

Model Reduction of Variable-Geometry Interconnects Using Variational Spectrally-Weighted Balanced Truncation

Payam Heydari

Department of Electrical and Computer Engineering
University of California
Irvine, CA 92697

Massoud Pedram

Department of Electrical Engineering-Systems
University of Southern California
Los Angeles, CA 90089

Abstract - This paper presents a spectrally-weighted balanced truncation technique for RLC interconnects, a technique needed when the interconnect circuit parameters change as a result of variations in the manufacturing process. The salient features of this algorithm are the inclusion of parameter variations in the RLC interconnect, the guaranteed stability of the reduced transfer function, and the availability of provable frequency-weighted error bounds for the reduced-order system. This paper shows that the balanced truncation technique is an effective model-order reduction technique when variations in the circuit parameters are taken into consideration. Experimental results show that the new variational spectrally-weighted balanced truncation attains, on average, 20% more accuracy than the variational Krylov-subspace-based model-order reduction techniques while the run-time is also, on average, 5% faster.

1. INTRODUCTION

The problem of interconnect model reduction has gained considerable attention in the EDA community in recent years. Model reduction techniques enable us to capture the interconnect effects with a much shorter computational time than that required for the simulation of the full circuit. On the other hand, as the minimum feature sizes shrink to the sub-quarter microns, geometrical variations in the line width, metal height, and dielectric thickness due to process variations have more pronounced effects on the reliability and performance of VLSI circuits [1]. As a consequence, it is crucial to assess the impact of these process variations on model-order reduction techniques.

Among various classes of model reduction techniques, the explicit moment matching algorithms (AWE [2], RICE [3]) and Krylov-subspace-based methods (Pact [4], PVL [5], PRIMA [6]) have received the most attention for generating reduced order models of the interconnects. The computational complexity of these model-order reduction techniques is due to the matrix-vector products. However, these methods do not provide a provable error bound for the reduced system.

An alternative to these model reduction techniques that has seen renewed interest and consideration are the balanced realization techniques [7], [8], [9]. Balanced realization techniques have not received the same attention as Krylov-based and Pade-based model reduction techniques partly because these methods normally require time-consuming computations [7]. A reduced system using the balanced realization technique is guaranteed to be stable. Moreover, a provable error bound does exist for this class of model-order reduction techniques.

In the balanced realization-based model reduction methods, each state is equally controllable and observable, and the reduced-order model of the original transfer function is derived by minimizing the Hankel-norm of the error between the transfer functions of the original and the approximated system. Paper [8] uses the truncated balanced realization as well as the Schur decomposition method to develop an efficient numerical method for the order reduction of a large linear time-invariant (LTI) system. The bottleneck in balanced truncation methods is the computational complexity incorporated in solving the Lyapunov equations. Papers [9] and [10] propose efficient algorithms to solve the two Lyapunov equations in order to obtain the controllability and the observability grammians. The algorithms are based on the Alternated Direction Implicit (ADI) method that was proposed in [11].

A shortcoming of these model reduction techniques is that they do not reshape the frequency spectrum to emphasize error minimization in some frequency range of interest. Furthermore, they also do not address the numerical difficulties when the system is nearly uncontrollable or unobservable. In paper [12], a new numerically stable, frequency-weighted balanced truncation technique was presented. The proposed method gives a definite *a priori* error bound and is guaranteed to be stable even when both input and output weightings are utilized at the same time.

Analyzing the interconnect without taking into account the rather large variations of the interconnect geometries is not useful in practice. These

variations are especially large in the inter-layer dielectric (ILD) thickness and the metal line width and height. These process-dependent geometrical variations have a definite impact on the total line and inter-wire coupling parasitics, which in turn results in variations in the signal delay and the coupling noise. Liu *et al.* in [13] studied the effect of interconnect parameter variations on Krylov-subspace model-order reduction techniques. The paper basically combines the matrix perturbation theory [14] and the Krylov-subspace-based model reduction method [4], [6]. The authors allow two-dimensional variations on the projection matrices. To compute the corresponding sensitivities of the susceptance and conductance matrices to each dimensional variation, some sample points were picked up and the dominant eigenvalues/eigenvectors were calculated.

The goal of this paper is to consider the effects of process variations and spectral shaping on model-order reduction using balanced truncation technique and to propose an efficient order reduction technique that includes these effects. The main contribution of this paper is our use of the method proposed in [12] combined with a new variational balanced truncation approach that accounts for process variations in order to obtain a new variational spectrally-weighted balanced truncation method.

Section 2 gives a brief overview of the balanced realization method. A discussion about the effect of process variations on interconnect modeling is also provided in this section. Section 3 reviews the formulation of the frequency-weighted model reduction proposed in [12]. In section 4, the new variational balanced truncation with spectral shaping is illustrated, and a theoretical comparison between this work and the work presented in paper [13] is made. In section 5, The new model reduction technique is compared to the work presented in [13] by running simulations on a number of global interconnect line and clock tree scenarios. Other experimental results in this section verify the accuracy of our technique. Finally, section 6 presents the conclusions of our paper.

2. BACKGROUND

In this section, first, we briefly explain the conventional balanced truncation method. Next, a brief discussion about the effect of process variations on interconnect modeling will be provided.

2.a. An overview of balanced realization

Consider a network consisting of inductances, capacitances, and resistances. Modified Nodal Analysis (MNA) can be used to obtain the following system of equations:

$$L \dot{\mathbf{x}} = -\mathbf{G}\mathbf{x} + \mathbf{B}_N \mathbf{u} \quad (1)$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} \quad (2)$$

where the state vector \mathbf{x} represents the vector of N node voltages across circuit capacitances and voltage sources and the vector of M currents flowing through inductors and current sources. In addition:

$$L = \begin{bmatrix} \mathbf{C}_{cap} & 0 \\ 0 & \mathbf{L} \end{bmatrix}, \quad G = \begin{bmatrix} \mathbf{G} & \mathbf{E} \\ -\mathbf{E}^T & \mathbf{0} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \mathbf{v} \\ \mathbf{i} \end{bmatrix}, \quad \mathbf{E} = [e_{ij}]$$

$$\text{where } e_{ij} = \begin{cases} 1 & \text{If branch } j \text{ is incident at node } i \text{ and oriented away from it.} \\ 0 & \text{If branch } j \text{ is not incident at node } i. \\ -1 & \text{If branch } j \text{ is incident at node } i \text{ and oriented towards it.} \end{cases}$$

The MNA equations can be rewritten in the form of the standard state-space representation by introducing the following matrices:

$$\mathbf{A} = -L^{-1}G, \quad \mathbf{B} = L^{-1}\mathbf{B}_N \quad (3-a)$$

Hence:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \quad (3)$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} \quad (4)$$

