Block oriented model order reduction of interconnected systems

INTERNATIONAL JOURNAL OF NUMERICAL MODELLING: ELECTRONIC NETWORKS, DEVICES AND FIELDS
Int. J. Numer. Model. 2009; **submission**:1–19 *Prepared using* jnmauth.cls *[Version: 2002/09/18 v1.01]*

Block oriented model order reduction of interconnected systems

Jorge Fernández Villena[1,*], Wil H. A. Schilders[2,3], and L. Miguel Silveira[1,4]

[1]*INESC ID / Instituto Superior Técnico, TU Lisbon, Rua Alves Redol 9, 1000-029 Lisbon,Portugal*
[2]*CASA - Dept. of Mathematics and Computer Science, TU Eindhoven, 5600 MB Eindhoven, The Netherlands*
[3]*NXP Semiconductors Research, High Tech Campus 37, 56656 AE Eindhoven, The Netherlands*
[4]*Cadence Research Labs, Rua Alves Redol 9, 1000-029 Lisbon, Portugal*

SUMMARY

Unintended and parasitic coupling effects are becoming more relevant in currently designed, small-scale/high-frequency RFICs. Electromagnetic (EM) based procedures must be used to generate accurate models for proper verification of system behaviour. But these EM methodologies may take advantage of structural sub-system organization as well as information inherent to the IC physical layout, to improve their efficiency. Model order reduction techniques, required for fast and accurate evaluation and simulation of such models, must address and may benefit from the provided hierarchical information. System-based interconnection techniques can handle some of these situations, but suffer from some drawbacks when applied to complete EM models. We will present an alternative methodology, based on similar principles, that overcomes the limitations of such approaches. The procedure, based on structure-preserving model order reduction techniques, is proved to be a generalization of the interconnected system based framework. Further improvements that allow a trade off between global error and block size, and thus allow a better control on the reduction, will be also presented.

KEY WORDS: Model Order Reduction; Structure Preserving; EM Simulation; Interconnected Systems

1. INTRODUCTION

Model Order Reduction (MOR) techniques have achieved a considerable level of maturity in the last few years, and their application is widespread and well established, from the automotive and aerospace industries to the electronic realm [1, 2, 3]. They are the basic tools for avoiding detailed simulations, far too costly in time and computational resources, providing smaller yet accurate models able to be processed in the simulation flow. However, these method must not only provide reliable models in the input-output behavioral sense, but they must also maintain some of the physical inherent characteristics

*Correspondence to: Jorge Fernández Villena, Room 133, INESC ID, Rua Alves Redol 9, 1000-029 Lisbon, Portugal
E-mail: jorge.fernandez@inesc-id.pt

Contract/grant sponsor: EU/IST/6FP STRP CHAMELEON RF (http://www.chameleon-rf.org); contract/grant number: 027378

Contract/grant sponsor: EC/Marie Curie Actions SERA COMSON (http://www.comson.org); contract/grant number: 019417

Received SOMEWHEN in 2009
Revised SOMEWHEN in 2009

of the original systems, such as the stability and passivity, with obvious implications to the simulation itself.

Within the electronics industry, the most popular methods are either the Moment Matching based methods [4, 5, 6, 7] or the Balanced Truncation framework [8, 10, 9]. Recently, a new approach that links the rational Krylov procedures to the *Truncated Balanced Realization* (TBR), named *Poor's Man* TBR (PMTBR), has been presented [11].

However, the ever increasing range of frequencies, together with the shrinking feature size, led to the necessity of taking into account formerly negligible Electromagnetic (EM) interactions between the multiple sub-systems that compose a full system. New EM modeling methodologies [12, 13] take advantage of such hierarchical information, and MOR stages should be able to include and maintain such a valuable information.

The present document aims at providing some basis for efficient techniques able to address the above mentioned issue, and it is structured as follows: in Section 2 a quick review of the main MOR techniques, including their advantages and drawbacks, will be presented. In Section 3 a first system-based approach for handling and reducing interconnected systems will be revisited, giving the basis for a hierarchical representation. In Section 4 the structure-preserving model reduction approaches will be introduced, to further link them to the interconnected-system based, and present some improvements. Next, some simulation results will be presented in Section 5. The document will be closed with some conclusions in Section 6.

2. BACKGROUND IN MODEL ORDER REDUCTION

The main techniques for MOR are geared towards the reduction of a state space linear time-invariant system, obtained by some modeling methodology, and representing a physical system. In such state space representation, the output is related to the input via some inner states satisfying,

$$\begin{aligned} C\dot{x} + Gx &= Bu \\ y &= Lx + Du. \end{aligned} \tag{1}$$

where $C, G \in \mathbb{R}^{n\times n}$ are respectively the dynamic and static matrices, $B \in \mathbb{R}^{n\times m}$ is the matrix that relates the input vector $u \in \mathbb{R}^m$ to the inner states $x \in \mathbb{R}^n$ 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 matrix $D \in \mathbb{R}^{p\times m}$ represents possible direct algebraic relation between inputs and outputs.

This time-domain description yields a frequency response modeled via the transfer function

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

for which we seek to generate a reduced order approximation, able to accurately capture the input-output behavior.

$$\widehat{H}(s) = \widehat{L}(s\widehat{C} + \widehat{G})^{-1}\widehat{B} + D, \tag{3}$$

in which $\widehat{C}, \widehat{G} \in \mathbb{R}^{q\times q}$, $\widehat{B} \in \mathbb{R}^{q\times m}$, and $\widehat{L} \in \mathbb{R}^{p\times q}$ are the reduced set of matrices, with $q \ll n$.

In general, one attempts to generate a *Reduced Order Model* (ROM) whose structure is as similar to the original as possible, in order to facilitate further simulations.

2.1. Moment Matching

Moment matching techniques have gained a well deserved fame due to their simplicity and efficiency. Proof of this is the large number of variants existing (some examples are [4, 5, 6, 7]). They rely

on the computation of the so called moments, i.e. the coefficients of the Taylor Series expansion of the transfer function in (2) around some frequency point. Perhaps, currently the most well known of these techniques is PRIMA [4], where a Krylov subspace is computed in an iterative fashion, via an iterative Block Arnoldi procedure, to latter project the original system matrices in (1), via a congruence transformation, into that subspace.

$$colsp\{V\} \equiv Kr\{A, R, q\}, \quad with \quad A = -G^{-1}C, \quad R = G^{-1}B, \tag{4}$$

$$\widehat{G} = V^T G V \quad \widehat{C} = V^T C V \quad \widehat{B} = V^T B \quad \widehat{L} = LV \tag{5}$$

where $Kr\{A, R, q\}$ is the Krylov subspace of q block moments $(R, AR, \ldots A^{q-1}R)$, and V is the projector applied in the congruence transformation (5) for the reduction.

This projection avoids numerical problems, and provides further benefits, such as the preservation of passivity under certain conditions (usually fulfilled in the case of electric based models). The iterative nature of the method and easy implementation turn these procedures into an efficient framework when applied to very large systems. On the other hand, the models obtained are sometimes larger than the necessary, in particular for large frequency ranges, and the reduced matrices are full, losing all initial sparsity. Furthermore, the only known techniques for a priori order selection are expensive and cumbersome to implement.

2.2. Balanced Truncation

Another wide spread approach is the Truncated Balanced Realization (TBR) [8, 9], which relies on the analysis of the Controllability and Observability of the inner states, via the computation of the Gramians. The system is balanced under a similarity transformation (which makes both Gramians equal and diagonal). That balancing maintains invariant the input-output properties of the state-space model, such as the transfer function and the eigenvalue space of the product of both the Controllability and Observability Gramians.

