

# Issues in Model Reduction of Power Grids

João M. S. Silva<sup>1</sup> and L. Miguel Silveira<sup>2</sup>

<sup>1</sup> INESC-ID, R. Alves Redol 9, 1000-029 Lisboa, Portugal, [jmss@algos.inesc-id.pt](mailto:jmss@algos.inesc-id.pt)

<sup>2</sup> INESC-ID, R. Alves Redol 9, 1000-029 Lisboa, Portugal, [lms@inesc-id.pt](mailto:lms@inesc-id.pt)

**Abstract.** *Power grid analysis has recently risen to prominence due to the widespread use of lower supply voltages by power-conscious designs. Low supply voltages imply smaller noise margins and make the voltage drop across the power grid very critical since it can lead to overall slower circuits, signal integrity issues and ultimately to circuit malfunction. Verifying proper behavior of a power grid is a difficult task due to the sheer size of such networks. The usual solution to this problem is to apply reduced-order modeling techniques to generate a smaller macro-model. These techniques are typically based on projections to subspaces whose dimension is determined by the input space. Unfortunately power grids are characterized by a massive number of network ports, which limits the amount of compression achievable. Recently, new algorithms have been proposed for solving this problem which may provide efficient alternatives. In this paper we discuss the main issues related to model reduction of power grid networks and compare several methods for such reduction, providing some insight into the problem and how it can be tackled.*

## 1 Introduction

Power dissipation is widely recognized as the greatest challenge to the continuing trend for higher performance fueled by technology scaling, increased functionality, and competitive designs. Increased chip functionality results in the need for huge power distribution networks, also referred to as power grids. A common technique to lower power consumption in such designs is to scale down the supply voltages, since chip power is roughly proportional to the square of the supply voltage. However, lower supply voltages imply smaller noise margins and make the voltage drop across the power grids very critical since it can lead to reduced noise margins and overall slower circuits. Once voltage drops exceed designer-specified thresholds, signal integrity violation occurs and circuit functionality is compromised with obvious yield consequences. Reduced noise margins may induce false switching and higher logic gate delays. This may directly cause chip failures or simply slow down the circuit enough so that timing requirements cannot be met.

Verifying proper behavior under realistic operating conditions requires accurate power grid analysis. However, analyzing power grids is a monumental task due to their sheer size which all but precludes direct usage in standard

simulation environments. A possible solution to this problem is to compress the model using a model reduction technique. Model order reduction (MOR) algorithms are the backbone of contemporary parasitic and interconnect modeling technologies. These algorithms take as input a linear interconnect model and produce as output a smaller model that is suitable for simulation in conjunction with nonlinear circuit elements. The effectiveness of the model reduction algorithm is judged by the decrease in the reduced circuit simulation time, compared to simulation with the full model, assuming acceptable error is incurred in the modeling process. MOR algorithms rely on the fact that on a variety of contexts only an accurate approximation to the input-output behavior of a dynamic linear system is necessary [1, 2, 3]. This is true for instance for delay analysis since only the waveforms at the gate inputs and outputs matter. Therefore even if one has to account for interconnect effects, the precise time-variation at any interconnection point is not relevant unless such point is a gate input or output. It is quite typical for MOR techniques to be able to reduce large RC(L) interconnect networks with just a few ports to models with very few states and still produce very accurate approximations of frequency- and time-domain behavior. In other words, even if the number of internal states,  $n$ , is very high, the description of the multi-port network is an  $q \times p$  matrix valued transfer function where  $p, q \ll n$  and typically only a few states are necessary for the required accuracy. The compression ratio is therefore quite high. Of course, it is reasonable to expect that when the number of ports increases, then the number of states to be retained must also increase since, in a simplified sense, that means we now care for an increasing number of internal points/states (i.e.  $p$  or  $q$  above increase). Ultimately, however, as the number of ports increases, the model must be able to accurately characterize the interaction between all input and output ports. If the number of retained states keeps increasing, this appears to leave little room for compression as the size of the matrix transfer function that characterizes all port interactions,  $\mathcal{O}(q \times p)$ , also increases and may approach the complexity of working with the original network equations. In Section 4 we will verify this relation in a precise manner and discuss its implications. Nevertheless, it is important to understand the reasons behind this loss of efficiency since knowledge of the specific scenarios where each method may produce better results is an important asset when determining how to perform the reduction.

Recently, the efficient reduction of systems with a large number of ports has been addressed and several methods have been proposed [4, 5, 6, 7]. In this paper we discuss the main issues related to order reduction of power grid networks and compare several methods for solving this problem, providing some insight into the problem and how it can be tackled. In Section 2 we present the standard model-order reduction methods that are now in widespread use in many applications in several fields, including electronic design automation. In Section 3 we discuss the newly proposed methods for handling massively-coupled linear dynamic systems as well as alternative approaches which are not based on projection schemes [8]. Then in Section 4 we present the problem

of power grid reduction and discuss some of its characteristics. We analyze the conditions in which it can be successfully reduced and the impact of an increasing number of ports. We also discuss scenarios in which the reduction might lead to better or worse compression ratios. In Section 5 we show results from applying the various methods, in a variety of settings to the power grid problem. Finally conclusions are drawn in Section 6.

## 2 Background

Model reduction algorithms are the backbone of contemporary parasitic and interconnect modeling technologies. Projection-based Krylov subspace algorithms, in particular, provide a general-purpose, rigorous framework for deriving interconnect modeling algorithms. Another class of methods that is sometimes used for model reduction and which finds its roots in systems and control theory are related to balancing transformations of the system state description. All of these algorithms take as input a linear interconnect model, and produce as output a smaller model that is suitable for simulation in conjunction with nonlinear circuit elements. The effectiveness of the model reduction algorithm is judged by the decrease in final circuit simulation time, compared to simulation with the full model, assuming acceptable error is incurred in the modeling process.

