View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by PORTO Publications Open Repository TOrino

Compressed Passive Macromodeling

S. B. Olivadese, S. Grivet-Talocia
Dipartimento di Elettronica e Telecomunicazioni, Politecnico di Torino
Corso Duca degli Abruzzi 24, 10129 Torino, Italy
Ph +39 011 0904104, Fax +39 011 0904099
e-mail stefano.grivet@polito.it

***Abstract*—This paper presents an approach for the extraction of passive macromodels of large-scale interconnects from their frequency-domain scattering responses. Here, large-scale is intended both in terms of number of electrical interface ports and required dynamic model order. For such structures, standard approaches based on rational approximation via Vector Fitting and passivity enforcement via model perturbation may fail due to excessive computational requirements, both in terms of memory occupation and runtime. Our approach addresses this complexity by first reducing the redundancy in the raw scattering responses through a projection and approximation process based on a truncated Singular Value Decomposition. Then, we formulate a compressed rational fitting and passivity enforcement framework, that is able to obtain speedup factors up to 2-3 orders of magnitude with respect to standard approaches, with full control over the approximation errors. Numerical results on a large set of benchmark cases demonstrate the effectiveness of the proposed technique.**

I. Introduction and motivation

Macromodeling techniques have become a standard practice in system design and verification flows. Such methods allow to convert external characterizations of linear and time-invariant structures such as passive devices and electrical interconnects into compact closed-form mathematical expressions or circuit equivalents. This conversion is needed to allow system-level transient simulations and verifications starting from a native characterization that is typically available in the frequency domain in form of tabulated scattering responses, the latter being determined from direct measurements or full-wave numerical solutions.

The above considerations led to major developments of macromodeling algorithms over the last few decades. We can safely state that the main result that fostered these developments is the introduction of the Vector Fitting (VF) algorithm [1]. Despite the lack of a theoretical result proving or disproving its convergence [2], the VF scheme formulates the problem of fitting a rational function to a set of frequency samples as an iterative solution of linear least squares and eigenvalue problems. Experience shows that convergence indeed occurs in very few iterations, with excellent accuracy and robustness. Since the first paper [1], many developments have been reported to enhance applicability, scalability, and performance. See, e.g., [3]-[11].

The basic VF scheme suffers two main problems. On one hand, the computational requirements may become excessive when the number of ports of the structure under modeling is large. Despite the smart formulation of [8], which substantially reduces memory consumption, and the subsequent parallel implementation in [10], [11], which allows major speedup on parallel computing platforms, there is still significant room for efficiency improvements.

The second problem of VF is its inability to guarantee the passivity of the resulting macromodels. Passivity is an essential property that guarantees stable and reliable system-level simulations [12], [13], [14]. For this reason, several techniques for a posteriori passivity enforcement have been proposed [15]-[27]. Such methods apply small perturbations to the model coefficients so that the modified model becomes passive. As for the rational fitting phase, also passivity enforcement schemes suffer from excessive computational requirements for large-scale models characterized by many ports and by a large dynamic order. Significant improvements were documented in [17], [26], including parallelization efforts [28]. However, the computational cost remains the main factor limiting applicability of passive macromodeling techniques to large-scale structures and devices.

In this paper, we present an approach for improving the efficiency of both rational fitting and passivity enforcement for medium and large-scale structures. We specifically address problems characterized by possibly hundreds of ports and requiring thousands of internal states for their models. Requirements for models of such complexity arise, for instance, in power bus modeling and optimization, chip-package co-design, and mixed-signal system design.

Our main approach is based on the fundamental idea that there is often a lot of redundancy in the frequency responses of coupled multiport structures. Following the approach preliminary documented in [29], we show in Sec. II that a simple projection based on a truncated Singular Value Decomposition (SVD) [30], [31] leads to drastic compression of scattering responses, which can be cast as a linear combination of few carefully selected "basis functions". The rational fitting of these basis functions leads to a compressed macromodel, which can be determined with reduced computational effort. The structure of this compressed model is exploited in Sec. III and IV to enforce asymptotic and global passivity at a reduced computational cost.

The effectiveness of the proposed approach is illustrated on a comprehensive set of benchmark cases. Numerical results and examples are reported at the end of each section in order to document each separate macromodeling step. A synoptic view of these results is presented and discussed in Sec. V.

Throughout this paper x, $\boldsymbol{x}$, and $\mathbf{X}$ denote a generic scalar,

vector, and matrix, respectively. Superscripts *, ^{T}, and ^{H} will stand for the complex conjugate, transpose, and conjugate (Hermitian) transpose, respectively. With $\mathbf{1}_L$ and $\mathbf{I}_L$, we denote respectively the column vector of ones and the identity matrix of size L (omitted when clear from the context). The set of eigenvalues of matrix $\mathbf{X}$ is denoted as $\lambda(\mathbf{X})$, whereas $\sigma(\mathbf{X})$ stands for the set of its singular values. The 2-norm $\|\cdot\|_2$ is defined as $\|\boldsymbol{x}\|_2^2 = \sum_\ell |x_\ell|^2$ for vectors (euclidean norm) and $\|\mathbf{X}\|_2 = \max \sigma(\mathbf{X})$ for matrices (spectral norm).

II. Compressed Rational Approximation

We consider a linear and time-invariant P-port interconnect system. We suppose that the scattering matrix $\mathbf{H}_\ell \in \mathbb{C}^{P\times P}$ at a suitable set of frequency points ω_ℓ with $\ell = 1,\dots,L$ is known. We want to derive a rational macromodel in form

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

where the poles p_n, the residue matrices $\mathbf{R}_n$, and the direct coupling matrix $\mathbf{R}_\infty$ are determined via some fitting or approximation process. A very effective and popular methodology to obtain macromodel (1) is to apply some formulation of the Vector Fitting (VF) algorithm [1]-[11], which computes all model parameters by an iterative solution of linear least squares and eigenvalue problems, providing a linearization of the global nonlinear optimization

$$\min_{\{p_n,\mathbf{R}_n,\mathbf{R}_\infty\}} \sum_{\ell=1}^{L}\sum_{i,j=1}^{P} \left|H_{ij}(\jmath\omega_\ell) - (\mathbf{H}_\ell)_{ij}\right|^2 . \tag{2}$$

The computational cost of VF in terms of CPU and memory occupation may grow excessively large for complex structures characterized by many ports and possibly many frequency samples over an extended frequency band, and requiring a possibly large number of poles in the rational approximation. Therefore, before resorting to the VF scheme, we try to eliminate any redundancy in the raw data, in order to reduce the size of the "independent" data points to be fed by the rational approximation engine. As pointed in [29], there may be a lot of redundancy in the scattering responses of typical electrical interconnects. Many responses look similar, and it is very likely that a high degree of compression can be achieved by smarter data representation. In the remainder of this section, we recall and complete the basic results of [29], in order to set the notation for later developments. Section II-A addresses data compression, while Sec. II-B exploits this compression to derive a reduced-complexity Vector Fitting scheme.

A. Data Compression

We start by collecting the P^2 elements of the scattering matrix $\mathbf{H}_\ell$ at single frequency ω_ℓ into a single row vector

$$\boldsymbol{x}_\ell = \mathrm{vec}(\mathbf{H}_\ell)^T\,, \tag{3}$$

where operator $\mathrm{vec}(\cdot)$ stacks all columns of its matrix element into a single column vector [32]. Equivalently, $(\boldsymbol{x}_\ell)_k = (\mathbf{H}_\ell)_{ij}$ with mapping $k \leftrightarrow (i,j)$ defined as

$$k = i + (j-1)P\,, \qquad \begin{cases} i = 1 + \mathrm{mod}(k-1,P)\,, \\ j = \lceil k/P \rceil\,, \end{cases} \tag{4}$$

where mod denotes the remainder of integer division and $\lceil\cdot\rceil$ rounds its argument to the nearest larger integer. Then, all the vectors $\boldsymbol{x}_\ell$ corresponding to different frequencies ω_ℓ are collected in matrix $\mathbf{X} \in \mathbb{C}^{L\times P^2}$, defined as

$$\mathbf{X} = \begin{bmatrix} \longleftarrow & \boldsymbol{x}_1 & \longrightarrow \\ \vdots & \vdots & \vdots \\ \longleftarrow & \boldsymbol{x}_L & \longrightarrow \end{bmatrix} = \begin{bmatrix} \uparrow & \cdots & \uparrow \\ \boldsymbol{z}_1 & \cdots & \boldsymbol{z}_{P^2} \\ \downarrow & \cdots & \downarrow \end{bmatrix} \tag{5}$$

Note that each column $\boldsymbol{z}_k$ of this matrix collects all frequency samples of a single scattering response $(\boldsymbol{z}_k)_\ell = S_{ij}(\jmath\omega_\ell)$.

