

A Compression Strategy for Rational Macromodeling of Large Interconnect Structures

S. Grivet-Talocia, S. B. Olivadese, P. Triverio

Dip. Elettronica, Politecnico di Torino

C. Duca degli Abruzzi 24, 10129 Torino, Italy

Ph +39 011 5644104, Fax +39 011 5644099

e-mail stefano.grivet@polito.it

Abstract—Rational macromodeling via Vector Fitting algorithms is a standard practice in Signal and Power Integrity analysis and design flows. However, despite the robustness and reliability of the Vector Fitting scheme, some challenges remain for those applications requiring models with a very large port count. Fully coupled signal and/or power distribution networks may require concurrent modeling of hundreds of simultaneously coupled ports over extended frequency bands. Direct rational fitting is impractical for such structures due to a large computational cost. In this work, we present a compression strategy aimed at representing the dynamic behavior of the structure through few carefully selected “basis functions”. We show that model accuracy can be traded for complexity, with full control over approximation errors. Application of standard Vector Fitting to the obtained low-dimensional compressed system leads to the construction of a global state-space macromodel with significantly reduced runtime and memory consumption. Several benchmarks demonstrate the effectiveness of the approach.

I. INTRODUCTION AND MOTIVATION

The importance of Signal and Power Integrity has increased significantly in the last two decades, due to the growing miniaturization and faster switching times achieved in digital and mixed-signal systems. Signal and power degradation typically arises in interconnect and power delivery networks where crosstalk, unwanted couplings, and resonances are more significant. Broadband simulation models are of paramount importance for system-level analysis, design, and verification.

A typical description of linear interconnect structures is through frequency-domain scattering responses obtained via full-wave simulations. The Vector Fitting (VF) algorithm [1] is a popular technique for casting these sampled scattering parameters into broadband simulation models for Signal and Power Integrity assessments. A possible issue of VF is the poor scalability with the complexity of the structure under modeling, in terms of model order N , which is directly related to the modeling bandwidth, port count P , and number of frequency samples L . It can be shown that the number of required operations per iteration scales as $O(P^2LN^2)$, and becomes infeasible when the number of ports or the model order is large.

Recently, it was shown that the VF computational cost can be significantly reduced by means of QR decomposition [2] and parallel computing [3]. However, the latter approach requires availability of high-performance hardware. In this

paper, we propose a new alternative for reducing the computational cost, based on a compression strategy. This solution is readily applicable to any version of VF, and makes the macromodeling of structure with hundreds of ports possible even on commodity hardware, avoiding the cost of a high-performance computing cluster.

We define the problem and the main idea of our compression strategy in Section II. The compression algorithm will be presented in Section III, followed in Section IV by a procedure to form the state-space realization of the compressed macromodel. Numerical results and comparisons on several benchmark cases will be finally presented in Section V.

II. PROBLEM STATEMENT

We consider a generic P -port electrical interconnect structure characterized through tabulated scattering frequency samples $\mathbf{S}_l \in \mathbb{C}^{P \times P}$ at frequencies ω_l , with $l = 1, \dots, L$. This raw data is usually available from field simulations or direct measurements. The VF algorithm is routinely used to fit these data samples with a rational model

$$\mathbf{S}(s) = \mathbf{S}_\infty + \sum_{n=1}^N \frac{\mathbf{R}_n}{s - p_n}, \quad (1)$$

where p_n are the poles of the macromodel, \mathbf{R}_n are the associated residue matrices, and \mathbf{S}_∞ is the direct coupling term. Standard formulations of the VF algorithm minimize the global model error

$$\sum_{l=1}^L \|\mathbf{S}(j\omega_l) - \mathbf{S}_l\|_F^2, \quad (2)$$

where $\|\cdot\|_F$ denotes the Frobenius norm, through an iterative sequence of linear least squares solutions, see [1].