Considering an RC network, the nodal analysis formulation leads to

$$\begin{aligned} \mathbf{C}\dot{\mathbf{v}} + \mathbf{G}\mathbf{v} &= \mathbf{M}\mathbf{u} \\ \mathbf{y} &= \mathbf{N}^T\mathbf{v} \end{aligned} \quad (1)$$

where  $\mathbf{C}, \mathbf{G} \in \mathbb{R}^{n \times n}$  are the capacitance and conductance matrices, respectively,  $\mathbf{M} \in \mathbb{R}^{n \times p}$  is a matrix that relates the inputs,  $\mathbf{u} \in \mathbb{R}^p$  to the states,  $\mathbf{v} \in \mathbb{R}^n$ , that describe the node voltages,  $\mathbf{N} \in \mathbb{R}^{n \times q}$  its counterpart with respect to the outputs,  $\mathbf{y} \in \mathbb{R}^q$ ,  $n$  is the number of states,  $p$  the number of inputs and  $q$  the number of outputs. The matrix transfer function of the network is then given by

$$\mathbf{H}(s) = \mathbf{N}^T(\mathbf{G} + s\mathbf{C})^{-1}\mathbf{M} \quad (2)$$

The goal of model-order reduction is, generically, to determine a new model,

$$\mathbf{H}_r(s) = \hat{\mathbf{N}}^T(\hat{\mathbf{G}} + s\hat{\mathbf{C}})^{-1}\hat{\mathbf{M}} \quad (3)$$

that closely matches the input-output behavior of the original model, and where the state description is given by  $\mathbf{z} = \mathbf{V}^T\mathbf{v} \in \mathbb{R}^r, r \ll n$ . Note however, even if  $r \ll n$ , the reduced-order model may still fail to provide relevant compression. This may happen because, for large networks, the matrices  $\mathbf{C}, \mathbf{G}$  are very sparse, having a number of non-zeros entries of order  $\mathcal{O}(n)$ . So, if the number of non-zero entries in the reduced-order model increases for instance with the number of ports, the benefits of reduction may vanish with increasingly large  $p$  and  $q$ .

In the following we review the standard model-order reduction techniques in order to understand their basic modes of operation.

## 2.1 Projection-Based Framework

Projection-based algorithms such as PRIMA [3], or PVL [9], have been shown to produce excellent compression in many scenarios involving on- and off-chip interconnect and packaging structures. The PRIMA algorithm [3] reduces a state-space model in the form of (1) by use of a projection matrix  $\mathbf{V}$ , through the operations

$$\hat{\mathbf{G}} = \mathbf{V}^T \mathbf{G} \mathbf{V} \quad \hat{\mathbf{M}} = \mathbf{V}^T \mathbf{M} \quad \hat{\mathbf{C}} = \mathbf{V}^T \mathbf{C} \mathbf{V} \quad \hat{\mathbf{N}} = \mathbf{V}^T \mathbf{N} \quad (4)$$

to obtain a reduced model in the form of (3). In the standard approach, the  $\mathbf{V}$  matrix is chosen as an orthogonal basis of a block Krylov subspace,  $\mathbf{K}_m(\mathbf{A}, \mathbf{p}) = \text{span}\{\mathbf{p}, \mathbf{A}\mathbf{p}, \dots, \mathbf{A}^{m-1}\mathbf{p}\}$ . A typical choice is  $\mathbf{A} = \mathbf{G}^{-1}\mathbf{C}$ ,  $\mathbf{p} = \mathbf{G}^{-1}\mathbf{M}$ . The construction of the projection matrix  $\mathbf{V}$  is done in an iterative block fashion, with each block  $i$  being the result of back-orthogonalizing  $\mathbf{A}^{i-1}\mathbf{p}$  with respect to all previously computed blocks. When the projection matrix is constructed in this way, the moments of the reduced model match the moments of the original model at least to order  $m$  (in PVL,  $2m+1$  moments are matched). The difficulty with these algorithms is that the model size is proportional to the number of moments matched multiplied by the number of ports. For example, consider the application of such an algorithm to a network with a large set of input ports. If only two (block) moments are to be matched at each port, and the network has 1000 ports, the resulting model will have at least 2000 states, and the reduced system matrices will be dense. Therefore such methods are almost impractical for networks with large numbers of input/output ports, that is, for networks with many columns in the matrices defining the inputs. This is often the case for such “massively coupled” parasitics networks as occur in substrate and package modeling, as well as power grids.

## 2.2 Multi-Point Rational Approximation

An evolution of Krylov-subspace schemes are methods that construct the projection matrix  $\mathbf{V}$  from a rational, or multi-point, Krylov subspace [10, 11, 12]. Compared to the single-point Krylov-subspace projectors, for a given model order, the multi-point approximants tend to be more accurate, but are usually more expensive to construct. Given  $N$  complex frequency points,  $s_i$ , a projection matrix may be constructed whose  $i$ -th column is

$$\mathbf{z}_i = (\mathbf{G} + s_i \mathbf{C})^{-1} \mathbf{M} \quad (5)$$

This leads to multi-point rational approximation. Multi-point projection is known to be an efficient reduction algorithm in that the number of columns, which determines the final model size, is usually small for a given allowable approximation error, at least compared to pure moment matching approaches. Of course there are many practical questions to ponder in an actual implementation, namely how many points  $s_i$  should be used, and how should the  $s_i$  be chosen. Lack of an automatic procedure to solve these problems has limited the applicability of the methods.

### 2.3 Truncated Balanced Realization (TBR)

An alternative class of reduction algorithms are based on Truncated Balanced Realization (TBR) [13, 14]. The TBR algorithm first computes the observability and controllability Gramians,  $\mathbf{X}$ ,  $\mathbf{Y}$ , from the Lyapunov equations

$$\mathbf{G}\mathbf{X}\mathbf{C}^T + \mathbf{C}\mathbf{X}\mathbf{G}^T = \mathbf{M}\mathbf{M}^T, \quad (6)$$

$$\mathbf{G}^T\mathbf{Y}\mathbf{C} + \mathbf{C}^T\mathbf{Y}\mathbf{G} = \mathbf{N}^T\mathbf{N} \quad (7)$$

and then reduces the model by projection onto the space associated with the dominant eigenvalues of the product  $\mathbf{X}\mathbf{Y}$  [13]. Model size selection and error control in TBR is based on the eigenvalues of  $\mathbf{X}\mathbf{Y}$ , also known as the the Hankel singular values,  $\sigma_r$ . In the proper case, there is a theoretical bound on the frequency-domain error in the order  $r$  TBR model, given by [14]

$$\|\mathbf{H} - \mathbf{H}_r\| \leq 2 \sum_{i=r+1}^n \sigma_i \quad (8)$$

The existence of such an error bound is an important advantage of the TBR-like class of algorithms. Unfortunately there is no counterpart in the projection-based class of algorithms. Note that the model selection criteria does not depend *directly* on the number of inputs. However, as we shall see, there is an indirect dependence in most problems. In principle, it is possible to have a 1000-port starting model, and obtain a good reduced model of only, say, 10 states, if the  $\mathbf{G}$ ,  $\mathbf{C}$ ,  $\mathbf{M}$ ,  $\mathbf{N}$  matrices are such that all but the the first few (10) Hankel singular values are small. In practice, solution of the Lyapunov equations is too computationally intensive for large systems as encountered in interconnect analysis. Therefore, a variety of approximate methods [12, 15, 16] have been proposed.