Following [29], we compute the truncated SVD [30], [31]

$$\begin{bmatrix} \mathbb{R}\mathrm{e}\{\mathbf{X}\} \\ \mathbb{I}\mathrm{m}\{\mathbf{X}\} \end{bmatrix} \simeq \bar{\mathbf{U}}\bar{\mathbf{\Sigma}}\bar{\mathbf{V}}^T, \tag{6}$$

where $\bar{\mathbf{\Sigma}} \in \mathbb{R}^{\rho\times\rho}$ collects in its diagonal the first ρ singular values σ_q sorted in descending order, and where $\bar{\mathbf{U}} \in \mathbb{R}^{2L\times\rho}$, $\bar{\mathbf{V}} \in \mathbb{R}^{P^2\times\rho}$, with $\bar{\mathbf{U}}^T\bar{\mathbf{U}} = \mathbf{I}$ and $\bar{\mathbf{V}}^T\bar{\mathbf{V}} = \mathbf{I}$. We are interested in enforcing the condition

$$\rho \ll \min\{2L, P^2\}\,, \tag{7}$$

which ensures that (6) is a low-rank approximation with "tall and thin" matrices $\bar{\mathbf{U}}, \bar{\mathbf{V}}$. If (7) holds and the approximation error in (6) is small, then the assumption of redundancy in raw data is true. We will show that this is indeed the case through several numerical examples. Defining now

$$\bar{\mathbf{W}} = \begin{bmatrix} \mathbf{I}_L & \jmath\mathbf{I}_L \end{bmatrix} \bar{\mathbf{U}}\bar{\mathbf{\Sigma}}\,, \tag{8}$$

we can rewrite (6) as

$$\mathbf{X} \simeq \bar{\mathbf{X}} = \bar{\mathbf{W}}\bar{\mathbf{V}}^T\,. \tag{9}$$

Equivalently, if we extract the k-th column of $\mathbf{X}$, we obtain

$$\boldsymbol{z}_k \simeq \sum_{q=1}^{\rho} v_{kq}\bar{\boldsymbol{w}}_q\,, \tag{10}$$

where $\bar{\boldsymbol{w}}_q \in \mathbb{C}^L$ denotes the q-th column of $\bar{\mathbf{W}}$. We will repeatedly denote $\bar{\boldsymbol{w}}_q$ as "basis functions" in the following. This denomination is motivated by the fact that with a suitable choice of coefficients $v_{kq} \in \mathbb{R}$, any scattering response $\boldsymbol{z}_k$ can be approximated by a linear combination of such ρ basis functions. The coefficients v_{kq} are the elements of matrix $\bar{\mathbf{V}}$ collecting the first ρ right singular vectors of (6).

We now list two results that will be useful in the following.

Lemma 1: The euclidean norm of the q-th basis function $\bar{\boldsymbol{w}}_q$ is $\|\bar{\boldsymbol{w}}_q\|_2 = \sigma_q$.

Lemma 2: The error in the approximation (9) is bounded by

$$\mathcal{E}_2 = \left\|\bar{\mathbf{X}} - \mathbf{X}\right\|_2 \le \sqrt{2}\sigma_{\rho+1}\,, \tag{11}$$

where $\sigma_{\rho+1}$ is the largest neglected singular value.

The proof of these two lemmas is omitted, being a direct consequence of standard properties of the SVD decomposition [30], [31], see also [29]. These two lemmas are quite important for our application. In fact, Lemma 1 guarantees that the most significant contributions appear first in the linear superposition (10). Lemma 2 provides an explicit bound for the approximation error through the magnitude of the first neglected term.

B. Compressed Macromodeling

Instead of building a global rational macromodel by fitting directly the raw data as in (2), we will fit the basis functions $\bar{\boldsymbol{w}}_q$. To this end, we define

$$\boldsymbol{w}(s) = \begin{pmatrix} w_1(s) & w_2(s) & \dots & w_\rho(s) \end{pmatrix} , \tag{12}$$

where each component is a rational function

$$w_q(s) = r_{q\infty} + \sum_{n=1}^{N_w} \frac{r_{qn}}{s - p_n} . \tag{13}$$

The unknown poles p_n, residues r_{qn} and direct coupling constants $r_{q\infty}$ are computed by applying the VF scheme to solve

$$\min_{\{p_n, r_{qn}, r_{q\infty}\}} \sum_{\ell=1}^{L} \sum_{q=1}^{\rho} \left| w_q(\jmath\omega_\ell) - (\bar{\boldsymbol{w}}_q)_\ell \right|^2 . \tag{14}$$

Only ρ basis functions are concurrently fitted with (14) instead of the P^2 responses in (2). Therefore, the computational cost that will be required for the rational fitting stage is expected to be drastically reduced. Moreover, since we use a set of common poles p_n for all basis functions, due to (10) each scattering response will be modeled as a rational function with the same poles, thus matching the general form (1).

We now construct a state-space realization for the resulting compressed macromodel. First, we define a state-space realization for the basis function models, collected in a column vector as

$$\begin{aligned} \boldsymbol{w}(s)^T &= \mathbf{C}_w (s\mathbf{I} - \mathbf{A}_w)^{-1} \boldsymbol{b}_w + \boldsymbol{d}_w \\ &\leftrightarrow \left(\begin{array}{c|c} \mathbf{A}_w & \boldsymbol{b}_w \\ \hline \mathbf{C}_w & \boldsymbol{d}_w \end{array} \right) \end{aligned} \tag{15}$$

with $\mathbf{A}_w \in \mathbb{R}^{N_w \times N_w}$ storing the poles p_n in its main diagonal, $\boldsymbol{b}_w = \mathbf{1}_{N_w}$ column vector of ones, $\mathbf{C}_w \in \mathbb{R}^{\rho \times N_w}$ collecting all residues r_{qn} and $\boldsymbol{d}_w \in \mathbb{R}^\rho$ collecting the direct coupling constants $r_{q\infty}$. In case of complex conjugate pole/residue terms, the above state-space matrices are complex-valued, but a standard similarity transformation [38] can be applied to obtain a purely real realization in form (15).

A global rational macromodel can be obtained by defining

$$\mathbf{H}(s) = \mathrm{mat}(\bar{\mathbf{V}} \boldsymbol{w}^T(s)) , \tag{16}$$

where the $\mathrm{mat}(\cdot)$ operator reconstructs a $P \times P$ matrix of rational functions starting from its $P^2 \times 1$ vector argument. Following [29], we can show that a state-space realization of $\mathbf{H}(s)$ is obtained as

$$\mathbf{H}(s) \leftrightarrow \left(\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array} \right) \tag{17}$$

with

$$\begin{aligned} \mathbf{A} &= \mathbf{I}_P \otimes \mathbf{A}_w , & \mathbf{B} &= \mathbf{I}_P \otimes \boldsymbol{b}_w , \\ \mathbf{C} &= \boldsymbol{\Psi}(\mathbf{I}_P \otimes \mathbf{C}_w) , & \mathbf{D} &= \boldsymbol{\Psi}(\mathbf{I}_P \otimes \boldsymbol{d}_w) , \end{aligned} \tag{18}$$

where $\otimes$ denotes the Kronecker matrix product [32] and

$$\boldsymbol{\Psi} = \begin{pmatrix} \bar{\mathbf{V}}_1 & \bar{\mathbf{V}}_2 & \cdots \bar{\mathbf{V}}_P \end{pmatrix} \tag{19}$$

TABLE I
BENCHMARK STRUCTURES: L IS THE NUMBER OF RAW FREQUENCY SAMPLES, P THE NUMBER OF PORTS, ρ THE NUMBER OF BASIS FUNCTIONS (TO BE COMPARED WITH P^2); N_x AND N_w DENOTE THE NUMBER OF POLES USED FOR FULL AND COMPRESSED FITTING, RESPECTIVELY.

Case	L	P	P^2	ρ	N_w	N_x
1	471	12	144	17	20	22
2	690	48	2304	24	27	28
3	1001	56	3136	30	30	30
4	572	25	625	5	5	5
5	71	92	8464	22	22	23
6	570	34	1156	40	57	58
7	1001	24	576	13	12	12
8	1228	83	6889	31	30	31
9	100	8	64	6	29	29
10	197	245	60025	14	45	29
11	13	52	2704	3	3	3
12	40	800	640000	8	8	8
13	572	41	1681	10	11	11
14	141	542	293764	16	21	0
15	1000	34	1156	10	10	15
16	501	28	784	9	12	16
17	364	20	400	40	58	59
18	367	181	32761	6	24	39

with $\bar{\mathbf{V}}_j \in \mathbb{R}^{P \times \rho}$ collecting the P rows $\{j(P-1)+1, \dots, jP\}$ of matrix $\bar{\mathbf{V}}$

$$\bar{\mathbf{V}} = \begin{bmatrix} \bar{\mathbf{V}}_1 \\ \vdots \\ \bar{\mathbf{V}}_P \end{bmatrix} . \tag{20}$$

In (18), the size of the various matrices is $\mathbf{A} \in \mathbb{R}^{N \times N}$, $\mathbf{B} \in \mathbb{R}^{N \times P}$, $\mathbf{C} \in \mathbb{R}^{P \times N}$, $\mathbf{D} \in \mathbb{R}^{P \times P}$, where $N = N_w P$ denotes the global dynamic order of the realization. The transfer matrix of the compressed macromodel associated to (18) reads

$$\mathbf{H}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D} . \tag{21}$$

The final approximation error accounting for both compression and fitting can be characterized as follows. We denote with $\widehat{\mathbf{W}}$ and $\widehat{\mathbf{X}}$ the matrices collecting, respectively, the responses of compressed macromodel (15) and those of the reconstructed global macromodel (21) at the same frequencies ω_ℓ. We have

$$\left\| \mathbf{X} - \widehat{\mathbf{X}} \right\|_2 \leq \underbrace{\left\| \mathbf{X} - \bar{\mathbf{X}} \right\|_2}_{\leq \sqrt{2}\sigma_{\rho+1}} + \underbrace{\left\| \bar{\mathbf{X}} - \widehat{\mathbf{X}} \right\|_2}_{\|\delta\mathbf{X}\|_2} , \tag{22}$$

where the individual contributions of SVD truncation $\sqrt{2}\sigma_{\rho+1}$ and VF approximation $\|\delta\mathbf{X}\|_2$ are explicit. We remark that, due to the orthonormality of the columns in $\bar{\mathbf{V}}$, we have

$$\|\delta\mathbf{X}\|_2 = \left\| \bar{\mathbf{X}} - \widehat{\mathbf{X}} \right\|_2 = \left\| \bar{\mathbf{W}} - \widehat{\mathbf{W}} \right\|_2 , \tag{23}$$

so that the global fitting error can be controlled directly during the compressed fitting stage.

