

Macromodel Generation for BioMEMS Components Using a Stabilized Balanced Truncation Plus Trajectory Piecewise-Linear Approach

Dmitry Vasilyev, *Student Member, IEEE*, Michał Rewieński, *Member, IEEE*, and Jacob White, *Associate Member, IEEE*

Abstract—In this paper, we present a technique for automatically extracting nonlinear macromodels of biomedical microelectromechanical systems devices from physical simulation. The technique is a modification of the recently developed trajectory piecewise-linear approach, but uses ideas from balanced truncation to produce much lower order and more accurate models. The key result is a perturbation analysis of an instability problem with the reduction algorithm, and a simple modification that makes the algorithm more robust. Results are presented from examples to demonstrate dramatic improvements in reduced model accuracy and show the limitations of the method.

Index Terms—Biomedical microelectromechanical devices (bioMEMS), microelectromechanical devices (MEMS), model order reduction, nonlinear dynamical systems, perturbation methods, piecewise linear models.

I. INTRODUCTION

THE application of micromachining to biological applications, such as labs-on-a-chip [1]–[3], require complicated combinations of individual biomedical microelectromechanical systems (bioMEMS) devices that process fluids, cells, and molecules (e.g., mixers, separators, and pumps). In order to simulate systems of these devices, models have been developed for common components, such as mixers and separators [4]–[7]. The wide variety of devices currently in development, and the need to rapidly assess the impact of candidate device performance on system behavior, will accelerate the demand for techniques that more automatically extract models of these bioMEMS devices from detailed physical simulation. The required automatic techniques may include approaches similar to the robust nonlinear model-order reduction (MOR) strategies being developed for nonlinear circuit model reduction [8]–[12], though bioMEMS devices can be more challenging because

Manuscript received March 5, 2005; revised June 28, 2005. This work was supported by National Science Foundation, by the Defense Advanced Research Projects Agency (DARPA) under the NeoCAD program, by the Semiconductor Research Corporation, and by the Singapore–Massachusetts Institute of Technology (MIT) alliance. This paper was recommended by Associate Editor J. Zeng.

D. Vasilyev and J. White are with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: vasilyev@mit.edu; white@mit.edu).

M. Rewieński was with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. He is now with Synopsys Inc., Mountain View, CA 94043 USA (e-mail: Michal.Rewienski@synopsys.com).

Digital Object Identifier 10.1109/TCAD.2005.857389

they are both nonlinear and typically much less damped than circuits.

In this short paper, we describe an effective model reduction algorithm for bioMEMS devices that is a modification of the trajectory piecewise-linear (TPWL) MOR algorithm [10]. In the following section, we describe the TPWL MOR algorithm, and then in Section III, we present an improvement on that algorithm based on using truncated balanced realization (TBR) [13]. In Section IV, we describe several example problems, and in Section V, we use those examples to demonstrate both the effectiveness of our TPWL-TBR algorithm, as well as an instability problem. Also, we demonstrate a fundamental difficulty of the TPWL approach when applied to traveling-wave problems. In Section VI, we describe a perturbation analysis of the instability problem, and give a second algorithm modification that resolves this problem. Conclusions end the paper.

II. TPWL NONLINEAR MODEL REDUCTION

After spatial discretization of the coupled partial differential equations (PDEs) that describe a bioMEMS component, the dynamic behavior of the component can often be represented using the standard state-space form

$$\begin{cases} \dot{x}(t) = f(x(t), u(t)) \\ y(t) = Cx(t) \end{cases} \quad (1)$$

where $x(t) \in R^N$ is a vector of states (e.g., mechanical displacements, fluid velocities) at time t , $f: R^N \times R^M \rightarrow R^N$ is a nonlinear vector-valued function, $u: R \rightarrow R^M$ is an input signal, C is an $N \times K$ output matrix and $y: R \rightarrow R^K$ is the output signal.

We assume nonlinear function f being differentiable for all values of x and u :

$$f(x, u) = f(x_0, u_0) + A(x - x_0) + B(u - u_0) + \text{h.o.t.} \quad (2)$$

where matrices A and B [which are dependent on the linearization point (x_0, u_0)] contain derivatives of f with respect to the components of the state and input signals, respectively.

The goal of applying MOR to (1) is to construct a macromodel capable of approximately simulating the input–output behavior of the systems in (1), but at a significantly reduced computational cost.

In order to achieve this goal, one needs to reduce the dimensionality of the state-space vector, which is usually achieved by employing projections. However, only projecting the nonlinear system (1) is not a complete solution to the nonlinear model reduction problem, because direct evaluation of the projected function is still directly proportional to the size of the unreduced system, and is too computationally expensive.¹

To reduce the cost of the projected-function evaluation, consider the following generalized quasi-piecewise-linear approximate representation of the nonlinear function f , which has been proposed, in a slightly simpler form, in [10]

$$f(x, u) \approx \sum_{i=1}^s \tilde{w}_i(x, u) (f(x_i, u_i) + A_i(x - x_i) + B_i(u - u_i)) \quad (3)$$

where x_i 's and u_i 's ($i = 1, \dots, s$) are selected linearization points (samples of state and input values), A_i and B_i are derivatives of f with respect to x and u , evaluated at (x_i, u_i) , and finally, $\tilde{w}_i(x, u)$'s are state- and input-dependent weights that satisfy

$$\sum_{i=1}^s \tilde{w}_i(x, u) = 1 \quad \forall (x, u),$$

$$\tilde{w}_i(x, u) \rightarrow 1 \quad \text{as } (x, u) \rightarrow (x_i, u_i). \quad (4)$$

Equation (4) implies that the TPWL approximation in (3) is simply a convex combination of samples of f and f 's derivatives.

Projecting the piecewise-linear approximation in (3) using biorthonormal projection bases V and W yields the following reduced-order nonlinear dynamical system:

$$\begin{cases} \dot{z} = \gamma \cdot w(z, u) + (\sum_{i=1}^s w_i(z, u) A_{ir}) z \\ \quad + (\sum_{i=1}^s w_i(z, u) B_{ir}) u \\ y = C_r z \end{cases} \quad (5)$$

where $z(t) \in R^q$ is the q -dimensional vector of states

$$\gamma = [W^T (f(x_1, u_1) - A_1 x_1 - B_1 u_1) \dots W^T (f(x_s) - A_s x_s - B_s u_s)].$$

Here, $w(z, u) = [w_1(z, u) \dots w_s(z, u)]^T$ is a vector of weights, $A_{ir} = W^T A_i V$, $B_{ir} = W^T B_i$, and $C_r = C V$. One should note that $\sum_{i=1}^s w_i(z, u) = 1$ for all (z, u) , $w_i \rightarrow 1$ as $(z, u) \rightarrow (W^T x_i, u)$, and that the evaluation of the right-hand side of (5) requires at most $O(sq^2)$ operations, where s is the number of linearization points.

