Electromagnetic Simulations for Nanoscale RF Blocks

Wil H. A. Schilders, L. Miguel Silveira, *Senior Member IEEE*, and Jorge Fernández Villena

Abstract— Next-generation nano-scale RFIC designs have an unprecedented complexity and performance that will inevitably lead to costly re-spins and loss of market opportunities. In order to cope with this, efficient and accurate models of interconnects, integrated inductors, the substrate and devices, together with their mutual interactions, need to be developed. The key idea is that integrated devices can no longer be treated in isolation as the EM interactions due to proximity effects are becoming more relevant in the behavior of the complete system. EM simulations must also address these interactions, so new procedures and models able to be included in coupled simulation must be developed. But these simulations may become very expensive as the complexity of the system increases, so Model Order Reduction techniques able to treat these coupling effects are necessary in order to obtain a better performance. In this work some solutions for efficient simulation of such problems are introduced.

Index Terms— EM simulation, model order reduction, interconnected systems, structured model order reduction.

I. INTRODUCTION

Next generation designs will be challenged by an increased number of trouble spots, many of which negligible at lower frequencies but representing a significant limitation for future designs. These trouble spots will have to be accounted for during the design phase in order to avoid costly mishaps that can originate potential failures, and additional design and silicon iterations, and must be addressed in future design automation tools.

New coupling and loss mechanisms, including EM field coupling and substrate noise as well as process-induced variability, are becoming too strong and too relevant to be neglected, whereas more traditional coupling and loss mechanisms are more difficult to describe given the wide frequency range involved and the greater variety of structures to be modeled. All this will cause extra design iterations, over-dimensioning or complete failures, unless appropriate solutions are found to resolve these design issues.

The performance of each device in the circuit is strongly affected by the environmental situation surrounding it. In other words, the response of each circuit part depends not only on its own physical and electrical characteristics, but to a great extent also on its positioning in the IC, i.e. on the devices to which it is connected. Therefore complete RF blocks must be considered as one entity, and be treated as such by the design automation tools. Today, it is not possible to perform such analyses of complete RF blocks. Electromagnetic simulation of the complete circuit is far too costly. Accurate models for active and passive devices must be done separately, but then the interaction between these devices may be lost.

A solution is to obtain a two-level hierarchical model for the system, where not only the devices are modeled, but as well the interactions between them. Hence the devices can be treated as single models inside a complete higher level RF model. However, the models provided by the extractors are usually too large, so heavy computation effort is needed to simulate them. One solution is to apply Model Order Reduction (MOR) procedures. These techniques are aimed at the reduction of the system representation by maintaining in an accurate way the input-output response of the system, at the same time that certain physical properties of the circuit inherent to the model must be taken into account and be guaranteed, namely stability and passivity. In the past decades, several robust techniques for performing efficient reduction of electrical and passive circuits were developed. Nonetheless, these techniques are not able to handle efficiently these interactions between systems.

The structure of the paper is as follows: in Section II we will discuss some of the considerations needed to be taken into account in a two-level hierarchical EM simulation. In the Section III, an overview of the standard MOR techniques will be performed and in Section IV, a review of the possible procedures, including a proposed methodology for overcoming the interconnection limitation, will be explained. Finally, in Section V, some results are presented and conclusions are drawn in Section VI.

Wil H. A. Schilders is with NXP Semiconductors Research, High Tech Campus 37, 56656 AE, Eindhoven, The Netherlands (e-mail: wil.schilders@nxp.com). He is also with the Department of Mathematics and Computer Science, Technical University of Eindhoven, 5600 MB Eindhoven, The Netherlands (e-mail: w.h.a.schilders@tue.nl).

Jorge Fernández Villena is with INESC-ID/IST, Technical University of Lisbon, Rua Alves Redol 9, 1000-029 Lisbon, Portugal (e-mail: jorge.fernandez@inesc-id.pt).