For stable LTI systems, the controllability and observability grammians are defined as follows:

$$\mathbf{P} = \int_0^{\infty} e^{\mathbf{A}t} \mathbf{B} \mathbf{B}^T e^{\mathbf{A}^T t} dt \quad ; \quad \mathbf{Q} = \int_0^{\infty} e^{\mathbf{A}^T t} \mathbf{C}^T \mathbf{C} e^{\mathbf{A}t} dt \quad (5)$$

The \mathbf{P} and \mathbf{Q} matrices (known as controllability and observability grammians) satisfy the Lyapunov equations [7]:

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = \mathbf{0} \quad ; \quad \mathbf{A}^T \mathbf{Q} + \mathbf{Q}\mathbf{A} + \mathbf{C}^T \mathbf{C} = \mathbf{0} \quad (6)$$

The controllability and observability grammians give some interesting insights about the system characteristics. A particularly interesting property is that the Hankel singular values of the system transfer function, $\mathbf{H}(s)$, are the square-roots of the eigenvalues of $\mathbf{P}\mathbf{Q}$:

$$\sigma_i(\mathbf{H}(s)) = \{\lambda_i(\mathbf{P}\mathbf{Q})\}^{1/2} \quad (7)$$

Each Hankel singular value represents the energy exerted by the corresponding state variable in the controllability and observability map of the balanced system. It can be proved that there exists a similarity transformation matrix \mathbf{T} such that the controllability and observability grammians of the new system $(\mathbf{A}_r, \mathbf{B}_r, \mathbf{C}_r)$ are equal and diagonal:

$$\mathbf{P}_r = \mathbf{Q}_r = \Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n) \quad (8)$$

where $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n > 0$. The balanced transformation of an LTI system allows one to choose the state variable set that would provide a significant amount of information in the external representation of the system. In fact, on the basis of the computed energy for each state variable x_k , we settle on a criterion for evaluating the possibility of eliminating x_k in the reduced model scheme. If Σ is partitioned into two submatrices:

$$\Sigma = \begin{bmatrix} \Sigma_1 & \mathbf{0} \\ \mathbf{0} & \Sigma_2 \end{bmatrix} \quad (9)$$

where $\Sigma_1 \in \mathfrak{R}^{k \times k}$, $\Sigma_2 \in \mathfrak{R}^{(n-k) \times (n-k)}$. The new coordinate transformed system $(\mathbf{A}_r, \mathbf{B}_r, \mathbf{C}_r)$ is also partitioned in conformity with Σ as:

$$\mathbf{A}_r = \mathbf{T}\mathbf{A}\mathbf{T}^{-1} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \quad \mathbf{B}_r = \mathbf{T}\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix}, \quad \mathbf{C}_r = \mathbf{C}\mathbf{T}^{-1} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \end{bmatrix}$$

The reduced order model based on $(\mathbf{A}_{11}, \mathbf{B}_1, \mathbf{C}_1)$ is stable, and the L^∞ -error is bounded by [7]:

$$\|\mathbf{H}(s) - \mathbf{H}^k(s)\|_\infty \leq 2 \left(\sum_{i=k+1}^n \sigma_i \right) \quad (10)$$

2.b. Modeling of interconnect variations

Due to process variations, interconnect technology parameters are varying substantially. These parameters can have as much as a 30% variation off their nominal values [1]. Therefore, the effect of the process variations on the interconnect delay and crosstalk should be taken into consideration. A common approach to anticipate these variations in the design is the conventional skew-corner, worst-case modeling. This method, however, is too conservative because the probability of all 3- σ process corner values occurring simultaneously is very small. As a consequence, statistically-based worst-case interconnect modelings using Monte-carlo simulation have been proposed [16], [17]. These approaches, however, fail to handle large circuits that exist in reality. To alleviate the problem of having large computational complexity (as also mentioned in [13]) the effect of process variations must be taken into account in model-order reduction algorithms. Furthermore, the resulting variational reduced-order model needs to converge to the reduced-order model of the nominal network when all the parameter variations are zero.

3. BALANCED TRUNCATION WITH SPECTRAL SHAPING

We have seen that balanced realization is an attractive model reduction technique due to the fact that it gives a provable L^∞ -error bound for the reduced-order system. Even more attractive is a balanced truncation technique extended to include weighting on the input and/or output.

To determine the state-space characteristics of the new augmented system, we are faced with two basic questions that must be answered:

1. What set of points in the x -state space could be a part of the zero initial condition response for the weighted input denoted by $\eta(t)$?

2. What set of points in the x -state space as initial conditions could produce a weighted output denoted by $\xi(t)$?

Consider the state-space representation of a set of tightly coupled RLC interconnects given by equations (3) and (4). The goal of the frequency-weighted balanced realization technique is to calculate $\mathbf{H}^k(s)$ of degree k ($k < n$), making

$$\|\mathbf{W}_o(s) (\mathbf{H}(s) - \mathbf{H}^k(s)) \mathbf{W}_i(s)\|_\infty \quad (13)$$

as small as possible. To obtain such a reduced system we first calculate the grammians of the augmented system and go through the same steps that were taken for a unity-weighted system. First, we write the Laplace transformation of the input and output weighting functions:

$$\mathbf{W}_i(s) = \mathbf{C}_i(s\mathbf{I} - \mathbf{A}_i)^{-1} \mathbf{B}_i + \mathbf{D}_i \quad (14)$$

$$\mathbf{W}_o(s) = \mathbf{C}_o(s\mathbf{I} - \mathbf{A}_o)^{-1} \mathbf{B}_o + \mathbf{D}_o \quad (15)$$

According to definition [18], the controllable subspace of the augmented system, $\mathbf{H}(s)\mathbf{W}_i(s)$, is the solution set to the first question. A controller-form realization of the augmented system $\mathbf{H}(s)\mathbf{W}_i(s)$ is as follows:

$$\bar{\mathbf{A}}_i = \begin{bmatrix} \mathbf{A} & \mathbf{B}\mathbf{C}_i \\ \mathbf{0} & \mathbf{A}_i \end{bmatrix}, \quad \bar{\mathbf{B}}_i = \begin{bmatrix} \mathbf{B}\mathbf{D}_i \\ \mathbf{B}_i \end{bmatrix}, \quad \bar{\mathbf{C}}_i = \begin{bmatrix} \mathbf{C} & \mathbf{0} \end{bmatrix} \quad (16)$$

Similarly, according to definition [18], the observable subspace of the augmented system, $\mathbf{W}_o(s)\mathbf{H}(s)$, is the solution set to the second question. An observer-form realization of the augmented system $\mathbf{W}_o(s)\mathbf{H}(s)$ is as follows:

$$\bar{\mathbf{A}}_o = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{B}_o\mathbf{C} & \mathbf{A}_o \end{bmatrix}, \quad \bar{\mathbf{B}}_o = \begin{bmatrix} \mathbf{C} \\ \mathbf{0} \end{bmatrix}, \quad \bar{\mathbf{C}}_o = \begin{bmatrix} \mathbf{D}_o\mathbf{C} & \mathbf{C}_o \end{bmatrix} \quad (17)$$

