

SPARE - a Scalable algorithm for passive, structure preserving, Parameter-Aware model order REduction

Jorge Fernández Villena
INESC ID / IST
TU Lisbon, Portugal
jorge.fernandez@inesc-id.pt

L. Miguel Silveira
INESC ID / IST / Cadence Research Labs
TU Lisbon, Portugal
lms@inesc-id.pt

Abstract

In this paper we describe a flexible and efficient new algorithm for model order reduction of parameterized systems. The method is based on the reformulation of the parametric system as a parallel interconnection of the nominal transfer function and the non-parametric transfer function sensitivities with respect to the parameter variations. Such a formulation reveals an explicit dependence on each parameter which is exploited by reducing each component system independently via a standard non-parametric structure preserving algorithm. Therefore, the resulting smaller size interconnected system retains the structure of the original with respect to parameter dependence. This allows for better accuracy control, enabling independent adaptive order determination with respect to each parameter and adding flexibility in simulation environments. It is shown that the method is efficiently scalable and preserves relevant system properties such as passivity. The new technique can handle fairly large parameter variations on systems whose outputs exhibit smooth dependence on the parameters. Several examples show that besides the added flexibility and control, when compared with competing algorithms, the proposed technique can, in some cases, produce smaller reduced models with potential accuracy gains.

1 Introduction

As we step toward the nano-scale and higher frequency eras, parameter variability can no longer be disregarded as it directly impacts system behavior and performance. In recent years, interconnect has become a dominant factor in such performance. Interconnect models are usually represented by large systems of frequency dependent linear equations. Accounting for the effects of manufacturing or operating variability, such as geometric parameters, temperature, etc., leads to parametric interconnect models whose complexity must be tackled both during the design and verification phases. For this purpose, Model Order Reduction (MOR) techniques that can handle parameterized descriptions are being considered as essential in the determination of correct system behavior. The systems generated after *Parametric MOR* (pMOR) must retain the ability to model the effects of both geometric and operating variability as well as variations caused by small random fluctuations, in order to accurately predict behavior and optimize designs.

Several pMOR techniques have been developed for modeling large-scale parameterized systems. The most common and effective ones appear to be extensions of the basic MOR algorithms [7, 9] to handle parameterized descriptions. An example of these are multiparameter moment-matching pMOR methods which match, via different approaches, generalized moments of the parametric transfer function, and build an overall projector. This can be accomplished by accounting for the functional dependence on both the frequency as well as the parameters and matching moments of the joint space [1, 5], or simply matching the moments of the individual parameter spaces [3]. However, the structure of such methods may present some computational problems, and the resulting system models usually suffer from oversize when the number of moments to match is high, either because high accuracy (order) is required or because the number of parameters is large. Sample-based techniques have been proposed in order to contain the large growth in model order for multiparameter, high accuracy systems [8, 11]. They rely on sampling of the multi-dimensional frequency and parameters space. This approach allows the inclusion of *a priori* knowledge of the parameter variation, and provides some error estimation. However, the issue of sample selection becomes particularly relevant, as the sampling must now be done in a potentially high-dimensional space.

A different approach based on a Taylor-series representation of the effect of parameters on the output of the system has been proposed in [6]. This approach is interesting as it directly captures the parametric dependence in an explicit sense, being able to tackle fairly large parameter variations in some scenarios. Unfortunately the technique does not guarantee desirable model characteristics, such as passivity, and the parametric dependence is lost in the resulting model. Therefore, the method is not efficiently scalable if higher order approximation on the parameters is required.

In this paper we present a new algorithm for model order reduction of parameterized systems. Similar to [6], the method is based on the reformulation of the parametric system revealing an explicit dependence on each parameter. Unlike [6], however, here this dependence is directly exploited by reducing each component system independently. Therefore, in the proposed approach, the resulting reduced model retains the structure of the original with respect to parameter dependence. This allows for better accuracy control, enabling independent adaptive order determination with respect to each parameter and adding flexibil-

ity in simulation environments. Furthermore, the procedure is shown to guarantee passivity and is efficiently scalable if higher accuracy, thus higher order is required.