As proposed in [10] and [12], linearization points (x_i, u_i) used in system (5) are usually selected from a "training trajectory" of the initial nonlinear system, corresponding to some appropriately determined "training input." The choice of the training input is an important aspect of the reduction procedure, since this choice directly influences accuracy. As the general rule, the training signal should be as close as possible to the

signals for which the reduced system will be used. Additionally, this input signal should be rich enough to collect all "important" states in the set of linearization points (x_i, u_i) [12].

In order to obtain a reduced system in the form of (5), biorthonormal projection bases V and W must also be determined. This issue is addressed below.

III. CHOICE OF LINEAR REDUCTION METHOD

Consider a simple linearization of (1) about the initial state (x_0, u_0)

$$\begin{cases} \dot{x} = f(x_0) - \hat{A}x_0 - \hat{B}u_0 + \hat{A}x + \hat{B}u \\ y = \hat{C}x \end{cases} \quad (6)$$

For the system in (6), a projection basis can be obtained using one of the many projection-based linear MOR procedures. One common choice is to reduce using the projection basis spanning the Krylov subspace [10], [14], [15]

$$\text{span}(V) = \text{span}\{\hat{A}^{-1}\hat{B}, \dots, \hat{A}^{-q}\hat{B}\}.$$

Reduction of (6) using the Krylov-subspace projection is not guaranteed to provide a stable reduced model, even in this linearized case [16], [17]. Therefore, TPWL macromodels obtained using Krylov projection are not guaranteed to be stable even if the original system is nearly linear.

Alternatively, one can apply a balanced truncation model reduction (TBR) procedure [18]–[20]:

Algorithm 1: TBR

Input: System matrices \hat{A} , \hat{B} , and \hat{C} .

Output: Projection bases V and W .

- (1) Find observability gramian P : $\hat{A}P + P\hat{A}^T = -\hat{B}\hat{B}^T$;
- (2) Find controllability gramian Q : $\hat{A}^T Q + Q\hat{A} = -\hat{C}^T \hat{C}$;
- (3) Compute q dominant eigenvectors of PQ : $(PQ)V = V\Sigma^2$, where $\Sigma^2 = \text{diag}(\Lambda_q^{\text{dom}}(PQ))$;
- (4) Compute q dominant eigenvectors of QP : $(QP)W = W\Sigma^2$ and scale columns of W such that $W^T V = I_{q \times q}$.

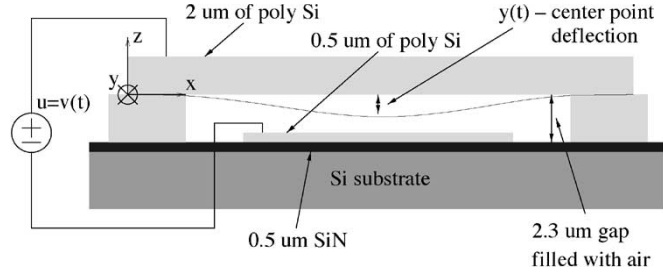
The projection bases V and W obtained using algorithm 1 can then be used to compute the reduced TPWL approximation in (5).

TBR reduction can be more accurate than Krylov-subspace reduction as it possesses a uniform frequency error bound [21], and TBR preserves the stability of the linearized model. This superior performance for the linear cases suggests that TPWL approximation models obtained using TBR will be stable and accurate as well. This is not necessarily the case, as will be shown below.

IV. EXAMPLES OF NONLINEAR SYSTEMS

In this section, we consider two examples of nonlinear systems that arise in the modeling of bioMEMS devices that have nonlinear dynamical behaviors, which make good test cases for reduction algorithms.

¹This issue is discussed in detail in [10].

Fig. 1. Micropump example (following Hung *et al.* [22]).

The first example is a fixed–fixed beam structure, which might be used as part of a micropump or valve, shown in Fig. 1. Following Hung *et al.* [22], the dynamical behavior of this coupled electromechanical–fluid system can be modeled with a one-dimensional (1-D) Euler’s beam equation and the two-dimensional (2-D) Reynolds’ squeeze film damping equation [22]:

$$\begin{cases} \hat{E}I \frac{\partial^4 w}{\partial x^4} - S \frac{\partial^2 w}{\partial x^2} = F_{\text{elec}} + \int_0^d (p - p_0) dy - \rho \frac{\partial^2 w}{\partial t^2} \\ \nabla \cdot ((1 + 6K)w^3 p \nabla p) = 12\mu \frac{\partial(pw)}{\partial t}. \end{cases} \quad (7)$$

Here, the axes x , y , and z are as shown on Fig. 1, \hat{E} is the Young’s modulus, I is the moment of inertia of the beam, S is the stress coefficient, K is the Knudsen number, d is the width of the beam in the y -direction, $w = w(x, t)$ is the height of the beam above the substrate, and $p(x, y, t)$ is the pressure distribution in the fluid below the beam. The electrostatic force is approximated assuming nearly parallel plates and is given by $F_{\text{elec}} = \epsilon_0 d v^2 / 2w^2$, where v is the applied voltage.

Spatial discretization of (7) using a standard finite-difference scheme leads to a nonlinear dynamical system in the form of (1), with $N = 880$ states. After discretization, the state vector x consists of the concatenation of: heights of the beam above the substrate w , values of $\partial(w^3)/\partial t$, and values of the pressure below the beam. For the considered example, the output $y(t)$ was selected to be the deflection of the center of the beam from the equilibrium point (cf. Fig. 1).

The remarkable feature of this example is that the system is strongly nonlinear, and no feasible Taylor expansion made at the initial state can correctly represent the nonlinear function f , especially in the so-called pull-in region² [22].

The exact actuation mechanism of the real micropumps may be quite different than the above simple structure, but this example is illustrative in that it combines electrical actuation with the structural dynamics and is coupled to fluid compression. We expect model reduction methods that are effective for this example problem to be extendable to realistic micropumps.

