On Symbolic Model Order Reduction

Guoyong Shi, *Member, IEEE*, Bo Hu, *Student Member, IEEE*, and C.-J. Richard Shi, *Fellow, IEEE*

***Abstract*—Symbolic model order reduction (SMOR) is a macromodeling technique that generates reduced-order models while retaining the parameters in the original models. Such symbolic reduced-order models can be repeatedly simulated with a greater efficiency for varying model parameters. Although the model-order-reduction concept has been extensively developed in literature and widely applied in a variety of problems, model order reduction from a symbolic perspective has not been well studied. Several methods developed in this paper include symbol isolation, nominal projection, and first-order approximation. These methods can be applied to models having only a few parametric elements and to models having many symbolic elements. Of special practical interest are models that have slightly varying parameters such as process related variations, for which efficient reduction procedures can be developed. Each technique proposed in this paper has been tested by circuit examples. Experiments show that the proposed methods are efficient and effective for many circuit problems.**

***Index Terms*—Arnoldi algorithm, first-order approximation, Krylov subspace, model order reduction, moment matching, Monte Carlo simulation, nominal projection, symbol isolation, symbolic model.**

I. Introduction

THE CONCEPT of model order reduction originally developed in a control theory has been gaining popularity in the electronic design community. After about ten years of research in its applications to circuit-related problems, model order reduction is becoming a standard methodology for interconnect modeling [1], [2], compact package modeling [3], and timing analysis [4] and is being extended to the general area of circuit macromodeling [5]. As an efficient means of automatic generation of compact behavioral models for both circuit blocks and structures, model order reduction makes the simulation of complex systems possible and enables system-level design and verification [6]. In particular, Krylov subspace-based model-order-reduction techniques are receiving much more attention because of their numerical efficiency and robustness [7], [8]. A comprehensive review on the application of Krylov subspace techniques to reduced-order modeling can be found in a recent survey paper [9].

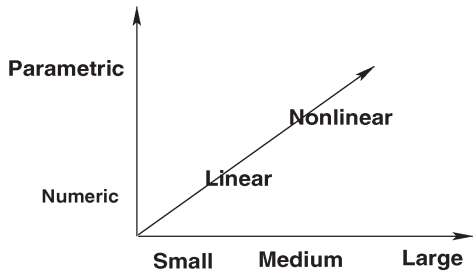


Fig. 1. Three dimensions of model order reduction.

While model-order-reduction techniques for linear models are reaching maturity, their extensions to nonlinear models and to parametric models are still underdeveloped due to their intrinsically different nature of difficulty (see the model complexity diagram in Fig. 1). In many electronic design problems, one frequently runs into linear (linearized) models involving parameters such as process variation parameters, geometric parameters, design parameters, and even artificially introduced parameters for design optimization, etc. Such models are typically found in the simulation of parasitics, interconnects, and three-dimensional (3-D) electromagnetic (EM) structures, where models are of high order and parameters are often used for design and analysis. One frequently needs to carry out a Monte Carlo simulation to investigate the parameter-dependent performances or other issues related to optimization. In such cases, reduced-order models retaining the same set of parameters as those in the original models would undoubtedly yield more efficiency than by using the full-order models directly. We formulate such type of problems in the framework of the symbolic model order reduction (SMOR), where parameters are treated as symbols.

A few researchers have attempted to address the problem of parametric model order reduction. A multivariate moment-matching technique is used in [10], where parameters are assumed to be linearly separable. A variational analysis approach is taken in [11] for resistor–capacitor–inductor (RCL) interconnect modeling with statistically varying parameters. An interpolation technique is proposed in [12] for parametric interconnect analysis. All these methods pose specific assumptions on the models, and their potential for general applications are limited. There are a few publications in the control literature that deal with parametric model order reduction, such as [13] and the references therein. Halevi *et al.* [13] propose a series expansion approach for an approximate computation of the optimal projection matrices. The projection matrices are assumed in the Taylor expansion form with respect to the parameters, and the expansion coefficient matrices are solved

Manuscript received December 29, 2004; revised March 28, 2005. This work was supported in part by the Defense Advanced Research Projects Agency (DARPA) Automated Design Tools for Integrated Mixed Signal Microsystems (NeoCAD) Program under Grant N66001-01-8920 from the Navy Space and Naval Warfare Systems Command (SPAWAR), by the Semiconductor Research Corporation (SRC) under Contract 2001-TJ-921, and by the National Science Foundation (NSF) Career Award under Grant 9985507. This paper was recommended by Associate Editor L. M. Silveira.

G. Shi was with the Department of Electrical Engineering, University of Washington, Seattle, WA 98195 USA. He is now with the School of Microelectronics, Shanghai Jiao Tong University, Shanghai 200030 China (e-mail: shiguoyong@ic.sjtu.edu.cn).

B. Hu and C.-J. R. Shi are with the Department of Electrical Engineering, University of Washington, Seattle, WA 98195 USA (e-mail: hubo@ee.washington.edu; cjshi@ee.washington.edu).

Digital Object Identifier 10.1109/TCAD.2005.855887

recursively from a sequence of matrix equations. It is well known that the cost of solving matrix equations is rather high, especially for high-order models. Hence, this method is applicable only to small-scale models and is adequate for many control system problems. Despite all these research works in literature, a generic solution for SMOR is still not available.

The primary goal of this paper is to formulate the SMOR problem and to propose a set of general-purpose solutions for certain applications. The two basic requirements in the SMOR framework are: 1) the computational procedure for constructing symbolic reduced-order models should be sufficiently efficient so that it is worthwhile to generate such models and 2) the evaluation of the reduced-order symbolic models should be sufficiently efficient as well. In other words, a practical SMOR methodology should be able to produce symbolic reduced-order models without much effort, and the produced models should be simple enough to evaluate. Other mathematical and physical properties such as accuracy, passivity, stability, etc., are naturally required as in the ordinary model-order-reduction methodologies. SMOR is, in general, a challenging problem. We hope that the preliminary results developed in this paper serve as an initial effort towards an ultimate solution for SMOR. Some partial results of this paper have been reported earlier in [14].

For a self-contained exposition, we first review in Section II some basic Krylov subspace techniques widely used in traditional model-order-reduction algorithms and also in this paper. Then, three approaches to symbolic model reduction are developed. The first method for SMOR, introduced in Section III, is the so-called symbol-isolation method, which is applicable to circuit models with a few symbolic elements so that they can be isolated by using a port formulation. Although the Krylov subspace approach to SMOR involves symbolic inversion of matrices, a computationally expensive operation, this difficulty can be overcome for certain special cases such as models with small parameter variations. For such models, we attempt to use the nominal models from which nominal orthonormalized Krylov subspaces are computed and approximate symbolic reduced-order models are then constructed. This technique is based on the observation that the nominal orthonormalized Krylov subspaces have a certain degree of robustness that warrants an approximate moment matching even when the model parameters have been perturbed slightly. This nominal-projection method is our second method for SMOR and is presented in Section IV, where two specific methods for creating robust projection matrices are discussed in detail. Since the robustness of the nominal projections is not able to tolerate large perturbations, we further develop the first-order modification to the nominal-projection method in Section V. At the price of a slightly added computation, the first-order method is able to provide much better symbolic reduced-order models while the symbolic matrix inversion is avoided. The effectiveness and the efficiency of the proposed methods are demonstrated by experiments and the experimental results are collected in Section VI. Finally, conclusions and future research issues of SMOR are described in Section VII.

II. Preliminaries

Consider a linear circuit model that can be described by the following equations:

$$C\frac{dx}{dt} + Gx = Bu \tag{1a}$$

$$y = Fx \tag{1b}$$

where $u \in \mathbb{R}^m$ is the external stimulus to the system, $B \in \mathbb{R}^{n\times m}$ is the input matrix, $x \in \mathbb{R}^n$ is the state vector, $F \in \mathbb{R}^{p\times n}$ is the output matrix, and $y \in \mathbb{R}^p$ is the output of the model. $C \in \mathbb{R}^{n\times n}$ is the susceptance matrix and $G \in \mathbb{R}^{n\times n}$ is the conductance matrix.

Congruence transformation is commonly used for the reduction of circuit models with port formulation because it preserves the passivity [8]. Let $V \in \mathbb{R}^{n\times q}$ be the transformation matrix with $q \ll n$. By defining

$$C_r = V^{\mathrm{T}}CV,\ \ G_r = V^{\mathrm{T}}GV,\ \ B_r = V^{\mathrm{T}}B,\ \ F_r = FV \tag{2}$$

where the superscript $^{\mathrm{T}}$ indicates transpose of a matrix or a vector and restricting the state x to the subspace spanned by the columns of V, i.e., $x = Vz$ for some $z \in \mathbb{R}^q$, the reduced-order model of order q can be written as

$$C_r\frac{dz}{dt} + G_r z = B_r u \tag{3a}$$

$$\tilde{y} = F_r z \tag{3b}$$

where $z \in \mathbb{R}^q$ becomes the new state vector of the reduced-order model.

One popular method for generating the transformation matrix V is by moment matching. Let $X(s) = T(s)U(s)$ be the Laplace transform of the state space model, where $T(s) = (Cs + G)^{-1}B$. For moment matching, we expand $T(s)$ in Taylor expansion at $s = 0$, i.e.,

$$\begin{aligned} T(s) &= (Cs+G)^{-1}B \\ &= \sum_{i=0}^{\infty}(-G^{-1}C)^i G^{-1}Bs^i \end{aligned} \tag{4}$$

where the coefficients of s^i are called the moments. Moment matching is directly connected to the Krylov subspace formed by the pair of matrices $(G^{-1}C, G^{-1}B)$ [7]. The Krylov subspace is spanned by the column vectors in the following collection of matrices:

$$\left\{G^{-1}B, (G^{-1}C)G^{-1}B, \ldots, (G^{-1}C)^i G^{-1}B, \ldots\right\} \tag{5}$$

where the column vectors are called the Krylov vectors. The qth-order Krylov subspace is denoted by

$$\mathcal{K}_q(G^{-1}C, G^{-1}B) \tag{6}$$

which is spanned by the leading q linearly independent Krylov vectors in (5).

Let $V \in \mathbb{R}^{n\times q}$ be any matrix whose columns span the Krylov subspace $\mathcal{K}_q(G^{-1}C, G^{-1}B)$. If the columns of V are

orthonormalized and B is a column vector, it can be shown that the following identities hold [8]:

$$(G^{-1}C)^i G^{-1}B = V\left(G_r^{-1}C_r\right)^i G_r^{-1}B_r \tag{7}$$

for $i = 0, 1, \ldots, q-1$. These identities can be used to verify that at least the q leading moments of the full-order and reduced-order transfer functions are matched [15].

As a common practice, the block vectors forming the Krylov subspace are orthonormalized by using the Arnoldi algorithm for a numerical stability. A block Arnoldi algorithm for a multi-column B is described in [8]. Since we shall be using the Arnoldi algorithm in SMOR algorithms, we describe the following algorithm for a single-column input matrix, i.e., $B = b$, where $b \in \mathbb{R}^n$, so that the algebraic operations involved can be seen.

Arnoldi Algorithm

i) LU factorize matrix G: $G = LU$.
ii) Solve $\tilde{v}_1$ from: $G\tilde{v}_1 = b$.
iii) Compute $h_{11} = \|\tilde{v}_1\|$ and $v_1 = \tilde{v}_1/h_{11}$.
iv) For $j = 2, \ldots, q$:
Solve $\tilde{v}_j$ from: $G\tilde{v}_j = Cv_{j-1}$.
For $i = 1, \ldots, j-1$: $h_{ij} = v_i^{\mathrm{T}}\tilde{v}_j$.
$w_j = \tilde{v}_j - \sum_{i=1}^{j-1} v_i h_{ij}$
$h_{jj} = \|w_j\|$, $v_j = w_j/h_{jj}$.