For a complete explanation of controller and observer form realizations, see chapter 3 of reference [18]. Since all controllable and observable modes of these two augmented systems are determined by $n \times n$ upper left corner submatrices of $\bar{\mathbf{A}}_i$ and $\bar{\mathbf{A}}_o$, the desired controllability and observability grammians are given by the corresponding $n \times n$ upper left corner submatrices of $\bar{\mathbf{P}}$ and $\bar{\mathbf{Q}}$:

$$\bar{\mathbf{P}} = \begin{bmatrix} \mathbf{P} & \mathbf{P}_{12} \\ \mathbf{P}_{12}^T & \mathbf{P}_{22} \end{bmatrix}, \quad \bar{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q} & \mathbf{Q}_{12} \\ \mathbf{Q}_{12}^T & \mathbf{Q}_{22} \end{bmatrix} \quad (18)$$

where $\bar{\mathbf{P}}$ and $\bar{\mathbf{Q}}$ must satisfy the following Lyapunov equations:

$$\bar{\mathbf{A}}_i \bar{\mathbf{P}} + \bar{\mathbf{P}} \bar{\mathbf{A}}_i^T + \bar{\mathbf{B}}_i \bar{\mathbf{B}}_i^T = \mathbf{0} \quad (19)$$

$$\bar{\mathbf{A}}_o^T \bar{\mathbf{Q}} + \bar{\mathbf{Q}} \bar{\mathbf{A}}_o + \bar{\mathbf{C}}_o^T \bar{\mathbf{C}}_o = \mathbf{0} \quad (20)$$

Expanding the $n \times n$ upper left corner block of the Lyapunov equations yields:

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{B}\mathbf{C}_1\mathbf{P}_{12} + \mathbf{P}_{12}^T \mathbf{C}_1^T \mathbf{B}^T + \mathbf{B}\mathbf{D}_1^T \mathbf{B}^T = \mathbf{0} \quad (21)$$

$$\mathbf{A}^T \mathbf{Q} + \mathbf{Q}\mathbf{A} + \mathbf{Q}_{12}\mathbf{B}_o\mathbf{C} + \mathbf{C}^T \mathbf{B}_o^T \mathbf{Q}_{12}^T + \mathbf{C}^T \mathbf{D}_o^T \mathbf{D}_o\mathbf{C} = \mathbf{0} \quad (22)$$

From equations (21) and (22) the new variables are defined as follows:

$$\mathbf{X} = \mathbf{B}\mathbf{C}_1\mathbf{P}_{12} + \mathbf{P}_{12}^T \mathbf{C}_1^T \mathbf{B}^T + \mathbf{B}\mathbf{D}_1^T \mathbf{B}^T \quad (23)$$

$$\mathbf{Y} = \mathbf{Q}_{12}\mathbf{B}_o\mathbf{C} + \mathbf{C}^T \mathbf{B}_o^T \mathbf{Q}_{12}^T + \mathbf{C}^T \mathbf{D}_o^T \mathbf{D}_o\mathbf{C} \quad (24)$$

It is readily seen that \mathbf{X} and \mathbf{Y} are symmetric matrices. As a consequence, there exist orthogonal matrices \mathbf{U} and \mathbf{V} and diagonal matrices \mathbf{S} and \mathbf{Z} such that:

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{U}^T \quad (25)$$

$$\mathbf{Y} = \mathbf{V}\mathbf{Z}\mathbf{V}^T \quad (26)$$

where $\mathbf{S} = \text{diag}(s_1, s_2, \dots, s_n)$, and $\mathbf{Z} = \text{diag}(z_1, z_2, \dots, z_n)$, and

$|s_1| \geq |s_2| \geq \dots \geq |s_n| \geq 0$, $|z_1| \geq |z_2| \geq \dots \geq |z_n| \geq 0$. Suppose that $\text{rank}(\mathbf{X})=i$ and $\text{rank}(\mathbf{Y})=j$, where $1 \leq i, j \leq n$. We can write:

$$\bar{\mathbf{B}} = \mathbf{U} \text{diag}(|s_1|^{1/2}, \dots, |s_i|^{1/2}, 0, \dots, 0) \quad (27)$$

$$\bar{\mathbf{C}} = \text{diag}(|z_1|^{1/2}, \dots, |z_j|^{1/2}, 0, \dots, 0) \mathbf{V}^T \quad (28)$$

Let $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$ denote the solutions of the following Lyapunov equations:

$$\mathbf{A}\hat{\mathbf{P}} + \hat{\mathbf{P}}\mathbf{A}^T + \bar{\mathbf{B}}\bar{\mathbf{B}}^T = \mathbf{0} \quad (29)$$

$$\mathbf{A}^T \hat{\mathbf{Q}} + \hat{\mathbf{Q}}\mathbf{A} + \bar{\mathbf{C}}^T \bar{\mathbf{C}} = \mathbf{0} \quad (30)$$

We find that the transformation matrix $\hat{\mathbf{T}}$ that simultaneously diagonalizes $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$ is as follows:

$$\hat{\mathbf{T}}\hat{\mathbf{P}}\hat{\mathbf{T}}^T = (\mathbf{T}^{-1})^T \hat{\mathbf{Q}}\mathbf{T}^{-1} = \text{diag}(\sigma_1, \dots, \sigma_r, \sigma_{r+1}, \dots, \sigma_n) \quad (31)$$

Similar to the unit-weighted balanced truncation method, this transformation matrix, $\hat{\mathbf{T}}$, is used to map the original system to a new coordinate transformed system. The reduced-order system is then obtained from the transformed system. Note that $\hat{\mathbf{T}}$ contains the characteristics of the weighting functions.

The input and output weightings are determined based on the range of frequencies where we would like to have the maximum accuracy. The

weighting functions should emphasize the frequency ranges where more accuracy is required. Similarly, they must de-emphasize the range of frequencies where the noise resulting from the order reduction has very minimal energy or is out of the desired frequency bound.

