplied by the number of moments matched. At low frequencies, the PACT [26] algorithm can lead to smaller numbers of states than PRIMA, since it does not rely on matching (block) moments. However, PACT still leads to matrices that are dense, and whose size is still bounded from below by the number of ports.

The key to a more efficient procedure lies in noting that in many practical problems, the inputs to an interconnect network are not arbitrary. Often it is necessary to retain all the input ports if the full impact of parasitic effects is to be correctly estimated [27], but there may be relations between the inputs (or output) at different network ports that can be exploited to give a smaller model. Suppose a correlation matrix [22] K for the input relations is known (that is, the correlation function discussed above $R_u(t_1, t_2)$ is $R_u = \delta(t_1 - t_2)K$). The appropriate Gramian for this restricted problem is given by

$$AX_c + X_c A^T = -BKB^T.$$

The key insight is, for symmetric positive definite K, the eigenvalues of X_c decay faster than the eigenvalues of X from (4), if the eigenvalues of K exhibit some decay. In other words, X_c is closer to a low-rank matrix than X if the inputs exhibit some correlated behavior, as in the perfectly uncorrelated case the eigenvalues of K are identical. This is equivalent to saying that we have partial information about the relation between the inputs, the case $K = I$ corresponding to zero information. Thus, for a given truncation criterion for the singular values, using X_c for a model reduction procedure will lead to smaller models. This is a special case of the more general observations made above. Under the hypothesis that K is a suitably representative model of the possible inputs, no accuracy will be lost. In practical problems, such fidelity can be guaranteed if we are suitably conservative in the specification of the correlation matrix ($K = I$, corresponding to the ultimate degree of safety, total ignorance). As in the frequency-selective case, though the physical interpretation as an absolute error bound no longer applies, the eigenvalues of the Gramian can still be used for error control, as they can be given an interpretation associated with the likelihood of error in the probabilistic input model.

To estimate input correlations, consider taking a set of N samples of input waveforms, u_k^l for input k, $k = 1 \ldots p$, where l is the sample index. The correlation matrix can be estimated as

$$K_{ij} = \frac{1}{N} \sum_{l=1}^{N} u_i^l u_j^l.$$

As is the usual case, the actual correlation matrix need not be formed. Instead, we can take the SVD of the matrix $\mathcal{U}$ whose columns are the input samples u_k, i.e.

$$\mathcal{U} = V_K S_K U_K^T$$

with U_K, V_K orthonormal.

Algorithm 3: **Input Correlated TBR**

1. Construct the SVD of inputs $\mathcal{U} = V_K S_K U_K^T$
2. Do until bored:
3. Draw a vector $r \in \mathbb{R}^p$ by taking p draws from a normal distribution, variances given by Σ_K.
4. Select a frequency point s_i.
5. Compute $z_i = [\, s_i I - A]^{-1} B U_K r$.
6. Form the matrix of columns $Z_{(i)}^{ic} = [\, z_1, z_2, \ldots, z_i]$.
7. Construct the SVD of $Z_{(i)}^{ic}$.
   If the error is satisfactory, set $Z = Z_{(i)}^{ic}$, go to Step 8.
   Otherwise, go to Step 3.
8. Construct the projection space V from the orthogonalized column span of Z, dropping columns whose associated singular values fall below a desired tolerance.

Fig. 4. Input-correlated TBR algorithm.

Note that, in addition, from this information, we can also obtain estimates of the frequency profile of the inputs. These estimates can be used to select the frequency points s_i for the PMTBR procedure. The final algorithm is shown as Algorithm 3 in Fig. 4.

V. Practical Implementation

A. Descriptor Systems

Usually, in circuit analysis it is inconvenient, and possibly prohibitively expensive, to translate to the form in (1). In the more general case, with the state-evolution equation given by $E dx/dt = Ax + Bu$, the controllability Gramian can be obtained from

$$AXE^T + EXA^T + BB^T = 0. \tag{19}$$

Not surprisingly, the frequency domain equation is

$$X = \int (j\omega E - A)^{-1} BB^T (j\omega E - A)^{-H} d\omega \tag{20}$$

and the above procedure follows exactly with the change that the columns of Z are given by

$$z_k = (s_k E - A)^{-1} B. \tag{21}$$

Note that the complications present in applying standard TBR to problems with singular E-matrices vanish in PMTBR.

B. Error Estimation

The above arguments can be extended to a generalized process of error estimation. The singular values obtained from the weighted Gramians can be interpreted as gains between filtered inputs and weighted outputs. Singular values from truncated modes can be interpreted as errors on the filtered system, i.e., finite-bandwidth or weighted errors. The singular value information can be used in three ways to guide model order control.

First, if enough samples are taken that good estimates of the true Gramians are obtained, then the singular values obviously provide error bounds, through the connection to TBR.

Second, the singular values can guide the process of point selection. With reasonable spaced sampling of points, as projection vectors are added to the ZW-matrix, convergence of the singular values indicates convergence of the error, which guides when to stop adding vectors to ZW.

Third, we have found that, again assuming a sampling density consistent with the weighting $w(\omega)$, the singular values usually give a fairly good guide to model order well before convergence is achieved. Our experiments indicate that when, for a number of samples in excess (e.g., twice) of the model order, the singular value distribution exhibits a small "tail" (that is, for a "small" ε, $\exists k$: $\varepsilon > \sum_{i=k+1}^{\infty} \sigma_i$), then sufficient order and point placement has been achieved. Again, this is, as one would expect, strikingly similar to the usual TBR concepts.

Of course, there is always the question of whether pathological cases, such as systems with very narrow resonances, can arise. In practical applications, fundamental physics usually establish limits on resonance width. When these limits are known up-front they can be used to guide minimum sampling densities. With appropriate weightings, adaptive schemes (e.g., bisection of frequency intervals) can also be used. Full discussion of point selection is beyond the scope of this paper, but in all systems we have encountered, approximation by multi-point projection is sufficiently powerful that we have not found point selection to be problematic in obtaining quality results.