## 3 Massively-Coupled Problems

In the previous section we briefly summarized the main techniques for model order reduction of linear interconnect networks currently in use. As discussed, the projection-based techniques, like PVL or PRIMA present two problems when dealing with networks with a large number of ports. First, the cost associated with model computation is directly proportional to the number of inputs,  $p$ , i.e. to the number of columns in the matrices defining the inputs. This is easy to see by noting that the number of columns in the projection matrix  $\mathbf{V}$  in (4) is directly proportional to  $p$  (a direct result of the block construction procedure). This implies that model construction for systems with large number of ports is costly. Furthermore, the size of the reduced model is also proportional to  $p$ , as was discussed earlier and can directly be seen from (4). While the cost of model construction can perhaps be amortized in later simulations, the large size of the model is more problematic since it implies a direct penalty for such simulations.

This is often the case for such “massively coupled” parasitic networks as occur in substrate, package, power grids or clock distribution networks. Massively-coupled problems are problems for which the system description contains a very large number of ports. In this section, we summarize two recent methods aimed at solving some of the issues related to reduction of such systems.

### 3.1 Singular Value Decomposition MOR (SVD MOR)

The SVD MOR [4] algorithm was developed to address the reduction of systems with a large number of ports, like power-grids. While the size of a reduced model produced via PRIMA is directly proportional to the number of ports in the circuit, SVD MOR theoretically overcomes this problem using singular value decomposition (SVD) analysis in order to truncate the system to any desired order.

The main idea behind SVD MOR is to assume that there is a large degree of correlation between the various inputs and outputs. SVD MOR further assumes that such input-output correlation can be captured quite easily from observation of some system property, involving matrices  $\mathbf{M}$  and  $\mathbf{N}$ . The method can, for instance, use an input-output correlation matrix, like the one given by the zero-th order moment matrix  $\mathbf{S}_{DC} = \mathbf{N}^T \mathbf{G}^{-1} \mathbf{M}$ , which contains only DC information. Alternatively more complicated response correlations can be used such as a zero-th order,  $s_j$ -shifted moment,  $\mathbf{S}_{DC}^{(s_j)} = \mathbf{N}^T (\mathbf{G} + s_j \mathbf{C})^{-1} \mathbf{M}$ , a more generic  $k$ -order moment,  $\mathbf{S}_r = \mathbf{N}^T (\mathbf{G}^{-1} \mathbf{C})^k \mathbf{G}^{-1} \mathbf{M}$ , or even combinations of these. If we let  $\mathbf{B}$  be the appropriate correlation matrix, and if the basic correlation hypothesis holds true, then  $\mathbf{B}$  can be approximated by a low-rank matrix. This low rank property can be revealed by computing the SVD of  $\mathbf{B}$ ,

$$\mathbf{B} = \mathbf{U} \Sigma \mathbf{W}^T \quad (9)$$

where  $\mathbf{U}, \mathbf{W}$  are orthogonal matrices and  $\Sigma$  is the diagonal matrix containing the ordered singular values. Assuming correlation, there will be only a small number,  $r \ll p + q$ , of dominant singular values. Therefore

$$\mathbf{B} \approx \mathbf{U}_r \Sigma_r \mathbf{V}_r^T \quad (10)$$

where truncation is performed leaving the  $r$  most significant singular values. The method then approximates:

$$\begin{aligned} \mathbf{M} &\approx \mathbf{b}_m \mathbf{V}_r^T = \mathbf{M} \mathbf{V}_r (\mathbf{V}_r^T \mathbf{V}_r)^{-1} \mathbf{V}_r^T \\ \mathbf{N} &\approx \mathbf{b}_n \mathbf{U}_r^T = \mathbf{N} \mathbf{U}_r (\mathbf{U}_r^T \mathbf{U}_r)^{-1} \mathbf{U}_r^T \end{aligned} \quad (11)$$

where  $\mathbf{b}_m$  and  $\mathbf{b}_n$  are obtained using the Moore-Penrose pseudo-inverse, resulting in:

$$\mathbf{H}(s) \approx \mathbf{U}_r \underbrace{\mathbf{b}_n^T (\mathbf{G} + s \mathbf{C})^{-1} \mathbf{b}_m}_{\mathbf{H}_r(s)} \mathbf{V}_r^T \quad (12)$$

Standard MOR methods, like SyMPVL [17] or PRIMA, can now be applied to  $\mathbf{H}_r(s)$ , resulting in the final reduced model:

$$\mathbf{H}(s) \approx \mathbf{H}_r(s) = \mathbf{U}_r \tilde{\mathbf{H}}_r(s) \mathbf{V}_r^T \quad (13)$$

In our implementation we used PRIMA to obtain  $\mathbf{H}_r(s)$ . The final reduced system is  $p \times q$  with a number of nonzero elements of order  $\mathcal{O}(r^2)$ .

### 3.2 Poor Man's TBR (PMTBR)

The PMTBR algorithm [7, 6] was motivated by a connection between frequency-domain projection methods and approximation to truncated balanced realization. The method is less expensive in terms of computation, but tends to TBR when the order of the approximation increases. The actual mechanics of the algorithm are akin to multi-point projection, summarized in Section 2.2. In a multi-point rational approximation, the projection matrix columns are computed by sampling in several frequency points along a desired frequency interval

$$\mathbf{z}_i = (\mathbf{G} + s_i \mathbf{C})^{-1} \mathbf{M} \quad (14)$$

where  $s_i$ ,  $i = 1, 2, \dots, N$ , are  $N$  frequency sample points. The frequency-sampled matrix thus obtained can then be used to project the original system in order to obtain a reduced model.

