# Model Order Reduction for Efficient Modeling and Simulation of Interconnect Networks

*Min Ma*

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

September 2007

---

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

*Your file Votre référence*
*ISBN: 978-0-494-38610-1*
*Our file Notre référence*
*ISBN: 978-0-494-38610-1*

# Abstract

As operating frequency increases and device sizes shrink, the complexity of current state-of-the-art designs has increased dramatically. One of the main contributors to this complexity is high speed interconnects. At high frequencies, interconnects become dominant contributors to signal degradation, and their effects such as delays, reflections, and crosstalk must be accurately simulated. Time domain analysis of such structures is however very difficult because, at high frequencies, they must be modeled as distributed transmission lines which, after discretization, result in very large networks. In order to improve the simulation efficiency of such structures, model order reduction has been proposed in the literature. Conventional model order reduction methods based on Krylov subspace have a number of limitations in many practical simulation problems. This restricts their usefulness in general commercial simulators.

In this thesis, a number of new reduction techniques were developed in order to address the key shortcomings of current model order reduction methods. Specifically a new approach for handling macromodels with a very large number of ports was developed, a multi-level reduction and sprasification method was proposed for regular as well as parametric macromodels, and finally a new time domain reduction method was presented for the macromodeling of nonlinear parametric systems. Using these approaches, CPU speedups of 1 to 2 orders of magnitude were obtained.

## Résumé

À mesure que la fréquence d'opération augmente et la taille des dispositifs se rétrécit, la complexité de conceptions avancées à la pointe de la technologie a augmenté considérablement. Une des causes principales de cette complexité sont les interconnexions haute-vitesse. À hautes fréquences, ces interconnexions deviennent les causes dominantes de la dégradation de signal et leurs effets tels que les retards, les réflexions et la diaphonie doivent être simulés précisément. L'analyse du domaine temporel de telles structures est cependant trés difficile, car à hautes fréquences, elles doivent être modelées comme lignes de transmission distribuées, ce qui entraîne de trés grands réseaux aprés discrétisation. Afin d'améliorer l'efficacité de simulation de telles structures, quelques travaux proposent la réduction d'ordre de modèle. Les méthodes conventionnelles de réduction d'ordre de modéle basées sur le sous-espace de Krylov causent plusieurs limitations dans beaucoup des problèmes de simulation pratiques. Ceci limite ainsi leur utilité dans les simulateurs commerciaux généraux.

Dans cette thèse, un certain nombre de nouvelles techniques de réduction ont été développées afin d'adresser les imperfections principales des méthodes courantes de réduction d'ordre de modèle. Spécifiquement une nouvelle approche pour manipuler des macromodèles avec un très grand nombre de portes a été développée. De plus, une méthode de réduction et de sparcification à multiniveaux a été proposée pour macromodèles généralaux ainsi que paramétriques. Finalement une nouvelle méthode de réduction dans le domaine temporel a été présentée pour le macromodeling des systèmes paramétriques non-linéaires. En utilisant ces techniques, des améliorations de vitesse de CPU à l'ordre de 1 à 2 ordres de grandeur ont été obtenues.

# Acknowledgments

First and foremost, I would like to express my deepest gratitude to my thesis advisor Professor Roni Khazaka, for his guidance and support through the course of this research work. The topic and direction of this dissertation are direct products of his suggestions and ideas. I am also grateful for his patience and encouragement for all ups and downs over three years of my Ph.D. study.

I would like to thank my thesis committee members, Professor Ramesh Abhari, Professor Dennis Giannacopoulos, Professor Gordon Roberts, Professor Fabrice Labeau and Professor El Mostapha Aboulhamid for their valuable time and suggestions. I would also like to thank Professor Nicholas Rumin for his support and help during my graduate study at McGill University.

During my years at McGill University, I have been lucky enough to be associated with an excellent research group of former and current graduate students at the Microelectronics and Computer System (MACS) lab. I wish to extend my thanks to Asanee Suntives, Zikai Zhang, Dani Tannir, Mouna Safi-Harab, Arash Khajooeizadeh, Kasra Payandehjoo, Mourad Oulmane, Mingyang Zhang, Jane Yu, Jeffrey Ho, Patricia Lee, Guangran Zhu, Toma Roseanu and many others for broadening my horizon in many aspects of life and making my graduate study at this University a joyful experience. In particular, I would like to thank Simon Hong for translating the French abstract.

Financial support by the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged, as well as CAD software support by the Canadian Microelectronics Corporation.

Finally, I am indebted to my parents for their love and support over the years back in China and across the Ocean. I am especially thankful to my husband Hailong and my son Junsong for their unconditional and unyielding support and encouragement throughout the years. This accomplishment is theirs to share with me.

# Contents

# List of Figures

# List of Tables

# List of Acronyms

| | |
|---|---|
| AWE | Asymptotic Waveform Evaluation |
| CAD | Computer Aided Design |
| CFH | Complex Frequency Hopping |
| CPU | Central Processing Unit |
| EDA | Electronic Design Automation |
| LTI | Linear Time Invariant |
| KCL | Kirchhoff's Current Law |
| KVL | Kirchholff's Voltage Law |
| MEMS | Microelectromechanical System |
| MNA | Modified Nodal Analysis |
| MOR | Model Order Reduction |
| NORM | Nonlinear Model Order Reduction Method |
| ODE | Ordinary Differential Equation |
| PMR | Parametric Model Reduction |
| PRIMA | Passive Reduced Order Interconnect Macromodel Algorithm |
| RLC | Resistor, Inductor and Capacitor |
| SPICE | Simulation Program with Integrated Circuits Emphasis |
| SVD | Singular Value Decomposition |
| TBR | Truncated Balanced Realization |
| VLSI | Very Large Scale Integrated Circuit |

# Chapter 1

# Introduction

## 1.1 Background and Motivation

In recent years, the complexity of the state-of-the-art microsystems has increased dramatically. Device miniaturization and technology scaling have led to larger designs containing orders of magnitude more components, and have resulted in more complicated models for both active and passive elements. This rise in complexity has made electronic design automation (EDA) tools a more essential part of the design cycle, while at the same time pushing the capabilities of existing tools and computing resources to their limits. In fact, the performance of high end designs is increasingly limited by the capabilities of EDA tools rather than by what can be optimally achieved using the available technology. One example of increased complexity in the digital Very Large Scale Integrated (VLSI) circuits and mixed signal areas is the signal integrity issues stemming from the high speed interconnects and packages [1]–[6].

Due to the increasing operating frequencies coupled with smaller feature sizes, lower power consumption, and the use of mixed analog/digital circuits, interconnect effects such as delay, crosstalk, attenuation, dispersion, reflection, ringing have become prominent and are causing a significant degradation in signal quality. These high speed interconnect effects, if not detected at early design stages, would severely degrade the system performance. It is therefore imperative

for circuit designers to simulate interconnects as accurately and efficiently as possible and correct the signal integrity problems as early as possible in the design cycle. Interconnects are therefore a critical component of the design process and can be a bottleneck for system performance. Efficient simulation of interconnects is critical from a design perspective. However high-speed interconnects also present significant difficulties for simulation and optimization, and have also become a bottleneck from a simulation and design automation perspective as well.

The main difficulty in the simulation of interconnect networks stems from the fact that, at high frequencies, they must be considered as transmission lines, and modeled using the Telegrapher's equations, which are partial differential equations in space and time. These equations cannot be directly integrated in a Spice like time domain simulator which is based on nonlinear ordinary differential equations. Such nonlinear transient simulators including the nonlinear receivers and drivers are, however, a necessary part of the design process. In order to link the transmission lines to the rest of the circuit, a time domain model in the form of ordinary differential equations is necessary. Obtaining such a model requires some form of discretization of the partial differential equations [2], [7]–[13]. Approaches based on cutting the interconnects into small sections with each section small enough to be modeled using lumped resistors, capacitors, and inductors (RLCG) provide a brute force way for the discretization [2]. Other more efficient discretization methods have been proposed. They include compact difference based discretization [8], interpolation based discretization [10], and matrix rational approximation [12], [13]. Regardless of discretization methods used, the resulting model after discretization contains a large system of equations, which significantly increases the CPU cost.

Model Order Reduction (MOR) techniques were therefore proposed in the literature to address the problem of simulation of high speed interconnects [5], [14]–[20]. The goal of model order reduction is to find a reduced order macromodel, which is much smaller than the original system, but still captures the essential input/output behaviors of the original system. Once available, these reduced order macromodels can replace the original large circuits and thus greatly reduce computational cost in system level simulation and designs. A number of model order reduction

methods for high-speed interconnect networks have been proposed in the literature [14]–[25]. They can be roughly classified into two general categories. One category is based on direct moment-matching. Reducing the original system using a single Padé expansion is known as asymptotic waveform evaluation (AWE) [18], [21], while complex frequency hopping (CFH) extends the moment-matching to multiple expansions [16]. However these methods have an inherent ill-conditioned problem. Moreover, there is no guarantee that the reduced order model is passive. Passivity is an important property for interconnect networks. A non-passive model, even if it is stable, may result in a unstable system when connected to other passive networks [5]. To overcome these difficulties, another category of model order reduction techniques known as indirect moment-matching was proposed in the literature [19], [20], [22], [23], [26], [27]. The key feature of these methods is that the reduced basis can be computed using robust well-conditioned algorithms such as Arnoldi and Lanczos [28]. The reduced system can preserve a certain number of moments of the original system implicitly, thus making the reduced macromodel more accurate as the number of moments is increased. This provides an accuracy versus size trade off up to a relatively high order. In addition, after some modification, indirect methods based on congruence transformation were shown to be passive by construction [19], [22], [23]. Due to robust algorithms and the guarantee of the passivity, congruence transformation based model order reduction has become a standard approach for generating macromodels. However, they still suffer from two important limitations. The first is related to the number of ports. As the number of ports increases, the size of the macromodel grows rapidly. The second limitation is the result of properties of the Krylov subspace used in the reduction. Such a subspace is known to capture significantly more poles than what is necessary in order to conserve the responses of the original system [29], [30]. Another indirect model order reduction technique is based on truncated balanced realization [24], [25]. These methods provide a good global error bound and the stability and the passivity of the original system can be preserved for reduced models. However high computational cost associated with solving Lyapunov equations limits these approaches to small or medium size problems.

Traditional model order reduction methods work for specific circuits, and require a new reduction each time a circuit parameter is modified. This can be very inefficient in many practical cases such as when the optimization or parameter sweeping is performed which requires repeated simulation for different parameter values. In order to address this issue, parametric model order reduction (PMR) techniques were proposed in [31], [32] to reduce the order of large systems, simultaneously with respect to frequency as well as design parameters. These techniques avoid repeated generation of the macromodels for different parameters, thus making it an efficient simulation tool for performing optimization and analyzing designs. However currently existing parametric model reduction methods face two difficulties. One is that these methods are based on multi-dimensional moment-matching which produces a larger parametric reduced model than what is necessary. Another difficulty is that the reduced parametric model is dense. A dense model would significantly limit the efficiency of the simulation.

The model order reduction methods described above are frequency domain methods, which cannot directly apply to nonlinear systems. However interconnect circuits are generally nonlinear systems containing drivers and receivers. The general approach used so far would therefore require partitioning of a system into linear and nonlinear subsections and performing model order reduction on the linear parts. However there were attempts to extend the linear macromodeling techniques to the nonlinear circuits in time domain. For example, a nonlinear system could be approximated with a set of linear subsystems, then followed by linear reduction utilizing methods described above. One straightforward way for the approximation is based on polynomial (Taylor) expansion of the nonlinearity [33]–[37]. The main problem about this group of methods is that they generate macromodels valid only around the expansion point. This limits the application to only weakly nonlinear systems and/or small input signals. In order to address this problem, a trajectory piece-linear model order reduction approach was proposed in [38]. The key observation in trajectory model order reduction is that the nonlinearity is represented by collections of linearized systems at different expansion points instead of one expansion point in order to overcome the weakly nonlinear limitation. The final reduced model is then obtained by taking

a weighted combination of the resulting reduced order linearized models. However, finding a good weighting function is not an easy task for a broad class of nonlinear systems. Moreover, if the weighting function is not well chosen, the resulting macromodel would give poor results and even suffer from the instability [39]. A more practical and robust methodology is thus required for obtaining nonlinear macromodels.

## 1.2 Contributions of the Thesis

In this thesis, a number of advanced novel model order reduction methods have been developed that address the key shortcomings of the current model order reduction techniques, thus improving their CPU efficiency and extending their applicability to new applications. More specifically, the main contributions are listed as follows.

1. Model order reduction for systems with large number of ports (see Chapter 3): One of the difficulties with model order reduction methods is that the size of the reduced order model increases rapidly with the number of ports as shown in Fig. 1.1. In this thesis, a new method for the computation of the congruence transformation matrix has been proposed as well as a new parametric port formulation has been developed which allows us to embed the load parameters without any modification to the vector of unknowns in the modified nodal analysis formulation [40]–[43]. It has been shown that the block moments with respect to frequency as well as the block moments with respect to load parameters are conserved [44]. As demonstrated in the examples, the size of the proposed reduced macromodels is significantly less sensitive to the number of ports than those obtained from the traditional methods. This proposed method therefore significantly extends the range of applicability of model order reduction methods to systems with large number of ports when information about the types of loads on the ports is available.

2. Multi-Level Reduction (see Section 4.2): Traditional Krylov subspace based model order reduction techniques result in reduced order models which are much larger than necessary

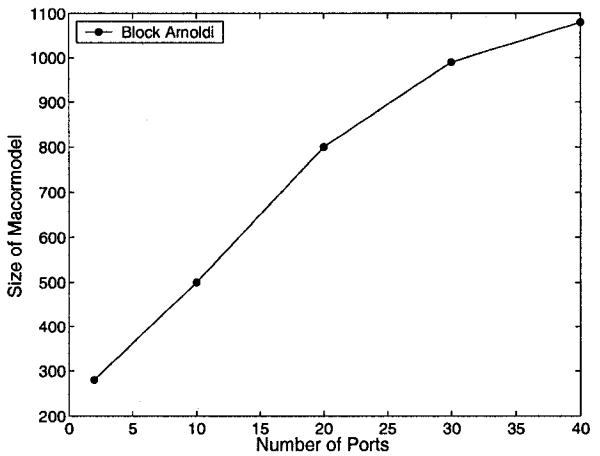


**Fig. 1.1** Size of the reduced macromodels obtained from traditional MOR versus the number of ports

and are thus not optimal. In this thesis, a multi-level reduction method with a second level of reduction based on singular value decomposition has been developed [45]–[47]. The new singular value decomposition based reduction can reduce by more than half the size of the macromodel obtained from the traditional model order reduction techniques. The major advantage of the proposed algorithm is that it is simple to implement, as compared with other methods such as Balanced Truncations.

3. Multi-level Parametric Reduction (see Section 4.3): Parametric model order reduction (PMR) methods were introduced for applications such as optimization and design space exploration. All these applications require repeated simulations at different values of certain design parameters. Parametric model order reduction techniques were therefore proposed in the literature to produce a macromodel which is valid over a range of parameter values [31], [32]. However, conventional parametric model order reduction methods are based on moment-matching and would result in a macromodel which contains many redundant poles. In this thesis, a multi-level reduction based on singular value decomposition has been proposed for parametric systems [48]. The macromodel obtained using the proposed approach is typically about one third the size of the macromodel obtained using

the traditional parametric model order reduction method.

4. Sparse parametric macromodel (see Section 4.4): The conventional parametric model order reduction techniques result in very dense reduced macromodels which significantly reduce the efficiency of the simulation. In this thesis, a new parametric formulation has been developed, which allows for the sparsification of the reduced parametric macromodel [49]. The reduction procedure and the sparsification are then performed on the macromodel in the space of the new formulation. This approach significantly improves the CPU efficiency of the simulation due to the sparsification.

5. Sparse Multi-level Parametric Reduction (see Section 4.5): In this thesis, a sparse multi-level parametric reduction method based on the sparsification and the multi-level reduction has been developed. The resulting macromodels are very small as well as sparse. They can achieve up to 350 times faster than the original systems to obtain the transient responses and therefore significantly improve the simulation efficiency.

6. Macromodeling of nonlinear networks (see Section 5.3 and Section 5.4): Traditional model order reduction methods are frequency domain techniques which are limited to the macromodeling of linear networks. In this thesis, a time domain nonlinear macromodeling technique, which is able to deal with arbitrary nonlinearity, has been developed [50] as well as a nonlinear formulation suitable for sparsification has been proposed [51]. This results in an efficient reduced order nonlinear macromodel which is sparse, and is valid over a predefined range of input waveforms and load conditions.

7. Macromodeling of parametric nonlinear networks (see Section 5.4.4): In this thesis, a new nonlinear model order reduction approach has been proposed for parametric nonlinear systems [52]. This new technique produces a sparse parametric nonlinear macromodel. This macromodel only needs to be created once, and can be reused many times over different input waveforms, different load conditions, as well as different internal circuit parameters. As demonstrated in the examples, the proposed nonlinear macromodel can achieve up to 40

times faster than the original nonlinear system to obtain the transient responses. This has many applications for interconnect networks containing nonlinear loads and internal circuit parameters, but also has the potential to provide an automatic and systematic approach to develop dynamic macromodels for nonlinear drivers.

## 1.3 Organization

The thesis is organized as follows. After the introduction in Chapter 1, the system formulation of circuit equations is outlined as well as existing model order reduction techniques for both linear and nonlinear systems are reviewed in Chapter 2. Chapter 3 presents a new model order reduction method to deal with a system with large number of ports by taking advantage of prior information on the ports. A new sparse multi-level parametric model order reduction technique is proposed in Chapter 4. In Chapter 5, a new sparse macromodeling technique for parametric nonlinear systems is proposed. Finally, a summary of current work and possible directions of future work is given in Chapter 6.

# Chapter 2

# Review of Model Order Reduction Techniques

As operation frequencies increase and device sizes shrink, the complexity of the current state-of-the-art designs has increased dramatically. One of the promising methods to deal with such increasing complexity is the use of model order reduction [5], [14]–[21], [23], [26], [29], [30], [53]–[62]. The general idea behind model order reduction is that, although the original networks may contain a large number of poles, only a fraction of those poles significantly contribute to the responses of the original network. We could therefore replace the original network with a reduced order macromodel, which captures the dominant poles and thus captures the essential characteristics of the original system. The focus of this thesis work is to develop advanced model order reduction methods with applications to the signal integrity analysis of modern microsystems. In this chapter, the main model order reduction methods for both linear and nonlinear systems currently available in the literature are presented in order to provide the necessary background as well as the motivation for the work.

This chapter is organized as follows. Section 2.1 presents the system formulation for interconnect circuits. Model order reduction based on direct methods is discussed in Section 2.2. Model order reduction based on indirect methods is described in Section 2.3. Parametric model order

reduction techniques are reviewed in Section 2.4, followed by discussing the truncated balanced realization in Section 2.5. Finally, model order reduction for nonlinear systems is reviewed in Section 2.6.

## 2.1 System Formulation

Before reviewing the currently existing model order reduction methods, it is useful to introduce the generic system formulation which is used throughout this thesis for representing circuit equations as well as macromodels for linear and nonlinear subsections. This formulation is a mathematical representation of the system that can be used for the analysis and simulation of the behavior of the system. For the electrical circuits containing resistors, capacitors, inductors, transmission lines as well as nonlinear elements such as drivers and receivers, the mathematical model is a set of nonlinear differential algebraic equations, which are obtained by using Kirchhoff's Current Law and Kirchholff's Voltage Law. The methodology for obtaining such circuit equations is known as Modified Nodal Analysis (MNA). It is important to note that in this context the transmission lines have been discretized into lumped sections using one of the discretization methods proposed in the literature [2], [7]–[13].

### 2.1.1 Modified nodal analysis formulation for circuits with lumped elements

Consider a lumped linear and nonlinear network, the modified nodal analysis (MNA) formulation of this network consists of nodal equations based on Kirchhoff's Current Law at each independent node, as well as additional equations to deal with voltage sources, inductors, and other special elements. The general MNA formulation can be written as [63], [64]

$$\boldsymbol{G}\boldsymbol{x}(t) + \boldsymbol{C}\dot{\boldsymbol{x}}(t) + \boldsymbol{f}(\boldsymbol{x}(t)) = \boldsymbol{b}(t) \tag{2.1}$$

1. $\boldsymbol{x} \in \mathbb{R}^n$ is a vector of node voltages appended by independent voltage source currents and linear inductor currents, nonlinear capacitor charges and nonlinear inductor fluxes.

2. $\boldsymbol{G} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{C} \in \mathbb{R}^{n\times n}$ are constant matrices containing the contributions of lumped memoryless and memory elements respectively.

3. $\boldsymbol{f}(\boldsymbol{x}(t)) \in \mathbb{R}^n$ is a vector containing algebraic functions describing the nonlinear elements of the circuit.

4. $\boldsymbol{b}(t) \in \mathbb{R}^n$ is a vector with entries determined by the independent voltage/current sources.

5. $n$ is the total number of the variables in the formulation.

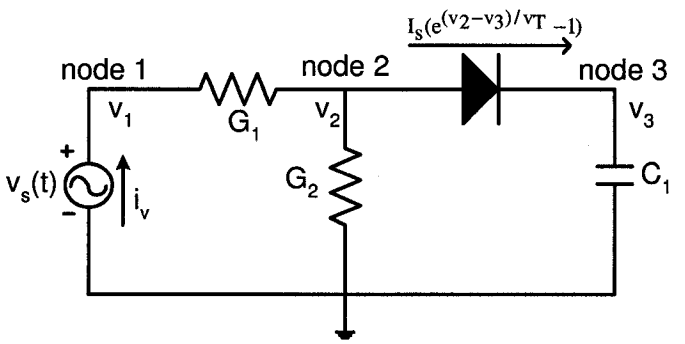


**Fig. 2.1** A simple nonlinear example circuit

The modified nodal analysis formulation in (2.1) is illustrated using a simple nonlinear example circuit shown in Fig. 2.1. By applying Kirchhoff's Current Law (KCL) at each node of the example circuit, we obtain

$$\begin{aligned} G_1(v_1 - v_2) - i_v &= 0 \\ G_1(v_2 - v_1) + G_2 v_2 + I_s\left(e^{\frac{v_2 - v_3}{v_T}} - 1\right) &= 0 \\ C_1\dot{v}_3 - I_s\left(e^{\frac{v_2 - v_3}{v_T}} - 1\right) &= 0 \\ v_1 &= v_s(t) \end{aligned} \tag{2.2}$$

Equation (2.2) can be recast in the matrix form (MNA formulation)

$$\begin{bmatrix} G_1 & -G_1 & 0 & -1 \\ -G_1 & G_1+G_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ i_v \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & C_1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{v}_1 \\ \dot{v}_2 \\ \dot{v}_3 \\ \dot{i}_v \end{bmatrix}$$

$$+ \begin{bmatrix} 0 \\ I_s\left(e^{\frac{v_2-v_3}{v_T}}-1\right) \\ -I_s\left(e^{\frac{v_2-v_3}{v_T}}-1\right) \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ v_s(t) \end{bmatrix} \tag{2.3}$$

where $I_s$ is the reverse bias saturation current of the diode and $v_T$ is the thermal voltage. Note that in addition to three nodal voltages $(v_1, v_2, v_3)$ as unknown variables and three nodal equations, one more unknown variable (voltage source current, $i_v$) and one more equation are added in the modified nodal analysis (MNA) formulation to deal with the independent voltage source. One of the main advantages of MNA formulation is that the resulting equations are sparse, which significantly increases the CPU efficiency of the simulation. The detailed information about modified nodal analysis formulation can be found in [65].

### 2.1.2 Modified nodal analysis formulation for linear subsections

In this section, we consider linear multi-port subsections which can contain lumped passive elements (resistors, inductors, and capacitors) as well as discretized transmission lines. The modified nodal analysis (MNA) formulation for this subsection, after discretization of the distributed elements, is written as [29], [63]

$$\boldsymbol{G}\boldsymbol{x}(t) + \boldsymbol{C}\dot{\boldsymbol{x}}(t) = \boldsymbol{R}\boldsymbol{v}(t) \tag{2.4a}$$

$$\boldsymbol{i} = \boldsymbol{R}^T\boldsymbol{x}(t) \tag{2.4b}$$

where $\boldsymbol{G} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{C} \in \mathbb{R}^{n\times n}$ are constant matrices obtained from lumped elements and discretization of the distributed elements. $\boldsymbol{R} \in \mathbb{R}^{n\times p}$ is a selector matrix that maps port currents to the node space of the subnetwork. $\boldsymbol{v}(t) \in \mathbb{R}^p$ is a vector containing the port voltages $v_1, v_2, \cdots, v_p$. $\boldsymbol{i}(t) \in \mathbb{R}^p$ is a vector containing the port currents $i_1, i_2, \cdots, i_p$. $n$ is the total number of the variables in the formulation. $p$ is the number of ports as shown in Fig. 2.2.

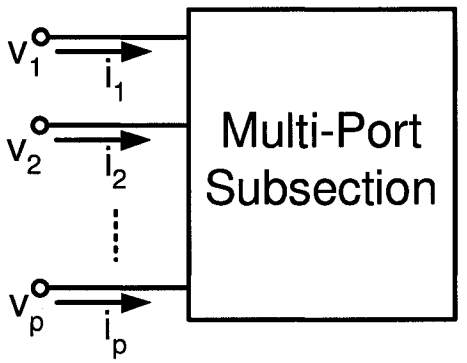


**Fig. 2.2** A subsection with $p$ ports

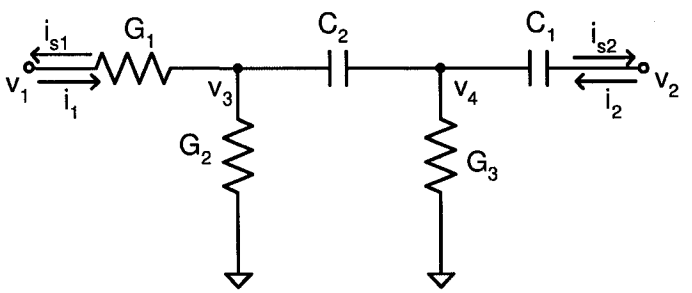


**Fig. 2.3** A two-port example circuit as a subsection

To illustrate the modified nodal analysis (MNA) formulation for a linear subsection, a simple two-port linear circuit is shown in Fig. 2.3. This two-port circuit is described as first order

differential equations in (2.4), where $\boldsymbol{G}, \boldsymbol{C}, \boldsymbol{x}(t), \boldsymbol{R}, \boldsymbol{v}(t), \boldsymbol{i}$ are in the form of

$$\boldsymbol{G} = \begin{bmatrix} G_1 & 0 & -G_1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -G_1 & 0 & G_1+G_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & G_3 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \end{bmatrix}; \boldsymbol{C} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & C_1 & 0 & -C_1 & 0 & 0 \\ 0 & 0 & C_2 & -C_2 & 0 & 0 \\ 0 & -C_1 & -C_2 & C_1+C_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{2.5}$$

$$\boldsymbol{x}(t) = \begin{bmatrix} v_1 & v_2 & v_3 & v_4 & i_{s1} & i_{s2} \end{bmatrix}^T \tag{2.6}$$

$$\boldsymbol{R} = \begin{bmatrix} 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}^T \tag{2.7}$$

$$\boldsymbol{v}(t) = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}; \quad \boldsymbol{i} = \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} \tag{2.8}$$

Note that the last two rows in the modified nodal analysis formulation are port equations, which are $v_1 = v_1$ and $v_2 = v_2$. In other words, the voltages across port1 and port2 are arbitrarily set by the boundary conditions. In general, for a $p$-port network, the last $p$ rows of the MNA formulation are port equations.

### 2.1.3 Modified nodal analysis formulation for nonlinear subsections

For a subsection containing nonlinear elements, such as nonlinear resistors, nonlinear capacitors, nonlinear inductors, and inverters, the modified nodal analysis (MNA)formulation can be written as

$$\boldsymbol{G}\boldsymbol{x}(t) + \boldsymbol{C}\dot{\boldsymbol{x}}(t) + \boldsymbol{f}(\boldsymbol{x}(t)) = \boldsymbol{R}\boldsymbol{v}(t) + \boldsymbol{b}(t) \tag{2.9a}$$

$$\boldsymbol{i} = \boldsymbol{L}^T\boldsymbol{x}(t) \tag{2.9b}$$

where $\boldsymbol{f}(\boldsymbol{x}(t)) \in \mathbb{R}^n$ is a vector containing algebraic functions describing the nonlinear elements of the circuit. $\boldsymbol{b}(t) \in \mathbb{R}^n$ is a vector containing the independent sources inside the subsection (e.g. the DC bias voltage source for an inverter). $\boldsymbol{R} \in \mathbb{R}^{n\times p}$ is a selector matrix that maps the port voltages into the node space of the subsection. $\boldsymbol{L} \in \mathbb{R}^{n\times p}$ is a selector matrix that maps the port currents into the node space of the subsection. $\boldsymbol{i}(t) = [\boldsymbol{i}_1, \cdots, \boldsymbol{i}_p]^t$ is a vector containing the port currents. This nonlinear subsection formulation is illustrated by a simple example shown in Fig. 2.4. This two-port subsection is described by first order differential equations in (2.9), where $\boldsymbol{G}$, $\boldsymbol{C}$, $\boldsymbol{x}(t)$, $\boldsymbol{f}(\boldsymbol{x}(t))$, $\boldsymbol{R}$, $\boldsymbol{v}(t)$ and $\boldsymbol{i}$ are as follows

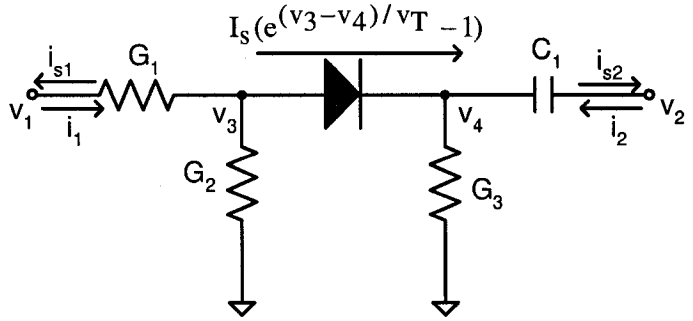


**Fig. 2.4** A two-port example nonlinear circuit as a subsection

$$\boldsymbol{G} = \begin{bmatrix} G_1 & 0 & -G_1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -G_1 & 0 & G_1+G_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & G_3 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \end{bmatrix}; \quad \boldsymbol{C} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & C_1 & 0 & -C_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -C_1 & 0 & C_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{2.10}$$

$$\boldsymbol{x}(t) = \begin{bmatrix} v_1 & v_2 & v_3 & v_4 & i_{s1} & i_{s2} \end{bmatrix}^T \tag{2.11}$$

$$\boldsymbol{R} = \begin{bmatrix} 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}^T \tag{2.12}$$

$$\boldsymbol{f}(\boldsymbol{x}(t)) = \begin{bmatrix} 0 & 0 & I_s\left(e^{\frac{v_3-v_4}{v_T}}-1\right) & -I_s\left(e^{\frac{v_3-v_4}{v_T}}-1\right) & 0 & 0 \end{bmatrix}^T \tag{2.13}$$

$$\boldsymbol{v}(t) = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}; \quad \boldsymbol{i} = \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} \tag{2.14}$$

Note that the last two rows of the modified nodal analysis formulation for the circuit in Fig. 2.4 are general port equations.