The second example, suggested in [23], is the injection of a (marker) fluid into a U-shaped three-dimensional microfluidic channel. The fluid is driven electrokinetically as depicted in Fig. 2, and the channel has a rectangular cross section of height d and width w . In this example, the electrokinetically driven

²If the beam is deflected by more than $\approx 1/3$ of the initial gap, the beam will be pulled-in to the substrate.

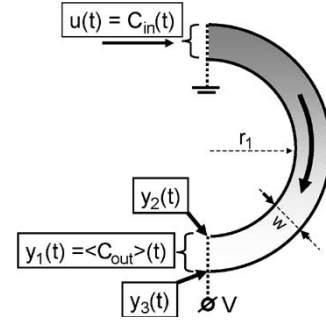


Fig. 2. Microfluidic channel.

flow of a buffer (carrier) fluid was considered to be steady, with the fluid velocity directly proportional to the electric field as in

$$\vec{v}(x, y, z) = -\mu \nabla \Phi(\vec{r})$$

where μ is the electroosmotic mobility of the fluid. The electric field can be determined from Laplace’s equation

$$\nabla^2 \Phi(\vec{r}) = 0$$

with Neumann boundary conditions on the channel walls [24]. If the concentration of the marker is not small, the electroosmotic mobility can become dependent on the concentration, i.e., $\mu \equiv \mu(C(\vec{r}, t))$, where $C(\vec{r}, t)$ is the concentration of the marker fluid. Finally, the marker can diffuse from the areas with high concentration to the areas with low concentration. The total flux of the marker, therefore, is

$$\vec{J} = \vec{v}C - D\nabla C \quad (8)$$

where D is the diffusion coefficient of the marker. Again, as the concentration of the marker grows, the diffusion will be governed not only by the properties of the carrying fluid, but also by the properties of the marker fluid, therefore, D can depend on concentration. Conservation applied to the flux (8) yields a convection–diffusion equation [25]

$$\begin{aligned} \frac{\partial C}{\partial t} = -\nabla \cdot \vec{J} = \nabla \Phi \cdot (C\nabla \mu(C) + \mu(C)\nabla C) \\ + \nabla D(C) \cdot \nabla C + D(C)\nabla^2 C. \end{aligned} \quad (9)$$

The standard approach is to enforce zero normal flux at the channel-wall boundaries, but since \vec{v} has a zero normal component at the walls, zero normal flux is equivalent to enforcing zero normal derivative in C . The concentration at the inlet was determined by the input, and the normal derivative of C was assumed 0 at the outlet.

Note that (9) is nonlinear with respect to marker concentration unless both electroosmotic mobility and diffusion coefficient are concentration-dependent.

A state-space system was generated from (9) by applying a second-order three-dimensional coordinate-mapped

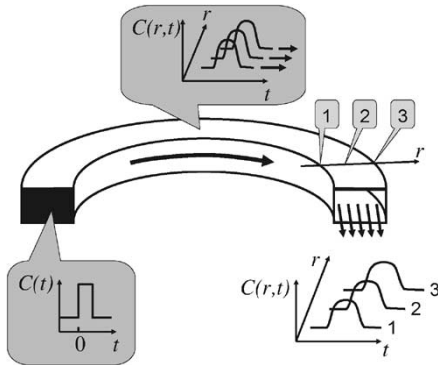


Fig. 3. Propagation of the square impulse of concentration of the marker. Due to the difference in lengths of the inner and outer arc, the marker reaches different points at the outlet with different delay.

finite-difference spatial discretization to (9) on the half-ring domain in Fig. 2. The states were chosen to be concentrations of the marker fluid at the spatial locations inside the channel. The concentration of the marker at the inlet of the channel is the input signal, and there are three output signals: the first being the average concentration at the outlet, and the second and third signals being the concentrations at the inner and outer radii of the outlet of the channel, respectively.

Fig. 3 illustrates the way an impulse of concentration at the inlet propagates through the channel: Diffusion spreads the pulse, and due to the curvature of the channel, the front of the impulse becomes tilted with respect to the channel's cross section. That is, the marker first reaches the points at the inner radius (point 1).

V. COMPUTATIONAL RESULTS

In this section, results are first presented for the linear microfluidic channel model, in order to emphasize the efficiency of the TBR linear reduction. Then, results are presented for the micromachined-pump model. The most challenging example was a nonlinear microfluidic channel.

A. Microchannel—Linear Model

First, in order to demonstrate the effectiveness of TBR linear reduction, consider applying a balanced-truncation algorithm to the linearized microchannel model. This corresponds to the problem of a very diluted solution of a marker in the carrier liquid (a widely used approximation in the literature). The values used for the electroosmotic mobility and diffusion coefficients are from [23]: $\mu = 2.8 \times 10^{-8} \text{ m}^2 \cdot \text{V}^{-1} \cdot \text{s}^{-1}$, $D = 5.5 \times 10^{-10} \text{ m}^2 \cdot \text{s}^{-1}$. Physical dimensions of the channel were chosen to be $r_1 = 500 \text{ } \mu\text{m}$, $w = 300 \text{ } \mu\text{m}$, and $d = 300 \text{ } \mu\text{m}$. Finite-difference discretization led to a linear time-invariant system (A, B, C) of order $N = 2842$ (49 discretization points by angle, 29 by radius, and 2 by height). Since algorithm 1 requires $O(n^3)$ computation, the discretized system was too costly to reduce using the original TBR algorithm. Instead, we used a fast-to-compute approximation to the TBR called modified approximate implicit subspace iteration with alternate directions (AISIAD) [26], [27].

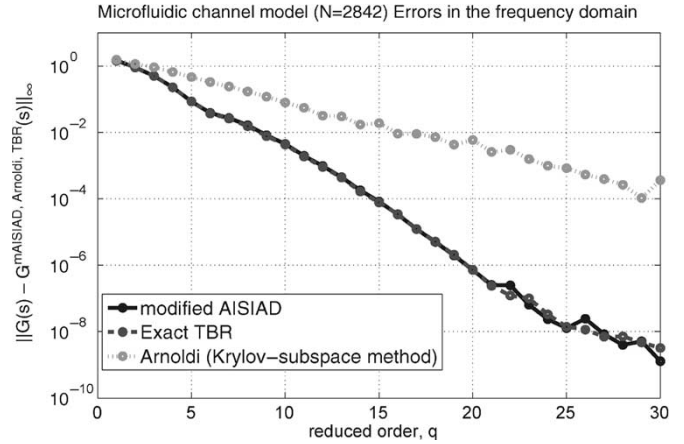


Fig. 4. H-infinity errors (maximal discrepancy over all frequencies between transfer functions of original and reduced models) for the Krylov, TBR, and modified AISIAD reduction algorithms.