C. Examples

We present here all benchmark cases that will be analyzed throughout this work. Table I lists a total of 18 interconnect structures, characterized by different number of ports P and raw frequency samples L. These structures include high-speed

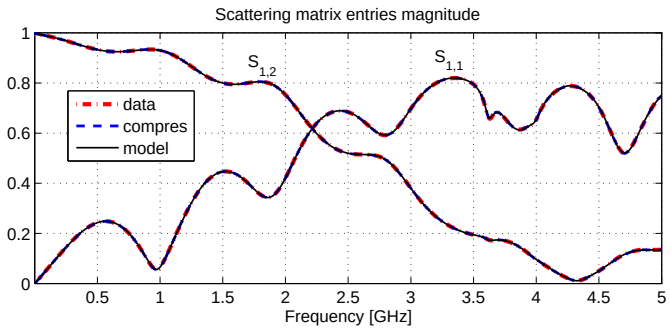

Fig. 1. Two sample scattering responses for case 6 before (dash-dotted lines) and after (dashed lines) compression, compared to the compressed rational fitted model responses (solid lines).

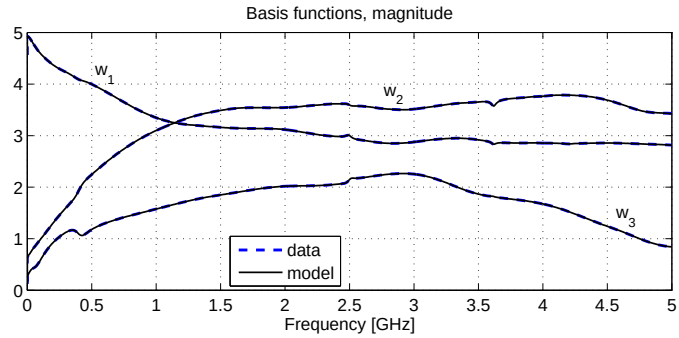

Fig. 2. First three basis functions for case 6. Original frequency samples $\tilde{\boldsymbol{w}}_q$ (dashed lines) and rational model $w_q(s)$ (solid lines).

connectors (cases 2, 3, 7), PCB interconnects (cases 9, 17), package interconnects (cases 5, 8, 13, 15, 16), power or mixed signal/power distribution networks (cases 1, 4, 6, 10, 11, 14, 18), and Through Silicon Via (TSV) fields (case 12). All raw frequency samples were obtained from 2D or 3D field characterizations. All numerical tests in this work were performed with a laptop (2 GHz clock and 4 GB memory).

The last column in Table I shows the number of poles N_x that were required by a standard application of Vector Fitting to fit the full set of responses $\mathbf{X}$ with a global model-vs-data deviation $\|\delta\mathbf{X}\|_2 < \epsilon_{\rm VF}$. Details on how to choose the threshold $\epsilon_{\rm VF}$ will be postponed to Sec. V. The publicly available VF code [9] based on the formulation [8] was used for these tests and applied by iteratively increasing the number of poles until the above accuracy condition was met.

In this section, we are interested in comparing the performance of standard and compressed VF. To this end, we use the threshold $\epsilon_{\rm SVD}$ to control the compression error $\mathcal{E}_2$, defined in (11), and $\epsilon_{\rm VF}$ to control the approximation error achieved by the compressed VF. This choice results in a number of basis functions ρ and in a number of poles for the basis functions N_w, also reported in Table I. These results show collectively that

- the number of basis functions always results $\rho \ll P^2$, therefore the computational complexity of the compressed VF run always results much less than the standard full VF;
- the number of poles required for the compressed and the full macromodels is comparable, $N_w \simeq N_x$, showing that the compression strategy does not create spurious or artificial components in the basis functions that would require an excessive number of poles for their fitting;
- the size of compressed macromodel $N_w P$ is comparable to the size of full macromodel $N_x P$ (assuming full-rank residue matrices, which was verified in all examples).

Figure 1 compares the compressed data and the compressed macromodel results to the raw scattering responses for benchmark case 6, showing that an excellent accuracy is obtained. Figure 2 shows some of the corresponding basis functions together with their rational fitted models.

Table II reports the execution time in seconds that was required by SVD algorithm [31] for compression, denoted as $T_{\rm SVD}$, for fitting the ρ basis functions and constructing the compressed macromodel, denoted as $T_{\rm VFW}$, and for applying standard VF to the full set of raw responses, denoted as $T_{\rm VFX}$. The overall speedup reported in the last column demonstrates how effective can the compressed macromodeling approach be for those cases that are characterized by a large port count or a large number of frequency samples. For case 14, standard VF could not even be applied due to an excessive memory occupation.

TABLE II
CPU TIME IN SECONDS REQUIRED FOR DATA COMPRESSION ($T_{\rm SVD}$) AND COMPRESSED FITTING ($T_{\rm VFW}$) COMPARED TO FULL FITTING ($T_{\rm VFX}$).

Case	$T_{\rm SVD}$ [s]	$T_{\rm VFW}$ [s]	$T_{\rm VFX}$ [s]	Speedup
1	0.03	0.66	4.2	6.03
2	0.8	1.7	183.5	70.5
3	1.3	3.7	419.7	82.4
4	0.28	0.02	1.42	4.6
5	0.7	0.23	59.4	63
6	0.33	10.6	355.2	32.1
7	0.37	0.28	11.6	17.8
8	3.2	4.6	1273	160
9	0.004	0.2	0.94	4.44
10	2.4	1.2	1609	437.1
11	0.01	0.006	0.2	12
12	12.8	0.04	592	45.8
13	1.7	0.3	17.8	8.8
14	9.2	0.8	-	-
15	4.8	1.5	39	6.1
16	0.3	0.154	12	24.2
17	0.15	8.05	77.3	9.4
18	2.2	0.4	2074	760.4

D. Passivity

There is no guarantee that the global macromodel (21) with state-space matrices (18) is passive. We can however explicitly enforce model (asymptotic) stability by constraining the poles p_n to have a strictly negative real part, a standard practice in VF applications [1]. Under this assumption, the macromodel is passive if and only if [12], [13], [14]

$$\min \lambda\{\mathbf{\Phi}(\jmath\omega)\} \geqslant 0\,, \quad \forall\omega\,, \tag{24}$$

where $\mathbf{\Phi}(\jmath\omega) = \mathbf{I}_P - \mathbf{H}^H(\jmath\omega)\mathbf{H}(\jmath\omega)$.

The passivity condition (24), which can be checked either via adaptive frequency sampling [26] or through identification of imaginary eigenvalues of the associated Hamiltonian

matrix [15], can be violated over finite or infinite frequency bands. In particular, this second case occurs if the model is not asymptotically passive, i.e., $\min \lambda\{\mathbf{\Phi}(\infty)\} < 0$. In this situation, asymptotic passivity can be recovered by perturbing just the direct coupling matrix $\mathbf{D}$. This will be the subject of Sec. III. Then, we will describe in Sec. IV a global passivity compensation scheme for enforcing (24) at all frequencies.

III. Asymptotic passivity enforcement

The macromodel (21) is asymptotically (strictly) passive if

$$\|\mathbf{D}\|_2 \leq \nu < 1\,, \tag{25}$$

where ν is some desired passivity threshold. In case (25) is not verified, we modify matrix $\mathbf{D}$ so that this condition is met. We want to operate directly on the compressed macromodel (15), so we add some perturbation vector $\boldsymbol{\eta}_w$ to the corresponding direct coupling vector $\boldsymbol{d}_w$, preserving the projection coefficients in matrix $\mathbf{\Psi}$. The perturbed matrix results

$$\mathbf{D}_p = \mathbf{\Psi}[\mathbf{I}_P \otimes (\boldsymbol{d}_w + \boldsymbol{\eta}_w)]\,, \tag{26}$$

with

$$\mathbf{D}_p - \mathbf{D} = \mathbf{\Psi}(\mathbf{I}_P \otimes \boldsymbol{\eta}_w)\,. \tag{27}$$

We want to achieve asymptotic passivity by a minimal perturbation of (27), which we measure in the standard 2-norm. This leads to the following formulation

$$\min_{\boldsymbol{\eta}_w} \|\mathbf{\Psi}(\mathbf{I}_P \otimes \boldsymbol{\eta}_w)\|_2 \quad \text{s.t.} \quad \|\mathbf{D}_p\|_2 \leq \nu\,. \tag{28}$$

The solution of (28) is addressed using various different approaches in Sections III-A–III-C, with results presented and compared in Sec. III-D.

Once a solution $\boldsymbol{\eta}_w$ of (28) is available, an asymptotically passive macromodel is constructed by

1) constructing the vector $\boldsymbol{d}_p = \boldsymbol{d}_w + \boldsymbol{\eta}_w$;
2) subtracting the q-th component $d_{p,q}$ of this vector from the frequency samples of the q-th basis function $\bar{\boldsymbol{w}}_q$ by redefining
$$\bar{\boldsymbol{w}}_q \leftarrow \bar{\boldsymbol{w}}_q - d_{p,q}\mathbf{1}_L \tag{29}$$
3) fitting the resulting frequency samples with a strictly proper rational function
$$w_q(s) = \sum_{n=1}^{N_w} \frac{r_{q,n}}{s - p_n}\,, \tag{30}$$
where the poles p_n are kept fixed to the poles of the original unperturbed macromodel (13);
4) defining the state-space realization of the compressed macromodel as in (15), but with $\boldsymbol{d}_w$ replaced by $\boldsymbol{d}_p$.