### 2.1.4 Modified nodal analysis stamp for linear and nonlinear subsections

In this section, we present the general stamps used to include linear and nonlinear subsections such as the ones defined in (2.4) and (2.9) in the overall modified nodal analysis formulation. Consider a circuit $\psi$ containing linear and nonlinear lumped elements, as well as $n_\psi$ subsections. The modified nodal analysis formulation for this overall circuit can be obtained by combining the lumped formulation with various subsection formulations as follows

$$\begin{bmatrix} \boldsymbol{G}_\psi & \boldsymbol{D}_1\boldsymbol{R}_1^T & \cdots & \boldsymbol{D}_{n_\psi}\boldsymbol{R}_{n_\psi}^T \\ -\boldsymbol{R}_1\boldsymbol{D}_1^T & \boldsymbol{G}_1 & \cdots & 0 \\ \vdots & 0 & \ddots & 0 \\ -\boldsymbol{R}_{n_\psi}\boldsymbol{D}_{n_\psi}^T & 0 & 0 & \boldsymbol{G}_{n_\psi} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_\psi(t) \\ \boldsymbol{x}_1(t) \\ \vdots \\ \boldsymbol{x}_{n_\psi}(t) \end{bmatrix} +$$

$$\begin{bmatrix} \boldsymbol{C}_\psi & 0 & \cdots & 0 \\ 0 & \boldsymbol{C}_1 & \cdots & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & \boldsymbol{C}_{n_\psi} \end{bmatrix} \begin{bmatrix} \dot{\boldsymbol{x}}_\psi(t) \\ \dot{\boldsymbol{x}}_1(t) \\ \vdots \\ \dot{\boldsymbol{x}}_{n_\psi}(t) \end{bmatrix} + \begin{bmatrix} \boldsymbol{f}_\psi(\boldsymbol{x}_\psi(t)) \\ \boldsymbol{f}_1(\boldsymbol{x}_1(t)) \\ \vdots \\ \boldsymbol{f}_{n_\psi}(\boldsymbol{x}_{n_\psi}(t)) \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}_\psi(t) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{2.15}$$

where $\boldsymbol{G}_\psi$, $\boldsymbol{C}_\psi$, $\boldsymbol{f}_\psi(\boldsymbol{x}_\psi(t))$, and $\boldsymbol{b}_\psi(t)$ are obtained from lumped elements as defined in (2.1). $\boldsymbol{G}_1, \cdots, \boldsymbol{G}_{n_\psi}$, $\boldsymbol{C}_1, \cdots, \boldsymbol{C}_{n_\psi}$, $\boldsymbol{f}_1(\boldsymbol{x}_1(t)), \cdots, \boldsymbol{f}_{n_\psi}(\boldsymbol{x}_{n_\psi}(t))$, $\boldsymbol{R}_1, \cdots, \boldsymbol{R}_{n_\psi}$ are obtained from $n_\psi$

subsections described in (2.9). $\boldsymbol{D}_1, \cdots, \boldsymbol{D}_{n_\psi}$ are selector matrices with maximum one nonzero in each row or column. These selector matrices map the port voltages or currents to the node space of the circuit $\psi$.

An example is given to illustrate the formulation in (2.15). Consider a circuit shown in Fig. 2.5, which contains one subsection. The port voltages and currents of this subsection are $v_{n1}$, $v_{n2}$ and $i_1$, $i_2$. This two-port subsection is represented by nonlinear first order differential equations as

$$\boldsymbol{G}_1\boldsymbol{x}_1(t) + \boldsymbol{C}_1\dot{\boldsymbol{x}}_1(t) + \boldsymbol{f}(\boldsymbol{x}_1(t)) = \boldsymbol{R}_1\boldsymbol{v}_1(t) \tag{2.16a}$$

$$\boldsymbol{i}_1 = \boldsymbol{R}_1^T\boldsymbol{x}_1(t) \tag{2.16b}$$

where the modified nodal analysis (MNA) matrices for the nonlinear subsection, $\boldsymbol{G}_1$, $\boldsymbol{C}_1$, $\boldsymbol{R}_1$, $\boldsymbol{x}_1$, $\boldsymbol{i}_1$ and $\boldsymbol{v}_1$ are described in (2.10), (2.11), (2.12), (2.13), and (2.14). The modified nodal analysis formulation for the overall circuit containing one nonlinear subsection as well as lumped elements can be written according to the unified formulation in (2.15).

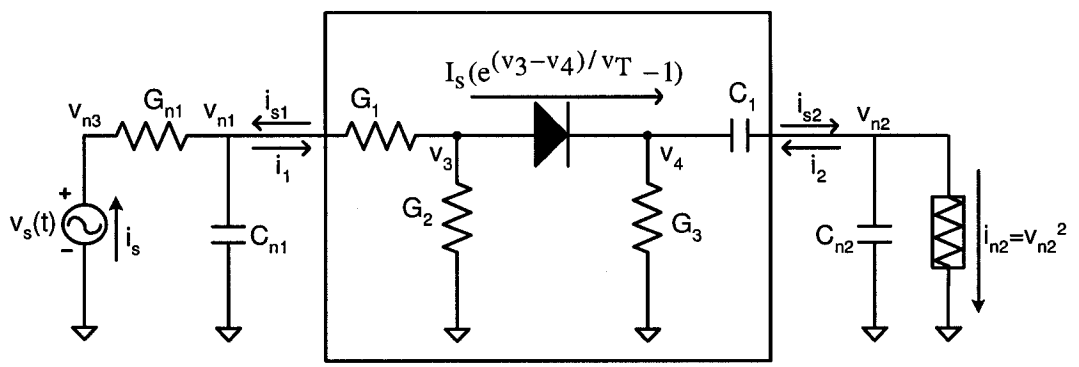


**Fig. 2.5** An example of circuit with one subsection

$$\begin{bmatrix} \boldsymbol{G}_\psi & \boldsymbol{D}_1\boldsymbol{R}_1^T \\ -\boldsymbol{R}_1\boldsymbol{D}_1^T & \boldsymbol{G}_1 \end{bmatrix} + \begin{bmatrix} \boldsymbol{C}_\psi & 0 \\ 0 & \boldsymbol{C}_1 \end{bmatrix} \begin{bmatrix} \dot{\boldsymbol{x}}_\psi(t) \\ \dot{\boldsymbol{x}}_1(t) \end{bmatrix} + \begin{bmatrix} \boldsymbol{f}_\psi(\boldsymbol{x}_\psi(t)) \\ \boldsymbol{f}_1(\boldsymbol{x}_1(t)) \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}_\psi(t) \\ 0 \end{bmatrix} \tag{2.17}$$

where $\boldsymbol{G}_\psi$, $\boldsymbol{C}_\psi$, $\boldsymbol{f}_\psi$, $\boldsymbol{b}_\psi(t)$, $\boldsymbol{x}_\psi$ are obtained from the lumped elements without the subsection as

$$\boldsymbol{G}_\psi = \begin{bmatrix} G_{n1} & 0 & -G_{n1} & 0 \\ 0 & 0 & 0 & 0 \\ -G_{n1} & 0 & G_{n1} & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}; \quad \boldsymbol{C}_\psi = \begin{bmatrix} C_{n1} & 0 & 0 & 0 \\ 0 & C_{n2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{2.18}$$

$$\boldsymbol{x}_\psi = \begin{bmatrix} v_{n1} & v_{n2} & v_{n3} & i_s \end{bmatrix}^T \tag{2.19}$$

$$\boldsymbol{f}_\psi(\boldsymbol{x}_\psi(t)) = \begin{bmatrix} 0 & v_{n2}^2 & 0 & 0 \end{bmatrix}^T \tag{2.20}$$

$$\boldsymbol{b}_\psi = \begin{bmatrix} 0 & 0 & 0 & v_s(t) \end{bmatrix}^T \tag{2.21}$$

$$\boldsymbol{D}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}^T \tag{2.22}$$

## 2.2 Model Order Reduction based on Direct Moment Matching

In the circuit simulation area, the initial attempts for model order reduction mainly focused on direct moment-matching methods, which are based on extracting dominant poles from a large network [14]–[18], [66]–[71]. To that end, the Taylor series coefficients also known as moments of the original system are first computed and then a direct moment-matching technique is used to obtain a low order rational approximation of the transfer function for the original system. Such rational approximation in the frequency domain can be used to find poles and residues of the system and obtain a time domain macromodel. This provides an efficient way to estimate the transient responses of a large system over a predefined range of frequency. Reducing the original system using a single Padé expansion is known as asymptotic waveform evaluation (AWE) [18], while complex frequency hopping (CFH) extends the moment-matching to multiple expansions [16].

### 2.2.1 Asymptotic waveform evaluation

Asymptotic waveform evaluation (AWE) provides a general approach for the approximation of the responses of linear RLC circuits. The basic idea of this approach is to reduce the original linear circuit with a large number of poles to a model with a small number of dominant poles. The dominant poles are the poles close to the imaginary axis and significantly influence the behavior of the system. Asymptotic waveform evaluation is based on single Taylor series expansion (typically at $s = 0$). It consists of two main steps, moment computation and moment matching.

To simplify the illustration of asymptotic waveform evaluation, consider a single input-single output system, the frequency domain expression is written as

$$\boldsymbol{G}\boldsymbol{x}(s) + s\boldsymbol{C}\boldsymbol{x}(s) = \boldsymbol{b}(s) \tag{2.23}$$

Taking Taylor series expansion of variables $\boldsymbol{x}(s)$ at $s = 0$ for the above system with the impulse input yields

$$\boldsymbol{G}(\boldsymbol{m}_0 + \boldsymbol{m}_1 s + \cdots + \boldsymbol{m}_n s^n) + s\boldsymbol{C}(\boldsymbol{m}_0 + \boldsymbol{m}_1 s + \cdots + \boldsymbol{m}_n s^n) = \boldsymbol{b} \tag{2.24}$$

where the Taylor coefficients of $\boldsymbol{m}_0, \boldsymbol{m}_1, \cdots, \boldsymbol{m}_n$ are known as moments, which can be calculated from equating the coefficients of equal power of $s$ in (2.24)

$$\boldsymbol{m}_0 = \boldsymbol{G}^{-1}\boldsymbol{b} \tag{2.25}$$

$$\boldsymbol{m}_i = -\boldsymbol{G}^{-1}\boldsymbol{C}\boldsymbol{m}_{i-1} \tag{2.26}$$

The $q$th moment $\boldsymbol{m}_q$ can therefore be related to the zeroth moment $\boldsymbol{m}_0$ as

$$\boldsymbol{m}_q = (-\boldsymbol{G}^{-1}\boldsymbol{C})^q \boldsymbol{m}_0 \tag{2.27}$$

Once the moments are computed, the next step is to find the impulse response $h(s)$ of the re-

duced order system, which is approximated by a Padé rational function with order $k/q$. Note that the moments of the impulse response correspond to a particular element in the vector $\boldsymbol{m}_i, i = 1, 2, \cdots$, which are denoted as $m_j, j = 1, 2, \cdots$. The impulse response of the system can therefore be expressed as

$$\begin{aligned} h(s) &= m_0 + m_1 s + m_2 s^2 + \cdots \\ &\approx \frac{a_0 + a_1 s + a_2 s^2 + \cdots + a_k s^k}{1 + b_1 s + b_2 s^2 + \cdots + b_q s^q} \end{aligned} \tag{2.28}$$

By cross multiplying equation (2.28) and equating the coefficients of equal power of $s$, the coefficients $a_0, a_1, \cdots, a_k, b_1, b_2, \cdots, b_q$ are obtained

$$\begin{bmatrix} m_{k-q+1} & m_{k-q+2} & \cdots & m_k \\ m_{k-q+2} & m_{k-q+3} & \cdots & m_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ m_k & m_{k+1} & \cdots & m_{k+q-1} \end{bmatrix} \begin{bmatrix} b_q \\ b_{q-1} \\ \vdots \\ b_1 \end{bmatrix} = - \begin{bmatrix} m_{k+1} \\ m_{k+2} \\ \vdots \\ m_{k+q} \end{bmatrix} \tag{2.29}$$

and

$$\begin{aligned} a_0 &= m_0 \\ a_1 &= m_1 + b_1 m_0 \\ a_2 &= m_2 + b_1 m_1 + b_2 m_0 \\ &\vdots \\ a_k &= m_k + \sum_{i=1}^{min(k,q)} b_i m_{k-i} \end{aligned} \tag{2.30}$$

Although low order rational functions can be obtained using asymptotic waveform evaluation, the accuracy of these methods deteriorates as we move away from the expansion point. In general, less than eight dominant poles can be extracted from asymptotic waveform evaluation. In addition, asymptotic waveform evaluation provides no estimation for error bounds [5]. To overcome

these difficulties, complex frequency hopping was introduced in the literature [16].

### 2.2.2 Complex frequency hopping

Due to the limitations of the single Padé approximation, Complex Frequency Hopping (CFH) was proposed in the literature to extend the moment-matching at multiple expansion points (hops) [16]. The expansion points are on or near the imaginary axis of the complex plane. A binary search algorithm is used in complex frequency hopping to find the expansion points and minimize the number of expansions. By doing this, enough information is obtained to generate the reduced system that can match the original system up to the predefined highest frequency. The comparison for the dominant poles extracted by CFH to those from AWE is illustrated in Fig. 2.6. In addition, CFH provides error criterion for the selection of accurate poles and transfer function. Detailed information on this method can be found in [16], [72]. Early CFH techniques were targeted at single input-single output systems. In [73], these techniques have been extended to multi-port systems known as block CFH.

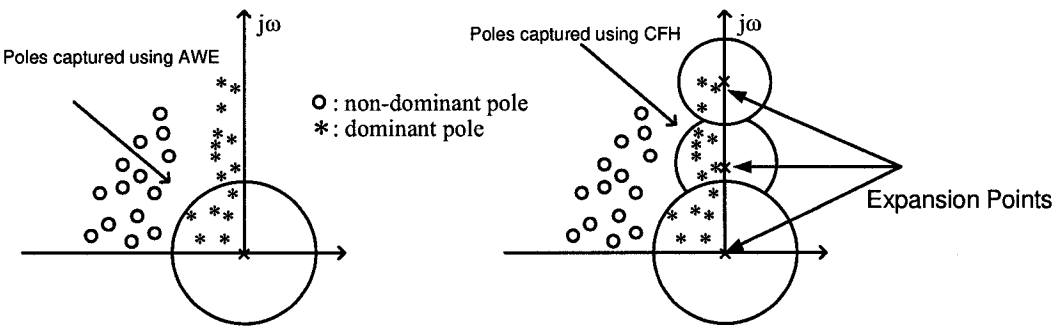


**Fig. 2.6** Poles extracting from CFH compare to those extracted from AWE

## 2.3 Model Order Reduction Based on Indirect Moment Matching

The direct moment-matching methods described in the previous section suffer from two main problems. First, there is no guarantee that the reduced model is passive. This is a significant

limitation because the passivity is an important property for interconnect networks. Violating this requirement would result in the instability of the simulation [5]. In addition, due to inherently ill-conditioned matrix in (2.29) associated with direct moment-matching, only less than eight dominant poles can be extracted from any expansion point. In order to address these difficulties, another class of model order reduction techniques known as indirect moment-matching techniques were developed in the literature [19], [20], [23], [26], [27], [55].

### 2.3.1 Moment matching based on Arnoldi

Consider a linear multi-port circuit described in subsection 2.1.2. In the Laplace domain, the MNA formulation (2.4) becomes

$$\boldsymbol{G}\boldsymbol{x}(s) + s\boldsymbol{C}\boldsymbol{x}(s) = \boldsymbol{R}\boldsymbol{V}(s) \tag{2.31a}$$

$$\boldsymbol{I}(s) = \boldsymbol{R}^T\boldsymbol{x}(s) \tag{2.31b}$$

where $\boldsymbol{G} \in \mathbb{R}^{n\times n}$, $\boldsymbol{C} \in \mathbb{R}^{n\times n}$, $\boldsymbol{R} \in \mathbb{R}^{n\times p}$, $\boldsymbol{x} \in \mathbb{R}^n$, $n$ is the number of unknown variables in the circuit. $p$ is the number of ports. Premultiplying (2.31a) by $\boldsymbol{G}^{-1}$, we can obtain

$$s\boldsymbol{A}\boldsymbol{x}(s) = \boldsymbol{x}(s) - \boldsymbol{B}\boldsymbol{V}(s) \tag{2.32}$$

where

$$\boldsymbol{A} = -\boldsymbol{G}^{-1}\boldsymbol{C}; \quad \boldsymbol{B} = \boldsymbol{G}^{-1}\boldsymbol{R} \tag{2.33}$$

The reduced order macromodel can be found by changing the variables as

$$\boldsymbol{x} = \boldsymbol{Q}\hat{\boldsymbol{x}} \tag{2.34}$$

where $\boldsymbol{Q} \in \mathbb{R}^{n \times q}$ ($q \ll n$) is an orthogonal matrix, $\hat{\boldsymbol{x}} \in \mathbb{R}^q$ is the variable vector of the reduced system. Substituting (2.34) in (2.32) results in

$$s\boldsymbol{A}\boldsymbol{Q}\hat{\boldsymbol{x}}(s) = \boldsymbol{Q}\hat{\boldsymbol{x}}(s) - \boldsymbol{B}\boldsymbol{V}(s) \tag{2.35a}$$

$$\boldsymbol{I}(s) = \boldsymbol{R}^{\boldsymbol{T}}\boldsymbol{Q}\hat{\boldsymbol{x}}(s) \tag{2.35b}$$

Premultiplying both sides of (2.35a) with $\boldsymbol{Q}^{\boldsymbol{T}}$, we obtain

$$s\boldsymbol{Q}^{\boldsymbol{T}}\boldsymbol{A}\boldsymbol{Q}\hat{\boldsymbol{x}}(s) = \boldsymbol{Q}^{\boldsymbol{T}}\boldsymbol{Q}\hat{\boldsymbol{x}}(s) - \boldsymbol{Q}^{\boldsymbol{T}}\boldsymbol{B}\boldsymbol{V}(s) \tag{2.36a}$$

$$\boldsymbol{I}(s) = \boldsymbol{R}^{\boldsymbol{T}}\boldsymbol{Q}\hat{\boldsymbol{x}}(s) \tag{2.36b}$$

Using the orthogonal definition, which is

$$\boldsymbol{Q}^{\boldsymbol{T}}\boldsymbol{Q} = \boldsymbol{I} \tag{2.37}$$

The reduced system can therefore be written as

$$s\hat{\boldsymbol{A}}\hat{\boldsymbol{x}}(s) = \hat{\boldsymbol{x}}(s) - \hat{\boldsymbol{B}}\boldsymbol{V}(s) \tag{2.38a}$$

$$\boldsymbol{I}(s) = \hat{\boldsymbol{R}}^{T}\hat{\boldsymbol{x}}(s) \tag{2.38b}$$

and the transfer function of the reduced system is

$$\hat{\boldsymbol{Y}}(s) = \hat{\boldsymbol{R}}^{T}(\boldsymbol{I} - s\hat{\boldsymbol{A}})^{-1}\hat{\boldsymbol{B}} \tag{2.39}$$

where $\hat{\boldsymbol{A}} \in \mathbb{R}^{q \times q} = \boldsymbol{Q}^{T}\boldsymbol{A}\boldsymbol{Q}$ is an upper Hessenberg matrix, $\hat{\boldsymbol{B}} \in \mathbb{R}^{q \times p} = \boldsymbol{Q}^{T}\boldsymbol{B}$, $\hat{\boldsymbol{R}} \in \mathbb{R}^{q \times p} = \boldsymbol{Q}^{T}\boldsymbol{R}$, and $\hat{\boldsymbol{x}} \in \mathbb{R}^q$ is the variable vector of the reduced system. $q \ll n$, the reduced system is therefore much smaller than the original system. It has been shown in [74] that the eigenvalues of $\hat{\boldsymbol{A}}$ are a good approximation of the leading eigenvalues of $\boldsymbol{A}$. In other words, the transfer

function of the reduced system is a good approximation of the transfer function of the original system. The congruence transformation $\boldsymbol{Q}$ is defined to span the subspace formed by the system moments.

$$colsp[\boldsymbol{Q}] = colsp[\boldsymbol{K}] \tag{2.40}$$

where

$$\boldsymbol{K} = [\boldsymbol{M}_0, \boldsymbol{M}_1, \cdots, \boldsymbol{M}_q] \tag{2.41}$$

$$= [\boldsymbol{B}, \boldsymbol{AB}, \cdots, \boldsymbol{A}^q\boldsymbol{B}] \tag{2.42}$$

It follows from (2.40) that any vector that is a linear combination of the moments in $\boldsymbol{K}$ is also a linear combination of the columns in $\boldsymbol{Q}$. Note that the column space spanned by the moments is simply the Krylov subspace. Although the matrix $\boldsymbol{K}$ spans the Krylov subspace, it is very ill-conditioned as the power order $q$ increases. This is due to the fact that the higher order moments quickly converge to an eigenvector corresponding to the largest eigenvalue [5]. It is therefore difficult to explicitly generate the reduction subspace from $\boldsymbol{K}$. To overcome this difficulty, an orthogonal basis $\boldsymbol{Q}$ is implicitly constructed to span the Krylov subspace. This orthogonal matrix can be computed using the Arnoldi process such as modified Gram-Schimdt algorithm [74]. In contrast to direct moment-matching techniques, the reduction based on the Arnoldi process can produce high order subspace which is well conditioned and can therefore achieve good accuracy. Further improvement for obtaining $\boldsymbol{Q}$, such as double orthogolization [75], were also proposed in the literature. An important indicator for the accuracy of the reduced order model is the total number of moments it can preserve. It can be shown that if the transformation matrix $\boldsymbol{Q}$ spans the Krylov subspace of the $q$th order moment vector, the reduced system preserves the first $q$ moments of the original system [19]. Consequently, the larger the order $q$ is, the more accurate the reduced system is. However, a larger order $q$ results in a larger reduced macromodel.

### 2.3.2 Projection based passive Krylov methods

Since the original interconnect network is passive by nature, it is therefore necessary that the reduced order system should also be passive. By definition, a passive circuit is one that cannot generate more energy than it absorbs. Preserving the passivity of the reduced system is important because a non-passive system, even if it is stable, may result in a non-stable system when connected to other passive networks [5].

Although the size of the reduced system in (2.38) is much smaller than the original system, it does not preserve the passivity. In [19], a general technique for the passive reduction was proposed based on the Arnoldi algorithm. Instead of performing reduction to the matrix $\boldsymbol{G}^{-1}\boldsymbol{C}$, the algorithm in [19] applies reduction to the conductance matrix $\boldsymbol{G}$ and the susceptance matrix $\boldsymbol{C}$ separately. Referring to the original system in (2.31), the reduced system can be found by applying the change of variables $\boldsymbol{x} = \boldsymbol{Q}\hat{\boldsymbol{x}}$

$$\boldsymbol{G}\boldsymbol{Q}\hat{\boldsymbol{x}}(s) + s\boldsymbol{C}\boldsymbol{Q}\hat{\boldsymbol{x}}(s) = \boldsymbol{R}\boldsymbol{V}(s) \tag{2.43a}$$

$$\boldsymbol{I}(s) = \boldsymbol{R}^T\boldsymbol{Q}\hat{\boldsymbol{x}}(s) \tag{2.43b}$$

Premultiplying both side of (2.43a) by $\boldsymbol{Q}^T$, we have

$$\hat{\boldsymbol{G}}\hat{\boldsymbol{x}}(s) + s\hat{\boldsymbol{C}}\hat{\boldsymbol{x}} = \hat{\boldsymbol{R}}\boldsymbol{V}(s) \tag{2.44a}$$

$$\boldsymbol{I}(s) = \hat{\boldsymbol{R}^T}\hat{\boldsymbol{x}}(s) \tag{2.44b}$$

where

$$\hat{\boldsymbol{G}} = \boldsymbol{Q}^T\boldsymbol{G}\boldsymbol{Q}; \quad \hat{\boldsymbol{C}} = \boldsymbol{Q}^T\boldsymbol{C}\boldsymbol{Q}; \quad \hat{\boldsymbol{R}} = \boldsymbol{Q}^T\boldsymbol{R} \tag{2.45}$$

The transformation performed on the original system is known as congruence transformation. It can be proved that the reduced order model in (2.44) is passive provided that the congruence transformation $\boldsymbol{Q}$ is a real matrix and the conductance and susceptance matrices $\boldsymbol{G}$ and $\boldsymbol{C}$ are

symmetric nonnegative definite [19]. The modified nodal analysis formulation for RLC interconnect networks can be formulated such that the above conditions are satisfied. For example, matrix $C$ is symmetric nonnegative, matrix $G$ can also be symmetric nonnegative definite after slight modification as the form of $G = \begin{bmatrix} N & E \\ -E^T & 0 \end{bmatrix}^T$, where $N$ is symmetric nonnegative definite. It is important to note that although $Q$ is obtained from the Arnoldi process, the passive reduction technique proposed in [19] is not limited to a particular Arnoldi process. For example, one of the important methods to obtain $Q$ is based on singular value decomposition (SVD) as described in Chapter 4. Due to the passivity by construction and the significant flexibility in choosing $Q$, the algorithm in [19] has become prevalent over the past decade for the reduction of linear systems. It is also a key technique widely used in the nonlinear macromodeling approaches [25], [37], [38], [64], [76].

## 2.4 Parametric Model Order Reduction

The model order reduction methods reviewed in the previous section work for specific circuits, and require a new reduction each time a circuit parameter is modified. This can be very inefficient in many practical cases such as when the optimization or parameter sweeping is performed which requires repeated simulation for different parameter values. In order to improve the efficiency, parametric model order reduction (PMR) was proposed in the literature [31], [32], [77].

Consider a large interconnect network consisting of many distributed elements as well as lumped components. After discretization, the parametrized modified nodal analysis circuit equations [31], [63], [77] can be written as:

$$\boldsymbol{G}\boldsymbol{x} + s\boldsymbol{C}\boldsymbol{x} + \left(\sum_{i=1}^{r} f_i(\lambda_i)\boldsymbol{D}_i\right)\boldsymbol{x} + s\left(\sum_{j=1}^{c} g_j(\phi_j)\boldsymbol{E}_j\right)\boldsymbol{x} = \boldsymbol{R}\boldsymbol{u}; \quad \boldsymbol{i} = \boldsymbol{R}^T\boldsymbol{x} \tag{2.46}$$

where:

1. $\boldsymbol{x}$ is a vector of node voltages appended by independent voltage source currents and linear inductor currents;

2. $\boldsymbol{G}$ is a matrix corresponding to the network's memoryless elements, excluding those that are memoryless parameters;

3. $\boldsymbol{C}$ is a matrix corresponding to the lumped memory elements of the network, also excluding the memory parameters;

4. $\boldsymbol{R}$ is a selector matrix that maps port voltages to the node space of the modified nodal analysis equations;

5. $\boldsymbol{u}$ is a vector that contains the port voltages;

6. $\boldsymbol{i}$ is a vector that contains the port currents;

7. $\boldsymbol{D}_1, ..., \boldsymbol{D}_r$ are matrices each containing the modified nodal analysis stamp of a particular memoryless element acting as a parameter;

8. $\boldsymbol{E}_1, ..., \boldsymbol{E}_c$ are matrices each containing the modified nodal analysis stamp of a particular memory element acting as a parameter;

9. $\lambda_1, ..., \lambda_r$ are input scalars that correspond to the values of the variable parameters represented by $\boldsymbol{D}_1, ..., \boldsymbol{D}_r$ respectively;

10. $\phi_1, ..., \phi_c$ are input scalars that correspond to the values of the variable parameters represented by $\boldsymbol{E}_1, ..., \boldsymbol{E}_c$ respectively.

The traditional parametric reduced order macromodel is obtained from (2.46) by using a congruence transformation which results in:

$$\hat{\boldsymbol{G}}\hat{\boldsymbol{x}} + s\hat{\boldsymbol{C}}\hat{\boldsymbol{x}} + \left(\sum_{i=1}^{r} f_i(\lambda_i)\hat{\boldsymbol{D}}_i\right)\hat{\boldsymbol{x}} + s\left(\sum_{j=1}^{c} g_j(\phi_j)\hat{\boldsymbol{E}}_j\right)\hat{\boldsymbol{x}} = \hat{\boldsymbol{R}}\boldsymbol{u}; \quad \hat{\boldsymbol{i}} = \hat{\boldsymbol{R}}^T\hat{\boldsymbol{x}} \tag{2.47}$$

where:

$$\hat{\boldsymbol{G}} = \boldsymbol{Q}^T\boldsymbol{G}\boldsymbol{Q}; \quad \hat{\boldsymbol{C}} = \boldsymbol{Q}^T\boldsymbol{C}\boldsymbol{Q}; \quad \hat{\boldsymbol{D}}_i = \boldsymbol{Q}^T\boldsymbol{D}_i\boldsymbol{Q}; \quad \hat{\boldsymbol{E}}_j = \boldsymbol{Q}^T\boldsymbol{E}_j\boldsymbol{Q}; \quad \hat{\boldsymbol{R}} = \boldsymbol{Q}^T\boldsymbol{R} \tag{2.48}$$

The congruence transformation matrix $\boldsymbol{Q}$ is chosen as an orthonormal basis of a subspace which spans the moments of the system with respect to frequency, and with respect to the parameters $\lambda_1, ..., \lambda_r, \phi_1, ..., \phi_c$, as well as the cross moments.

$$colsp\left[\boldsymbol{Q}\right] = colsp\left[\begin{array}{ccccccccc} \boldsymbol{M}_{s_k} & ... & \boldsymbol{M}_{\lambda_{ik}} & ... & \boldsymbol{M}_{\phi_{jk}} & ... & \boldsymbol{M}_{s_a\phi_{bd}} & ... & \boldsymbol{M}_{s_p\phi_{mn}} \quad ... \end{array}\right] \tag{2.49}$$

where $\boldsymbol{M}_{s_k}$ is the $k^{th}$ moment with respect to frequency, $\boldsymbol{M}_{\lambda_{ik}}$ is the $k^{th}$ moment with respect to the $i^{th}$ memoryless parameter, $\boldsymbol{M}_{\phi_{jk}}$ is the $k^{th}$ moment with respect to the $j^{th}$ memory parameter, $\boldsymbol{M}_{s_a\phi_{bd}}$ is the cross moment between the $a^{th}$ frequency moment and the $d^{th}$ moment of the $b^{th}$ memoryless parameter, $\boldsymbol{M}_{s_p\phi_{nm}}$ is the cross moment between the $p^{th}$ frequency moment and the $m^{th}$ moment of the $n^{th}$ memory parameter.