We illustrate the main idea of our compression scheme through an example. Figure 1 depicts several scattering responses of a high-speed connector. We see that the various responses that are depicted look very similar, with only marginal differences. Of course, these differences may be important, so they should be preserved in the final macromodel. However, it is conceivable that all these responses can be represented as a linear superposition of selected “representative” responses or, more formally, “basis functions”. We will then look for

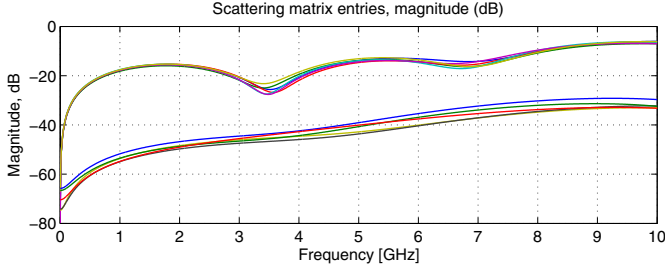


Fig. 1. Various scattering responses of a high-speed connector (top curves: reflection coefficients, bottom curves: crosstalks).

expansions of the form

$$S_{ij}(s) \simeq \sum_{q=1}^{\rho} \alpha_q^{(i,j)} w_q(s) \quad (3)$$

with constant coefficients $\alpha_q^{(i,j)}$ and frequency-dependent “basis functions” $w_q(s)$ to be determined. It is clear that if the number of required basis functions $w_q(s)$ is much smaller than the total number of responses, $\rho \ll P^2$, we can achieve a significant computational cost reduction by applying VF to the few functions $w_q(s)$, rather than to the complete set of P^2 raw scattering responses.

III. SVD-BASED COMPRESSION

We start with the set of raw scattering samples $\mathbf{S}_l, \forall l$. For each selected frequency ω_l , we stack all elements of the scattering matrix into a single row-vector $\mathbf{x}_l \in \mathbb{C}^{P^2}$, constructed as $\mathbf{x}_l = \text{vec}(\mathbf{S}_l)^T$. We recall that the $\text{vec}()$ operator stacks all columns of its matrix element into a single column vector. More precisely, element $(\mathbf{S}_l)_{ij}$ with $1 \leq i, j \leq P$ corresponds to element $(\mathbf{x}_l)_k$ for $1 \leq k \leq P^2$ through

$$\begin{aligned} k &= i + (j - 1)P \\ i &= 1 + \text{mod}(k - 1, P) \\ j &= \lceil k/P \rceil \end{aligned} \quad (4)$$

where $\text{mod}(a, b)$ returns the remainder of the integer division a/b and $\lceil c \rceil$ is the ceil operator that returns the smallest integer not less than c . The mapping $(i, j) \leftrightarrow k$ in (4) will be used consistently throughout this paper.

We now collect all vectors \mathbf{x}_l corresponding to different frequencies ω_l as rows in a matrix $\mathbf{X} \in \mathbb{C}^{L \times P^2}$

$$\mathbf{X} = \begin{bmatrix} \leftarrow & \mathbf{x}_1 & \rightarrow \\ \vdots & \vdots & \vdots \\ \leftarrow & \mathbf{x}_L & \rightarrow \end{bmatrix} = \begin{bmatrix} \uparrow & \dots & \uparrow \\ \mathbf{z}_1 & \dots & \mathbf{z}_{P^2} \\ \downarrow & \dots & \downarrow \end{bmatrix} \quad (5)$$

Each row \mathbf{x}_l of this matrix corresponds to a single frequency ω_l , while each column \mathbf{z}_k collects all frequency samples of a single scattering response $(\mathbf{z}_k)_l = S_{ij}(j\omega_l)$.

We make the hypothesis that the P^2 scattering responses can be represented as an approximate sum of few basis functions. This implies that the column span of matrix \mathbf{X} can be safely approximated by projection onto a subspace \mathcal{W} having a dimension $\rho \ll P^2$. Several alternatives are available for

constructing this subspace. In this work, we adopt the Singular Value Decomposition (SVD), since it provides a full control over the approximation error.

A direct application of SVD to matrix \mathbf{X} leads to

$$\mathbf{X} = \tilde{\mathbf{U}} \tilde{\Sigma} \tilde{\mathbf{V}}^H = \tilde{\mathbf{W}} \tilde{\mathbf{V}}^H \quad (6)$$

where $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ are complex unitary matrices collecting the left and right singular vectors and $\tilde{\Sigma}$ collects the sorted real and positive singular values $\tilde{\sigma}_q$ on its main diagonal. Matrix $\tilde{\mathbf{W}} = \tilde{\mathbf{U}} \tilde{\Sigma}$ is orthogonal with each column $\tilde{\mathbf{w}}_q$ scaled by the corresponding singular value, $\|\tilde{\mathbf{w}}_q\| = \tilde{\sigma}_q$. The k -th column of \mathbf{X} is thus represented, using (6), as

$$\mathbf{z}_k = \sum_q \tilde{v}_{kq}^* \tilde{\mathbf{w}}_q. \quad (7)$$

This expression is exact, with no approximation error, if all singular values/vectors are considered in the expansion. Each sampled scattering response is thus represented as a superposition of “basis vectors” $\tilde{\mathbf{w}}_q$, whose norm decreases uniformly with increasing q .