A. Direct scaling

The easiest way to enforce the asymptotic passivity is through the following rescaling

$$\boldsymbol{d}_p = \boldsymbol{d}_w \frac{\nu}{\|\mathbf{D}\|_2}\,, \quad \mathbf{D}_p = \mathbf{\Psi}(\mathbf{I}_P \otimes \boldsymbol{d}_p)\,. \tag{31}$$

This definition imposes asymptotic passivity by construction, but does not guarantee that the asymtpotic model perturbation $\|\mathbf{\Psi}(\mathbf{I}_P \otimes \boldsymbol{\eta}_w)\|_2$ is minimized, as required by (28). However, since the compressed macromodel will be re-generated via a new constrained vector fitting run (30), the asymptotic perturbation will have a significant effect only well beyond the last available frequency point, resulting in a quite acceptable accuracy within the modeling band. These statements will be validated through numerical examples in III-D. Therefore, this scaling method is actually quite competitive with the more precise approaches that follow due to its simplicity.

B. Linearization

The method described in this section is based on two simplifications of (28). First, the norm of $\boldsymbol{\eta}_w$ is minimized instead of the norm of $\mathbf{D}_p - \mathbf{D}$. Second, the constraint $\|\mathbf{D}_p\|_2 \leq \nu$ is replaced by an approximate constraint on $\boldsymbol{\eta}_w$ based on a linearization process. These two conditions lead to a problem of smaller size with respect to (28), which should require less computational effort for its solution.

We start with a SVD decomposition of $\mathbf{D} = \mathbf{L}\mathbf{\Sigma}_D\mathbf{R}^T$. Denoting the singular values as ς_i, $i = 1, \ldots, P$ with the associated left and right singular vectors $\boldsymbol{l}_i$ and $\boldsymbol{r}_i$, we have

$$\varsigma_i = \boldsymbol{l}_i^T \mathbf{D} \boldsymbol{r}_i\,. \tag{32}$$

Let us now apply the same projection to the perturbed direct coupling matrix $\mathbf{D}_p$. We obtain

$$\boldsymbol{l}_i^T \mathbf{D}_p \boldsymbol{r}_i = \varsigma_i + \boldsymbol{l}_i^T \mathbf{\Psi}(\mathbf{I}_P \otimes \boldsymbol{\eta}_w)\boldsymbol{r}_i\,. \tag{33}$$

Note that this quantity is not equal to the i-th singular value $\varsigma_{p,i}$ of $\mathbf{D}_p$, but it provides only a first-order approximation. Thus, condition

$$\boldsymbol{l}_i^T \mathbf{D}_p \boldsymbol{r}_i \leq \nu \tag{34}$$

corresponds to a linearized projection of constraint $\|\mathbf{D}_p\|_2 \leq \nu$. Using (33), after some straightforward algebraic manipulations, we obtain

$$(\boldsymbol{r}_i^T \otimes \boldsymbol{l}_i^T)\bar{\mathbf{V}}\boldsymbol{\eta}_w \leq \nu - \varsigma_i\,. \tag{35}$$

Collecting the various constraints (35) for all i leads to the linear underdetermined system

$$\mathbf{M}\boldsymbol{\eta}_w = \boldsymbol{\phi}\,, \tag{36}$$

where the number of rows in $\mathbf{M}$ defines the number of singular values of $\mathbf{D}$ being perturbed. Among all vectors $\boldsymbol{\eta}_w$ satisfying (36), we seek the minimum-norm solution, which is available in closed form as

$$\boldsymbol{\eta}_w = \mathbf{M}^\dagger \boldsymbol{\phi}\,, \tag{37}$$

with $\mathbf{M}^\dagger$ denoting the Moore-Penrose pseudoinverse of $\mathbf{M}$.

Due to the approximate nature of (35), the solution (37) of (36) does not guarantee that $\|\mathbf{D}_p\|_2 \leq \nu$. Therefore, we can iterate the process until this condition is achieved. At each iteration, two slightly different constraints can be used, leading to different numerical schemes

1) system (36) si formed by collecting all P singular values, setting at the right hand side
$$\phi_i = \begin{cases} \nu - \varsigma_i & \varsigma_i > \nu\,, \\ 0 & \varsigma_i \leq \nu\,. \end{cases} \tag{38}$$

This choice tries to explicitly preserve those singular values that are already below the threshold ν.

2) only constraints with $\varsigma_i > \nu$ are formed, so that only the singular value terms exceeding the threshold ν are explicitly perturbed.

C. Linear Matrix Inequalities

The problem stated in (28) can be cast as a Linear Matrix Inequality (LMI) [33], [34]. In fact, introducing the slack variable γ, minimization of the objective function in (28) can be restated as

$$\min \gamma \quad \text{s.t.} \begin{bmatrix} \gamma \mathbf{I}_P & \mathbf{\Psi}(\mathbf{I}_P \otimes \boldsymbol{\eta}_w) \\ (\mathbf{I}_P \otimes \boldsymbol{\eta}_w^T)\mathbf{\Psi}^T & \gamma \mathbf{I}_P \end{bmatrix} \succ 0 \,, \quad (39)$$

whereas the asymptotic passivity constraint is equivalent to

$$\begin{bmatrix} \nu \mathbf{I}_P & \mathbf{D} + \mathbf{\Psi}(\mathbf{I}_P \otimes \boldsymbol{\eta}_w) \\ \mathbf{D}^T + (\mathbf{I}_P \otimes \boldsymbol{\eta}_w^T)\mathbf{\Psi}^T & \nu \mathbf{I}_P \end{bmatrix} \succ 0 \,. \quad (40)$$

Expressions (39) and (40) form a system of LMI's. This formulation is based on convex constraints with a convex objective function. Therefore, its solution can be achieved numerically within arbitrary precision and with a finite number of steps using some specialized software. All results documented in the following were obtained with the SeDuMi package [35].

D. Numerical Results

Table III compares the asymptotic passivity enforcement results obtained by the various schemes presented in Sections III-A–III-C for those cases that resulted non-asymptotically passive after the compressed fitting stage. The maximum singular value $\|\mathbf{D}\|_2$ of the direct coupling matrix is reported for convenience in the second column. The four schemes are compared in terms of direct coupling perturbation amount $\mathbf{\Delta} = \mathbf{D}_p - \mathbf{D}$ measured in the spectral norm, number of iterations (when applicable), and total runtime. The latter includes not only the direct coupling perturbation, but also the computation of the perturbed residues and the construction of the global state-space realization, as described in Sec. III.

The direct scaling method requires no iterations. Only the computation of the norm $\|\mathbf{D}\|_2$ is required. Scaling requires negligible time, so that the total runtime is practically used for recomputing the updated residue matrices. The linarization and the LMI methods instead require several iterations and require significantly larger runtime. These three methods fail for the largest cases 12 and 14 due to excessive memory occupation (LMI) or lack of convergence (linearization) within a maximum number of 600 iterations. If converging, the linearization methods are faster than the LMI approach. However, the liearization methods are not guaranteed to attain the optimal solution, as does the LMI approach. This is confirmed by the amount of perturbation, which is smallest for the LMI case among all other methods. We see however that the simplistic direct scaling approach provides final perturbation errors that are comparable with the LMI scheme. Due to its efficiency, we indicate the direct scaling approach as most competitive. Of course, in case the resulting perturbation is excessive, one can resort to the LMI scheme, which is guaranteed to be optimal though slow.

IV. Global passivity enforcement

We now address the enforcement of global passivity for the macromodel (21) characterized by the state-space realization (18), assumed to be asymptotically stable and asymptotically passive. We will therefore assume that (24) is violated at some frequencies $\omega \in \Omega$, where Ω is the union of finite-width frequency bands.

In order to enforce passivity, we can follow one of the standard perturbation approaches. The main difference in the present framework with respect to published results is that the system perturbation should not be arbitrary but structured, according to the form of (18). We choose to perturb only the state-to-output map

$$\mathbf{C}_p = \mathbf{C} + \mathbf{\Delta}_C \,, \quad (41)$$

where the perturbation term $\mathbf{\Delta}_C$ is defined as

$$\mathbf{\Delta}_C = \mathbf{\Psi}(\mathbf{I}_P \otimes \mathbf{\Delta}_{C_w}) \,. \quad (42)$$

As for the asymptotic passivity enforcement of Sec. III, we preserve the expansion/projection coefficients in matrix $\mathbf{\Psi}$ and we perturb only the lower-dimensional compressed macromodel (15) using a local eigenvalue perturbation strategy [16].

A. Passivity enforcement

Let us consider a single frequency ω_0 at which condition (24) is violated by some negative eigenvalue $\lambda_i < 0$, and let the corresponding eigenvector of $\mathbf{\Phi}(\jmath\omega_0)$ be $\boldsymbol{\zeta}_i$, normalized such that $\|\boldsymbol{\zeta}_i\|_2 = 1$. Applying (41) leads to a first-order approximation of the perturbed eigenvalue [36]

$$\lambda_{p,i} \simeq \lambda_i + \boldsymbol{\zeta}_i^H \mathbf{\Delta}_{\mathbf{\Phi}} \boldsymbol{\zeta}_i \,, \quad (43)$$

where

$$\mathbf{\Delta}_{\mathbf{\Phi}} \simeq -\mathbf{K}_0^H \mathbf{\Delta}_C \mathbf{H}_0 - \mathbf{H}_0^H \mathbf{\Delta}_C^T \mathbf{K}_0 \quad (44)$$

and

$$\mathbf{H}_0 = \mathbf{D} + \mathbf{C}\mathbf{K}_0 \,, \quad \mathbf{K}_0 = (\jmath\omega_0 \mathbf{I} - \mathbf{A})^{-1}\mathbf{B} \,. \quad (45)$$

Standard manipulations lead to

$$\lambda_{p,i} \simeq \lambda_i + \boldsymbol{t}_i \,\mathrm{vec}(\mathbf{\Delta}_C) \,, \quad (46)$$

where the row-vector $\boldsymbol{t}_i$ is defined as

$$\boldsymbol{t}_i = -2\mathrm{Re}\{(\mathbf{K}_0\boldsymbol{\zeta}_i)^T \otimes (\mathbf{H}_0\boldsymbol{\zeta}_i)^H\} \,. \quad (47)$$