These eigenvalues, also known as Hankel singular values, can be associated to an energetic interpretation of the system, and more precisely, of the states. By truncating the "weak" states associated to the low value eigenvalues, the remaining "strong" states, associated with highly energetic eigenvalues, lead to a good reduced approximation of the original system. Furthermore, this framework is purported to give quasi-optimal reduced models, and the error can be controlled via an *a posteriori* theoretical error bound on the frequency domain, given by the sum of the truncated Hankel singular values [10]. On the other hand, the computation of the Gramians requires solving a dual pair of Lyapunov equations, which can be hard to implement and demands high computational effort. These facts limit the applicability of the TBR frameworks to small to medium size models.

2.3. PMTBR

Poor's Man TBR (PMTBR) [11] is a projection MOR technique that exploits the direct relation between the multipoint rational projection framework [7] and the truncated balanced realization, via a statistical interpretation of the system Gramians. This new approach can take advantage of some *a priori* knowledge of the system properties, and is less expensive in terms of computation, but tends to TBR when the order of the approximation increases.

The system Gramians can be expressed in the frequency domain (after applying *Parseval's Theorem*). In the case of the controllability Gramian

$$X = \int_{-\infty}^{\infty} (j\omega C + G)^{-1} B B^T (j\omega C + G)^{-H} dw \tag{6}$$

where ω is the frequency. A quadrature rule can be applied in the frequency space to approximate the Gramian via numerical computation

$$\widetilde{X} = \sum_{i}^{P} w_i z_i z_i^H \tag{7}$$

The samples are given by $z_i = (G + s_i C)^{-1} B$, where $s_i = j\omega_i$ (with $i = 1, 2, \ldots, P$) are P frequency sample points. The s_i and w_i can be interpreted as nodes and weights of the quadrature scheme. The actual mechanics of the algorithm are akin to multi-point projection. Let Z be a matrix whose columns are the z_i, and W the diagonal matrix of the square root of the weights. Eqn. (7) can be written more compactly as:

$$\widetilde{X} = ZW^2Z^H \tag{8}$$

If the quadrature rule applied is accurate, $\widetilde{X}$ will converge to X, which implies the dominant eigenspace of $\widetilde{X}$ converges to the dominant eigenspace of X. Computing the *Singular Value Decomposition* (SVD) of ZW, $ZW = VSU$ (with S real diagonal, and V, U unitary matrices), it is easy to see that V converges to the eigenspace of X, and the Hankel singular values are obtained directly from the entries of S. V can then be used as the projection matrix in a model order reduction scheme. The congruence projection scheme provides this method with the same advantages that in the case of PRIMA. Furthermore, the connection with the TBR framework and the fact that the singular values converge to the Hankel singular values, makes it possible to use them in a direct *a posteriori* error bound in the same way the Hankel singular values were used.

3. INTERCONNECTED SYSTEMS

An Interconnected System (IS) is a global model composed of several sub-systems. Inside the global IS, the inner sub-systems interact by some relations between their *local* inputs and outputs. The global IS can be represented in terms of the *global* input-output behavior, but the hierarchy of the interconnection makes the inner models still recognizable. The schematic of an example is shown in Figure 1.

From this starting point, the reduction of the complete system may follow different paths. One is to reduce the complete system by focusing on the global input-output response. This approach leads to a more compressed result, as it just focuses on the global behavior. On the other hand it may destroy the inner structure, making the sub-systems no longer distinguishable. Another possibility is to reduce every inner model individually, taking into account its local behavior. If each model response is accurately captured, the same relations between local inputs and outputs will lead to a reduced IS with similar characteristics. Despite maintaining the inner structure of the interconnection, some complex sub-systems may require large order models to accurately capture some individual behavior with no effect on the global response. This means that, from a global point of view, the reduction is not optimal.

Yet another possibility was already pursued in [14], which also gives a good system-viewpoint formulation for IS. This approach consists in reducing each individual model but by taking into account its effect in the global input-output response. Therefore, it is able to maintain the inner structure while a better global compression is achieved. In the next subsections the basis and reduction methods for this last approach will be briefly presented.

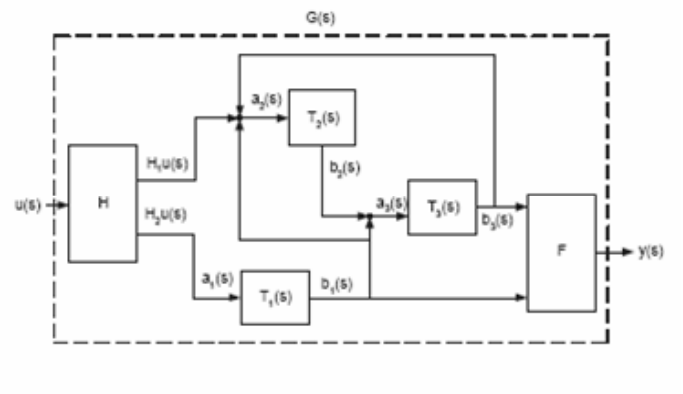

Figure 1. IS Global System composed from a set of interconnected sub-systems.

3.1. Basis

Under a system viewpoint, the interconnections are treated as signals which flow between the inner models, defined as sub-systems, with their local inputs and outputs. A new set of matrices is presented, defined as interconnection matrices, which relates the local inputs and outputs between them, and to the global inputs and outputs:

$$\begin{aligned} a_i &= u_i + \textstyle\sum_{i=1}^{N_b} K_{ij} b_j \\ u_i &= H_i u \\ y &= \textstyle\sum_{i=1}^{N_b} F_i b_i \end{aligned} \tag{9}$$

where N_b is the number of sub-systems, $u \in \mathbb{R}^m$ and $y \in \mathbb{R}^p$ are the global inputs and outputs respectively, $a_i \in \mathbb{R}^{m_i}$ and $b_i \in \mathbb{R}^{p_i}$ are the local inputs and outputs of the i-th sub-system, $H_i \in \mathbb{R}^{m_i \times m}$ is the matrix that relates the global to the local inputs of the i-th sub-system , $F_i \in \mathbb{R}^{p \times p_i}$ is the matrix that relates the local outputs of the i-th sub-system to the global outputs, and $K_{ij} \in \mathbb{R}^{p_i \times m_j}$ is the matrix that relates the local outputs of the j-th sub-system to the local inputs of the i-th sub-system. These relations, in combination with the state space matrices of each sub-system, yield in a global transfer function between the inputs u and outputs y, which can be reformulated in a global state-space system (see Figure 1 for a graphical depiction)

$$\begin{aligned} C_G \dot{x} + G_G x &= B_G u \\ y &= L_G x + D_G u \end{aligned} \tag{10}$$

where (for further details see [14])

$$\begin{aligned} C_G &= C_D & C_D &= diag\{C_1, C_2, \dots C_{N_b}\} \\ C_G &= G_D - BK(I - DK)^{-1}L & G_D &= diag\{G_1, G_2, \dots G_{N_b}\} \\ B_G &= B_D(I - DK)^{-1}H & B_D &= diag\{B_1, B_2, \dots B_{N_b}\} \\ L_G &= F(I - KD)^{-1}L_D & L_D &= diag\{L_1, L_2, \dots L_{N_b}\} \\ D_G &= FD_D(I - DK)^{-1}H & D_D &= diag\{D_1, D_2, \dots D_{N_b}\} \end{aligned} \tag{11}$$

$$H = \left[{H_1}^T {H_2}^T \dots {H_{N_b}}^T\right]^T \qquad K = \begin{bmatrix} K_{11} & \dots & K_{1N_b} \\ & \ddots & \\ K_{N_b 1} & \dots & K_{N_b N_b} \end{bmatrix}$$

$$F = [L_1, L_2, \dots L_{N_b}] \tag{12}$$

Although providing a nice and clear systemic formulation for the treatment of interconnected systems, this formulation retains some drawbacks that make this approach hard to model EM effects. The signal based interconnection only allows input-output relations, and, as seen in (11), has effect only in the algebraic part of the system (i.e. G_G, B_G, L_G and D_G). This means that the modeled interactions accounted for in this formulation are restricted to conductive connections, with an associated current flowing through the contact. However, some relevant EM effects occur without physical contact, via frequency dependent inductive and capacitive couplings (which may be desired or unintended). As an example let us take two close integrated spiral inductors. This situation generates a considerable amount of electromagnetic interaction, modeled as capacitive and inductive couplings. This case can not be properly modeled following the formulation presented. Nevertheless, this system based formulation can be very useful in some cases. The work in [14] presents two procedures to address the reduction of the system in (10), one based in Krylov projection methods and other in the TBR framework, which will be briefly summarized in the next two sub-sections (for further details see the referred literature).