The coefficients \tilde{v}_{kq}^* are complex-valued constants. Since we want to have real expansion coefficient in order to guarantee the causality and the realness of each element in the expansion (3), we slightly modify the SVD by splitting real and imaginary parts $\mathbf{X} = \mathbf{X}' + j\mathbf{X}''$ where $\mathbf{X}', \mathbf{X}'' \in \mathbb{R}^{L \times P^2}$, or equivalently

$$\mathbf{X} = [\mathbf{I}_L \quad j\mathbf{I}_L] \begin{bmatrix} \mathbf{X}' \\ \mathbf{X}'' \end{bmatrix} \quad (8)$$

where \mathbf{I}_L is the identity matrix of size L . Then, we perform a truncated SVD decomposition, where only the first ρ singular values are retained

$$\begin{bmatrix} \mathbf{X}' \\ \mathbf{X}'' \end{bmatrix} = \mathbf{U} \Sigma \mathbf{V}^T \simeq \bar{\mathbf{U}} \bar{\Sigma} \bar{\mathbf{V}}^T, \quad (9)$$

where $\bar{\mathbf{U}} \in \mathbb{R}^{2L \times \rho}$, $\bar{\Sigma} \in \mathbb{R}^{\rho \times \rho}$, $\bar{\mathbf{V}} \in \mathbb{R}^{P^2 \times \rho}$ with $\rho \ll r = \min\{2L, P^2\}$, and $\bar{\mathbf{V}}$ is orthonormal, $\bar{\mathbf{V}}^T \bar{\mathbf{V}} = \mathbf{I}$. Defining now

$$\bar{\mathbf{W}} = [\mathbf{I}_L \quad j\mathbf{I}_L] \bar{\mathbf{U}} \bar{\Sigma} \quad (10)$$

we obtain the low-rank approximation

$$\mathbf{X} \simeq \bar{\mathbf{X}} = \bar{\mathbf{W}} \bar{\mathbf{V}}^T \quad (11)$$

Equivalently,

$$\mathbf{z}_k \simeq \sum_{q=1}^{\rho} v_{kq} \bar{\mathbf{w}}_q, \quad (12)$$

which is similar to (7) but has guaranteed real coefficients v_{kq} . The q -th column $\bar{\mathbf{w}}_q \in \mathbb{C}^L$ of $\bar{\mathbf{W}}$, collects all frequency samples that define the q -th basis function.

We estimate now the error between the original matrix \mathbf{X} collecting all scattering data and its low-rank approximation $\bar{\mathbf{X}}$. Using the spectral norm, we have

$$\begin{aligned} \mathcal{E}_2 &= \|\bar{\mathbf{X}} - \mathbf{X}\| = \left\| [\mathbf{I}_L \quad j\mathbf{I}_L] \left[\bar{\mathbf{U}} \bar{\Sigma} \bar{\mathbf{V}}^T - \mathbf{U} \Sigma \mathbf{V}^T \right] \right\| \\ &\leq \left\| [\mathbf{I}_L \quad j\mathbf{I}_L] \right\| \left\| \bar{\mathbf{U}} \bar{\Sigma} \bar{\mathbf{V}}^T - \mathbf{U} \Sigma \mathbf{V}^T \right\| \\ &\leq \sqrt{2} \sigma_{\rho+1}, \end{aligned} \quad (13)$$

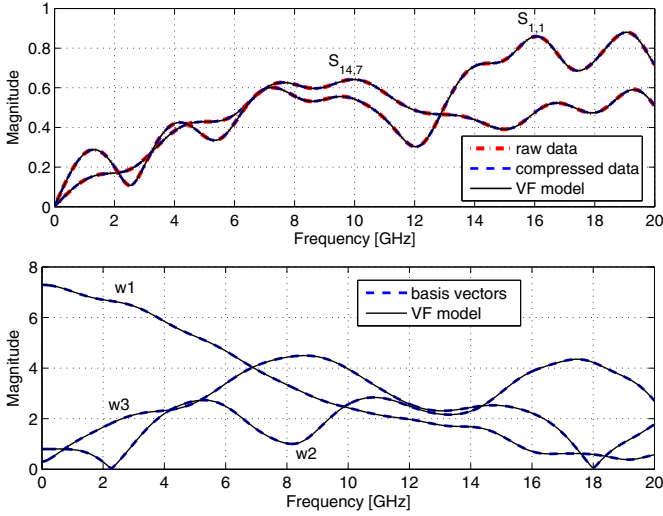


Fig. 2. Top: raw scattering responses of a high-speed connector before compression (red dashed line), its compressed ($\rho = 30$) approximation (blue dashed line), and its low-rank rational approximation computed by VF (black line). Bottom: first three vectors \bar{w}_q (blue dashed lines) in expansion (12) and corresponding VF approximation (black line).