A combination of the Arnoldi process and standard QR decomposition is used to accurately compute $\boldsymbol{Q}$. It has been shown in [31] that the moments with respect to frequency and the moments with respect to parameters are conserved. In addition, the parametric reduced macromodel is passive by construction.

## 2.5 Model Order Reduction Using Truncated Balanced Realization

The concept of balanced truncation of a system was first introduced in the area of control theory to generate a reduced order model [61]. The general idea is to obtain a reduced order system by retaining the most controllable-observable states and truncating the rest. These states have more impact on the input and output behavior of the original system. Consider a classic state

formulation for a linear time invariant (LTI) system as

$$\dot{x} = Ax + Bu \tag{2.50a}$$

$$y = Cx + Du \tag{2.50b}$$

where $x$ is the state vector, $u$ and $y$ are vectors containing inputs and outputs of the system. The algorithm for standard truncated balance realization can be summarized as follows [24], [78].

1. Solve Lyapunov equation $AW_c + W_cA^T + BB^T = 0$ for the controllability grammian $W_c$.
2. Solve Lyapunov equation $A^TW_o + W_oA + C^TC = 0$ for the observability grammian $W_o$.
3. Since $W_c$ and $W_o$ are always symmetric and positive definite matrices, Cholesky factorization can be computed as $W_c = L_cL_c^T$, $W_o = L_oL_o^T$ for $L_c$ and $L_o$, where $L_c$ and $L_o$ are known as square roots of the grammians $W_c$ and $W_o$ respectively;
4. Compute singular value decomposition (SVD) of the product $U\sum V^T = L_o^TL_c$, $\sum$ is a diagonal matrix, with singular values in decreasing order. These singular values are called Hankel singular values of the system;
5. Compute the balanced transformation matrix $T = L_cV(\sum)^{-1/2}$, $T^{-1} = (\sum)^{-1/2}U^TL_o^T$;
6. Compute the balanced realization as $\hat{A} = T^{-1}AT$, $\hat{B} = T^{-1}B$, $\hat{C} = CT$, $\hat{D} = D$.
7. Partition $\hat{A}$, $\hat{B}$, $\hat{C}$, $\hat{D}$ and truncate them to desired orders.

The reduced model obtained from the standard truncated balanced realization can be made passive by modifying step 1 and step 2 in the standard algorithm [24]. Specifically, instead of solving Lyapunov equations, Lur'e equations and its dual are used for the controllability grammian $W_c$ and the observability grammian $W_o$. Truncated balance realization (TBR) offers theoretically provable error bounds and also produces better global accuracy than projection-based

approaches. However comparing with Krylov projection based approaches whose computational cost is $O(nq^2)$ [5], truncated balanced realization is significantly more expensive to implement. The computational cost is $O(n^3)$ with two large Lyapunov equations to solve and one singular value decomposition to perform [24], where $n$ is the dimension of the original system, $q$ is the dimension of the reduced system, and $q \ll n$. This often excludes it from being used directly on large problem sizes and requires the employment of projection based methods to obtain an initial system reduction in a prior step.

## 2.6 Model Reduction for Nonlinear Networks

Compared to the model order reduction for linear networks, the problems of finding a reduced order macromodel for nonlinear networks have been less studied and explored. This is mainly due to the fact that it is significantly more difficult and complex to deal with nonlinear systems than linear counterparts. The primary challenge facing the computer aided design (CAD) community is to find an efficient macromodel while maintaining good model accuracy for arbitrary nonlinearity.

### 2.6.1 Nonlinear circuit reduction based on congruence transformation

The main objective of the reduction algorithm in [64] is to significantly reduce the size of the original nonlinear system in (2.1) by using congruence transformation. The resulting reduced nonlinear model is therefore relatively inexpensive to solve. In order to achieve this goal, reduction is performed by changing the variables

$$\boldsymbol{x} = \boldsymbol{Q}\hat{\boldsymbol{x}} \tag{2.51}$$

in (2.1), and then premultiplying by $\boldsymbol{Q}^T$ to result in

$$\hat{\boldsymbol{G}}\hat{\boldsymbol{x}}(t) + \dot{\hat{\boldsymbol{C}}}\hat{\boldsymbol{x}}(t) + \hat{\boldsymbol{f}}(\hat{\boldsymbol{x}}(t)) = \hat{\boldsymbol{b}}(t) \tag{2.52}$$

where

$$\hat{G} = Q^T G Q; \quad \hat{C} = Q^T C Q;$$
$$\hat{f}(\hat{x}) = Q^T f(Q\hat{x}); \quad \hat{b}(t) = Q^T b(t) \tag{2.53}$$

In (2.53), the congruence transformation matrix $Q \in \mathbb{R}^{n \times q}$ is an orthogonal basis in time domain for the subspace spanned by the first $q$ time domain derivatives of $x$. In order to obtain such derivatives, we take time domain Taylor series expansion of $x$

$$x(t) = \sum_{k=0} a_k (t - t_0)^k \tag{2.54}$$

where $a_k = x^{(k)}/k!, k = 1, 2, \cdots$, are normalized time domain derivatives evaluated at $t = t_0$, and the time domain expansion for of $f(x(t))$ and $b(t)$ is given by

$$f(x(t)) = \sum_{k=0} f_k (t - t_0)^k \tag{2.55}$$

and

$$b(t) = \sum_{k=0} b_k (t - t_0)^k \tag{2.56}$$

where $f_k$ and $b_k$ are normalized time domain derivatives of $f(x(t))$ and $b(t)$ evaluated at $t = t_0$. Substituting (2.54), (2.55) and (2.56) in (2.1), we have

$$G a_k + (k+1) C a_{k+1} + f_k = b_k \tag{2.57}$$

Assuming the initial condition $a_0$ is known, the time domain coefficients can be computed recursively using (2.57). The subspace, constructed by time derivatives, is given by

$$K = [a_0, a_1, a_2, \cdots, a_q] \tag{2.58}$$

The transformation matrix $\boldsymbol{Q}$ is chosen as the orthogonal basis of the time domain subspace in (2.58), which is

$$colsp(\boldsymbol{Q}) = colsp(\boldsymbol{K}) \tag{2.59}$$

This is achieved by performing orthogonal decomposition on $\boldsymbol{K}$

$$\boldsymbol{K} = \boldsymbol{QR} \tag{2.60}$$

It can be shown that the reduced order circuit in (2.52) preserves the first $q$ time derivatives of the original system [64]. It is to be noted that the reduced order circuit is accurate only near the expansion point, which is $t = t_0$. The accuracy deteriorates as we move away from the expansion point. In other words, the expansion point needs to be updated as time marches ahead. The derivatives $(\boldsymbol{a}_0, \boldsymbol{a}_1, \cdots, \boldsymbol{a}_q)$ are computed again for each expansion point. The reduced order nonlinear circuit is shown to be much smaller than the original circuit, thus resulting in significant CPU savings [64]. However this technique is basically a circuit reduction based approach. It does not produce a macromodel, which is valid over different input waveforms and output conditions. In other words, this reduction cannot be performed on a nonlinear subsection described in (2.9). In addition, the reduced matrices in (2.52) are typically dense. The simulation of corresponding dense systems is relatively more expensive compared to sparse ones. These two issues will be addressed in Chapter 5.

### 2.6.2 Nonlinear macromodeling techniques for weakly nonlinear circuits

Weakly nonlinear systems are often referred to systems whose nonlinear terms can be sufficiently characterized by polynomial series. Typical example circuits include mixers, op-amps, etc. Consider a multi-port nonlinear system described by the modified nodal analysis formulation

$$\boldsymbol{C}\dot{\boldsymbol{x}}(t) + \boldsymbol{f}(\boldsymbol{x}(t)) = \boldsymbol{Ru}(t) \tag{2.61a}$$

$$\boldsymbol{y}(t) = \boldsymbol{L}^T\boldsymbol{x}(t) \tag{2.61b}$$

where $\boldsymbol{x}(t) \in \mathbb{R}^n$ is the unknown vector. $\boldsymbol{u}(t) \in \mathbb{R}^n$ is the input to the system. $\boldsymbol{f}(\boldsymbol{x}(t))$ is a vector containing algebraic function describing the nonlinear elements of the network. $\boldsymbol{R}$ and $\boldsymbol{L}$ are select matrices that map the input and output to the node space of the system respectively. $n$ is the total number of variables in the formulation.

The early approach towards the weakly nonlinear macromodeling was proposed in [34], where the nonlinearity is represented by Taylor series expansion. For example, we can expand the nonlinear terms $\boldsymbol{f}(\boldsymbol{x}(t))$ in (2.61) in the following form

$$\boldsymbol{f}(\boldsymbol{x}(t)) = \boldsymbol{F}_1\boldsymbol{x}(t) + \boldsymbol{F}_2\boldsymbol{x}(t) \otimes \boldsymbol{x}(t) + \boldsymbol{F}_3\boldsymbol{x}(t) \otimes \boldsymbol{x}(t) \otimes \boldsymbol{x}(t) + \cdots \tag{2.62}$$

where $\boldsymbol{x}(t) \in \mathbb{R}^n$ is a small signal response around DC bias point. The symbol $\otimes$ is the Kronecker (tensor) product operator. $\boldsymbol{F}_i \in \mathbb{R}^{n \times n^i}$ is the $i$th order tensor and $\boldsymbol{F}_1$ is the Jacobian of $\boldsymbol{f}$ evaluated at the expansion point. In order to extend the linear model order reduction approaches to nonlinear systems, variational method is used to decompose the original nonlinear system into series of linear systems. This variational approach was widely used in nonlinear system theory to find various Volterra kernels [79]. Consider a nonlinear system with inputs of the form $\alpha\boldsymbol{u}(t)$

$$\boldsymbol{C}\dot{\boldsymbol{x}}(t) + \boldsymbol{f}(\boldsymbol{x}(t)) = \boldsymbol{R}(\alpha\boldsymbol{u}(t)) \tag{2.63a}$$

$$\boldsymbol{y}(t) = \boldsymbol{L}^T\boldsymbol{x}(t) \tag{2.63b}$$

where $\alpha$ is an arbitrary scalar. The system response is then a function of the parameter $\alpha$ and it can be expanded into power series as follows

$$\boldsymbol{x}(t) = \sum_{i=1}^{\infty} \alpha^i \boldsymbol{x}_i(t) \tag{2.64}$$

Substituting (2.64) in (2.63a) and replacing the nonlinear vector $\boldsymbol{f}(\boldsymbol{x}(t))$ with its Taylor series

expansion in (2.62), we can write

$$\boldsymbol{C}(\alpha\dot{\boldsymbol{x}}_1+\alpha^2\dot{\boldsymbol{x}}_2+\alpha^3\dot{\boldsymbol{x}}_3+\cdots)+\boldsymbol{F}_1(\alpha\boldsymbol{x}_1+\alpha^2\boldsymbol{x}_2+\cdots)+\boldsymbol{F}_2(\alpha^2\boldsymbol{x}_1\otimes\boldsymbol{x}_1+\cdots) = \boldsymbol{R}(\alpha\boldsymbol{u}(t)) \quad (2.65)$$

Equating the coefficients of powers of $\alpha$, we can obtain the responses for different order system. For example, the first to the third order responses, $\boldsymbol{x}_1(t)$, $\boldsymbol{x}_2(t)$ and $\boldsymbol{x}_3(t)$ are given by

$$\boldsymbol{C}\dot{\boldsymbol{x}}_1 + \boldsymbol{F}_1\boldsymbol{x}_1 = \boldsymbol{R}\boldsymbol{u} \quad (2.66a)$$

$$\boldsymbol{C}\dot{\boldsymbol{x}}_2 + \boldsymbol{F}_1\boldsymbol{x}_2 + \boldsymbol{F}_2(\boldsymbol{x}_1 \otimes \boldsymbol{x}_1) = 0 \quad (2.66b)$$

$$\boldsymbol{C}\dot{\boldsymbol{x}}_3 + \boldsymbol{F}_1\boldsymbol{x}_3 + \boldsymbol{F}_2(\boldsymbol{x}_1 \otimes \boldsymbol{x}_2 + \boldsymbol{x}_2 \otimes \boldsymbol{x}_1) + \boldsymbol{F}_3(\boldsymbol{x}_1 \otimes \boldsymbol{x}_1 \otimes \boldsymbol{x}_1) = 0 \quad (2.66c)$$

To simplify the notation, the time augments for the first order to the third order responses have been dropped in (2.65) and (2.66). Given the fact that systems in (2.66a), (2.66b) and (2.66c) are linear, we could therefore use Krylov subspace based techniques such as the algorithm in [19] to reduced them separately [34]. For example, the first order reduced system from (2.66a) is given by

$$\hat{\boldsymbol{C}}_1\dot{\hat{\boldsymbol{x}}}_1(t) + \hat{\boldsymbol{F}}_1\hat{\boldsymbol{x}}_1(t) = \hat{\boldsymbol{R}}\boldsymbol{u}(t) \quad (2.67)$$

where

$$\hat{\boldsymbol{C}}_1 = \boldsymbol{Q}_1^T\boldsymbol{C}\boldsymbol{Q}_1; \quad \hat{\boldsymbol{F}}_1 = \boldsymbol{Q}_1^T\boldsymbol{F}_1\boldsymbol{Q}_1; \quad \hat{\boldsymbol{R}} = \boldsymbol{Q}_1^T\boldsymbol{R} \quad (2.68)$$

$\boldsymbol{Q}_1 \in \mathbb{R}^{n\times q_1}$ is chosen as the orthonormal basis of the subspace spanned by the system moments of (2.66a). After solving $\hat{\boldsymbol{x}}_1$, the first order response is approximated by

$$\boldsymbol{x}_1(t) \approx \boldsymbol{Q}_1\hat{\boldsymbol{x}}_1(t) \quad (2.69)$$

Substituting (2.69) in the second order system in (2.66b), we have

$$\boldsymbol{C}\dot{\boldsymbol{x}}_2 + \boldsymbol{F}_1\boldsymbol{x}_2 = \boldsymbol{R}_2\boldsymbol{u}_2(t) \tag{2.70}$$

where $\boldsymbol{R}_2 = -\boldsymbol{F}_2(\boldsymbol{Q}_1 \otimes \boldsymbol{Q}_1) \in \mathbb{R}^{n\times q_1^2}$, $\boldsymbol{u}_2(t) = \hat{\boldsymbol{x}}_1(t) \otimes \hat{\boldsymbol{x}}_1(t) \in \mathbb{R}^{q_1^2}$. It follows that the system describing the second order response is also a linear system with $\boldsymbol{u}_2(t)$ as the input. Similarly, an orthogonal basis of the subspace spanned by the system moments of (2.70) is constructed to perform the reduction. The procedure for the reduction of higher order linearized systems follows analogously. The main difficulty associated with this method is the exponentially increasing size of the subspace basis as the order of the linearized systems grows. For example, if the first order linearized system in (2.66a) has been reduced into the size of $q_1$, *i.e.*, $\boldsymbol{Q}_1 \in \mathbb{R}^{n\times q_1}$. The reduction matrix $\boldsymbol{Q}_2$ for the second order system is implicitly constructed to span the block Krylov subspace which is defined by the block moments. Each block moment contains $q_1^2$ vectors. While for the third order system, the reduction subspace is constructed to span the block moments of the system, with each block moment containing $(q_1^3 + q_1q_2)$ vectors, where $q_2$ is the size of the second order reduced system [37].

In order to generate a more efficient reduced macromodel , it was proposed in [36] to construct a single projection basis based on separate subspaces $(\boldsymbol{Q}_1, \boldsymbol{Q}_2, \cdots)$. Since these subspaces are not linearly independent, singular value decomposition is therefore used to remove the redundant information.

$$\boldsymbol{Q} = svd[\boldsymbol{Q}_1, \boldsymbol{Q}_2, \cdots] \tag{2.71}$$

The final reduced nonlinear model is therefore given by a unit form

$$\hat{\boldsymbol{C}}\dot{\hat{\boldsymbol{x}}} + \hat{\boldsymbol{F}}_1\hat{\boldsymbol{x}} + \hat{\boldsymbol{F}}_2\hat{\boldsymbol{x}} \otimes \hat{\boldsymbol{x}} + \hat{\boldsymbol{F}}_3\hat{\boldsymbol{x}} \otimes \hat{\boldsymbol{x}} \otimes \hat{\boldsymbol{x}} + \cdots = \hat{\boldsymbol{R}}\boldsymbol{u}(t) \tag{2.72a}$$

$$\boldsymbol{y}(t) = \hat{\boldsymbol{L}}^T\hat{\boldsymbol{x}}(t) \tag{2.72b}$$

where $\hat{\boldsymbol{C}} = \boldsymbol{Q}^T\boldsymbol{C}\boldsymbol{Q}$, $\hat{\boldsymbol{F}}_1 = \boldsymbol{Q}^T\boldsymbol{F}_1\boldsymbol{Q}$, $\hat{\boldsymbol{F}}_2 = \boldsymbol{Q}^T\boldsymbol{F}_2(\boldsymbol{Q} \otimes \boldsymbol{Q})$, $\hat{\boldsymbol{F}}_3 = \boldsymbol{Q}^T\boldsymbol{F}_3(\boldsymbol{Q} \otimes \boldsymbol{Q} \otimes \boldsymbol{Q})$,

$\hat{R} = Q^T R, \hat{L} = Q^T L.$

The common difficulty for variational reduction methods is the large size of the reduced system due to the exponentially increasing size of the input vectors for higher order systems. In order to address this difficulty, an efficient nonlinear reduction scheme for weakly nonlinear systems known as NORM was proposed in [37]. Unlike methods in [34], [36], a general matrix-form nonlinear transfer function and associated nonlinear transfer function moments were developed in [37]. This development allows us to generate a minimum order of the Krylov projection basis by investigating the interdependence between the Krylov subspace and moments to be matched in a nonlinear context, and thus resulting in a compact reduced macromodel.

The nonlinear macromodeling techniques described so far are all based on Taylor series expansion of nonlinearities around a fixed state. They share the common problem that Taylor series expansion with two or three terms is only valid around the equilibrium point (expansion point). If the operation point of the original system varies significantly from the expansion point, the reduced order model would be very inaccurate, thus the reduction methods only work for weakly nonlinear systems and/or limited range of input signals. Furthermore, the exponentially increasing computational costs with the number of expansion terms included makes these methods only limited in practice to cubic expansion.

In addition to Taylor series based approaches, another closely related method for the weakly nonlinear macromodeling is based on bilinear forms of a nonlinear system. The detailed reduction procedure of this technique can be found in [35], [76]. Although the bilinear-based method results in smaller macromodels than the original nonlinear time-varying systems, it suffers from the same difficulties with Taylor series approaches.

### 2.6.3 Trajectory piecewise linear methods

A trajectory piecewise linear approach [38], [80], [81] was proposed to address strong nonlinearity issues. The key observation of this method is to represent a nonlinear system with a small set of linearized systems at different expansion points about the state trajectory, and then reduce each

set with Krylov subspace projection methods. The trajectory is the trace of the solutions $\boldsymbol{x}(t)$ to the nonlinear differential equations excited by input signals. The main steps of this algorithm are summarized as follows. First, given a nonlinear system, run simulations to find the solution $\boldsymbol{x}(t)$ with certain input signals $\boldsymbol{u}(t)$. This solution constructs a state trajectory of the original system. Then, the system is linearized around sampling points, which are properly picked up on the state trajectory. The nonlinear system is therefore decomposed into collections of linearized systems. Next, reduce the order of each of lineariztions using projection based methods. The final nonlinear macromodel is obtained by weighted combinations of all the reduced order linear models.

## 2.7 Limitations of Existing Projection-based Model Order Reduction Methods

In this section, main difficulties facing the existing moment-matching techniques based on Krylov subspace are discussed. These difficulties tend to increase the size of the reduced macromodel, thus severely limit the CPU efficiency of such macromodel. These problems are addressed in later chapters using advanced projection based model order reduction techniques.

### 2.7.1 Unnecessarily large reduced macromodel

The subspace for projection based reduction is obtained from the implicit moment-matching technique, which captures the dominant poles in a radius around the expansion point (for example $s = 0$). As can be seen from Fig. 2.7, in order to extract all the dominant poles of the system, the reduced system contains not only the dominant poles but many non-dominant poles as well. In other words, The passive reduction technique described in Section 2.3.2 would result in a macromodel whose size is much larger than what is necessary.

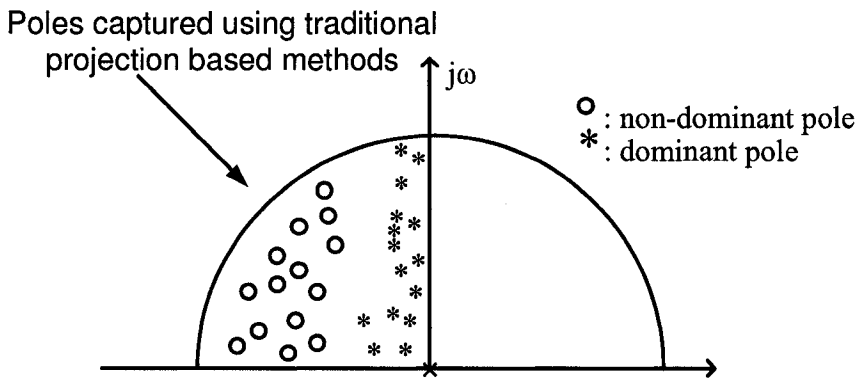


**Fig. 2.7** Poles captured using traditional Krylov subspace based reduction

### 2.7.2 Model order reduction for a system with large number of ports

When using Krylov subspace based methods to obtain the reduction subspace, the size of the reduced order macromodel grows rapidly as the number of ports increases. This is mainly due to the fact that the reduction matrix is implicitly constructed to span the Krylov subspace which is defined by the block moments as shown in Fig. 2.8, each block moment contains $p$ vectors, where $p$ is the number of ports. In order to illustrate this point, an example interconnect network was considered. Fig. 2.9 shows a typical circuit example when the size of the reduced macromodel increases as the number of ports is increased in the network. As can be seen from the figure, the size of the macromodel obtained using the standard Krylov subspace reduction process increases rapidly with the number of ports.

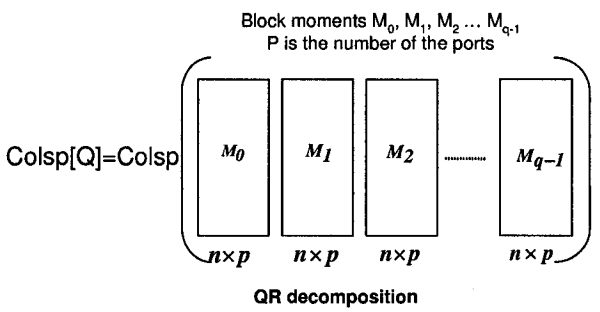


**Fig. 2.8** Reduction matrix spanning the same subspace as block moments

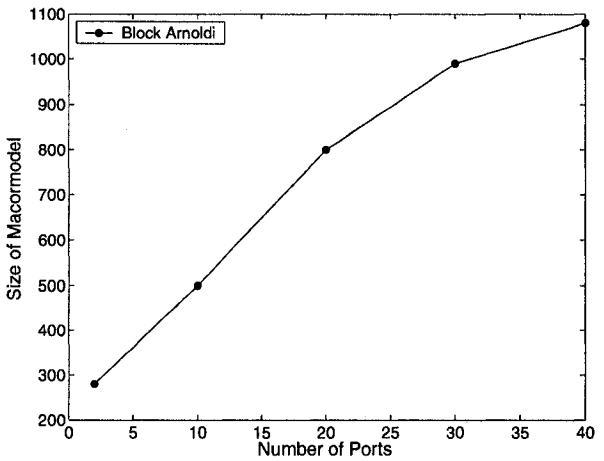


**Fig. 2.9** The size of the reduced macromodel obtained from the standard Krylov subspace method versus the number of ports

### 2.7.3 Dense parametric reduced model

Parametric reduction methods were introduced for application such as optimization and design space exploration. All these applications require repeated simulations at different parametric values. The parametric model order reduction techniques were therefore proposed in the literature to produce a macromodel which is valid over a range of parameter values [31], [32], [77]. The parametric techniques, however, result in very dense reduced macromodels which significantly reduce the efficiency of the simulation.

### 2.7.4 Nonlinear macromodeling technique

Projection based model order reduction using Krylov subspace methods are frequency domain methods, and cannot directly apply to nonlinear systems. In [64], the concept of projection based techniques has been extended to time domain, and the reduced nonlinear circuit has been shown to be much smaller than the original nonlinear circuit. However, this method is a circuit reduction based approach, and cannot be used for a nonlinear subsection.

There have been some nonlinear macromodeling techniques proposed in the literature recently for a nonlinear subsection [25], [33], [37], [38], [76]. These methods first linearize the

nonlinear system and then perform the reduction, which utilizes linear model order reduction methods. However, current nonlinear macromodeling methods face two main challenges. The first challenge for the nonlinear macromodeling is to capture arbitrary nonlinearities, both weak and strong nonlinearities. This is very important to a wide range of applications. Secondly, although the size of the reduced model can be much smaller than the original system, the nonlinear model order reduction would generally produce a dense model, which severely reduces the efficiency of the simulation.

# Chapter 3

# Model Order Reduction with Parametric Port Formulation

## 3.1 Introduction

As we have discussed in Section 2.7, one of the major difficulties associated with Krylov subspace based model order reduction methods is that the size of the reduced macromodel grows rapidly as the number of ports increases. In order to address this issue, a new method for the computation of the congruence transformation matrix, which is significantly less sensitive to the number of ports is proposed [40], [41]. The new approach is based on taking advantage of prior information regarding the port conditions. In order to achieve this goal, a systematic method for taking into account information about the possible loads that may be connected to the ports was developed. Specifically, resistive, reactive, resistive/reactive, transmission line loads as well as nonlinear loads are considered. A new parametric port formulation was developed which allows us to embed the load parameters without any modification to the vector of unknowns in the MNA formulation [63]. Parametric model order reduction method [77] is then used implicitly to obtain an efficient subspace $Q$ which takes into account the load parameters. Once the new real reduction subspace is computed, a congruence transformation on the original system is done using

the standard traditional approach [19] which results in a passive reduced model. Note that since the load conditions were taken into account, the reduction subspace is significantly smaller than those obtained from the block moment method. This provides a computationally efficient tool for the analysis of circuits with a large number of ports.

The rest of the chapter is organized as follows. Section 3.2 introduces the conventional model order reduction. Section 3.3 presents the proposed model order reduction with parametric port formulation followed by moments conservation proof in Section A. Finally, examples and results are given in Section 3.4.

## 3.2 Conventional Model Order Reduction

Consider a multi-port interconnect network containing $p$ ports. After discretization, the MNA [63] circuit equations can be written as:

$$\begin{aligned} (\boldsymbol{G} + s\boldsymbol{C})\boldsymbol{x} &= \boldsymbol{R}\boldsymbol{V} \\ \boldsymbol{I} &= \boldsymbol{R}^T\boldsymbol{x} \end{aligned} \tag{3.1}$$

where $s$ represents the complex frequency, $\boldsymbol{V}$ and $\boldsymbol{I}$ are vectors containing port voltages and currents. $\boldsymbol{G} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{C} \in \mathbb{R}^{n\times n}$ contain the contributions of the memoryless and memory elements respectively. $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$ refers to unknown node voltages and unknown currents. $n$ is the total number of variables in the MNA equations. $\boldsymbol{R} \in \mathbb{R}^{n\times p}$ is a selector matrix that maps the port voltages and currents into the node space of the circuit. The reduced order model is constructed through congruence transformation to obtain:

$$(\hat{\boldsymbol{G}} + s\hat{\boldsymbol{C}})\hat{\boldsymbol{x}} = \hat{\boldsymbol{R}}\boldsymbol{V} \tag{3.2}$$

where

$$\hat{\boldsymbol{G}} = \boldsymbol{Q}^T\boldsymbol{G}\boldsymbol{Q}; \quad \hat{\boldsymbol{C}} = \boldsymbol{Q}^T\boldsymbol{C}\boldsymbol{Q}; \quad \hat{\boldsymbol{R}} = \boldsymbol{Q}^T\boldsymbol{R} \tag{3.3}$$

In the case of traditional Krylov subspace methods, the matrix $\boldsymbol{Q} \in \mathbb{R}^{n \times N_q}$ is a real orthonormal basis of the subspace spanned by the block moments [19], [75].

$$\text{colsp}[\boldsymbol{Q}] = \text{colsp}[\boldsymbol{M}_0, \boldsymbol{M}_1, \boldsymbol{M}_2, \ldots, \boldsymbol{M}_{q-1}] \tag{3.4}$$

where $\boldsymbol{M}_0 \in \mathbb{R}^{n \times p}, \ldots, \boldsymbol{M}_{q-1} \in \mathbb{R}^{n \times p}$ are block moments, therefore the size of reduced matrix $\boldsymbol{Q}$ ($N_q = p \times q$) grows rapidly with the number of ports. Note that $\boldsymbol{Q}$ is typically implicitly computed using the Arnoldi process.

## 3.3 Model Order Reduction with Parametric Port Formulation

The method proposed in this section differs from traditional approaches in the way the congruence transformation matrix is computed. Using conventional Krylov techniques, the reduction matrix $\boldsymbol{Q}$ spans the column space of the block moments of all ports. This can lead to a large subspace when the number of ports is high. The subspace found in this section spans the block moments of the unconstrained ports only as well as the parametric moments with respect to the conductances and capacitances parameters for the constrained ports with resistive and reactive loads. This results in a significantly smaller subspace which is obtained by taking advantage of this parametric port information.

The proposed macromodel is obtained in three steps. The first two steps are mainly concerned with the computation of the subspace $\boldsymbol{Q}$. The last step is a congruence transformation based reduction step and is similar to the conventional reduction methods. More specifically, the first step consists of transforming the original multi-port MNA equations into a corresponding parametric model by taking the resistive and reactive loads on the port as parameters. This is outlined in Section 3.3.1, Section 3.3.2 and Section 3.3.3. The second step (outlined in Section 3.3.4) is to apply parametric model order reduction techniques to find a parametric reduction subspace. The final step is to obtain the reduced order model using a real congruence transformation on the original network as outlined in Section 3.3.5. Note that using such a real congruence transformation

matrix guarantees the passivity of the macromodel [19].