Luis Miguel Silveira is with CADENCE Laboratories and INESC-ID/IST, Technical University of Lisbon, Rua Alves Redol 9, 1000-029 Lisbon, Portugal (e-mail: lms@inesc-id.pt).
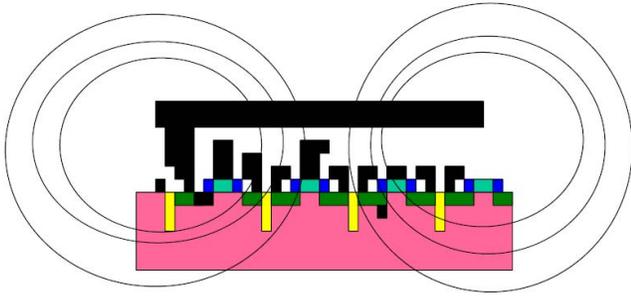
Fig. 1. *Active and Passive devices scale and coupling considerations in an electromagnetic environment.*

II. TWO LEVEL ELECTROMAGNETIC EXTRACTION

A. Component Level

The first consideration that must be addressed with the electromagnetic field aspects of coupling comes at the component level. The high level of integration that is being reached in nowadays RFIC design leads to proximity effects between the devices, both active and integrated passive, as well as with the interconnects. Therefore the EM extraction stage must be capable of generating compact models that incorporate those coupling effects. It is also important to notice that the EM fields can be spread over a large range. Another consideration that must be taken into account when modeling these two different kinds of devices is that the scale applicable in either cases is quite different. Fig. 1 shows a typical IC Manhattan profile, where the active device is much smaller than the spiral inductor over it. This scale difference also exists between different back-end (passive) devices. In these cases, an adaptive 3D mesh is necessary in order to obtain some coherent models for the devices.

B. Circuit Level

The second consideration refers to modeling the global interactions between the physical (on-chip) realizations of the circuit elements from the schematic. These elements include the active as well as the passive devices. The couplings to be modeled at this extraction level will attach to the specific connectors of the compact models that are the outcome of the device or component level extraction.

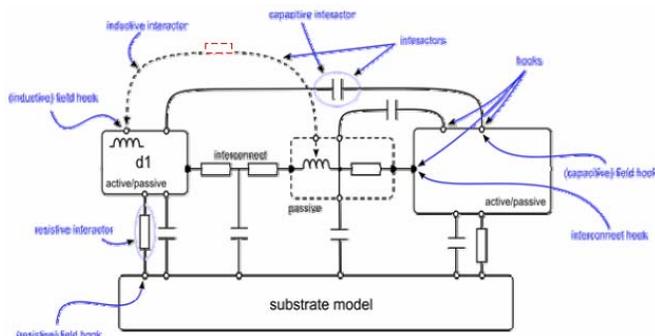Fig. 2 shows an example of the circuit level extraction information, where the interconnections and the couplings are included as a part of the schematic. At this point some topological information about the device situation must be kept, which is indicated by the interconnections and coupling interactions between the several devices that form the RF block. This step is crucial, as it gives the basis for an interconnected simulation of the complete block, which will determine the global behavior of the circuit.

III. REDUCTION STAGE

The two-level extraction step has provided us with two types of information. First, the compact models for linear (passive) components and active components. Second, the interconnection and coupling information that, in conjunction with the previous compact models, forms the RF block. It is possible now to apply model order reduction techniques over the passive devices, which will enable the generation of compressed representations of interconnected sets of compact models, including the interaction effects.

From this starting point, one approach could be to treat the whole circuit as a single model, i.e. obtain a global model for all the devices and their interactions, and from this global model, perform the reduction. Although this method would capture accurately the input-output behavior of the complete RF circuit, this is not the best approach, due to the fact that the global model is expected to be a very large one, and the computational effort needed for reducing it would be too big. What is more, if a direct reduction is made, the structure and hierarchy of the interconnection will be lost.