C. On-the-Fly Order Control

To minimize computational effort, it is desirable to run the algorithm until a desired accuracy level is obtained, and at that point cease to add points into the sample space of PMTBR. We would make this decision by looking at a small number of the trailing singular values of the ZW matrix, and stopping when the sum drops below a given threshold. This means that estimates of the singular values of ZW must be available each time a sample (or set of samples, depending on the updating scheme) become available. In the previous development we utilized the SVD decomposition because of its direct connection to eigenvalue analysis, and the utility of having available the singular values for comparative analysis of the TBR and PMTBR algorithms. However, the SVD is not the most appropriate tool for such an adaptive order control procedure, since no fast update procedure is known. Thus when using the SVD, the computational complexity of adding a single new vector into the sample space is about the same as performing a new SVD of the entire set of column samples. However, in our case, we do not need the actual singular values themselves, only the ability to estimate the magnitude of a trailing few, and to obtain a basis for the dominant subspaces. Other rank revealing factorizations that possess better updating properties may be more appropriate, for example the RRQR [28], [29] and UTV factorizations [30].

We should clarify two potential points of confusion in implementation. First, the singular value estimates do not necessarily indicate convergence of the integral form to the Gramians (this is obviously the case over finite frequency ranges). Usually, good models are obtained well before the singular value estimates converge to the true Hankel singular values. A good example is a very high-Q two-state resistance–inductance–capacitance (*RLC*) circuit. Only two sample vectors are required to obtain the exact model, which PMTBR will correctly predict. However, obtaining the exact Gramian by numerical integration could require a very large number of quadrature points, especially if the points are placed in a naive manner such as a uniform distribution.

Second, in the SVD, the columns of the basis matrices U and V are potentially complex. However, they are only unique to within a phase factor which can be factored out to obtain real valued projection matrices and, thus, real-valued models as needed for implementation in time-domain analysis.

D. Cross-Gramian Method

In the general case, computation of both Gramians X and Y is required in order to perform the TBR procedure. The reduction depends on the dominant eigenspaces of the product XY, which is an invariant of the system, whereas X and Y individually depend on the chosen coordinates. Motivated by [31], we suggest as basis for a procedure the cross-Gramian X_{CG} that encodes controllability and observability information into a single matrix.

Analogous to the Lyapunov equations for X, Y, X_{CG} can be obtained from the Sylvester equation

$$AX_{\mathrm{CG}} + X_{\mathrm{CG}}A = -BC.$$

We propose performing model reduction by projection onto the dominant eigenspaces of X_{CG}. In the case of symmetric models, including single-input, single-output models as a special case, $XY = X_{\mathrm{CG}}^2$ and reduction based on eigenvalues of the cross-Gramian is identical to TBR. In the general case, since the eigenvalues of X_{CG} are the eigenvalues of the Hankel operator, not the singular values, the procedures are not the same. However, since the "sum-of-the-tail" of eigenvalues X_{CG} is a bound on the sum of the tail of the singular values [31], if the trailing eigenvalues of X_{CG} are small, we may still expect good models. It is easy to construct counterexamples, but we have obtained good results on many real systems.

The Sylvester equation is related to the time-domain expression

$$X_{\mathrm{CG}} = \int_0^\infty e^{At}BCe^{At}dt \tag{22}$$

and the frequency domain representation

$$X_{\mathrm{CG}} = \int_{-\infty}^{\infty} (j\omega I - A)^{-1}BC(j\omega I - A)^{-1}d\omega. \tag{23}$$

In the PMTBR context, we need to construct two sets of sample vectors, one corresponding to the controllability subspaces previously discussed, represented by a matrix Z^R with column vectors

$$z_k^R = (s_k I - A)^{-1}B$$

and the other to the observability subspaces, represented by a matrix Z^L whose columns are

$$z_k^L = (s_k I - A^T)^{-1}C^T.$$

The PMTBR procedure proceeds by performing the eigendecomposition of $Z^L(Z^R)^T$. Note that this matrix is nonsymmetric, and is of the same dimension n as the unreduced system. Therefore, explicit construction followed by eigenanalysis of this matrix is by hypothesis infeasible. We propose a procedure that compresses the eigenvalue problem. Suppose we compute a matrix Q, with orthonormal columns, whose columns span the joint column space of Z^R and Z^L. Then, $Z^R = QR^R$, $Z^L = QR^L$ for some (nonsingular) R^R, R^L. The eigenvalue problem becomes

$$QR^R(R^L)^TQ^TS = S\Lambda$$

or

$$R^R(R^L)^TR^S = R^S\Lambda$$

with $R^S = Q^TS$. This reduced eigenvalue problem can be solved, R^S truncated to the significant eigenvectors R_r^S, and the necessary projection matrix obtained in the original coordinates from QR_r^S.

E. Passivity and Stability

Because we do not compute the full Gramians, as developed here, PMTBR does not possess the stability and/or passivity [12] preserving properties of the full-blown TBR algorithms. A full remedy to this situation is possible but beyond the scope of this paper. However, the usual solution followed for integrated circuit problems, reduction via congruence transforms, can still be utilized. In this case, stability and passivity will be guaranteed for suitably formulated *RLC* circuit networks.

VI. Computational Experiments

In this section, we show results from applying PMTBR to a set of examples. The section is structured in three parts. First, we illustrate the relevant properties of PMTBR, its relation to TBR and compare it to alternate reduction methods. Then, we show an example that illustrates the frequency-selectivity capabilities enabled by PMTBR. Finally, examples are shown to illustrate the input-correlated variant of PMTBR and the dramatic reduction that it can provide.