3.2. Krylov Approaches

Under Krylov based MOR, some theorems are presented and proved in [14], which present some interesting moment matching properties inside the IS framework. First, if the i-th sub-system is reduced with a projector that spans the Krylov subspace of q moments obtained with its local state-space matrices (i.e. $colsp\{V_i\} \equiv Kr\{A_i, R_i, q\}$, with $A_i = -G_i^{-1}C_i$ and $R_i = G_i^{-1}B_i$), then the transfer function that results from the interconnection of the all the non reduced systems and the i-th (reduced) system matches q moments of the original one. This can be useful in order to reduce some sub-systems in *stand alone mode*, but, as it has been already explained at the beginning of Section 3, may not be the best reduction methodology to achieve a good global compression.

Second, if a projector whose columns spans the Krylov subspace of the global system is computed (i.e. $colsp\{V_G\} \equiv Kr\{A_G, R_G, q\}$, with $A_G = -G_G^{-1}C_G$ and $R_G = G_G^{-1}B_G$), then every single sub-system can be projected block wise (i.e. $V_G = \left[V_1^T \dots V_{N_b}^T\right]^T$, and V_i applied in a congruence transformation over the i-th sub-system). The transfer function generated from the interconnection of the resulting reduced sub-systems matches at least q moments of the original non reduced transfer function. This theorem opens some possibilities more in accordance with the proposed scheme, as it provides an independent sub-system reduction but focusing on capturing the global input-output behavior.

3.3. Balanced Truncation Approaches

The work in [14] is centered in a TBR-based scheme for MOR of the IS. The Interconnected System Balance Truncation (ISBT) presented there pursues exactly the reduction of the individual sub-systems in terms of the global I/O response.

This objective is achieved by computing the controllability and observability Gramians of the global IS system. Once these global Gramians are obtained, they are split into block Gramians (by following the system hierarchy), and the balancing is performed block wise, over the states related to each sub-system in an independent fashion. Therefore, the diagonal blocks of the controllability and observability Gramians become equal and diagonal. The truncation of the "weak" states can be also done block wise, but the information used for that truncation (i.e. the Hankel singular values) was obtained from the global system, and thus those truncated states are the ones that have a "weak" effect on the global response (see [14] for details and a proof).

To summarize, every sub-system is truncated independently but using the controllability and observability information of the global input-output response. The methodology still maintains the advantages of the TBR procedures, i.e. it provides quasi-optimal models and allows some degree of error control via the Hankel singular values. Furthermore, the truncation is applied locally (at each sub-system), but with respect to the global behavior. On the other hand, the drawbacks of the TBR framework are increased here, as the scheme needs to compute the global Gramians, and as a consequence, to solve a pair of dual Lyapunov equations with the global matrices. It should be noticed that the size of the global system is the sum of the individual sizes of the multiple sub-systems. Therefore this step becomes extremely computationally expensive, implying that the reduction of large size IS becomes very inefficient.

4. BLOCK STRUCTURE PRESERVING

A block structured system is one that has a well defined structure inside the state-space descriptor, i.e. the states can be split into several sets depending on their nature (e.g. if the states models voltage or currents in a simple MNA formulation) or other characteristics. Therefore, the matrices of the state-space descriptor in (1) have a relevant block hierarchy

$$G=\begin{bmatrix} G_{11} & \dots & G_{1N_b} \\ \vdots & \ddots & \vdots \\ G_{N_b1} & \dots & G_{N_bN_b} \end{bmatrix} \quad C=\begin{bmatrix} C_{11} & \dots & C_{1N_b} \\ \vdots & \ddots & \vdots \\ C_{N_b1} & \dots & C_{N_bN_b} \end{bmatrix}$$
$$B = \begin{bmatrix} {B_1}^T & \dots & {B_{N_b}}^T \end{bmatrix}^T \qquad L = [L_1 \ \dots \ L_{N_b}] \tag{13}$$

In certain situations it may be useful, or even necessary, to maintain the inner structure of this system after reduction. It is well known that projection techniques lead to full reduced matrices, in which the inner hierarchy and sparsity pattern of the original matrices are lost.

Some techniques for maintaining the block structure of a system were already pursued. The first work to address the problem was [15], where a two-block structure was kept in order to separate voltage from current states in a MNA formulation. Later, in [16], the same idea was extended to an arbitrary number of blocks in order to model several sub-circuits, and in [17] the same approach was used in the reduction of second order systems. In the following a brief review of the basis of the method will be done.

4.1. Basis

The main idea for this procedure is to retain the system block structure, i.e. the multi-level hierarchy, after reduction via projection, allowing for a more efficient reduction and the maintenance of certain system properties, such as the sparsity block pattern, and the block hierarchical structure. *Block Structure Preserving* (BSP) relies on expanding the projector of the global system (obtained via any classical MOR projection technique) into a block diagonal matrix, with block sizes equal to the sizes of its N_b individual component blocks (13).

A basis that spans a suitable subspace for reduction via projection is then computed (for example a Krylov subspace). The projector built from that basis can be split and restructured into a block diagonal

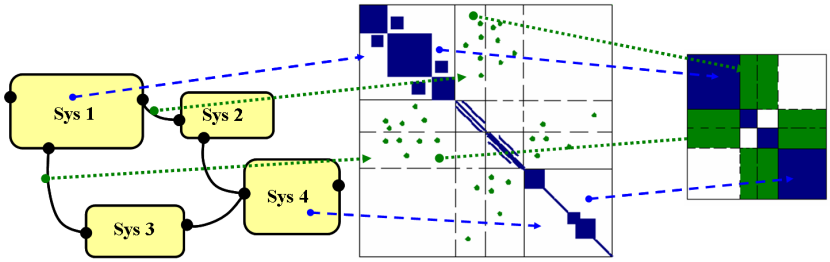

Figure 2. Interconnected system represented as a Block Structured system, and the effect of the subsequent Block Structure Preserving reduction.

one so that the structure is preserved under congruence transformation.

$$\begin{bmatrix} V_1 \\ \vdots \\ V_{N_b} \end{bmatrix} \equiv colsp\left[Kr\left\{A,R,q\right\}\right] \rightarrow \begin{bmatrix} V_1 & & \\ & \ddots & \\ & & V_{N_b} \end{bmatrix} = \breve{V} \tag{14}$$

with $A=-G^{-1}C$, $R=G^{-1}B$, and q the reduced order.

The block-wise congruence transformation leads to

$$\begin{array}{ll} \widehat{G}_{ij}=V_i^T G_{ij} V_j & \widehat{B}_i=V_i^T B_i \\ \widehat{C}_{ij}=V_i^T C_{ij} V_j & \widehat{L}_j=L_j V_j \end{array} \tag{15}$$

It is clear that inside the blocks the structure is lost, and up to the authors' knowledge, there are no known techniques to avoid this effect. Furthermore, the size of the reduced system is increased, as the number of columns of the projector after the expansion is $N_b \times q$.

On the other hand, the block structure is retained, the level of sparsity is increased (any empty block remains a empty reduced block), and it can be shown that under certain circumstances the number of moments matched can be increased (up to $N_b \times q$). In the worst case, the number of moments matched is q, i.e. the same as in the non-structure-preserving reduction.