In the PMTBR algorithm, a similar procedure is used. The connection to TBR methods is made by noting that an approximation  $\hat{\mathbf{X}}$  to the Gramian  $\mathbf{X}$  can be computed as

$$\hat{\mathbf{X}} = \sum_i w_i \mathbf{z}_i \mathbf{z}_i^H \quad (15)$$

where  $s_i = j\omega_i$  and the  $\omega_i$  and  $w_i$  can be interpreted as nodes and weights of a quadrature scheme applied to a frequency-domain interpretation of the Gramian matrix (see [7] for details). If we let  $\mathbf{Z}$  be a matrix whose columns are the  $\mathbf{z}_i$ , and  $\mathbf{W}$  is now the diagonal matrix of the square root of the weights, Eqn. (15) can be written more compactly as

$$\hat{\mathbf{X}} = \mathbf{Z} \mathbf{W}^2 \mathbf{Z}^H \quad (16)$$

If the quadrature rule applied is accurate,  $\hat{\mathbf{X}}$  will converge to  $\mathbf{X}$ , which implies the dominant eigenspace of  $\hat{\mathbf{X}}$  converges to the dominant eigenspace of  $\mathbf{X}$ . If we compute the singular value decomposition of  $\mathbf{Z} \mathbf{W}$ .

$$\mathbf{Z} \mathbf{W} = \mathbf{V}_Z \mathbf{S}_Z \mathbf{U}_Z \quad (17)$$

with  $\mathbf{S}_Z$  real diagonal,  $\mathbf{V}_Z$  and  $\mathbf{U}_Z$  unitary matrices, it is easy to see that  $\mathbf{V}_Z$  converges to the eigenspaces of  $\mathbf{X}$ , and the Hankel singular values are obtained directly from the entries of  $\mathbf{S}_Z$ .  $\mathbf{V}_Z$  can then be used as the projection matrix

in a model order reduction scheme. The method was shown to perform quite well in a wide variety of settings [16].

An interesting additional interpretation, and quite relevant for our purposes, was recently presented [6]. It has been shown that if further information revealing time-domain correlation between the ports is available, a variant of PMTBR can be used that can lead to significant efficiency improvement. This idea is akin to the basic assumptions in SVDMOR and relate to exploiting correlation between the inputs. Unlike SVDMOR, however, it is assumed that the correlation information is not contained in the circuit information directly, but rather in its inputs. In this variant of PMTBR, a correlation matrix  $\mathbf{K}$  is formed by columns which are samples of port values along the time-steps of some interval. Those samples, should characterize as well as possible the values expected at the inputs of the system, i.e.  $\mathbf{K}$  should be a suitably representative model of the possible inputs. An SVD is then performed over  $\mathbf{K}$  in order to retain only the most significant components of the input correlation information:

$$\mathbf{K} \approx \mathbf{U}_K \Sigma_K \mathbf{V}_K^T \quad (18)$$

With this additional correlation information, the samples relative to multi-point approximation become:

$$\mathbf{z}_i = (\mathbf{G} + s_i \mathbf{C})^{-1} \mathbf{M} \mathbf{U}_K \Sigma_K \quad (19)$$

Using the  $\mathbf{z}_i$  above as columns of the  $\mathbf{Z}$  matrix in (16) leads to the input-correlated TBR algorithm (ICTBR). See [16] for more details and a more thorough description of the probabilistic interpretation of both PMTBR as well as ICTBR.

### 3.3 Time Constant Equilibration Reduction (TICER)

TICER [8] is an RC model reduction method that behaves in a very efficient way. Model extraction tools usually obtain lumped element parasitics based on local changes in geometry. The resulting models have a huge variety of dynamics which can be reduced by TICER. This method analyzes the time constant associated with each extracted net and eliminates the ones with a time constant outside a given interval. This way, a realizable RC circuit which maintains the original network topology is obtained.

The time constant associated with a node  $N$  of a circuit is given by:

$$\tau_N = \frac{\chi_N}{\gamma_N} = \frac{\sum_k c_{kN}}{\sum_k g_{kN}} \quad (20)$$

where  $\chi_N$  and  $\gamma_N$  are, respectively, the equivalent capacitance and conductance seen by node  $N$ .  $\chi_N$  is obtained by adding the capacitances between node  $N$  and its neighbors,  $c_{kN}$ , and  $\gamma_N$  is obtained by adding the conductances between node  $N$  and its neighbors,  $g_{kN}$ .

From the point of view of the capacitors connected to a node  $N$ , if  $s\chi_N \gg \gamma_N$ , it is as if that node is floating and thus is considered a *slow* node. On the other hand, if  $s\chi_N \ll \gamma_N$ , the voltages of that node are at all times determined to be in DC equilibrium with its neighbors, i.e., that node is always fully relaxed, and is said to be a *quick* node.

With this in mind, TICER node elimination operates in two parts:

- If a node  $N$  is a *slow* node, it is eliminated and any pair of nodes  $(i, j)$  previously connected to  $N$  are now connected by a conductance  $g_{ij} = \frac{g_{iN}g_{jN}}{\gamma_N}$ . Moreover, if nodes  $i$  and  $j$  had capacitances connected to  $N$ , connect these nodes with a capacitance  $c_{ij} = \frac{c_{iN}c_{jN}}{\gamma_N}$ .
- If a node  $N$  is a *quick* node, it is eliminated and any pair of nodes  $(i, j)$  previously connected to  $N$  are now connected by a conductance  $g_{ij} = \frac{g_{iN}g_{jN}}{\gamma_N}$ . Moreover, if node  $i$  had a conductance to  $N$  and node  $j$  had a capacitance to  $N$  connect these nodes by a capacitance  $c_{ij} = \frac{g_{iN}c_{jN}}{\gamma_N}$ .