A. General Properties of PMTBR

1) Convergence to TBR: In our first example, we consider an *RC* circuit model of a clock distribution network. This circuit, to a good approximation, is finite bandwidth. We use this example to illustrate the asymptotic equivalence of the TBR and PMTBR methods. Fig. 5 illustrates the singular values of the ZW matrix resulting from a moderate number (50) of sample points w_k. It can be seen that the estimated singular values, while not exact, are good approximations, and follow the general trend of the exact solution. It is interesting that the approximate singular values continue to rapidly decrease over nearly fifteen orders of magnitude, even with a relatively low accuracy approximation of the Gramians. PMTBR appears to capture the fact that this *RC* model is intrinsically low order. Of course, adding sample points would increase the accuracy of the singular value approximations, as we will show later.

More critical for model reduction is the estimation of the projection subspaces. Fig. 6 shows convergence of one *angle* between projection subspaces. In this case, we chose the second principal vector to estimate within the first four leading

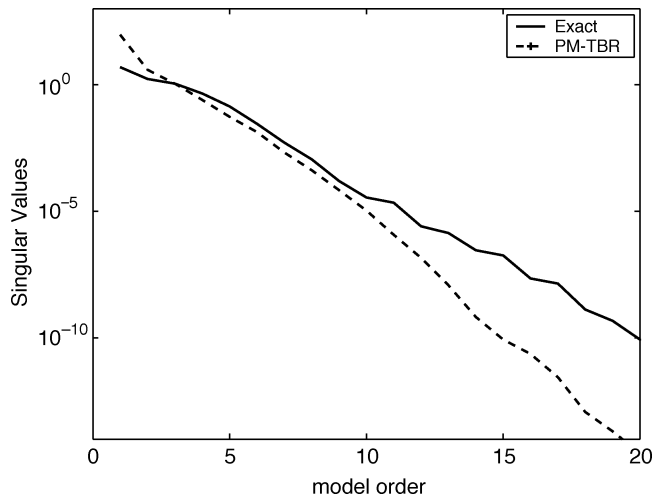

Fig. 5. Hankel singular values as computed from the exact Gramians (solid line) and the estimated from PMTBR (dashed line).

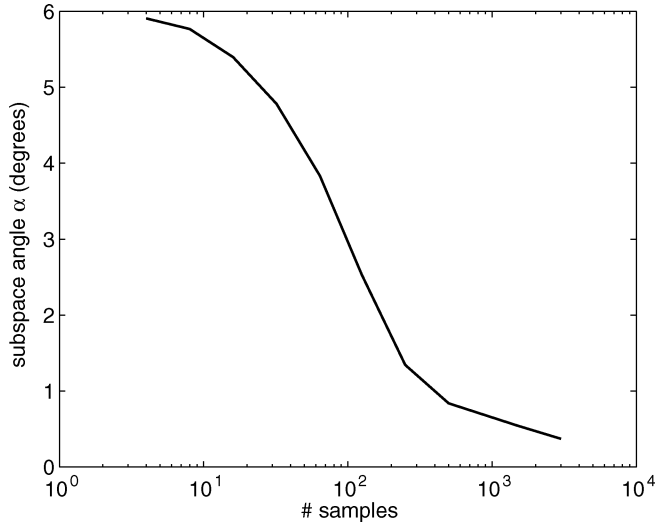

Fig. 6. Angle between the second principle vector and the PMTBR singular subspaces.

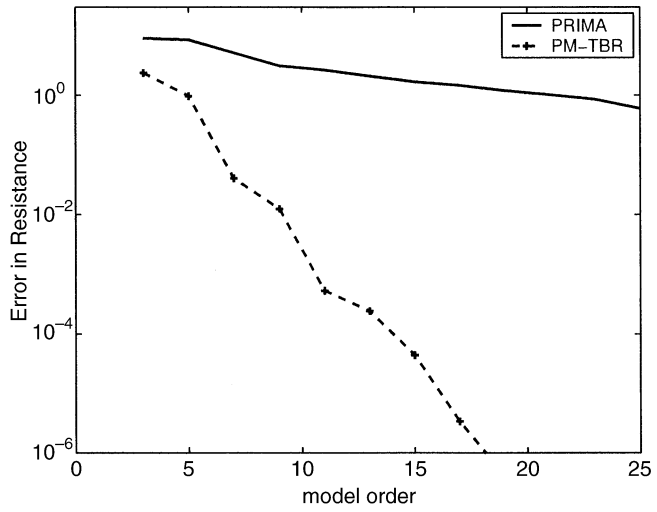

Fig. 7. Error of approximations of resistance obtained with PRIMA and PMTBR for increasing order models on the spiral inductor example.

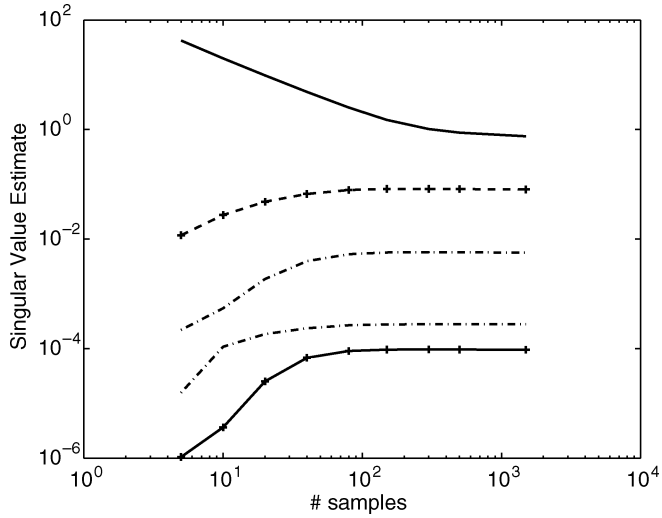

Fig. 8. Spiral inductor example, convergence of singular values of ZW.

subspaces of PMTBR. Even for small numbers of sample points, the subspaces are fairly closely aligned, and alignment increases with increasing number of samples. This indicated convergence of PMTBR to TBR for this example. The leveling out of the curve is due to the fact that the model under study has nonzero response outside the finite bandwidth used to compute the PMTBR results. By increasing the bandwidth over which we perform PMTBR, we could continue to decrease the subspace angles, but at some point the accuracy obtained thereby ceases to be of interest, because in a real problem, negligible signal strength exists outside finite bandwidths.