4.2. Interconnected Systems Viewpoint

Let us consider now the case of EM modeling, where the linear components considered include designed-in passives, interconnect, etc. The system description has an interesting structure (see Figure 2), where the diagonal blocks correspond to the individual block matrices, whereas the off-diagonal blocks correspond to the conductive interconnections (in the G matrix) or capacitive and inductive couplings (C matrix). Standard model order reduction techniques can be applied to this global system and, while the resulting reduced model will usually be able to accurately capture the input-output behavior of the complete set of blocks, the generated reduced matrices are full and, furthermore, the original block hierarchy can no longer be recovered. In this context, the BSP technique presented in the previous section can be very helpful, as it can allow us to maintain the block hierarchy existing in the system (see again Figure 2). This structure has a close similarity with the one presented in the IS formulation (see Section 3), and in fact this is not coincidental, as we will show in a more formal manner. Let us take for simplicity the case of two components or blocks, which have no coupling effects other than conductive interconnections. Therefore the global system matrix C is block

diagonal, whereas the G matrix has off diagonal blocks related to the interconnections between the two components. The notation will be slightly changed in order to give a clearer relation with the IS formulation, and D will be taken as zero for simplicity, but without loss of generality. The global system representation is

$$\begin{gathered} C\dot{x} + Gx = Mu \\ y = Nx \end{gathered} \tag{16}$$
$$C = \begin{bmatrix} C_{11} & 0 \\ 0 & C_{22} \end{bmatrix} \quad G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \quad M = \begin{bmatrix} M_1^T & M_2^T \end{bmatrix}^T \quad N = [N_1 \;\; N_2]$$

The system can be rewritten in the following expressions

$$C = \begin{bmatrix} C_{11} & 0 \\ 0 & C_{22} \end{bmatrix} = C_D$$

$$G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} = \begin{bmatrix} G_{11} & 0 \\ 0 & G_{22} \end{bmatrix} + \begin{bmatrix} 0 & B_1 K_{12} L_2 \\ B_2 K_{21} L_1 & 0 \end{bmatrix} = G_D + BKL \tag{17}$$

$$M = \begin{bmatrix} M_1^T & M_2^T \end{bmatrix}^T = \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \end{bmatrix} \begin{bmatrix} H_1^T & H_2^T \end{bmatrix}^T = BH$$

$$N = [N_1 \;\; N_2] = [F_1 \;\; F_2] \begin{bmatrix} L_1 & 0 \\ 0 & L_2 \end{bmatrix} = FL$$

where

$$\begin{array}{ll} G_{12} = B_1 K_{12} L_2 & G_{21} = B_2 K_{21} L_1 \\ M_1 = B_1 H_1 & M_2 = B_2 H_2 \\ N_1 = F_1 L_1 & N_2 = F_2 L_2 \end{array} \tag{18}$$

The simplest way to achieve this decomposition is to set B_i and L_i equal to the identity, and thus $K_{12} = G_{12}$, $K_{21} = G_{21}$, $H_i = M_i$ and $F_i = N_i$. A different approach could be to perform a Singular Value Decomposition (SVD) of these blocks, i.e. $G_{12} = U_{12}\Sigma_{12}V_{12}$, with U and V orthonormal matrices containing the singular vectors associated to the singular values contained in the diagonal matrix Σ, and thus $B_1 = U_{12}$, $L_2 = V_{12}$, and $K_{12} = \Sigma_{12}$. This would also allow a certain degree of compression in the number of ports of each sub-system if a low rank approximation, based on the singular values, is applied. Of course there may be many other solutions depending on the matrices.

Note now that the description in (16) is exactly the same representation achieved via the IS interconnection of the systems

$$\begin{array}{ll} C_{11}\dot{x} + G_{11}x = B_1 a_1 & C_{22}\dot{x} + G_{22}x = B_2 a_2 \\ b_1 = L_1 x & b_2 = L_2 x \end{array} \tag{19}$$

where a_i and b_i are respectively the local inputs and outputs, with the interconnection matrices defined as

$$K = \begin{bmatrix} 0 & K_{12} \\ K_{21} & 0 \end{bmatrix} \quad H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} \quad F = [F_1 \;\; F_2] \tag{20}$$

This sketch of a proof, presented here for the simple case of two sub-systems, can be extended to an arbitrary number of subsystems, and shows that the interconnection of any circuit can be presented as a block structured system.

Let us see what happens when a Krylov-based Block Structure Preserving reduction is applied to this interconnected system. Consider a projector V that spans a suitable subspace for the reduction of the global (structured) system. As shown in Section 4.1, this projector can be split and further expanded following the system structure, in order to preserve such block structure after the reduction. This leads to

$$V \to \breve{V} = \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix}$$

$$\widehat{C} = \breve{V}^T C \breve{V} = \begin{bmatrix} V_1^T C_{11} V_1 & 0 \\ 0 & V_2^T C_{22} V_2 \end{bmatrix} \qquad \widehat{M} = \breve{V}^T M = \begin{bmatrix} M_1^T V_1 & M_2^T V_2 \end{bmatrix}^T \tag{21}$$

$$\widehat{G} = \breve{V}^T G \breve{V} = \begin{bmatrix} V_1^T G_{11} V_1 & V_1^T G_{12} V_2 \\ V_2^T G_{21} V_1 & V_2^T G_{22} V_2 \end{bmatrix} \qquad \widehat{N} = N \breve{V} = [N_1 V_1 \;\; N_2 V_2]$$

If now the decomposition presented in (17) is applied, it is straightforward that the reduction in (21) is equivalent to the independent reduction of the systems in (19), with V_1 and V_2 respectively, followed by interconnecting them again with the matrices in (20).

This proves that the BSP reduction of the system is equivalent to the Krylov-based reduction of the interconnected systems as presented in [14] (this fact was proved numerically via several simulations, but the results are not presented as we feel the theoretical proof sketched here is stronger and more appealing).

However, it is important to recall that the BSP scheme allows the reduction of systems when there are couplings between the sub-system blocks (i.e. there are non-zero off-diagonal blocks in the global C matrix). For this reason, it can be concluded that the BSP reduction framework is, in fact, a generalization of the Krylov-based IS scheme.

4.3. PMTBR connection

The results in the previous section show a direct relation between both BSP and Krylov-based IS frameworks. In Sections 2.1 and 2.2 the advantages and drawbacks of both moment matching and TBR were presented, which can be extended to the current case.

The Krylov approach is highly efficient, but it has poor control over the accuracy and order of the blocks. Furthermore, all the blocks must be reduced to the same size, what can become a serious drawback when the complexity of the models differ a lot (and therefore very different reduced orders are needed for their accurate modeling). The balanced truncation approach seems to have better performance, in particular in the case of interconnected systems, in which the trade off between the order and the accuracy can be done sub-system wise. However, the high cost required for the computation leads one to outright discard this method in an EDA framework.

It is important to notice that any projection-based MOR procedure can be extended in the BSP manner to maintain the hierarchical structure of a system. This includes the case of the PMTBR algorithm, where additional characteristics of the procedure can be further taken advantageous of in the current framework. In Section 2.3 it was shown that if the quadrature scheme (7) is accurate enough, then the estimated Gramian $\widetilde{X}$ converges to the original one X, which implies that the dominant eigenspace of $\widetilde{X}$ converges to the dominant eigenspace of X. If the system has some internal structure, then the matrix Z computed from the vector samples of the global system can be split into several

blocks. The estimated Gramian can be written block-wise as

$$\begin{bmatrix} Z_1 \\ \vdots \\ Z_{N_b} \end{bmatrix} \rightarrow ZZ^H = \begin{bmatrix} Z_1 Z_1^H & \dots & Z_1 Z_{N_b}^H \\ \vdots & \ddots & \vdots \\ Z_{N_b} Z_1^H & \dots & Z_{N_b} Z_{N_b}^H \end{bmatrix} = \widetilde{X} \tag{22}$$

But if we expand the matrix Z into diagonal blocks

$$\breve{Z} = \begin{bmatrix} Z_1 & & \\ & \ddots & \\ & & Z_{N_b} \end{bmatrix} \rightarrow \breve{Z}\breve{Z}^H = \begin{bmatrix} Z_1 Z_1^H & & \\ & \ddots & \\ & & Z_{N_b} Z_{N_b}^H \end{bmatrix} = \breve{X}. \tag{23}$$

From (22) it can be seen that $Z_i Z_i^H = \bar{X}_{ii}$, i.e. the matrix $\breve{X} = \breve{Z}\breve{Z}^H$ is a block diagonal matrix whose entries are the block diagonal entries of the matrix $\widetilde{X}$. Under a good quadrature scheme, the matrix $\widetilde{X}$ converges to the original X, and therefore $\breve{X}$ will converge to the block diagonals of X.