Note that the Arnoldi algorithm terminates when $h_{jj} = 0$, which means that the subsequent vectors belong to the subspace already generated. The Arnoldi algorithm is basically a Gram–Schmidt procedure for orthonormalizing the Krylov vectors.

There are two fundamental operations involved in the Arnoldi algorithm that cannot be extended easily to the symbolic context. One is the inversion of matrix G, which is numerically implemented by the LU factorization, and the other is the orthonormalization process. Symbolically inverting a high-dimensional sparse matrix G is computationally not feasible for most realistic models due to the exhaustive memory usage and the exponential number of terms. Even if a symbolic G^{-1} is available, the repeated multiplication by the symbolic matrix $(G^{-1}C)$ for generating and orthonormalizing the symbolic Krylov subspace in (5) immediately becomes unwieldy. The methods to be developed are targeted at developing computationally tractable SMOR while avoiding highly expensive symbolic manipulation.

III. Symbol-Isolation Method

In this section, we introduce the first SMOR method called symbol isolation. This method is motivated by occasions in a circuit simulation where a large network has to be simulated by a sweeping analysis over a few critical elements. In this case, the simulation efficiency would be improved greatly if the large network excluding those critical elements is replaced by a compact reduced-order model.

The underlying idea of symbol isolation is rather straightforward: Isolate those symbolic elements and treat them as ports (see the illustration in Fig. 2), then reduce the rest of the network by a standard model-reduction algorithm. Most model-order-reduction algorithms in the state space do not change the input/output port structure. Hence, the isolated symbolic elements can be incorporated again after the subnetwork is reduced.

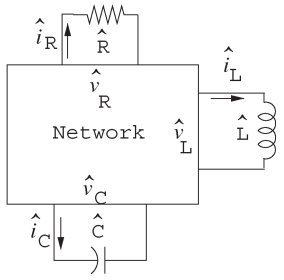


Fig. 2. Isolation of symbolic elements from a network.

Although the partition idea used in symbol isolation is rather straightforward, we would like to formalize it to see that what the symbol-isolation method does is actually equivalent to a block congruence transform.

We consider only a network with a single resistor (R), a single inductor (L), and a single capacitor (C) isolated for a symbolic analysis (see Fig. 2). Each two-terminal element is treated as a port connected to the main network indicated by the rectangular block. Extending the following formulation to multiple symbolic RCL elements should be straightforward.

To facilitate the modified nodal analysis (MNA) formulation [16], we introduce some notations for the network model as shown in Fig. 2. Let x be the state of the network inside the block, including the nodal voltages and the necessary currents inside the block. Augmented with the variables for the isolated symbolic elements, the full state vector becomes

$$x_f^{\mathrm{T}} = [x^{\mathrm{T}}, \hat{i}_C, \hat{v}_C, \hat{i}_L, \hat{i}_R]^{\mathrm{T}} \tag{8}$$

where $\hat{i}_R$, $\hat{i}_L$, and $\hat{i}_C$ are currents through the isolated elements R, L, and C, respectively, and $\hat{v}_C$ is the voltage across the isolated capacitor C. For easy identification, we use the variables marked by hat to indicate the symbolic elements and the state variables associated with them. The model equations now become

$$\left[\begin{array}{cc|ccc} C & & & & \\ & 0 & & & \\ \hline & & \hat{C} & & \\ & & & \hat{L} & \\ & & & & 0 \end{array}\right] \frac{d}{dt} \left[\begin{array}{c} x \\ \hat{i}_C \\ \hline \hat{v}_C \\ \hat{i}_L \\ \hat{i}_R \end{array}\right]$$

$$= -\left[\begin{array}{cc|ccc} G & E_C & 0 & E_L & E_R \\ -E_C^{\mathrm{T}} & 0 & 1 & 0 & 0 \\ \hline 0 & -1 & 0 & 0 & 0 \\ -E_L^{\mathrm{T}} & 0 & 0 & 0 & 0 \\ -E_R^{\mathrm{T}} & 0 & 0 & 0 & \hat{R} \end{array}\right] \times \left[\begin{array}{c} x \\ \hat{i}_C \\ \hline \hat{v}_C \\ \hat{i}_L \\ \hat{i}_R \end{array}\right] + \left[\begin{array}{c} B \\ \hline 0 \\ 0 \\ 0 \\ 0 \end{array}\right] u \tag{9a}$$

$$y = [F \quad 0 \,|\, 0 \quad 0 \quad 0] x_f \tag{9b}$$

where $\hat{R}$, $\hat{L}$, $\hat{C}$ are the resistance, inductance, and capacitance of the symbolic elements. In (9), E_R, E_L, and E_C are, respectively, three column vectors containing all zeros but 1 and -1 at the two locations corresponding to the port node indices with respect to the isolated R, L, and C. Here, we assume that the variables with hat are not part of the output. In model (9), the symbolic elements in the coefficient matrices have been isolated to the trailing part of the state model by ordering the state variables appropriately. Note that, although $\hat{i}_C$ is a state variable associated with the symbolic $\hat{C}$, it is grouped with the nonsymbolic state variables of the model.

We rewrite the model of (9) in the block form as

$$\begin{bmatrix} \bar{C} & \\ & \hat{D} \end{bmatrix} \frac{d}{dt} \begin{bmatrix} \bar{x} \\ \hat{x} \end{bmatrix} = - \begin{bmatrix} \bar{G} & E \\ -E^{\mathrm{T}} & \hat{J} \end{bmatrix} \begin{bmatrix} \bar{x} \\ \hat{x} \end{bmatrix} + \begin{bmatrix} \bar{B} \\ 0 \end{bmatrix} u \quad (10a)$$

$$y = [\, \bar{F} \quad 0 \,] \begin{bmatrix} \bar{x} \\ \hat{x} \end{bmatrix} \quad (10b)$$

where

$$\bar{x} = \begin{bmatrix} x \\ \hat{i}_C \end{bmatrix}, \qquad \hat{x} = \begin{bmatrix} \hat{v}_C \\ \hat{i}_L \\ \hat{i}_R \end{bmatrix}$$

and the submatrices can be identified easily. After the partition, the isolated symbols only appear in the equations for $d\hat{x}/dt$, which is of order 3. The nonsymbolic part of the model is extracted as

$$\bar{C} \frac{d\bar{x}}{dt} = - \bar{G}\bar{x} + \tilde{B}\tilde{u} \quad (11a)$$

$$y = \bar{F}\bar{x} \quad (11b)$$

where

$$\tilde{B} = [\, -E \quad \bar{B} \,], \qquad \tilde{u} = \begin{bmatrix} \hat{x} \\ u \end{bmatrix}$$

are the augmented input matrix and input vector, respectively.

The model in (11) is still a large network but without symbols. The input dimension of this model has been augmented by the symbolic state variable $\hat{x}$. This model can be reduced by the standard reduction methods using the block Arnoldi algorithm or the block Lanczos algorithm such as that used for Pade-via-Lanczos (PVL) and passive reduced-order interconnect macromodeling algorithm (PRIMA) [7], [8]. Suppose that we have obtained a projection matrix $V \in \mathbb{R}^{n_1 \times q}$ using the matrix triple $(\bar{C}, \bar{G}, \tilde{B})$, where n_1 is the model order of (11) and q is the reduced model order. Let $\bar{x} = Vz$ and premultiply the first equation in (11) by V^{T}. We obtain the following reduced-order model from (11):

$$\bar{C}_r \frac{dz}{dt} = - \bar{G}_r z + \tilde{B}_r \bar{u} \quad (12a)$$

$$y = \bar{F}_r z \quad (12b)$$

where

$$\bar{C}_r = V^{\mathrm{T}}\bar{C}V, \quad \bar{G}_r = V^{\mathrm{T}}\bar{G}V, \quad \tilde{B}_r = V^{\mathrm{T}}\tilde{B}, \quad \bar{F}_r = \bar{F}V. \quad (13)$$

Note that after the reduction, the network inside the block in Fig. 2 is replaced by a smaller sized model while the port structure is unaltered. As a result, we obtain a reduced-order model of the original network while retaining the isolated symbolic elements.

It is interesting to note that given the block structure of model (9), the reduction procedure described above is in fact equivalent to a block transformation. Using the notation introduced above, the reduced symbolic model is

$$\begin{bmatrix} V^{\mathrm{T}}\bar{C}V & \\ & \hat{D} \end{bmatrix} \frac{d}{dt} \begin{bmatrix} z \\ \hat{x} \end{bmatrix} = - \begin{bmatrix} V^{\mathrm{T}}\bar{G}V & V^{\mathrm{T}}E \\ -E^{\mathrm{T}}V & \hat{J} \end{bmatrix} \begin{bmatrix} z \\ \hat{x} \end{bmatrix} + \begin{bmatrix} V^{\mathrm{T}}\bar{B} \\ 0 \end{bmatrix} u \quad (14a)$$

$$y = [\bar{F}V \quad 0] \begin{bmatrix} z \\ \hat{x} \end{bmatrix}. \quad (14b)$$

This means that this reduced-order model is actually obtained from the block model (10) by applying the block congruence transformation $\begin{bmatrix} V & \\ & I_3 \end{bmatrix}$.

Although extending the symbol-isolation procedure above to models involving multiple symbolic RCL elements is straightforward, as the number of ports increases, the column dimension of the input matrix in model (11) increases as well. Recall that the Krylov subspace has a dimension of at least the number of columns of B (assumed column full rank). If the column dimension of B is large, then not only the model cannot be reduced to a very low order using the conventional Krylov subspace approach, but also the computational cost becomes high. For this reason, the symbol-isolation method is useful only for a few symbols, regardless of the element type.

The computational cost for generating a reduced-order symbolic model using the symbol-isolation method is mainly due to the reduction of the nonsymbolic block with an increased number of input ports. Suppose that the original model has m inputs, and we consider n_s symbolic elements connected as extra ports. Then, according to the formulation above, we have to reduce a nonsymbolic model of order $n - n_s$ with $m + n_s$ inputs, where n is the full model order. Using the block Arnoldi or Lanczos algorithm is the most efficient and numerically stable way for reducing nonsymbolic linear models. Its computational cost consists of one LU factorization and a sequence of matrix-vector multiplications, for which there are finely tuned dedicated packages taking the full advantage of a matrix sparsity [7]. The computational procedure can be further optimized for the case of multiple ports [17].

IV. Nominal-Projection Method

Parametric models frequently appear in very large scale integrated (VLSI) circuit design and simulation. For example, RC(L) models for interconnect analysis can have R, C, or L as parameters because of the uncertainty in the process. Geometric parameters used as design parameters can also be introduced in the models. For many parametric models, directly treating parametric elements as symbolic ones as we did in the symbol-isolation method would not be efficient, because many elements could depend on only a few parameters and the large number of ports arising from symbol isolation is a problem.

If one would like to use the projection method for SMOR, one might have to construct symbolic Krylov subspaces. However, due to the computational complexity involved, direct construction of symbolic Krylov subspace may not be viable. However, we note that very often in practice the model parameters do not change drastically; hence, the parameter variations can be treated as small perturbations. With this observation, we propose the second method for SMOR called nominal-projection method. The basic idea is to compute a sufficiently robust nominal Krylov subspace so that the same subspace can be used for models with slightly perturbed parameters. Meanwhile, we shall investigate issues on the computation of robust Krylov subspaces and the effectiveness by using nominal projections. The efficiency should be obvious because of the reuse of the nominal projections.