Enforcing now $\lambda_{p,i} \geqslant 0$ leads to the following linear inequality constraint

$$\boldsymbol{t}_i \,\mathrm{vec}(\mathbf{\Delta}_C) \geqslant -\lambda_i \,. \quad (48)$$

We also include the additional constraint

$$\boldsymbol{t}_i \,\mathrm{vec}(\mathbf{\Delta}_C) \leqslant 1 - \lambda_i \quad (49)$$

to guarantee that the perturbed eigenvalue remains bounded by one, as required by the assumed scattering representation. The above constraints are built for all $\mathcal{I}$ eigenvalues λ_i to be perturbed, possibly at multiple frequencies [16], and formulated as

$$\min \theta \quad \text{s.t.} \quad \begin{cases} \|\mathrm{vec}(\mathbf{\Delta}_C)\|_2^2 < \theta \\ \mathbf{T}\,\mathrm{vec}(\mathbf{\Delta}_C) \geqslant \boldsymbol{b} \end{cases} \quad (50)$$

TABLE III
COMPARISON OF ASYMPTOTIC PASSIVITY ENFORCEMENT METHODS IN TERMS OF PERTURBATION AMOUNT $\|\mathbf{\Delta}\|_2$, NUMBER OF ITERATIONS $\sharp$ it, AND RUNTIME IN SECONDS. SEE TEXT FOR ADDITIONAL DETAILS.

		Scaling		Linearization 1			Linearization 2			LMI		
Case	$\|\mathbf{D}\|_2$	$\|\mathbf{\Delta}\|_2$	Time [s]	$\|\mathbf{\Delta}\|_2$	$\sharp$ it	Time [s]	$\|\mathbf{\Delta}\|_2$	$\sharp$ it	Time [s]	$\|\mathbf{\Delta}\|_2$	$\sharp$ it	Time [s]
5	1.02	0.03	1	0.11	154	12	1.9	424	14	0.03	24	26.2
9	1.041	0.05	0.04	0.051	3	0.007	0.05	3	0.005	0.05	11	0.17
11	1.26	0.28	0.15	0.29	5	0.3	0.29	5	0.01	0.28	17	2.7
12	2.74	0.3	54.3	–	–	–	–	–	–	–	–	–
14	1.04	0.07	114	–	–	–	–	–	–	–	–	–
18	1	0.01	5	1.2	600	146	1.45	136	13.6	0.01	32	434

where θ is a slack variable. The last row collects in a compact form all constraints (48)-(49).

We now impose the perturbation structure (42). Using (19), it is easy to show that

$$\mathbf{\Delta}_C = \left(\bar{\mathbf{V}}_1 \mathbf{\Delta}_{C_w}, \ldots, \bar{\mathbf{V}}_P \mathbf{\Delta}_{C_w}\right) . \tag{51}$$

Applying the $\mathrm{vec}(\cdot)$ operator to the i-th column block in (51) leads to

$$\mathrm{vec}(\bar{\mathbf{V}}_i \mathbf{\Delta}_{C_w}) = (\mathbf{I}_{N_w} \otimes \bar{\mathbf{V}}_i)\, \mathrm{vec}(\mathbf{\Delta}_{C_w}) , \tag{52}$$

so that (51) can be written in "vectorized" form as

$$\mathrm{vec}(\mathbf{\Delta}_C) = \mathbf{\Theta}\, \mathrm{vec}(\mathbf{\Delta}_{C_w}) , \tag{53}$$

where $\mathbf{\Theta} \in \mathbb{R}^{PN \times \rho N_w}$ is defined as

$$\mathbf{\Theta} = \begin{bmatrix} \mathbf{I}_{N_w} \otimes \bar{\mathbf{V}}_1 \\ \vdots \\ \mathbf{I}_{N_w} \otimes \bar{\mathbf{V}}_P \end{bmatrix} \tag{54}$$

Finally, defining $\mathbf{T}_w \in \mathbb{R}^{2\mathcal{I} \times \rho N_w}$ as

$$\mathbf{T}_w = \mathbf{T}\mathbf{\Theta} , \tag{55}$$

we can formulate the structured and compressed passivity enforcement problem as

$$\min \theta \quad \text{s.t.} \quad \begin{cases} \|\mathrm{vec}(\mathbf{\Delta}_{C_w})\|_2^2 < \theta \\ \mathbf{T}_w\, \mathrm{vec}(\mathbf{\Delta}_{C_w}) \geqslant \boldsymbol{b} \end{cases} \tag{56}$$

Note that matrix $\mathbf{\Theta}$ is never constructed in practice, since all constraints in (56) and in particular matrix $\mathbf{T}_w$ can be built directly using optimized code.

If we compare the standard formulation (50) with the compressed and structured formulation (56), we see that the latter is much more convenient, since the number of decision variables is reduced by a factor

$$\frac{\#\{\mathbf{\Delta}_{C_w}\}}{\#\{\mathbf{\Delta}_C\}} = \frac{\rho N_w}{PN} = \frac{\rho}{P^2} \ll 1 . \tag{57}$$

This makes the cost for the solution of (56) practically negligible with respect to all other macromodeling steps. Note that the converse is typically the case, since passivity enforcement is usually the most demanding part of state of the art schemes. This big advantage is due to the particular state-space structure in (18).

B. Accuracy control

The formulations in (50) and (56) aim at finding the minimum norm of the perturbation terms $\mathbf{\Delta}_C$ or $\mathbf{\Delta}_{C_w}$ that are compatible with the passivity constraints. This condition however does not ensure that the energy (squared $\mathscr{L}^2$-norm) of the transfer matrix perturbation is minimized. To this end, we need to seek the minimum of

$$\|\mathbf{\Delta}_{\mathbf{H}}\|_{\mathscr{L}^2}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{tr}\{\mathbf{\Delta}_{\mathbf{H}}(\jmath\omega)\mathbf{\Delta}_{\mathbf{H}}^H(\jmath\omega)\} d\omega . \tag{58}$$

However, it is well known that this norm can be characterized as [37]

$$\|\mathbf{\Delta}_{\mathbf{H}}\|_{\mathscr{L}^2}^2 = \mathrm{tr}\{\mathbf{\Delta}_C \mathbf{P}_C \mathbf{\Delta}_C^T\} \tag{59}$$

where $\mathbf{P}_C$ is the controllability Gramian associated to (18), found as the unique, symmetric and positive definite solution of the Lyapunov equation

$$\mathbf{A}\mathbf{P}_C + \mathbf{P}_C \mathbf{A}^T = -\mathbf{B}\mathbf{B}^T . \tag{60}$$

If we compute the Cholesky factorization $\mathbf{P}_C = \mathbf{Q}_C^T \mathbf{Q}_C$ and define

$$\mathbf{\Xi} = \mathbf{\Delta}_C \mathbf{Q}_C^T , \quad \boldsymbol{\xi} = \mathrm{vec}(\mathbf{\Xi}) = (\mathbf{Q}_C \otimes \mathbf{I}_P)\, \mathrm{vec}(\mathbf{\Delta}_C) , \tag{61}$$

we have

$$\|\mathbf{\Delta}_{\mathbf{H}}\|_{\mathscr{L}^2}^2 = \mathrm{tr}\{\mathbf{\Xi}\mathbf{\Xi}^T\} = \|\boldsymbol{\xi}\|_2^2 . \tag{62}$$

Therefore, problem (50) can be cast as a minimum $\mathscr{L}^2$-norm formulation by performing the change of variable (61), obtaining

$$\min \theta \quad \text{s.t.} \quad \begin{cases} \|\boldsymbol{\xi}\|_2^2 < \theta \\ \mathbf{\Gamma}\boldsymbol{\xi} \geqslant \boldsymbol{b} \end{cases} \tag{63}$$

where $\mathbf{\Gamma} = \mathbf{T}(\mathbf{Q}_C^{-1} \otimes \mathbf{I}_P)$.

Let us now apply the same procedure to (56). We compute the controllability Gramian associated to the compressed state-space realization (15) as

$$\mathbf{A}_w \mathbf{P}_{C_w} + \mathbf{P}_{C_w} \mathbf{A}_w^T = -\boldsymbol{b}_w \boldsymbol{b}_w^T , \tag{64}$$

together with its Cholesky factorization $\mathbf{P}_{C_w} = \mathbf{Q}_{C_w}^T \mathbf{Q}_{C_w}$. Note that the numerical solution of (15) requires only $O(N_w)$ operations due to the sparse (diagonal or tridiagonal) realization of $\boldsymbol{w}(s)^T$. This cost is negligible with respect to all other macromodeling steps in the proposed framework. Defining

$$\begin{aligned} \mathbf{\Xi}_w &= \mathbf{\Delta}_{C_w} \mathbf{Q}_{C_w}^T , \\ \boldsymbol{\xi}_w &= \mathrm{vec}(\mathbf{\Xi}_w) = (\mathbf{Q}_{C_w} \otimes \mathbf{I}_\rho)\, \mathrm{vec}(\mathbf{\Delta}_{C_w}) , \end{aligned} \tag{65}$$

and denoting as $\boldsymbol{\Delta}_{\boldsymbol{w}^T}(s)$ the induced perturbation on the compressed macromodel, we have

$$\|\boldsymbol{\Delta}_{\boldsymbol{w}^T}\|^2_{\mathscr{L}^2} = \|\boldsymbol{\xi}_w\|^2_2 \,, \tag{66}$$

so that substitution into (56) leads to

$$\min \theta \quad \text{s.t.} \quad \left\{ \begin{array}{l} \|\boldsymbol{\xi}_w\|^2_2 < \theta \\ \boldsymbol{\Gamma}_w \boldsymbol{\xi}_w \geqslant \boldsymbol{b} \end{array} \right. \tag{67}$$

where $\boldsymbol{\Gamma}_w = \mathbf{T}_w(\mathbf{Q}^{-1}_{C_w} \otimes \mathbf{I}_\rho)$. The solution of (67) thus provides the minimum $\mathscr{L}^2$-norm perturbation of the compressed macromodel $\boldsymbol{w}^T(s)$.