4. VARIATIONAL SPECTRALLY-WEIGHTED BALANCED TRUNCATION

Characterization of the interconnect geometry variation is an important issue in deep-submicron VLSI technology. In order to accurately assess the performance of an interconnect system, it is essential to characterize the interconnect geometry, which in turn specifies the interconnect parasitics [13]. From a designer point of view, one important source of the IC performance variability is the physical source of variability [19]. For the purpose of design performance evaluation, we are concerned with two possible cases here. The first one includes the case where the interconnect (or device) parameters are constant within a die but vary within a wafer or a lot. In the second case, the device and interconnect parameters vary within a die. The inter-die variability can be minimized by using several techniques and corrections during the fabrication process. Due to its relatively low spatial frequency and smoothness, simple models can be used to describe the wafer level variations. As a result, the wafer level variation can be modeled locally (within the die) as a linear function of the position within the die:

$$\Phi(x, y) = \Phi_0 + \Delta\Phi_x x + \Delta\Phi_y y \quad (32)$$

where the model parameters Φ_0 , $\Delta\Phi_x$, and $\Delta\Phi_y$ are random variables and have joint probability density functions. For instance, for a given metal wire, if we know that there is a width variation of w_1 and a height variation of w_2 , then the resistance and capacitance of that particular metal wire are:

$$r(w_1, w_2) = r_0 + \Delta r_1 w_1 + \Delta r_2 w_2 = r_0 + \Delta r$$

$$c(w_1, w_2) = c_0 + \Delta c_1 w_1 + \Delta c_2 w_2 = c_0 + \Delta c$$

or in general, for the susceptance and conductance matrices of the interconnect system that are exposed to the process variations, we have:

$$\hat{G}_{var}(w_1, w_2) = \hat{G}_0 + \Delta \hat{G}_1 w_1 + \Delta \hat{G}_2 w_2 = \hat{G} + \Delta \hat{G} \quad (33)$$

$$\hat{L}_{var}(w_1, w_2) = \hat{L}_0 + \Delta \hat{L}_1 w_1 + \Delta \hat{L}_2 w_2 = \hat{L} + \Delta \hat{L} \quad (34)$$

To obtain a balanced truncation technique that takes the process variations into account, we first find the new system matrix \mathbf{A}_{var} (the so-called perturbed system matrix) of the interconnect that is affected by process variation in terms of the ideal system matrix, \mathbf{A} , and the perturbed susceptance and conductance matrices $\Delta \hat{L}$ and $\Delta \hat{G}$. Lemma 1 helps us determine this new system matrix.

Lemma 1. Given an LTI system whose state-space representation is provided by equations (3) and (4), let the susceptance and conductance matrices, \hat{L} and \hat{G} , vary according to equations (33) and (34). If $\|\Delta \hat{L}\|_2 \ll \|\hat{L}\|_2$, then:

$$\mathbf{A}_{var} \approx \mathbf{A} + \Delta \mathbf{A} \quad (35)$$

where

$$\Delta \mathbf{A} = \Delta \mathbf{A}_1 + \Delta \mathbf{A}_2 \quad (36)$$

$$\Delta \mathbf{A}_1 = -L^{-1} \Delta \hat{G} + L^{-1} \Delta \hat{L} L^{-1} \hat{G}$$

$$\Delta \mathbf{A}_2 = L^{-1} \Delta \hat{L} L^{-1} \Delta \hat{G}$$

This lemma enables us to obtain the relationship between the new perturbed system matrix and the original one as well as the incremental variations of the susceptance and conductance matrices.

The balanced realization approach directly utilizes the balancing transformation to project the existing system to a new system whose controllability and observability grammians are identical and diagonalized. The diagonal elements represent all the singular values of the system corresponding to the state variables of the system. The k -th order truncated balanced realization is then obtained by considering the first k singular values. This approach gives more intuitive information about the energy exerted by each state variable and thus the contribution of each state variable on the external behavior of the system, as opposed to the Krylov-subspace methods that involve more abstract computations. From a mathematical viewpoint, the Hankel singular values of the system transfer function are indeed the eigenvalues of the symmetric matrix, \mathbf{PQ} . To realize the effect of interconnect parameter variations on the observability and controllability grammians of the system, we prove the following Theorem.

Theorem 1. Consider a stable LTI system with the state-space representation given by equations (3) and (4). Suppose that the system matrix $\mathbf{A} \in \mathfrak{R}^{n \times n}$ is perturbed by $\Delta \mathbf{A} \in \mathfrak{R}^{n \times n}$. The resulting perturbed system has the following controllability and observability grammians:

$$\mathbf{P}_{var} \leq \Delta \mathbf{Y} \mathbf{P} \Delta \mathbf{Y}^T \quad (37)$$

$$\mathbf{Q}_{var} \leq \Delta \mathbf{Y}^T \mathbf{Q} \Delta \mathbf{Y} \quad (38)$$

where

$$\Delta \mathbf{Y} = \mathbf{I} + (n-1) \cdot \Delta \mathbf{A} \cdot \mathbf{A}^{-1}$$

According to Theorem 1, any perturbation in the system matrix manifests itself as a congruence transformation, $\Delta \mathbf{Y}$, that maps the observability and controllability grammians of the system to the ones for the new perturbed system demonstrated by equations (37) and (38).

To account for the effect of process variations in our proposed model-order reduction technique, we directly utilize Theorem 1 in our algorithm. Moreover, our proposed technique also reshapes the estimated error in the frequency domain using the technique described in section 3 and thus gives rise to a reduced order system with more accurate time and frequency domain responses. Comparing our proposed algorithm with the work in [13] on the variational Krylov-subspace-model-reduction, we can theoretically prove the superiority of the truncated balanced realization technique in terms of its computational complexity. More specifically, the PRIMA-based variational order-reduction proposed in [13] involves the calculation of an expanded Krylov-subspace whose projection matrix is as follows:

$$\mathbf{X}(w_1, w_2) = \mathbf{X}_0 + \Delta \mathbf{X}_{11} w_1 + \Delta \mathbf{X}_{21} w_2 + \Delta \mathbf{X}_{12} w_1^2 + \Delta \mathbf{X}_{22} w_2^2$$

where w_1 and w_2 are dimensional variations. $\Delta \mathbf{X}_{ij}$'s are computed by choosing a set of sample points. A variational reduced-order model can then be constructed by inserting the resulting variational Krylov-subspace into the PRIMA equations. The geometrical variations of interconnects are random processes. So are the corresponding variations on the interconnect electrical parameters. Therefore, we have to consider an interval for the dimensional variations. Hence the whole model-reduction technique would be very time-consuming and even impossible to use when handling very large circuits.