2) Comparison to PRIMA: In our next example, we use a model for an on-chip spiral inductor to demonstrate how PMTBR can outperform the standard model order reduction method PRIMA. Particularly for the real part of the inductor's impedance (i.e., the resistance), PRIMA converges slowly on this example. Fig. 7 shows a comparison of the error in the inductor's resistance for approximations obtained with PRIMA and PMTBR of increasing sizes. From the plot, one can see that the PMTBR approximation produces a more accurate approximation at any given order, and converges more quickly. As 30 frequency samples were used to compute the PMTBR model, at each order, more work was also required to compute the approximations for a given order. However, as about 60 PRIMA vectors are required to obtain 1% accuracy in the resistance, the overall work is still less.

Next, we demonstrate the order-control and error-estimation capabilities of PMTBR. Fig. 8 shows the convergence of the five largest singular values of $\hat{X}$ as the number of frequency-domain sample points (also referred to as quadrature nodes w_k) is increased. In this example, we used a very crude uniform sampling/weighting that would correspond to the "rectangle rule" in quadrature. We see that the largest five singular values have mostly converged by the time we reach 100 sample points. Fig. 9 shows the error versus order for PMTBR models using 100 sample basis points, as well as the error estimates computed using the singular values. First, we can see that increasing the order of the approximation beyond ten or twelve benefits very little, as the corresponding singular values are certainly below the relevant error and quickly approaching machine precision. Second, we see that, for the orders corresponding to the well-estimated singular values, the error estimates are very good. Esti-

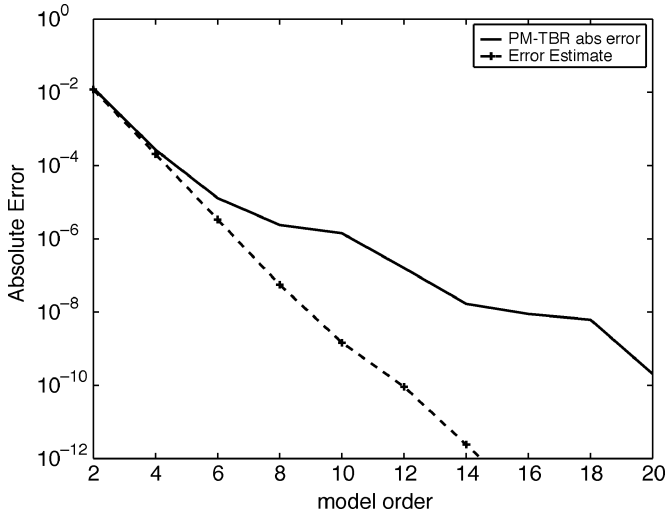

Fig. 9. Admittance transfer function error and error estimates for the spiral inductor example.

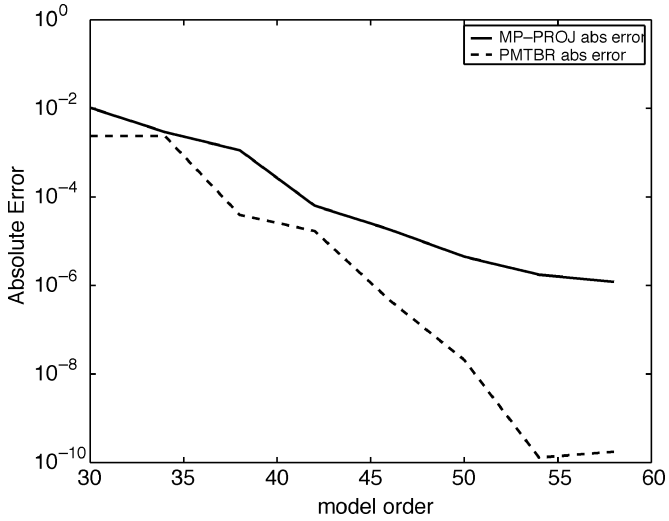

Fig. 10. Comparison of error in multipoint projection method (MPPROJ) and PMTBR for the PEEC example.

mates for higher orders are not as good, but do indicate correctly that the actual error is small and rapidly approaching zero.

Note that as indicated above for this example a much smaller number of samples was required to achieve excellent error performance on this model. Singular value plots with this sample size (not shown) show that a model of size 5–7 is indeed enough to achieve acceptable accuracy with this number of sample points.

3) Comparison to Multipoint Projection: Of course, an obvious question is, since the PMTBR technique uses the same information as multipoint projection, whether there is any advantage in using PMTBR over multipoint projection. To answer this, we show results from an example introduced in [1], a lumped-element equivalent circuit for a three-dimensional problem modeled via PEEC.

Fig. 10 shows a comparison of the errors incurred with approximations of increasing order obtained using a multipoint projection method and PMTBR for this PEEC example. The plots clearly show the superior accuracy of PMTBR for similar size models (equivalently, PMTBR is able to generate more compact size models for the same accuracy). Furthermore it is interesting to note that for high accuracy this difference actually increases, as the error of the projection method goes down very slowly with order increase. This is clearly due to the ability of PMTBR to prune out redundant information from the model. Note that in [1] an order 60th approximation computed with PVL was needed to obtain good approximation of the transfer function of the equivalent circuit.