where the last row follows from standard properties of the SVD decomposition. We see that the quality of the approximation is fully controlled by the first neglected singular value $\sigma_{\rho+1}$.

The top panel in Fig. 2 depicts two scattering responses of the same connector already considered in Fig. 1, together with the corresponding low-rank approximation. The difference is hardly visible. Bottom panel reports the first three basis vectors \bar{w}_q in the corresponding expansion (12).

IV. COMPRESSED MACROMODELING

Once expansion (12) is available, we compute a rational approximation of each basis vector \bar{w}_q . We introduce a row-vector of scalar functions of frequency

$$\mathbf{w}(s) = (w_1(s) \quad w_2(s) \quad \dots \quad w_\rho(s)), \quad (14)$$

with each element assumed in rational form

$$w_q(s) = w_{q,\infty} + \sum_{n=1}^{N_w} \frac{R_{q,n}}{s - p_n}. \quad (15)$$

The unknown poles p_n , residues $R_{q,n}$ and direct coupling constants $w_{q,\infty}$ are computed by applying a standard VF run. Since only ρ independent responses are concurrently fitted instead of P^2 , it is expected that the runtime of the VF process is drastically reduced. This is indeed the case, as we will show in Section V. We remark that we adopt common poles p_n for all basis functions in $\mathbf{w}(s)$, since these will be used to reconstruct the original scattering matrix through (12), thus obtaining a global rational macromodel in form (1).

A successful fitting process with stable poles is guaranteed by the realness of the expansion coefficients in (12). If we

post-multiply (11) by $\bar{\mathbf{V}}$, since $\bar{\mathbf{V}}^T \bar{\mathbf{V}} = \mathbf{I}$, we obtain

$$\bar{\mathbf{w}}_q \simeq \sum_{k=1}^{P^2} v_{kq} \mathbf{z}_k, \quad (16)$$

which shows that each basis vector can be represented as a linear combination of the raw scattering responses with real coefficients. This is sufficient to conclude that if the original responses are causal, each of the basis functions will be causal. Therefore, the rational approximation (15) is guaranteed to have stable poles p_n , see [4].

A state-space realization can be constructed from (15) using standard techniques. For later convenience, we construct this realization for the transpose system, which has a Single-Input Multiple-Output structure, as

$$\mathbf{w}(s)^T \leftrightarrow \left(\begin{array}{c|c} \mathbf{A}_w & \mathbf{B}_w \\ \hline \mathbf{C}_w & \mathbf{D}_w \end{array} \right) \quad (17)$$

with $\mathbf{A}_w \in \mathbb{R}^{N_w \times N_w}$, $\mathbf{B}_w \in \mathbb{R}^{N_w \times 1}$, $\mathbf{C}_w \in \mathbb{R}^{\rho \times N_w}$, $\mathbf{D}_w \in \mathbb{R}^{\rho \times 1}$. We then define a (reshaped) global rational macromodel according to the expansion (11), as

$$\begin{aligned} \mathbf{X}^T(s) &= \bar{\mathbf{V}} \mathbf{w}^T(s) = \\ &= \bar{\mathbf{V}} \mathbf{D}_w + \bar{\mathbf{V}} \mathbf{C}_w (s \mathbf{I}_{N_w} - \mathbf{A}_w)^{-1} \mathbf{B}_w, \end{aligned} \quad (18)$$

where $\mathbf{X}^T(s)$ is a column vector of P^2 rational responses. Finally, a global rational macromodel for the original scattering representation is obtained with a simple reshape operation

$$\mathbf{S}(s) = \text{mat}(\mathbf{X}^T(s)), \quad (19)$$

where the $\text{mat}(\cdot)$ operator reconstructs a $P \times P$ matrix starting from the corresponding $P^2 \times 1$ vector $\text{vec}(\mathbf{S})$. It is easy to show that a state-space realization of $\mathbf{S}(s)$ can be obtained as