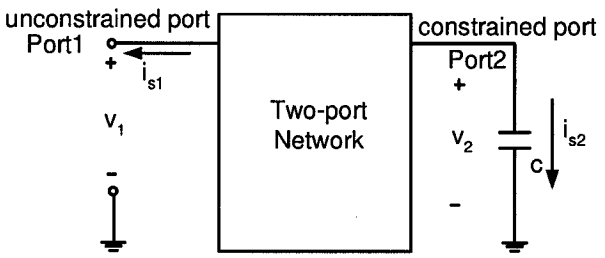


**Fig. 3.1** A large interconnect network with two ports

### 3.3.1 Parametric port model for reactive loads

For clarity of presentation, we will first consider a two-port system shown in Fig. 3.1, then the formulation for general multi-port networks will be presented. The ports are divided into unconstrained ports and constrained ports. A capacitor $c$ is connected to the constrained port2. The voltages on the unconstrained port (port1) and the constrained port (port2) in Fig. 3.1 are $v_1$ and $v_2$ respectively, and the currents for port1 and port2 are $i_{s1}$ and $i_{s2}$ respectively. If we examine the MNA equations of such a system as defined in (3.1), we note that the last two rows of $\boldsymbol{G}$ matrix, corresponding to port1 and port2 equations, in (3.1) are:

$$\boldsymbol{G} = \begin{bmatrix} \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & \cdots & 0 & 0 \end{bmatrix} \tag{3.5}$$

and

$$\boldsymbol{x} = \begin{bmatrix} v_1 \\ v_2 \\ \cdots \\ i_{s1} \\ i_{s2} \end{bmatrix}, \boldsymbol{R} = \begin{bmatrix} 0 & 0 \\ \cdots & \cdots \\ -1 & 0 \\ 0 & -1 \end{bmatrix}, \boldsymbol{V} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \tag{3.6}$$

Using this general port representation, the port equations in the last two rows are simply $v_1 = v_1$ and $v_2 = v_2$. In other words, the port voltage can be arbitrarily set by the boundary conditions. Note that port1 is designated as an unconstrained port and port2 is designated as a constrained port with reactive loading. In this case the equation at port1 remains unchanged but the equation at port2 becomes:

$$i_{s2} - scv_2 = 0 \tag{3.7}$$

Incorporating equation (3.7) into the MNA stamp has the effect of converting the multi-port network in (3.1) into a parametric "single-port" network while keeping the vector of unknowns $\boldsymbol{x}$ unchanged. This is done by modifying the row corresponding to port2 equation in the MNA formulation in (3.1) as follows:

$$\boldsymbol{G}_\phi \boldsymbol{x} + s\boldsymbol{C}\boldsymbol{x} + sc\boldsymbol{r}_2\boldsymbol{l}_2^T\boldsymbol{x} + \boldsymbol{r}_2\boldsymbol{r}_2^T\boldsymbol{x} = \boldsymbol{r}_1 v_1 \tag{3.8}$$

where $\boldsymbol{G}_\phi$ is obtained by setting the last row of $\boldsymbol{G}$ matrix in (3.5) to all zeros. $\boldsymbol{r}_1$ is the first column of $\boldsymbol{R}$. $\boldsymbol{r}_2$ is the second column of $\boldsymbol{R}$. $\boldsymbol{l}_2$ is a selector vector containing only one non-zero, corresponding to $v_2$, the node voltage of port2. Therefore the resulting elements in $\boldsymbol{r}_2\boldsymbol{l}_2^T$ are all zeros except for the last row and the column, corresponding to the node voltage of port2.

$$\boldsymbol{r}_2\boldsymbol{l}_2^T = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & -1 & \cdots & 0 & 0 \end{bmatrix} \tag{3.9}$$

The resulting elements in $\boldsymbol{r}_2\boldsymbol{r}_2^T$ are all zeros except for the last row and the last column, corresponding to $i_{s2}$, the port2 current.

$$\boldsymbol{r}_2\boldsymbol{r}_2^T = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \tag{3.10}$$

Note that there are two parameters in the parametric MNA equations in (3.8), the frequency parameter, $s$, and the parameter, $c$, corresponding to the reactive load.

To extend the above method into $M$ unconstrained ports and $N$ constrained ports networks, $(M+N=p)$, the parametric formulation becomes:

$$\boldsymbol{G}_{\phi}\boldsymbol{x} + s\boldsymbol{C}\boldsymbol{x} + s\left(\sum_{i=1}^{N} c_i \boldsymbol{r}_{M+i}\boldsymbol{l}_{M+i}^T\right)\boldsymbol{x} + \left(\sum_{i=1}^{N} \boldsymbol{r}_{M+i}\boldsymbol{r}_{M+i}^T\right)\boldsymbol{x} = \boldsymbol{R}_M\boldsymbol{u}_M \tag{3.11}$$

where $\boldsymbol{G}_{\phi}$ is obtained by setting $N$ rows of $\boldsymbol{G}$ in (3.1) into all zeros. $\boldsymbol{R}_M \in \mathbb{R}^{n\times M}$ is a selector matrix that maps currents and node voltages at the unconstrained ports into the node space of the circuit. $\boldsymbol{u}_M \in \mathbb{R}^{M\times 1}$ represents $M$ unconstrained voltage sources. $\boldsymbol{r}_{M+i}$ is a selector vector that maps the currents on the constrained port $(M+i)$ into the node space of the circuit. $\boldsymbol{l}_{M+i}$ is a selector vector mapping the node voltages on the port $(M+i)$ into the node space of the circuit. The parameters $c_i$ represent the reactive loads that can be connected to the constrained ports.

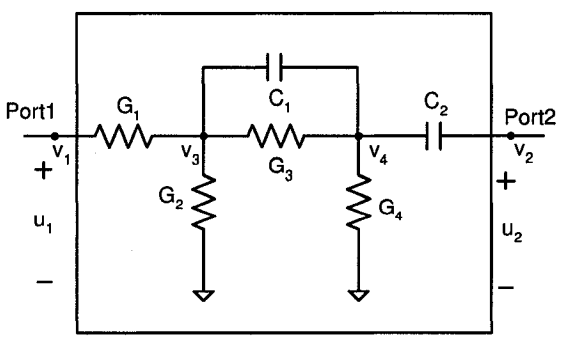


**Fig. 3.2** An example 2-port network

In order to simplify the notation, (3.11) is recast as follows:

$$\boldsymbol{G}_{\phi}\boldsymbol{x} + s\boldsymbol{C}\boldsymbol{x} + s\boldsymbol{E}_n\boldsymbol{R}_N\boldsymbol{L}_N^T\boldsymbol{x} + \boldsymbol{R}_N\boldsymbol{R}_N^T\boldsymbol{x} = \boldsymbol{R}_M\boldsymbol{u}_M \tag{3.12}$$

where $\boldsymbol{E}_n \in \mathbb{R}^{n\times n}$ is a diagonal matrix. The first $n-N$ diagonal elements are zero. While the last $N$ diagonal elements are $c_i$, $i = 1, \cdots, N$. $c_i$ represents the reactive loads that can be connected to the constrained ports. $\boldsymbol{L}_N \in \mathbb{R}^{n\times N}$ is a selector matrix with one nonzero in each column and a maximum of one nonzero in each row that maps voltages at the constrained ports into the node space $\mathbb{R}^{n\times 1}$ of the circuit. It is given by:

$$\boldsymbol{V}_N = \boldsymbol{L}_N^T\boldsymbol{x} \tag{3.13}$$

where $\boldsymbol{V}_N \in \mathbb{R}^{N\times 1}$ is a vector containing constrained port voltages. $\boldsymbol{R}_N \in \mathbb{R}^{n\times N}$ is a selector matrix with one nonzero in each column and a maximum of one nonzero in each row that maps currents at the constrained ports into the node space $\mathbb{R}^{n\times 1}$ of the circuit. It is given by:

$$\boldsymbol{I}_N = \boldsymbol{R}_N^T\boldsymbol{x} \tag{3.14}$$

where $\boldsymbol{I}_N \in \mathbb{R}^{N\times 1}$ is a vector containing constrained port currents. In order to illustrate the above formulation, we take a simple circuit shown in Fig. 3.2 as an example. The MNA equations for the simple circuit are given by:

$$\begin{bmatrix} G_1 & 0 & -G_1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -G_1 & 0 & G_1+G_2+G_3 & -G_3 & 0 & 0 \\ 0 & 0 & -G_3 & G_3+G_4 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ i_{s1} \\ i_{s2} \end{bmatrix}$$

$$+s\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & C_2 & 0 & -C_2 & 0 & 0 \\ 0 & 0 & C_1 & -C_1 & 0 & 0 \\ 0 & -C_2 & -C_1 & C_1+C_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ i_{s1} \\ i_{s2} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{3.15}$$

Assume port2 is the constrained port connecting to a capacitor $c$, and port1 is the unconstrained

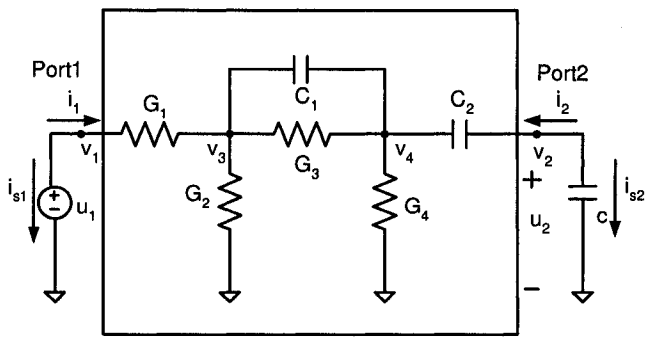


**Fig. 3.3** An example 2-port network with parametric port condition

port connecting to a voltage source $u_1$, which is shown in Fig. 3.3. By incorporating the port2 equation into the above MNA stamp, the parametric equations are as follows:

$$\begin{bmatrix} G_1 & 0 & -G_1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -G_1 & 0 & G_1+G_2+G_3 & -G_3 & 0 & 0 \\ 0 & 0 & -G_3 & G_3+G_4 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ i_{s1} \\ i_{s2} \end{bmatrix} +s\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & C_2 & 0 & -C_2 & 0 & 0 \\ 0 & 0 & C_1 & -C_1 & 0 & 0 \\ 0 & -C_2 & -C_1 & C_1+C_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ i_{s1} \\ i_{s2} \end{bmatrix}$$

$$+sc\begin{bmatrix}0&0&0&0&0&0\\0&0&0&0&0&0\\0&0&0&0&0&0\\0&0&0&0&0&0\\0&0&0&0&0&0\\0&0&0&0&0&1\end{bmatrix}\begin{bmatrix}0\\0\\0\\0\\0\\-1\end{bmatrix}\begin{bmatrix}0&1&0&0&0&0\end{bmatrix}\begin{bmatrix}v_1\\v_2\\v_3\\v_4\\i_{s1}\\i_{s2}\end{bmatrix}$$

$$+\begin{bmatrix}0\\0\\0\\0\\0\\-1\end{bmatrix}\begin{bmatrix}0&0&0&0&0&-1\end{bmatrix}\begin{bmatrix}v_1\\v_2\\v_3\\v_4\\i_{s1}\\i_{s2}\end{bmatrix}=\begin{bmatrix}0\\0\\0\\0\\-1\\0\end{bmatrix}\begin{bmatrix}u_1\end{bmatrix} \tag{3.16}$$

Also the constrained port voltage $\boldsymbol{V}_2$ is given by:

$$\boldsymbol{V}_2=\boldsymbol{v}_2=\begin{bmatrix}0&1&0&0&0&0\end{bmatrix}\begin{bmatrix}v_1\\v_2\\v_3\\v_4\\i_{s1}\\i_{s2}\end{bmatrix} \tag{3.17}$$

The constrained port current $\boldsymbol{I}_2$ is given by:

$$\boldsymbol{I}_2 = \boldsymbol{i}_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ i_{s1} \\ i_{s2} \end{bmatrix} \tag{3.18}$$

Note that the parametric equations in (3.16) are obtained by modifying the last row of the MNA equations in (3.15), corresponding to the port2 equation and changing the right hand side of the MNA equations into a vector indicating one input $u_1$. The other equations remain unchanged.

### 3.3.2 Parametric port model for resistive loads

If a resistor is connected to port2 in Fig. 3.1, then the port2 equation becomes:

$$i_{s2} - gv_2 = 0 \tag{3.19}$$

Here, we use conductance $g$ to represent the value of the resistive load in order to have the similar parametric MNA formula with that for reactive loads. Incorporating equation (3.19) into the MNA formulation in (3.1) by modifying the row corresponding to the port2 equation in the formulation as follows:

$$\boldsymbol{G}_\phi \boldsymbol{x} + s\boldsymbol{C}\boldsymbol{x} + g\boldsymbol{r}_2\boldsymbol{l}_2^T\boldsymbol{x} + \boldsymbol{r}_2\boldsymbol{r}_2^T\boldsymbol{x} = \boldsymbol{r}_1 v_1 \tag{3.20}$$

Note that the above formulation is similar to those in (3.8) except for the parameter $g$, which represents the conductance. As we did in the reactive loads, extending the two-port formulation into $M$ unconstrained ports and $N$ constrained ports networks, the above parametric formulation

becomes:

$$\boldsymbol{G_\phi x} + s\boldsymbol{C x} + \boldsymbol{D_n R_N L_N^T x} + \boldsymbol{R_N R_N^T x} = \boldsymbol{R_M u_M} \tag{3.21}$$

where $\boldsymbol{D_n} \in \mathbb{R}^{n\times n}$ is a diagonal matrix. The first $n-N$ diagonal elements are zero, while the last $N$ diagonal elements are $g_i, i = 1, \cdots, N$. $g_i$ represents the resistive loads that can be connected to the constrained ports.

### 3.3.3 Parametric port model for parallel reactive and resistive loads

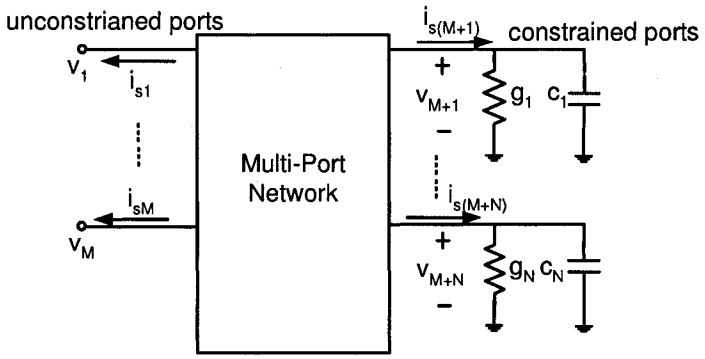


**Fig. 3.4** Multi-port network connected to combined reactive and resistive loads

If both a capacitor with capacitance $c$ and a resistor with conductance $g$ are connected to port2 in Fig. 3.1, the port2 equation becomes

$$i_{s2} - (g + sc)v_2 = 0 \tag{3.22}$$

Incorporating equation (3.22) into the MNA formulation in (3.1) would result in a parametric formulation based on the combination of (3.12) and (3.21) for a two-port network. Extending the two-port formulation into $M$ unconstrained ports and $N$ constrained ports networks as shown in Fig. 3.4, the parametric formulation is given by combining the formulations in (3.12) and (3.21):

$$\boldsymbol{G_\phi x} + s\boldsymbol{C x} + (s\boldsymbol{E_n} + \boldsymbol{D_n})\boldsymbol{R_N L_N^T x} + \boldsymbol{R_N R_N^T x} = \boldsymbol{R_M u_M} \tag{3.23}$$

The last $N$ diagonal elements of $\boldsymbol{E}_n$ and $\boldsymbol{D}_n$ are $c_i$ and $g_i$ respectively, which represent the reactive and resistive loads that can be connected to the constrained ports. Note that the parametric port formulation in (3.23) is not limited to a physical resistor or capacitor. It is valid for general resistive and reactive loads as will be shown in the next subsections. In other words, we do not make any assumption about the physical nature of the loads.

### 3.3.4 Reduction subspace

Using parametric model order reduction techniques established in [31], [77], the subspace $\boldsymbol{Q}_s$ spanning the moments with respect to frequency is implicitly computed using the Arnoldi process. The cross moments $(\boldsymbol{m}_{sc_i})_{k,j}$ with respect to frequency and the parameters $c_i$ can be computed by the procedure elaborated in [31], [82]. The corresponding subspace denoted by $\boldsymbol{Q}_{sc_i}$, $i = 1, \ldots N$, is obtained by performing a standard QR decomposition [74] on the cross moments. The subspace $\boldsymbol{Q}_{g_1}$, $\boldsymbol{Q}_{g_2}$, $\cdots$, $\boldsymbol{Q}_{g_N}$ spanning the moments with respect to conductance $g_1, g_2, \cdots, g_N$ is implicitly computed using the Arnoldi process.

Once all the required subspaces are evaluated, the resulting multidimensional subspace denoted by $\boldsymbol{Q}$ is as follows:

$$\text{colsp}[\boldsymbol{Q}] = \text{colsp}[\boldsymbol{Q}_s, \boldsymbol{Q}_{sc_1}, \boldsymbol{Q}_{sc_2}, \cdots, \boldsymbol{Q}_{sc_N}, \boldsymbol{Q}_{g_1}, \boldsymbol{Q}_{g_2}, \cdots, \boldsymbol{Q}_{g_N}] \tag{3.24}$$

This can be achieved by using another standard QR decomposition [74] on the various subspaces in (3.24).

### 3.3.5 Reduced order macromodel

Having the multidimensional subspace $\boldsymbol{Q}$, the reduced system is then produced by performing a congruence transformation on the original system in (3.1):

$$(\hat{\boldsymbol{G}} + s\hat{\boldsymbol{C}})\hat{\boldsymbol{x}}(s) = \hat{\boldsymbol{R}}\boldsymbol{V} \tag{3.25}$$

where

$$\hat{G} = Q^T G Q; \quad \hat{C} = Q^T C Q; \quad \hat{R} = Q^T R \tag{3.26}$$

Note that although the congruence matrix $Q$ is obtained from the system containing the information on the constrained ports, the model order reduction procedure is applied to the original multi-port system in (3.1) so that the ports can be connected with any value of capacitors and resistors. This is possible because the parametric modification in Section 3.3.1, Section 3.3.2 and Section 3.3.3 does not affect the unknown variables in the vector $x$ in (3.1). It can be demonstrated that the reduced system conserves the moments with respect to frequency as well as the load values. The proof of conservation of moments with respect to the load values $g_h$ and frequency is shown in the appendix. An experimental verification is done using a twenty-port network, nineteen of which are constrained ports. The size of the original network after discretization is $4890 \times 4890$. The 2-norm of relative errors for the moments of the reduced model and original system are below $1e-7$. Furthermore, since the proposed macromodel is obtained by applying real congruence transformation to the original system in (3.1), and the system formulation for the original system in (3.1) is in PRIMA compatible form [19], the resulting macromodel is passive by construction.

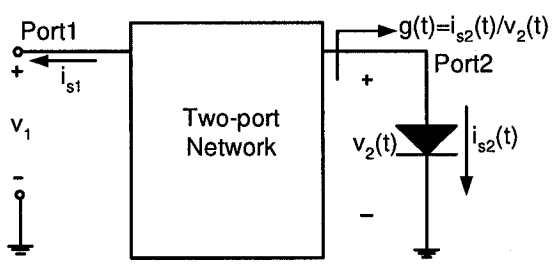


**Fig. 3.5** A diode is connected to port2

### 3.3.6 Model validity for general loads

As was discussed in the previous section, the parametric port formulation in (3.23) can be used to conduct a subspace for a reduced order macromodel which is valid for a certain range of loads.

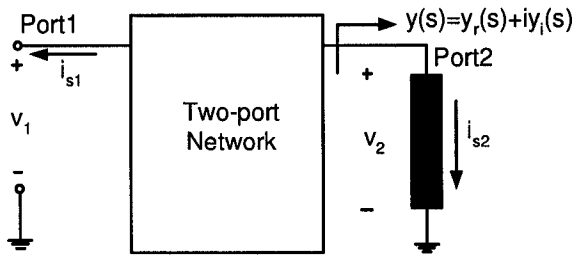


**Fig. 3.6** A transmission line is connected to the port2

Namely, the real and imaginary parts of the admittance seen by the port have to fall within a certain range. Note that no assumption was made about the physical nature of this load. The macromodel is therefore valid for linear loads consisting of lumped components as long as their overall admittance seen by the port fall in the validity range. But also, it is valid for other types of loads such as nonlinear loads and distributed loads as shown in Fig. 3.5 and Fig. 3.6. In the case of nonlinear loads, the admittance seen by the port is a function of time, and in the case of distributed transmission line loads, it is a function of frequency. However, as long as the values of this admittance are always bounded by the validity range of the macromodel, the macromodel will be valid for nonlinear and distributed loads as will be demonstrated in the examples.

## 3.4 Results

Four cases studies are presented. The reduced macromodel was obtained using the proposed method, and then was sparsified based on the standard diagonalization schemes [75]. The transient simulations were conducted using a spice-class simulator implemented in MATLAB.

### 3.4.1 Example 1

The first example is a 10-port network (nine of which are considered to be constrained ports) which we refer to as circuit1 as shown in Fig. 3.7. The circuit contains 18 non-coupled transmission lines and 1 nine-coupled transmission line system. After discretization of the interconnects, the resulting size of the MNA matrices was 4340 $\times$ 4340. Using the proposed approach, the

size of the reduced macromodel was 324 × 324. Using traditional block Arnoldi macromodeling approach, the size of the reduced order model was 460 × 460 in order to match the original system up to 4GHz. A summary of macromodel size comparisons for example 1 is shown in Table 3.1. The nominal load values (expansion point) for generating the reduced macromodel are conductance 0S (corresponds to open circuit) and capacitance $0pF$ at all constrained ports. This macromodel was tested to be valid for resistive capacitive loads at the ports when the conductance ranges from 0 to 1S (which corresponds to 1 Ω to open circuit), and capacitance from 0 to $90pF$. This range of validity depends on the number of moments matched, as well as the expansion points.

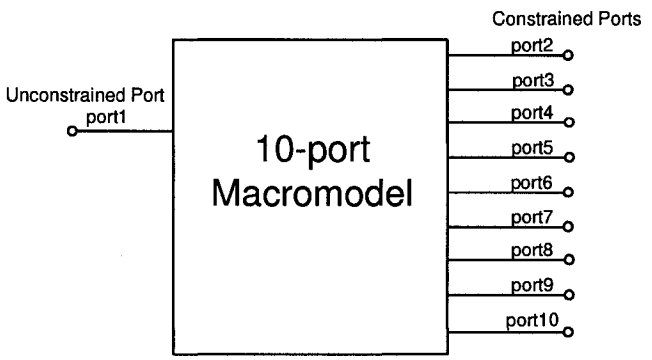


**Fig. 3.7** A 10-port interconnect network as circuit1.

The macromodel was tested by varying the value of one load resistor (or capacitor) while the other values of loads were kept constant. Fig. 3.8 shows a comparison of the frequency responses obtained from the proposed macromodel with the responses from the original system as the value of the resistor for constrained port2 varies from 1Ω to 20000Ω, while the value of each resistor connected to the other 8 constrained ports is 20Ω and the value of each capacitor at 9 constrained ports is $10pF$. Fig. 3.9 shows the frequency response comparisons as the value of the capacitor for constrained port2 varies from $0.0001pF$ to $90pF$, while the values of the capacitors for the other 8 constrained ports is $10pF$ and the values of the resistors are 20Ω. As can be seen from Fig. 3.8 and Fig. 3.9, they are all good approximation of the original network.

This macromodel was also tested with 100 different cases with resistive loads, capacitive

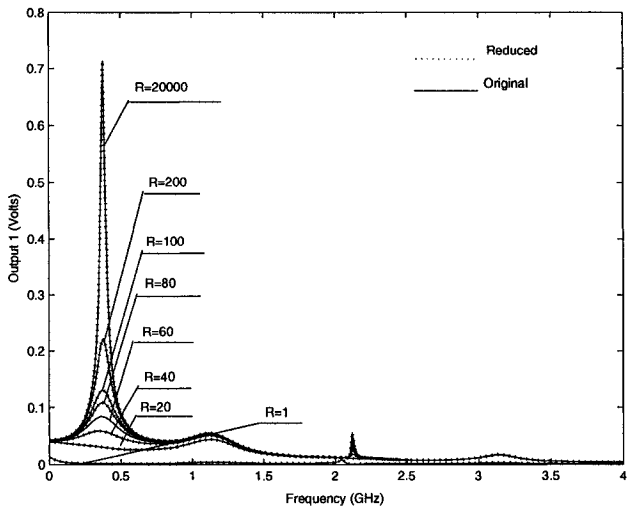


**Fig. 3.8** Frequency responses for the reduced macromodel with resistive/capacitive loads as the value of the resistor at the constrained ports varies from $1\Omega$ to $20000\Omega$, while the values of other resistor are $20\Omega$ and the values of capacitors are $10pF$

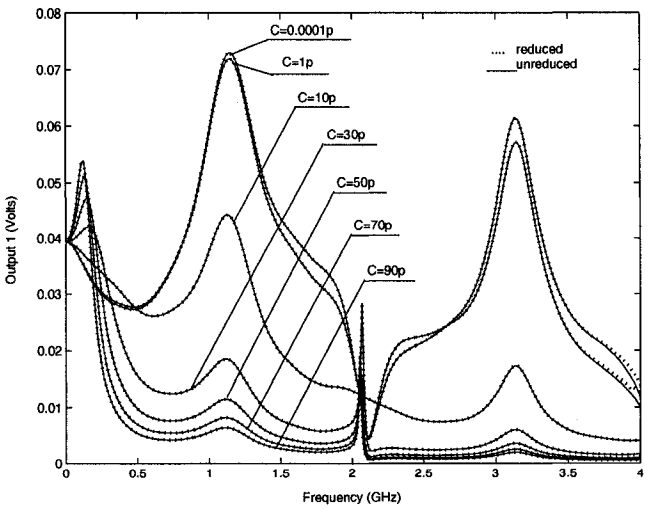


**Fig. 3.9** Frequency responses for the reduced macromodel with resistive/capacitive loads as the value of the capacitor at port2 varies form $0.0001pF$ to $90pF$, while the values of other capacitors are $10pF$ and the values of the resistors are $20\Omega$

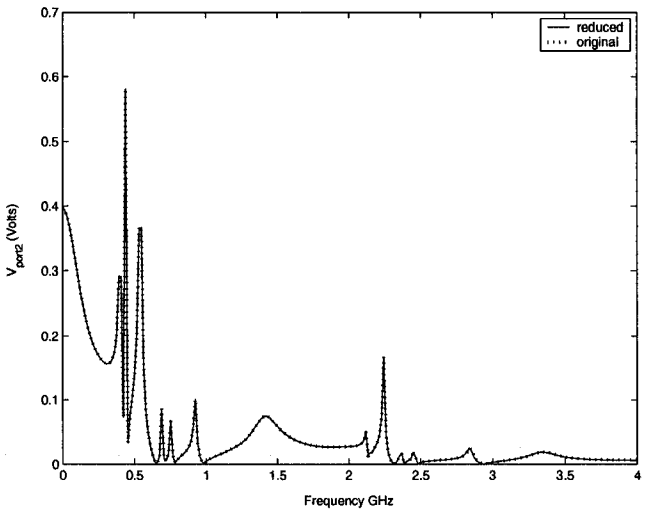


**Fig. 3.10** Frequency response of port2 for circuit1 with capacitive loads at the constrained ports (comparison between the original system and the proposed macromodel).

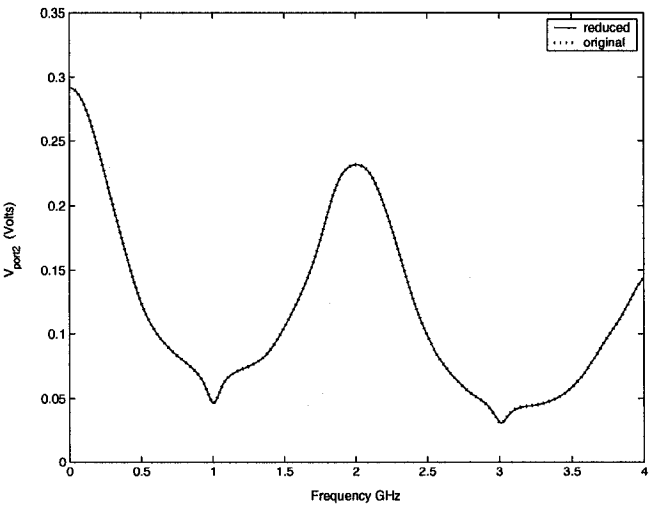


**Fig. 3.11** Frequency response of port2 for circuit1 with resistive loads at the constrained ports (comparison between the original system and the proposed macromodel).

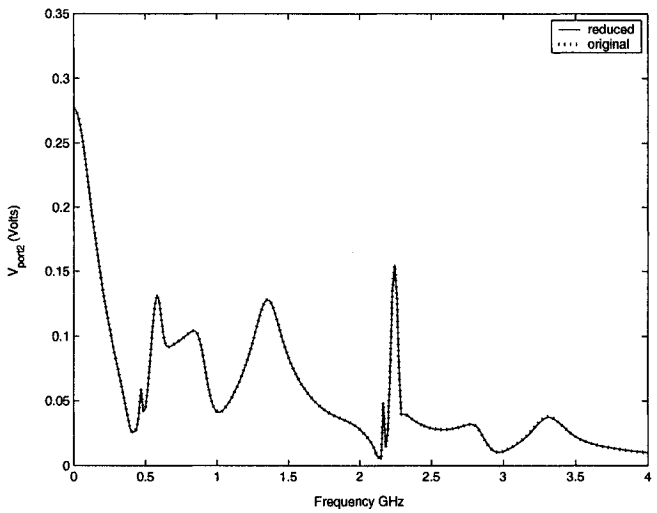


**Fig. 3.12** Frequency response of port2 for circuit1 with resistive/capacitive loads at the constrained ports (comparison between the original system and the proposed macromodel).

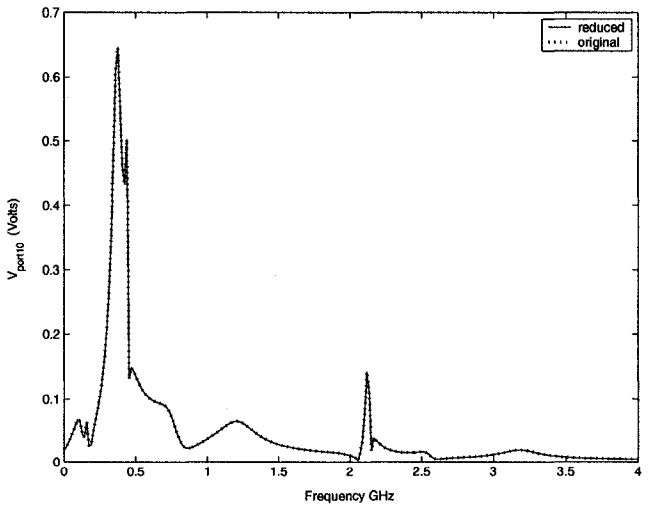


**Fig. 3.13** Frequency response of the constrained port10 for circuit1 with resistive/capacitive loads at the constrained ports (comparison between the original system and the proposed macromodel).