This means that the dominant eigenspace of $\breve{X}$ converges to the dominant eigenspace of the block diagonals of X. We can then apply an SVD to each block of the Z matrix

$$Z_i = V_i S_i U_i \;\rightarrow\; \breve{X}_{ii} = \widetilde{X}_{ii} = V_i S_i^2 V_i^T \tag{24}$$

where S_i is real diagonal, and V_i and U_i are unitary matrices. The dominant eigenvectors of V_i corresponding to the dominant eigenvalues of S_i can be used as a projection matrix in a congruence transformation over the system matrices for model order reduction. The elements of S_i can also be used for a simple *a posteriori* error estimation in a fashion similar to how Hankel singular values are used in TBR procedures.

Using these block projectors V_i, a structure preserving projector for the global system ($\breve{V}$) can be built (14) which will capture the most relevant behavior of each block (revealed by the SVD) with respect to the global response (recall that Z is composed of sample vectors of the complete system).

This approach provides us with more flexibility when reducing a complete system composed of multiple blocks and the interactions between them, as it allows us to individually control the reduced size of each sub-system, i.e. block-wise, via an error estimation based on the global input-output response (and unlike the Krylov-based procedure, only the strictly necessary order for the reduction of each sub-system is needed). The procedure is analogous to the ISBT one presented in Section 3.3 (for further details see [14]), which provides another link between the BSP and IS frameworks, but unlike the latter, the PMTBR-based BSP does not incur the excessively large cost of the ISBT when applied to large systems. Furthermore, the existence of EM couplings is no longer an issue, since the original system is formulated as a global block structured state space description (and thus the projection is applied on the global system). Passivity is guaranteed (as long as the PRIMA conditions for passivity are fulfilled by the global structured system), since the reduction is done via a congruence transformation. Further advantages inherent to the PMTBR framework, such as weighting and frequency selection, are also extendable in a straightforward manner.

5. RESULTS

In this section we show the results of applying the technique presented to several systems.

Table I. Characteristics of the first example: SPIRAL and CMIM

Block	Original	PMTBR	BS PMTBR	PRIMA	BS PRIMA
	Size |SR |#I/O	Size |SR |#I/O	Size |SR |#I/O	Size |SR |#I/O	Size |SR |#I/O
1st (RC line)	101 | 0.029 | 2	–	5 | 1.00 | 2	–	14 | 1.00 | 2
2nd (Spiral)	4961 | 9e-4 | 2	–	39 | 1.00 | 2	–	60 | 1.00 | 2
3rd (RC line)	101 | 0.029 | 2	–	5 | 1.00 | 2	–	12 | 1.00 | 2
4th (CMIM)	6044 | 8e-4 | 2	–	18 | 1.00 | 2	–	60 | 1.00 | 2
Global	11207 | 4e-4 | 2	44 | 1.00 | 2	67 | 0.636 | 2	60 | 1.00 | 2	146 | 0.567 | 2

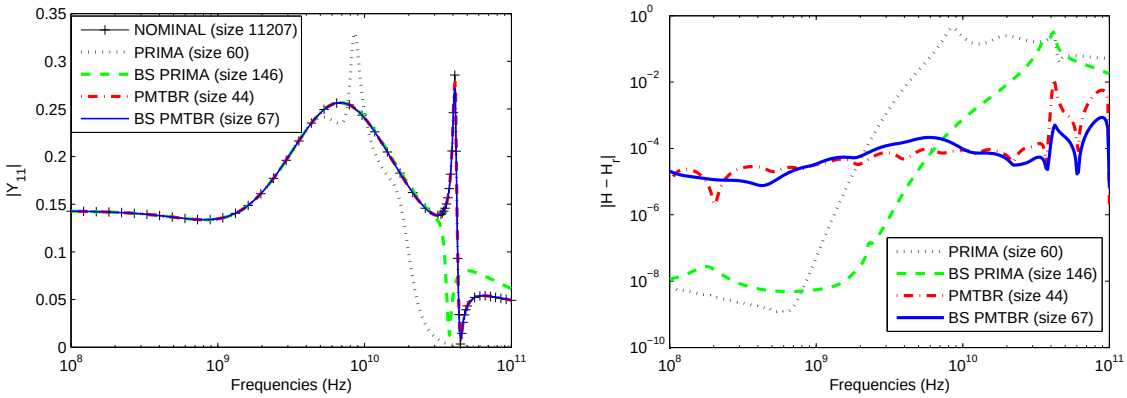

Figure 3. SPIRAL and CMIM example: (Left) $|Y_{11}|$ versus the frequency for the original model and the several ROMs. (Right) Infinity Error ($\|H - H_r\|_\infty$) of the ROMs versus the frequency.

5.1. Spiral and CMIM

This first example we discuss is composed of four sub-systems: the first one is a RC lumped model of an interconnect line, the second is a Full-Wave EM model of a integrated Spiral, including surrounding substrate and air, the third is another RC lumped model of an interconnect line, and the forth is another Full-Wave EM model, in this case of a Metal-Insulator-Metal (MIM) capacitor. These four models are connected, generating a four block global model. Table I shows the relevant characteristics of the system, in which each sub-system is characterized by its number of states (Size), its sparsity ratio (SR), and its number of terminals (#I/O). These features are also presented for the global system (Global), and for each of the reduction procedures that are benchmarked.

For the reduction, we are going to benchmark standard MOR techniques based on Moment Matching (PRIMA [4]) and on TBR approximation (PMTBR [11]), against Block Structure Preserving (BS) approaches with the same underlying reduction procedures (i.e. BS PRIMA and BS PMTBR). It is important to recall here that the BS approaches and the IS approaches are exactly the same here, since there are no EM couplings between subsystems, only electric interconnections.

For the PRIMA approaches, 30 block moments are computed, whereas PMTBR approaches are reduced with 15 frequency samples and a tolerance of 1e-3. In the case of BS PMTBR this tolerance is applied block-wise. From Table I we can see the block sizes in the BS approaches (in the standard

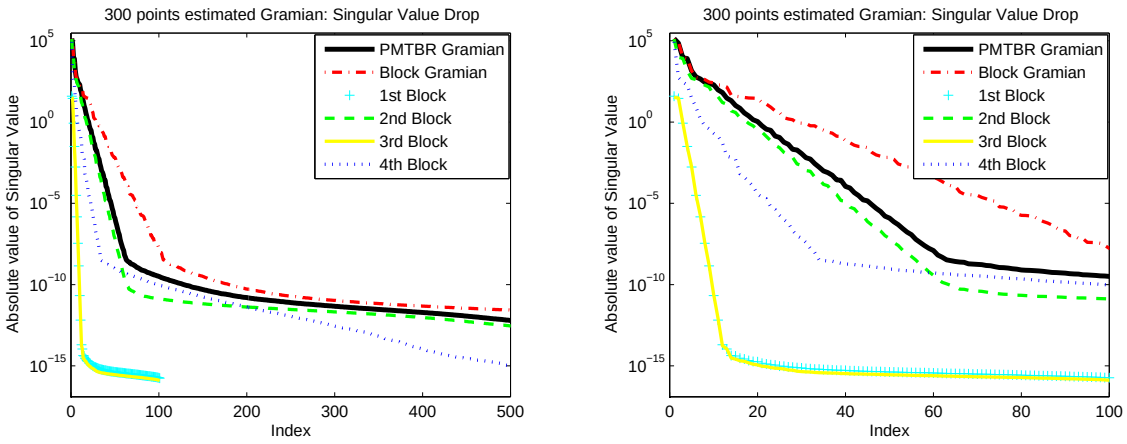

Figure 4. SPIRAL and CMIM example: (Left) drop of the singular values for the PMTBR estimated Gramian, the BS PMTBR estimated Gramian, and the several Block Gramians. (Right) detail of the leading Singular Values.