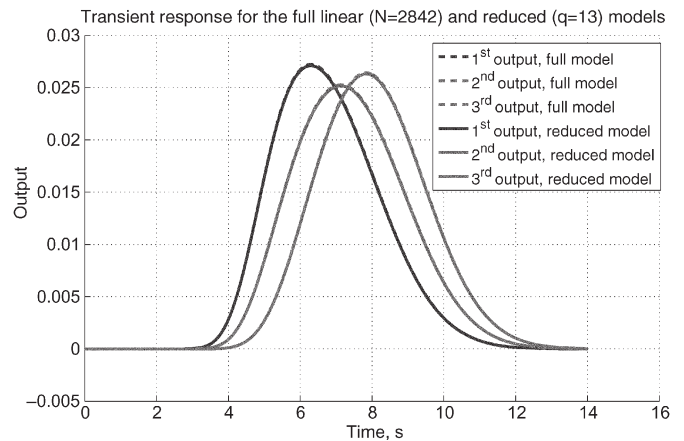


Fig. 5. Transient responses of the original linear (dashed lines) and reduced (solid lines) model (order $q = 13$). Input signal: unit pulse with duration 0.1 s. The maximum error between these transients is $\approx 1 \times 10^{-4}$, therefore, the difference is barely visible. The different outputs correspond to the different locations along the channel's outlet (from left to right: innermost point, middle point, outermost point).

As shown in Figs. 4 and 5, applying reduction to the spatial discretization of (9) demonstrates the excellent efficiency of the TBR reduction algorithm. The reduction error decreases exponentially with increasing reduced model order, both in frequency- and in time-domain measurements. For example, in the time-domain simulations, the maximum error in the unit step response for the reduced model of order $q = 20$ (over a 100-times reduction) was lower than 10^{-6} for all three output signals.

The modified AISIAD method was compared with Krylov-subspace-based reduction (Arnoldi method [14]) and the original TBR method in both time and frequency domains. As shown in Fig. 4, TBR and modified AISIAD are much more accurate than the Krylov method, and are nearly indistinguishable. However, the modified AISIAD model is much faster to compute.

To demonstrate the time-domain accuracy of the reduced model, we first redefined the outputs of the model as concentrations at the points 1–3 on Fig. 3, and then performed approximate TBR reduction using the modified AISIAD method.

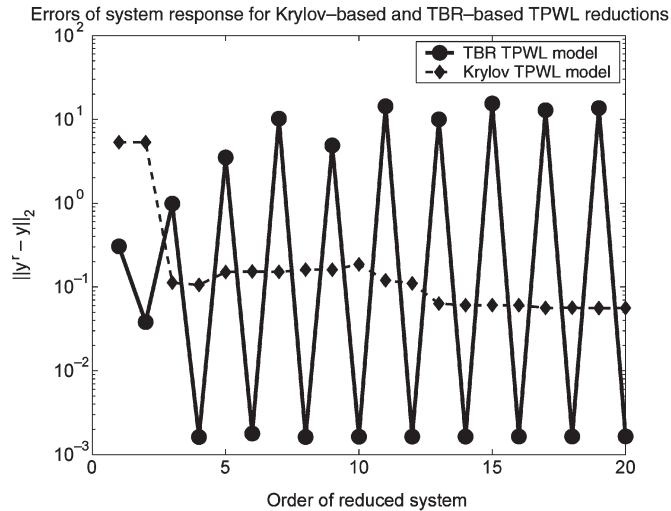


Fig. 6. Errors in output computed by TPWL models generated with different MOR procedures (micromachined-pump example); $N = 880$; 5.5-V step testing and training input voltage.

In Fig. 5, the output produced by a 0.1-s unit pulse is shown. The results for the 2842-state model and modified AISIAD reduced model of order 13 are compared in Fig. 5. One can clearly see that the reduced model nearly perfectly represents different delay values and the spread of the outputs.

B. Micromachined-Pump Example

The TBR TPWL MOR strategy was applied to generate macromodels for the micromachined-pump example described in Section IV. The reduced basis was generated using the linearized model of system (1) only at the initial state, and the initial state was included in the bases V and W .

Surprisingly, unlike in several nonlinear circuit examples [13], the output error did not decrease monotonically as the order q of the reduced system grew. Instead, macromodels with odd orders behaved very differently than macromodels with even orders. Models of even orders were substantially more accurate than models of the same order generated by Krylov reduction—cf., Fig. 6. However, if q was odd, inaccurate and unstable reduced-order models were obtained. This phenomenon is reflected in the error plot shown in Fig. 6. Fig. 7 illustrates that a fourth-order (even) reduced model accurately reproduces transient behavior.

This “even–odd” phenomenon was observed in [28] and explained in the very general sense in [29]. The main result in [29] is described in the following section. However, there is also an insightful but less general way of looking at this phenomenon.

The “even–odd” phenomenon can be viewed by examining eigenvalues of the reduced-order Jacobians from different linearization points. For the pump example, the initial nonlinear system is stable and Jacobians of f at all linearization points are also stable. Nevertheless, in this example, the generated reduced-order basis provides a truncated balancing transformation only for the linearized system from the initial state x_0 . Therefore, only the reduced Jacobian from x_0 is guaranteed to

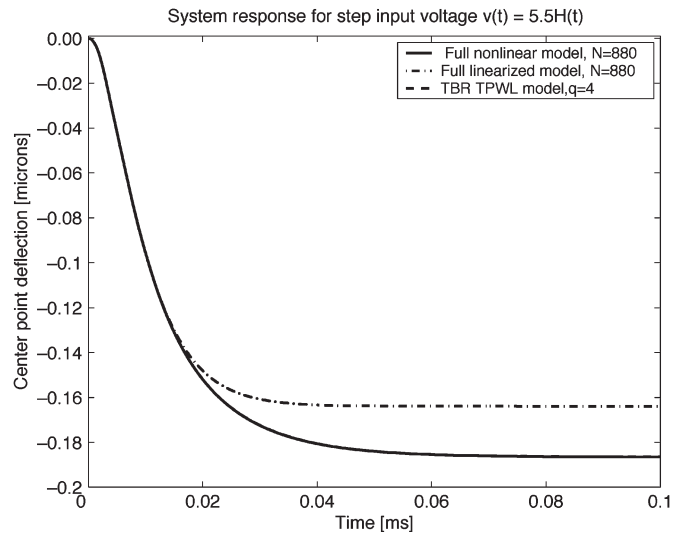


Fig. 7. Comparison of system response (micromachined-pump example) computed with both nonlinear and linear full-order models, and TBR TPWL reduced-order model (seven models of order $q = 4$); 5.5-V step testing and training input voltage. Note: solid and dashed lines almost fully overlap.

be stable. Other Jacobians, reduced with the same projection bases, may develop eigenvalues with positive real parts.