It is interesting to observe that, just as the multipoint projection method did, the PMTBR technique is able to circumvent any difficulties involved with having singular A or E matrices. This is a matter of practical relevance as many systems obtained applying the modified nodal formulation to some discretized model often have singular A and/or E matrices (typically, in those cases A would be related to the conductance matrix and E to the capacitance matrix). Applying standard TBR to such systems involves some complicated preprocessing, see [12] for a discussion. Note also that PMTBR was quite accurate on this example, despite the fact that it contains sharp resonances that cause difficulty for quadrature (compare Fig. 5 where the singular value estimates are not exact). PMTBR does not produce the exact same singular value estimates as TBR, as it weights the contributions to the projection subspaces differently, the subspaces produced contain the system information relevant to the point selection chosen.

B. Frequency Selectivity

An important question to resolve is whether there is any advantage to PMTBR compared to a standard projection technique followed by standard TBR. We show results from an 18 pin shielded connector structure that was previously used to illustrate a PEEC formulation based on PRIMA that generates passive reduced-order models [6]. While the resulting model was indeed provably passive, disappointing reductions were reported, which were attributed to limitation in the PRIMA algorithm in dealing with the "relevant modes of the system." In order to address this issue, in [13], the same example was used to illustrate a two-step algorithm for *RLC* order reduction based on PRIMA followed by TBR. Significant order reductions were reported after the second step of reduction as TBR is able to determine that those modes are not observable nor controllable. Therefore, it is a good model system on which to compare PMTBR and TBR.

Fig. 11 shows a plot of the exact transfer function of the connector, as well as approximations obtained with TBR and PMTBR. For this particular example, we were interested in testing the ability of the PMTBR algorithm to produce approximations on a finite bandwidth. We decided to illustrate approximation over a finite range of 0–8 GHz. Samples were generated to cover the frequency range from dc to 8 GHz, and these samples were used to generate an order 18 PMTBR approximation. At the same time, a TBR approximation of order 30 was also generated (we found that 30 was the minimum order required for TBR to provide reasonable representation of *any* features in the 0–8 GHz range). From the figure, we can see that the PMTBR approximation does indeed show

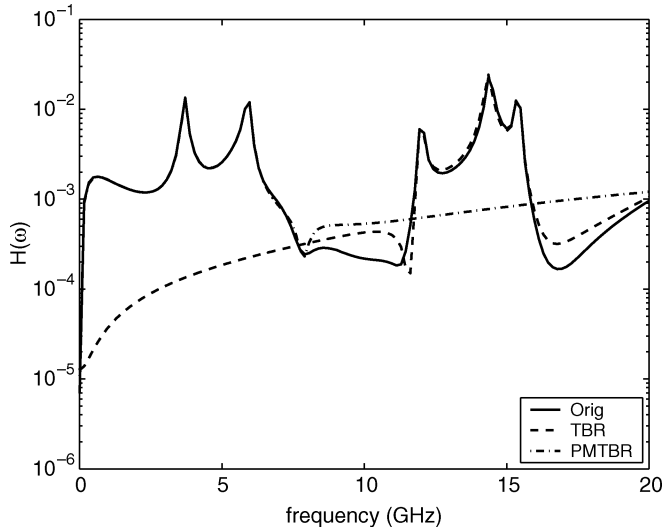

Fig. 11. Transfer function approximations for the connector example. Note PMTBR has better finite-bandwidth performance than TBR.

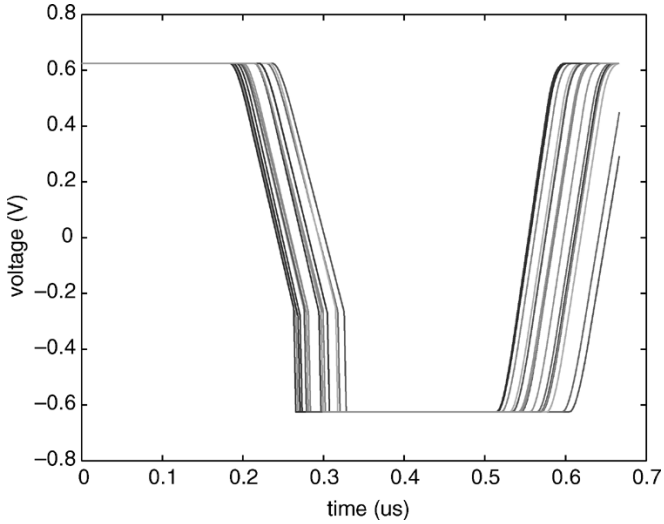

Fig. 12. Set of waveform samples for one input on RC network example.

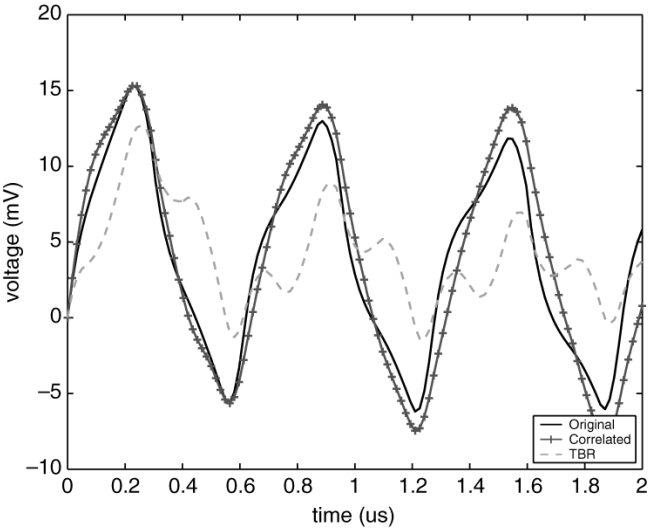

Fig. 13. Simulation results for one output on *RC* network example. PMTBR with correlation information out-performs TBR.

very good accuracy in the frequency range of interest. The figure also shows, rather dramatically, the inability of the TBR approximation to produce an accurate model at the frequency of interest, even with the higher order approximation. Furthermore, the TBR approximation seems to be accurately picking some features of the system but these happen to fall out of the bandwidth of interest. We believe TBR concentrates effort around 15 GHz because of the relative amplitude of the transfer function. PMTBR is easily focused on the 8 GHz and below range merely by selection of sampling points, and does not waste effort with approximation at higher frequencies.