The paper is structured as follows: in Section 2 an overview of pMOR and the existing techniques is presented, along with a discussion of their pros and cons. In Section 3 the new scheme will be introduced, starting with a description of the underlying representation and the methodology for reduction. We argue that the reduction effort and the *Reduced Order Model* (ROM) size are fairly independent from the variation range envisioned. We also show that the procedure preserves passivity of the model. In Section 4 several examples are shown that illustrate the efficiency of the proposed technique, and in Section 5 conclusions are drawn.

2 Background

2.1 Parametric System Formulation

Variability in actual fabrication of physical devices leads to a dependence of the extracted circuit elements on several parameters, of electrical or geometrical origin, that must be accounted for. This leads to parametric state-space system representations, which in descriptor form can be written as

$$\begin{aligned} C(\lambda)\dot{x}(t, \lambda) + G(\lambda)x(t, \lambda) &= Bu(t) \\ y(t, \lambda) &= Lx(t, \lambda) \end{aligned} \quad (1)$$

where $C, G \in \mathbb{R}^{n \times n}$ are respectively the dynamic and static matrix descriptors, $B \in \mathbb{R}^{n \times m}$ is the matrix that relates the input vector $u \in \mathbb{R}^m$ to the state $x \in \mathbb{R}^n$, and $L \in \mathbb{R}^{p \times n}$ is the matrix that links those inner states to the outputs $y \in \mathbb{R}^p$. The elements of C and G , as well as the states x , depend on a set of M parameters $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_M]$ which model the effects of the mentioned uncertainty. Usually the system is formulated so that the matrices related to the inputs and outputs (B and L) do not depend on the parameters. This time-domain descriptor yields a parametric dependent frequency response modeled via the transfer function

$$H(s, \lambda) = L(sC(\lambda) + G(\lambda))^{-1}B \quad (2)$$

for which we seek to generate a reduced order approximation, able to accurately capture the input-output behavior of the system for any point in the joint frequency-parameter space.

$$\hat{H}(s, \lambda) = \hat{L}(s\hat{C}(\lambda) + \hat{G}(\lambda))^{-1}\hat{B} \quad (3)$$

In general, one attempts to generate a ROM whose structure is as similar to the original as possible, i.e. exhibiting a similar parametric dependence. The reasoning behind this is that such models are better suited and allow more control within analysis and optimization frameworks. The most common procedure to achieve this goal is to use some form of projection scheme. Once a suitable subspace basis is computed, the system can be projected into that subspace, and a reduced model such as (3) can be obtained, that captures the behavior of the system under parameter variations.

2.2 Parametric Model Order Reduction

In the past few years several techniques have emerged in order to tackle this problem. Extensions to nominal moment-matching techniques (e.g. [7]), in order to

match generalized moments of the parametric transfer function (2), and build an overall projector, have been proposed. This can be accomplished by accounting for the functional dependence on both the frequency as well as the parameters and matching moments of the joint space [1, 5], or simply matching the moments of the individual parameter spaces [3]. These methods, which rely in local matching, usually suffer from oversize of the models when the number of moments to match is high, either because high accuracy is required or because the number of parameters is large, and their variation can lead to vastly different dynamics changes in different frequency ranges. Techniques that rely on sampling of the multi-dimensional frequency, parameter space, are potentially less sensitive to these issues [8]. They allow the inclusion of some knowledge of the parameter variation in order to guide the reduction. A technique that is concerned with the issue of sampling and bridges into the moment-based realm has been presented in [11]. However, when the range of parameter variations is large, the sampling effort can be considerable.