approaches the inner structure is lost, and thus the fields in the table are empty). Smaller sizes in blocks 1 and 3 for the PRIMA based framework are due to deflation in the basis orthonormalization. The block sizes of the BS PMTBR procedure are determined automatically for the fixed tolerance. The blocks are full after the reduction, but the complete system still retains a certain degree of sparsity, and the sub-systems are still distinguishable.

Figure 3 shows the frequency based transfer function of the original model and the ROMs. It also presents the error, measured as the infinite norm $\|H - H_r\|_\infty$ at each frequency point. It can be seen that, while accurate at lower frequencies, local expansion Moment Matching approaches have less overall accuracy. Sampling based approaches behave better along the whole frequency range. Multipoint moment matching approaches may be applied here with improvements on the accuracy. Now let us study the error control features of the BS PMTBR approach. Figure 4 shows a plot of the magnitude of the singular values (SV) for the PMTBR approach, the BS PMTBR approach, and for each of the individual blocks (note that since systems in blocks 1 and 3 are the same, the singular values are quite similar, and thus the curves almost overlap one another). These SV are obtained by estimating the (global) gramian with 300 samples. It can be seen that the PMTBR SV "drop" is limited by the maximum among the blocks' SV (i.e., the curve is above the maximum of the individual blocks' SV but close to them). On the other hand, the BS PMTBR SV "dropt" is exactly the same as the contribution of the blocks' SVs (this is logical, as these are obtained from the SVD of the block diagonal Gramian), and thus the number of SV retained for the same tolerance is increased, but the curve exhibits the same behaviour as the one of the PMTBR. Furthermore, a detailed study shows that the first (dominant) SVs are the same for both approaches. Figure 5 shows the convergence of the Singular Values as we increase the number of sampling points used in the estimation of the Gramian. It is clear that both approaches converge at a similar ratio.

Figure 6 shows how the error in the transfer function approximation decreases along with the tolerance settled for the reduction, with an automatic order selection (order is presented along the horizontal axis). Curves are shown for both approaches, PMTBR and BS PMTBR. It is clear that PMTBR generates models with higher compression, but the interesting thing is how the error can be

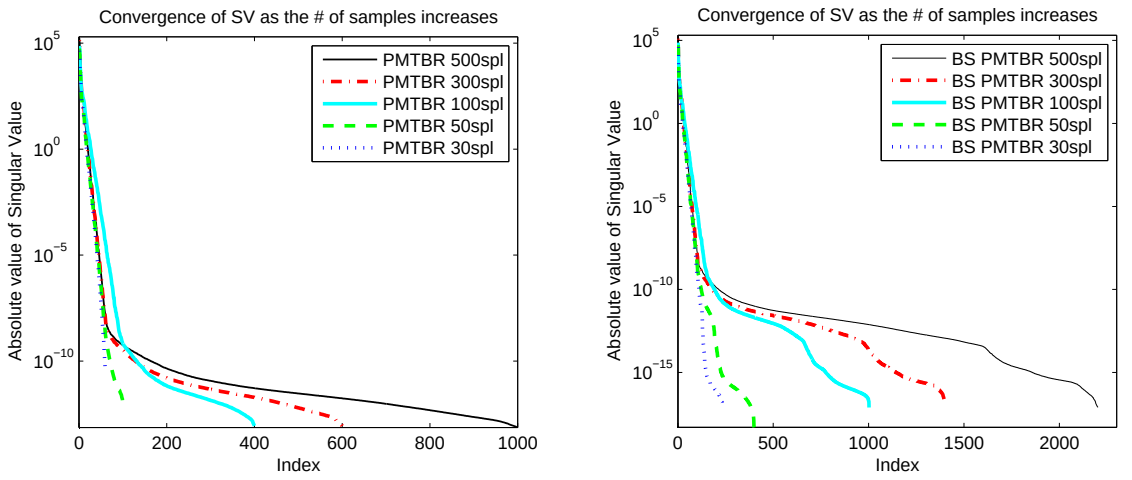

Figure 5. SPIRAL and CMIM example: Convergence of the Singular Value with the increase in the number of sampling points: (Left) for standard PMTBR, and (Right) for BS PMTBR.

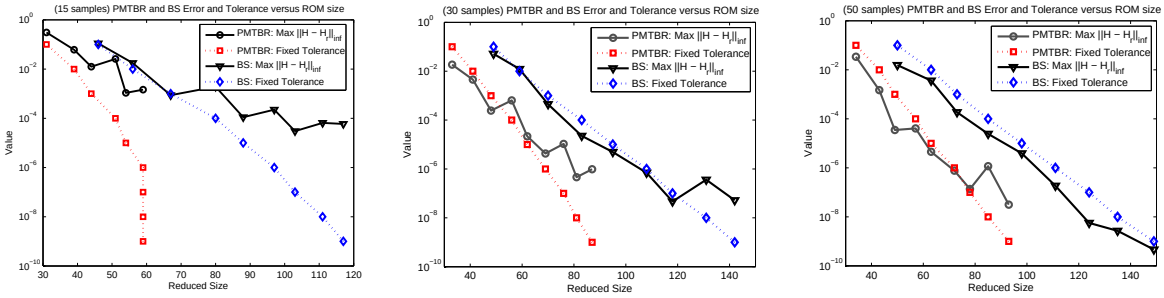

Figure 6. SPIRAL and CMIM example: Comparison of the behaviour of the tolerance settled and the error incurred with such tolerance with an automatic order selection, versus the ROM size. (Left) For 15 samples, (Center) for 30 samples, and (Right) for 50 samples.

controlled with the blocks' singular values for the BS PMTBR approaches, and thus generating an almost optimal block-wise compression based on global I/O relevance. Furthermore, the convergence of the BS PMTBR approach for small tolerances seems faster. This is because for a small number of samples, the standard PMTBR approach does not have enough vectors to generate a good model for the global system, whereas the block-wise approach only needs to generate basis for the blocks, which can be approximated with a smaller number of vectors. This fact potentially reduces the number of samples needed for achieving a desired error.

5.2. SPIRAL over N-Well

The second example is a system taken from an industrial circuit, which has an integrated Spiral Inductor over an N-Well. The system is modeled via Full-Wave Finite Integration Technique (FIT), with Domain Decomposition (for further details see [13]). The complete domain is split into two sub-domains, the

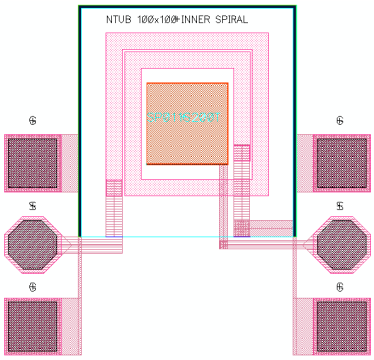

Figure 7. Layout configuration of the Spiral over N-Well example.