Formally, parametric linear time-invariant models can be described by

$$C(p)\frac{dx}{dt} + G(p)x = B(p)u \tag{15a}$$

$$y = F(p)x \tag{15b}$$

where $C(p)$, $G(p)$, $B(p)$, and $F(p)$ are model matrices depending on parameter vector p, which may contain a number of parameters. With the projection matrix $V(p_0)$ computed from a set of nominal parameters p_0 (e.g., using the Arnoldi algorithm), the reduced parametric matrices are written as

$$\begin{aligned} C_r(p) &= V(p_0)^{\mathrm{T}} C(p) V(p_0) \\ G_r(p) &= V(p_0)^{\mathrm{T}} G(p) V(p_0) \\ B_r(p) &= V(p_0)^{\mathrm{T}} B(p) \\ F_r(p) &= F(p) V(p_0). \end{aligned} \tag{16}$$

In the following, we discuss two methods for computing the nominal-projection matrix $V(p_0)$, which are believed to be more robust but computationally efficient.

We assume that all the model parameters only perturb around some nominal values in a certain design task. The ranges of parameter perturbation are problem dependent. Since parameter perturbation can be viewed as a model uncertainty, reduction to a lower dimensional state space generally suppresses such uncertainty, which is a well-known fact in statistical analysis literature [18].

We call the model with all parameters fixed to their nominal values the nominal model, from which nominal projections are computed. A nominal-projection matrix is required to be robust enough to tolerate the model perturbation. Two computational methods are introduced below for robust nominal-projection construction.

A. Mixed Moment Matching

The first method for nominal-projection computation is to combine the moments from both the zero and the infinity frequency points. More specifically, we consider the nominal model as in (15) with $p = p_0$. Let

$$C_0 = C(p_0), \qquad G_0 = G(p_0), \qquad B_0 = B(p_0).$$

The Krylov subspace for the moment matching at $s = 0$ (DC), i.e., expanding the transfer function $T(s)$ in terms of s^i, is

$$\mathcal{K}_q\left(G_0^{-1}C_0, G_0^{-1}B_0\right). \tag{17}$$

Similarly, the Krylov subspace for the moment matching at $s = \infty$, i.e., expanding the transfer function $T(s)$ in terms of $1/s^i$, is

$$\mathcal{K}_p\left(C_0^{-1}G_0, C_0^{-1}B_0\right). \tag{18}$$

Here, we assume that matrices C_0 and G_0 are both nonsingular.

The nominal-projection matrix $V(p_0)$ is then constructed from the Krylov vectors partially from (17) and (18). More specifically, we choose the leading q_1 Krylov vectors from (17) and the leading q_2 vectors from (18) to form $q = q_1 + q_2$ vectors for a nominal projection. Since the Krylov subspace formed by the low- and high-frequency moments captures both the steady-state and the transient behaviors of the time-domain response, better robustness is expected than a Krylov subspace formed solely by either low- or high-frequency components. We shall demonstrate the better robustness of the mixed-moment-matching method by an example in Section VI.

B. Real Rational Krylov Subspace

Since model-order-reduction has been widely used in interconnect design and analysis, a few comments on moment matching are worthwhile. As the operating frequency is now entering the gigahertz scale, the conventional RC models for interconnect is becoming inadequate and the inductance effect of interconnect must be addressed explicitly [19]. Since an RC circuit does not have resonance, moment matching at the a low frequency can sufficiently capture the frequency response, meaning that very compact reduced-order models can be obtained by moment matching. However, for RCL models with a low loss, capturing the resonance behavior (multiple peaks in the frequency response) at the high-frequency band usually requires a high-dimensional Krylov subspace. Although the moment-matching technique proposed above by matching both the low- and high-frequency bands can possibly produce a compact model, we found in our experiment that this method could not help improve the frequency-response accuracy in the band of resonance. Moreover, the Krylov subspace formed at $s = 0$ often resulted in unstable or singular reduced-order models for highly inductive circuits.

To solve this problem, we propose the second method for nominal-projection computation, where we use real rational Krylov subspace. A real rational Krylov subspace originates from the expansion of the transfer function $T(s) = (Cs + G)^{-1}B$ at a real point σ, i.e.,

$$\begin{aligned} T(s) &= (Cs + G)^{-1}B \\ &= \left[C(s - \sigma) + (C\sigma + G)\right]^{-1} B \\ &= \sum_{i=0}^{\infty} \left[-(C\sigma + G)^{-1}C\right]^i (C\sigma + G)^{-1}B(s - \sigma)^i. \end{aligned} \tag{19}$$

Similar to the expansion at $s = 0$, if we would like to have a reduced-order model that matches the leading q moments in this expansion, i.e., the coefficients of the terms $(s-\sigma)^i$ for $i = 0, \ldots, q-1$, we use the Krylov subspace

$$\mathcal{K}_q\left((C\sigma+G)^{-1}C, (C\sigma+G)^{-1}B\right). \tag{20}$$

By taking $\sigma = 0$, this Krylov subspace (20) reduces to the one in (17). Here, we assume that a real σ has been chosen so that $(C\sigma+G)^{-1}$ exists.

The proper choice of σ has been addressed in many papers. Grimme [15] extensively studied the rational Krylov subspace approach to model order reduction. Chiprout and Nakhla [20] used a complex σ for localized moment matching in the frequency domain. Shi and Shi further provided a new interpretation of the real σ from the waveform matching perspective [12] and the dominant subspace computation perspective [21]. The latter new developments have provided us substantial evidence to use the real rational Krylov subspaces for robust nominal subspace computation.

An example used in the experimental section will demonstrate that the real Krylov subspace has sufficient robustness for reducing inductive circuit models. We have observed that as the parameters are perturbed, the reduced-order model obtained by using the nominal projection computed from real rational Krylov subspace is able to capture the perturbed resonance modes very well.

In practice, the use of the nominal-projection method involves storing the original full-order model and the projection matrix constructed from the original full model. When the parameters in the full-order model are changed, the new reduced-order model can be constructed by using an efficient matrix-vector multiplication to compute the reduced coefficient matrices in (16). If the parameter perturbation is too far away from the nominal value, then the new nominal projection should be recomputed.

V. First-Order-Approximation Method

In this section, we propose the third method for SMOR called the first-order approximation. The basic idea is to update the first-order terms of the Krylov vectors when the model data are perturbed. The Krylov subspace created this way is partially symbolic.

As seen from the Arnoldi algorithm, the main algebraic operations involved in the Krylov subspace computation are matrix inversion and matrix-vector multiplication. Since direct inversion of a symbolic matrix is computationally costly, we take a partial symbolic approach to symbolic matrix inversion under the condition that parameter variations are small.

Let the matrices G and C be perturbed as follows:

$$G' = G_0 + \Delta G, \qquad C' = C_0 + \Delta C \tag{21}$$

where G_0 and C_0 are the matrices with the nominal values. Under the assumption that ΔG is small in a certain matrix norm, the inverse of G' can be approximated by the first-order expression

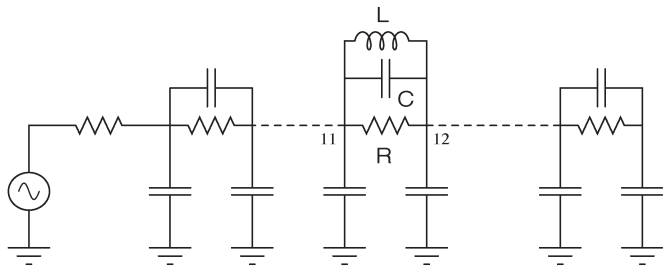


Fig. 3. Test circuit for symbol isolation.

$$(G_0+\Delta G)^{-1} \approx G_0^{-1} - G_0^{-1}(\Delta G)G_0^{-1}. \tag{22}$$

Since the nominal matrix G_0 is numeric and remains constant for small perturbations, its inverse has to be computed only once in advance. The parameter variations are symbolically characterized by the variation matrix ΔG. Clearly, the symbolic inversion in the form of (22) is greatly simplified.

Given the perturbation matrices ΔG and ΔC, the exact basis vectors of the Krylov subspace are computed from

$$\left[(G_0+\Delta G)^{-1}(C_0+\Delta C)\right]^k (G_0+\Delta G)^{-1}B \tag{23}$$

for $k = 0, 1, \ldots, q-1$. Here we assume that the B matrix is not perturbed. Again, for reducing the computational complexity, each expression in (23) can be approximated by keeping the terms up to the first order of ΔG and ΔC. It can be verified by induction that by the first-order approximation

$$\begin{aligned}&\left[(G_0+\Delta G)^{-1}(C_0+\Delta C)\right]^k \\ &\quad\approx (A_0)^k - \sum_{i=1}^{k-1}(A_0)^i(G^{-1}\Delta G)(A_0)^{k-i} \\ &\qquad + \sum_{j=0}^{k}(A_0)^j(G^{-1}\Delta C)(A_0)^{k-j}\end{aligned} \tag{24}$$

where $A_0 = G_0^{-1}C_0$. Furthermore

$$\begin{aligned}&\left[(G_0+\Delta G)^{-1}(C_0+\Delta C)\right]^k (G_0+\Delta G)^{-1}B \\ &\quad\approx \left\{A_0^k - \sum_{i=1}^{k-1} A_0^i\left(G_0^{-1}\Delta G\right)A_0^{k-i} + \sum_{j=0}^{k} A_0^j\left(G_0^{-1}\Delta C\right)A_0^{k-j}\right\} \\ &\qquad \times \left\{G_0^{-1} - G_0^{-1}(\Delta G)G_0^{-1}\right\}B \\ &\quad\approx A_0^kG_0^{-1}B - \left\{A_0^kG_0^{-1}(\Delta G) - \sum_{i=1}^{k-1} A_0^i\left(G_0^{-1}\Delta G\right)A_0^{k-i}\right. \\ &\qquad \left. + \sum_{j=0}^{k} A_0^j\left(G_0^{-1}\Delta C\right)A_0^{k-j}\right\}G_0^{-1}B.\end{aligned} \tag{25}$$

Note that in the last expression the term $A_0^kG_0^{-1}B$ is the matrix used for nominal Krylov vectors, while the rest three terms involving ΔG and ΔC are the symbolic first-order correction terms. Since the symbols only appear in the perturbation