The second approach is to perform the reduction of the several models that are part of the global system in a hierarchical fashion. This means to reduce each model independently without taking into account the rest of the models or the environment. This has the advantages that every single model is reduced separately, which gives us a lot of flexibility. These models are usually smaller than the global system, so the reduction is computationally cheaper, and the hierarchy and structure of the global system is maintained. On the other hand, to apply MOR on each model means to capture its single behavior, not the global one. This can lead to the existence of effects of the global response that are not captured, or that too much effort is spent in capturing some model behavior that is not relevant at all for the global response (maybe filtered by another model). Another drawback is that if the number of interactions between models is high, the stand alone reduction becomes inefficient, because these interactions are translated into a lot of ports for every device. The existing MOR techniques have their efficiency reduced when the number of ports of the model grows, so it is preferable to have as few ports as possible. On the other hand the global system may have only a few ports, which makes the reduction easier. This can be seen in the next sub-sections.
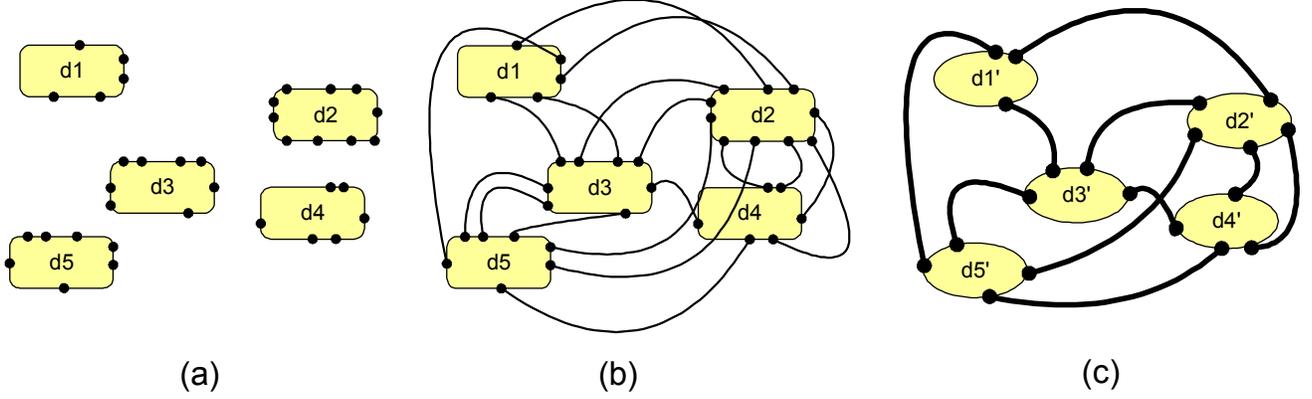

Fig. 2. *Example of circuit topology with interconnections and coupling effects between devices.*

Fig. 3 *Illustration of the proposed modeling and model order reduction procedures. (a) shows compact models of devices, such as transistors or inductors, able to account for the interaction, which is added in (b). This model is transformed into a simpler but still accurate model using structured model order reduction procedures in (c).*

A. Review of model order reduction

The main techniques in MOR are geared towards the reduction of state space linear time-invariant systems, generally representing some formulation of a physical system or the result of some previous extraction step. These descriptions represent the devices as a system where the *p* outputs are related to the *m* inputs via some inner *n* "states" in a differential algebraic system:

$$C \cdot \dot{x}(t) + G \cdot x(t) = B \cdot u(t)$$
$$y(t) = L \cdot x(t) + D \cdot u(t) \tag{1}$$

where $C \in \Re^{nxn}$, $G \in \Re^{nxn}$, $B \in \Re^{nxm}$, $L \in \Re^{pxn}$, $D \in \Re^{pxm}$ are the state space matrices, $u(t) \in \Re^{m}(t)$ is the vector of inputs, $y(t) \in \Re^{p}(t)$ is the vector of outputs and $x(t) \in \Re^{n}(t)$ is the vector of inner states. We can assume $D = [0]$ without loss of generality. The previous state-space formulation corresponds to the input-output transfer function in the frequency domain *s*

$$H(s) = L \cdot (C \cdot s + G)^{-1} \cdot B . \tag{2}$$

The two most common techniques for Model Order Reduction are based on either truncated balanced realization [2], [3] or projection schemes [4]-[8]. Of the former, the most common are the moment matching techniques, in which PRIMA [4] is usually reported as the standard.

B. Introduction to PRIMA algorithm

PRIMA is a projection method based on the computation of a basis for a Krylov subspace, which is afterwards applied over the original system in a congruence transformation to obtain the Reduced Order Model (ROM). The columns of the projection matrix are obtained from the block moments of the Transfer Function in the frequency domain. This means that the moments of the transfer function in (2) around the

expansion point are implicit matched.