There are, however, some problems that need to be taken into account:

- We still need to solve the Lyapunov equations to obtain the grammians of the system. To efficiently solve the Lyapunov equations we make use of the ADI procedure [11], which is an iterative method for solving a Lyapunov equation,

$$\mathbf{A} \mathbf{P} + \mathbf{P} \mathbf{A}^T + \mathbf{X} = 0$$

The system is first reduced to tridiagonal form with Gaussian similarity transformation, \mathbf{T}_{trid} , as follows:

$$\mathbf{S} = \mathbf{T}_{trid} \mathbf{A} \mathbf{T}_{trid}^{-1}$$

$$\mathbf{Z} = \mathbf{T}_{trid} \mathbf{P} \mathbf{T}_{trid}^T$$

$$\mathbf{X}_s = \mathbf{T}_{trid} \mathbf{X} \mathbf{T}_{trid}^T$$

Reducing a matrix to tridiagonal form has an $O(n^3)$ order of complexity. The resulting system is solved with the ADI iteration [11]:

$$\mathbf{Z}_0 = 0$$

$$(\mathbf{S} + p_j \mathbf{I}) \mathbf{Z}_{j-\frac{1}{2}} = \mathbf{X}_s - [(\mathbf{S} - p_j \mathbf{I}) \mathbf{Z}_{j-1}]^T,$$

$$(\mathbf{S} + p_j \mathbf{I}) \mathbf{Z}_j = \mathbf{X}_s - [(\mathbf{S} - p_j \mathbf{I}) \mathbf{Z}_{j-\frac{1}{2}}]^T$$

for $j = 1, 2, \dots, J$

An iterative solution of the reduced Lyapunov equation is accomplished in $O(12Jn^2)$ flops [11]. In the above iteration, $\{p_j\}_{j=1}^J$ are the ADI parameters. We use the same optimal ADI parameters that were presented in [11].

- The calculations required to construct the balancing transformation \mathbf{T} are complicated and sensitive to numerical errors. In particular, the balancing transformation \mathbf{T} may be poorly conditioned when matrix \mathbf{PQ} has a high condition number¹. Paper [15] proposes an algorithm where the balancing is avoided altogether, and as a result the numerical difficulties are never encountered. However, this algorithm involves solving an eigenvalue problem whose dimension is as large as the order of the original transfer function.

We can avoid the problem of solving a large eigenvalue problem by using the Krylov subspace-based methods. In fact revisiting the balanced realization method reveals that it is only necessary to find the first k largest eigenvalues of the matrix product \mathbf{PQ} and their corresponding left and right eigenvectors. Based on this observation, a modified version of Safanov's algorithm is utilized here. Recall that \mathbf{PQ} is a large symmetric matrix, which will also be a positive definite

1. Recall that the condition number $cond(M) = \sigma_{max}(M)/\sigma_{min}(M)$ provides a measure of the distance of M to the set of singular matrices.

matrix. Thus the problem is to efficiently obtain the largest eigenvalues of a symmetric matrix. This problem is solved by using the *Lanczos* method. A complete explanation of the Lanczos method can be found in [5] and [14]. In our algorithm, we utilize a modified version of Safanov's algorithm using the Lanczos method to obtain the balancing transformation.

- To account for the effect of process variations on the interconnect delay and performance, Theorem 1 is directly utilized. However, there is a matrix inverse involved in equations (37) and (38). To compute the inverse of a matrix more efficiently, we elaborate the problem in the form of solving a set of linear algebraic equations rather than explicit inverse formation.

Now we proceed with describing our new algorithm for variational balanced truncation.

Given $\mathbf{H}(s)$, $\mathbf{W}_i(s)$, $\mathbf{W}_o(s)$:

1. Using lemma 1, Determine matrix $\Delta \mathbf{A}$.
2. Using Theorem 1, compute perturbing congruence transformation, $\Delta \mathbf{Y}$.
3. Using the ADI iteration, compute the controllability and observability grammians of $\bar{\mathbf{P}}$ and $\bar{\mathbf{Q}}$ of the original spectrally-weighted system by solving the Lyapunov equations (19) and (20).
4. Apply perturbing congruence transformation $\Delta \mathbf{Y}$ on the $n \times n$ upper left corner blocks of $\bar{\mathbf{P}}$ and $\bar{\mathbf{Q}}$ specified by matrices \mathbf{P} and \mathbf{Q} in Eq. (18), and construct new perturbed controllability and observability grammians, \mathbf{P}_{var} and \mathbf{Q}_{var} .
5. Use equations (23) and (24) to compute \mathbf{X}_{var} and \mathbf{Y}_{var} from \mathbf{P}_{var} and \mathbf{Q}_{var} .
6. Decompose \mathbf{X}_{var} and \mathbf{Y}_{var} using the eigenvalue decomposition technique into $\mathbf{U}_{var} \mathbf{S}_{var} \mathbf{U}_{var}^T$ and $\mathbf{V}_{var} \mathbf{Z}_{var} \mathbf{V}_{var}^T$.
7. Use equations (27) and (28) to compute $\bar{\mathbf{B}}_{var}$ and $\bar{\mathbf{C}}_{var}$.
8. Using the ADI iteration, solve the mapped Lyapunov equations (29) and (30) for the new perturbed system to compute $\hat{\mathbf{P}}_{var}$ and $\hat{\mathbf{Q}}_{var}$.
9. Using the Lanczos algorithm, obtain the reduced order left and right eigen-matrices, $\hat{\mathbf{V}}_{L,var}$ and $\hat{\mathbf{V}}_{R,var}$, associated with the k largest singular values of the matrix product $\hat{\mathbf{P}}_{var} \hat{\mathbf{Q}}_{var}$.
10. Let $\hat{\mathbf{E}}_{var} = \hat{\mathbf{V}}_{L,var}^T \hat{\mathbf{V}}_{R,var}$ and compute the singular-value decomposition of $\hat{\mathbf{E}}_{var}$, i.e., $\hat{\mathbf{E}}_{var} = \hat{\mathbf{U}}_{E,var} \hat{\mathbf{\Sigma}}_{E,var} \hat{\mathbf{V}}_{E,var}^T$.
11. Let $\hat{\mathbf{S}}_{L,var} = \hat{\mathbf{V}}_{L,var} \hat{\mathbf{U}}_{E,var} \hat{\mathbf{\Sigma}}_{E,var}^{-1/2} \in \mathfrak{R}^{n \times k}$ and $\hat{\mathbf{S}}_{R,var} = \hat{\mathbf{V}}_{R,var} \hat{\mathbf{V}}_{E,var} \hat{\mathbf{\Sigma}}_{E,var}^{-1/2} \in \mathfrak{R}^{n \times k}$.
12. Compute the reduced order space realization using the matrices $\hat{\mathbf{S}}_{L,var}$ and $\hat{\mathbf{S}}_{R,var}$ as follows:

$$\begin{bmatrix} \hat{\mathbf{A}}_{var} & \hat{\mathbf{B}}_{var} \\ \hat{\mathbf{C}}_{var} & \hat{\mathbf{D}}_{var} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{S}}_{L,var} \mathbf{A}_{var} \hat{\mathbf{S}}_{R,var} & \hat{\mathbf{S}}_{L,var}^T \mathbf{B} \\ \mathbf{C} \hat{\mathbf{S}}_{R,var} & \mathbf{D} \end{bmatrix} \quad (39)$$

Now we prove that this new variational, spectrally-weighted balanced truncation has a provable L^∞ -error bound.