C. Input-Correlated TBR Examples

In this section, we complete the experimental demonstration of PMTBR with the examples from [25] that show how increased information about input form can reduce model size for models with large numbers of inputs/outputs.

1) Simple Multi-Input RC *Circuit:* To illustrate the basic characteristics of the proposed reduction method, we will first consider a 32-port *RC* interconnect network. To simulate the situation where there is some degree of information about the relation between inputs, we drive the network with a set of square waves with uncertain delays That is, each input is driven with a waveform as shown in Fig. 12, a square wave with timings randomly dithered about 10% of the period. This is intended to mimic the situation where signals incident on the network have some correlation for example because they originate from the same functional block (mixer, oscillator, etc.) or are time-correlated, due to a common clock, but the signals themselves can be known only approximately before the reduction procedure.

Fig. 13 shows results from setting the SVD tolerance set to 10^{-3} in Algorithm 3, and extracting a 15-state reduced model. The results from the input-correlated TBR method are acceptable. For comparison, we also show the 15-state TBR model: the accuracy of this model is clearly unacceptable. For equivalent accuracy, TBR requires about a 45-state model. Note that PRIMA *matching only one moment*, would require a 32-state model. For this example, PRIMA requires at least two moments for acceptable accuracy, i.e., 64 states. A PACT model incorporating poles up to only the sinusoid frequency would have over seventy states.

Now, we demonstrate the effect of drawing inputs from *outside* the class assumed for model construction. If the inputs venture far from the distribution assumed when the model was built, accuracy will deteriorate and more states will be required in the model. To illustrate this, we reran the same example, again using square waves for inputs, but completely changing the phase relation between the inputs (as opposed to the low-level dither introduced in Fig. 13). Fig. 14 shows the results from the same 15-state models as used previously. The accuracy of the input-correlated reduction procedure degrades noticeably. Recovering accuracy requires a model of many more states, so without some degree of information about the input correlation, there is no advantage over using TBR.

2) Large Substrate Network: Finally, we consider application of the method to a real circuit (a data converter) with an

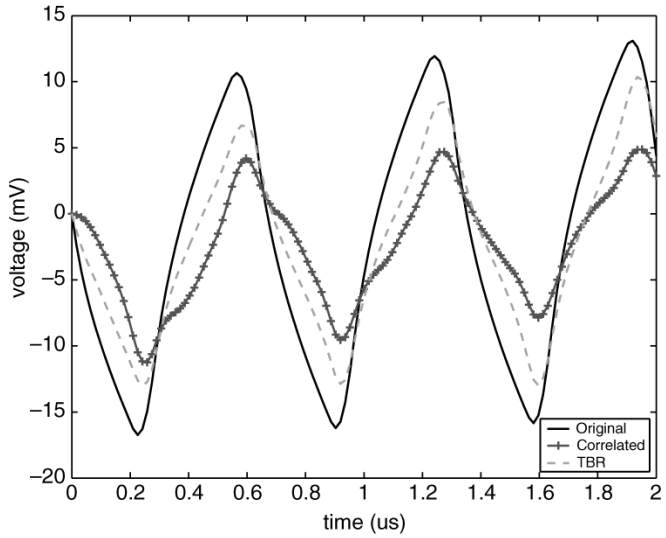

Fig. 14. Simulation results for one output on *RC* network example, with re-randomized phase relation. PMTBR with correlation information breaks down.

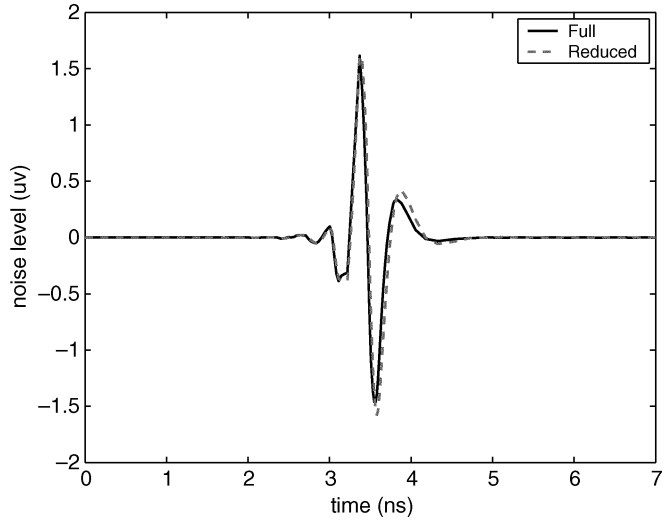

Fig. 15. Simulation results for data converter example, 150 port substrate models, full versus 4-state reduced model.

extracted substrate network. First, for purposed of assessing the actual error performance of the model reduction algorithm, we extracted only a small portion of the substrate network connection the bulk nodes of the MOS transistors. 150 ports of the substrate network were extracted using a boundary-element procedure. Both resistive and capacitive terms were retained, leading to a 150-state model. To obtain estimates of the input-signal correlations at the input, we use the MOS transistor bulk current signals from simulating the circuit *without* the substrate network as inputs to the input-correlated TBR procedure.[5] We then compared the results of simulation with the reduced model to simulation with the full model. These results are shown in Fig. 15. In this case, fair agreement with the full model was obtained using only four states, and excellent agreement obtained with eight

[5]Note that, should the substrate network result in such *large* changes to the circuit operation that these estimates were completely unrepresentative, we would have to iterate this procedure to obtain a self-consistent estimate. This would probably indicate that the circuit ceased to function as designed.