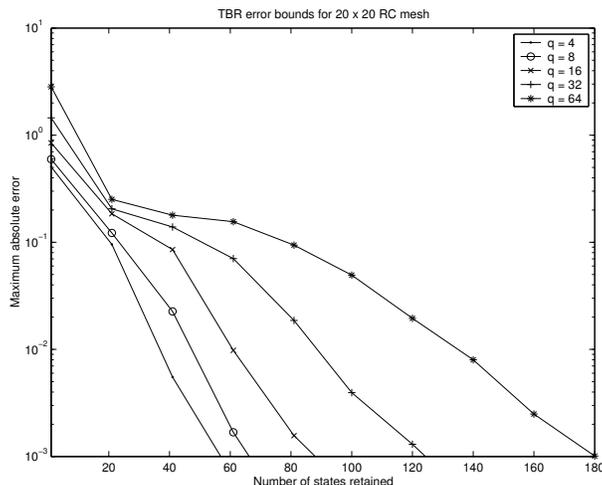
In the elimination process, ground node is treated like the remaining neighbors of the node to be eliminated.

The reduced model is an RC circuit and the output is passive, so stability and DC characteristics are exactly preserved. Notwithstanding, the drawback of this method is that when a node with  $n$  connections is eliminated those connections disappear but  $\frac{n(n-1)}{2}$  new connections appear. So, with TICER node elimination, the number of elements grows quadratically while the number of nodes decreases linearly. The method itself has linear complexity on the number of nodes.

## 4 Power-Grid Reduction

Both the standard model order reduction as well as the methods described in the previous section can be applied to massively coupled systems. Methods like SVDMOR are reported to provide significant advantages over the standard algorithms if certain conditions are met, namely that significant port correlation exists and can be ascertained in a practical way. PMTBR is a more general algorithm for model reduction, which can nonetheless be applied to large systems, given its reduced computational complexity. TICER, on the other hand, acts in a way that is similar to node elimination in a direct solver procedure. It can be used irrespective of the number of ports, but the resulting model tends to become denser as more nodes are eliminated.

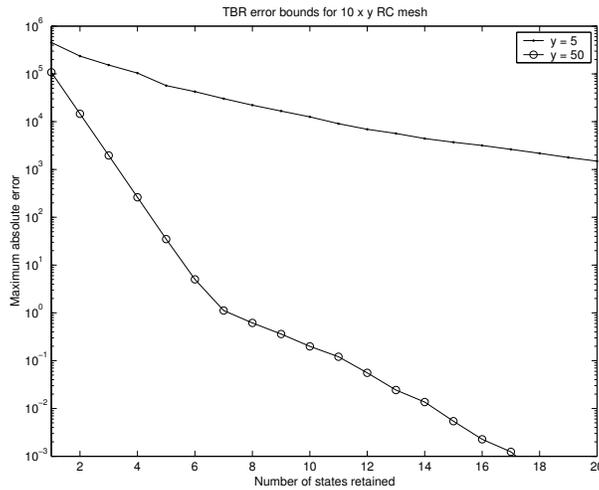
As stated previously, the difficulty with standard projection algorithms like PRIMA or multi-point projection schemes, is that the models produced have size proportional to the number of ports. This limits their applicability to problems such as power grids, where the number of network ports is likely to be very large. An interesting question that might be raised is whether this restriction is inherent to the system, given the number of ports, or an artifact of the computation scheme chosen. In other words, one might ask whether accurate modeling



**Fig. 1.** TBR error bounds for a  $20 \times 20$  RC grid as a function of the number of inputs,  $q$ .

and analysis of a power grid, modeled as a large RC grid, does indeed require so much dynamic information. This questions is all the more relevant as there is a common popular belief that only a few poles are required to accurately model an RC circuit. The roots of this problem are ancient and can be traced back to other domains like timing simulation. Here one asked the question of whether localized approximations of a node's behavior could be used for speeding up circuit simulation. It is now widely accepted that in certain settings that is indeed the case, but this conclusion is not general (see [18] for a discussion regarding simple RC models). Here a similar question is asked but now with respect to the number of ports.

To get some insight into the problem, it is interesting to consider a simplified scenario of a power grid and examine its behavior as the number of ports increases. Consider then a  $20 \times 20$  elements RC grid, representing a power network, and consider that the grid's inputs are positioned along the left side of the grid. Furthermore consider increasing the number of inputs by attaching more sources to the various grid nodes (i.e. adding more columns to  $\mathbf{M}$ ), again all located at the left and assume that the same nodes are observed (i.e.  $\mathbf{N}^T = \mathbf{M}$ ). As a proxy for system complexity, Figure 1 shows the TBR error bound from (8) obtained from the Hankel singular values as a function of the number of inputs. From the figure, we can see that indeed the order of the model required for acceptable accuracy grows with the number of inputs. Even in this simple setup, for the 64-input case, low-accuracy (say 20%) still requires at least a model with 120 states. A similar conclusion had been reached in [16] for the simpler case of an RC line. This result, seems to put into doubt the possibility of being able to perform model compression in such networks. Indeed, if 120 states (out

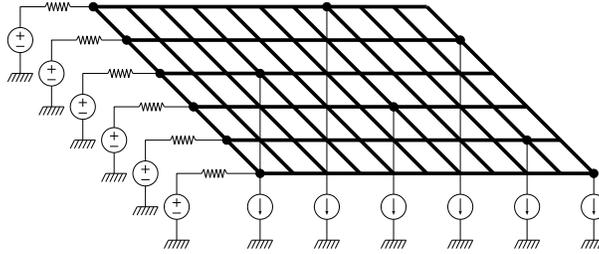


**Fig. 2.** TBR error bounds for a  $10 \times 5$  and a  $10 \times 50$  RC grid with separated inputs and output.

of a possible 400) are required for accuracy, then the chances of being able to perform significant reduction are small.

Consider now Figure 2 where the TBR error bound is again plotted, but now for two cases corresponding to a “thin”  $10 \times 5$  and a “fat”  $10 \times 50$  RC grids where a fixed number of inputs was used. Obviously neither grid is realistic in any way, but they serve the purpose of illustrating an important issue. Clearly the “fatter” grid, where the inputs are further away from the outputs, is much more compressible than the “thinner” grid. Indeed, for the “fatter” grid, only a handful of states are required even for high accuracy. The “thinner” grid shows the same behavior as before and seems fairly incompressible.