Theorem 2. The L^∞ -error of the variational model reduction with spectral shaping is:

$$\|\mathbf{W}_o(s)(\mathbf{H}_{var}(s) - \mathbf{H}_{var}^k(s))\mathbf{W}_i(s)\|_\infty \leq k \left(\sum_{i=k+1}^n \sigma_i \right) \quad (40)$$

where $k = \|\mathbf{W}_o(s)\mathbf{L}\|_\infty \|\mathbf{K}\mathbf{W}_i(s)\|_\infty$ and

$$\mathbf{K} = \text{diag}(|s_1|^{-1/2}, \dots, |s_n|^{-1/2}, 0, \dots, 0) \mathbf{U}^T \mathbf{B}$$

$$\mathbf{L} = \mathbf{C} \mathbf{V} \text{diag}(|z_1|^{-1/2}, \dots, |z_n|^{-1/2}, 0, \dots, 0)$$

As mentioned earlier, process variations will have a definite impact on the on-chip interconnect parasitics. These variations have two adverse effects on the system matrix as also demonstrated by Eq. (36). Ignoring second-order variations, any incremental increase in the values of parasitic

conductances and resistances reduces $\Delta \mathbf{A}$, whereas any incremental increase in the values of parasitic inductances and capacitances increases $\Delta \mathbf{A}$.

As an interesting special case, consider the tightly coupled RLC interconnects. It is easily proved that, for this case, the system matrix \mathbf{A} is a symmetric positive-definite matrix. Under these circumstances, the following theorem proves useful in finding the upper and lower limits of variations in the poles of the system transfer function of tightly coupled RLC interconnects that are subject to the process variations:

Theorem 3 [14]. Given an LTI system whose state-space representation is given by equations (3) and (4), if the system matrix $\mathbf{A} \in \mathfrak{R}^{n \times n}$ is perturbed by $\Delta \mathbf{A} \in \mathfrak{R}^{n \times n}$ due to the process variation, then the following inequalities hold:

$$|\lambda_k(\mathbf{A} + \Delta \mathbf{A}) - \lambda_k(\mathbf{A})| \leq \|\Delta \mathbf{A}\|_2 \quad (41)$$

where $\Delta \mathbf{A}$ is given by lemma 1.

According to Theorem 3, the magnitude of difference between poles of the perturbed system and those of the original system is limited by the 2-norm of the perturbed matrix, $\Delta \mathbf{A}$.

5. EXPERIMENTAL RESULTS

In this section the variational spectrally-weighted balanced truncation model-order reduction technique (referred to as VSWBT) is evaluated by performing experiments on some global interconnect lines such as clock trees, and coupled lines. First, the spectrally-weighted balanced truncation (SWBT) technique is demonstrated and the result of applying this algorithm is compared with those obtained by utilizing our implementations of PRIMA [6] and the truncated balanced realization using Vector ADI (VADI) [9]. We then apply the VSWBT method to study the impact of the interconnect process variations on the timing performance of the clock trees and coupled busses. Finally the accuracy of VSWBT is validated by using it to reduce the order of an arbitrary stable LTI system that is subject to the perturbation.

5.a. Two capacitively coupled interconnects

We compare the performances of the model reduction of two capacitively-coupled transmission lines. The order of the reduced order system is 3. The circuit is depicted in Fig. 1. The Bode diagram of the original system along with the reduced systems obtained by using three approaches (PRIMA, VADI, SWBT) are also shown in Fig. 2. The output weighting function is:

$$W_o(s) = \frac{10}{s + 0.2}$$

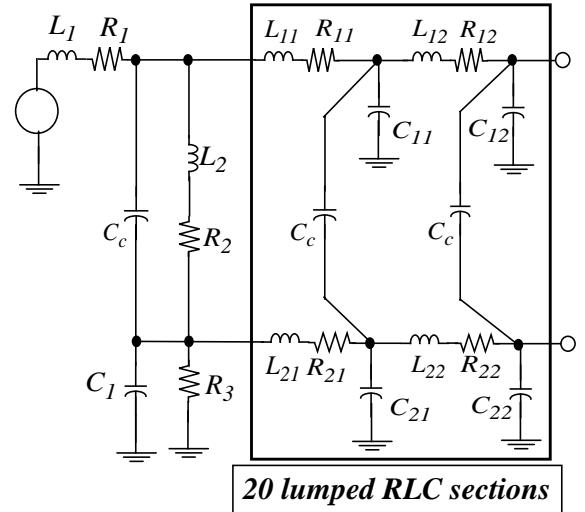


Fig. 1. Two capacitively coupled RLC interconnects consisting of 20 RLC lumped sections.

From Fig. 2, we observe that the frequency response of the reduced order system obtained by SWBT closely follows the frequency response of the original system, especially in the low frequency range. Comparing the bode diagrams clearly indicates that SWBT is much more accurate than PRIMA and also VADI in generating the reduced-order model.

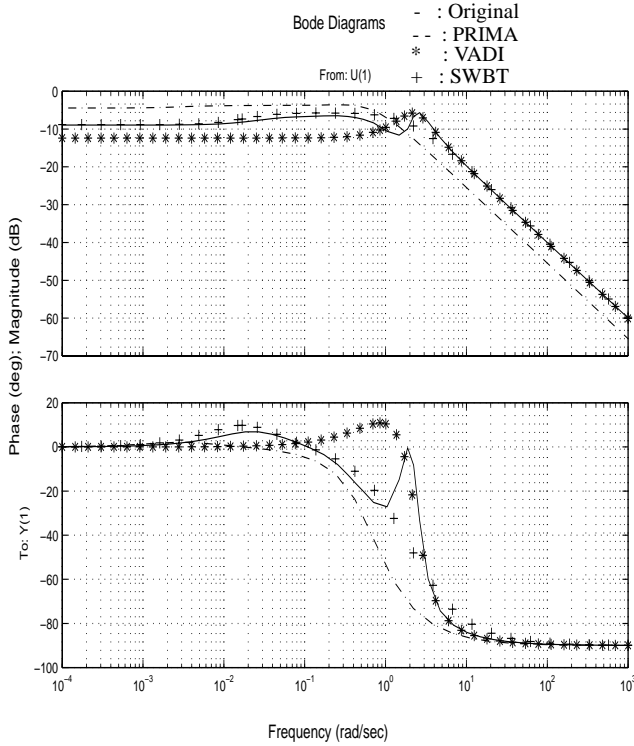


Fig. 2. The Bode diagram of the system in Fig. 1.

5.b. H-tree clock distribution

The clock tree is routed using the TSMC 0.25 μ digital CMOS technology library. The clock tree is an H-tree clock distribution with tapered buffers at the root of the tree as shown in Fig. 3. The design target of the on-chip clock frequency is 1.0 GHz. The clock tree is modeled by a large RLC circuit. Each wire segment in Fig. 3 is modeled by a large RLC π network whose R, L, and C values are changed due to metal width and ILD thickness variations. Based on the data reported in paper [17] on the interconnect-dominated test circuits, the typical variational distribution for metal interconnects as well as ILD thicknesses is a normal distribution. The widths of metal and ILD layers are changing up to 30% of their nominal values (which is the magnitude of the percentage of 3- σ variations).