$$\begin{aligned} \mathbf{A} &= \mathbf{I}_P \otimes \mathbf{A}_w, & \mathbf{B} &= \mathbf{I}_P \otimes \mathbf{B}_w, \\ \mathbf{C} &= \Psi (\mathbf{I}_P \otimes \mathbf{C}_w), & \mathbf{D} &= \Psi (\mathbf{I}_P \otimes \mathbf{D}_w), \end{aligned} \quad (20)$$

where \otimes denotes the Kronecker matrix product [5] and

$$\Psi = ((\bar{\mathbf{V}})_1 \quad (\bar{\mathbf{V}})_2 \quad \dots \quad (\bar{\mathbf{V}})_P) \quad (21)$$

with $(\bar{\mathbf{V}})_j \in \mathbb{R}^{P \times \rho}$ collecting the P rows $\{j(P-1) + 1, \dots, jP\}$ of matrix $\bar{\mathbf{V}}$. Once the state-space matrices (20) are available, standard methods [7], [8], [9] can be applied in order to check and enforce model passivity.

Figure 2 shows that the obtained rational approximation of both basis functions (bottom panel) and scattering responses (top panel) is very accurate.

Once the rational approximation (15) is available, we evaluate $\mathbf{w}(s)$ at each raw frequency point ω_l and we collect the results as rows in matrix $\widehat{\mathbf{W}} \in \mathbb{C}^{L \times \rho}$, which in turn is used to reconstruct the samples of the global rational macromodel, collected in matrix $\widehat{\mathbf{X}} = \widehat{\mathbf{W}} \bar{\mathbf{V}}^T$. Due to the orthonormality of the columns of $\bar{\mathbf{V}}$, we have

$$\|\widehat{\mathbf{X}} - \widehat{\mathbf{X}}\|_2 = \|\widehat{\mathbf{W}} \bar{\mathbf{V}}^T - \widehat{\mathbf{W}} \bar{\mathbf{V}}^T\|_2 = \|\widehat{\mathbf{W}} - \widehat{\mathbf{W}}\|_2. \quad (22)$$

This implies that the construction of a global rational model starting from the rational basis functions is well-behaved, since

TABLE I
SUMMARY OF THE COMPRESSED MACROMODELING RESULTS FOR SEVERAL BENCHMARKS. SEE TEXT FOR DETAILS

test	P	L	ρ	N_w	N	\mathcal{E}_2	δ_2	VF(X) [s]	SVD [s]	VF(W) [s]	SpeedUp
1	12	471	17	20	22	0.07	0.102	3.05	0.36	0.5	3.5×
2	24	1001	13	12	12	0.01	0.03	8.6	1.1	0.2	6.6×
3	34	570	40	57	58	0.072	0.096	263.1	1.9	5.1	36.9×
4	41	572	10	11	11	0.05	0.05	18.5	1.82	0.26	8.8×
5	48	690	24	27	28	0.06	0.102	119.6	3.7	1.09	24.9×
6	52	13	3	3	3	0.01	0.01	0.66	0.04	0.07	5.9×
7	56	1001	30	30	30	0.06	0.08	198.46	7.3	1.7	22.0×
8	92	71	22	22	23	0.06	0.106	32.8	1.4	0.1	21.8×
9	245	197	14	99	93	0.07	0.09	12885.2	24.5	1.29	499.2×
10	800	40	8	8	8	0.02	0.05	432.8	34.7	0.97	12.4×

it results in a fitting error that is identical to the fitting error achieved in the construction of the low-rank system $\mathbf{w}(s)$.

The global approximation error between raw scattering samples and global rational macromodel can thus be characterized as

$$\begin{aligned} \delta_2 &= \left\| \mathbf{X} - \widehat{\mathbf{X}} \right\|_2 \leq \left\| \mathbf{X} - \bar{\mathbf{X}} \right\|_2 + \left\| \bar{\mathbf{X}} - \widehat{\mathbf{X}} \right\|_2 \\ &\leq \sqrt{2}\sigma_{\rho+1} + \left\| \bar{\mathbf{W}} - \widehat{\mathbf{W}} \right\|_2, \end{aligned} \quad (23)$$

where the individual contributions of SVD truncation and VF approximation are explicit.