Table II. Characteristics of the second example: SPIRAL over N-Well

Block	Original Size |SR |#I/O	PMTBR Size |SR |#I/O	BS PMTBR Size |SR |#I/O	PRIMA Size |SR |#I/O	BS PRIMA Size |SR |#I/O
Top	34595 | $1.9e$-4 | 466	–	36 | 1.00 | 466	–	80 | 1.00 | 466
Bottom	16397 | $3.2e$-4 | 464	–	24 | 1.00 | 464	–	80 | 1.00 | 464
Complete	50992 | $1.4e$-4 | 2	36 | 1.00 | 2	60 | 1.00 | 2	80 | 1.00 | 2	160 | 1.00 | 2

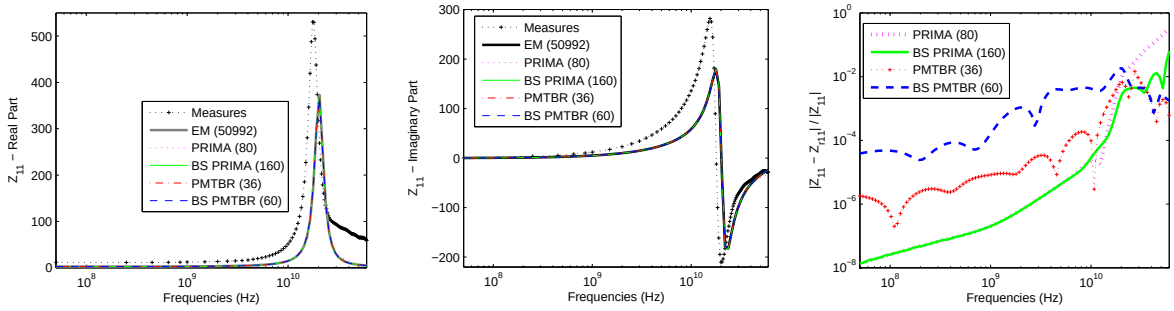

Figure 8. SPIRAL over N-Well example: (Left) Real part and (Centre) imaginary part of Z_{11}, for the measurements, the original EM model and the ROMs. (Right) Relative Error in $|Z_{11}|$ for the ROMs with respect to the original EM model.

Top one consists of the spiral, air and the more critical part of the substrate (including the N-Well), whereas the Bottom domain includes the lower homogeneous substrate bulk. The global ports are in the Top domain, one related to the spiral (one terminal connected and the other grounded), and the other one to the N-Well (See Figure 7 for the layout topology), which means that the Bottom domain does not have any external port. This domain is only related to the Top domain via EM couplings. The characteristics of the example system are shown in Table II, where the couplings between domains are included as terminals inside the corresponding block (464 coupling terminals connect both domains).

The BS methodologies generate very compressed models, with excellent accuracy in comparison

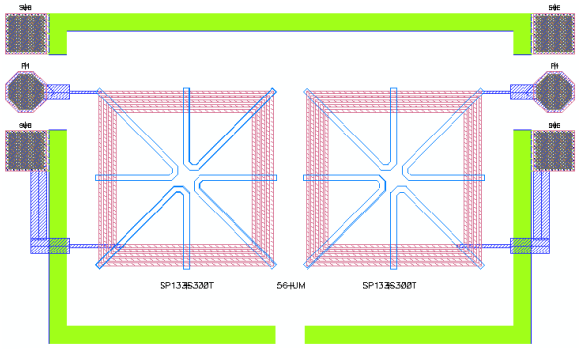

Figure 9. Layout configuration of the Double Spiral example.

Table III. Characteristics of the third example: Double Spiral

Block	Original			BS PMTBR			BS PRIMA		
	Size	SR	#I/O	Size	SR	#I/O	Size	SR	#I/O
Left	85616	$6.3e$-5	938	54	1.00	938	600	1.00	938
Middle	8027	$4.4e$-3	1874	59	1.00	1874	600	1.00	1874
Right	85629	$9.1e$-5	938	87	1.00	938	600	1.00	938
Complete	179272	$5.9e$-5	2	200	0.88	2	1800	0.89	2

with the original EM model. Figure 8 shows the value of the impedance at the spiral port as a function of the frequency. Experimental measurements are also shown in order to validate the EM modeling technique. It is important to notice that the BS PRIMA approach is more accurate around the expansion point (lower frequency), losing accuracy at higher frequencies for the given number of moments matched (40). The PMTBR based approach offers more compressed models, with independent compression for each block (36 and 24 in this example) based on its complexity and relevance on the global ports, and a better accuracy overall, although it requires a higher computational cost (due to its multi-point nature). For details on the reduction sizes, see Table II. Another interesting aspect of the methodology that Table II presents, is that the BS approaches manage to reduce the blocks to a size smaller than the number of local connectors of each sub-system, while preserving the structure of the system. Therefore the sub-system representation can be recovered (see Section 4.2 for a theoretical interpretation of such phenomena). This is possible because only global ports are taken into account, and is an interesting advantage, since individual projection based reduction of such sub-systems could never achieve such a degree of compression.

5.3. Double Spiral

Our third example is another industrial circuit, composed of two coupled integrated planar spirals. Each of the spiral ends represent one port, having one terminal voltage excited (intentional terminal, IT) and one terminal connected to ground (see Figure 9 for a layout topology). The complete domain, of size 179272, includes substrate and upper air, and is partitioned into three sub-domains, each of them connected to the others via a set of connectors (modeling both electric and magnetic interactions). In