**Table 3.1** Macromodel size comparisons between the proposed method and block Arnoldi

| | Example 1 | Example 2 | Example 3 |
|---|---|---|---|
| Number of Ports | 10 | 20 | 30 |
| Original System | 4340 | 4890 | 4900 |
| Block Arnoldi | 460 | 800 | 990 |
| Proposed Method | 324 | 494 | 554 |

**Table 3.2** CPU comparison of transient analysis for circuit1

| | Size | Time | Speed-up |
|---|---|---|---|
| Reduced System | 324 | $28s$ | 27 |
| Original System | 4340 | $759s$ | - |

loads and parallel resistive and capacitive loads. The values of the capacitors are randomly chosen from 0.1pF to 90pF, and the values of resistors were randomly chosen from $1\Omega$ to $10000\Omega$, which fall within the range of validity of the macromodel indicated above. For 100 testing cases, the average relative error in the output frequency responses obtained from the proposed model compared to the original network is 0.38%. Fig. 3.10, Fig. 3.11, Fig. 3.12 and Fig. 3.13 show four sample frequency response results for different loads. As can be seen, they match well to the original network. The comparison of transient responses for the reduced macromodel and the original system using a 1-V pulse with a rise time of $0.1ns$ and a pulse width of $2ns$ is shown in Fig. 3.14. The CPU cost for the transient analysis is shown in Table 3.2, indicating a speed-up of 27 with respect to the solution obtained from the original circuit1.

This macromodel was also tested with transmission line loads and nonlinear loads as shown in Fig. 3.15. Fig. 3.16 shows the frequency response for the transmission line load obtained from the proposed macromodel compared with that from the original system. It matches the original system very well. The transient responses for the transmission line loads, diode and inverter are

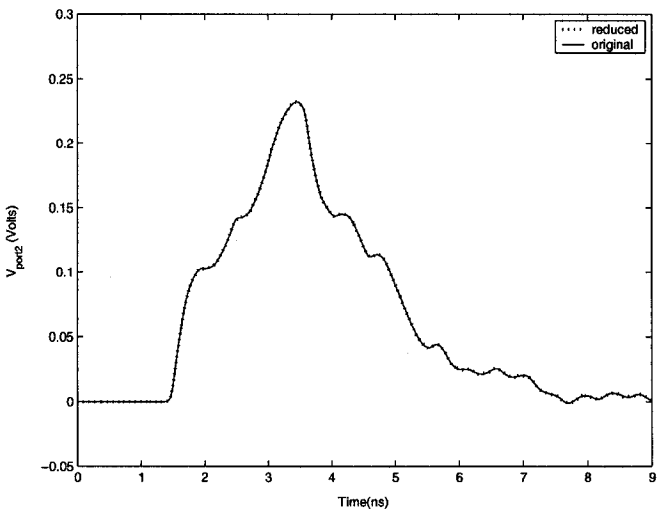


**Fig. 3.14** Transient results obtained from the proposed macromodel for resistive and capacitive loads and from the original system for circuit1.

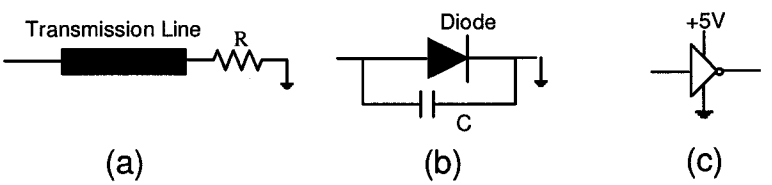


**Fig. 3.15** Transmission line, diode and inverter loads

shown in Fig. 3.17, Fig. 3.18 and Fig. 3.19 respectively. The transient responses are due to a pulse with a rise time of 500ps and a pulse width of 2ns. The amplitudes of the pulses are 1V, 5V and 15V respectively. As can be seen, they are accurate. Note that the reduced macromodel was generated once and was reused for performing frequency and transient responses in all the tests in Example 1.

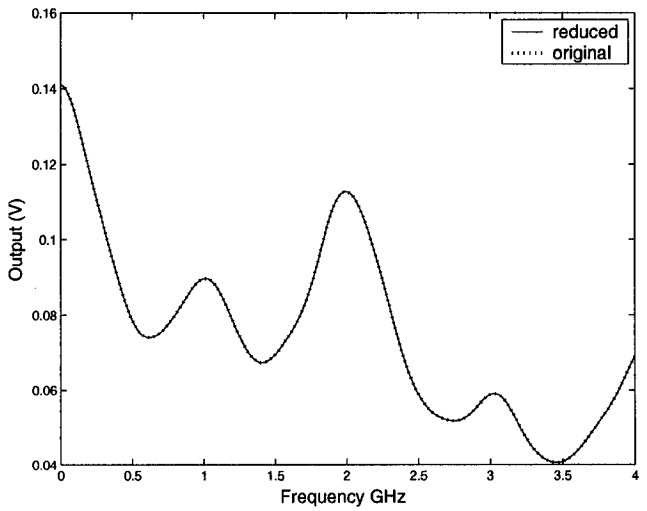


**Fig. 3.16** Frequency response for transmission line loads for example1 (comparison between the original system and the proposed macromodel).

### 3.4.2 Example 2

The second example is a 20-port network (nineteen of which are considered to be constrained ports) which we refer to as circuit2 as shown in Fig. 3.20. The circuit contains 36 non-coupled transmission lines and 1 nine-coupled transmission line system. After discretization of the interconnects, the resulting size of the MNA matrices was 4890 $\times$ 4890. The size of the reduced macromodel was 494 $\times$ 494 if using the proposed approach. Using traditional block Arnoldi macromodeling approach, the size of the reduced order model was 800 $\times$ 800 in order to match the original system up to 4GHz. A summary of macromodel size comparisons between the proposed method and block Arnoldi for example 2 is shown in Table 3.1. The nominal load values

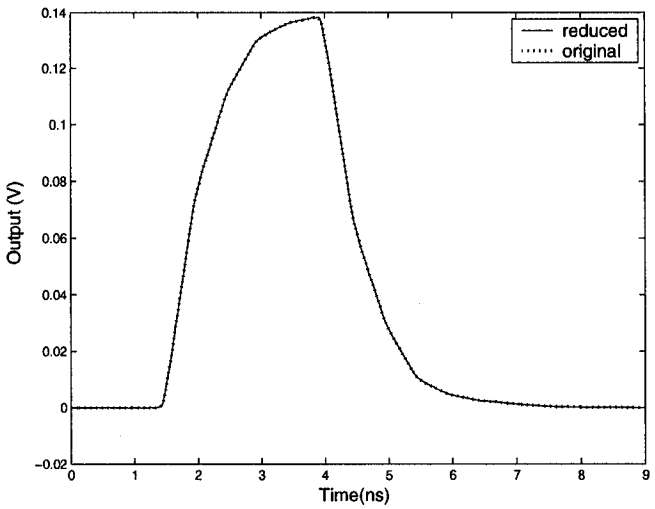


**Fig. 3.17** Transient response for transmission line loads for example 1 (comparison between the original system and the proposed macromodel).

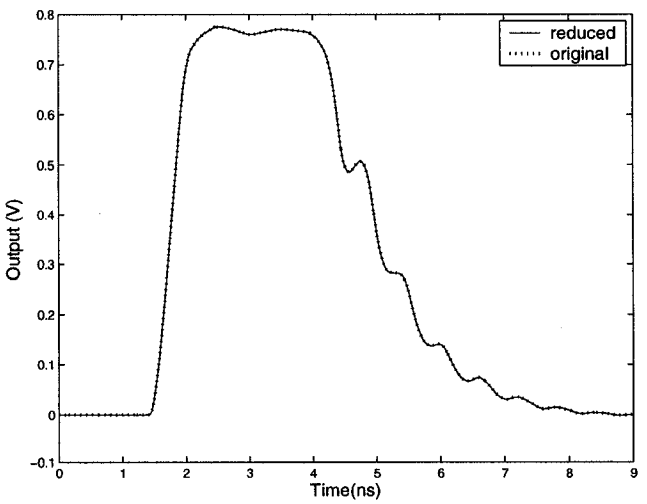


**Fig. 3.18** Transient response for diode loads for example 1 (comparison between the original system and the proposed macromodel).

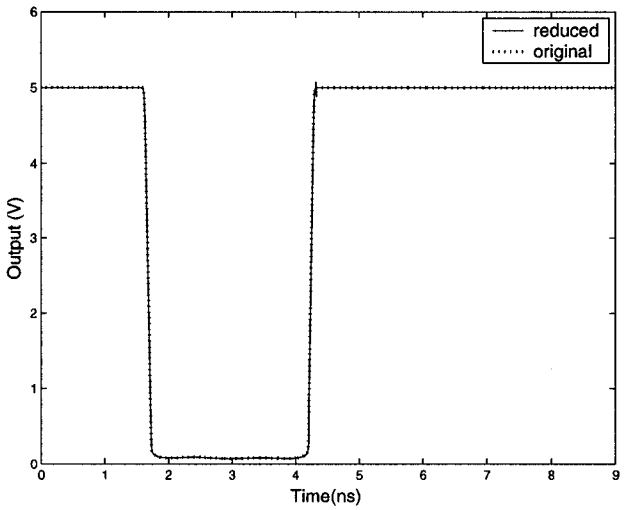


**Fig. 3.19** Transient response for inverter load for example 1 (comparison between the original system and the proposed macromodel).

(expansion point) for generating the reduced macromodel are conductance 0S (corresponds to open circuit) and capacitance $0pF$ at all constrained ports. This macromodel for example 2 was tested to be valid for resistive capacitive loads at the ports when the conductance ranges from 0 to 10S (which corresponds to 0.1 $\Omega$ to open circuit), and capacitance from 0 to 100pF.

The macromodel for example 2 was generated once using the proposed approach and then was reused by 100 testing cases with different resistive and capacitive loads. The values of the capacitors are randomly chosen from 0.01pF to 100pF, and the values of resistors were randomly chosen from 0.1$\Omega$ to 10000$\Omega$, which fall within the range of validity of the macromodel indicated above. The average relative error in the frequency responses for all testing cases is 0.43%. Fig. 3.21, Fig. 3.22, Fig. 3.23 and Fig. 3.24 show four sample frequency response results. As can be seen, they are accurate compared to the original network.

The transient response for circuit2 with resistive and capacitive loads is shown in Fig. 3.25. CPU cost for the transient analysis is shown in Table 3.3, indicating a speed-up of 28 with respect to the solution obtained from the original circuit2.

This macromodel was also tested with transmission line loads and diode loads as shown

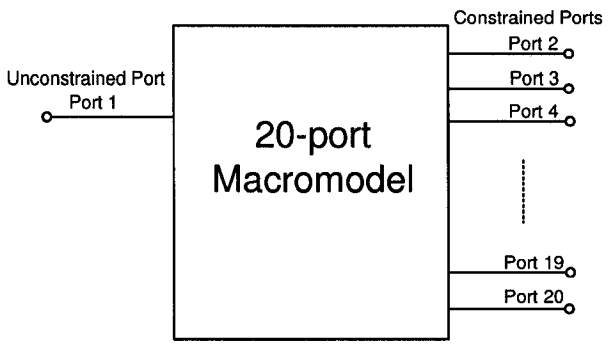


**Fig. 3.20** A 20-port interconnect network as circuit2.

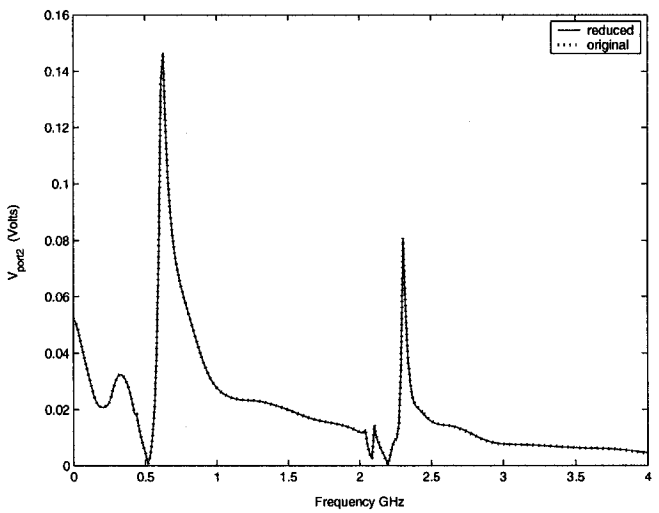


**Fig. 3.21** Frequency response of port2 for circuit2 with capacitive loads at the constrained ports (comparison between the original system and the proposed macromodel).

**Table 3.3** CPU comparison of transient analysis for circuit2

| | Size | Time | Speed-up |
|---|---|---|---|
| Reduced System | 494 | $36s$ | 28 |
| Original System | 4890 | $1022s$ | - |

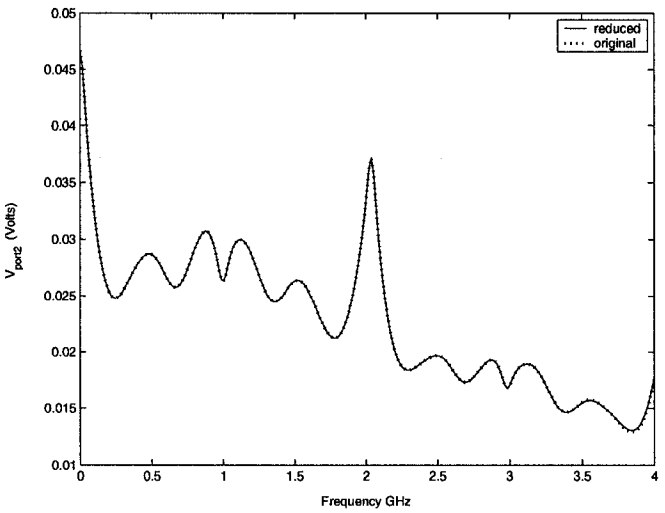


**Fig. 3.22** Frequency response of port2 for circuit2 with resistive loads at the constrained ports (comparison between the original system and the proposed macromodel).

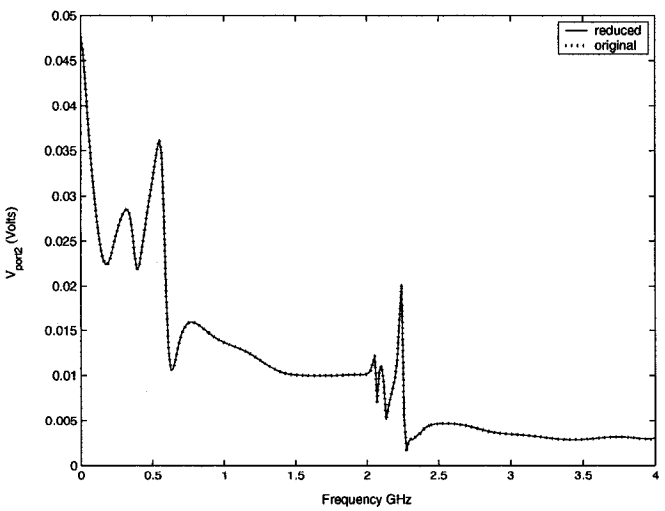


**Fig. 3.23** Frequency response of port2 for circuit2 with resistive/capacitive loads at the constrained ports (comparison between the original system and the proposed macromodel).

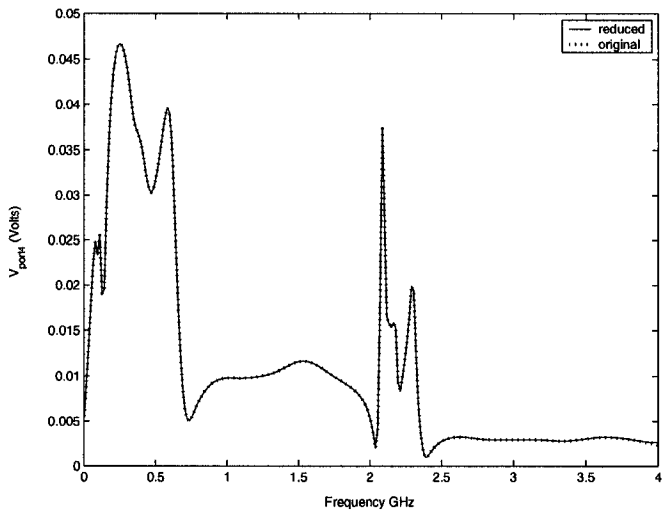


**Fig. 3.24** Frequency response of port4 for circuit2 with resistive/capacitive loads at the constrained ports (comparison between the original system and the proposed macromodel).

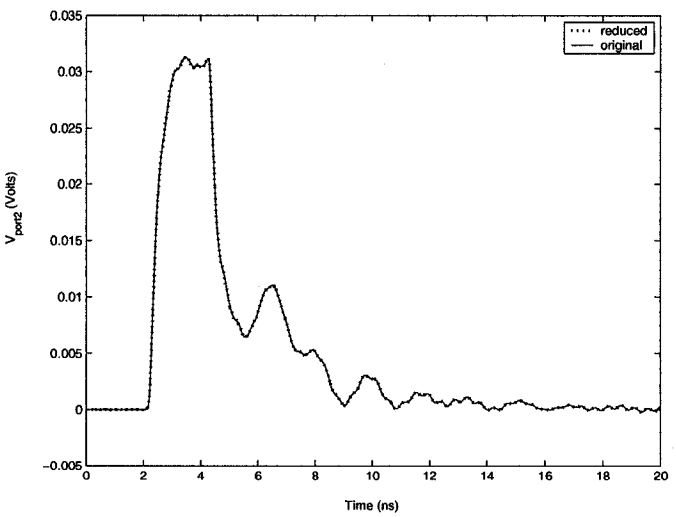


**Fig. 3.25** Comparison of the transient results obtained from the proposed macromodel and the original system in circuit2 with constrained ports connected to resistive and capacitive loads.

in Fig. 3.15(a) and (b). Fig. 3.26 shows the frequency response for the transmission line load compared with that of the original system. It matches the original system. The transient responses for the transmission line loads and diode loads are shown in Fig. 3.27 and Fig. 3.28 respectively. The transient responses are due to a pulse with a rise time of 500ps and a pulse width of 2ns. The amplitudes for the pulses are 1V and 50V respectively. As can be seen, they match the original system very well.

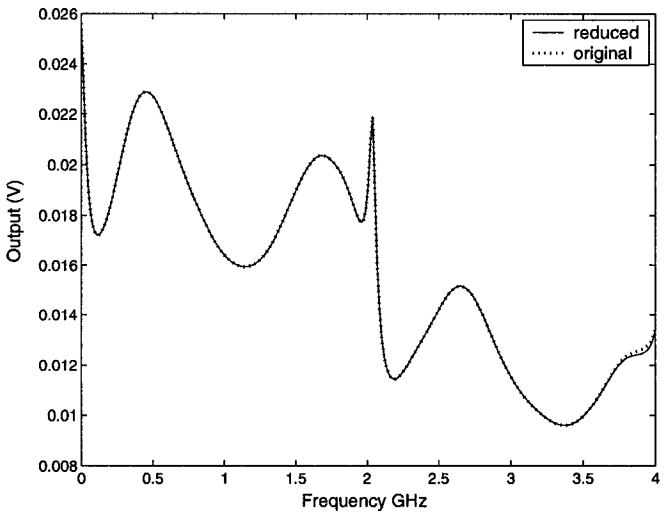


**Fig. 3.26** Frequency response for transmission line loads for example 2 (comparison between the original system and the proposed macromodel).

### 3.4.3 Example 3

The third example is a 30-port network(twenty-nine of which are constrained ports) which we refer to as circuit3 as shown in Fig. 3.29. The circuit contains 36 non-coupled transmission lines and 1 nine-coupled transmission line system. After discretization of the interconnects, the resulting size $n$ of the MNA matrices was 4900 $\times$ 4900. Using traditional block Arnoldi macromodeling approach, the size of the reduced order model was 990 $\times$ 990. Using the proposed approach, the size of the reduced macromodel was 554 $\times$ 554 in order to match the original system up to 4GHz. A summary of macromodel size comparisons between the proposed method

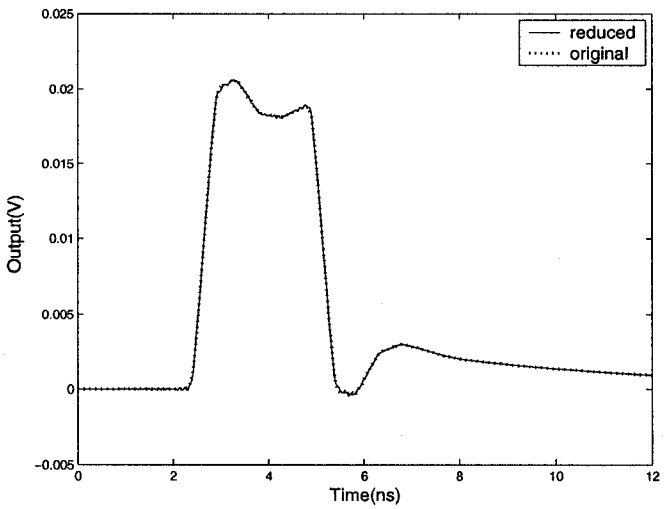


**Fig. 3.27** Transient response for transmission line loads for example 2 (comparison between the original system and the proposed macromodel).

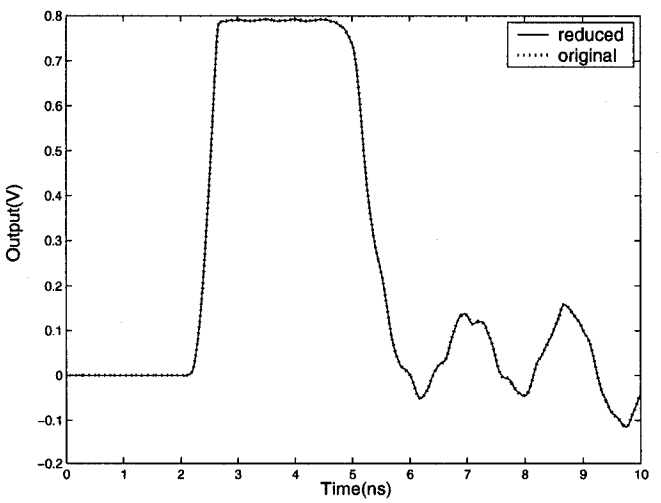


**Fig. 3.28** Transient response for diode loads for example 2 (comparison between the original system and the proposed macromodel).

and block Arnoldi for example 3 is shown in Table 3.1. The nominal load values (expansion point) for generating the reduced macromodel are conductance 0S (corresponds to open circuit) and capacitance $0pF$ at all constrained ports. This macromodel for example 3 was tested to be valid for resistive capacitive loads at the ports when the conductance ranges from 0 to 1S (which corresponds to 1 $\Omega$ to open circuit), and capacitance from 0 to 50pF.

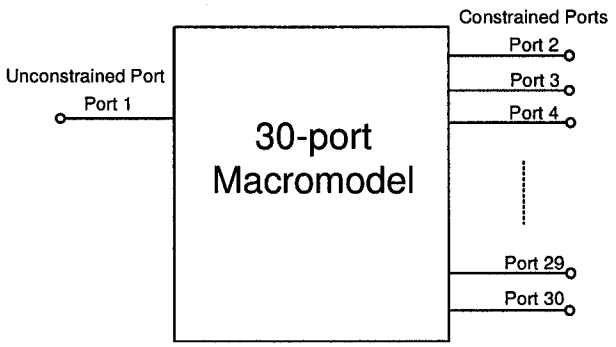


**Fig. 3.29** A 30-port interconnect network as circuit3

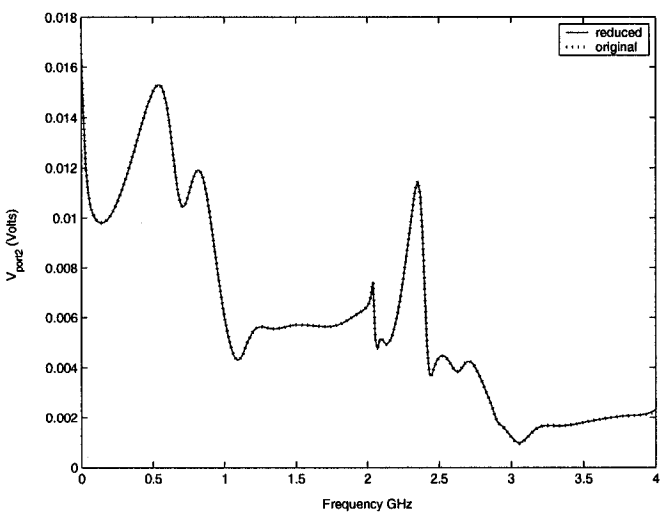


**Fig. 3.30** Frequency response for circuit3 with resistive/capacitive loads at the constrained ports (comparison between the original system and the proposed macromodel).

Once again, the proposed macromodel was generated once and was tested by 30 different cases. The values of the capacitors are randomly chosen from 0.1pF to 50pF, and the values of

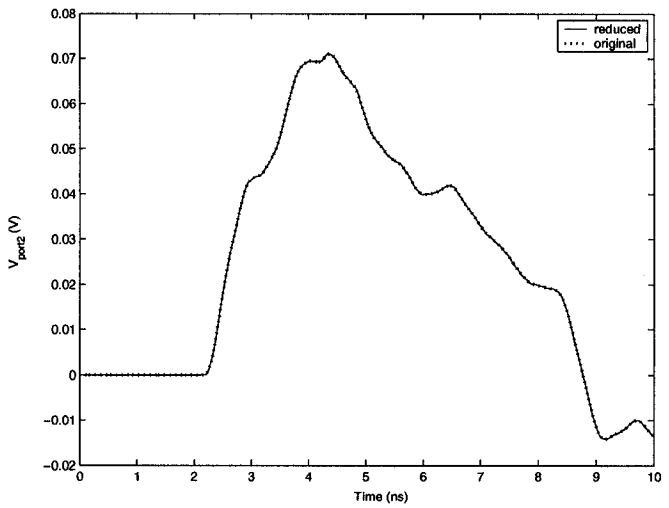


**Fig. 3.31** Comparison of the transient results obtained from the proposed macromodel and the original system in circuit3 with constrained ports connected to capacitive loads.

**Table 3.4** CPU comparison of transient analysis for circuit3

| | Size | Time | Speed-up |
|---|---|---|---|
| Reduced System | 554 | $39s$ | 17 |
| Original System | 4900 | $662s$ | - |

resistors were randomly chosen from 1Ω to 10000Ω, which fall within the range of validity of the macromodel indicated above. The average relative error for those testing cases in the frequency responses compared to the original network is 0.39%. One sample frequency response is shown in Fig. 3.30. As can be seen, it matches the original network. The transient response for circuit3 with capacitive loads is shown in Fig. 3.31. CPU cost comparison for the transient analysis is shown in Table 3.4, indicating a speed-up of 17 with respect to the solution obtained from the original circuit3.

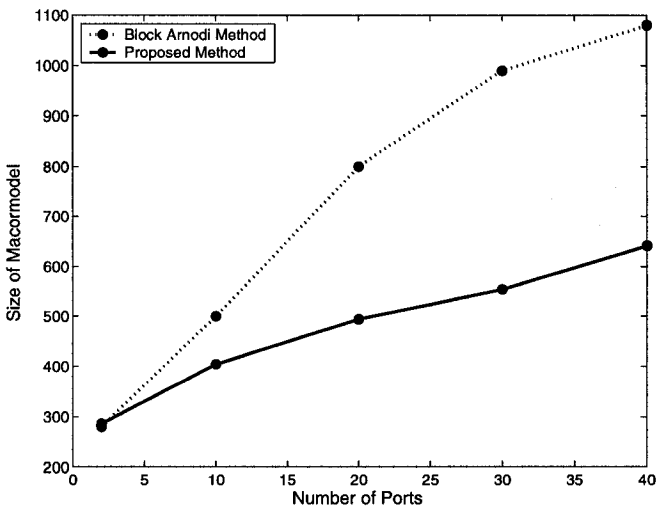


**Fig. 3.32** Macromodel size versus number of ports (comparison between the proposed method and the traditional MOR method).

### 3.4.4 Example 4

The fourth example illustrates the relationship between the size of the reduced macromodel and the number of ports. Consider an interconnect network with $p$ ports. $p$ is varied between 2 to 40. The comparison between the proposed approach and the traditional model order reduction method is shown in Fig. 3.32. As can be seen, the size of reduced order macromodels obtained from the proposed algorithm is much less sensitive to the number of ports than those obtained from traditional block Arnoldi method. This proposed method therefore significantly extends the

range of applicability of model order reduction methods to systems with large number of ports when information about the types of loads on the ports is available.

# Chapter 4

# Sparse Parametric Multi-level Reduction

## 4.1 Introduction

Krylov projection methods have proven to be very effective in reducing the size of interconnect networks while, at the same time, preserving the passivity. However, these methods still suffer from two difficulties. First, these methods generate a reduced order macromodel which contains many redundant poles. This results in a relatively large macromodel and significantly reduces the efficiency of the simulation. In order to address the problem associated with a large macromodel, a projection method based on two levels of reduction is presented [45]. The proposed method is guaranteed passive and typically results in a macromodel which is half the size of that obtained using conventional Krylov techniques.

Another difficulty for the traditional model order reduction methods is that the reduced model is valid for a specific circuit, and therefore has to be regenerated each time a circuit parameter is modified. This is a very inefficient process for many practical cases such as optimization or parameter sweeping. In order to address this issue, parametric model order reduction techniques were proposed [31], [32], [77], [83]. These techniques improved the simulation time for parametric interconnect networks since the model only needs to be created once and then it can be used many times with different parameter values. However, the traditional parametric model re-

duction methods are based on moment-matching. Such approaches are known to achieve a larger macromodel than what is necessary. Furthermore, the resulting parametric macromodel obtained from parametric model reduction is generally dense. A dense macromodel significantly reduces the efficiency of the simulation. To overcome these two problems, sparse multi-level parametric model order reduction is proposed to produce a parametric macromodel which is very small as well as sparse [48], [49]. In the proposed approach, a new parametric system model suitable for sparsification is first developed. The formulation for this new model replaces the stamps of the parametric elements with constrained port representation. Then two-level reduction is done on the new model. This results in a reduced order parametric macromodel which is very small but in general dense. Sparsification is therefore performed on this dense macromodel. Finally, the sparse reduced macromodel is brought back to the traditional representation without losing sparsity.

The rest of the chapter is organized as follows. A new multi-level model order reduction method based on singular value decomposition is proposed in Section 4.2. In addition, a new multi-level parametric model order reduction is described in Section 4.3. Furthermore, a novel sparsification technique for a parametric system is outlined in Section 4.4. Finally, sparse multi-level parametric model order reduction is proposed in Section 4.5.