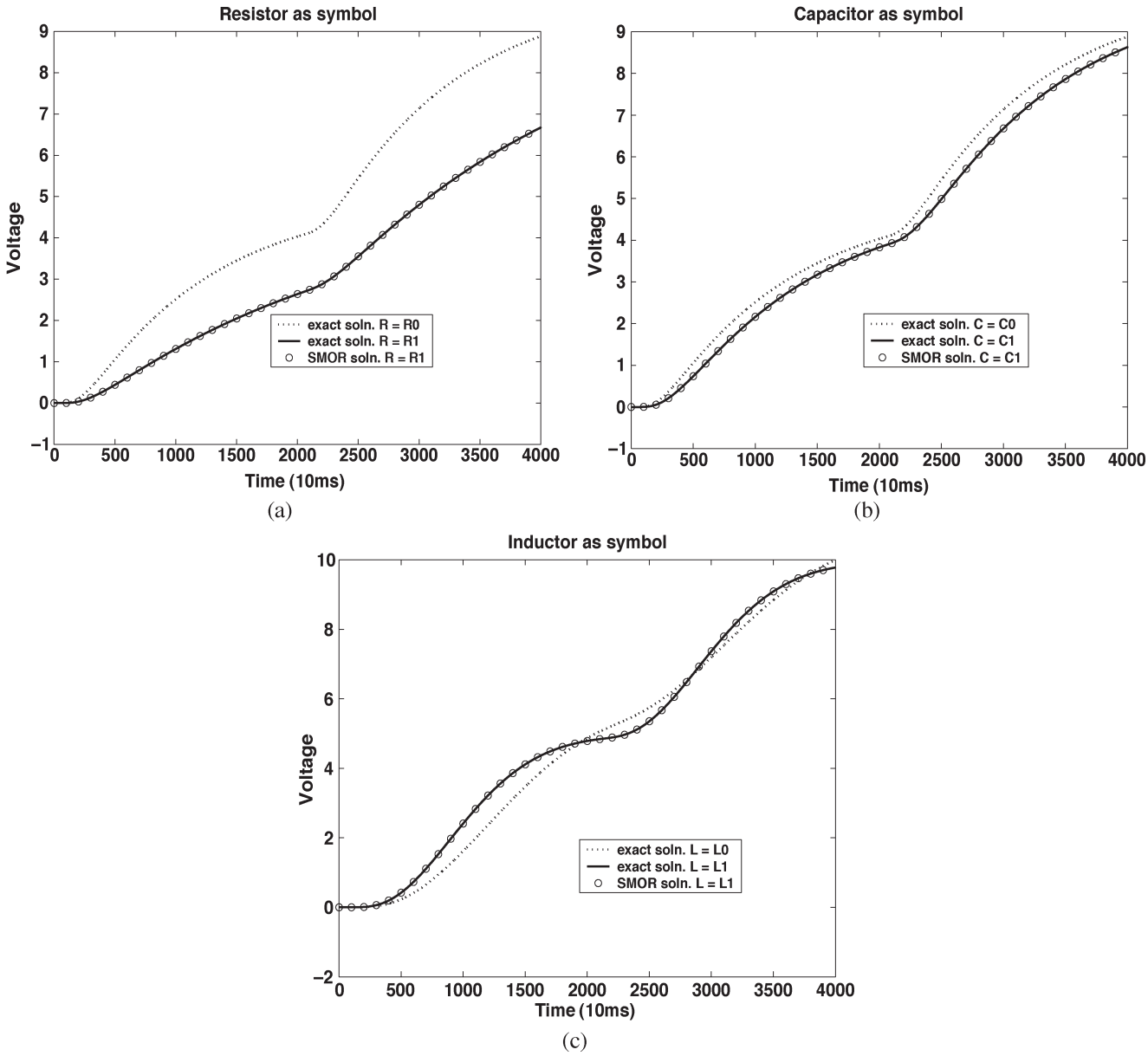


Fig. 4. Transient responses with a symbolic element between nodes 11 and 12. (a) Resistor R. (b) Capacitor C. (c) Inductor L.

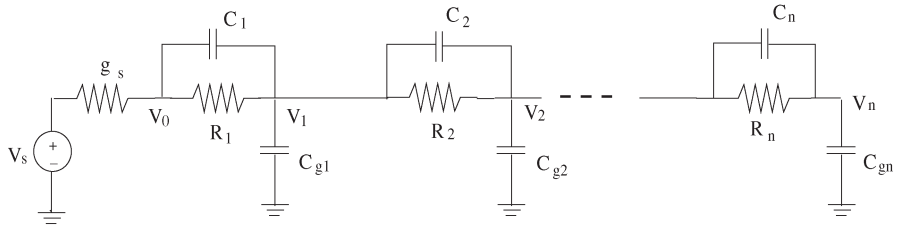


Fig. 5. RC ladder circuit.

matrices ΔG and ΔC, all other numeric matrices can be computed *a priori*. Without performing direct symbolic matrix inversion, it is now possible to obtain an approximate symbolic representation of the Krylov basis vectors.

Note that, although the symbolic procedure outlined above is feasible for a set of symbolic Krylov basis vectors, the orthonormalization of these vectors is nontrivial because, unlike the numeric case, the memory required for symbolic orthonormalization is high. To reduce the computational complexity, the symbolic basis vectors are not orthonormalized in our implementation. However, one should be aware that nonorthonormalized projection matrices may yield singular reduced-order models due to potentially bad numerical conditioning.

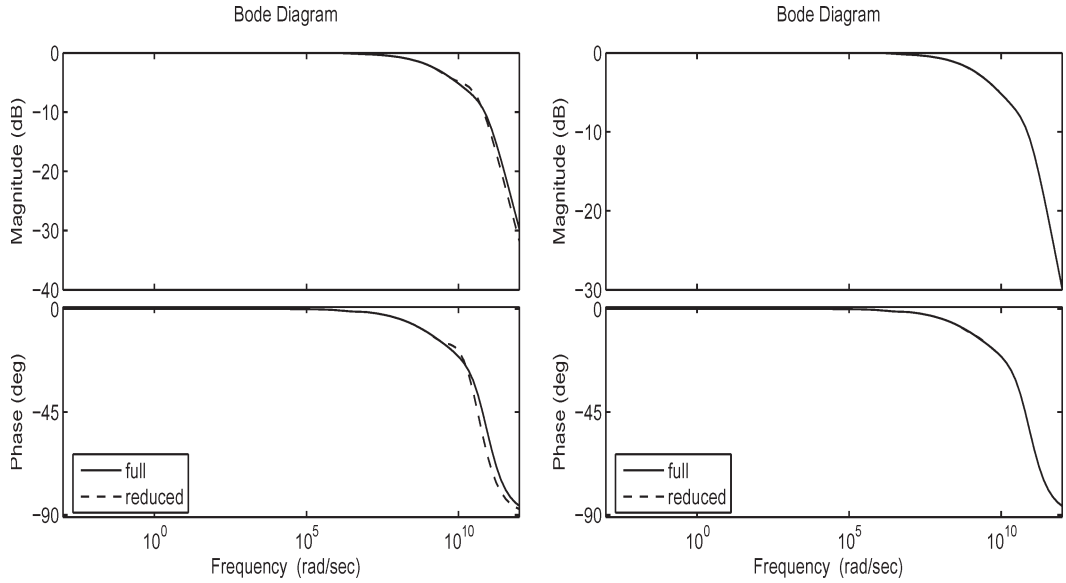


Fig. 6. Comparison of the reduction effect in the frequency domain. (a) By moment matching at $s = 0$. (b) By mixed moment matching at both low and high frequencies (note the two curves overlap).

The memory required for storing the reduced symbolic matrices

$$C_r = V^{\mathrm{T}}(C_0 + \Delta C)V, \qquad G_r = V^{\mathrm{T}}(G_0 + \Delta G)V \quad (26)$$

is roughly estimated as follows. First, it is worth noting that the parameters in the original problem have been recombined in the model and represented by the entries in the two perturbation matrices ΔG and ΔC. Because of the linear approximation, each entry of the symbolic Krylov vectors in (25) is a linear combination of the entries of ΔG and ΔC, plus a constant term. The matrices G and C are sparse in most applications. Hence, we assume that there is a total number of N_s symbols from ΔG and ΔC, where $N_s \ll n^2$. This means that each entry of the projection matrix V has the form of

$$v_{ij} = \rho_{ij} + \sum_{\ell=1}^{N_s} \rho_{ij}^{(\ell)} \Pi_\ell \quad (27)$$

where ρ_{ij}'s and $\rho_{ij}^{(\ell)}$'s are constants and Π_ℓ's are the symbols. Consequently, the memory required for storing C_r and G_r in (26) is approximately $O(N_s^3 q^2)$, where N_s is the total number of symbols and q is the order of the reduced model. For small N_s and q, the memory requirement is low.

The computational cost for generating a symbolic reduced-order model using the first-order model is mostly for computing the first-order symbolic Krylov vectors, that is, for computing all the coefficients $\rho_{ij}^{(\ell)}$ in (27) for the symbolic projection matrix V. This part of the time complexity depends on N_s. Assuming that the cost for computing each coefficient takes a constant time, the time complexity for computing V is about $O(N_s q n)$, where N_s is the number of symbols in ΔG and ΔC; q is the number of Krylov vectors, i.e., the reduced model order; and n is the original model order.

VI. Experimental Results

We have described three methods for SMOR in the preceding sections: the symbol isolation, the nominal projection, and the first-order approximation. Collected in this section are experimental results for testing these methods. At the end of this section, we comment on the comparison of these three methods, their advantages, applicable cases, efficiencies, etc.

A. Symbol Isolation

The RC ladder circuit shown in Fig. 3 is used to test the symbol-isolation method. The circuit has 100 stages with the element (either R, C, or L) between the nodes 11 and 12 used as a symbolic element. The full-order model is reduced to a model of the 12th order. We simulated three cases by changing the value R, C, or L separately. The transient responses to a two-level input voltage stimulus for the nominal value $R_0 = 1.7$ kΩ and a new value $R_1 = 10R_0$ are shown in Fig. 4(a). The dashed curve indicates the nominal response of the full-order model, while the dotted solid curves (overlapped) are the responses of the full-order and the reduced-order models for the new resistance value R_1. Other transient responses with the symbolic element replaced by C or L are shown in Fig. 4(b) and (c), respectively, where C is changed from the nominal value of $C_0 = 5$ μF to a new value $C_1 = 10C_0$ and L is changed from the nominal value of $L_0 = 400$ kH to $L_1 = 2L_0$. As expected, the symbol-isolation method can generate models with a reliable time-domain response.

The symbol-isolation method is very useful for the sweeping analysis of a few critical circuit elements. Here, we use the same circuit in Fig. 3 to demonstrate the efficiency achievable by reducing the nonsymbolic part of the model. For a full-order model of size 103, the total sweeping time for analyzing 300 values of the resistance R between the nodes 11 and 12 took 54 s by simulating the full-order model directly, with each