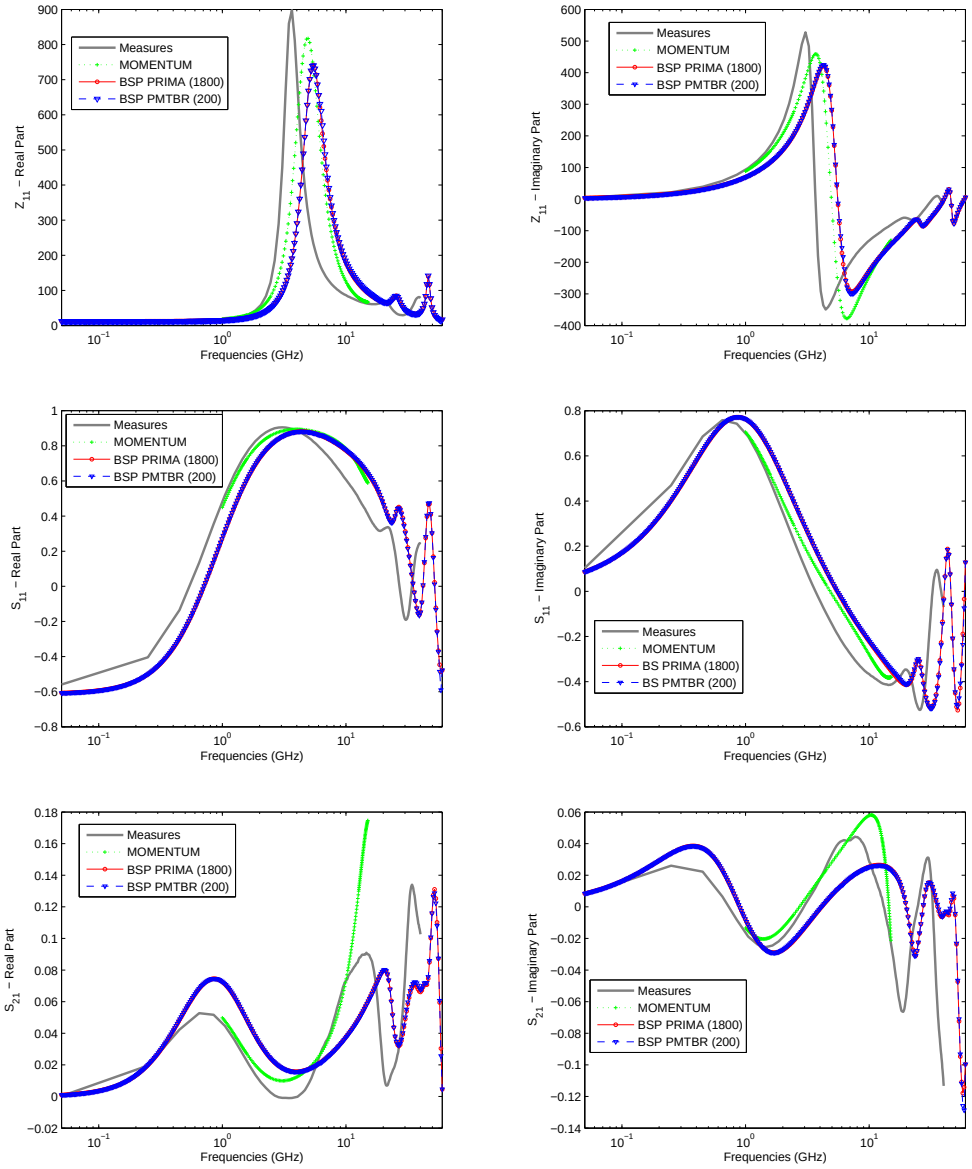

Figure 10. Coupled Double Spiral: Comparison of ROMs versus Experimental measurements and MOMENTUM results. (Top) Real (left) and Imaginary (Right) part of Z_{11}. (Center) Real (left) and Imaginary (Right) part of S_{11}. (Bottom) Real (left) and Imaginary (Right) part of S_{21}.

the left sub-domain we have the left spiral, in the middle domain the area between spirals, whereas in the right domain we have the right spiral. Each sub-domain includes the corresponding substrate and upper air layer. The Full Wave EM model was obtained via Finite Integration Technique (FIT) [13], and

its matrices present a Block Structure that follows the domain partitioning (see Table III for details).

Notice that the number of connectors or hooks for each domain is 938, whereas the global ports are only 2. Trying to reduce the domains independently with standard projection methodologies would lead to huge models (each block vector generated would have 938 vectors), and would not provide us with any simulation advantage. On the other hand, the BS methodologies manage to obtain very compressed models. In fact, in Table III it can be seen that the sizes of the blocks are much smaller than the number of connectors for each domain. This may look strange, but in fact we have maintained the block structure, and we are able to recover the original connection between sub-domains, while at the same time achieving a better compression by using the global ports. It is important to notice that, although the number of connectors between sub-domains is high, the off-diagonal blocks in the global reduced matrix (related to the couplings or interconnections) are of the adequate size and full. This is coherent with the results in Section 4.2.

The BS PMTBR based approach generates different orders for each domain, according to its complexity and relevance on the external ports, to yield a 200-dofs ROM. For the Krylov based approach, 300 moments are matched at a single expansion point, to generate a 1800-dofs ROM. Both methodologies preserve the block structure (provided by the domain decomposition) after the reduction stage. The frequency results for the impedance at one port, and the S-parameters, can be seen in Figure 10, in which the experimental measurements up to 40GHz, and the EM results using the commercial MOMENTUM tool (up to 15GHz) are included. The proposed methodology provides a very good agreement, and both BSP PRIMA and and BSP PMTBR provide highly compressed and accurate models.

6. CONCLUSIONS

In this paper we have presented guidelines for the reduction of a complete system described by a set of hierarchy aware EM models. The systems can be represented in a Block Structure manner, with entries related to the sub-systems in the diagonal blocks of the matrices, and interactions between sub-systems (i.e. interconnections or couplings) in the off-diagonal blocks. In order to preserve the initial hierarchical structure of the complete system, Block Structure Preserving Model Order Reduction methodologies were advocated. We have shown that these methodologies are generalizations of the systemic approaches based on Interconnected Systems' representation and reduction.

A novel approach that combines the BSP reduction with the PMTBR framework was introduced. The presented scheme, similarly to the TBR based Interconnected systems reduction (ISBT), allows us a individual reduction of the multiple sub-systems by using global I/O information. The reduction order of each sub-system can be controlled individually in terms of its global I/O relevance, by a simple *a posteriori* analysis of the Singular Values. The use of a BSP projection scheme allows us to model coupling effects, while maintaining the hierarchy and block sparsity patterns after the reduction step. Some examples were also presented, which showed the performance of the advocated approach.

ACKNOWLEDGEMENTS

The authors would like to thank Wim Schoenmaker (MAGWEL), Nick van der Meijs (T.U. Delft), Kees-Jan van der Kolk (T.U. Delft), Gabriela Ciuprina (LMN / P.U. Bucharest), Daniel Ioan (LMN / P.U. Bucharest), and Henk van der Vorst (Utrecht University) for many fruitful discussions and useful comments.

REFERENCES

1. Antoulas AC. *Approximation of Large-Scale Dynamical Systems.* Advances in Design and Control 6, SIAM, 2005.
2. Schilders WHA, ter Maten EJW. *Numerical Methods in Electromagnetics.* Amsterdam: Elsevier Science Publishers (North-Holland).
3. Achar R, Nakhla MS. Simulation of high-speed interconnects. *Inv. In Proc. of the IEEE*, May 2001; **89**(5):693–728.
4. Odabasioglu A, Celik M, Pileggi LT. PRIMA: passive reduced-order interconnect macromodeling algorithm. *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, Aug. 1998; **17**(8):645–654.
5. Feldmann P, Freund RW. Efficient linear circuit analysis by pad approximation via the Lanczos process. *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, May 1995; **14**(5):639–649.
6. Silveira LM, Kamon M, Elfadel I, White J. A coordinate-transformed Arnoldi algorithm for generating guaranteed stable reduced-order models of RLC circuits. *In International Conference on Computer Aided-Design*, San Jose , CA., USA, Nov. 1996; 288–294.
7. Elfadel IM, Ling DL. A block rational Arnoldi algorithm for multipoint passive model-order reduction of multiport RLC networks. *In International Conference on Computer Aided-Design*, San Jose , CA., USA, Nov. 1997; 66–71.
8. Moore B. Principal component analysis in linear systems: controllability, observability and model reduction. *IEEE Trans. on Automatic Control*, Feb. 1981; **AC-26**(1):17–32.
9. Phillips JR, Daniel L, Silveira LM. Guaranteed passive balancing transformations for model order reduction. *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, Aug. 2003; **22**(8):1027–1041.
10. Glover K. All optimal Hankel-norm approximations of linear multivariable systems and their L^{∞}-error bounds. *Int. J. Control*, June 1984; **39**(6):1115-1193.
11. Phillips JR, Silveira LM. Poor man's TBR: A simple model reduction scheme. *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, Jan. 2005; **24**(1):43–55.
12. loan D, Schilders W, Ciuprina G, Meijs N, Schoenmaker W. Models for integrated components coupled with their EM environment. *International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, 2008; **27**(4):820–829.
13. Ciuprina G, loan D, Mihalache D. Magnetic Hooks in the Finite Integration Technique: a Way towards Domain Decomposition. *IEEE Conference on Electromagnetic Field Computation*, Athens, Greece, May. 2008;
14. Vandendorpe A, Van Dooren P. Model reduction of interconnected systems. *in Proc. of 16th International Symposium on Mathematical Theory of Networks and Systems* (MTNS 2004), Leuven, Belgium, July 2004; **THP3-4**.
15. Freund RW. SPRIM: structure-preserving reduced-order interconnect macro-modeling. *In IEEE Proc. International Conference on Computer Aided-Design*, San Jose, CA., USA, Nov. 2004; 80–87.
16. Yu H, He L, Tan SXD. Block structure preserving model order reduction. *In IEEE Behavioral Modeling and Simulation Wokshop*, Sept. 2005; 1–6.
17. Li R-C, Bai Z. Structure-preserving model order reduction using a Krylov subspace projection formulation. In *Communications in Mathematical Sciences*, Jin S. (eds). International Press, 2005; **3**(2):179–199.