A compression strategy for rational macromodeling of large interconnect structures

Original

Availability:
This version is available at: 11583/2474980 since:

Publisher:
IEEE

Published
DOI:10.1109/EPEPS.2011.6100184

Terms of use:
This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2011 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)
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 a 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 $S_l \in \mathbb{C}^{P \times P}$ at frequencies $\omega_l$, with $l = 1, \ldots, L$. 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

$$S(s) = S_\infty + \sum_{n=1}^{N} \frac{R_n}{s - p_n},$$

where $p_n$ are the poles of the macromodel, $R_n$ are the associated residue matrices, and $S_\infty$ is the direct coupling term. Standard formulations of the VF algorithm minimize the global model error

$$\sum_{l=1}^{L} \|S(\omega_l) - S_l\|_F^2,$$

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
Each row of the scattering matrix entries, magnitude (dB) with constant coefficients \( \alpha_{ij}^{(i,j)} \) 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 \( S_l \), \( \forall l \). For each selected frequency \( \omega_l \), we stack all elements of the scattering matrix into a single row-vector \( x_l \in \mathbb{C}^{P^2} \), constructed as \( x_l = \text{vec}(S_l)^T \). We recall that the vec() operator stacks all columns of its matrix element into a single column vector. More precisely, element \((S_l)_{ij}\) with \( 1 \leq i, j \leq P \) corresponds to element \((x_l)_k\) for \( 1 \leq k \leq P^2 \) through

\[
\begin{align*}
    k & = i + (j - 1)P \\
    i & = 1 + \text{mod}(k - 1, P) \\
    j & = \lceil k/P \rceil
\end{align*}
\]

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 \( x_l \) corresponding to different frequencies \( \omega_l \) as rows in a matrix \( X \in \mathbb{C}^{L \times P^2} \)

\[
X = \begin{bmatrix}
    x_1 & \cdots & x_L
  \end{bmatrix}
\]

Each row \( x_j \) of this matrix corresponds to a single frequency \( \omega_j \), while each column \( z_k \) collects all frequency samples of a single scattering response \((x_k)_l = S_{ij}(\omega_l)\).

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 \( X \) 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 \( X \) leads to

\[
X = \bar{U}\bar{\Sigma}\bar{V}^H = \tilde{W}\tilde{V}^H
\]

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

\[
z_k = \sum_q \tilde{v}_{kq} w_q.
\]

This expression is exact, with no approximation error, if all singular vectors are considered in the expansion. Each sampled scattering response is thus represented as a superposition of “basis vectors” \( \bar{w}_q \), 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 \( X = X^r + jX^i \), where \( X^r, X^i \in \mathbb{R}^{L \times P^2} \), or equivalently

\[
X = [I_L \ jI_L] \begin{bmatrix}
    X^r \\
    X^i
  \end{bmatrix}
\]

where \( I_L \) is the identity matrix of size \( L \). Then, we perform a truncated SVD decomposition, where only the first \( \rho \) singular values are retained

\[
\begin{bmatrix}
    X^r' \\
    X^i'
  \end{bmatrix} = U \Sigma V^T \simeq \bar{U} \bar{\Sigma} \bar{V}^T,
\]

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

\[
\tilde{W} = [I_L \ jI_L] \bar{U}\bar{\Sigma}
\]

we obtain the low-rank approximation

\[
X \simeq \bar{X} = \tilde{W}\tilde{V}^T
\]

Equivalently,

\[
z_k \simeq \sum_{q=1}^{\rho} v_{kq} w_q,
\]

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

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

\[
E_2 = \|X - \bar{X}\| = \|I_L \ jI_L\| \|\bar{U}\bar{\Sigma}V^T - U\Sigma V^T\| \leq \|I_L \ jI_L\| \|\bar{U}\bar{\Sigma}V^T - U\Sigma V^T\| \leq \sqrt{2}\bar{\sigma}_{\rho+1},
\]

where \( \bar{\sigma}_{\rho+1} \) is the \( \rho+1 \)-th singular value of \( \bar{U}\bar{\Sigma} \).

Fig. 1. Various scattering responses of a high-speed connector (top curves: reflection coefficients, bottom curves: crosstalks).
instead of $p$ is drastically reduced. This is indeed the case, as we will show for all basis functions in $\omega_m$.