A different approach is explored in [6]. Here an explicit moment matching with respect to the parameters is first done, via Taylor-series expansion, followed by an implicit moment matching in frequency (via projection). The advantage of this technique is that the subspace basis is generated from the augmented system, so the frequency moments are captured for both the nominal and first (or higher) order sensitivities with respect to the parameters. It not only captures deviations in the output caused by the parameter, (to the prescribed order), but they are captured for any variation range, and for all frequency space. This approach is specially suited for systems where the output dependence on the parameters is smooth. However, in the approach of [6], the parametric dependence is lost after reduction, making model evaluation for different parameter values expensive and decreasing flexibility. Also, the method is not scalable, as the expansion order with respect to the parameters must be decided a-priori. Furthermore, accuracy considerations for the frequency moment matching based on a single expansion point may require high orders in large frequency ranges. This, however, can be improved with the usage of a sampling scheme as that of [11].

3 Scalable Parametric Aware Reduction

In this section we outline SPARE, a scalable and flexible new method for passive, structure preserving, parametric MOR based on a formulation similar to that of [6], which overcomes most of the difficulties of that method. The algorithm exhibits the following properties:

- **explicit structural dependence on the parameter set is maintained** which implies that re-evaluation of the model for different parameter values is very efficient (only the weighted sum of the parameter related transfer functions must be done, using existing values).
- **model evaluation is efficient** as the reduced model is described by a sparser block lower triangular form, that can be computed via recursive procedures.
- **model is accurate** for smooth output dependence on the parameters, as any range of variation is perfectly

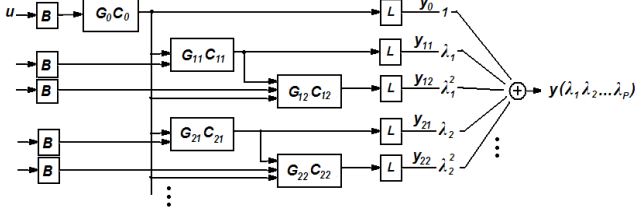


Figure 1. Parametric System representation as a cascade of non-parametric systems (Second order representation).

important implications. A graphical depiction of the above system representation is shown in Figure 1. The state-space descriptor matrices in (11) have a very characteristic structure that can be further exploited: the matrices can be stored block-wise, therefore, when the system is needed, it can be easily “assembled” for computation. Furthermore, if we only need to work with a subset of the parameter variations, it is not necessary to build the complete system, just the structure related to the varying parameters up to the required order. Additionally, the block lower triangular structure enables recursive procedures to be used for model evaluation of the individual responses, from which the output can be obtained through linear combination of terms. The effect of a change in the parameter values can be efficiently addressed by a weighted sum of such terms, sparing from extra evaluations. This adds flexibility, making it possible to, on the fly, generate a model as comprehensive as necessary, and is an interesting side benefit of the method. Preserving the structure is in itself a worthwhile endeavor, as previously discussed in [2, 10]. However it turns out that in this case it can lead to further advantages in terms of reduction and simulation.

3.2 Proposed Reduction

Let us now describe a projection-based MOR methodology for this formulation that maintains the explicit parameter dependence and avoids high computational costs. Such a methodology preserves the structure of the system, not only in terms of the structure of the underlying electrical circuit (thus in the sense described in [2]) but also in terms of the interconnection of systems prescribed in (6) and Figure 1. Due to the special structure of this representation, it is possible to apply a *Block Structure Preserving* (BSP) technique [2, 10] in order to maintain the block structure of the complete system. The expanded basis leads to larger reduced models, but this reduction has additional advantages, notably in terms of accuracy (see [10] for details).