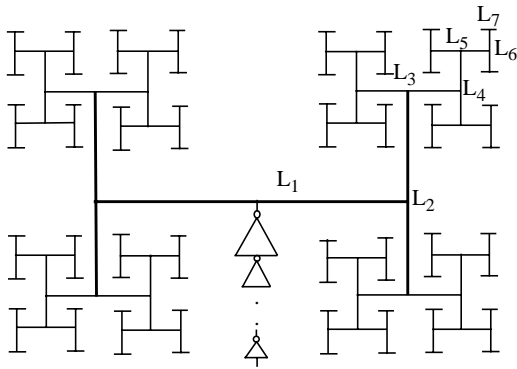


Fig. 3. The H-tree clock distribution driven by a tapered buffer.

We performed 20 experiments where, in each experiment, a set of normally-distributed numbers for the metal and ILD width variations were generated. The 50% delay at an arbitrarily chosen fan-out (leaf) node was computed by VSWBT and compared with the result obtained by [13]. The results of this comparison for different experiments are provided in Table 1. Here, we assume that the geometrical variations are mutually independent.

Table 1: COMPARISON BETWEEN SIMULATION RESULTS ON AN H-TREE CLOCK NET WITH PROCESS VARIATIONS (Star-HSPICE LEVEL 49, 0.25 μ CMOS PROCESS). DELAYS ARE GIVEN IN psec

3- σ variations (%)		50% delay of the fan-out node		
ILD width	Metal Layer width	Paper [13] (psec)	VSWBT (psec)	Star-HSPICE (psec)
12.3	-10.3	195.7	196.0	196.3
5.8	19.8	215.5	216.4	216.2
-21.4	-2.3	199.3	200.1	200.1
-10.2	-8.2	205.0	208.2	206.6
2.45	22.1	201.2	204.1	203.2
19.2	10.6	193.1	195.0	194.2
-24.3	-17.8	198.2	200.3	200.0
28.1	11.4	220.3	222.5	221.7
-12.4	-5.9	219.5	221.9	220.9
-14.9	-2.5	194.1	196.2	195.6
-1.3	12.6	211.6	213.2	212.5
25.7	-7.2	203.2	207.3	206.4
7.6	1.6	210.7	217.7	215.3
-2.3	-29.1	223.2	233.4	231.8
20.1	-18.5	206.4	211.8	210.5
-18.8	13.7	196.2	201.0	200.3
22.4	-4.5	200.0	204.7	203.3
-27.6	13.6	191.7	193.1	192.8
5.3	14.3	216.5	218.2	217.8
-13.4	-10.6	195.5	198.1	197.5

From Table 1, it is clear that our approach can predict the 50% delays more accurately for all possible 3- σ variations of the metal and ILD layers that are indicated in Table 1. By comparing the number of flops from MATLAB simulation, the CPU-time for our method is, on average, 5% faster than the work reported in [13].

5.c. Two capacitively coupled interconnects with variational parameters

In this example we apply the VSWBT model reduction technique on two capacitively-coupled microstrip lines whose electrical parameters are subject to process variations. The schematic of these two coupled lines along with their nominal geometrical parameters are depicted in Fig. 4.

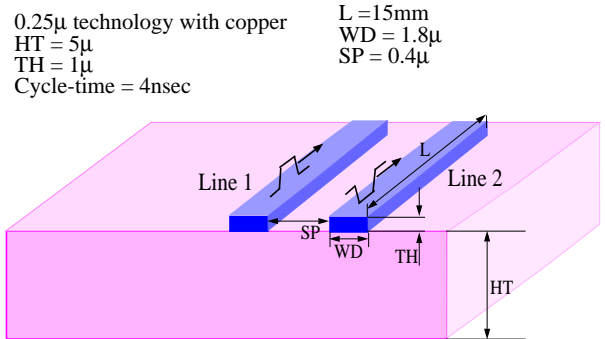


Fig. 4. Two parallel microstrip lines in 0.25 μ CMOS technology.

The height and width of the metal layer are subject to $\pm 30\%$ variations. Again we performed 20 experiments with normally distributed random variations. The result of these 20 experiments is shown in Table 2. The MATLAB-reported flop usage of our method is 10% less than that of [13]. From Table 2, it can be seen that, for all variations demonstrated in Table 2, the 50% delays predicted by VSWBT are more accurate than those predicted by paper [13].

Table 2: COMPARISON BETWEEN SIMULATION RESULTS ON TWO CAPACITIVELY COUPLED MICROSTRIP LINES (Star-HSPICE LEVEL 49, 0.25 μ CMOS PROCESS). DELAYS ARE GIVEN IN psec

3- σ variations (%)		50% delay of the far-end node of Line 2 (psec)		
Metal W	Metal H	Paper [13] (psec)	VSWBT (psec)	Star-HSPICE (psec)
-0.4	20.1	1592	1597	1596
15.8	-1.9	1636	1640	1639
-9.3	13.2	1519	1525	1524
-1.75	10.8	1607	1612	1610
15.8	-14.4	1690	1700	1697
-13.2	-0.9	1726	1734	1731
19.1	-29.1	1711	1725	1722
29.3	2.5	1531	1538	1536
-1.8	7.9	1603	1610	1607
11.5	-4.2	1673	1683	1680
-15.3	19.4	1650	1665	1662
14.7	25.3	1557	1567	1563
1.1	-10.4	1662	1667	1665
-15.6	22.1	1704	1717	1711
18.2	-23.5	1622	1631	1629
-13.2	-0.5	1699	1705	1703
-9.4	21.7	1640	1644	1643
0.7	-11.1	1708	1713	1712
10.8	-0.7	1621	1630	1627

5.d. A general example

To demonstrate the accuracy and validity of VSWBT on any arbitrary stable LTI system exposed to perturbations, we started from an arbitrary state-space specification of an LTI system with 20 poles and 13 zeros, which is shown in the pole-zero map of Fig. 5. Suppose that each and every element of the system matrix \mathbf{A} experiences a normally distributed random perturbation with a standard deviation of 30% around the nominal value. The order of the reduced order system is 3. We would also like to have high accuracy in high frequencies, therefore the following weighting function is introduced: $W_i(s) = 14/(s + 14)$

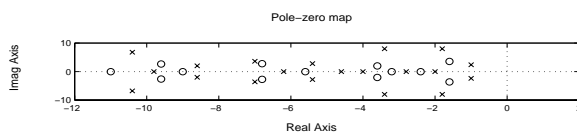


Fig. 5. The pole-zero map of the LTI system specified in section 5.d.