Fig. 8 shows spectra of the reduced-order Jacobians for models of order $q = 7$ and $q = 8$. One may note that, for $q = 8$, the spectra of the Jacobians from a few first linearization points are very similar. They also follow the same pattern: two of the eigenvalues are real, and the rest form complex-conjugate pairs. Increasing or decreasing the order of the model by 2 creates or eliminates a complex-conjugate pair of stable eigenvalues from the spectra of the Jacobians. If the order of the model is increased or decreased by 1 [cf., Fig. 8 (left)], the situation is very different. A complex-conjugate pair will be broken, and a real eigenvalue will form. At the first linearization point, this eigenvalue is a relatively small negative number. At the next linearization point, the corresponding eigenvalue shifts significantly to the right half plane to form an unstable mode of the system. An obvious workaround for this problem in the considered example is to generate models of even order. Nevertheless, a true solution to this problem would involve investigating how perturbations in the model affect the balanced reduction, and this is examined in Section VI.

C. Nonlinear Microfluidic Example

Consider introducing a mild nonlinearity into the mobility and diffusion coefficients in (9) as follows:

$$\mu(C) = (28 + C \cdot 5.6) \times 10^{-9} \text{ m}^2 \cdot \text{V}^{-1} \cdot \text{s}^{-1}$$

$$D(C) = (5.5 + C \cdot 1.1) \times 10^{-10} \text{ m}^2 \cdot \text{s}^{-1}.$$

Our experiments showed that even such a small nonlinearity creates a challenging problem for the TPWL algorithm. For this problem, the choice of training input significantly affects the set of the input signals for which the reduced model produces accurate outputs. For the case of a pulsed marker, this example has, in effect, a traveling-wave solution. Therefore, linearizing

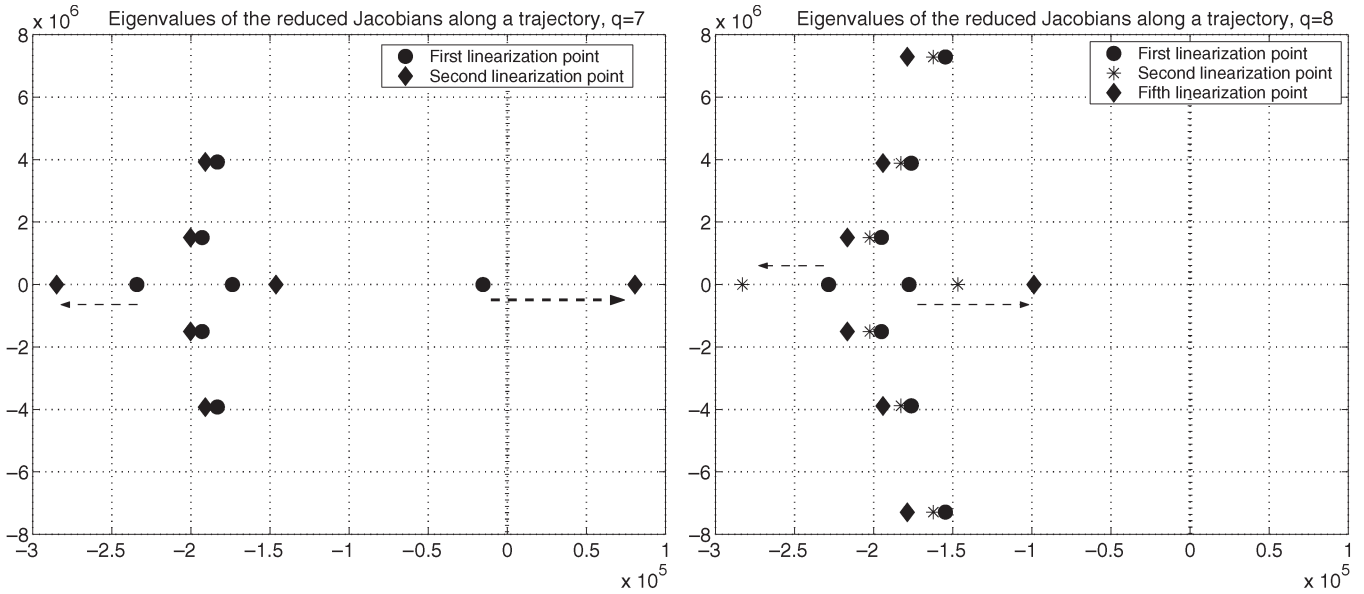


Fig. 8. Eigenvalues of the Jacobians from the first few linearization points (micromachined-pump example, Krylov-TBR TPWL reduction). Order of the reduced system $q = 7$ (left), $q = 8$ (right).

at different timepoints implies linearizing different spatially local regions of the device, and many linearizations will be needed to cover the entire device.

Our experiments showed that a workable choice of projection matrices V and W for this example is an aggregation of the TBR basis and some of the linearization states x_i . Therefore, the projection used was a mix between TBR and snapshots-based projection [30]. For example, the reduced model whose transient step response is presented in Fig. 9 was obtained using an aggregation of an order-15 TBR basis and 18 linearization states. The resulting system size was $q = 33$, and the number of linearization points was 23 (the initial model size was $N = 2842$). The linearization points were generated using the same step input for which the reduced simulation was performed. Although the results from the reduced model match when the input is the same as the training input, the errors become quite large if other inputs are used. For these nonlinear-wave-propagation problems, one needs to use a richer set of training inputs, which will result in a larger set of TPWL linearization points. In addition, instability in this simulation is still an issue, which makes the exact choice of projection basis an *ad hoc* procedure.