If we denote the basis whose columns spans the Krylov subspace of the system at frequency $s = 0$ as $V \in \Re^{nxq}$, i.e.

$$Kr(A, R, q) = colsp[R, A \cdot R, A^2 \cdot R, ...] = V$$
$$k = \left\lfloor \frac{q}{m} \right\rfloor \quad A = -G^{-1} \cdot C \quad R = G^{-1} \cdot B \tag{3}$$

Then the reduced system whose matrices are

$$\hat{C} = V^T \cdot C \cdot V \quad \hat{B} = V^T \cdot B$$
$$\hat{G} = V^T \cdot G \cdot V \quad \hat{L} = L \cdot V \tag{4}$$

matches *k* block moments of the original system. It is important to notice that the number of columns of every block moment is equal to the number of inputs, *m*, so for a higher number of inputs a higher number of columns (and therefore a larger ROM) is needed for matching the same number of moments. Another important property of PRIMA is that under certain conditions (which in the case of electrical circuits are almost always true) the ROM is passive if the original system is passive (for further details see [4]).

However, this kind of reduction based on projection of the matrices leads to the lost of the structure of the system and full reduced matrices. If it is applied over the global system, the original two-level structure with the interconnections and couplings can no longer be recovered. This implies a loss of some advantages at the simulation stage. However, if it is applied on the stand alone devices, the structure of the interconnections is still kept, although inside the device matrices the inner structure is lost. Some effort has been spent in trying to maintain the inner structure of the systems [10] but it is only possible to maintain certain block structure.

IV. PROPOSED APPROACH: STRUCTURED REDUCTION

From the above discussion, it seems that neither stand alone reduction nor global reduction is the best choice when trying to reduce a two-level structure system as the one we have.

Another option is to reduce every single device separately but oriented to capture the global input-output response (as shown in Fig. 3). This approach will provide us with more control in the reduction stage while the structure of the interconnections is maintained. The transfer function to match is the global one, so the most relevant behaviors for the complete RF block are captured. What is more, only the global inputs and outputs of the complete RF block are relevant, so the inefficiencies caused by the large number of ports of the several devices are avoided.

A. Interconnected systems

This approach was already pursued in [1], but the work developed there had a system viewpoint, where the interconnections were signal flows between the systems. This fact leads to some relevant drawbacks in our context. In [1] the component systems are defined as sub-systems, with their local inputs and outputs. A new set of matrices are presented, defined as interconnection matrices, which relates the local inputs and outputs between them, and to the global inputs and outputs. If there is a set of N_S interconnected sub-systems, each one with α_i local inputs $a_i(s)$, and β_i outputs $b_i(s)$, and the global interconnected system has the external inputs $u(s)$ and outputs $y(s)$, the generic relation between the set of inputs and outputs is :

$$a_i(s) = u_i(s) + \sum_{j=1}^{N} K_{i,j} \cdot b_j(s)$$

$$u_i(s) = H_i \cdot u(s) \tag{5}$$

$$y(s) = \sum_{i=1}^{N} F_i \cdot b_i(s)$$

where $H_i \in \Re^{\alpha_i x m}$ is the matrix that relates the global input ports, i.e. $u(s)$ to the input ports, i.e. $a_i(s)$, of the *i-th* sub-system, $F_i \in \Re^{p x \beta_i}$ relates output ports of the *i-th* sub-system, i.e. $b_i(s)$, to the global outputs, i.e. $y(s)$ and $K_{i,j} \in \Re^{\alpha_i x \beta_j}$ relates the *j-th* sub-system output ports, i.e. $b_j(s)$, to the *i-th* sub-system input ports, i.e. $a_i(s)$.

From these matrices and the state space matrices of each sub-system, a global relation between inputs $u(s)$ and outputs $y(s)$ can be obtained:

$$y(s) = F \cdot (I - T(s) \cdot K)^{-1} \cdot T(s) \cdot H \cdot u(s). \tag{6}$$

Here $T(s)$ is the transfer function of the diagonal system obtained with the sub-system matrices, i.e.

$$b(s) = T(s) \cdot a(s)$$

$$T(s) = L \cdot (C \cdot s + G)^{-1} \cdot B \tag{7}$$

where, $a(s)$ and $b(s)$ are vectors with all the local inputs and outputs of all the sub-systems, and C, G, B, and L are block diagonal matrices with the *i-th* sub-system matrices C_i, G_i, B_i, and L_i on the *i-th* diagonal block.