Figure 2 indicates that there is indeed hope for some reasonable reduction to be achieved. It also indicates that whenever inputs and outputs are widely separated, significant compression is possible. This is akin to the ideas of the multi-pole algorithm developed for electromagnetic modeling and used for instance in capacitance and inductance extraction. The effect on any point of a cluster of faraway input sources is individually indistinguishable. The system is therefore functionally similar to another one with just a few inputs. Therefore, only a few states are necessary to capture the various dynamics and the compression achievable is much greater. Unfortunately that situation is too restrictive for power grids in general, where ports are usually located all over the grid. Furthermore, the more likely scenario is that one will at least want to observe the potential at all grid nodes where inputs are connected (and thus where current spikes may appear). In this case, it is expected that the compression ratio will be small. Nevertheless, it is possible that high accuracy is only really needed to model the effect of nearby sources, while far away sources



**Fig. 3.** Setup for grid B.

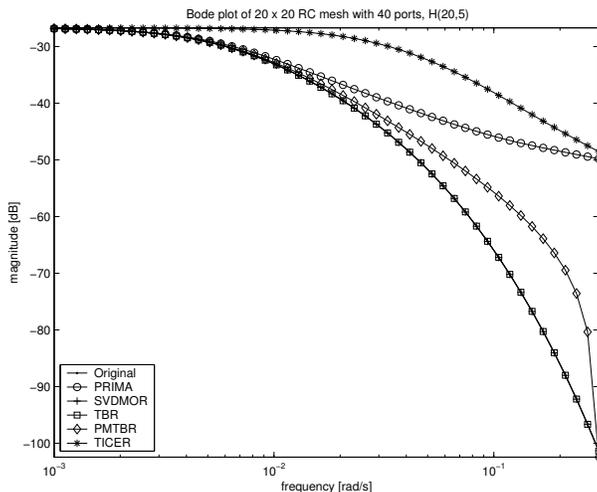
can be modeled in a coarser way. In this case smaller or at least sparser models should suffice.

## 5 Results

In this section we present results for reduction of power grids. Two types of topologies were tested: a mesh with voltage inputs on the left side and current outputs on the right one, which we term grid A, and a mesh with voltage ports along the left side and current ports randomly distributed over the remaining nodes, such as shown in Figure 3. We call this second setup grid B. There are two main differences between these two setups. The first one concerns formulation. While in grid A matrices  $M$  and  $N$  in Eqn. (1) are distinct ( $M$  yields input information and  $N$  yields output information) in grid B we have  $M = N$ , thus all ports are controllable and observable. The second main difference consists in the separation between ports. Relating back to the discussion in Section 4, in grid A the separation between inputs and outputs is maximal, while in grid B not only every port is both input and output, but also the geometric proximity between ports is reduced. We thus expect grid A to be fairly compressible, but smaller reductions to be seen for grid B. Grid A is similar to the one used in [4], while grid B was created in order to illustrate a more realistic setup.

The electric model of all grids is the following: every connection between nodes is purely resistive and in every node there is a capacitance to ground. Resistance and capacitance values were randomly generated in the interval  $(0.9, 1.1)$ .

In the following set of experiments the size of the reduced model is the same for all methods and was pre-determined. The correlation matrix of SVD MOR is a DC shifted moment with a shift of  $s = 0.1$  rad/s in normalized frequency, e.g.,  $\mathbf{M}_s = \mathbf{N}^T(\mathbf{G} + 0.1 \times \mathbf{C})^{-1}\mathbf{M}$ . For this method, after computing the SVD and choosing how many singular values to keep, a number of PRIMA iterations is performed in order to generate a model of the required size. The number of frequency samples of PMTBR was set such that we can draw a model of the same size from matrix  $\mathbf{Z}$ . Samples were chosen uniformly in the frequency range shown in the plots, and an additional sample added at DC. Concerning



**Fig. 4.** Bode plot of arbitrarily selected entry of  $20 \times 20$  transfer function matrix corresponding to grid A ( $r = 40$ ).

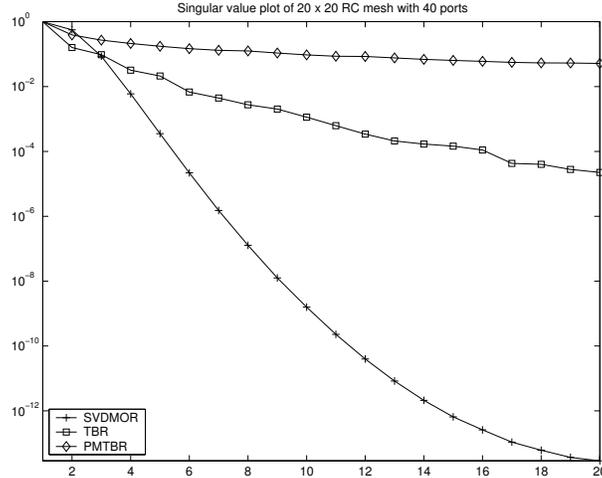
TICER, we computed the mean of the time constants of all nodes and began elimination of the nodes which are farther from the mean, until we reached the required model size.

### 5.1 Grid A

Grid A was originally used in [4] to illustrate the SVD MOR algorithm. We applied all previously discussed methods to reduce this grid. The Bode plot of an arbitrarily selected transfer function is presented in Figure 4. The number of retained states was forced at  $r = 40$ . In the case of SVD MOR, 4 singular values were kept and 10 PRIMA iterations were run, yielding the reduced model of  $4 \times 10 = 40$  states. We can observe that SVD MOR and TBR show good results, better than PMTBR, while PRIMA and TICER show a large error (using larger orders it is possible to produce an accurate approximation). In order to understand the reason for these results the plot of the singular values of all relevant methods is presented in Figure 5. We see that the singular values (s.v.) of  $\mathbf{M}_s$ , used by SVD MOR to guide the reduction, decay quite fast. Therefore keeping just the first 4 yields a good approximation. On the other hand the TBR Hankel s.v. and the PMTBR s.v. decay very slowly.