VI. PERTURBATION ANALYSIS OF TBR REDUCTION ALGORITHM

For the micropump example, the even-odd behavior of the model reduction can be analyzed using perturbation analysis. Assume the projection bases V and W are computed using TBR reduction from a single linearization point. The key issue is whether or not the TBR basis obtained at one linearization point is still suitable for reducing piecewise-linear models further along the trajectory. To understand this issue, consider two linearizations of the nonlinear system (1) (A_0, B, C) (initial) and (A, B, C) (perturbed). Suppose TBR reduction is performed

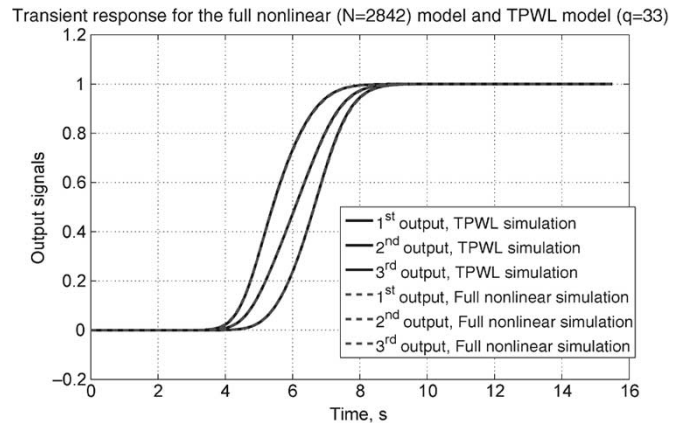


Fig. 9. Step response of reduced and initial microfluidic model. Solid lines: order-33 TPWL reduced model obtained by using step training input. Dashed lines: full nonlinear model, $N = 2842$. Note: solid and dashed lines almost fully overlap. The leftmost lines are the second input, which corresponds to the concentration closer to the center of the channel’s curvature. The middle lines correspond to the first output signal (average concentration at the outlet). The rightmost lines correspond to the concentration at the outlet’s points away from the center of curvature.

for both of these models, resulting in projection bases V, W and \tilde{V}, \tilde{W} , respectively. If these two bases are not significantly different, then perhaps V and W can be used to reduce the perturbed system, as is done for TPWL macromodels. This is true given some care, as will be made clear below.

A. Effect of Perturbation on Gramians

Consider the case for the controllability gramian P only, the results are valid for Q as well. Let $A = A_0 + \delta A$, $P = P_0 + \delta P$, where P_0 is an unperturbed gramian corresponding to unperturbed matrix A_0 , and δA is relatively small so that δP is also small.

Using the perturbed values of A and P in the Lyapunov equation and neglecting $\delta P \delta A$ yields

$$A_0 \delta P + \delta P A_0^T + (\delta A P_0 + P_0 (\delta A)^T) = 0. \quad (10)$$

Note that (10) is a Lyapunov equation with the same matrix A_0 as for the unperturbed system. This equation has a unique solution, assuming that the initial system is stable. The solution to (10) can be expressed using the following integral formula

$$\delta P = \int_0^{\infty} e^{A_0^T t} (\delta A P_0 + P_0 (\delta A)^T) e^{A_0 t} dt. \quad (11)$$

Assuming A is diagonalizable, δP can be bounded as

$$\|\delta P\| \leq 2(\text{cond}(T))^2 \|\delta A\| \|P_0\| \int_0^{\infty} e^{2\Re(\lambda_{\max}(A_0))t} dt \quad (12)$$

where T is the matrix that diagonalizes A .

Since A is stable, the integral in (12) exists and yields an upper bound on infinitesimal perturbations of the gramian:

$$\|\delta P\| \leq \frac{1}{|\Re(\lambda_{\max}(A_0))|} (\text{cond}(T))^2 \|P_0\| \|\delta A\|. \quad (13)$$

Equation (13) shows that the bound on the norm of δP increases as the maximal eigenvalue of A_0 approaches the imaginary axis. In addition, note that perturbations in A will result in small perturbations in the gramian P as long as the system remains “stable enough,” i.e., its eigenvalues are bounded away from the imaginary axis.

B. Effect of Perturbations on the Balancing Transformation

The balancing transformation in algorithm 1 can be viewed essentially as a symmetric eigenvalue problem [21]:

$$RPR^T = U \text{diag}(\Sigma^2) U^T \quad T = \Sigma^{-1} U^T R \quad (14)$$

where $R^T R = Q$ (R is a Cholesky factor of Q) and T is the coordinate transformation that diagonalizes both gramians. In algorithm 1, matrix W consists of the first q rows of T , and matrix V consists of the first q columns of T^{-1} .

Applying the same perturbation analysis to the Cholesky factors, it can be shown that the perturbations in the Cholesky factors due to the perturbations in the original gramian are also small, provided that the system remains “observable enough,” that is, the eigenvalues of Q are bounded away from 0. Therefore, we can state that the perturbation properties of the TBR algorithm are dictated by the symmetric eigenvalue problem $RPR^T = U \Sigma^2 U^T$.

The perturbation theory for the eigenvalue problem has been developed quite thoroughly [31], and one of the first observations is that small perturbations of a symmetric matrix can

lead to large changes in the eigenvectors if there are subsets of eigenvalues in the initial matrix that are very near to each other.

Below, we summarize a perturbation theory for a symmetric eigenvalue problem with a nondegenerate spectrum.

Consider a symmetric matrix $M = M_0 + \delta M$, where M_0 is the unperturbed matrix with known eigenvalues and eigenvectors, and no repeated eigenvalues. Eigenvectors of M can be represented as a linear combination of eigenvectors of M_0

$$x_k = \sum_{i=1}^N c_i^k x_i^0$$

where x_k is the k th eigenvector of the perturbed matrix M and x_i^0 is the i th eigenvector of the unperturbed matrix. Coefficients c_i^k show how the eigenvectors of matrix M_0 are intermixed due to the perturbation δM , as in

$$\begin{aligned} (M_0 + \delta M) \sum_{i=1}^N c_i^k x_i^0 &= \lambda_k \sum_{i=1}^N c_i^k x_i^0 \\ \Rightarrow \sum_{i=1}^N c_i^k \delta M_{ji} &= (\lambda_k - \lambda_j^0) c_j^k \end{aligned}$$

where λ_k and λ_k^0 are the k th eigenvalues of M and M_0 , respectively, and $\delta M_{ij} = (x_i^0)^T \delta M x_j^0$ is a matrix element of the perturbation in the basis of the unperturbed eigenvectors.