The relation between inputs and outputs in (6) can be expressed via the state space descriptor

$$C_G \cdot \dot{x}(t) + G_G \cdot x(t) = B_G \cdot u(t)$$

$$y(t) = L_G \cdot x(t) \tag{8}$$

where $C_G = C$

$$G_G = G - B \cdot K \cdot (I - D \cdot K)^{-1} \cdot L$$

$$B_G = B \cdot (I - K \cdot D)^{-1} \cdot H$$

$$L_G = F \cdot (I - D \cdot K)^{-1} \cdot L .$$

This means that we can obtain a global system representation for the complete RF block starting from the description of the several sub-systems and the interconnections.

The work in [1] presents two reduction approaches. The first one is based on Balanced Truncation (BT) procedures, by reducing every single system independently, starting from the global Gramian. The second relies on Moment Matching approaches, and presents a series of theorems that guarantees the number of moments matched with respect the global transfer function (6) when the single devices are reduced (for further details see [1]). The BT approaches, although purportedly being able to produce quasi-minimal realizations, are very expensive for large systems, so it seems inefficient to apply them in this case. On the other hand the Krylov approaches are very efficient. However, this interconnected formulation has one drawback that limits its application inside EM simulations: although giving a nice theoretical basis, the system viewpoint limits the interactions between the sub-systems to signal flows, i.e. physical connections, which makes it difficult to apply these techniques to the case where capacitive couplings between the systems take place.

B. Block Structure Preserving

The Block Structure preserving (BSP) technique was first presented in [10] and later generalized in [11]. The main idea there was to maintain some block structure after the reduction via projection. This allows a more efficient treatment in the reduction and the maintenance of certain system properties, such as some degree of sparsity, and the block hierarchical structure after a reduction.

The procedure relies on expanding the projector (obtained via any classical MOR projection technique) of the global system into a block diagonal matrix, with block sizes equal to the size of the block.

The system is supposed to have some hierarchical structure so that it is possible to divide it in Nb blocks.
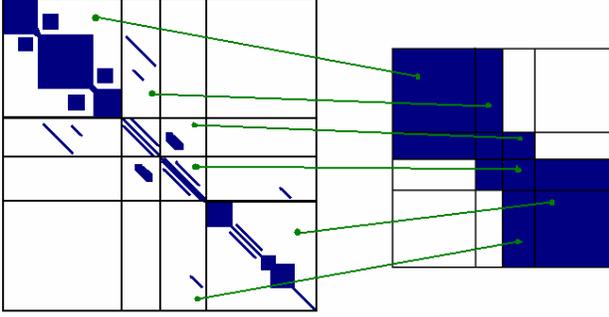
Fig. 4. Block Structure of the matrices, where the diagonal blocks are related to the device, and the off-diagonal blocks are related to the interconnections or couplings, and effect of the BSP reduction.

$$G = \begin{bmatrix} G_{1,1} & \cdots & G_{1,Nb} \\ \vdots & \ddots & \vdots \\ G_{Nb,1} & \cdots & G_{Nb,Nb} \end{bmatrix} \quad B = \begin{bmatrix} B_1^T & \cdots & B_{Nb}^T \end{bmatrix}^T$$

$$C = \begin{bmatrix} C_{1,1} & \cdots & C_{1,Nb} \\ \vdots & \ddots & \vdots \\ C_{Nb,1} & \cdots & C_{Nb,Nb} \end{bmatrix} \quad L = \begin{bmatrix} L_1 & \cdots & L_{Nb} \end{bmatrix}$$

(9)

In this way we can obtain the Krylov subspace that spans the combination of k block moment vectors generated by different sources in the circuit, as shown in (3). Any basis that spans this subspace is a suitable projector. Let us denote by $V \in \Re^{nxq}$ the usual orthonormal basis of the Krylov Subspace. We can split it as follows:

$$V = \begin{bmatrix} V_1 \\ \vdots \\ V_{Nb} \end{bmatrix} = colsp[Kr(A, R, q)] \tag{10}$$