Notwithstanding, a Bode plot shows only one transfer function from the transfer matrix. Table 1 shows the infinity norm of the transfer matrix error,  $\|H(s) - H_r(s)\|_\infty$ . Analysis of the table indicates that in the overall model, TBR behaves better than SVD MOR for this grid setup.

With respect to TICER, its main advantage over the remaining methods is that it directly generates a realizable reduced model. Consequently, this method



**Fig. 5.** Normalized plot of singular values for grid A: SVD MOR matrix, TBR-Hankel singular values and PMTBR samples matrix.

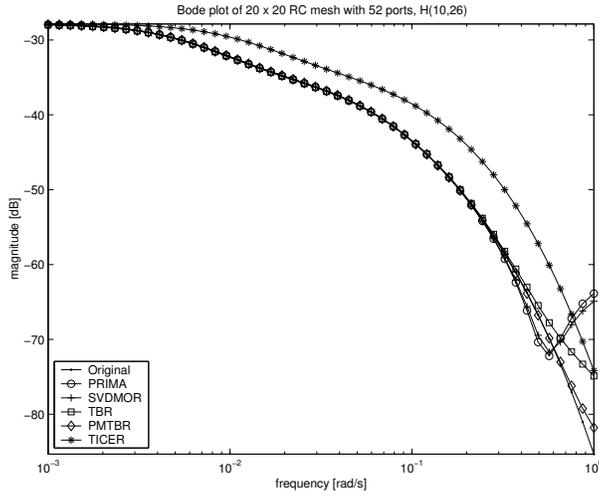
**Table 1.** Infinity norm of  $H(s) - H_r(s)$  for  $20 \times 20$  mesh with 20 inputs on the left side and 20 outputs on the right side. SVD MOR used 4 singular values.

$r = 40$	PRIMA	SVD MOR	TBR	PMTBR	TICER
$\ H - H_r\ _\infty$	2.391e-01	3.552e-04	1.320e-07	5.901e-02	8.085e-01

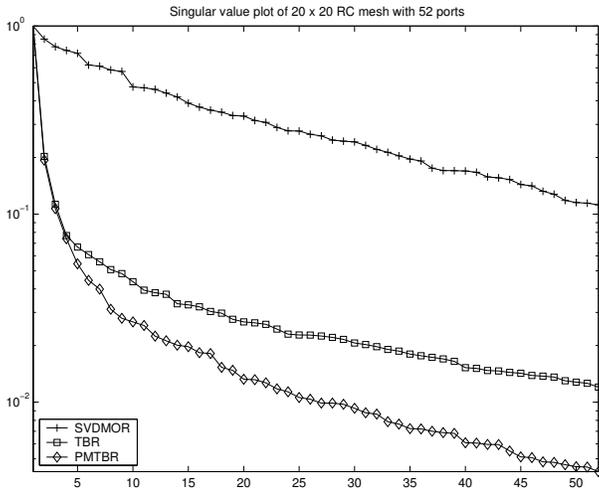
should be used whenever such a model is strictly necessary or in conjunction with other MOR method, since by itself it fails to obtain a reasonable approximation for small model sizes.

## 5.2 Grid B

In grid B the objective was to emulate a more realistic situation whereby potentially many devices, modeled as current sources, are attached to the power grid and can draw or sink current from/to it when switching. The number of current sources was chosen to be around 10% of the number of nodes. We have 32 current sources and 20 voltage sources. This is a harder problem to reduce, due to port proximity, and thus interaction, and the results show it. Again the Bode plot of an arbitrarily selected transfer function is presented in Figure 6. The number of retained states was now forced at  $r = 104$  (two times the number of ports) already showing smaller reduction than for grid A. In this case, the approximation produced by SVD MOR is less accurate. TBR and PMTBR produce the most accurate models. PRIMA shows a reasonable approximation while TICER fails to model accurately the behavior of the cutoff frequency. This was expected from inspection of Figure 7, where we see that the TBR Hankel s.v and the PMTBR s.v decay very fast, while the s.v. of  $\mathbf{M}_s$ , used by SVD MOR



**Fig. 6.** Bode plot of arbitrarily selected entry of  $20 \times 20$  transfer function matrix corresponding to grid B ( $r = 104$ ).



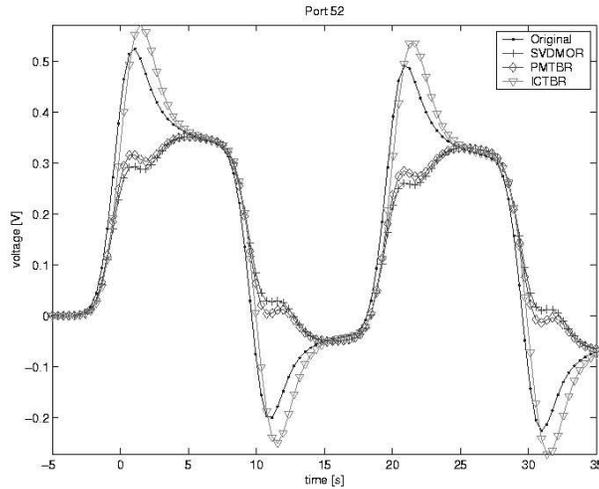
**Fig. 7.** Normalized plot of singular values for grid B: SVD MOR matrix, TBR-Hankel singular values and PMTBR samples matrix.

for reduction, decay very slowly. Clearly, the assumption of highly correlated ports is not valid here. The results concerning the error of the transfer matrix are in Table 2.

The matrices of the reduced models of both experiences are full with the exception of TICER. However, given the lack of accuracy of the TICER-generated models for this model size, such an advantage is of no consequence.

**Table 2.** Infinity norm of  $H - H_r$  for  $20 \times 20$  mesh with 20 ports on the left side and 32 randomly distributed ports over the mesh.