Now, assume small perturbations and represent $\lambda_k = \lambda_k^0 + \lambda_k^{(1)} + \lambda_k^{(2)} + \dots$ and $c_k^n = \delta_{kn} + c_k^{n(1)} + c_k^{n(2)} + \dots$, where each subsequent term represents smaller orders in magnitude. The first-order terms are

$$\lambda_k^{(1)} - \lambda_k^0 = \delta M_{jj} \quad (15)$$

and

$$c_k^n = \frac{\delta M_{kn}}{\lambda_n^0 - \lambda_k^0}, \quad k \neq n. \quad (16)$$

Equation (16) implies that the greater the separation between eigenmodes, the less they tend to intermix due to small perturbations. If a pair of modes have eigenvalues that are close, they change rapidly with perturbation. The following recipe for choosing an order of projection basis exploits this observation.

C. Recipe for Using TBR With TPWL

Pick a reduced order to ensure that the remaining Hankel singular values are small enough, and the last-kept and first-removed Hankel singular values are well separated.

The above recipe yields a revised TBR-based TPWL algorithm:

TBR-based TPWL with the linearization at the initial state

- 1) Perform the TBR linear reduction at the initial state x_0 . Add x_0 to the projection matrices V and W by using biorthonormalization.

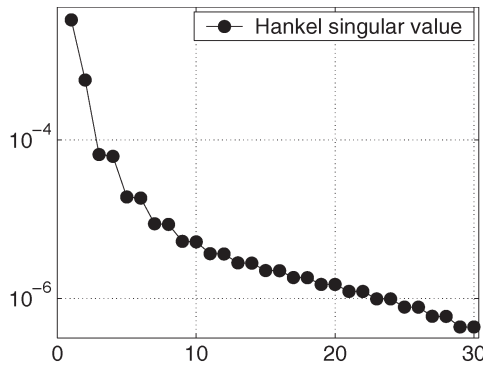


Fig. 10. Hankel singular values of the balancing transformation at the initial state, micromachined-pump example.

- 2) Choose the reduced-order q such that the truncated Hankel singular values are:
 - Small enough to provide sufficient accuracy;
 - Separated enough from the Hankel singular values that are kept.
- 3) Simulate the training trajectory and collect linearizations.
- 4) Reduce linearizations using the projection matrices obtained in step 1).

D. Even–Odd Behavior Explained

The perturbation analysis suggests that the sensitivity of TBR projection basis is strongly dependent on the separation of the corresponding Hankel singular values. The Hankel singular values for the linearization point of the micromachined-pump example are shown in Fig. 10.

As one can clearly see, the Hankel singular values for the micropump example are arranged in pairs of values, and evidently, even-order models violate the recipe for the choice of reduction basis.

VII. REMARK: SYSTEM-LEVEL AND INDIVIDUAL MODEL STABILITY

The above approach attempts to create individual reduced models of devices, which are stable. However, interconnections of stable models do not, in general, result in a stable system (even for linear systems). In order to ensure system-level stability, it may be necessary to enforce stronger criteria on the individual models. One possibility would be to ensure an individual system's dissipativity (also termed passivity). That is, the energy in the output signal should not exceed energy in the input signal. This is a much more challenging problem than the one considered in this paper. In addition, particular constraints associated with the system's dissipativity are problem dependent: They depend on the physical nature of the input and output signals. On the other hand, the proposed method is generic, that is, it does not assume any particular physical nature in the signals under consideration.

VIII. CONCLUSION

In this short paper, we demonstrated that replacing Krylov-subspace methods with TBR as the linear reduction method

in a TPWL algorithm dramatically improves reduced model accuracy for a given order, or substantially reduces the order needed for a given accuracy. In addition, we discovered, analyzed, and resolved an instability problem with the TPWL-TBR approach. In particular, we gave a perturbation analysis that showed that when TBR is used in combination with TPWL, one should not truncate at an order that splits nearly equal Hankel singular values. Finally, we also demonstrated that the TPWL-TBR approach has much more difficulty when applied to problems with nonlinear wave propagation.

REFERENCES

- [1] P. G. Glavina, D. J. Harrison, and A. Manz, "Towards miniaturized electrophoresis and chemical analysis systems on silicon: An alternative to chemical sensors," *Sens. Actuators B, Chem.*, vol. 10, no. 1, pp. 107–116, 1993.
- [2] C. H. Ahn, J. W. Choi, G. Beaucage, J. Nevin, J. B. Lee, A. Puntambekar, and J. Y. Lee, "Disposable smart lab on a chip for point-of-care clinical diagnostics," *Proc. IEEE, Special Issue on Biomedical Applications for MEMS and Microfluidics*, vol. 92, no. 1, pp. 154–173, Jan. 2004.
- [3] F. Pourahmadi, K. Lloyd, G. Kovacs, R. Chang, M. Taylor, S. Sakai, T. Schafer, W. McMillan, K. Petersen, and M. A. Northrup, "Versatile, adaptable and programmable microfluidic platforms for DNA diagnostics and drug discovery assays," in *Proc. Micro Total Analysis Systems (MicroTAS) Symp.*, Enschede, The Netherlands, 2000, pp. 243–284.
- [4] Y. Wang, Q. Lin, and T. Mukherjee, "Applications of behavioral modeling and simulation on a lab-on-a-chip: Micro-mixer and separation system," in *Proc. IEEE Int. Behavioral Modeling and Simulation Conf. (BMAS)*, San Jose, CA, 2004, pp. 8–13.
- [5] —, "System-oriented dispersion models of general-shaped electrophoresis microchannels," *Lab chip*, vol. 4, no. 5, pp. 453–463, Oct. 2004.
- [6] —, "System simulations of complex electrokinetic passive micromixers," in *Tech. Proc. NSTI Nanotechnology Conf.*, Boston, MA, 2005, pp. 579–582.
- [7] T. Korsmeyer, J. Zeng, and K. Greiner, "Design tools for biomems," in *Proc. 41st. Conf. Design Automation*, San Diego, CA, 2004, pp. 622–627.
- [8] P. Li and L. T. Pileggi, "Norm: Compact model order reduction of weakly nonlinear systems," in *Proc. 40th Conf. Design Automation*, Anaheim, CA, 2003, pp. 472–477.
- [9] N. Dong and J. Roychowdhury, "Piecewise polynomial nonlinear model reduction," in *Proc. 40th Conf. Design Automation*, Anaheim, CA, 2003, pp. 484–489.
- [10] M. Rewieński and J. White, "A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 22, no. 2, pp. 155–170, Feb. 2003.
- [11] A. Oliveira, J. R. Phillips, J. Afonso, and L. M. Silveira, "Analog macro-modeling using kernel methods," in *Proc. Int. Conf. Computer-Aided Design*, San Jose, CA, 2003, pp. 446–453.
- [12] S. K. Tiwary and R. A. Rutenbar, "Scalable trajectory methods for on-demand analog macromodel extraction," in *Proc. 42nd Annu. Conf. Design Automation*, San Diego, CA. New York: ACM Press, 2005, pp. 403–408.
- [13] D. Vasilyev, M. Rewieński, and J. White, "A TBR-based trajectory piecewise-linear algorithm for generating accurate low-order models for nonlinear analog circuits and MEMS," in *Proc. 40th Conf. Design Automation*, Anaheim, CA. New York: ACM Press, 2003, pp. 490–495.
- [14] E. J. Grimme, "Krylov projection methods for model reduction," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Illinois, Urbana Champaign, 1997.
- [15] P. Feldman and R. Freund, "Efficient linear circuit analysis by padé approximation via lanczos process," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 14, no. 5, pp. 639–649, May 1995.
- [16] L. Miguel Silveira, M. Kamon, I. Elfadel, and J. White, "A coordinate-transformed arnoldi algorithm for generating guaranteed stable reduced-order models of RLC circuits," in *ICCAD: Proc. IEEE/ACM Int. Conf. Computer-Aided Design*, San Jose, CA. Washington, DC: IEEE Computer Society, 1996, pp. 288–294.
- [17] A. Odabasioglu, M. Celik, and L. T. 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.