A variational spectrally-weighted reduced-order model is derived using two approaches. The first approach is to calculate the perturbed system matrix and then directly apply the spectrally-weighted balanced truncation. The second approach is to use the VSWBT technique.

Fig. 6 shows and compares the bode diagram of the magnitude response, the bode diagram of phase response, the impulse response of the original perturbed system, the reduced-order system using direct order reduction technique, and the reduced-order system using VSWBT algorithm. As can be seen from Fig. 6, the difference between the two order reduction approaches is indistinguishable. Remember that applying the direct mode-order reduction technique on a large system of parameter-varying interconnects to generate the reduced order system is very time consuming. The reason for this is that the direct order reduction algorithm must be run on the system each time under different offset values for electrical parameters of the interconnect.

6. CONCLUSIONS

In this paper a variational spectrally-weighted balanced truncation technique for the model-order reduction of geometrically varying multiport RLC interconnects was proposed. It was shown that the balanced truncation technique is a highly effective approach when the variations in the circuit parameters need to be taken into consideration. Various experiments demonstrate that the computational requirement of the variational spectrally-weighted balanced truncation approach is 5% lower than that of the

variational Krylov-subspace-based model-order reduction techniques while the accuracy is also 20% higher, on average.

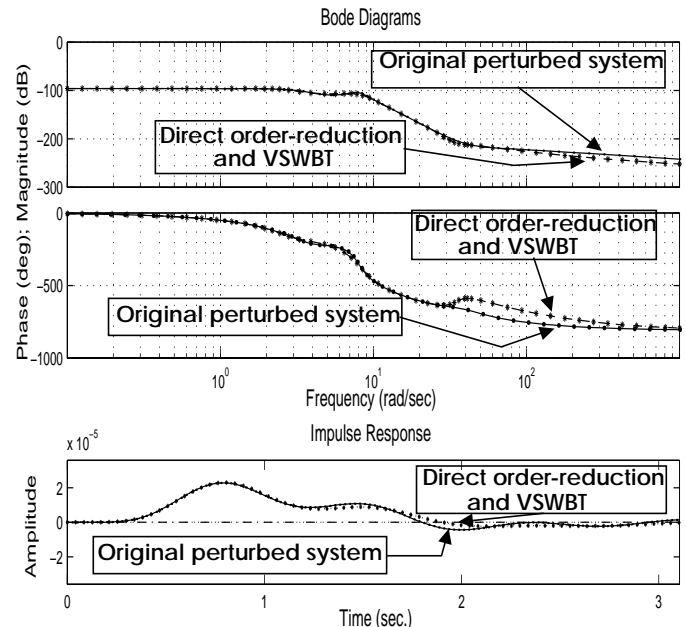


Fig. 6. The bode diagram and impulse response of the original perturbed system, the system obtained by direct spectrally-weighted balanced truncation, and the VSWBT algorithm.

7. REFERENCES

- [1] S. Oh, W. Jung, J. Kong, K. Lee, "Interconnect Modeling for VLSI," *International Conf. on Simulation of Semiconductor Processes and Devices*, pp. 203-206, 1999.
- [2] L. T. Pillage, R. A. Rohrer, "Asymptotic Waveform Evaluation for Timing Analysis," *IEEE Trans. CAD*, vol. 9, No. 4, pp. 352-366, April 1990.
- [3] C. Ratzlaff, L. T. Pillage, "RICE: Rapid Interconnect Circuit Evaluation Using AWE," *IEEE Trans. CAD*, vol. 13, No. 6, pp. 763-776, June 1994.
- [4] K. J. Kerns, A. T. Yang, "Stable and Efficient Reduction of Large, Multiport RC Networks by Pole Analysis via Congruence Transformations," *IEEE Trans. CAD*, vol. 16, pp. 734-744, 1997.
- [5] P. Feldmann, R. W. Freund, "Efficient Linear Circuit Analysis by Pade Approximation via the Lanczos Process," *IEEE Trans. CAD*, vol. 14, pp. 639-649, May 1995.
- [6] A. Odabasioglu, M. Celik, L. T. Pileggi, "PRIMA: Passive Reduced-Order Interconnect Macromodeling Algorithm," *IEEE Trans. CAD*, vol. 17, No. 8, pp. 645-654, Aug. 1998.
- [7] K. Glover, "All Optimal Hankel-norm Approximation of Linear Multivariable Systems and Their L^∞ -error Bounds," *Int. J. Control*, vol. 39, No. 6, pp. 1115-1193, 1984.
- [8] P. Rabiei, M. Pedram, "Model Order Reduction of Large Circuits Using Balanced Truncation," *IEEE Proc. ASP-DAC*, pp. 237-240, 1999.
- [9] J. Li, F. Wang, J. White, "An Efficient Lyapunov Equation-Based Approach Generating Reduced-Order Models of Interconnect," *36th ACM/IEEE Design Automation Conference*, pp. 1-6, 1999.
- [10] J. Li, J. White, "Efficient Model Reduction of Interconnect via Approximate System Grammians," *IEEE/ACM Proc. ICCAD*, pp. 380-384, 1999.
- [11] A. Lu, E. L. Wachspress, "Solution of Lyapunov Equations by Alternating Direction Implicit," *Computers and Mathematics with Applications*, vol. 21, No. 9, pp. 43-58, June 1991.
- [12] P. Heydari, M. Pedram, "Balanced Truncation with Spectral Shaping for RLC Interconnects," *IEEE Proc. ASP-DAC*, pp. 203-208, Jan. 2001.
- [13] Y. Liu, L. T. Pileggi, A. Strojwas, "Model Order-Reduction of RC(L) Interconnect including Variational Analysis," *36th IEEE/ACM Proc. Design Automation Conf.*, pp. 201-206, June 1999.
- [14] G. H. Golub, C. Van Loan, *Matrix Computation*, pp. 390-405, Johns Hopkins University Press, 1996.
- [15] M. G. Safanov, R. Y. Chiang, "A Schur Method for Balanced Truncation Model Reduction," *American Control Conference*, vol. 2, pp. 1036-1040, 1988.
- [16] O. S. Nakagawa, S. -Y. Oh, G. Ray, "Modeling of Pattern-Dependent On-Chip Interconnect Geometry Variation of Deep-Submicron Process and Design Technology," *IEEE International Technical Digest On Electron Devices Meeting*, pp. 137-141, 1997.
- [17] O. S. Nakagawa, N. Chang, S. Lin, D. Sylvester, "Circuit Impact and Skew Corner Analysis of Stochastic Process Variation in Global Interconnect," *IITC*, pp. 230-232, 1999.
- [18] P. J. Antsaklis, A. N. Micel, *Linear Systems*, New York: McGraw- Hill, 1997.
- [19] S. R. Nassif, "Modeling and Forecasting of Manufacturing Variations," *5th Int'l Workshop Statistical Metrology*, pp. 2-10, 2000.