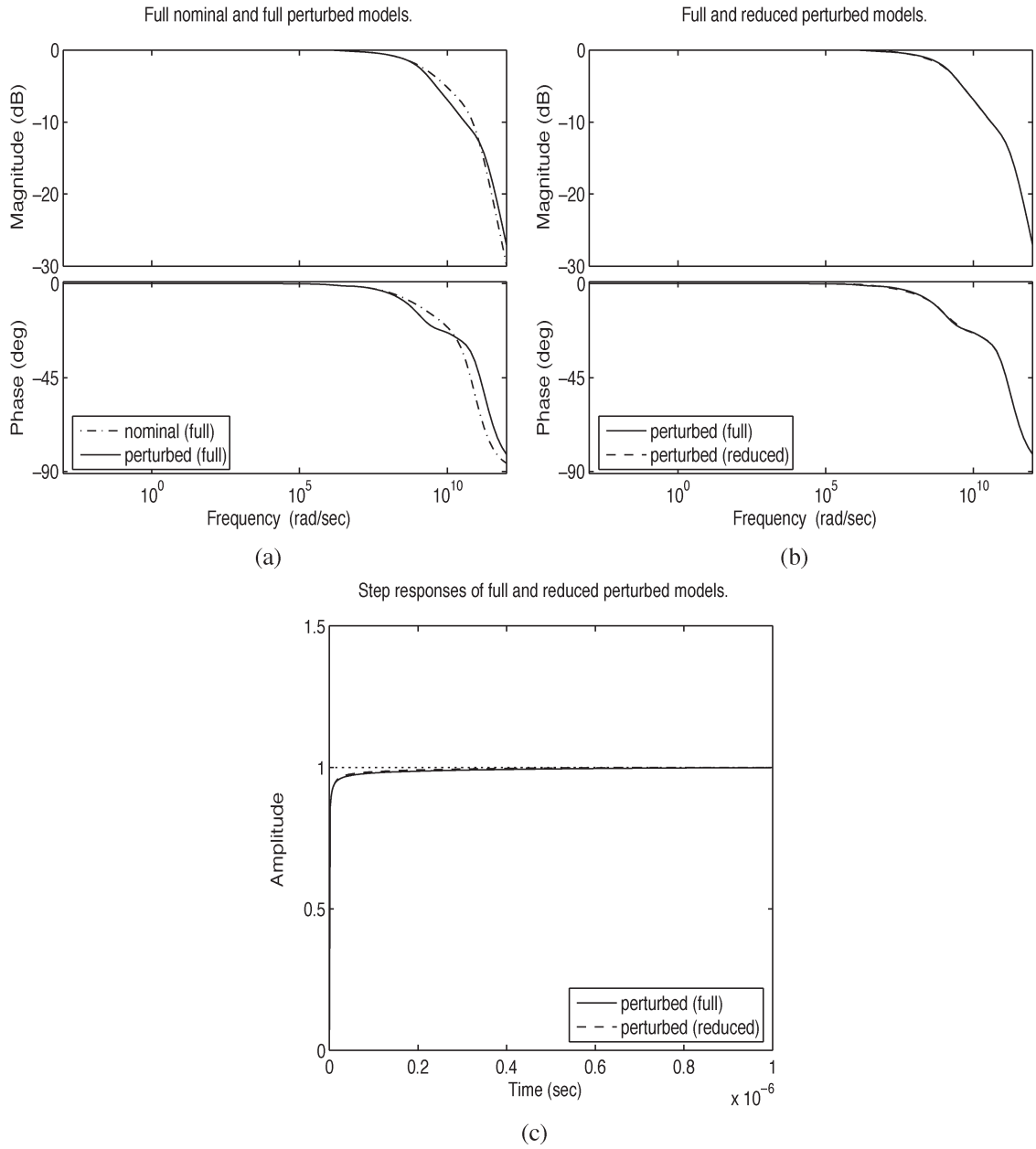


Fig. 7. Normal projection by mixed moment matching. (a) Bode plots of the full nominal and full perturbed models with perturbation up to 60%. (b) Bode plots of the full perturbed model and its reduced-order model reduced by the nominal projection. (c) Step responses of the full perturbed model and its reduced-order model (solid: full order; dashed: reduced order).

sweeping point meaning a transient simulation from 0 to 40 s. After reducing the nonsymbolic block to the 17th order, it took only 8 s to finish the same set of sweeping analysis.

B. Nominal Projection

The RC ladder circuit in Fig. 5 is used for testing the nominal-projection method. V_s is the input voltage and V_1 is the output voltage. For a demonstration purpose, we choose all $R_i = 20\ \Omega$, all $C_i = C_{gi} = 1$ pF, and $g_s = 1/50$ S for nominal values. These values are perturbed up to certain level to test the robustness of the nominal projection.

In this experiment, the full-order model is of the 200th order and will be reduced to the 10th order. There are many ways to generate a 200×10 nominal-projection matrix V. According to our discussion earlier, we prefer to use mixed moment matching. But before we proceed, we first demonstrate that using the Krylov vectors for matching the ten leading moments at $s = 0$ is not as good as using the Krylov vectors for mixed moment matching, i.e., matching five leading moments at $s = 0$ and another five leading moments at $s = \infty$.

For the RC ladder example that we have, we run two nominal reductions by using two different projection matrices mentioned above without considering the perturbation for now. The Bode plots shown in Fig. 6 clearly indicate that the mixed moment-matching method [Fig. 6(b)] gives the better frequency response matching at all frequencies of interest, while the result from moment matching only at $s = 0$ does not guarantee a good matching at the high-frequency part. We will be using the nominal-projection method later on for measuring the signal

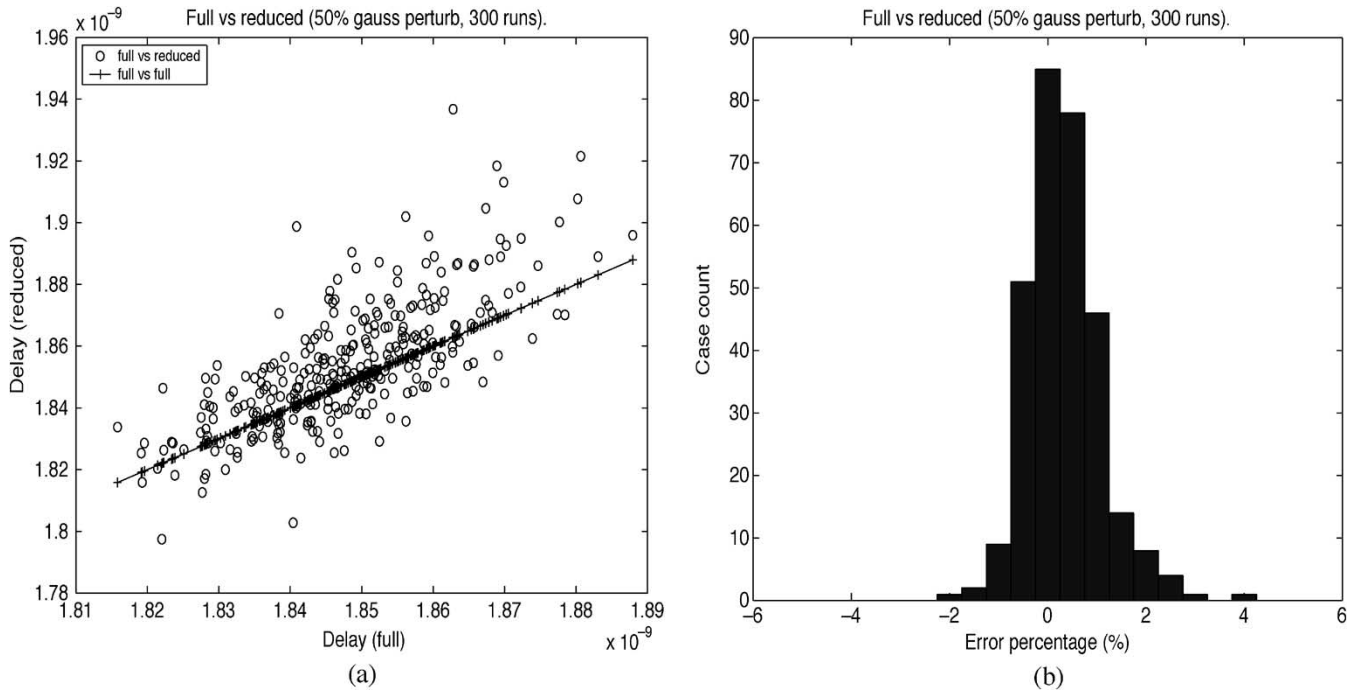


Fig. 8. Monte Carlo test of the RC ladder circuit reduction using the mixed moment matching. (a) Delay distribution of 300 cases with Gaussian perturbation up to 50%. (b) Histogram of the measured delay error percentage.

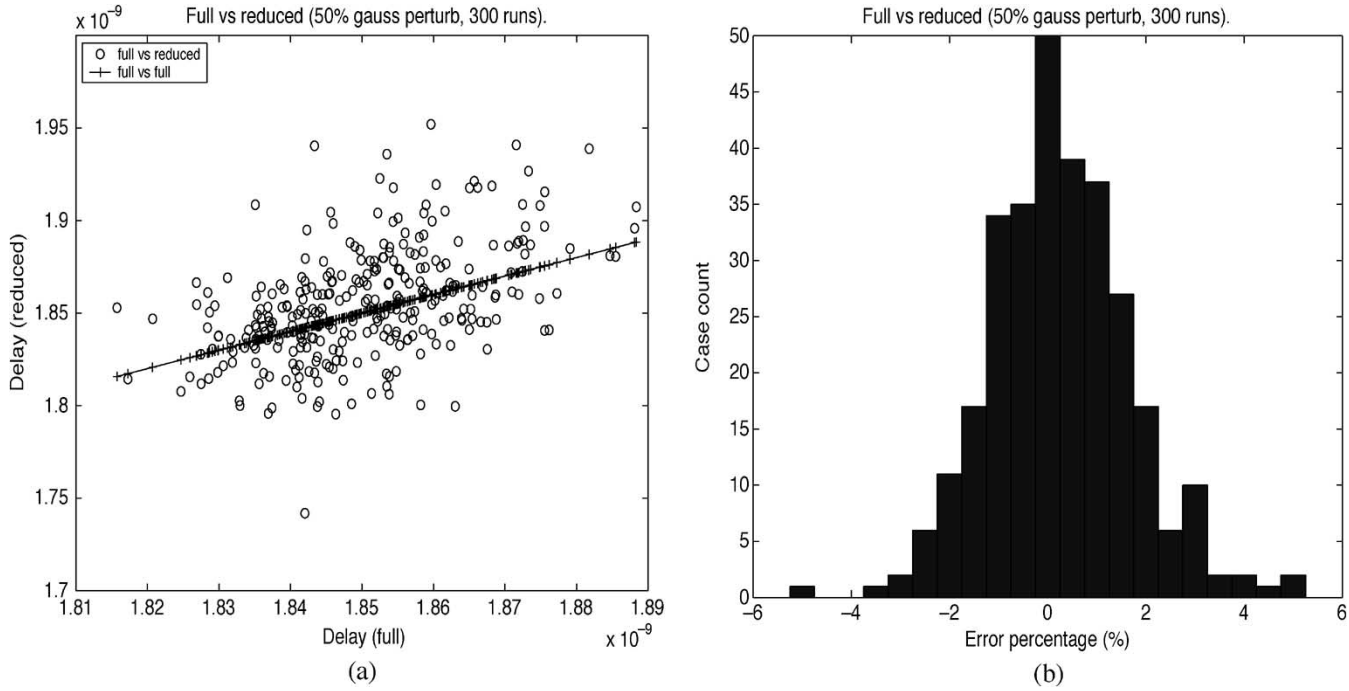


Fig. 9. Monte Carlo test of the RC ladder circuit reduction using the moment matching at $s = 0$. (a) Delay distribution of 300 cases with Gaussian perturbation up to 50%. (b) Histogram of the measured delay error percentage.

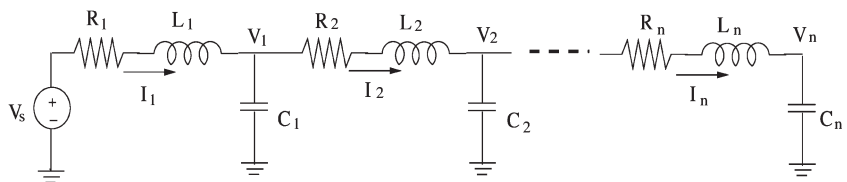


Fig. 10. RCL ladder circuit.