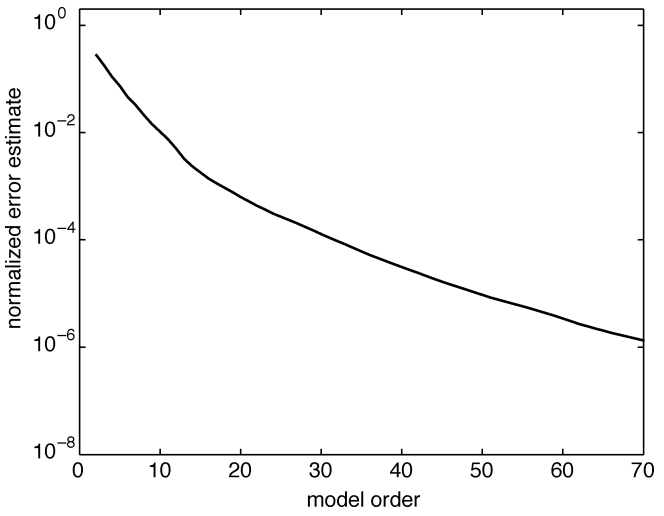

Fig. 16. Error estimate based on singular value analysis of Z-matrix from input-correlated TBR, for 1000-port substrate network with inputs from data converter example.

states. This is a 20X compression from the full model. Note that this network is, for most intents, unreducible with standard projection methods.

To illustrate the capabilities of the algorithm on larger networks, we also applied the proposed technique to a larger section of the extracted substrate network, this time comprising 1000 substrate ports. Fig. 16 shows the error estimate data obtained from the singular value analysis in Algorithm 3. In this case, a model size of 30 states is sufficient to achieve high accuracy. This represents a compression of over 30X in model size and, because of the superlinear complexity associated with factorizing dense matrix blocks, considerably more savings in time required for linear system solution in simulation.

VII. Conclusion

In this work, we discussed a connection between TBR model reduction methods and multipoint rational approximation/projection techniques. While primarily of theoretical interest, this connection leads to a potentially useful new algorithm: PMTBR. PMTBR was shown to have some advantages over existing algorithms, particularly in generating smaller reduced models, and possibly in order control and error estimation. In retrospect, the connection of TBR and PMTBR is not surprising: both the TBR procedure and the SVD used in PMTBR are principal components analyzes. TBR is a principal components analysis of the functionals defined by the state-space model, and naturally arises from time-domain theory of state-space systems. PMTBR arises naturally from a numerical approximation viewpoint of frequency domain data.

A potentially more important observation is that the existing model order reduction algorithms contain implicit assumptions about the inputs to the systems being modeled. To each set of assumptions, corresponds an implicit model of the inputs themselves. When correlated with actual information available from application domains, these input models seem unduly restrictive, implying that the assumptions implicit in the standard model order reduction schemes may be somewhat naive. Our

hope is that increased care in modeling the system inputs themselves can lead to more powerful modeling schemes.

Possible extensions of this work include integration of adaptive point selection estimation with error control, and extension of the PMTBR approach to the positive–real TBR [12] algorithms.

REFERENCES

[1] P. Feldmann and R. W. Freund, "Efficient linear circuit analysis by Padé approximation via the Lanczos process," *IEEE Trans. Computer-Aided Design Integr. Circuits Syst.*, vol. 14, no. 5, pp. 639–649, May 1995.

[2] A. Odabasioglu, M. Celik, and L. T. Pileggi, "PRIMA: passive reduced-order interconnect macromodeling algorithm," *IEEE Trans. Computer-Aided Design Integr. Circuits Syst.*, vol. 17, no. 8, pp. 645–654, Aug. 1998.

[3] ——, "Practical considerations for passive reduction of *RLC* circuits," in *Proc. Int. Conf. Computer Aided-Design*, San Jose, CA, Nov. 1999, pp. 214–219.

[4] M. Kamon, N. Marques, L. M. Silveira, and J. White, "Automatic generation of accurate circuit models of 3-D interconnect," *IEEE Trans. Compon., Packag., Manuf. Technol. B*, vol. 21, no. 3, pp. 225–240, Aug. 1998.

[5] M. Kamon, F. Wang, and J. White, "Generating nearly optimally compact models from Krylov-subspace based reduced-order models," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 47, no. 4, pp. 239–248, Apr. 2000.

[6] N. Marques, M. Kamon, J. White, and L. M. Silveira, "A mixed nodal-mesh formulation for efficient extraction and passive reduced-order modeling of 3-D interconnects," in *Proc. 35th ACM/IEEE Design Automation Conf.*, San Francisco, CA, Jun. 1998, pp. 297–302.

[7] I. M. Elfadel and D. L. Ling, "A block rational arnoldi algorithm for multipoint passive model-order reduction of multiport *RLC* networks," in *Proc. Int. Conf. Computer-Aided Design*, San Jose, CA, Nov. 1997, pp. 66–71.

[8] B. Moore, "Principal component analysis in linear systems: controllability, observability, and model reduction," *IEEE Trans. Autom. Control*, vol. AC-26, no. 1, pp. 17–32, Feb. 1981.

[9] J.-R. Li, F. Wang, and J. White, "An efficient lyapunov equation-based approach for generating reduced-order models of interconnect," in *Proc. 36th ACM/IEEE Design Automation Conf.*, New Orleans, LA, Jun. 1999, pp. 1–6.