This projector can be restructured so that

$$\tilde{V} = \begin{bmatrix} V_1 & & 0 \\ & \ddots & \\ 0 & & V_{Nb} \end{bmatrix} = colsp[Kr(A, R, q)] \tag{11}$$

where $\tilde{V} \in \Re^{nxNb \cdot q}$. Using it, a projection on the system matrices is performed via a congruence transformation

$$\begin{aligned} \hat{C} &= \tilde{V}^T \cdot C \cdot \tilde{V} & \hat{B} &= \tilde{V}^T \cdot B \\ \hat{G} &= \tilde{V}^T \cdot G \cdot \tilde{V} & \hat{L} &= L \cdot \tilde{V} \end{aligned} \tag{12}$$

It should be noticed that the projection matrix $\tilde{V} \in \Re^{nxNb \cdot q}$ used in this reduction context has Nb (number of blocks) times more columns than the original matrix $V \in \Re^{nxq}$. This leads to a Nb times larger reduced system. On the other hand, this technique maintains the structure of the original

system and gives us some flexibility when choosing the size of the reduced model depending on the block layout and relevance. The projector block V_i is only used in the reduction of the blocks $C_{i,x}$ and $C_{x,i}$ (the same in G), so each reduced block will match k block moments of the original one, and the complete system will be able to match up to $Nb \cdot k$ block moments of the original complete transfer function under the best conditions (i.e. with very weak entries in the off-diagonal blocks). Under the worst conditions, only k block moments are matched, i.e. the same number than in the "flat" reduction.

The global system obtained in the two-level extraction, composed of the compact models and the interactions, has an useful block structure, where the diagonal blocks correspond to the device matrices, whereas the off-diagonal blocks correspond to the interconnections (in the G matrix) or capacitive couplings (C matrix). It must be noticed that this point of view of the complete RF block allows the treatment of capacitive couplings between the systems. The BSP technique can be therefore applied so that the block structure is maintained. On the other hand, the inner structure of the device matrices is destroyed and lost, turning any non-empty block in the original system into a full block, but it is still possible to identify the blocks and relate them to the original device or interaction block. Nevertheless, if any block is empty in the global system matrix, it remains empty after reduction, increasing the sparsity. Fig. 4 shows the structure of the original system and the effects of the structured reduction.

V. RESULTS

In this section a simple simulation example will be presented. The scheme is the one shown in Fig. 5, where 6 passive devices (of 500 RC segments), with different R and C values, represent a system (of size 3004). The devices are interconnected (lines) and coupled via capacitive effects (every Cc block has 50 capacitive couplings, affecting different nodes in the systems). The global system has four inputs and four outputs.

For the reduction, PRIMA and Block Structure (BS) preserving via PRIMA were applied. Table I shows the characteristics of the ROM, as well as the computational effort required for the reduction. In Fig. 6 we plot the real part of the
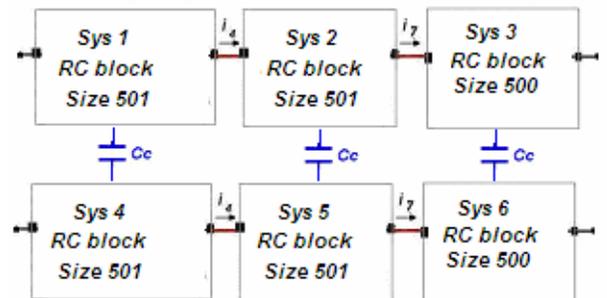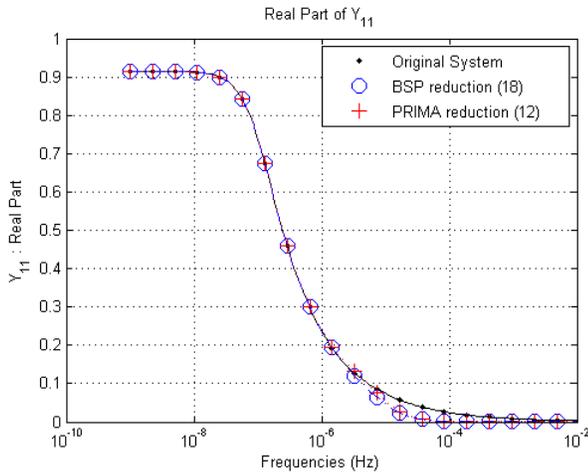