We have the following result

Lemma 3: Defining $\mathbf{P}_C$ and $\mathbf{P}_{C_w}$ as in (60) and (64), we have

$$\mathbf{P}_C = \mathbf{I}_P \otimes \mathbf{P}_{C_w} \,. \tag{68}$$

Proof: Suppose that $\mathbf{P}_{C_w}$ is the solution of (64). We see that $\mathbf{P}_C$ defined in (68) is a solution of (60) by direct substitution. Using (18),

$$\begin{aligned} &\mathbf{A}\mathbf{P}_C + \mathbf{P}_C\mathbf{A}^T \\ &= (\mathbf{I}_P \otimes \mathbf{A}_w)(\mathbf{I}_P \otimes \mathbf{P}_{C_w}) + (\mathbf{I}_P \otimes \mathbf{P}_{C_w})(\mathbf{I}_P \otimes \mathbf{A}_w^T) \\ &= \mathbf{I}_P \otimes (\mathbf{A}_w\mathbf{P}_{C_w} + \mathbf{P}_{C_w}\mathbf{A}_w^T) \\ &= \mathbf{I}_P \otimes (-\boldsymbol{b}_w\boldsymbol{b}_w^T) \\ &= -(\mathbf{I}_P \otimes \boldsymbol{b}_w)(\mathbf{I}_P \otimes \boldsymbol{b}_w^T) \\ &= -\mathbf{B}\mathbf{B}^T \,. \end{aligned}$$

Since both $\mathbf{A}$ and $\mathbf{A}_w$ are strictly negative definite, $\mathbf{P}_C$ and $\mathbf{P}_{C_w}$ are the unique solutions of Lyapunov equations (60) and (64), which implies (68). ■

We are now ready to state the main result of this section.

Theorem 1: Defining the compressed macromodel perturbation

$$\boldsymbol{\Delta}_{\boldsymbol{w}^T} \leftrightarrow \left(\begin{array}{c|c} \mathbf{A}_w & \boldsymbol{b}_w \\ \hline \boldsymbol{\Delta}_{C_w} & \mathbf{0} \end{array} \right) \tag{69}$$

and the corresponding global macromodel perturbation

$$\boldsymbol{\Delta}_{\mathbf{H}} \leftrightarrow \left(\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \boldsymbol{\Delta}_{C} & \mathbf{0} \end{array} \right) \,, \tag{70}$$

with state-space matrices constructed as in (18), we have

$$\|\boldsymbol{\Delta}_{\mathbf{H}}\|^2_{\mathscr{L}^2} = \|\boldsymbol{\Delta}_{\boldsymbol{w}^T}\|^2_{\mathscr{L}^2} \tag{71}$$

Proof: As a preliminary result, consider matrix $\bar{\mathbf{V}}$ in (6). Using (20), the orthogonality condition $\bar{\mathbf{V}}^T\bar{\mathbf{V}} = \mathbf{I}$ can be rewritten in terms of its constituent blocks $\bar{\mathbf{V}}_i$ as

$$\sum_{i=1}^{P}\sum_{m=1}^{P}(\bar{\mathbf{V}}_i)_{m\ell}(\bar{\mathbf{V}}_i)_{mn} = \delta_{n\ell} \,, \quad n,\ell = 1,\ldots,\rho \,, \tag{72}$$

where $\delta_{n\ell} = 1$ if $n = \ell$ and 0 otherwise. Considering now (51) and using (68), a straightforward algebraic manipulation leads to

$$\boldsymbol{\Delta}_C\mathbf{P}_C\boldsymbol{\Delta}_C^T = \sum_{i=1}^{P}\bar{\mathbf{V}}_i\boldsymbol{\Upsilon}_w\bar{\mathbf{V}}_i^T \,, \tag{73}$$

TABLE IV
COMPARISON OF FULL AND COMPRESSED PASSIVITY ENFORCEMENT SCHEMES IN TERMS OF NUMBER OF ITERATIONS ♯ it, RUNTIME, AND ACCURACY $\|\delta\mathbf{X}\|_2$. LAST TWO COLUMNS REPORT THE OVERALL SPEEDUP (SU) AND THE SPEEDUP PER ITERATION (SUit).

Case	Full / Compressed			SU	SUit
	♯ it	Time [s]	$\|\delta\mathbf{X}\|_2$		
1	6 / 7	2.42 / 1.52	0.22 / 0.26	1.6	1.8
2	2 / 1	9.63 / 1.85	0.22 / 0.11	5.2	2.6
3	12 / 7	255 / 3.87	2.61 / 2.61	66.1	38.5
4	2 / 1	3.7 / 0.36	0.04 / 0.04	10.2	5.1
5	12 / 9	687.5 / 22.6	0.16 / 0.21	30.4	22.8
6	50 / 30	324.3 / 12.1	0.53 / 0.41	26.8	16.1
7	2 / 2	1.45 / 0.36	0.05 / 0.06	4.1	4.1
8	28 / 10	510 / 15.9	1.43 / 1.26	32.1	11.4
9	2 / 26	5.83 / 1.64	4.15 / 4.21	3.5	3.1
10	9 / 8	3865 / 145	3.31 / 3.32	26.6	23.6
11	2 / 4	9.34 / 1.81	0.04 / 0.05	5.2	10.4
12	- / 32	-.- / 17344	-.- / 1.21	-.-	-.-
13	8 / 7	24.7 / 4.86	0.16 / 0.21	5.1	4.2
14	- / 13	-.- / 5049	-.- / 1.21	-.-	-.-
15	1 / 2	5.85 / 3.17	0.08 / 0.08	1.8	3.6
16	10 / 8	13.1 / 1.95	0.21 / 0.25	6.8	5.4
17	10 / 6	13.7 / 1.26	0.51 / 0.51	11.4	6.8
18	- / 5	-.- / 1621	-.- / 6.79	-.-	-.-

where $\boldsymbol{\Upsilon}_w = \boldsymbol{\Delta}_{C_w}\mathbf{P}_{C_w}\boldsymbol{\Delta}_{C_w}^T$. The $\mathscr{L}^2$ norm of the global macromodel perturbation is characterized as

$$\begin{aligned} \|\boldsymbol{\Delta}_{\mathbf{H}}\|^2_{\mathscr{L}^2} &= \mathrm{tr}\{\boldsymbol{\Delta}_C\mathbf{P}_C\boldsymbol{\Delta}_C^T\} \\ &= \sum_{m=1}^{P}\left(\sum_{i=1}^{P}\bar{\mathbf{V}}_i\boldsymbol{\Upsilon}_w\bar{\mathbf{V}}_i^T\right)_{mm} \\ &= \sum_{m=1}^{P}\sum_{i=1}^{P}\sum_{n=1}^{\rho}\sum_{\ell=1}^{\rho}(\bar{\mathbf{V}}_i)_{m\ell}(\boldsymbol{\Upsilon}_w)_{\ell n}(\bar{\mathbf{V}}_i)_{mn} \\ &= \sum_{n=1}^{\rho}\sum_{\ell=1}^{\rho}(\boldsymbol{\Upsilon}_w)_{\ell n}\sum_{m=1}^{P}\sum_{i=1}^{P}(\bar{\mathbf{V}}_i)_{m\ell}(\bar{\mathbf{V}}_i)_{mn} \\ &= \sum_{\ell=1}^{\rho}(\boldsymbol{\Upsilon}_w)_{\ell\ell} \\ &= \|\boldsymbol{\Delta}_{\boldsymbol{w}^T}\|^2_{\mathscr{L}^2} \,, \end{aligned}$$

which completes the proof. ■

The practical relevance of this theorem is that the solution of the small-size optimization problem (67), in addition to providing the minimum-energy perturbation of the compressed macromodel, will also provide as a byproduct the minimum-energy solution of the full-size passivity enforcement problem, which is our main objective in this Section. Global passivity enforcement is thus achieved with optimal accuracy and negligible cost through (67).

C. Examples

In this section, we compare the performance of the passivity enforcement schemes (63) and (67) for each of the benchmark cases of Table I. The results are summarized in Table IV, where the total execution time and number of iterations for both schemes are grouped in columns 2 and 3 for convenience. We see that the number of iterations for the compressed scheme is practically always less than for the full scheme.

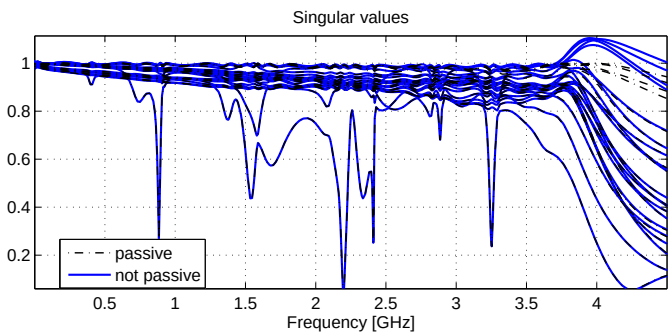

Fig. 3. Singular value plot before and after passivity enforcement for case 17.

This implies that, independent on the runtime required for a single iteration, the compressed scheme performs generally better. This consideration should be taken into account when interpreting the total runtime, reported in the second column. We observe that a dramatic reduction is achieved by the compressed scheme, which is able to complete the passivity enforcement also for those large cases (12, 14, and 18) for which the full scheme requires excessive memory.