[10] P. Heydari and M. Pedram, "Model reduction of variable-geometry interconnects using variational spectrally-weighted balanced truncation," in *Proc. Int. Conf. Computer-Aided Design*, San Jose, CA, Nov. 2001, pp. 586–591.

[11] I. M. Jaimoukha and E. M. Kasenally, "Oblique projection methods for large scale model reduction," *SIAM J. Matrix Anal. Appl.*, vol. 16, pp. 602–627, 1995.

[12] J. Phillips, L. Daniel, and L. M. Silveira, "Guaranteed passive balancing transformations for model order reduction," *IEEE Trans. Computer-Aided Design Integr. Circuits Syst.*, vol. 22, no. 8, pp. 1027–1041, Aug. 2003.

[13] L. M. Silveira, N. Marques, M. Kamon, and J. White, "Improving the efficiency of parasitic extraction and simulation of 3-D interconnect models," in *Proc. IEEE Int. Conf. Electron., Circuits, Syst.*, Pafos, Cyprus, September 1999, pp. 1729–1736.

[14] J. R. Phillips, "A statistical perspective on nonlinear model reduction," in *Proc. Behav. Modeling Simul. Workshop*, San Jose, CA, Oct. 2003, pp. 41–46.

[15] G. Wang, V. Sreeram, and W. Q. Liu, "A new frequency-weighted balanced truncation method and an error bound," *IEEE Trans. Autom. Control*, vol. 44, no. 9, pp. 1734–1737, Sep. 1999.

[16] K. Zhou, "Frequency-weighted L_∞ norm and optimal hankel norm model reduction," *IEEE Trans. Autom. Control*, vol. 40, no. 10, pp. 1687–1699, Oct. 1995.

[17] B. D. O. Anderson and Y. Liu, "Controller reduction: concepts and approaches," *IEEE Trans. Autom. Control*, vol. 34, no. 8, pp. 802–812, Aug. 1989.

[18] M. G. Safonov and R. Y. Chiang, "A Schur method for balanced truncation model reduction," *IEEE Trans. Autom. Control*, vol. 34, no. 7, pp. 729–733, Jul. 1989.

[19] K. Glover, "All optimal Hankel-norm approximations of linear multivariable systems and their $\mathbf{l}_\infty$ error bounds," *Int. J. Control*, vol. 36, pp. 1115–1193, 1984.

[20] E. Grimme, "Krylov projection methods for model reduction," Ph.D. dissertation, Coordinated-Sci. Lab., Univ. Illinois, Urbana–Champaign, 1997.

[21] G. W. Stewart and J. G. Sun, *Matrix Perturbation Theory*. Boston, MA: Academic, 1990.

[22] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1991.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[24] A. Barron and T. Cover, "Minimum complexity density estimation," *IEEE Trans. Inf. Theory*, vol. 37, no. 4, pp. 1034–1054, Jul. 1991.

[25] J. R. Phillips and L. M. Silveira, "Exploiting input information in a model reduction algorithm for massively coupled parasitic networks," in *Proc. Design Automation Conf.*, San Diego, CA, Jun 2004, pp. 385–388.

[26] K. J. Kerns and A. T. Yang, "Stable and efficient reduction of large, multiport *RC* networks by pole analysis via congruence transforms," *IEEE Trans. Computer-Aided Design Integr. Circuits Syst.*, vol. 16, no. 7, pp. 734–744, Jul. 1997.

[27] J. R. Phillips and L. M. Silveira, "Simulation approaches for strongly coupled interconnect systems," in *Proc. Int. Conf. Computer-Aided Design*, Nov. 2001, pp. 430–437.

[28] T. F. Chan, "Rank revealing QR factorizations," *Lin. Alg. Appl.*, vol. 88/89, pp. 67–82, 1987.

[29] C. H. Bischof and G. Quintana-Ortí, "Computing rank-revealing QR factorizations of dense matrices," *ACM Trans. Math. Software*, vol. 24, pp. 226–253, 1998.

[30] G. W. Stewart, "An updating algorithm for subspace tracking," *IEEE Trans. Signal Process.*, vol. 40, no. 6, pp. 1535–1541, Jun. 1992.

[31] D. C. Sorensen and A. C. Antoulas, "Projection Methods for Balanced Model Reduction," Rice Univ., Houston, TX, Tech. Rep. TR01-03, 2001.

Joel Phillips (S'91–M'97) received the S.B., M.S., and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in 1991, 1993, and 1997 respectively.

He joined Cadence Design Systems, San Jose, CA, in 1997 and is currently a Research Scientist with Cadence Berkeley Laboratories. His research interests are in the application of numerical simulation techniques to problems in electronic design automation.

Dr. Phillips received notable paper citations at DAC in 2000 and 2002, the SBCCI'02, and the 1995 Copper Mountain Conference on Multigrid Methods.

L. Miguel Silveira (S'85–M'95–SM'00) was born in Lisbon, Portugal. He received the Engineer's (*summa cum laude*) and Master's degrees in electrical and computer engineering from the Instituto Superior Técnico, Technical University of Lisbon, Lisbon, Spain, in 1986 and 1989, respectively, and the M.S., E.E., and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in 1990, 1991, and 1994, respectively.

He is currently an Associate Professor of Electrical and Computer Engineering at the Instituto Superior Técnico at the Technical University of Lisbon, a Senior Researcher at INESC ID, Systems and Computer Engineering Institute Research and Development, and a founding member of the Lisbon Center of the Cadence Laboratories. His research interests are in various aspects of computer-aided design of integrated circuits with emphasis on interconnect modeling and simulation, parallel computer algorithms, and the theoretical and practical issues concerning numerical simulation methods for circuit design problems.

Prof. Silveira is a Member of Sigma Xi.