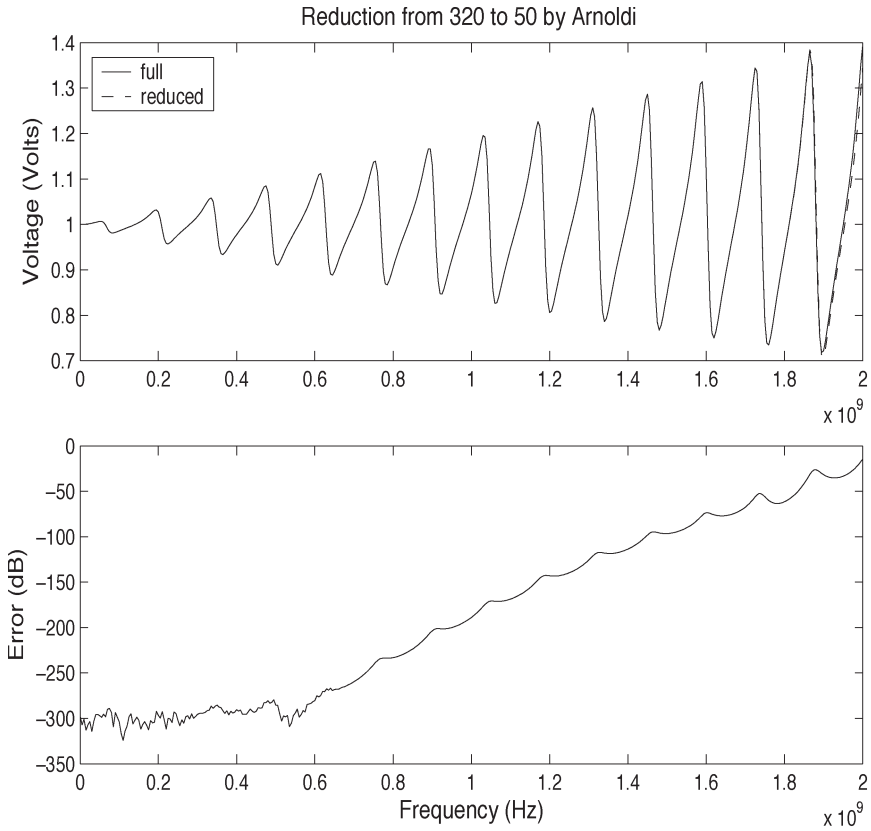


Fig. 11. Nominal model reduced by nominal projection.

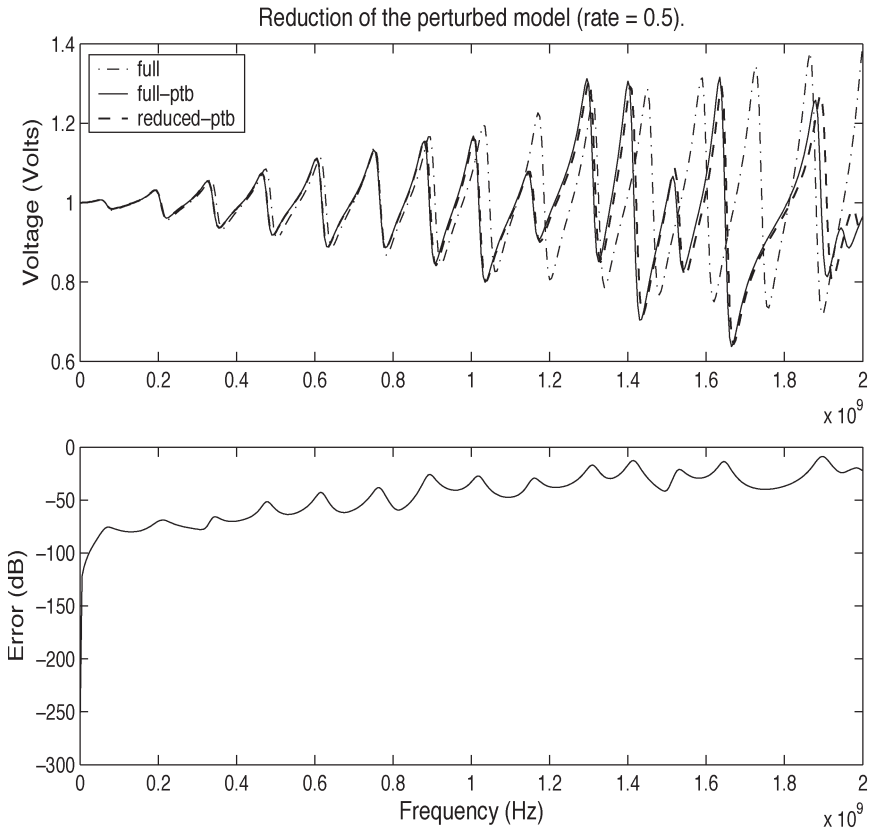


Fig. 12. Reduction of a perturbed model by nominal reduction.

delay time. We believe that matching at the high-frequency band is important because this part controls the signal rising behavior.

Then, we examine whether the robustness of mixed moment matching is good. For this purpose, we use a projection matrix generated by mixed moment matching from the nominal model and use it for the reduction of a perturbed model. The results are reported in Fig. 7. Fig. 7(a) shows the deviation of the perturbed model from the nominal model in the frequency responses after the circuit parameters have been perturbed up to 60% at random. Despite the deviation, the model reduced from the perturbed model by using the nominal projection still generates an accurate approximation of the full-order perturbed model, as seen from Fig. 7(b). Furthermore, the time-domain step response is shown in Fig. 7(c); again, we see a very close approximation.

The reliability of the nominal-projection method is further demonstrated by Monte Carlo tests. Shown in Fig. 8 are the Monte Carlo test results for the RC ladder circuit with 300 runs. The nominal parameters are perturbed individually by a Gaussian distribution up to 50%. We use the rise time up to 0.8 of a step response for the delay measurement. Fig. 8(a) shows the distribution of the measured delays out of 300 runs by using a nominal-projection matrix computed by the mixed-moment-matching method, with the x-axis for the delay of the full-order model and the y-axis for the delay of the reduced-order model. The diagonal line marked "+" is the equal-delay line for reference. The plot shows the clustering of the delay measurements surrounding the reference line. Noticing the small axes scale (10^{-9}) used in this plot, we conclude that the delays measured from the reduced-order models are mostly quite accurate. Fig. 8(b) shows a histogram of the relative-delay-error percentage computed from 300 runs. The relative delay error is defined by the formula

$$\frac{\text{Delay of reduced model} - \text{Delay of full model}}{\text{Delay of full model}}. \tag{28}$$

The relative-delay-error percentage mostly falls within 2%, meaning that the nominal projection has an adequate robustness given the fact that the parameters have been perturbed up to 50%.

To compare whether mixed moment matching has a better performance for delay measuring, we ran another Monte Carlo test with 300 runs but using the tenth-order Krylov subspace at $s = 0$ (i.e., DC moment matching). The test results are shown in Fig. 9. The histogram in Fig. 9(b) indicates a wider deviation of the relative delay error with a lower count at the center, comparing to that in Fig. 8(b). Evidently, the nominal projection computed from mixed moment matching has a better robustness. We note that, among the 300 runs, both methods did not encounter an unstable reduced-order model, which indicates that the nominal projection is quite reliable for RC network reduction.

The second method for nominal-projection computation is to use the real rational Krylov subspace. The robustness of such a nominal projection is tested by the inductive ladder circuit shown in Fig. 10. In this example, the nodal voltages and the currents passing the inductors are the state variables. We chose a model with the 320th order and reduced it to one of order 50 with a real rational Krylov subspace computed at $\sigma = 5 \times 10^9$. Uniform nominal RCL values were chosen with $R = 0.2\ \Omega$, $L = 1.0$ nH, and $C = 0.5$ pF. Fig. 11 shows the frequency responses of both the full-order and reduced-order models together with the error plot. The real rational Krylov subspace has achieved a good approximation at the resonance frequency band.

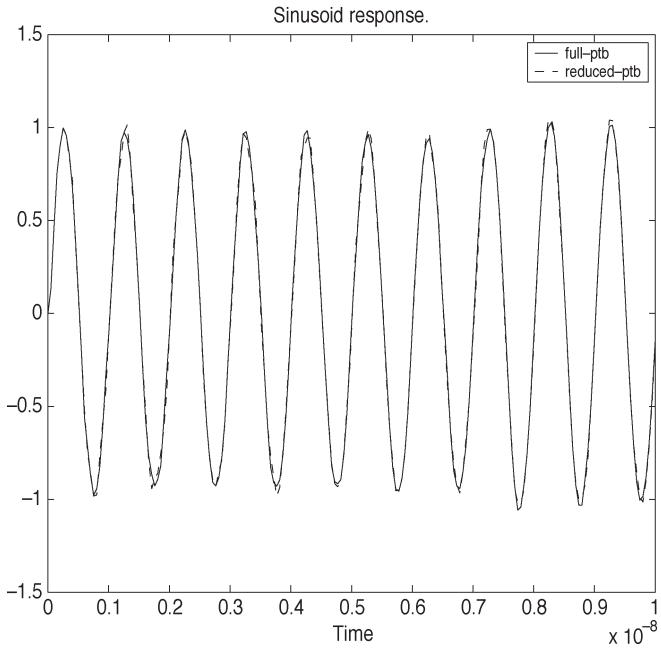


Fig. 13. Sinusoidal response at $f = 1$ GHz.

To test whether the nominal projection still works for perturbed models, we add random perturbations to each RCL value up to $\pm 50\%$ and then perform the reduction by using the nominal projection computed already. Shown in Fig. 12 is the frequency response result and the error plot. The frequency response of the nominal full-order model is also plotted (the dot-dashed curve) for reference. Clearly, the frequency response of the reduced-order model still captures the full-order frequency response quite well. Although the error becomes larger, it is still at an acceptable level.

The transient response is shown in Fig. 13, which displays the output of the perturbed full-order and reduced-order models in response to a sinusoidal input at the frequency of 1 GHz. Except for some slight distortions at the transient, the two waveforms match quite well. The reduction results for another set of perturbed parameters at random are shown in Figs. 14 and 15.

To demonstrate the efficiency achievable by using the nominal-projection method, we use the RC ladder circuit in Fig. 5 again and reduce a sequence of models with sizes 100, 200, 300, 400, and 500. For each model, we first compute the nominal-projection matrix V of dimension $n \times q$, where n is the full model size and q is the reduced model size. For each full model, we reduce it to $q = 10, 12, 14, 16, 18$, respectively. With each nominal-projection matrix V, we obtain a reduced-order symbolic model according to (16). Then, for each model size,

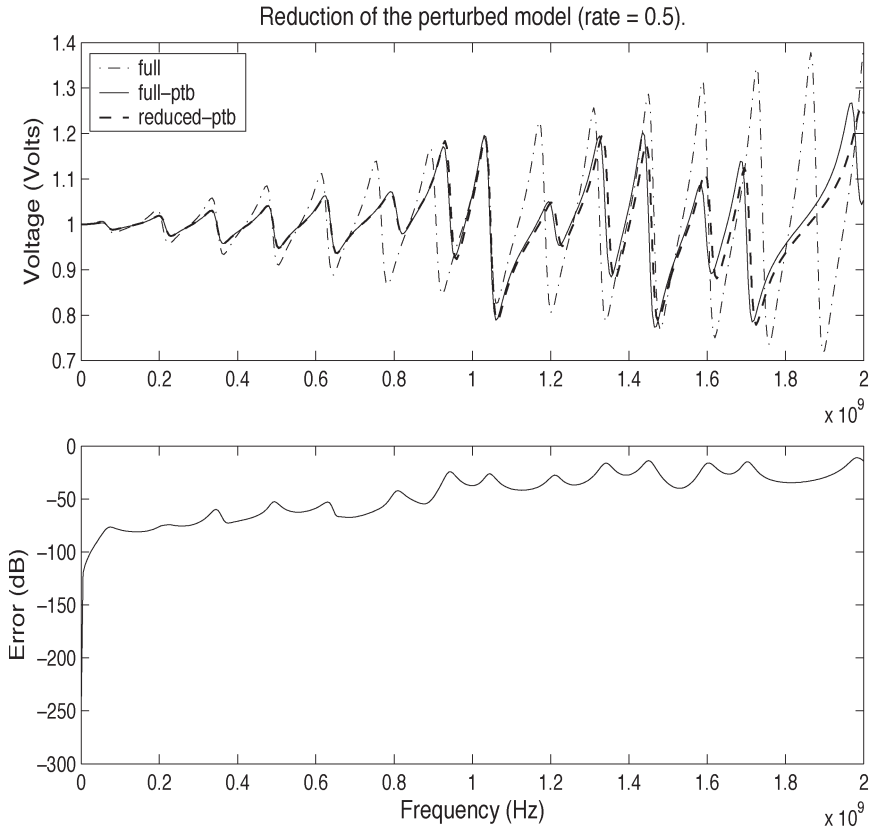


Fig. 14. Reduction of another perturbed model by nominal reduction.