V. RESULTS AND DISCUSSION

We demonstrate the effectiveness of the compressed rational macromodeling scheme through a rich set of benchmarks, known via scattering samples available as Touchstone files. Several combinations of port count P and number of available frequency samples L have been tried, in order to test the scalability of the algorithm in a wide parameter range. All details are available in Table I.

The number of basis functions ρ has been automatically determined in order to reach a cumulative compression error $\mathcal{E}_2 < 0.1$. Similarly, the number of poles N_w of the low-rank macromodel has been determined in order to achieve a cumulative VF error $\left\| \bar{\mathbf{X}} - \widehat{\mathbf{X}} \right\|_2 < 0.1$. The performance of the compressed scheme is compared to the performance of the fast VF scheme [2] in its public-domain implementation [6] applied to the complete set of P^2 scattering responses, which required N poles to reach the same approximation error threshold. A total of 3 VF iterations were used for all tests. The table shows that the compression error \mathcal{E}_2 and the final error δ_2 are well under control for all cases.

The last columns in the table report the runtime in seconds for the various steps of the proposed algorithm. Column ‘‘SVD’’ reports the time required by the SVD compression part implemented according to [10]. Columns ‘‘VF(W)’’ and ‘‘VF(X)’’ report the runtime for the low-rank and full VF runs, respectively. Finally, column ‘‘SpeedUp’’ reports the overall speedup factor for the computation of the global macromodel in terms of total runtime. We see a dramatic reduction in runtime with our proposed scheme.

VI. CONCLUSIONS

We have presented a compressed macromodeling strategy for the extraction of rational macromodels of interconnect structures with a large number of ports. The method represents the full set of scattering responses as a linear combination of few and automatically selected basis functions, which are fitted in negligible time. The numerical results show that structures with a large port count can greatly benefit from the proposed compression scheme, which produces accurate macromodels with limited computing resources.

REFERENCES

- [1] B. Gustavsen, A. Semlyen, ‘‘Rational approximation of frequency domain responses by vector fitting’’, *IEEE Trans. Power Del.*, vol. 14, no. 3, pp. 1052-1061, Jul. 1999
- [2] D. Deschrijver, M. Mrozowski, T. Dhaene, D. De Zutter, ‘‘Macromodeling of Multiport Systems Using a Fast Implementation of the Vector Fitting Method’’, *IEEE Microwave and Wireless Comp. Lett.*, vol. 18, no. 6, pp. 383-385, June 2008
- [3] A. Chineza, S. Grivet-Talocia, ‘‘A parallel Vector Fitting implementation for fast macromodeling of highly complex interconnects’’, *Proc. of 19th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS 2010)*, Austin (TX), USA, pp. 101-104, October 24-27, 2010
- [4] P. Triverio, S. Grivet-Talocia, M.S. Nakhla, F. Canavero, R. Achar, ‘‘Stability, Causality, and Passivity in Electrical Interconnect Models’’, *IEEE Transactions on Advanced Packaging*, vol. 30, no. 4, pp. 795-808, November 2007
- [5] Charles F. Van Loan, ‘‘The ubiquitous Kronecker product’’, *Journal of Computational and Applied Mathematics*, vol. 123, pp. 85-100, 2000
- [6] D. Deschrijver, B. Gustavsen, T. Dhaene, ‘‘Computer code for fast macromodeling of large multiport systems’’, *Proc. of IEEE 17th Topical Meeting on Electrical Performance of Electronic Packaging (EPEP 2008)*, San Jose (CA), USA, pp. 199-202, October 27-29, 2008
- [7] S. Grivet-Talocia, ‘‘Passivity Enforcement via Perturbation of Hamiltonian Matrices’’, *IEEE Trans. Circuits and Systems I*, vol. 51, no. 9, pp. 1755-1769, September 2004
- [8] D. Saraswat, R. Achar and M. Nakhla, ‘‘Global Passivity Enforcement Algorithm for Macromodels of Interconnect Subnetworks Characterized by Tabulated Data’’, *IEEE Transactions on VLSI Systems*, vol. 13, no. 7, pp. 819-832, July 2005
- [9] B. Gustavsen, A. Semlyen, ‘‘Enforcing passivity for admittance matrices approximated by rational functions’’, *IEEE Trans. Power Systems*, vol. 16, no. 1, pp. 97-104, Feb. 2001
- [10] R. M. Larsen, ‘‘Computing the SVD for Large and Sparse Matrices’’, SCCM, Stanford University, June 16 2000, available online: <http://sun.stanford.edu/~rmunk/PROPACK/index.html>