- [18] B. C. Moore, "Principal component analysis in linear systems: Controllability, observability, and model reduction," *IEEE Trans. Autom. Control*, vol. AC-26, no. 1, pp. 17–32, Feb. 1981.
- [19] K. Glover, P. J. Goddard, and Y. C. Chu, *Model Reduction for Classes of Uncertain, Multi-Dimensional, Parameter Varying and Non-Linear Systems*, vol. 240, Lecture Notes in Control and Information Sciences. New York: Springer-Verlag, 1998, pp. 269–280.
- [20] J.-R. Li, "Model reduction of large linear systems via low rank system gramians," Ph.D. dissertation, Dept. Math., Cambridge, MA, 2000.
- [21] K. Glover, "All optimal Hankel-norm approximations of linear multivariable systems and their l^∞ -error bounds," *Int. J. Control*, vol. 39, no. 6, pp. 1115–1193, 1984.
- [22] E. Hung, Y. Yang, and S. Senturia, "Low-order models for fast dynamical simulation of MEMS microstructures," in *Proc. IEEE Int. Conf. Solid State Sensors and Actuators*, Chicago, IL, 1997, pp. 1101–1104.
- [23] Z. Tang, S. Hong, D. Djukic, V. Modi, A. C. West, J. Yardley, and R. M. Osgood, "Electrokinetic flow control for composition modulation in a microchannel," *J. Micromech. Microeng.*, vol. 12, no. 6, pp. 870–877, Nov. 2002.
- [24] E. B. Cummings, S. K. Griffiths, and R. H. Nilson, "Irrotationality of uniform electroosmosis," in *SPIE Conf. Microfluidic Devices and Systems II*, Santa Clara, CA, 1999, vol. 3877, pp. 180–189.
- [25] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed., vol. 6. Oxford, U.K.: Butterworth-Heinemann, 1977.
- [26] Y. Zhou, "Numerical methods for large scale matrix equations with applications in LTI system model reduction," Ph.D. dissertation, Dept. Comput. Appl. Math., Rice Univ., Houston, TX, 2002.
- [27] D. Vasilyev and J. White, "A more reliable reduction algorithm for behavioral model extraction," in *Proc. IEEE/ACM Int. Conf. Computer-Aided Design (ICCAD)*, San Jose, CA, Nov. 2005, pp. 813–820.
- [28] M. I. Younis, E. M. Abdel-Rahman, and A. Nayfeh, "A reduced-order model for electrically actuated microbeam-based MEMS," *J. Microelectromech. Syst.*, vol. 12, no. 5, pp. 672–680, Oct. 2003.
- [29] D. Vasilyev, M. Rewieński, and J. White, "Perturbation analysis of TBR model reduction in application to trajectory-piecewise linear algorithm for MEMS structures," in *Proc. NSTI Nanotechnology Conf.*, Boston, MA, 2004, vol. 2, pp. 434–437.
- [30] K. Willcox and J. Peraire, "Balanced model reduction via the proper orthogonal decomposition," *AIAA J.*, vol. 40, no. 11, pp. 2323–2330, Nov. 2002.
- [31] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory*, 3rd ed, vol. 3. Oxford, U.K.: Butterworth-Heinemann, 1977, ch. 36, pp. 133–137.



Dmitry Vasilyev (S'03) received the M.S. degree in physics in 2000, from Saint-Petersburg Technical University, St. Petersburg, Russia, and is currently working towards the Ph.D. degree in electrical engineering and computer science at the Massachusetts Institute of Technology, Cambridge.

His research interests include problems of linear and nonlinear model order reduction (MOR).



Michał Rewieński (S'01–M'03) received the M.S. degree in computer science and the Ph.D. degree in microwave engineering from the Technical University of Gdansk, Gdansk, Poland, in 1998 and 2000, respectively and in the Ph.D. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 2003.

He is currently with Synopsys Inc., Mountain View, CA. His research interests include MOR algorithms, full-wave electromagnetic modeling, and simulation techniques applied to problems in elec-

tronic design automation.



Jacob White (A'88) received the B.S. degree in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, and the S.M. and Ph.D. degrees in electrical engineering and computer science from the University of California, Berkeley.

He worked at the IBM T. J. Watson Research Center from 1985 to 1987, and was an Analog Devices Career Development Assistant Professor at the Massachusetts Institute of Technology from 1987 to 1989. He is currently at the Massachusetts Institute

of Technology, where he is the C. H. Green Professor in Electrical Engineering and Computer Science and an Associate Director of the Research Laboratory of Electronics. His current research interests are in numerical algorithms for problems in circuits, interconnect, micromachining for biological applications, biomolecule design, and systems biology.

Dr. White was a 1988 Presidential Young Investigator, and was an Associate Editor of the *IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS* from 1992 to 1996. He was Chair of the International Conference on Computer-Aided Design in 1999.