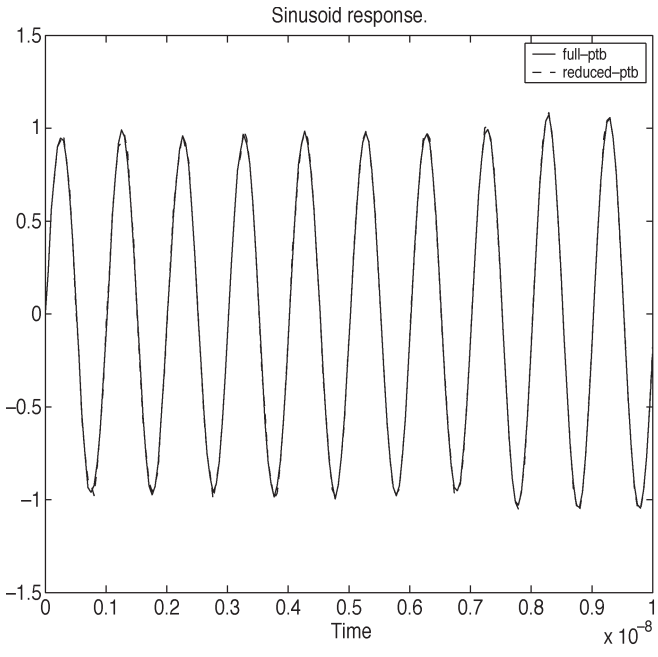


Fig. 15. Another sinusoidal response at $f = 1$ GHz.

we compare the time used for running a transient from 0 to 1 μs between the full-order parametric model and the reduced-order parametric model for 20 sets of randomly generated parameter values. To have a fair comparison, the time used for generating a nominal-projection matrix was also counted with the time used for the transient analysis of the reduced-order model. We used

TABLE I
SPEEDUP VERSUS MODEL SIZE: NOMINAL PROJECTION

Full model size	100	200	300	400	500
Reduced model size	10	12	14	16	18
Speedup	4.8	72.1	132.8	214.7	255.1
Rel. Delay Err. (%)	0.4	0.32	0.18	0.00	0.08

the MATLAB step response function for this simulation and the transient times are averaged over the 20 runs. In addition to the statistics for speedup, we also monitored the quality of the reduced-order models, for which we used the absolute values of the relative delay error defined in (28) also averaged over the 20 runs for each model size. The data are collected in Table I, where the average speedups are significant for the large-size models. Fig. 16 shows a visualization of the speedup data. We mention that for this set of statistical run, we did not encounter any unstable reduced-order models. We also comment that the speedup measured from the MATLAB experiment is very rough. If the algorithms are implemented in the C language and the modern sparse matrix technology is used, we believe the speedup would be more significant.

C. First-Order Approximation

The first-order-approximation method is an improvement over the nominal-projection method for better accuracy at the price of higher algebraic operation complexity. The RC circuit in Fig. 5 is used again to test the first-order-approximation method using an approximate symbolic Krylov subspace. The

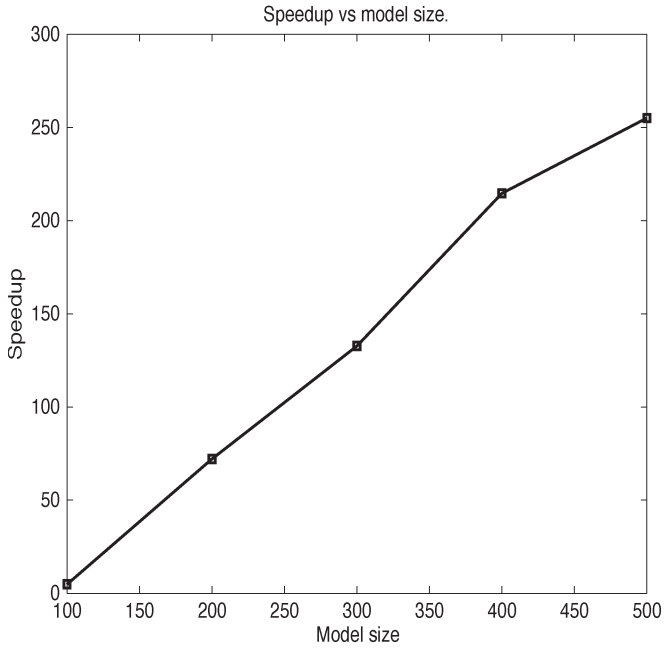


Fig. 16. Simulation speedup using nominal projection.

circuit consists of 100 stages of RC blocks and the model is of the 100th order. The nodal voltages are the state variables.

For demonstration purposes, we used the mathematical software Maple [22] to do symbolic algebra. The circuit is divided into three sections: nodes 1–30 for section 1, nodes 31–60 for section 2, and the rest for section 3. Four parameters are introduced as the symbols, including the two parameters representing the perturbations added to the capacitors and the resistors in the first section and the other two parameters representing the perturbations added to the capacitors and the resistors in the second section, respectively. The resistors and capacitors in the third section are not perturbed. The model is reduced from the 100th order to the fourth order. The effectiveness of the first-order-approximation method is tested by running a Monte Carlo test with the four parameters perturbed up to 30% according to the normal distribution. We simulated the unit step responses of both the full-order and reduced-order models and compared the 50% rise times. The statistical test results are shown in Fig. 17, with the delays measured shown in Fig. 17(a) and a histogram in Fig. 17(b). From these plots, we see that the first-order-approximation method was able to produce very accurate reduced-order models measured by the rise times, more accurate than the nominal-projection method. The extra prices are more complicated reduced-order models to be stored symbolically and more model evaluation time.

In addition to the accuracy achievable, we also demonstrate the efficiency of using a symbolic reduced-order model obtained from the first-order-approximation method. Again, we use the RC ladder circuit in Fig. 5 with model sizes of 100, 200, 300, and 400, all reduced to order 4. The $n \times 4$ projection matrices V were constructed symbolically by Maple using the first-order-approximation formula in (25).

Maple took 0.156 s to reduce the 100th order model to the fourth order. The symbolic reduced-order model was then evaluated in MATLAB. We ran 800 repeated transient analyses for different sets of the four parameter values; each transient spans the time interval from 0 to 12 s. It took 360 s to complete the whole batch of simulations using the full-order model while it only took 75 s for the same amount of simulation using the reduced-order symbolic model, in which 54 s was used for the transient simulation and 21 s for updating the symbolic model.

Similar to what we did in the nominal-projection case, we collected the speedup data in a transient analysis from 0 to 12 s for the four different sizes of models. The data are collected in Table II, and the speedup is also shown in Fig. 18. In Table II, we also show the time and the memory used by Maple for generating the symbolic reduced-order models. The speedups here were measured from simulating the full-order and reduced-order models for the given time period of transient analysis averaged over 800 runs.

D. *Efficiency of the Three Methods*

We briefly summarize the advantages and weaknesses of the three SMOR methods based on our experiences in experimentation.

There is no assumption of small parameter variation in the symbol-isolation method. Hence, if we are interested in studying the effect of a few elements in a large network, reducing the nonsymbolic part of network is highly recommended. In addition to its usefulness for sweeping analysis, this technique can be used for behavioral modeling of a large network with a few tuning elements.

The other two methods, the nominal projection and the first-order approximation, both assume small parameter variations. Although parameter variations are common phenomenon, drawing a line between the small and the large variations is not easy. We believe that uncontrollable statistical variations such as that caused by the uncertainty in the process technology can be treated as a small variation. However, there could be cases that small variations result in a large deviation of the circuit behavior. Then, the two methods are probably not applicable for such highly singular circuits.

While the nominal projection is computationally efficient, its accuracy is relatively weak. Hence, it is recommended for a rough statistical analysis rather than for an accurate nanoscale timing analysis. On the other hand, although the first-order-approximation method has a better accuracy owing to symbolic correction to the nominal Krylov vectors, its complexity is relatively higher. Also, because the symbolic Krylov vectors are not orthonormalized, it may not be reliable to generate symbolic reduced-order models with a higher order. Since most RC networks can be safely approximated by very-low-order models, the first-order-approximation method is highly recommended for the symbolic reduction of such networks.

VII. Conclusion and Future Work

Motivated by the practical needs for behavioral modeling in a parametric form, we have explored the possibility and the potential of SMOR in this paper. Three methods are proposed and tested, namely, symbol isolation, nominal projection, and first-order approximation, and each is applicable for certain

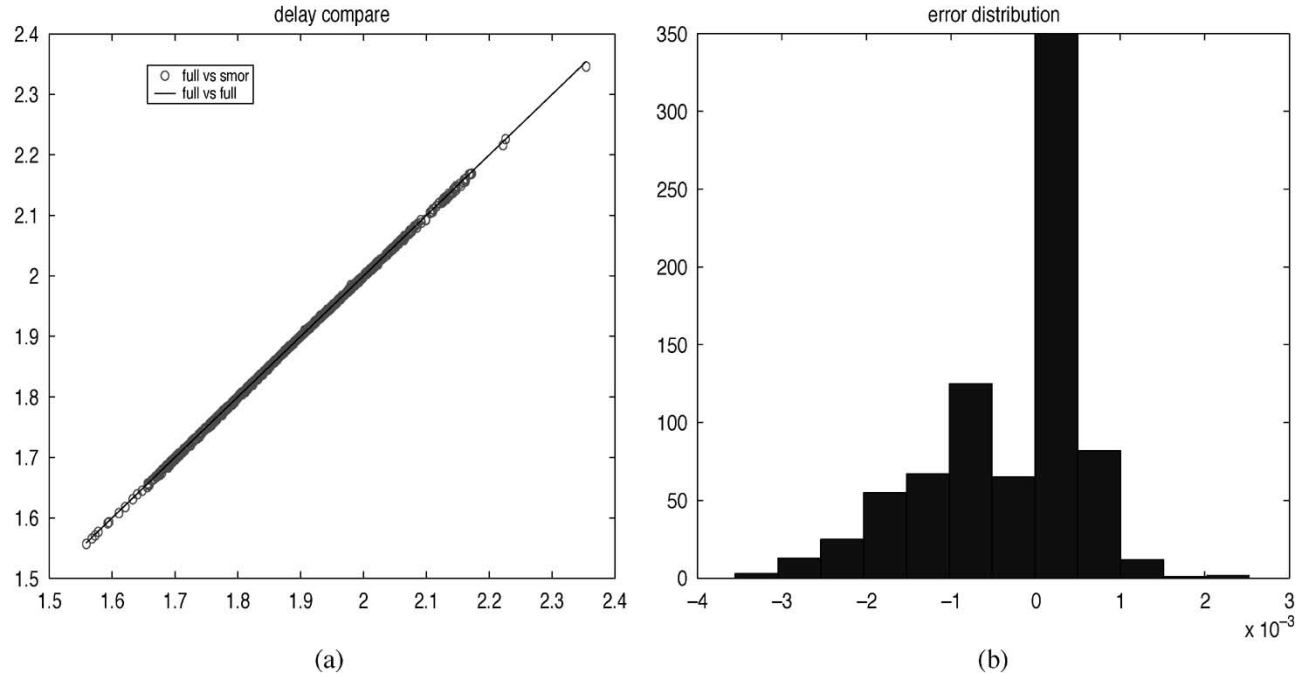


Fig. 17. Monte Carlo test of the symbolic first-order-approximation method. (a) Delay distribution of 800 cases with normal perturbation up to 30%. (b) Histogram of the measured delay error.