In the present case, the input of every system depends on previous states, and the output is the contribution of all the partial outputs weighted by the parameter values (see Figure 1). From this viewpoint, the BSP projector can be built block-wise in a recursive manner. Each block must span a basis that, after projection, captures the most relevant behavior of each block of states x_{ij} in (6), so the complete reduced system will capture the most relevant behavior of the parametric system for any parameter variation. We point out that any technique that produces a suitable projector can be used at this point, attesting to the generality and flexibility of the algorithm. For instance, one can compute a basis

for a space that spans the moments of that particular subsystem. This can be done efficiently and in a numerically stable fashion. It is in fact akin to the computation of the multiparameter moments as defined in [11] but here performed in an individual basis, which adds flexibility in that only the desired moments need to be computed. An alternative is to use a sampling-based technique such as [9]. For illustration, this is the approach that we now outline.

The PMTBR algorithm is based on the Gramian estimation via a frequency sampling scheme. This technique has been shown to enable considerable reductions while providing a good heuristic error bound, useful to control the reduction size as a function of accuracy. To apply PMTBR to the multiple interconnected, individual systems can be very expensive due to the sample effort required. However, careful consideration of the structure of the system matrices and of the computation involved can lead to considerable savings. Consider for instance the transfer function

$$H_{11}(s) = -L(G_0 + sC_0)^{-1}(G_{11} + sC_{11})(G_0 + sC_0)^{-1}B.$$

In order to sample this function in the PMTBR scheme we need to obtain the sample vectors

$$z_{11k} = (G_0 + s_k C_0)^{-1}(G_{11} + s_k C_{11})(G_0 + s_k C_0)^{-1}B.$$

The term $(G_0 + s_k C_0)^{-1}$ is common to the nominal and all sensitivity transfer functions. Thus it only needs to be generated once for each sample, factored and then reused in ensuing computations. This computation requires care in order to ensure numerical stability. In fact a single LU factorization at each sample point allows us to obtain the sampled vectors of all the transfer functions in (7) with much less computational effort. Once the samples are obtained for each transfer function (i.e. for every x_{ij}), the corresponding projector V_{ii} can be obtained. When all such blocks are available, a BSP-type projector can be readily assembled [10], V_A , and applied to the augmented system (11). Under a good sampling scheme, a small number of vectors is enough for accurately characterizing the system (whereas more samples are required to cover a multidimensional space, as seen in [8]). This approach also allows us to control the frequency sampling range, focusing on the more interesting areas or the more affected by the variation. Once the projector V_A is obtained, a reduced system can be computed as

$$\hat{H}(s, \lambda) = \lambda_A \hat{L}_A (s \hat{C}_A + \hat{G}_A)^{-1} \hat{B}_A = \lambda_A \hat{H}_A(s) \quad (12)$$

where $\hat{L}_A = L_A V_A$, $\hat{C}_A = V_A^T C_A V_A$, $\hat{G}_A = V_A^T G_A V_A$, $\hat{B}_A = V_A^T B_A$, and λ_A was defined in (10). The following result shows that passivity is preserved during the reduction.

Theorem 3.1 *Considering the system in (9), where L_A, C_A, G_A, B_A are defined in (11), then the projected system $\hat{H}_A(s, \lambda)$ in (12) is passive if the original system is passive.*

Proof. In Section III-B of [7], it was shown that for a system formulated generically as that of Eqn. (9), where $L_A = B_A^T$, a sufficient (but not necessary) condition for passivity preservation in a congruence-type projection is that G_A and C_A both be non-negative definite (strictly speaking

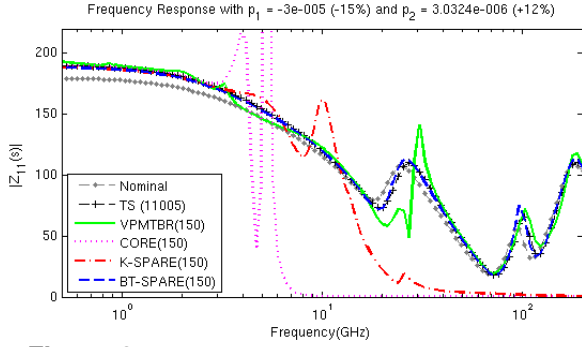


Figure 2. Spiral and CMIM example: $|Z_{11}|$ versus frequency for -15% variation in p_1 and $+12\%$ in p_2 .