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 $\rho$ 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 $w(s)$, 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 $\bar{V}$, since $\bar{V}^T \bar{V} = I$, we obtain

$$w_q \simeq \sum_{k=1}^{P^2} v_{kq} z_k,$$

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

$${w(s)}^T \leftrightarrow \begin{pmatrix} A_w & B_w \\ C_w & D_w \end{pmatrix},$$

with $A_w \in \mathbb{R}^{N_w \times N_w}$, $B_w \in \mathbb{R}^{N_w \times 1}$, $C_w \in \mathbb{R}^{P \times N_w}$, $D_w \in \mathbb{R}^{P \times 1}$. We then define a (reshaped) global rational macromodel according to the expansion (11), as

$${X^T(s)} = \bar{V} w(s) =$$

$$= \bar{V} D_w + \bar{V} C_w (\hat{S}_{N_w}^{-1} - A_w)^{-1} B_w,$$

where $X^T(s)$ 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

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

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

$$A = I_P \otimes A_w, \quad B = I_P \otimes B_w,$$

$$C = \Psi(I_P \otimes C_w), \quad D = \Psi(I_P \otimes D_w),$$

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

$$\Psi = \left( (\bar{V})_1 \ (\bar{V})_2 \ \cdots \ (\bar{V})_P \right)$$

with $(\bar{V})_j \in \mathbb{R}^{P \times P}$ collecting the $P$ rows $\{j(P - 1) + 1, \ldots, jP\}$ of matrix $\bar{V}$. 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 $w(s)$ at each raw frequency point $\omega$ and we collect the results as rows in matrix $\bar{W} \in \mathbb{C}^{L \times P}$, which in turn is used to reconstruct the samples of the global rational macromodel, collected in matrix $\bar{X} = \bar{W} \bar{V}^T$. Due to the orthonormality of the columns of $\bar{V}$, we have

$$\|\bar{X} - \hat{X}\|_2 = \|\bar{W} \bar{V}^T - \hat{W} \bar{V}^T\|_2 = \|\bar{W} - \hat{W}\|_2.$$
It results in a fitting error that is identical to the fitting error achieved in the construction of the low-rank system $w(s)$.

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

$$
\delta_2 = \| X - \tilde{X} \|_2 \leq \| X - \bar{X} \|_2 + \| \bar{X} - \tilde{X} \|_2 
= \sqrt{2} \sigma_{D+1} + \| \bar{W} - \tilde{W} \|_2, \tag{23}
$$

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 $\rho$ has been automatically determined in order to reach a cumulative compression error $\epsilon_2 < 0.1$. Similarly, the number of poles $N_w$ of the low-rank macromodel has been determined in order to achieve a cumulative VF error $\| X - \tilde{X} \|_2 < 0.1$. 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 $\epsilon_2$ and the final error $\delta_2$ 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


\begin{table}[!ht]
\centering
\small
\begin{tabular}{|c|c|c|c|c|c|c|c|c|}
\hline
\text{test} & $P$ & $L$ & $\rho$ & $N_w$ & $N$ & $\epsilon_2$ & $\delta_2$ & $\text{VF(X)}$ [s] & $\text{SVD}$ [s] & $\text{VF(W)}$ [s] & SpeedUp \\
\hline
1 & 12 & 471 & 17 & 20 & 22 & 0.07 & 0.102 & 3.05 & 0.36 & 0.5 & 3.5x \\
2 & 24 & 1001 & 13 & 12 & 12 & 0.01 & 0.03 & 8.6 & 1.1 & 0.2 & 6.6x \\
3 & 34 & 570 & 40 & 57 & 58 & 0.072 & 0.096 & 263.1 & 1.9 & 5.1 & 36.9x \\
4 & 41 & 572 & 10 & 11 & 11 & 0.05 & 0.05 & 18.5 & 1.82 & 0.26 & 8.8x \\
5 & 48 & 690 & 24 & 27 & 28 & 0.06 & 0.102 & 119.6 & 3.7 & 1.09 & 24.9x \\
6 & 52 & 13 & 3 & 3 & 3 & 0.01 & 0.01 & 0.66 & 0.04 & 0.07 & 5.9x \\
7 & 56 & 1001 & 30 & 30 & 30 & 0.06 & 0.08 & 198.36 & 7.3 & 1.7 & 22.0x \\
8 & 92 & 71 & 22 & 22 & 23 & 0.06 & 0.106 & 32.8 & 1.4 & 0.1 & 21.8x \\
9 & 245 & 197 & 14 & 99 & 93 & 0.01 & 0.09 & 12885.2 & 24.5 & 1.29 & 499.2x \\
10 & 300 & 40 & 8 & 8 & 8 & 0.02 & 0.015 & 432.8 & 34.7 & 0.97 & 124x \\
\hline
\end{tabular}
\caption{SUMMARY OF THE COMPRESSED MACROMODELING RESULTS FOR SEVERAL BENCHMARKS. SEE TEXT FOR DETAILS}
\end{table}