We report in the fourth column of Table IV two different speedup factors. The first is the overall speedup factor, obtained as the ratio of the total runtime required by the full and compressed schemes. The second is the average runtime per iteration, which provides a more precise metric for assessing the enhancement in efficiency that can be achieved with proposed approach. In any case, both speedup per iteration and overall speedup are between 1 and 2 orders of magnitude for the most challenging cases, except for the largest cases for which only the compressed scheme could achieve its goal.

Finally, the fourth column of Table IV reports the deviation of the obtained passive models with respect to the original raw data, showing that the accuracies of both full and compressed schemes are comparable. Figure 3 reports as an example the singular value plot for case 17, showing all singular values before and after compressed passivity enforcement. As expected, the singular values of the passive model are uniformly unitary bounded.

V. A SUMMARY OF NUMERICAL RESULTS

We now summarize the main results for all benchmark cases. Table V provides a detailed report on the accuracy of all intermediate steps of the proposed compressed passive macromodeling approach. The second column reports the thresholds $\epsilon_{\rm SVD}$ and $\epsilon_{\rm VF}$ that were used, respectively, to bound the approximation error for SVD truncation and compressed VF. Note that these thresholds are used to bound the spectral norm of error matrices $\|\delta\mathbf{X}\|_2$ collecting all responses at all frequencies. Since the relationship of these thresholds to the actual deviation that is achieved at a given frequency for a given response is not obvious, we also report the results in terms of the worst-case norm, defined as

$$\|\delta\mathbf{X}\|_{\max} = \max_{\ell k} |(\delta\mathbf{X})_{\ell k}| \ . \tag{74}$$

The last three columns of Table V report the spectral and worst-case accuracies (with respect to raw data) of compressed data $\delta\mathbf{X}_{\rm SVD}$, compressed fitted model $\delta\mathbf{X}_{\rm VF}$, and final model after compressed passivity enforcement $\delta\mathbf{X}_{\rm PAS}$, respectively. The table clearly shows that accuracy is well preserved through all modeling steps. For illustration, we also report in Figures 4 and 5, respectively, the responses characterized by the worst-case absolute error for case 17, and the responses characterized by the worst-case relative error for case 2. Similar results were obtained for all other cases and are not reported here.

TABLE V
ACCURACY WITH RESPECT TO RAW DATA OF COMPRESSED DATA ($\delta\mathbf{X}_{\rm SVD}$) AND COMPRESSED MACROMODEL BEFORE ($\delta\mathbf{X}_{\rm VF}$) AND AFTER ($\delta\mathbf{X}_{\rm PAS}$) PASSIVITY ENFORCEMENT.

Case	ϵ SVD / VF	$\delta\mathbf{X}_{\rm SVD}$ $\|\cdot\|_2$ / $\|\cdot\|_{\max}$	$\delta\mathbf{X}_{\rm VF}$ $\|\cdot\|_2$ / $\|\cdot\|_{\max}$	$\delta\mathbf{X}_{\rm PAS}$ $\|\cdot\|_2$ / $\|\cdot\|_{\max}$
1	0.1 / 0.1	0.07 / 0.0039	0.09 / 0.006	0.26 / 0.014
2	0.1 / 0.1	0.06 / 0.0045	0.09 / 0.007	0.11 / 0.007
3	0.1 / 0.1	0.06 / 0.0029	0.08 / 0.003	2.61 / 0.064
4	0.1 / 0.1	0.04 / 0.0015	0.04 / 0.002	0.04 / 0.002
5	0.1 / 0.1	0.06 / 0.0105	0.09 / 0.051	0.23 / 0.057
6	0.1 / 0.1	0.07 / 0.0041	0.09 / 0.006	0.42 / 0.015
7	0.1 / 0.1	0.01 / 0.0005	0.04 / 0.001	0.06 / 0.002
8	0.1 / 0.5	0.08 / 0.0027	0.48 / 0.016	1.05 / 0.014
9	0.1 / 0.1	0.05 / 0.0084	0.05 / 0.008	4.12 / 0.632
10	0.1 / 3.0	0.07 / 0.0061	2.21 / 0.048	2.53 / 0.048
11	0.1 / 0.1	0.01 / 0.0012	0.01 / 0.001	0.18 / 0.016
12	0.1 / 0.1	0.02 / 0.0002	0.05 / 0.001	1.22 / 0.011
13	0.1 / 0.1	0.04 / 0.0046	0.05 / 0.011	0.21 / 0.011
14	0.1 / 0.1	0.07 / 0.0213	0.08 / 0.021	1.26 / 0.031
15	0.1 / 0.1	0.06 / 0.0018	0.08 / 0.002	0.08 / 0.002
16	0.1 / 0.1	0.04 / 0.0147	0.08 / 0.015	0.25 / 0.015
17	0.1 / 0.4	0.07 / 0.0241	0.39 / 0.315	0.43 / 0.315
18	0.1 / 6.8	0.07 / 0.0055	6.79 / 0.212	6.91 / 0.218

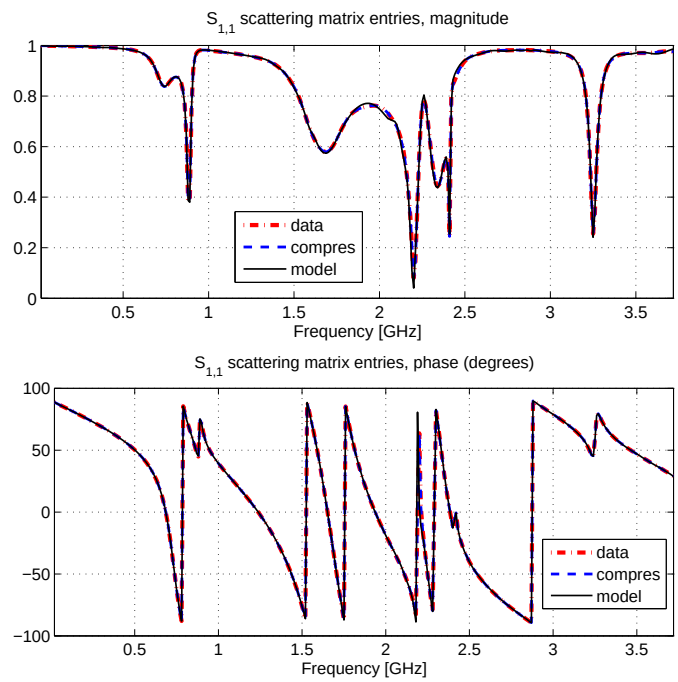

Fig. 4. A scattering response of a PCB interconnect (case 17) before (red dashed line) and after (blue dashed line) compression. The black line represents the response of the passive compressed macromodel.

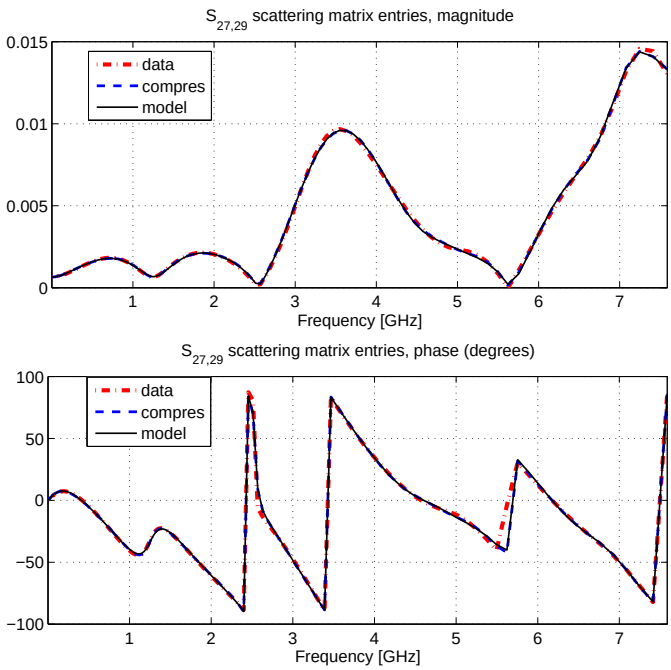

Fig. 5. As in Fig. 4, but for a high-speed connector (case 2).

VI. Conclusions

In this work, we have presented a comprehensive framework for compressed passive macromodeling of large-scale interconnect structures. The main enabling factor for this new approach is the observation that the whole set of P^2 scattering responses of P-port large-scale systems can be expressed through a much lower-dimensional set of $\rho \ll P^2$ basis functions. A singular value truncation is able to determine both the number of such basis functions and the corresponding expansion coefficients, with full control over the approximation error. Although this strategy was applied in this work to scattering representations, we expect that the same singular value truncation and approximation process can be applied to systems in impedance or admittance form without additional difficulties.

The above compressed data representation was used in the paper to derive reduced-complexity Vector Fitting and passivity enforcement schemes. The former generates a rational macromodel for the set of basis functions. The latter enforces global passivity constraints using a restricted set of perturbation variables. The overall result is a passive macromodeling scheme that has the potential to outperform state-of-the-art methods in terms of scalability, memory occupation, and CPU requirements, as illustrated through several challenging benchmark cases. Therefore, the results of this paper indicate that the proposed technique may become an enabling technology for massive macromodeling application to design and verification of signal and power distribution networks, 3D interconnects, and chip-package-board co-design.

One main difficulty remains, namely the size of the obtained macromodel. Even if the proposed scheme is able to compute this macromodel much faster, the number of states of the compressed macromodel is practically identical to the number of states of the macromodel that would be obtained by applying the standard Vector Fitting to the full set of raw responses. In case of very large number of ports P, this size may be problematic for further system-level simulations. We believe that, unless some further hypotheses or constraints are enforced, e.g., on the terminations to be employed in these simulations, it will be very difficult to further reduce the macromodel size without neglecting important dynamic contributions and affecting accuracy. The subject of optimal model order reduction for large-scale interconnects [40], [39], which is not addressed in this paper, is and will remain a very important research direction to try to overcome this difficulty [41].