that their symmetric parts, e.g. $G_A + G_A^T$ and $C_A + C_A^T$ be non-negative definite, but this has no relevance here). In our case, by construction, $L_A = B_A^T$, so it is sufficient to show that G_A and C_A are non-negative definite. Consider C_A as defined in (11). Clearly C_A is a block-lower triangular matrix with block diagonal C_0 . Therefore its eigenvalues are the same as those of C_0 (with multiplicity dependent on the number of parameters and the order of the expansion used). Since by construction C_0 resulting from an MNA formulation is non-negative definite, all its eigenvalues are non-negative. Therefore all eigenvalues of C_A are non-negative and this implies that C_A is non-negative definite. A similar reasoning applies to G_A , recalling that G_0 resulting from MNA is also non-negative definite. \square

4 Simulation Results

In this Section we benchmark the proposed algorithm against those presented in [8], denoted as VPMTBR, and [6], denoted as CORE. In general, we will denote our algorithm as SPARE and, when necessary, as K-SPARE or BT-SPARE to distinguish approximations where our projector was computed either by moment matching or the PMTBR approach. The non-reduced model response will be denoted as Nominal or Taylor Series (TS), depending on whether a variation of the parameters has been applied.

For our first example we use an EM model of a planar spiral inductor connected in series with a MIM (metal-insulator-metal) capacitor, including surrounding and substrate. The model, obtained by solving the full wave EM equations (via Finite Integration Technique, FIT [4]), has a size of 11005, and depends on two parameters. One is the length of the side of the square that forms the spiral, and the other is the insulator width. Sensitivity (1st order) information relative to these parameters is obtained for G and C . Table 1 shows the relevant characteristics of the models and algorithms used, namely the number of nonzero elements in the model (NNZ), the sparsity factor (ratio of NNZ), the cost of model generation (Gen. Effort) and evaluation (Speed-Up, a ratio of the time spent evaluating the reduced versus the original models). In the case of CORE, K-SPARE, and BT-SPARE, the ROMs are obtained for 1st order w.r.t. the parameters (linear approximation). The sizes of all the ROMs are set to 150, in order to compare them. The effort needed for generating such models varies: Krylov based techniques, such as CORE and

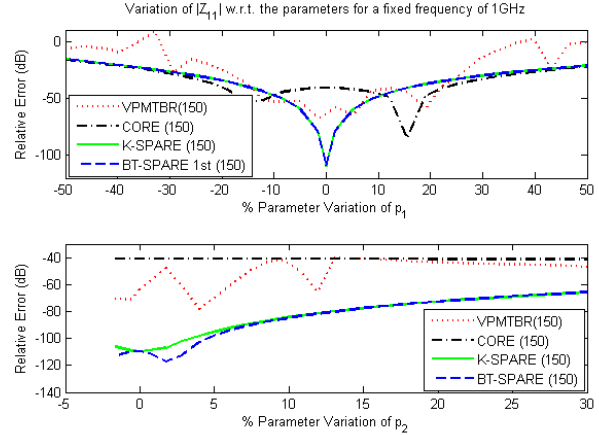


Figure 3. Spiral and CMIM example: Relative Error (dB) of the ROMs of Z_{11} versus the parameter variation at 1GHz.