## 4.2 Singular Value Decomposition Based Multi-level Reduction

The new model order reduction method consists of two levels of order reduction. In the first level, a passive reduced model is obtained through projection on the Krylov subspace using the Arnoldi process. The second level of reduction is also performed using a real congruence transformation in order to guarantee the passivity of the macromodel [84]. In this case the reduction basis is chosen by using singular value decomposition (SVD) to filter out redundant information as will be outlined in the sections below.

Recalling from Section 2.1.2, the MNA formulation for a multi-port interconnect network

containing $p$ ports can be written as [63]

$$\begin{aligned}(\boldsymbol{G}+s\boldsymbol{C})\boldsymbol{x} &= \boldsymbol{R}\boldsymbol{V}\\ \boldsymbol{I} &= \boldsymbol{R}^T\boldsymbol{x}\end{aligned} \tag{4.1}$$

where $s$ represents the complex frequency, $\boldsymbol{V}$ and $\boldsymbol{I}$ are vectors containing port voltages and currents. $\boldsymbol{G} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{C} \in \mathbb{R}^{n\times n}$ contain the contributions of the memoryless and memory elements respectively. $\boldsymbol{x}$ refers to unknown node voltages and unknown currents. $n$ is the total number of variables in the MNA equations. $\boldsymbol{R} \in \mathbb{R}^{n\times p}$ is a selector matrix that maps the port voltages and currents into the node space of the circuit.

### 4.2.1 First level of reduction

In the first level of reduction, a standard Krylov subspace is used [19]. The reduced order model is constructed through a congruence transformation $\boldsymbol{Q}$ on the original system in (4.1) to obtain:

$$(\hat{\boldsymbol{G}}+s\hat{\boldsymbol{C}})\hat{\boldsymbol{x}} = \hat{\boldsymbol{R}}\boldsymbol{V} \tag{4.2}$$

where $\hat{\boldsymbol{G}}, \hat{\boldsymbol{C}} \in \mathbb{R}^{N_q\times N_q}$ and $\tilde{\boldsymbol{R}} \in \mathbb{R}^{N_q\times p}$ are given by

$$\hat{\boldsymbol{G}} = \boldsymbol{Q}^T\boldsymbol{G}\boldsymbol{Q}; \quad \hat{\boldsymbol{C}} = \boldsymbol{Q}^T\boldsymbol{C}\boldsymbol{Q}; \quad \hat{\boldsymbol{R}} = \boldsymbol{Q}^T\boldsymbol{R} \tag{4.3}$$

The matrix $\boldsymbol{Q} \in \mathbb{R}^{n\times N_q}$ ($N_q = (q+1)\times p$) is a real orthonormal basis of the subspace spanned by the block moments

$$\text{colsp}[\boldsymbol{Q}] = \text{colsp}[\boldsymbol{M}_0, \boldsymbol{M}_1, \boldsymbol{M}_2, \ldots, \boldsymbol{M}_q] \tag{4.4}$$

where $\boldsymbol{M}_0 \in \mathbb{R}^{n\times p}, \ldots, \boldsymbol{M}_q \in \mathbb{R}^{n\times p}$ are block moments of the original system. The congruence matrix $\boldsymbol{Q}$ is accurately computed using the Arnoldi process. It can be shown that the reduced

macromodel is passive and preserves the first $q$ moments of the original system [19]. Although the reduced model is much smaller than the original system ($N_q \ll n$), it is not optimal and contains a lot of redundant poles [29], [30], [84]. A large macromodel significantly increases the CPU cost of simulation. This problem is addressed using a second level reduction as discussed in the following section.

### 4.2.2 Second level of reduction

The objective of doing the second level reduction is to remove unnecessary poles of the reduced model from the first level reduction, while at the same time preserving the passivity. In order to achieve this goal, the system in (4.2) is further reduced by performing another congruence transformation.

$$(\tilde{\boldsymbol{G}} + s\tilde{\boldsymbol{C}})\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{R}}\boldsymbol{V} \tag{4.5}$$

where $\tilde{\boldsymbol{G}}, \tilde{\boldsymbol{C}} \in \mathbb{R}^{N_k \times N_k}$ and $\tilde{\boldsymbol{R}} \in \mathbb{R}^{N_k \times p}$ are given by

$$\tilde{\boldsymbol{G}} = \boldsymbol{Q}_s^T \hat{\boldsymbol{G}} \boldsymbol{Q}_s; \quad \tilde{\boldsymbol{C}} = \boldsymbol{Q}_s^T \hat{\boldsymbol{C}} \boldsymbol{Q}_s; \quad \tilde{\boldsymbol{R}} = \boldsymbol{Q}_s^T \hat{\boldsymbol{R}} \tag{4.6}$$

$\boldsymbol{Q}_s \in \mathbb{R}^{N_q \times N_k}$ is a real orthonormal matrix, thus the reduced system is passive by construction [84]. In order to obtain the subspace $\boldsymbol{Q}_s$ for the second level of reduction, we first construct a subspace which spans the response of the reduced system in (4.2) over the frequency range of interest.

$$\boldsymbol{F} = [real(\hat{\boldsymbol{x}}_1), imag(\hat{\boldsymbol{x}}_1), real(\hat{\boldsymbol{x}}_2), imag(\hat{\boldsymbol{x}}_2), \ldots, real(\hat{\boldsymbol{x}}_N), imag(\hat{\boldsymbol{x}}_N)] \tag{4.7}$$

where $\hat{\boldsymbol{x}}_1 \in \mathbb{R}^{N_q \times p}$, $\ldots$, $\hat{\boldsymbol{x}}_N \in \mathbb{R}^{N_q \times p}$ are sampling frequency response points. They are computed using (4.2) and are given by the solution:

$$(\hat{\boldsymbol{G}} + j\omega_i \hat{\boldsymbol{C}})\hat{\boldsymbol{x}}_{\boldsymbol{i}} = \hat{\boldsymbol{R}} \tag{4.8}$$

It is to be noted that while the frequency response points are in general complex. $\boldsymbol{F}$ is collected by splitting the real and imaginary parts of the complex response in order to obtain a real basis. A real basis is required for preserving the passivity. As can be seen from (4.7), the matrix $\boldsymbol{F}$ spans sampling frequency response points and therefore conserves the system response. But this matrix cannot be directly used as the reduction subspace since various vectors in $\boldsymbol{F}$ are not completely linearly independent and therefore contains many redundant information. Thus, the next step is to find an orthonormal basis of the dominant directions in $\boldsymbol{F}$. In order to achieve this, the matrix $\boldsymbol{F}$ is decomposed using singular value decomposition process. It computes the matrices $\boldsymbol{U}$, $\boldsymbol{S}$, $\boldsymbol{V}$ from $\boldsymbol{F}$ such that:

$$\boldsymbol{F} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\boldsymbol{T}} \tag{4.9}$$

where $\boldsymbol{S}$ is a diagonal matrix whose entries are non-negative elements in decreasing order and $\boldsymbol{U}, \boldsymbol{V}$ are orthogonal matrices. The values on the diagonal of $\boldsymbol{S}$ provide a measure of the relative importance of the various vectors in the orthonormal subspace defined by the columns of $\boldsymbol{U}$. This provides a convenient way to filter out the redundant subspace. Thus, taking only the leading $k$ columns of $\boldsymbol{U}$ that correspond to large values in $\boldsymbol{S}$ will give us a compact subspace, $\boldsymbol{Q}_s$, which contains the responses of the system as shown in Fig. 4.1. The order $N_k$ is chosen using the following criteria:

$$\frac{s(N_k, N_k)}{s(1,1)} \leq err \tag{4.10}$$

$s(1,1)$ and $s(N_k, N_k)$ is the first and the $N_k{}^{th}$ elements of the diagonal elements of $\boldsymbol{S}$. $err$ is the error bound to choose the dominant subspace. Thus the dominant subspace $\boldsymbol{Q}_s$ for the second level of reduction is given by:

$$\boldsymbol{Q}_s = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{N_k}] \tag{4.11}$$

where $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{N_k}$ are the first $N_k$ columns of the matrix $\boldsymbol{U}$.

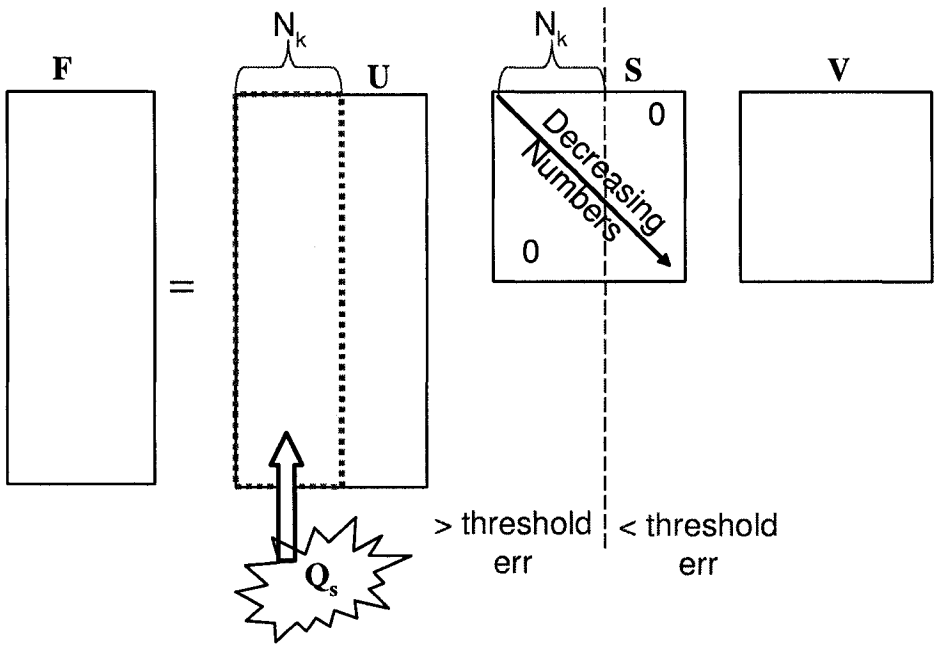


**Fig. 4.1** Reduction subspace obtained using singular value decomposition

### 4.2.3 Results

#### Example 1

The first example is an 18-port network. This network contains 1 coupled nine-transmission line network. After discretization of the interconnects, the resulting size $n$ of the MNA matrices is 2727. Using the block Arnoldi macromodeling approach, the size of the subspace that would match the original network up to 10GHz was 720. After the second level of reduction, the final size of the subspace was only 490. A summary of macromodel size comparison among the original system, the reduced macromodel obtained from traditional model order reduction based on block Arnoldi, and that from multi-level reduction is shown in Table 4.1.

The $Y$ parameters $Y_{11}$ and $Y_{12}$ are computed and shown in Fig. 4.2 and Fig. 4.3. As can be seen, the results of the proposed macromodel match the original circuit up to 10GHz. Fig. 4.4 shows the transient response due to a $1V$ pulse with a rise time of $0.3ns$ and a pulse width of $5ns$. As expected, it is also accurate. To further demonstrate the efficiency of the proposed

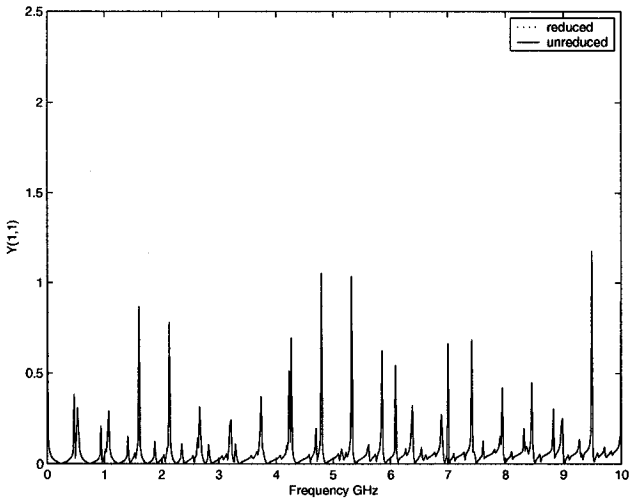


**Fig. 4.2** $Y_{11}(s)$ for exampl1 (comparison between the original system and the proposed macromodel)

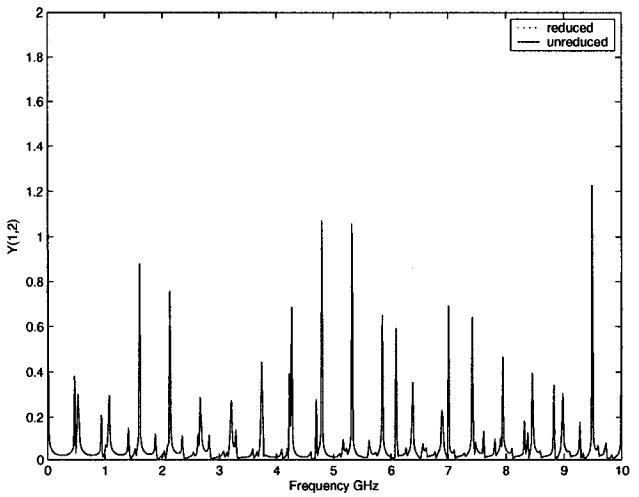


**Fig. 4.3** $Y_{12}(s)$ for example1 (comparison between the original system and the proposed macromodel)

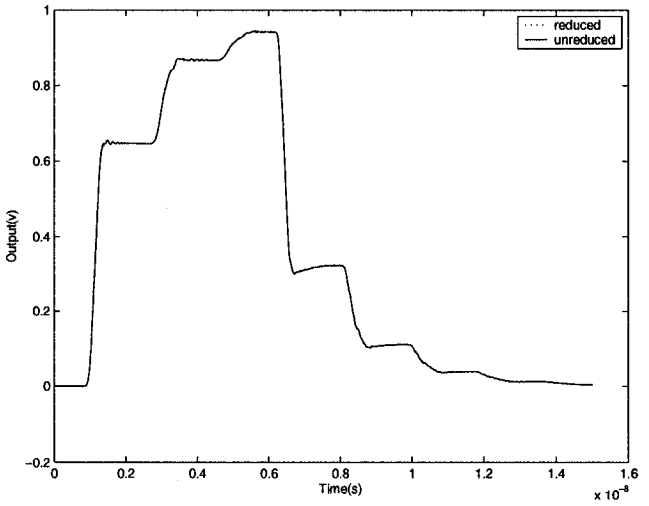


**Fig. 4.4** transient response for example1 (comparison between the original system and the proposed macromodel)

method, a summary of CPU cost comparison for example1 to obtain transient responses is shown in Table 4.2, indicating a speed-up of 15 with respect to the original circuit.

**Table 4.1** Macromodel size comparison

| | Example1 | Example2 | Example3 |
|---|---|---|---|
| Number of Ports | 18 | 10 | 20 |
| Original System | 2727 | 4340 | 4890 |
| Block Arnoldi | 720 | 920 | 1500 |
| Multi-level Reduction | 490 | 400 | 620 |

### Example2

The second example is a 10-port network. This network contains 18 non-coupled transmission lines and 1 coupled nine-transmission line. After discretization of the interconnects, the resulting size $n$ of the MNA matrices is 4340. Using traditional block Arnoldi macromodeling approach,

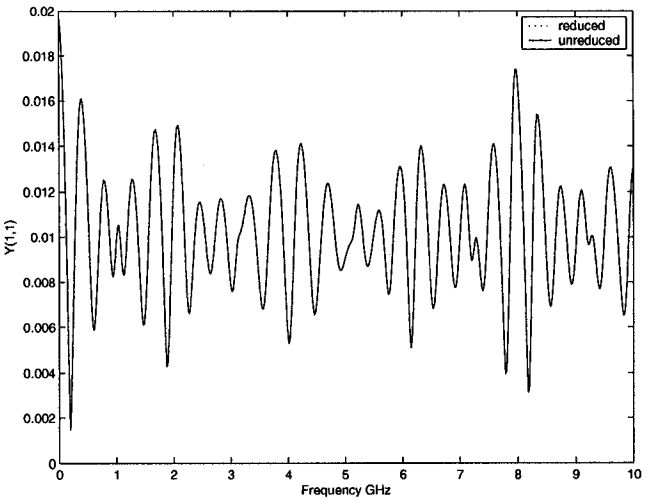


**Fig. 4.5** $Y_{11}(s)$ for example2 (comparison between the original system and the proposed macromodel)

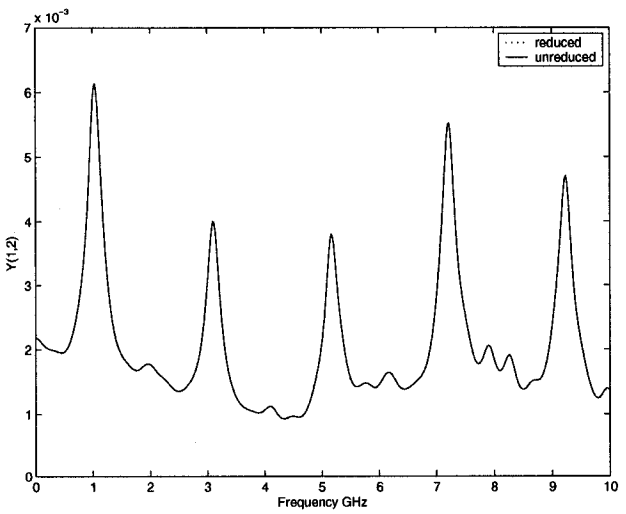


**Fig. 4.6** $Y_{12}(s)$ for example2 (comparison between the original system and the proposed macromodel)

the size of the subspace that would match the original network up to 10GHz was 920. After the second level of reduction, the final size of the subspace was only 400. A summary of macromodel size comparison is shown in Table 4.1. The $Y$ parameters $Y_{11}$ and $Y_{12}$ for example2 are computed

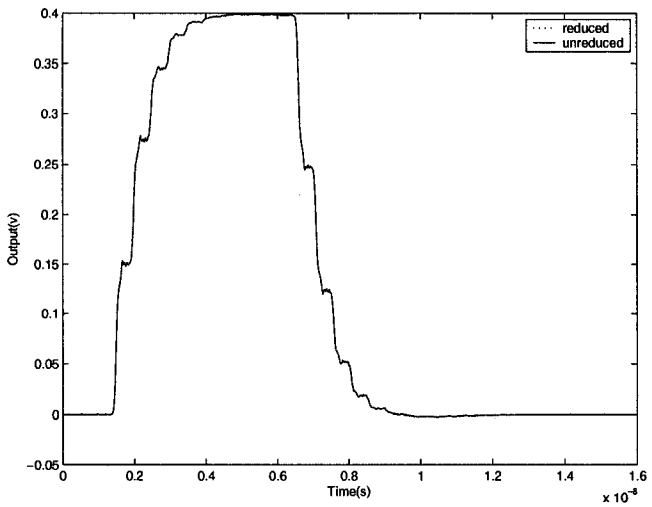


**Fig. 4.7** transient response for example2 (comparison between the original system and the proposed macromodel)

and shown in Fig. 4.5 and Fig. 4.6. As can be seen, the results are accurate up to 10GHz. Fig. 4.7 shows the transient response due to a $1V$ pulse with a rise time of $0.1ns$ and a pulse width of $5ns$. As expected, it is also accurate. A summary of CPU cost comparison for example2 to obtain transient responses is shown in Table 4.2.

**Table 4.2** Macromodel CPU comparison between the proposed reduced macromodel and the original system

| | Multi-level Reduction | Original System | Speed-up |
|---|---|---|---|
| Example1 | 27.4(s) | 411.4(s) | 15 |
| Example2 | 17.8(s) | 201.5(s) | 11 |
| Example3 | 43.3(s) | 633.2(s) | 14 |

### Example3

The third example is a 20-port network. This network contains 36 non-coupled transmission lines and 1 coupled nine-transmission line. After discretization of the interconnects, the resulting size $n$ of the MNA matrices is 4890. Using traditional block Arnoldi macromodeling approach, the size of the subspace that would match the original network up to 10GHz was 1500. After the second level of reduction, the final size of the subspace was only 620. A summary of macromodel size comparison is shown in Table 4.1.

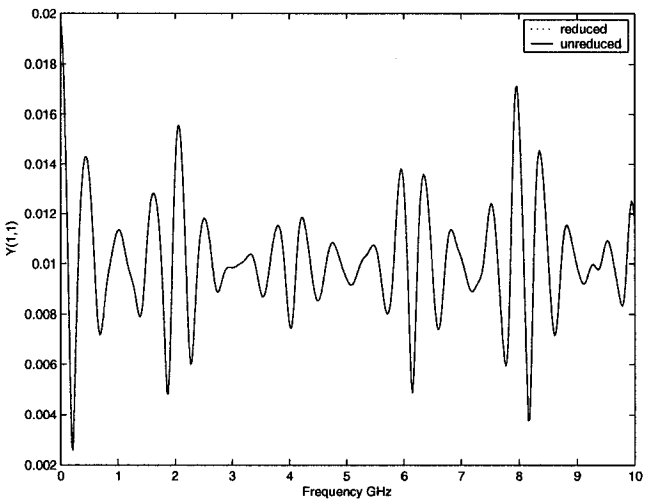


**Fig. 4.8** $Y_{11}(s)$ for example3 (comparison between the original system and the proposed macromodel)

The $Y$ parameters $Y_{11}$ and $Y_{12}$ for example3 are computed and shown in Fig. 4.8 and Fig. 4.9. Again, the results are accurate up to 10GHz. Fig. 4.10 shows the transient response due to a $1V$ pulse with a rise time of $0.5ns$ and a pulse width of $5ns$. As expected, accurate results are also obtained in this case. A summary of CPU cost comparison for circuit3 to obtain transient responses is shown in Table 4.2.

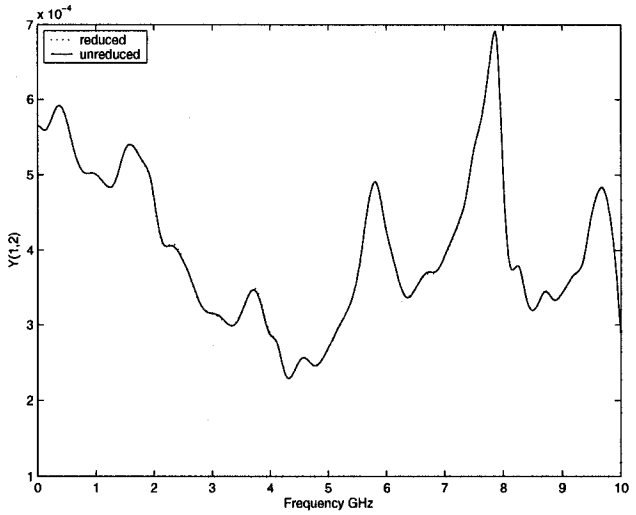


**Fig. 4.9** $Y_{12}(s)$ for example3 (comparison between the original system and the proposed macromodel)

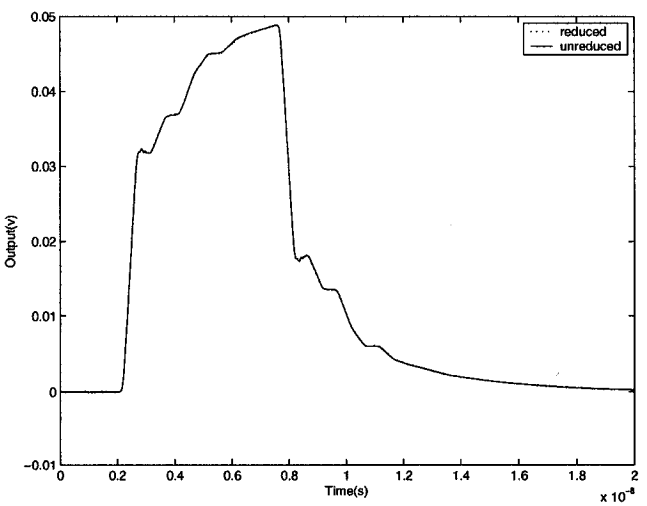


**Fig. 4.10** transient response for example3 (comparison between the original system and the proposed macromodel)

## 4.3 Parametric Singular Value Decomposition Based Model Order Reduction

The model order reduction methods presented in the previous section are performed on one parameter (frequency). However in many practical cases, it is necessary to account for the variations of some parameters when evaluating system performance. These variations could be the width and height of the transmission lines due to the environmental effects or manufacturing variations. Given the fact that the system equations for interconnect networks are very large due to discretization, it is not feasible to perform model order reduction each time as the parameter changes. In order to improve the efficiency, parametric model order reduction was proposed in the literature [31], [32], [77]. This technique produced a reduced order macromodel which is valid over a predefined range of parameters. In other words, we do not need to perform a new model order reduction each time as the parameter is modified. This results in significant CPU cost savings. However current parametric model order reduction is based on multi-dimensional moment matching, this method is known to result in a macromodel, which is not optimal [83]. In order to address this problem, a multi-level parametric model order reduction approach is proposed [48]. This new approach can remove the redundant information from using Krylov methods, while at the same time preserving the passivity of the system. The proposed method consists of two levels of reduction. The first is done using traditional parametric model order reduction, and in the second level, the singular value decomposition (SVD) [85] process is used to obtain the congruence transformation matrix and further reduces the system. The system after the second level of reduction is typically one-third the size of the first reduced model obtained using Krylov methods. Finally, since a real congruence transformation is used, the reduced model is provably passive by construction as will be demonstrated in Section 4.3.3.

Consider a large parametric interconnect network containing distributed elements as well as lumped components. After discretization, the parametrized modified nodal analysis (MNA) formulation for this network can be written as

$$\boldsymbol{G}\boldsymbol{x} + s\boldsymbol{C}\boldsymbol{x} + \left(\sum_{i=1}^{r} f_i(\lambda_i)\boldsymbol{D}_i\right)\boldsymbol{x} + s\left(\sum_{j=1}^{c} g_j(\phi_j)\boldsymbol{E}_j\right)\boldsymbol{x} = \boldsymbol{R}\boldsymbol{u}; \quad \boldsymbol{i} = \boldsymbol{R}^T\boldsymbol{x} \tag{4.12}$$

The above formulation is the same as the parametrized formulation described in (2.46). However it is repeated in this chapter for the sake of clarity and completeness.

### 4.3.1 First level of reduction

In the first level of reduction, traditional parametric model order reduction is used [31], [86]. The reduced order macromodel is obtained from (4.12) by using a congruence transformation which results in:

$$\hat{\boldsymbol{G}}\hat{\boldsymbol{x}} + s\hat{\boldsymbol{C}}\hat{\boldsymbol{x}} + \left(\sum_{i=1}^{r} f_i(\lambda_i)\hat{\boldsymbol{D}}_i\right)\hat{\boldsymbol{x}} + s\left(\sum_{j=1}^{c} g_j(\phi_j)\hat{\boldsymbol{E}}_j\right)\hat{\boldsymbol{x}} = \hat{\boldsymbol{R}}u; \quad \hat{\boldsymbol{i}} = \hat{\boldsymbol{R}}^T\hat{\boldsymbol{x}} \tag{4.13}$$

where $\hat{\boldsymbol{G}}$, $\hat{\boldsymbol{C}}$, $\hat{\boldsymbol{D}}_i, (i = 1, \cdots, r)$ and $\hat{\boldsymbol{E}}_j, (j = 1, \cdots, c) \in \mathbb{R}^{N_q \times N_q}$ and are given by

$$\hat{\boldsymbol{G}} = \boldsymbol{Q}^T\boldsymbol{G}\boldsymbol{Q}; \quad \hat{\boldsymbol{C}} = \boldsymbol{Q}^T\boldsymbol{C}\boldsymbol{Q}; \quad \hat{\boldsymbol{D}}_i = \boldsymbol{Q}^T\boldsymbol{D}_i\boldsymbol{Q}; \quad \hat{\boldsymbol{E}}_j = \boldsymbol{Q}^T\boldsymbol{E}_j\boldsymbol{Q}; \quad \hat{\boldsymbol{R}} = \boldsymbol{Q}^T\boldsymbol{R} \tag{4.14}$$

The congruence transformation matrix, $\boldsymbol{Q}$, is chosen as an orthonormal basis of a subspace which spans the moments of the system with respect to frequency, and with respect to the parameters $\lambda_1, ..., \lambda_r, \phi_1, ..., \phi_c$, as well as the cross moments.

$$colsp\left[\boldsymbol{Q}\right] = colsp\left[\begin{array}{ccccccccccc} \boldsymbol{M}_{s_k} & ... & \boldsymbol{M}_{\lambda_{ik}} & ... & \boldsymbol{M}_{\phi_{jk}} & ... & \boldsymbol{M}_{s_a\phi_{bd}} & ... & \boldsymbol{M}_{s_p\phi_{mn}} & ... \end{array}\right] \tag{4.15}$$

where $\boldsymbol{M}_{s_k}$ is the $k^{th}$ moment with respect to frequency, $\boldsymbol{M}_{\lambda_{ik}}$ is the $k^{th}$ moment with respect to the $i^{th}$ memoryless parameter, $\boldsymbol{M}_{\phi_{jk}}$ is the $k^{th}$ moment with respect to the $j^{th}$ memory parameter, $\boldsymbol{M}_{s_a\phi_{bd}}$ is the cross moment between the $a^{th}$ frequency moment and the $d^{th}$ moment of the $b^{th}$ memoryless parameter, $\boldsymbol{M}_{s_p\phi_{nm}}$ is the cross moment between the $p^{th}$ frequency moment and the

$m^{th}$ moment of the $n^{th}$ memory parameter.

A combination of the Arnoldi process and standard QR decomposition is used to accurately compute $\boldsymbol{Q}$. The reduced order model in (4.13) is much smaller than the original system in (4.12). However, this approach is based on moment matching and results in a much larger reduced model than is necessary [5], [83]. In the following section, a second level of reduction is described which addresses this issue.