References

[1] B. Gustavsen and A. Semlyen, "Rational approximation of frequency domain responses by vector fitting," *IEEE Trans. Power Delivery*, vol. 14, no. 3, pp. 1052–1061, Jul. 1999.

[2] P. Stoica, T. Soderstrom,"The Steiglitz-McBride identification algorithm revisited–Convergence analysis and accuracy aspects," *IEEE Trans. Automatic Control*, vol. 26, no. 3, pp. 712- 717, Jun 1981.

[3] B. Gustavsen, "Improving the pole relocating properties of vector fitting," *IEEE Trans. Power Delivery*, Vol. 21, N. 3, Aug. 2006, pp. 1587–1592.

[4] B. Gustavsen, "Computer code for rational approximation of frequency dependent admittance matrices," *IEEE Trans. Power Delivery*, Vol. 17, N. 4, October 2002, pp. 1093–1098.

[5] D. Deschrijver, B. Haegeman, T. Dhaene, "Orthonormal Vector Fitting: A Robust Macromodeling Tool for Rational Approximation of Frequency Domain Responses," *IEEE Trans. Adv. Packaging*, vol. 30, pp. 216–225, May 2007.

[6] S. Grivet-Talocia, M. Bandinu, "Improving the Convergence of Vector Fitting in Presence of Noise", *IEEE Transactions on Electromagnetic Compatibility*, vol. 48, n. 1, pp. 104-120, February, 2006.

[7] Y. S. Mekonnen, J. Schutt-Ainé, "Broadband macromodeling of sampled frequency data using z-domain vector-fitting method," em IEEE Workshop on Signal Propagation on Interconnects, 13-16 May 2007, pp. 45–48.

[8] 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 Components Letters*, Vol. 18, N. 6, June 2008, pp.383–385.

[9] B. Gustavsen, "The Vector Fitting Web Site," available online: http://www.energy.sintef.no/Produkt/VECTFIT

[10] A. Chinea and S. Grivet-Talocia, "A parallel vector fitting implementation for fast macromodeling of highly complex interconnects," in *IEEE* 19th *Topical Meeting on Electrical Performance of Electronic Packaging and Systems (EPEPS 2010), Austin, TX*, October 24–27, 2010.

[11] A. Chinea and S. Grivet-Talocia, "On the parallelization of vector fitting algorithms," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 1, pp. 1761–1773, November 2011.

[12] M. R. Wohlers, *Lumped and Distributed Passive Networks*, Academic Press, 1969.

[13] B. D. O. Anderson and S. Vongpanitlerd, *Network Analysis and Synthesis*, Prentice-Hall, Englewood Cliffs, NJ, 1973.

[14] 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, Nov. 2007.

[15] S. Grivet-Talocia, "Passivity enforcement via perturbation of Hamiltonian matrices", *IEEE Trans. CAS-I*, pp. 1755-1769, vol. 51, n. 9, September, 2004

[16] S. Grivet-Talocia, A. Ubolli, "A Comparative Study of Passivity Enforcement Schemes for Linear Lumped Macromodels, *IEEE Trans. Adv. Packaging*, vol. 31, No. 4, pp. 673–683, Nov. 2008.

[17] S. Grivet-Talocia, A. Ubolli "On the Generation of Large Passive Macromodels for Complex Interconnect Structures", *IEEE Trans. Adv. Packaging*, vol. 29, No. 1, pp. 39–54, Feb. 2006

[18] T. Dhaene, T. Deschrijver, N. Stevens, "Efficient Algorithm for Passivity Enforcement of S-Parameter-Based Macromodels," *IEEE Trans. on Microwave Theory and Techniques*, vol. 57, no. 2, pp. 415–420, Feb. 2009.

[19] D. Deschrijver, T. Dhaene, “Fast Passivity Enforcement of S-Parameter Macromodels by Pole Perturbation,” *IEEE Trans. on Microwave Theory and Techniques*, vol. 57, no. 3, pp. 620–626, March 2009.

[20] 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.

[21] C. P. Coelho, J. Phillips, L. M. Silveira, “A Convex Programming Approach for Generating Guaranteed Passive Approximations to Tabulated Frequency-Data”, *IEEE Trans. Computed-Aided Design of Integrated Circuits and Systems*, Vol. 23, No. 2, February 2004, pp. 293–301.

[22] H. Chen, J. Fang, “Enforcing Bounded Realness of S parameter through trace parameterization”, in *12th IEEE Topical Meeting on Electrical Performance of Electronic Packaging*, October 27–29, 2003, Princeton, NJ, pp. 291–294.

[23] B. Gustavsen, A. Semlyen, “Enforcing passivity for admittance matrices approximated by rational functions”, *IEEE Trans. Power Systems*, Vol. 16, N. 1, Feb. 2001, pp. 97–104.

[24] B. Gustavsen, “Computer Code for Passivity Enforcement of Rational Macromodels by Residue Perturbation,” *IEEE Trans. Adv. Packaging*, vol. 30, pp. 209–215, May 2007.

[25] D. Saraswat, R. Achar and M. Nakhla, “A Fast Algorithm and Practical Considerations For Passive Macromodeling Of Measured/Simulated Data,” *IEEE Transactions on Advanced Packaging*, Vol. 27, N. 1, pp. 57–70, Feb. 2004.

[26] S. Grivet-Talocia, “An adaptive sampling technique for passivity characterization and enforcement of large interconnect macromodels,” *IEEE Trans. Adv. Packaging*, vol. 30, pp. 226–237, May 2007.

[27] A Lamecki and M. Mrozowski, “Equivalent SPICE Circuits With Guaranteed Passivity From Nonpassive Models,” *IEEE Transactions on Microwave Theory And Techniques*, Vol. 55, No. 3, March 2007, pp. 526–532.

[28] L. Gobbato, A. Chinea, S. Grivet-Talocia, “A Parallel Hamiltonian Eigensolver for Passivity Characterization and Enforcement of Large Interconnect Macromodels,” in *DATE 2011 - Design, Automation and Test in Europe, Grenoble, France*, pp. 26–31, March 14–18, 2011.

[29] S. Grivet-Talocia, S. B. Olivadese, and P. Triverio, “A compression strategy for rational macromodeling of large interconnect structures,” in *IEEE* 20th *Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS 2011), San Jose, CA*, pp. 53–56, October 23–26, 2011.

[30] G. H. Golub, C. F. Van Loan, *Matrix Computations*. The Johns Hopkins University Press, 3rd edition, 1996.

[31] 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

[32] Charles F. Van Loan, “The ubiquitous Kronecker product,” *Journal of Computational and Applied Mathematics*, vol. 123, pp. 85–100, 2000.

[33] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[34] S. Boyd, L. El Ghaoui, E. Feron, V. Balakrishnan, *Linear matrix inequalities in system and control theory, SIAM studies in applied mathematics*, SIAM, Philadelphia, 1994.

[35] Jos F. Sturm, “Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones”, *Optimization Methods and Software*, No. 11–12, 1999, pp. 625-653.

[36] J. H. Wilkinson, *The algebraic eigenvalue problem*, Oxford University Press, London, 1965.

[37] K. Zhou, J. C. Doyle, K. Glover, *Robust and Optimal Control*, Prentice Hall, 1996.

[38] E.-P. Li, E.-X. Liu, L.-W. Li, M.-S. Leong, “A coupled efficient and systematic full-wave time-domain macromodeling and circuit simulation method for signal integrity analysis of high-speed interconnects,”, *IEEE Trans. on Advanced Packaging*, vol. 27, no. 1, pp. 213–223, Feb. 2004.

[39] W.H.A. Schilders, H.A. Van Der Vorst, and J. Rommes, *Model order reduction: theory, research aspects and applications*, Springer Verlag, 2008.

[40] A.C. Antoulas, *Approximation of large-scale dynamical systems*, Society for Industrial and Applied Mathematics, 2005.

[41] J.M.S. Silva, L.M. Silveira, L.M., “On the Compressibility of Power Grid Models,” 2007 IEEE Computer Society Annual Symposium on VLSI, 9-11 March 2007, pp. 186–191. .

PLACE PHOTO HERE

Salvatore Bernardo Olivadese received the Laurea degree (B.Sc) in Information Technology in 2007 and the Laurea Specialistica degree (M.Sc) in Electronic Engineering in 2009, both from Politecnico di Torino, Italy, where He is currently working towards his Ph.D. degree. In 2009, He visited the Interconnect and Packaging Analysis Group of IBM T.J. Watson Research Center, Yorktown, NY, developing an Adaptive Frequency Sampling method. His research interests concern modeling of interconnects, circuit simulation and parallel algorithms.

PLACE PHOTO HERE

Stefano Grivet-Talocia (M’98–SM’07) received the Laurea and the Ph.D. degrees in electronic engineering from Politecnico di Torino, Italy. From 1994 to 1996, he was with the NASA/Goddard Space Flight Center, Greenbelt, MD, USA. Currently, he is an Associate Professor of Circuit Theory with Politecnico di Torino. His research interests are in passive macromodeling of lumped and distributed interconnect structures, modeling and simulation of fields, circuits, and their interaction, wavelets, time-frequency transforms, and their applications. He is author of more than 120 journal and conference papers. He is co-recipient of the 2007 Best Paper Award of the IEEE Trans. Advanced Packaging. He received the IBM Shared University Research (SUR) Award in 2007, 2008 and 2009. Dr. Grivet-Talocia served as Associate Editor for the IEEE TRANSACTIONS ON ELECTROMAGNETIC COMPATIBILITY from 1999 to 2001.