Table 1. Characteristics of the pMOR methods applied for the $n = 11005$ -states LC example

| | NNZ (G C) & Sparsity | Gen.Effort | Speed-Up (1 eval) |
|----------------------------|--------------------------|---------------------------------------|----------------------|
| VPMTBR 150 | 22500 22500 1.00 1.00 | 60 Spl.(w+λ) SVD($n \times 150$) | 145× |
| CORE 150 | 22500 22500 1.00 1.00 | 75 BM QR($3n \times 150$) | 254× |
| K-SPARE 150 (50,50,50) | 12500 12500 0.55 0.55 | 25 BM 3xQR($n \times 50$) | 1021× |
| BT-SPARE 150 (50,50,50) | 12500 12500 0.55 0.55 | 20 Spl(w) 3xSVD($n \times 50$) | 1021× |

K-SPARE, are cheaper. However, the number of block moments (BM) computed for K-SPARE is one third of that computed in CORE (see Table 1), as the basis is expanded. The cost of BT-SPARE is lower than that of VPMTBR, as it only samples the frequency, and the SVD is applied on a much smaller number of vectors. The VPMTBR model is computed using the same frequency samples as BT-SPARE, plus parameter samples around those frequency points. The evaluation of the models is again much faster in the SPARE-based models, as their matrices are sparser and block lower triangular (see Table 1, where the ROM, and the blocks sizes in the SPARE case are shown below the method name) whereas CORE and VPMTBR yield full models. Furthermore, any change in the parameter values implies a complete reevaluation of the model for all techniques but SPARE. For SPARE this cost is negligible and, therefore, for N different parameter settings, the Speed-Up increases by a factor of N . With respect to the accuracy of the models, Figure 2 shows the frequency transfer function computed by all methods, for a given setting of the parameters. CORE and K-SPARE show good agreement for low frequencies, but with increasing error as frequency rises (more frequency moments would improve their accuracy). VPMTBR is not able to maintain the accuracy for parameter variations whereas BT-SPARE of 1st order exhibits the best accuracy for the whole frequency range and for a fairly large variation on the parameters. Figure 3, on the other hand, shows the transfer function variation with respect to parameters p_1 (Up) and p_2 (Down), for a fixed frequency point.

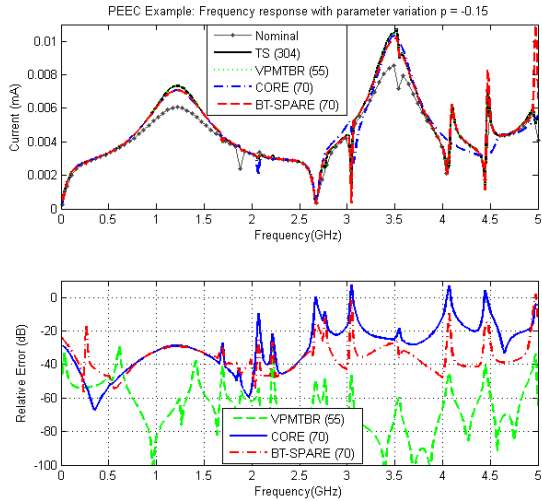


Figure 4. Frequency transfer function and relative error of the PEEC example with -0.15 parameter variation.

Table 2. PEEC example with $n = 304$: ROMs features

| Method | NNZ (G C) Sparsity | Gen. Effort | Speed-Up (1 eval) |
|---------------------|--------------------------|---|----------------------|
| VPMTBR 55 | 3025 3025 1.00 1.00 | 40 Spl.(w+ λ) SVD($nx55$) | 35.4 \times |
| CORE 70 | 4900 4900 1.0 1.0 | 70 BM QR($2nx70$) | 27.1 \times |
| SPARE 70 (32,38) | 3684 3684 0.751 0.751 | 30 Spl.(w) 2xSVD($nx38$) | 57.5 \times |

It is clear that the SPARE based models are able to capture the parametric behavior for fairly large variations. CORE exhibits a similar behavior, but loses accuracy as frequency rises. The accuracy of VPMTBR varies and deteriorates when the variations increase (for the same ROM size). p_2 has a less relevant effect, and all the algorithms are able to capture it accurately, although the SPARE based techniques display a *smoother* behavior.