### 4.3.2 Second level of reduction

The reduced model in (4.13) provides an efficient way to solve the original system, however, it is still much larger than is necessary [5], [29]. In the second level of reduction, this macromodel is further reduced using another congruence transformation matrix, $\mathbf{Q}_s$. This time, the change of variables is with respect to the reduced model in (4.13). Thus, substituting $\tilde{\mathbf{x}} = \mathbf{Q}_s\hat{\mathbf{x}}$, the second level of reduction gives:

$$\tilde{\boldsymbol{G}}\tilde{x} + s\tilde{\boldsymbol{C}}\tilde{\boldsymbol{x}} + \left(\sum_{i=1}^{r} f_i(\lambda_i)\tilde{\boldsymbol{D}}_i\right)\tilde{\mathbf{x}} + s\left(\sum_{j=1}^{c} g_j(\phi_j)\tilde{\boldsymbol{E}}_j\right)\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{R}}u; \quad \boldsymbol{i} = \tilde{\boldsymbol{R}}^T\tilde{\boldsymbol{x}} \tag{4.16}$$

where: $\tilde{\boldsymbol{G}} = \boldsymbol{Q}_s^T\hat{\boldsymbol{G}}\boldsymbol{Q}_s, \tilde{\boldsymbol{C}} = \boldsymbol{Q}_s^T\hat{\boldsymbol{C}}\boldsymbol{Q}_s, \tilde{\boldsymbol{D}}_i = \boldsymbol{Q}_s^T\hat{\boldsymbol{D}}_i\boldsymbol{Q}_s, \tilde{\boldsymbol{E}}_j = \boldsymbol{Q}_s^T\hat{\boldsymbol{E}}_j\boldsymbol{Q}_s, \tilde{\boldsymbol{R}} = \boldsymbol{Q}_s^T\hat{\boldsymbol{R}}$. and $\tilde{\boldsymbol{x}}$ is the solution to this second reduced model. For the second level of reduction, the congruence transformation matrix, $\boldsymbol{Q}_s$, is chosen such that it spans the response of the system in (4.13) over the desired range of frequency and parameters $\lambda_1, ..., \lambda_r, \phi_1, ..., \phi_c$. To that end, the response of the system is sampled as follows in order to define the subspace $\boldsymbol{K}$:

$$\mathbf{K} = \begin{bmatrix} real(\boldsymbol{x}_1) & imag(\boldsymbol{x}_1) & ... & real(\boldsymbol{x}_n) & imag(\boldsymbol{x}_n) \end{bmatrix} \tag{4.17}$$

where $\mathbf{X}_i$ are the system responses given by:

$$\boldsymbol{x}_i = (\hat{\boldsymbol{G}}_{all}(\lambda_i) + s\hat{\boldsymbol{C}}_{all}(\phi_i))^{-1}\hat{\boldsymbol{R}} \tag{4.18}$$

$$\hat{\boldsymbol{G}}_{all}(\lambda_i) = \hat{\boldsymbol{G}} + \left(\sum_{i=1}^{r} f_i(\lambda_i)\hat{\boldsymbol{D}}_i\right), \quad \hat{\boldsymbol{C}}_{all}(\phi_i) = \hat{\boldsymbol{C}} + \left(\sum_{j=1}^{c} g_j(\phi_j)\hat{\boldsymbol{E}}_j\right) \tag{4.19}$$

and evaluated at various frequency and parameter points within the range of interest.

Given the nature of the frequency response, the subspace $\boldsymbol{K}$ contains a lot of redundant information. The singular value decomposition (SVD) technique presented in [45], [85] is used to identify the dominant dimensions and obtain an optimal orthonormal basis spanning the system response. Singular value decomposition computes the matrices $\boldsymbol{U}, \boldsymbol{S}, \boldsymbol{V}$ from $\boldsymbol{K}$ such that $\boldsymbol{K} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$ where $\boldsymbol{S}$ is a diagonal matrix with non-negative elements and in decreasing order, and $\boldsymbol{U}$ and $\boldsymbol{V}$ are unitary matrices. The singular values along the diagonal of $\boldsymbol{S}$ measure the relative importance of the corresponding column of $\boldsymbol{U}$. This provides a convenient way to filter out the redundant directions. Taking only the first $N_k$ columns of $\boldsymbol{U}$, those corresponding to the highest singular values, produces the congruence transformation matrix, $\boldsymbol{Q}_s$, which now contains only the dominant directions of the system response. Thus:

$$\boldsymbol{Q}_s = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \dots & \boldsymbol{u}_{N_k} \end{bmatrix} \tag{4.20}$$

where $\boldsymbol{u}_i$ is the $i^{th}$ column of $\boldsymbol{U}$. The order $N_k$ is chosen using the following criteria:

$$\frac{s(N_k, N_k)}{s(1,1)} \leq err \tag{4.21}$$

where $s(1,1)$ and $s(N_k, N_k)$ are the first and the ${N_k}^{th}$ singular values along the diagonal of $\boldsymbol{S}$ respectively. $err$ is the error bound for the dominant subspace. Using this newly produced congruence transformation matrix, $\boldsymbol{Q}_s$, for the reduction in (4.16) produces a compact model valid over a range of values, determined by the sample points, for the parameters and frequency. It is to be noted that such a macromodel obtained from the second level reduction produces a macromodel which is passive by construction as will be proved in the following subsection.

### 4.3.3 Proof of preservation of passivity

By definition, a passive circuit is one that cannot generate more energy than it absorbs. Passivity is an important property for interconnect networks. Non-passive model, even if it is stable, may result in an unstable system when connected to other passive networks [5]. In [87], the necessary and sufficient conditions for the passivity are outlined using admittance matrix $\boldsymbol{Y}(s)$ as follows.

1. $\boldsymbol{Y}(s^*) = \boldsymbol{Y}^*(s)$ for all complex $s$, where $*$ is the complex conjugate operator.
2. $\boldsymbol{Y}(s)$ is a positive matrix, that is, $z^{*^T}(\boldsymbol{Y}(s) + \boldsymbol{Y}^T(s^*))z \geq 0$ for all complex values of $s$ satisfy $\boldsymbol{Re}(s) > 0$ and for any complex vector $z$.

In order to prove that the reduced system from two-level reduction is passive, we first show that the parametric model reduction based on a congruence transformation preserves the passivity of the system. We then demonstrate that two-level reduction in the proposed algorithm is equivalent to the reduction on the original parametric system based on a congruence transformation. Consider the MNA formulation for the original parametric system described in (4.12).

$$\boldsymbol{G}\boldsymbol{x} + s\boldsymbol{C}\boldsymbol{x} + \left(\sum_{i=1}^{r} f_i(\lambda_i)\boldsymbol{D}_i\right)\boldsymbol{x} + s\left(\sum_{j=1}^{c} g_j(\phi_j)\boldsymbol{E}_j\right)\boldsymbol{x} = \boldsymbol{R}\boldsymbol{u}; \quad \boldsymbol{i} = \boldsymbol{R}^T\boldsymbol{x} \tag{4.22}$$

After performing the first level reduction, the reduced model can be written as

$$\hat{\boldsymbol{G}}\hat{\boldsymbol{x}} + s\hat{\boldsymbol{C}}\hat{\boldsymbol{x}} + \left(\sum_{i=1}^{r} f_i(\lambda_i)\hat{\boldsymbol{D}}_i\right)\hat{\boldsymbol{x}} + s\left(\sum_{j=1}^{c} g_j(\phi_j)\hat{\boldsymbol{E}}_j\right)\hat{\boldsymbol{x}} = \hat{\boldsymbol{R}}u; \quad \hat{i} = \hat{\boldsymbol{R}}^T\hat{\boldsymbol{x}} \tag{4.23}$$

where $\hat{\boldsymbol{G}}$, $\hat{\boldsymbol{C}}$, $\hat{\boldsymbol{D}}_i, (i = 1, \cdots, r)$ and $\hat{\boldsymbol{E}}_j, (j = 1, \cdots, c) \in \mathbb{R}^{N_q \times N_q}$ and are given by

$$\hat{\boldsymbol{G}} = \boldsymbol{Q}^T\boldsymbol{G}\boldsymbol{Q}; \quad \hat{\boldsymbol{C}} = \boldsymbol{Q}^T\boldsymbol{C}\boldsymbol{Q}; \quad \hat{\boldsymbol{D}}_i = \boldsymbol{Q}^T\boldsymbol{D}_i\boldsymbol{Q}; \quad \hat{\boldsymbol{E}}_j = \boldsymbol{Q}^T\boldsymbol{E}_j\boldsymbol{Q}; \quad \hat{\boldsymbol{R}} = \boldsymbol{Q}^T\boldsymbol{R} \tag{4.24}$$

and $\boldsymbol{Q} \in \mathbb{R}^{n \times N_q}$ is the congruence transformation matrix for the first level reduction. It has been proved in [77] that the parametric reduced order system in (4.23) is passive by construction.

Next, we demonstrate that two-level reduction using congruence transformation $\boldsymbol{Q} \in \mathbb{R}^{n \times N_q}$ and $\boldsymbol{Q}_s \in \mathbb{R}^{N_q \times N_k}$ is equivalent to reducing the original system using a congruence transformation $\boldsymbol{Q}_{total} \in \mathbb{R}^{n \times N_k}$.

The reduced macromodel obtained from the second level reduction was described in (4.16) as

$$\tilde{\boldsymbol{G}}\tilde{\boldsymbol{x}} + s\tilde{\boldsymbol{C}}\tilde{\boldsymbol{x}} + \left(\sum_{i=1}^{r} f_i(\lambda_i)\tilde{\boldsymbol{D}}_i\right)\tilde{\mathbf{x}} + s\left(\sum_{j=1}^{c} g_j(\phi_j)\tilde{\boldsymbol{E}}_j\right)\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{R}}u; \quad \boldsymbol{i} = \tilde{\boldsymbol{R}}^T\tilde{\boldsymbol{x}} \tag{4.25}$$

where $\tilde{\boldsymbol{G}}, \tilde{\boldsymbol{C}}, \tilde{\boldsymbol{D}}_i, (i = 1, \cdots, r)$ and $\tilde{\boldsymbol{E}}_j, (j = 1, \cdots, c) \in \mathbb{R}^{N_k \times N_k}$ and are given by

$$\tilde{\boldsymbol{G}} = \boldsymbol{Q}_s^T\hat{\boldsymbol{G}}\boldsymbol{Q}_s; \quad \tilde{\boldsymbol{C}} = \boldsymbol{Q}_s^T\hat{\boldsymbol{C}}\boldsymbol{Q}_s; \quad \tilde{\boldsymbol{D}}_i = \boldsymbol{Q}_s^T\hat{\boldsymbol{D}}_i\boldsymbol{Q}_s; \quad \tilde{\boldsymbol{E}}_j = \boldsymbol{Q}_s^T\hat{\boldsymbol{E}}_j\boldsymbol{Q}_s; \quad \tilde{\boldsymbol{R}} = \boldsymbol{Q}_s^T\hat{\boldsymbol{R}} \tag{4.26}$$

and $\boldsymbol{Q}_s \in \mathbb{R}^{N_q \times N_k}$ is the congruence transformation matrix for the second level reduction. Substituting (4.24) in (4.26), we obtain

$$\begin{aligned}
\tilde{\boldsymbol{G}} &= \boldsymbol{Q}_s^T\boldsymbol{Q}^T\boldsymbol{G}\boldsymbol{Q}\boldsymbol{Q}_s \\
\tilde{\boldsymbol{C}} &= \boldsymbol{Q}_s^T\boldsymbol{Q}^T\boldsymbol{C}\boldsymbol{Q}\boldsymbol{Q}_s \\
\tilde{\boldsymbol{D}}_i &= \boldsymbol{Q}_s^T\boldsymbol{Q}^T\boldsymbol{D}_i\boldsymbol{Q}\boldsymbol{Q}_s \\
\tilde{\boldsymbol{E}}_j &= \boldsymbol{Q}_s^T\boldsymbol{Q}^T\boldsymbol{E}_j\boldsymbol{Q}\boldsymbol{Q}_s \\
\tilde{\boldsymbol{R}} &= \boldsymbol{Q}_s^T\boldsymbol{Q}^T\boldsymbol{R}
\end{aligned} \tag{4.27}$$

By setting $\boldsymbol{Q}_{total} = \boldsymbol{Q}\boldsymbol{Q}_s$, the above equations can be rewritten as

$$\begin{aligned} \tilde{\boldsymbol{G}} &= \boldsymbol{Q}_{total}^T \boldsymbol{G} \boldsymbol{Q}_{total} \\ \tilde{\boldsymbol{C}} &= \boldsymbol{Q}_{total}^T \boldsymbol{C} \boldsymbol{Q}_{total} \\ \tilde{\boldsymbol{D}}_i &= \boldsymbol{Q}_{total}^T \boldsymbol{D}_i \boldsymbol{Q}_{total} \\ \tilde{\boldsymbol{E}}_j &= \boldsymbol{Q}_{total}^T \boldsymbol{E}_j \boldsymbol{Q}_{total} \\ \tilde{\boldsymbol{R}} &= \boldsymbol{Q}_{total}^T \boldsymbol{R} \end{aligned} \tag{4.28}$$

It follows from (4.28) that the reduced macromodel based on two congruence transformations $\boldsymbol{Q}$ and $\boldsymbol{Q}_s$ is equivalent to a macromodel based on one congruence transformation $\boldsymbol{Q}_{total}$. Therefore the final reduced order macromodel obtained from two-level reduction is passive by construction.

### 4.3.4 Results

#### Example 1

The first example is an interconnect network with 9 non-coupled transmission lines, 1 nine-coupled transmission line system (see [88] for a description of the 9 coupled line system), and 30 resistors. This system is a parametric 2-port network with 3 resistor values, $R_1, R_2, R_3$ as parameters. The resulting size of the MNA matrices is 3522.

Using the traditional parametric method described in Section 4.3.1 for the first level of reduction, the resulting reduced system was size 360. After applying the second level of reduction, the size of the reduced macromodel became 110. Table 4.5 shows the size comparison of the original network (original system), macromodel obtained from traditional parametric model order reduction (traditional PMR) and macromodel obtained from multi-level parametric model order reduction (multi-level PMR). The reduced model was tested over the frequency and parameter range of interest and the results matched the original system within an error of 0.04%. The $Y_{2,2}$ parameters of three sample cases (see Table 4.3 for the parameter values) are shown in Fig. 4.11.

**Table 4.3** Three Sample Test Cases for Example 1

| | $R_1$ | $R_2$ | $R_3$ |
|---|---|---|---|
| Case 1 | 102Ω | 230Ω | 34Ω |
| Case 2 | 57Ω | 84Ω | 153Ω |
| Case 3 | 151Ω | 44Ω | 229Ω |

**Table 4.4** Three Sample Test Cases for Example 2

| | $R_1$ | $R_2$ | $C_1$ | $C_2$ |
|---|---|---|---|---|
| Case 1 | 211Ω | 168Ω | 1.70pF | 1.25pF |
| Case 2 | 131Ω | 209Ω | 0.94pF | 1.77pF |
| Case 3 | 50Ω | 4Ω | 2.07pF | 1.07pF |

The transient responses for the three sample cases are shown in Fig. 4.12. A summary of CPU cost comparison to obtain the transient responses is shown in Table 4.6.

### Example 2

The second example is another interconnect network, this time with 3 ports and parametric with respect to 2 resistor values, $R_1, R_2$, and 2 capacitor values, $C_1, C_2$. It contains 12 transmission lines, 145 capacitors, 138 inductors, and 175 resistors. The resulting size of the MNA matrices is 1989.

The first level reduction used the traditional parametric modified Krylov technique outlined in Section 4.3.1 and created a reduced model of size 300. Applying the second level of reduction created a smaller macromodel, having a size of 101. Table 4.5 compares the sizes of the original system (original system), macromodel obtained from traditional parametric model order reduction (traditional PMR) and macromodel obtained from multi-level parametric model order reduction (multi-level PMR). Again, the reduced model was tested over the range of interest for parameters and frequency and it matched the original within an error of 0.03%. Three sample

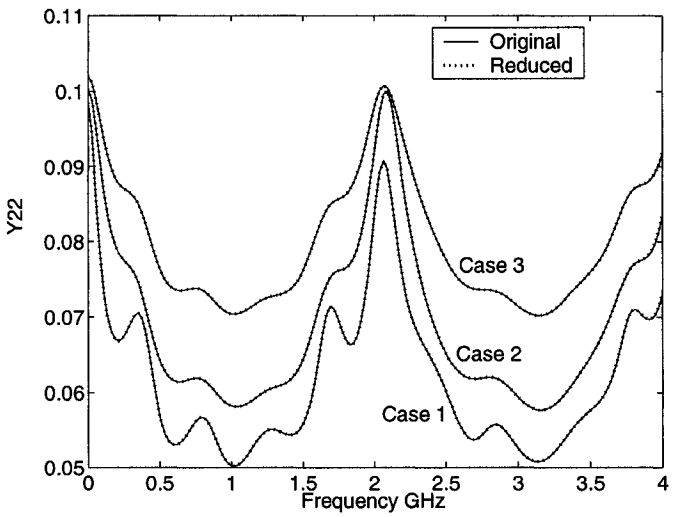


**Fig. 4.11** $Y_{22}$ parameters of three sampling test cases for example 1 (comparison between the original system and the proposed macromodel)

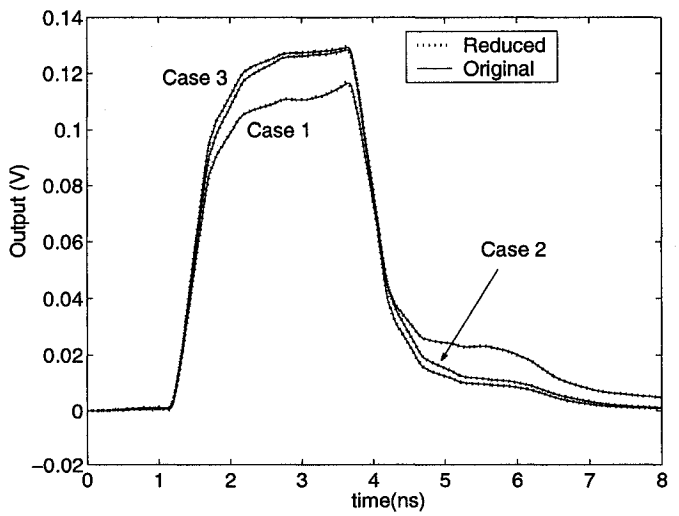


**Fig. 4.12** Transient responses of three sampling test cases for example 1 (comparison between the original system and the proposed macromodel)

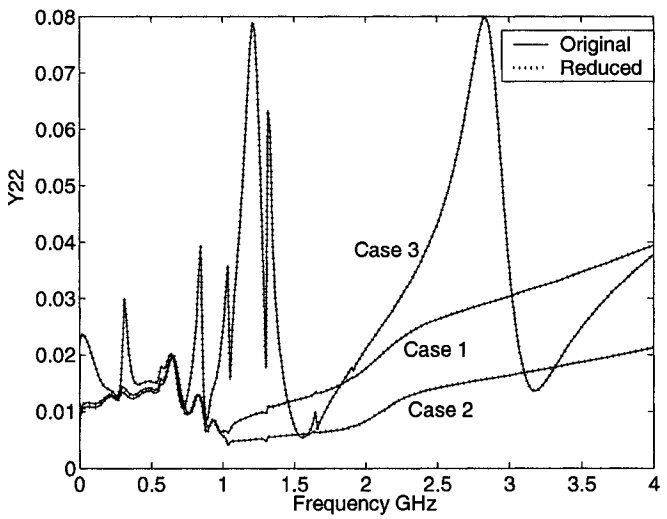


**Fig. 4.13** $Y_{22}$ of three sampling test cases for example 2 (comparison between the original system and the proposed macromodel)

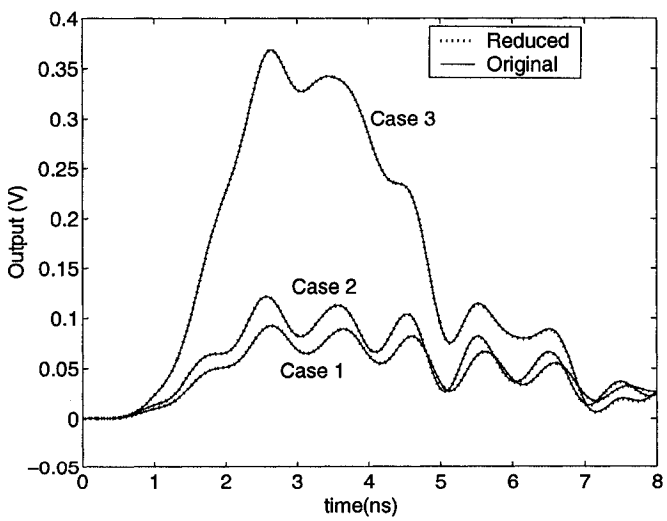


**Fig. 4.14** Transient responses of three sampling test cases for example 2 (comparison between the original system and the proposed macromodel)

**Table 4.5** Macromodel Size Comparison

| | Example 1 | Example 2 |
|---|---|---|
| Original system | 3522 | 1989 |
| Traditional PMR | 360 | 300 |
| Multi-level PMR | 110 | 101 |

**Table 4.6** CPU Cost Comparison for Example1 and Example2

| | Example1 | Example2 |
|---|---|---|
| Traditional PMR | 493.8s | 339.6s |
| Multi-level PMR | 16.8s | 24.1s |
| Speed-up | 29.4 | 14.1 |

cases (see Table 4.4 for the parameter values) of the $Y_{2,2}$ parameters are shown in Fig. 4.13. The transient responses for the three cases are also shown in Fig. 4.14. A summary of CPU cost comparison for example2 to obtain the transient responses is shown in Table 4.6.

## 4.4 Sparse Parametric Model Order Reduction

The size of the parametric macromodel obtained from parametric multi-level reduction discussed in the previous section is significantly smaller than the size of the original system, however, this reduced macromodel is generally dense. A dense macromodel significantly reduces the efficiency of the simulation. This problem was addressed in [75] for regular model order reduction methods. However, this approach is not applicable to parametric macromodels. In this section, a new technique is presented to produce a reduced order parametric macromodel which is sparse [49]. In order to achieve this goal, a new parametric formulation is proposed. This formulation replaces the stamps of the parametric elements with constrained port representations. The reduction and sparsification are performed on the macromodel in the space of the new formulation scheme, and the sparse reduced macromodel is then brought back to the traditional representation with-

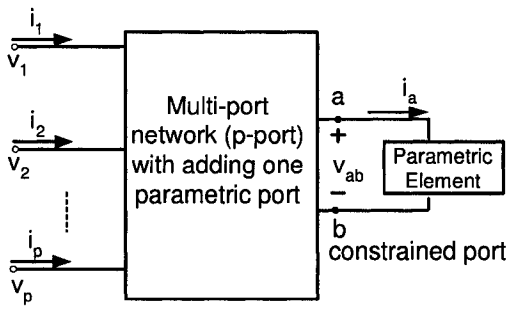


**Fig. 4.15** Adding a constrained port

out losing sparsity. It is to be noted that the newly introduced parametric ports do not have a significant impact on the size of the reduced macromodel. A significant CPU cost saving over traditional parametric model order reduction methods is however achieved due to the sparsity of the macromodel.

### 4.4.1 New formulation suitable for sparsification

As can be seen in Section 4.3, the parametric elements, represented by $f_{i(\lambda_i)}\boldsymbol{D}_i$ and $g_{j(\phi_j)}\boldsymbol{E}_j$ in (4.12) hinder the use of the procedure to sparse macromodels elaborated in [75]. In the first step of the proposed method, these elements are decoupled from the original formulation by adding constrained ports.

For instance, assume that (4.12) only contains one parametric element and it is connected to the node "a" and node "b" of the system. Then, one additional constrained port is added into the original multi-port network as shown in Fig. 4.15. Therefore (4.12) becomes

$$\boldsymbol{G}_1\boldsymbol{x}_1 + s\boldsymbol{C}_1\boldsymbol{x}_1 = \boldsymbol{R}_1\boldsymbol{u} + \boldsymbol{R}_2\boldsymbol{u}_2 \tag{4.29a}$$

$$\boldsymbol{i} = \boldsymbol{L}_1^T\boldsymbol{x}_1 \tag{4.29b}$$

$$\boldsymbol{f}_1(\boldsymbol{u}_2) = \boldsymbol{L}_2^T\boldsymbol{x}_1 \tag{4.29c}$$

where

$$\boldsymbol{G}_1 = \begin{bmatrix} \boldsymbol{G} & \boldsymbol{d}_a \\ -\boldsymbol{d}_a^T & 0 \end{bmatrix}, \boldsymbol{C}_1 = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{0} & 0 \end{bmatrix} \tag{4.30}$$

$$\boldsymbol{R}_1 = \begin{bmatrix} \boldsymbol{R} \\ 0 \end{bmatrix}, \boldsymbol{R}_2 = \begin{bmatrix} \boldsymbol{0} \\ -1 \end{bmatrix}$$

$$\boldsymbol{L}_1 = \begin{bmatrix} \boldsymbol{L} \\ 0 \end{bmatrix}, \boldsymbol{L}_2 = \begin{bmatrix} \boldsymbol{0} \\ -1 \end{bmatrix} \tag{4.31}$$

$$\boldsymbol{u}_2 = \begin{bmatrix} \boldsymbol{v}_{ab} \end{bmatrix}, \boldsymbol{x}_1 = \begin{bmatrix} \boldsymbol{x} \\ i_a \end{bmatrix} \tag{4.32}$$

and $v_{ab}$ and $i_a$ are the voltage and current on the newly introduced constrained port. $\boldsymbol{d}_a$ is a vector with all zeros elements except for the rows that corresponds to node "a" and "b". Equation (4.29c) represents the parametric constraint on the newly appended constrained port. Notice that (4.29) denotes the same network as (4.12) but with the parametric element decoupled and a port at node "a" and "b" added. This port is a constrained port since the type of load (memory or memoryless) that will be attached is known – it is the same as that of the removed element. This load information is used to parametrize the port as will be shown in the following section.

Similarly, if (4.12) contains $w$ parametric elements, then multiple constrained ports are added with each port connecting to one parametric element. In general, the formulation becomes

$$\boldsymbol{G}_w\boldsymbol{x}_w + s\boldsymbol{C}_w\boldsymbol{x}_w = \boldsymbol{R}_{w1}\boldsymbol{u} + \boldsymbol{R}_{w2}\boldsymbol{u}_{w2} \tag{4.33a}$$

$$\boldsymbol{i} = \boldsymbol{L}_{w1}^T\boldsymbol{x}_w \tag{4.33b}$$

$$\boldsymbol{f}_w(\boldsymbol{u}_{w2}) = \boldsymbol{L}_{w2}^T\boldsymbol{x}_w \tag{4.33c}$$

where

$$\boldsymbol{G}_w = \begin{bmatrix} \boldsymbol{G} & \boldsymbol{D}_\pi \\ -\boldsymbol{D}_\pi^T & \boldsymbol{0} \end{bmatrix}, \boldsymbol{C}_w = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \tag{4.34a}$$

$$\boldsymbol{R}_{w1} = \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix}, \boldsymbol{R}_{w2} = \begin{bmatrix} \boldsymbol{0} \\ -\boldsymbol{I} \end{bmatrix} \tag{4.34b}$$

$$\boldsymbol{L}_{w1} = \begin{bmatrix} \boldsymbol{L} \\ \boldsymbol{0} \end{bmatrix}, \boldsymbol{L}_{w2} = \begin{bmatrix} \boldsymbol{0} \\ -\boldsymbol{I} \end{bmatrix} \tag{4.34c}$$

$$\boldsymbol{u}_{w2} = \begin{bmatrix} \boldsymbol{v}_\pi \end{bmatrix}, \boldsymbol{x}_w = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{i}_\pi \end{bmatrix} \tag{4.34d}$$

and $\boldsymbol{v}_\pi$ and $\boldsymbol{i}_\pi$ are the voltages and currents respectively at newly introduced constrained ports. $\boldsymbol{D}_\pi$ is a matrix with all zero elements except for the rows and columns that correspond to the parametric nodes, and $\boldsymbol{I}$ is the identity matrix. Again, formulation (4.33a) does not contain any parametric elements, but instead contains constrained ports. In other words, the constrained port voltages and currents are not arbitrary but are subject to the parametric condition in (4.33c).

### 4.4.2 Model reduction with constrained ports

Having the modified formulation in (4.33), the reduced system is then obtained by using congruence transformation, resulting in

$$\hat{\boldsymbol{G}}_w \hat{\boldsymbol{x}}_w + s\hat{\boldsymbol{C}}_w \hat{\boldsymbol{x}}_w = \hat{\boldsymbol{R}}_{w1}\boldsymbol{u} + \hat{\boldsymbol{R}}_{w2}\hat{\boldsymbol{u}}_{w2} \tag{4.35a}$$

$$\boldsymbol{i} = \hat{\boldsymbol{L}}_{w1}^T \hat{\boldsymbol{x}}_w \tag{4.35b}$$

$$\boldsymbol{f}_w(\hat{\boldsymbol{u}}_{w2}) = \hat{\boldsymbol{L}}_{w2}^T \hat{\boldsymbol{x}}_w \tag{4.35c}$$

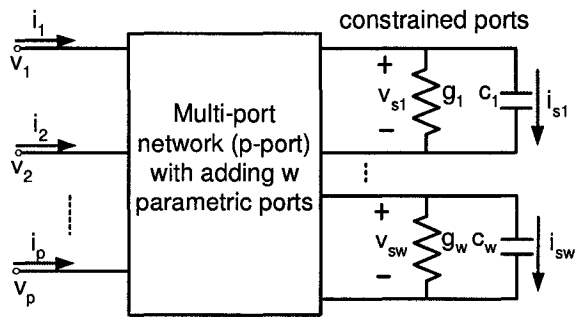


**Fig. 4.16** Adding w constrained ports

where

$$\hat{G}_w = Q^T G_w Q; \quad \hat{C}_w = Q^T C_w Q \tag{4.36a}$$

$$\hat{R}_{w1} = Q^T R_{w1}; \quad \hat{R}_{w2} = Q^T R_{w2} \tag{4.36b}$$

$$\hat{L}_{w1} = Q^T L_{w1}; \quad \hat{L}_{w2} = Q^T L_{w2} \tag{4.36c}$$

The congruence transformation matrix, $Q$, is chosen such that the size of the reduced system in (4.35) is not affected by the introduced constrained ports by exploiting the information on those ports. Since the type of elements (resistive or capacitive elements) on the constrained ports are known, the congruence transformation matrix can be obtained by taking the parametric elements on the ports as parameters and using parametric port formulation. To obtain such new formulation, consider a $p$ port parametric system with $w$ constrained ports as described in (4.33). The parametric elements are resistive and capacitive as shown in Fig. 4.16. In this case, the parametric condition in (4.33c) becomes

$$i_{s1} = (g_1 + sc_1)v_{s1} \tag{4.37a}$$

$$\vdots$$

$$i_{sw} = (g_{sw} + sc_{sw})v_{sw} \tag{4.37b}$$

Incorporating the constraint equations in (4.37) into (4.33a), we obtain the parametric port formulation

$$\check{\boldsymbol{G}}_w \boldsymbol{x}_w + s\boldsymbol{C}_w \boldsymbol{x}_w + \left(\sum_{i=1}^{w} g_i \boldsymbol{F}_i\right) \boldsymbol{x}_w + s\left(\sum_{i=1}^{w} c_i \boldsymbol{F}_i\right) \boldsymbol{x}_w = \boldsymbol{R}_{w1} \boldsymbol{u} \tag{4.38}$$

where $\check{\boldsymbol{G}}_w$ is obtained by modifying $w$ rows of $\boldsymbol{G}_w$ matrix with one non-zero in each row, corresponding to the constrained port current. The entries of $\boldsymbol{F}_i$ are no more than two non-zeros, corresponding to the two node voltages on the $i^{th}$ constrained port. The parameters $c_i$ and $g_i$ represent the capacitive and resistive elements connected to the constrained ports. The detailed process of constructing the above parametric port formulation can be found in Section 3.3. Note that the unknown variables in (4.38) is the same as those in (4.33a). In other words, embedding the parametric elements into the system equations is done in such a way that unknown variables do not change. It is also important to note that the new formulation in (4.38) is multi-dimensional with $g_i$ and $c_i$, $i = 1, \cdots, w$ as parameters. Parametric model order reduction techniques are therefore used to obtain the subspace $\boldsymbol{Q}$. This subspace spans the moments with respect to frequency, moments with respect to parameters as well as some cross moments. The subspace denoted as $\boldsymbol{Q}_s$ spanning the moments with respect to frequency can be implicitly computed using the Arnoldi process. The cross moments $(\boldsymbol{m}_{sc_i})_{k,j}$ with respect to frequency and the parameters $c_i$ can be computed by the procedure elaborated in [31], [82]. The corresponding subspace denoted by $\boldsymbol{Q}_{sc_i}$, $i = 1, \ldots w$, is obtained by performing a standard QR decomposition [74] on the cross moments. The subspace $\boldsymbol{Q}_{g_1}$, $\boldsymbol{Q}_{g_2}$, $\cdots$, $\boldsymbol{Q}_{g_w}$ spanning the moments with respect to conductance $g_1, g_2, \cdots, g_w$ is implicitly computed using the Arnoldi process.