Fig. 5. Scheme of the Simulation Example

Fig. 6. Real part of the Admittance at the port 1 versus the frequency, for the original system, for PRIMA and for BS-PRIMA reduction.

admittance Y_{11} as a function of frequency for the original and reduced models. It can be seen that although the ROM size with the BS method is larger, for the same degree of accuracy, the computational effort (Block Moments computed and QR orthonormalization of the subspace) is smaller than in the case of the "flat" (non-structured) reduction with PRIMA. In this last case the structure is lost and the matrices are full, whereas in the BS approach the block structure is maintained as well as some degree of sparsity (the number of non-zero entries is smaller). For the BS approach, there is also an extra degree of flexibility, as different sizes can be applied when reducing each block (we can control the number of columns of each block projector).

VI. CONCLUSIONS

In this document a procedure for reducing the systems obtained from a two-level EM extraction in the presence of interconnections and couplings is presented. The approach described takes advantage of the inner structure of the extracted system, and provides more control and efficiency in the reduction stage. The block structure of the original system is kept, allowing the identification of the blocks related to each device and to the interactions between devices. This gives a clear advantage and can be used by the simulation tools to obtain more accurate results and higher efficiency.

ACKNOWLEDGMENT

TABLE I
SYSTEM CHARACTERISTICS AFTER REDUCTION

MOR Method	ROM Size	NNZ G	NNZ C	Comp. Effort	Structure
NONE	3004	9004	3286	---	TOTAL
PRIMA	12	144	144	3 Block Moments QR: 3004x12	NONE
BS-PRIMA	18	126	108	1 Block Moment QR: 4x (501x3) 2x (500x3)	BLOCK

REFERENCES

[1] A. Vandendorpe and P. Van Dooren, "Model Reduction of Interconnected Systems", *in Proc. of 16th International Symposium on Mathematical Theory of Networks and Systems* (MTNS 2004), Leuven, Belgium, pp. paper THP3-4, July, 2004.

[2] B. Moore, "Principal Component Analysis in Linear Systems: Controllability, Observability and Model Reduction", *IEEE Trans. on Automatic Control*, pp. 17-32, Vol AC-26, No. 1, Feb. 1981.

[3] J. R. Phillips, L. Daniel and L. M. Silveira, "Guaranteed Passive Balancing Transformations for Model Order Reduction", *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, pp. 1027-1041,Vol 22, No. 8, Aug. 2003.

[4] A. Odabasioglu, M. Celik and L. T. Pileggi, "PRIMA: passive reduced-order interconnect macromodeling algorithm", *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, pp. 645-654,Vol 17, No. 8, Aug. 1998.

[5] P. Feldmann and R. W. Freund, "Efficient Linear Circuit Analysis by Padé Approximation via the Lanczos Process", *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, pp. 639-649,Vol 14, No. 5, May. 1995.

[6] L. M. Silveira, M. Kamon, I. Elfadel and J. White, "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, pp. 288-294, Nov. 1996.

[7] I. M. Elfadel and D. L. Ling, "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, pp. 66-71, Nov. 1997.

[8] J. R. Phillips and L. M. Silveira, "Poor man's TBR: A simple model reduction scheme", *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, pp. 43-55,Vol 24, No. 1, Jan. 2005

[9] R. Achar and M. S. Nakhla, "Simulation of High-Speed Interconnects", Inv. *In Proc. of the IEEE*, pp. 693-728, Vol. 89, No. 5, May 2001.

[10] R. W. Freund, "SPRIM: Structure-Preserving Reduced-Order Interconnect Macro-Modeling", *In IEEE Proc. International Conference on Computer Aided-Design, San Jose, CA., USA*, pp. 80-87, Nov. 2004.

[11] H. Yu, L. He and S. X. D. Tan, "Block Structure Preserving Model Order Reduction", *In IEEE Behavioral Modeling and Simulation Wokshop* pp. 1-6, Sept. 2005.

[12] W. H. A. Schilders and E. J. W. ter Maten, Numerical Methods in Electromagnetics, Amsterdam: Elsevier Science Publishers (North-Holland).

Copyright Information