For our next example, a PEEC model of size 304 will be used to test the methodology for small variation in highly non-linear environments. A synthetic single parameter is applied with different effects on the resistive, capacitive and inductive elements of the matrices. 1st order BT-SPARE will be benchmarked against CORE and VPMTBR. Figure 4 shows the frequency response for a chosen parameter value. SPARE of order 70 provides a better approximation than CORE, but in this situation is unable to match the accuracy of the VPMTBR algorithm. However, the generation cost and the speed-up in model evaluation is still better than for VPMTBR, even for a smaller model (see Table 2).

5 Conclusions

In this paper, a flexible and efficient parametric MOR algorithm was presented. The method is based on a reformulation of the original system as a parallel interconnection of the nominal transfer function and the non-parametric transfer function that describe the effect of the various parameters. This formulation reveals an explicit dependence on

the parameters which is exploited and preserved during reduction. This allows for better accuracy control, enabling independent adaptive order determination with respect to each parameter and adding flexibility in simulation environments. The new technique can handle fairly large parameter variations on systems whose outputs exhibit smooth dependence on the parameters and has been shown to preserve passivity. Examples show that besides the added flexibility and control, the proposed algorithm can produce smaller reduced models with potential accuracy gains in comparison with competing methods.

Acknowledgment

This work was supported by the EC IST/FP6 CHAMELEON-RF project. The authors gratefully acknowledge the support from the EC, and would like to thank the CHAMELEON RF consortium for many helpful discussions and for providing some of the examples.

References

- [1] L. Daniel, O. C. Siong, S. C. Low, K. H. Lee, and J. K. White. A multiparameter moment-matching model-reduction approach for generating geometrically parametrized interconnect performance models. *IEEE Trans. Computer-Aided Design*, 23:678–693, May 2004.
- [2] R. W. Freund. Sprim: Structure-preserving reduced-order interconnect macro-modeling. In *International Conference on Computer Aided-Design*, pages 80–87, San Jose, CA, U.S.A, November 2004.
- [3] P. Gunupudi, R. Khazaka, M. Nakhla, T. Smy, and D. Celo. Passive parameterized time-domain macromodels for high-speed transmission-line networks. *IEEE Trans. On Microwave Theory and Techniques*, 51(12):2347–2354, December 2003.
- [4] D. Ioan, G. Ciuprina, R. Radulescu, and E. Seebacher. Compact modeling and fast simulation of on-chip interconnect lines. *IEEE Transactions on Magnetics*, 42(4):547–550, April 2006.
- [5] P. Li, F. Liu, X. Li, L. Pileggi, and S. Nassif. Modeling interconnect variability using efficient parametric model order reduction. In *Proc. Design, Automation and Test in Europe Conference and Exhibition*, February 2005.
- [6] X. Li, P. Li, and L. Pileggi. Parameterized interconnect order reduction with Explicit-and-Implicit multi-Parameter moment matching for Inter/Intra-Die variations. In *International Conference on Computer Aided-Design*, pages 806–812, San Jose, CA, November 2005.
- [7] A. Odabasioglu, M. Celik, and L. T. Pileggi. PRIMA: passive reduced-order interconnect macromodeling algorithm. *IEEE Trans. Computer-Aided Design*, 17(8):645–654, August 1998.
- [8] J. Phillips. Variational interconnect analysis via PMTBR. In *International Conference on Computer Aided-Design*, pages 872–879, San Jose, CA, USA, November 2004.
- [9] J. R. Phillips and L. M. Silveira. Poor Man’s TBR: A simple model reduction scheme. *IEEE Trans. Computer-Aided Design*, 24(1):43–55, Jan, 2005.
- [10] H. Yu, L. He, and S. X. D. Tan. Block structure preserving model order reduction. In *BMAS - IEEE Behavioral Modeling and Simulation Workshop*, pages 1–6, September 2005.
- [11] Z. Zhu and J. Phillips. Random sampling of moment graph: a stochastic krylov-reduction algorithm. In *Proc. Design, Automation and Test in Europe Conference and Exhibition*, pages 1502–1507, Nice, France, April 2007.