$r = 104$	PRIMA	SVDMOR	TBR	PMTBR	TICER
$\ H - H_r\ _\infty$	9.8.0e-02	8.071e-02	1.828e-02	1.195e-02	7.297e+00

**Fig. 8.** Input waveforms used for ICTBR experiment on grid B.

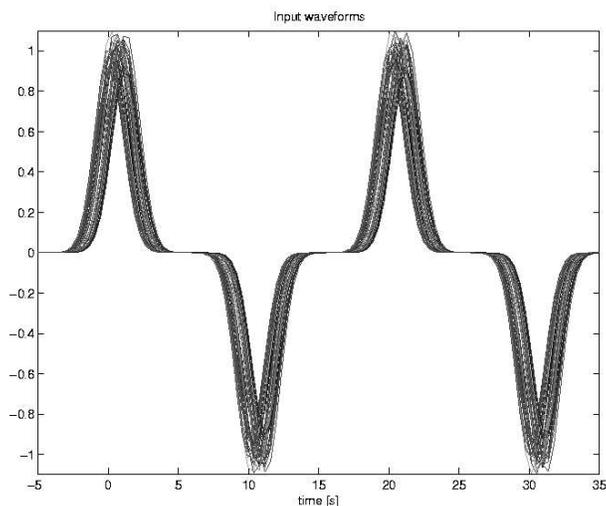
### 5.3 Time analysis: Input-Correlated TBR (ICTBR)

In this experiment, the ICTBR method, presented in Section 3.2, was used to generate a reduced model. We assumed that the grid inputs were correlated and had waveforms similar to those shown in Figure 8, which emulate transistor current signatures. The amplitude of the waveforms was randomly varied by 10%, while the phase shows a random 20% jitter.

Grid B was used for this experiment and the voltage resulting from the time analysis of one of the 32 ports connected to current sources is shown in Figure 9. The reduced models shown have size  $r = 40$  states (compare with size 104 used in Section 5.2). From the plot it is clear that only the 40-states ICTBR model can accurately mimic the voltage behavior of the port. This example shows that significant reduction can be obtained by exploiting input correlation.

## 6 Conclusions

In this paper we discuss several issues related to model order reduction of power grid networks and compare several standard and other recently proposed methods for solving this problem. We show that power grids present a strong challenge for model order reduction techniques and discuss scenarios in which this



**Fig. 9.** Time variation of randomly selected node voltage for the ICTBR experiment on grid B.

reduction might yield different compression ratios. We demonstrate through simple examples that achieving relevant compression requires a careful study of the grid characteristics and that no method produces the best solution in all scenarios. We also show that significant reductions can be achieved by exploiting known correlation between the input ports.

## 7 Acknowledgment

This work was partly supported by the Portuguese Foundation for Science and Technology under the grant SFRH/BD/10586/2002.

## References

1. Roland W. Freund and Peter Feldmann. Efficient Small-Signal Circuit Analysis and Sensitivity Computations with the PVL Algorithm. In *International Conference on Computer Aided-Design*, San Jose, California, November 1994.
2. L. Miguel Silveira, Mattan Kamon, and Jacob K. White. Efficient reduced-order modeling of frequency-dependent coupling inductances associated with 3-d interconnect structures. In *proceedings of the European Design and Test Conference*, pages 534–538, Paris, France, March 1995.
3. A. Odabasioglu, M. Celik, and L. T. Pileggi. PRIMA: passive reduced-order interconnect macromodeling algorithm. *IEEE Trans. Computer-Aided Design*, 17(8):645–654, August 1998.

4. Peter Feldmann. Model order reduction techniques for linear systems with large number of terminals. In *Proceedings of the Design, Automation and Test in Europe conference (DATE)*, volume 2, pages 944–947, Paris, France, February 2004.
5. Peter Feldmann and Frank Liu. Sparse and efficient reduced order modeling of linear subcircuits with large number of terminals. In *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, San Jose, California, U.S.A., November 2004.
6. L. Miguel Silveira and Joel Phillips. Exploiting input information in a model reduction algorithm for massively coupled parasitic networks. In *41<sup>st</sup> Design Automation Conference*, pages 385–388, San Diego, CA, USA, June 2004.
7. Joel Phillips and L. Miguel Silveira. Poor Man’s TBR: A simple model reduction scheme. In *DATE’2004 - Design, Automation and Test in Europe, Exhibition and Conference*, pages 938–943, Paris, France, February 2004.
8. Bernard N. Sheehan. TICER: Realizable reduction of extracted RC circuits. In *International Conference on Computer Aided-Design*, pages 200–203, Santa Clara, CA, November 1999.
9. Peter Feldmann and Roland W. Freund. Reduced-order modeling of large linear subcircuits via a block Lanczos algorithm. In *32<sup>nd</sup> ACM/IEEE Design Automation Conference*, pages 474–479, San Francisco, CA, June 1995.
10. I. M. Elfadel and David. L. Ling. A block rational arnoldi algorithm for multipoint passive model-order reduction of multiport rlc networks. In *International Conference on Computer Aided-Design*, pages 66–71, San Jose, California, November 1997.
11. Eric Grimme. *Krylov Projection Methods for Model Reduction*. PhD thesis, Coordinated-Science Laboratory, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL, 1997.
12. Jing-Rebecca Li, F. Wang, and J. White. An efficient lyapunov equation-based approach for generating reduced-order models of interconnect. In *36<sup>th</sup> ACM/IEEE Design Automation Conference*, pages 1–6, New Orleans, Louisiana, June 1999.
13. Bruce Moore. Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction. *IEEE Transactions on Automatic Control*, AC-26(1):17–32, February 1981.
14. K. Glover. All optimal Hankel-norm approximations of linear multivariable systems and their  $l_\infty$  error bounds. *International Journal of Control*, 36:1115–1193, 1984.
15. I. M. Jaimoukha and E. M. Kasenally. Oblique projection methods for large scale model reduction. *SIAM J. Matrix Anal. Appl.*, 16:602–627, 1995.
16. Joel R. Phillips and L. Miguel Silveira. Poor Man’s TBR: A simple model reduction scheme. *submitted to the IEEE Transactions on CAD special issue on DATE Conference*, 2004.
17. R. W. Freund and P. Feldmann. The SyMPVL algorithm and its applications to interconnect simulation. In *Proceedings of the 1997 International Conference on Simulation of Semiconductor Processes and Devices*, pages 113–116. IEEE, 1997.
18. Luís Miguel Silveira, Jacob K. White, Horácio Neto, and Luís Vidigal. On Exponential Fitting for Circuit Simulation. *IEEE Transactions on Computer-Aided Design of Integrated Circuits (TCAD)*, 11(5):566–574, May 1992.