TABLE II
SPEEDUP VERSUS MODEL SIZE: FIRST-ORDER METHOD

Full model size	100	200	300	400
Reduced model size	4	4	4	4
Reduction time (s)	12	82	276	746
Maple memory (MB)	11	26.2	53.6	90
Speedup	5	61	130	141
Rel. Delay Err. (%)	< 0.3	< 0.3	< 0.3	< 0.3

circuit models with typical features. The potential effectiveness of these methods has been demonstrated by experiments on some typical circuits that are representative of interconnect models.

Although numerous model-order-reduction algorithms are available in literature for nonsymbolic linear model order reduction, their extension to SMOR is nontrivial in general. The experimental study conducted in this paper reveals that certain parametric modeling problems can be addressed by SMOR and solved by using the proposed methods. Future research includes efficient model-order-reduction methods for uncertain models, error bounds for using the nominal projection, nonsubspace-based approaches, and applications of other identification-based approaches to SMOR.

We note that a research area closely related to SMOR is the symbolic circuit analysis [23]–[25]. The symbolic circuit analysis finds symbolic expressions of circuit transfer functions by various enumeration techniques such as topological tree enumeration [23] and determinant expansion [26]. Very often, simplified expressions that approximate the exact circuit transfer functions are sought. In this sense, the approximate symbolic circuit analysis shares the same objective as the SMOR, namely, finding simplified but equivalent models at the circuit ports. However, they differ in their underlying principles; the approximate symbolic circuit analysis finds the principal terms from a system matrix determinant [26], [27], while the model order reduction finds the principal eigenvalues or the dominant state space of a linear system. It will be interesting to study the connection between the subspace-based SMOR and the approximate symbolic circuit analysis in the frequency domain.

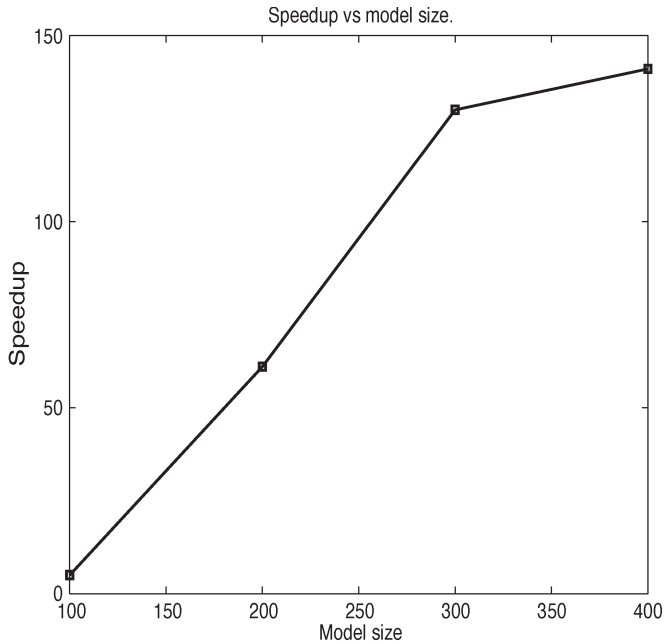


Fig. 18. Simulation speedup using the first-order SMOR models.

ACKNOWLEDGMENT

The authors would like to thank the three anonymous reviewers for their constructive comments.

REFERENCES

[1] L. Silveira, I. Elfadel, J. White, M. Chilukuri, and K. Kundert, "Efficient frequency-domain modeling and circuit simulation of transmission lines," *IEEE Trans. Compon., Packag., Manuf. Technol. B*, vol. 17, no. 4, pp. 505–513, Nov. 1994.

[2] L. Silveira, M. Kamon, and J. White, "Efficient reduced-order modeling of frequency-dependent coupling inductances associated with 3-D interconnect structures," *IEEE Trans. Compon., Packag., Manuf. Technol. B*, vol. 19, no. 2, pp. 283–288, May 1996.

[3] 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.

[4] L. Pillage and R. Rohrer, "Asymptotic waveform evaluation for timing analysis," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 9, no. 4, pp. 352–366, Apr. 1990.

[5] J. Roychowdhury, "Automated macromodel generation for electronic systems," in *Proc. IEEE Int Behavioral Modeling and Simulation (BMAS) Workshop*, San Jose, CA, 2003, pp. 11–16.

[6] E. Hung and S. Senturia, "Generating efficient dynamical models for MEMS systems from a few finite-element simulation runs," *J. Microelectromech. Syst.*, vol. 8, no. 3, pp. 280–289, 1999.

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

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

[9] Z. Bai, "Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems," *Appl. Numer. Math.*, vol. 43, no. 1–2, pp. 9–44, Oct. 2002.

[10] L. Daniel, C. S. Ong, S. C. Low, K. H. Lee, and J. White, "Geometrically parameterized interconnect performance models for interconnect synthesis," in *Proc. Int. Symp. Physical Design*, San Diego, CA, 2002, pp. 202–207.

[11] Y. Liu, L. Pileggi, and A. Strojwas, "Model order-reduction of RC(L) interconnect including variational analysis," in *Proc. 36th Design Automation Conf.*, New Orleans, LA, 1999, pp. 201–206.

[12] G. Shi and C.-J. Shi, "Parametric reduced order modeling for interconnect analysis," in *Proc. Asia South Pacific Design Automation Conf. (ASPDAC)*, Yokohama, Japan, 2004, pp. 774–779.

[13] Y. Halevi, A. Zlochevsky, and T. Gilat, "Parameter-dependent model order reduction," *Int. J. Control*, vol. 66, no. 3, pp. 463–485, Feb. 1997.

[14] B. Hu, G. Shi, and C.-J. R. Shi, "Symbolic model order reduction," in *Proc. IEEE Int. Workshop Behavioral Modeling and Simulation (BMAS)*, San Jose, CA, 2003, pp. 34–40.

[15] E. Grimme, "Krylov projection methods for model reduction," Ph.D. dissertation, Electr. Comput. Eng. Dept., Univ. Illinois, Urbana-Champaign, 1997.

[16] C.-W. Ho, A. E. Ruehli, and P. A. Brennan, "The modified nodal approach to network analysis," *IEEE Trans. Circuits Syst.*, vol. CAS-22, no. 6, pp. 504–509, Jan. 1975.

[17] P. Feldmann and F. Liu, "Sparse and efficient reduced order modeling of linear subcircuits with large number of terminals," in *Proc. Int. Conf. Computer-Aided Design*, San Jose, CA, 2004, pp. 88–92.

[18] T. Kohonen, *Self-Organization and Associative Memory*. Berlin, Germany: Springer-Verlag, 1988.

[19] Y. Ismail and E. Friedman, "Effects of inductance on the propagation delay and repeater insertion in VLSI circuits," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 8, no. 2, pp. 195–206, Apr. 2002.

[20] E. Chiprout and M. Nakhla, "Analysis of interconnect networks using complex frequency hopping (CFH)," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 14, no. 2, pp. 186–200, Feb. 1995.

[21] G. Shi and C.-J. Shi, "Model order reduction by dominant subspace projection: Error bound, subspace computation and circuit application," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 52, no. 5, pp. 975–993, May 2005.

[22] *Maple 7*, Waterloo, ON, Canada: Waterloo Maple Inc., 2001.

[23] P. Lin, *Symbolic Network Analysis*. New York: Elsevier, 1991.

[24] G. Gielen and W. Sansen, *Symbolic Analysis for Automated Design of Analog Integrated Circuits*. Norwell, MA: Kluwer, 1991.

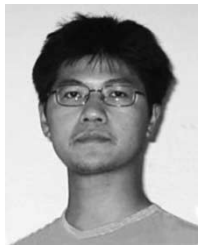
[25] F. Fernández, A. Rodríguez-Vázquez, J. Huertas, and G. Gielen, *Symbolic Analysis Techniques—Applications to Analog Design Automation*. New York: IEEE Press, 1998.

[26] C.-J. Shi and X.-D. Tan, "Canonical symbolic analysis of large analog circuits with determinant decision diagrams," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 19, no. 1, pp. 1–18, Jan. 2000.

[27] Q. Yu and C. Sechen, "A unified approach to the approximate symbolic analysis of large analog integrated circuits," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 43, no. 8, pp. 656–669, Aug. 1996.

Guoyong Shi (S'98–M'02) received the Bachelor's degree in applied mathematics from Fudan University, Shanghai, China, in 1987, the M.E. degree in electronics and information science from Kyoto Institute of Technology, Kyoto, Japan, in 1997, and the Ph.D. degree in electrical engineering from Washington State University, Pullman, in 2002.

He is currently a Professor in the School of Microelectronics, Shanghai Jiao Tong University, Shanghai, China. From August 2002 to June 2005, he was a Postdoctoral Research Fellow in Department of Electrical Engineering, University of Washington, Seattle, working on mixed-signal computer-aided design (CAD) projects supported by the Defense Advanced Research Projects Agency (DARPA), the National Science Foundation (NSF), and the Semiconductor Research Corporation (SRC). He is the author or coauthor of about 40 technical articles on a variety of subjects including artificial intelligence, operations research, control theory, and circuit simulation. His current area of interest is in research on multiple aspects of electronic design automation.



Bo Hu (S'02) received the Bachelor's degree in physics from the University of Science and Technology of China, Hefei, China, in 1999 and the M.S. degree in electrical engineering in 2004 from the University of Washington, Seattle, where he is currently working toward the Ph.D. degree.

His current research interests include symbolic circuit analysis, device modeling, circuit macromodeling, and simulation.

C.-J. Richard Shi (S'91–M'91–SM'99–F'06) received the Ph.D. degree in computer science from the University of Waterloo, Waterloo, ON, Canada in 1994.

From 1994 to 1998, he was with Analogy, Rockwell Semiconductor Systems, and the University of Iowa. In 1998, he joined the University of Washington, Seattle, where he is currently a Professor in electrical engineering. His research interests include several aspects of the computer-aided design (CAD) and test of integrated circuits and systems, with particular emphasis on analog/mixed-signal and deep-submicrometer circuit modeling, simulation, and design automation.

Dr. Shi is a key contributor to IEEE Std. 1076.1-1999 very-high-speed-integrated-circuit (VHSIC) hardware description language analog and mixed signal (VHDL-AMS) standard for the description and simulation of mixed-signal circuits and systems. He founded IEEE International Workshop on Behavioral Modeling and Simulation (BMAS) in 1997, and has served on the technical program committees of several international conferences. He received a Best Paper Award from the 1999 IEEE/ACM Design Automation Conference, a Best Paper Award from the 1998 IEEE VLSI Test Symposium, a National Science Foundation Career Award (2000), and a Doctoral Prize from the Natural Science and Engineering Research Council of Canada (1995). He has been an Associate Editor, as well as a Guest Editor, of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II, ANALOG AND DIGITAL SIGNAL PROCESSING. He is currently an Associate Editor of the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS, as well as the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-II: TRANSACTION BRIEFS.