Once all the required subspaces are evaluated, the resulting multidimensional subspace denoted by $\boldsymbol{Q}$ is as follows:

$$\text{colsp}[\boldsymbol{Q}] = \text{colsp}[\boldsymbol{Q}_s, \boldsymbol{Q}_{sc_1}, \boldsymbol{Q}_{sc_2}, \cdots, \boldsymbol{Q}_{sc_N}, \boldsymbol{Q}_{g_1}, \boldsymbol{Q}_{g_2}, \cdots, \boldsymbol{Q}_{g_N}] \tag{4.39}$$

This can be achieved by using another standard QR decomposition [74] on various required

subspaces in (4.39).

### 4.4.3 Sparsification of the macromodel

The reduced order macromodel in (4.35), while much smaller than the original system, is generally very dense. This significantly reduces the CPU efficiency of the simulation. Fortunately the macromodel is now expressed in a traditional state space form and the contributions of the parameters are expressed as port conditions. This allows us to use standard diagonalization schemes to make the reduced system sparse [75]. Equations in (4.35a) can be reformulated by premultiplying $\hat{\boldsymbol{G}}_w^{-1}$

$$\hat{\boldsymbol{x}}_w + s\hat{\boldsymbol{G}}_w^{-1}\hat{\boldsymbol{C}}_w\hat{\boldsymbol{x}}_w = \hat{\boldsymbol{G}}_w^{-1}\hat{\boldsymbol{R}}_{w1}\boldsymbol{u} + \tilde{\boldsymbol{G}}_w^{-1}\hat{\boldsymbol{R}}_{w2}\tilde{\boldsymbol{u}}_{w2} \tag{4.40}$$

By applying eigen-decomposition to $\hat{\boldsymbol{G}}_w^{-1}\hat{\boldsymbol{C}}_w$, we obtain

$$\hat{\boldsymbol{G}}_w^{-1}\hat{\boldsymbol{C}}_w = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{-1} \tag{4.41}$$

$\boldsymbol{V}$ and $\boldsymbol{D}$ are in general complex; however a real diagonalization can be derived from (4.41) in the form

$$\hat{\boldsymbol{G}}_w^{-1}\hat{\boldsymbol{C}}_w = \boldsymbol{S}_r\boldsymbol{D}_r{\boldsymbol{S}_r}^{-1} \tag{4.42}$$

where $\boldsymbol{D}_r$ is a real block diagonal matrix and

$$\boldsymbol{S}_r = \boldsymbol{V}\boldsymbol{P}^{-1} \tag{4.43}$$

$$\boldsymbol{D}_r = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1} \tag{4.44}$$

$\boldsymbol{P}$ is defined as [75]:

1. when the corresponding element of the diagonal matrix $\boldsymbol{D}$, $d_{i,i}$, is a real, the element of $\boldsymbol{P}$ matrix is $p_{i,i} = 1$,

2. when the corresponding elements of the diagonal matrix $\boldsymbol{D}$, $d_{i,i}$ and $d_{i+1,i+1}$ are complex

conjugate, the elements of $\boldsymbol{P}$ matrix are: $p_{i:i+1,i:i+1} = \begin{bmatrix} 1 & 1 \\ j & -j \end{bmatrix}$

By substituting (4.42) into (4.40), the sparse reduced system becomes

$$\breve{\boldsymbol{G}}_w \breve{\boldsymbol{x}}_w + s\breve{\boldsymbol{C}}_w \breve{\boldsymbol{x}}_w = \breve{\boldsymbol{R}}_{w1}\boldsymbol{u} + \breve{\boldsymbol{R}}_{w2}\breve{\boldsymbol{u}}_2 \tag{4.45a}$$

$$\boldsymbol{i} = \breve{\boldsymbol{L}}_{w1}^T \breve{\boldsymbol{x}}_w \tag{4.45b}$$

$$\boldsymbol{f}_w(\breve{\boldsymbol{u}}_{w2}) = \breve{\boldsymbol{L}}_{w2}^T \breve{\boldsymbol{x}}_w \tag{4.45c}$$

where $\breve{\boldsymbol{G}}_w$ is an identity matrix,

$$\breve{\boldsymbol{C}}_w = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1} \tag{4.46a}$$

$$\breve{\boldsymbol{R}}_{w1} = (\boldsymbol{V}\boldsymbol{P}^{-1})^{-1}\hat{\boldsymbol{G}}_w^{-1}\hat{\boldsymbol{R}}_{w1} \tag{4.46b}$$

$$\breve{\boldsymbol{R}}_{w2} = (\boldsymbol{V}\boldsymbol{P}^{-1})^{-1}\hat{\boldsymbol{G}}_w^{-1}\hat{\boldsymbol{R}}_{w2} \tag{4.46c}$$

$$\breve{\boldsymbol{L}}_{w1} = (\boldsymbol{V}\boldsymbol{P}^{-1})^{T}\hat{\boldsymbol{L}}_{w1} \tag{4.46d}$$

$$\breve{\boldsymbol{L}}_{w2} = (\boldsymbol{V}\boldsymbol{P}^{-1})^{T}\hat{\boldsymbol{L}}_{w2} \tag{4.46e}$$

$$\breve{\boldsymbol{x}}_w = (\boldsymbol{V}\boldsymbol{P}^{-1})^{-1}\hat{\boldsymbol{x}}_w \tag{4.46f}$$

and $\breve{\boldsymbol{G}}_w$ and $\breve{\boldsymbol{C}}_w$ are small yet sparse matrices. Note that this sparse model, similar to (4.35) also retains the constraints on the loads attached to the constrained ports.

### 4.4.4 Sparse reduced order macromodel

The final step is to reincorporate the nonlinear constraints in (4.45c) into the overall sparse macromodel equations in (4.45a). In order to achieve this, the reduced macromodel is treated as a subsection, which is represented by (4.45a) and (4.45b), then the parametric elements, which were removed before are connected back to the constrained ports of the subsection to form the sparse

reduced parametric macromodel.

$$\left(\begin{bmatrix} \boldsymbol{G}_\phi & \boldsymbol{D}\breve{\boldsymbol{L}}_w^T \\ -\breve{\boldsymbol{R}}_w\boldsymbol{D}^T & \breve{\boldsymbol{G}}_w \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^{r} f_{i(\lambda_i)}\breve{\boldsymbol{D}}_i \\ 0 \end{bmatrix} + s\begin{bmatrix} \boldsymbol{C}_\phi & 0 \\ 0 & \breve{\boldsymbol{C}}_w \end{bmatrix} + s\begin{bmatrix} \sum_{j=1}^{r} g_{j(\phi_i)}\breve{\boldsymbol{E}}_j \\ 0 \end{bmatrix}\right)\begin{bmatrix} \boldsymbol{x}_\phi \\ \breve{\boldsymbol{x}}_w \end{bmatrix} = \begin{bmatrix} \boldsymbol{R}_\phi \\ 0 \end{bmatrix}\boldsymbol{u} \tag{4.47}$$

where

$$\breve{\boldsymbol{L}}_w^T = \begin{bmatrix} \breve{\boldsymbol{L}}_{w1}^T \\ \breve{\boldsymbol{L}}_{w2}^T \end{bmatrix} \tag{4.48}$$

$$\breve{\boldsymbol{R}}_w = \begin{bmatrix} \breve{\boldsymbol{R}}_{w1} & \breve{\boldsymbol{R}}_{w2} \end{bmatrix} \tag{4.49}$$

$\boldsymbol{G}_\phi$ and $\boldsymbol{C}_\phi$ represent the general port formulation of unconstrained ports. $\boldsymbol{R}_\phi$ is a selector matrix. $\sum_{i=1}^{r} f_{i(\lambda_i)}\breve{\boldsymbol{D}}_i$ represents the parametric memoryless elements. $\sum_{j=1}^{r} g_{j(\phi_i)}\breve{\boldsymbol{E}}_j$ represents the parametric memory elements. $\boldsymbol{D}$ is a selector matrix with elements $d_{i,j} \in \{0, 1\}$ and a maximum of one non-zero in each row or column that maps the vectors of port voltages and port currents entering subsection into the node space of the network. $\breve{\boldsymbol{x}}_w$ is the unknown vectors of the reduced order macromodel in (4.45). In summary, the final reduced order macromodel based on (4.47) is in the form of

$$\left[\bar{\boldsymbol{G}} + s\bar{\boldsymbol{C}} + \sum_{i=1}^{r} f_{i(\lambda_i)}\bar{\boldsymbol{D}}_i + s\sum_{j=1}^{c} g_{j(\phi_j)}\bar{\boldsymbol{E}}_j\right]\bar{\boldsymbol{x}} = \bar{\boldsymbol{R}}\boldsymbol{u} \tag{4.50a}$$

$$\boldsymbol{i} = \bar{\boldsymbol{L}}^T\boldsymbol{x} \tag{4.50b}$$

where $\bar{\boldsymbol{G}}, \bar{\boldsymbol{C}}, \bar{\boldsymbol{D}}_i$, and $\bar{\boldsymbol{E}}_j$ are small yet sparse matrices.

### 4.4.5 Numerical results

#### Example 1

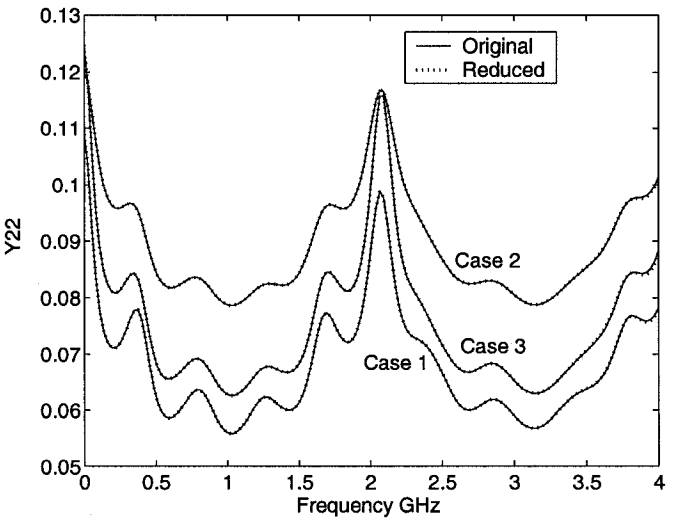


**Fig. 4.17** Frequency responses for example 1 (comparison between the original system and the proposed macromodel)

The first example is an interconnect network with 9 non-coupled transmission lines, 1 nine-coupled transmission line system (see [88] for a description of the 9 coupled line system), and 31 resistors. This system is a parametric 2-port network with 4 resistor values, $R_1, R_2, R_3$, and $R_4$, as parameters. The resulting size of the MNA matrices is 3522.

**Table 4.7** Method Comparison for Example 1

| | Size | G Density (%) | C Density (%) | Time (s) | Speed-up |
|---|---|---|---|---|---|
| Original System | 3522 | 0.0756 | 0.0866 | 2906.4 | - |
| Traditional PMR | 372 | 99.9 | 99.9 | 383.41 | 7.6 |
| Sparse PMR | 366 | 3.43 | 0.479 | 36.84 | 78.9 |

Using the sparse macromodeling technique proposed in this section, the size of the reduced

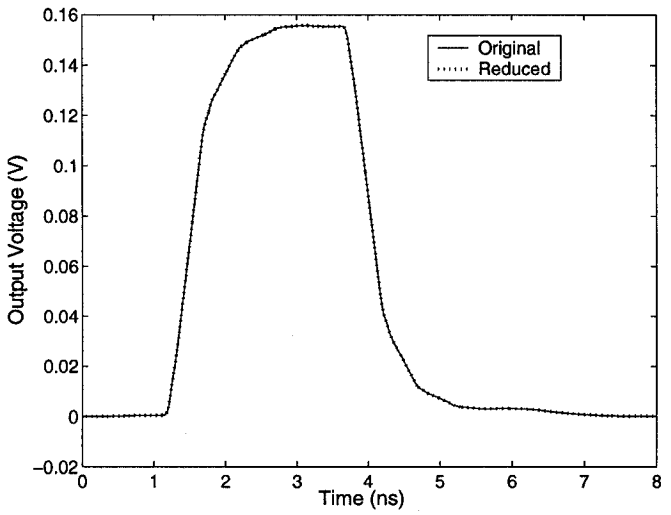


**Fig. 4.18** Transient response for example 1 (comparison between the original system and the proposed macromodel)

model is 366, with $\bar{G}$ only 3.43% filled, and $\bar{C}$ only 0.479% filled. Note also that the other matrices ($\bar{R}, \bar{L}, ...$) in this proposed model contain less than 4 elements whereas using the traditional approach, these other matrices are over 98.9% filled. Three sample cases of $Y_{2,2}$ parameters (with resistor values randomly chosen from $1 \sim 200\Omega$) are shown in Fig. 4.17. As can be seen from the figure, the results match very well (with average error 0.07%). Fig. 4.18 shows the transient responses due to a $1V$ pulse with a rise time of $0.5ns$ and a pulse width of $2ns$. As expected, it is also accurate. A summary of CPU cost comparison between the original network, traditional parametric model order reduction (Traditional PMR), and the proposed method (Sparse PMR) to obtain transient responses is shown in Table 4.7. It is not surprising that the proposed approach can achieve 78.9 CPU speed-up while the traditional parametric model order reduction is only 7.6 speed-up, since the reduced model is not only small but also sparse.

### Example 2

The second example is another interconnect network, which consists of 9 non-coupled transmission lines, 2 nine-coupled transmission line systems (see [88] for a description of the 9 coupled

line system), 3 capacitors, and 29 resistors. The MNA size of this 2-port system is 6223. This system is parametric with respect to 2 resistor values and 2 capacitor values, $R_1$, $R_2$, $C_1$, and $C_2$.

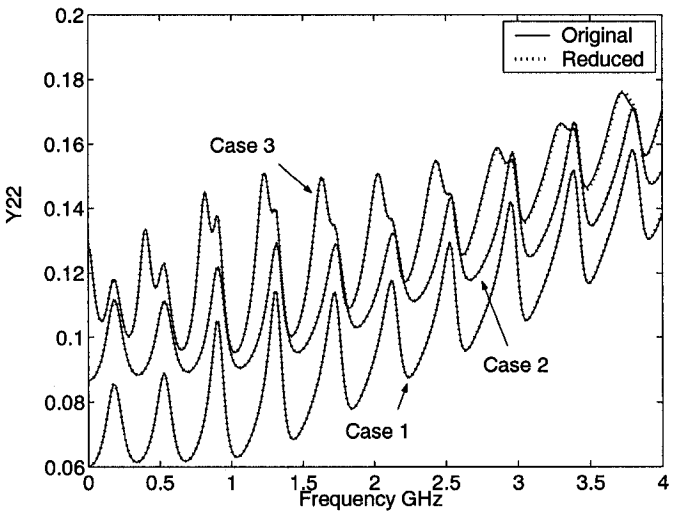


**Fig. 4.19** Frequency responses for example 2 (comparison between the original system and the proposed macromodel)

**Table 4.8** Method Comparison for Example 2

| | Size | G Density (%) | C Density (%) | Time (s) | Speed-up (%) |
|---|---|---|---|---|---|
| Original System | 6223 | 0.0428 | 0.0541 | 3272.1 | - |
| Traditional PMR | 602 | 99.9 | 99.9 | 1848.8 | 1.8 |
| Sparse PMR | 628 | 1.72 | 0.299 | 91.04 | 36.0 |

Applying the sparse parametric reduction procedure results in a reduced macromodel with size 628, where only 1.72% of $\bar{G}$ contains non-zero elements, and only 0.299% for $\bar{C}$. The other matrices ($\bar{R}, \bar{L}, ...$) contain less than 4 elements whereas using the traditional approach produces matrices which are over 99.9% filled. Three sample cases of the $Y_{2,2}$ parameters (with the values of resistor and capacitor randomly chosen from $1 \sim 200\Omega$ and $0.5 \sim 5$ pF respectively) are shown in Fig. 4.19. As can be seen, it is accurate with the average error 0.1%. Fig. 4.20

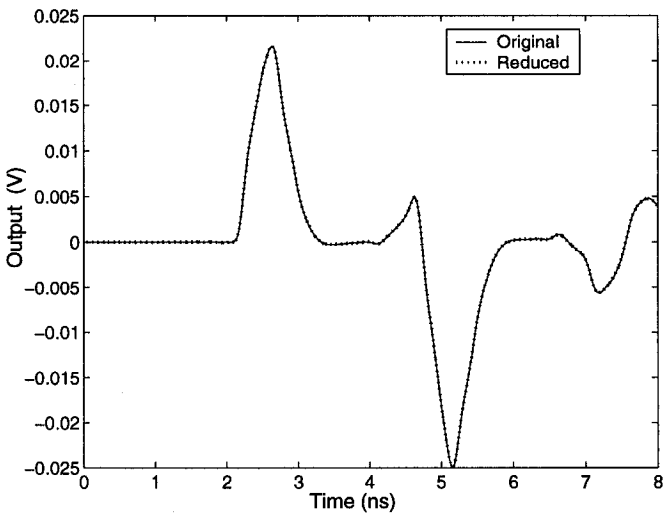


**Fig. 4.20** Transient response for example 2 (comparison between the original system and the proposed macromodel)

shows the transient responses due to a $1V$ pulse with a rise time of $0.5ns$ and a pulse width of $2ns$. As expected, it is also accurate. A summary of CPU cost comparison between the original network, the traditional parametric model order reduction, and the proposed method to obtain transient responses for example2 is shown in Table 4.8. In the example2, The CPU speed-up of the proposed model is 36.0 compared to 1.8 if using traditional parametric model order reduction, which again demonstrates the efficiency and accuracy of the proposed model.

## 4.5 Sparse Multi-level Parametric Model Order Reduction

### 4.5.1 Proposed algorithm

In this section, a sparse two-level parametric model order reduction (PMR) is proposed. This method is based on combining the multi-level PMR in section 4.3 with the sparse PMR in section 4.4. The proposed algorithm consists of five main steps as follows

1. In the first step, the parametric elements are decoupled from the original parametric formulation in (4.12) by adding constrained ports. The resulting formulation is a traditional state

space form as described in (4.33).

$$\boldsymbol{G}_w\boldsymbol{x}_w + s\boldsymbol{C}_w\boldsymbol{x}_w = \boldsymbol{R}_{w1}\boldsymbol{u} + \boldsymbol{R}_{w2}\boldsymbol{u}_{w2} \tag{4.51a}$$

$$\boldsymbol{i} = \boldsymbol{L}_{w1}^T\boldsymbol{x}_w \tag{4.51b}$$

$$\boldsymbol{f}_w(\boldsymbol{u}_{w2}) = \boldsymbol{L}_{w2}^T\boldsymbol{x}_w \tag{4.51c}$$

In the new formulation, the parametric element contributions appear as port constraints in (4.51c).

2. In the second step, first level reduction based on congruence transformation is performed on the new formulation in (4.33) to obtain the reduced model as described in (4.35).

$$\hat{\boldsymbol{G}}_w\hat{\boldsymbol{x}}_w + s\hat{\boldsymbol{C}}_w\hat{\boldsymbol{x}}_w = \hat{\boldsymbol{R}}_{w1}\boldsymbol{u} + \hat{\boldsymbol{R}}_{w2}\hat{\boldsymbol{u}}_{w2} \tag{4.52a}$$

$$\boldsymbol{i} = \hat{\boldsymbol{L}}_{w1}^T\hat{\boldsymbol{x}}_w \tag{4.52b}$$

$$\boldsymbol{f}_w(\hat{\boldsymbol{u}}_{w2}) = \hat{\boldsymbol{L}}_{w2}^T\hat{\boldsymbol{x}}_w \tag{4.52c}$$

where

$$\hat{\boldsymbol{G}}_w = \boldsymbol{Q}^T\boldsymbol{G}_w\boldsymbol{Q};\quad \hat{\boldsymbol{C}}_w = \boldsymbol{Q}^T\boldsymbol{C}_w\boldsymbol{Q};\quad \hat{\boldsymbol{R}}_{w1} = \boldsymbol{Q}^T\boldsymbol{R}_{w1} \tag{4.53a}$$

$$\hat{\boldsymbol{R}}_{w2} = \boldsymbol{Q}^T\boldsymbol{R}_{w2};\quad \hat{\boldsymbol{L}}_{w1} = \boldsymbol{Q}^T\boldsymbol{L}_{w1};\quad \hat{\boldsymbol{L}}_{w2} = \boldsymbol{Q}^T\boldsymbol{L}_{w2} \tag{4.53b}$$

The congruence transformation matrix $\boldsymbol{Q}$ is chosen such that the size of reduction subspace is not affected by the newly introduced constrained ports. This is achieved by taking the parametric elements as the parameters and using parametric prot formulation in (4.38). The details about model order reduction with constrained ports are discussed in section 4.4.2.

3. In the third step, second level reduction based on singular value decomposition is used to

further reduce the macromodel. The resulting macromodel is given by

$$\tilde{\boldsymbol{G}}_w\tilde{\boldsymbol{x}}_w + s\tilde{\boldsymbol{C}}_w\tilde{\boldsymbol{x}}_w = \tilde{\boldsymbol{R}}_{w1}\boldsymbol{u} + \tilde{\boldsymbol{R}}_{w2}\tilde{\boldsymbol{u}}_{w2} \tag{4.54a}$$

$$\boldsymbol{i} = \tilde{\boldsymbol{L}}_{w1}^T\tilde{\boldsymbol{x}}_w \tag{4.54b}$$

$$\boldsymbol{f}_w(\tilde{\boldsymbol{u}}_{w2}) = \tilde{\boldsymbol{L}}_{w2}^T\tilde{\boldsymbol{x}}_w \tag{4.54c}$$

where

$$\tilde{\boldsymbol{G}}_w = \boldsymbol{Q}_s^T\hat{\boldsymbol{G}}_w\boldsymbol{Q}_s; \quad \tilde{\boldsymbol{C}}_w = \boldsymbol{Q}_s^T\hat{\boldsymbol{C}}_w\boldsymbol{Q}_s; \quad \tilde{\boldsymbol{R}}_{w1} = \boldsymbol{Q}_s^T\hat{\boldsymbol{R}}_{w1} \tag{4.55a}$$

$$\tilde{\boldsymbol{R}}_{w2} = \boldsymbol{Q}_s^T\hat{\boldsymbol{R}}_{w2}; \quad \tilde{\boldsymbol{L}}_{w1} = \boldsymbol{Q}_s^T\hat{\boldsymbol{L}}_{w1}; \quad \tilde{\boldsymbol{L}}_{w2} = \boldsymbol{Q}_s^T\hat{\boldsymbol{L}}_{w2} \tag{4.55b}$$

The reduction subspace $\boldsymbol{Q}_s^T$ spans the dominant direction of the system responses in (4.52) over the range of frequency and parameters as presented in section 4.3.2.

4. In the fourth step, the reduced macromodel obtained from two-level reduction in (4.54) is sparsified as discussed in section 4.4.3.

$$\breve{\boldsymbol{G}}_w\breve{\boldsymbol{x}}_w + s\breve{\boldsymbol{C}}_w\breve{\boldsymbol{x}}_w = \breve{\boldsymbol{R}}_{w1}\boldsymbol{u} + \breve{\boldsymbol{R}}_{w2}\breve{\boldsymbol{u}}_2 \tag{4.56a}$$

$$\boldsymbol{i} = \breve{\boldsymbol{L}}_{w1}^T\breve{\boldsymbol{x}}_w \tag{4.56b}$$

$$\boldsymbol{f}_w(\breve{\boldsymbol{u}}_{w2}) = \breve{\boldsymbol{L}}_{w2}^T\breve{\boldsymbol{x}}_w \tag{4.56c}$$

5. The final sparse reduced macromodel is obtained by incorporating the parametric elements into the sparse model in (4.56) and in form of

$$\left[\bar{\boldsymbol{G}} + s\bar{\boldsymbol{C}} + \sum_{i=1}^{r} f_{i(\lambda_i)}\bar{\boldsymbol{D}}_i + s\sum_{j=1}^{c} g_{j(\phi_j)}\bar{\boldsymbol{E}}_j\right]\bar{\boldsymbol{x}} = \bar{\boldsymbol{R}}\boldsymbol{u} \tag{4.57a}$$

$$\boldsymbol{i} = \bar{\boldsymbol{L}}^T\boldsymbol{x} \tag{4.57b}$$

where $\bar{\boldsymbol{G}}, \bar{\boldsymbol{C}}, \bar{\boldsymbol{D}}_i$, and $\bar{\boldsymbol{E}}_j$ are small yet sparse matrices.

### 4.5.2 Numerical results

**Example 1**

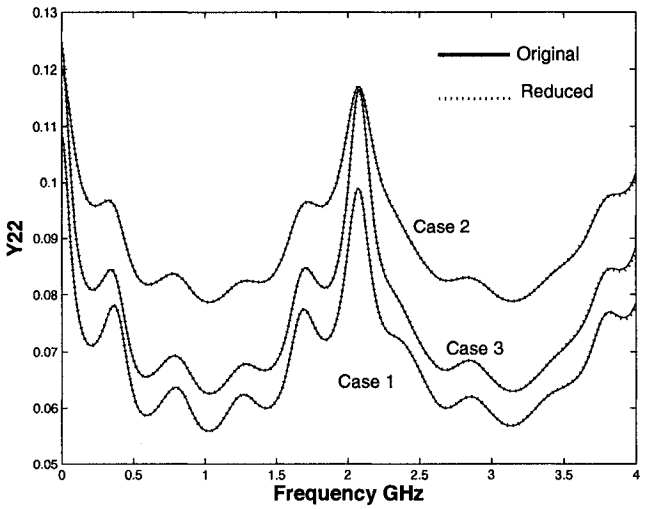


**Fig. 4.21** Frequency responses for example 1 (Y22) (comparison between the original system and the proposed macromodel)

The first example is an interconnect network with 9 non-coupled transmission lines, 1 nine-coupled transmission line system (see [88] for a description of the 9 coupled line system), and 31 resistors. This system is a parametric 2-port network with 4 resistor values, $R_1, R_2, R_3$, and $R_4$, as parameters. The resulting size of the MNA matrices is 3522.

Using the proposed sparse multi-level reduction method, the size of the reduced system after doing first level reduction is 366, while the macromodel is further reduced to 159 after doing the second level reduction. The size comparison is shown in Table 4.9. The values of the four parametric resistors are varied from $1 \sim 200\Omega$ as shown in Table 4.10. The frequency responses for three sampling parameter cases are shown in Fig. 4.21 and Fig. 4.22. As can be seen, the results from the reduced macromodel match those from the original system up to 4GHz. The average relative error of Y parameters for all testing cases is 0.07%. Fig. 4.23 shows the transient responses for three sampling test cases of parameters due to a $1V$ pulse with a rise time of $0.1ns$ and a pulse width of $2ns$. As expected, it is also accurate. A summary of CPU cost

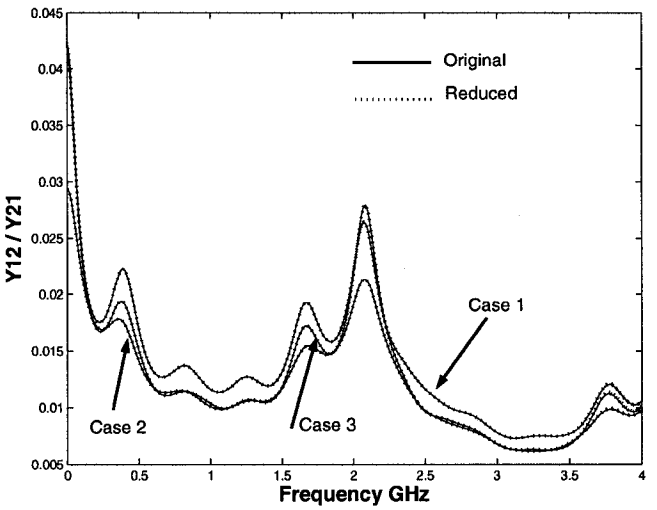


**Fig. 4.22** Frequency responses for example 1 (Y12/Y21) (comparison between the original system and the proposed macromodel)

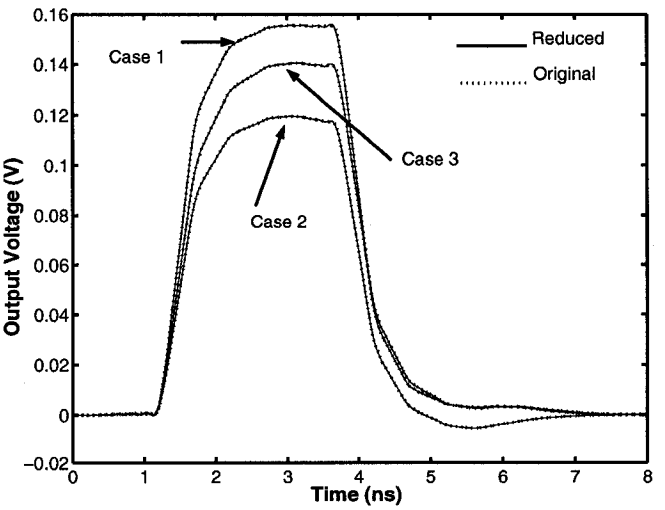


**Fig. 4.23** Transient responses for example 1 (comparison between the original system and the proposed macromodel)

comparison between the original network, traditional parametric model order reduction (PMR), and the proposed method to obtain transient responses is shown in Table 4.11. It is not surprising that the proposed approach can achieve 350.8 average CPU speed-up since the proposed model is not only very small but also sparse.

**Table 4.9** Size Comparison for Examples

| | Example1 | Example2 |
|---|---|---|
| Original System | 3522 | 6223 |
| Traditional PMR | 372 | 608 |
| Sparse Multi-level PMR | 159 | 326 |

**Table 4.10** Test Cases for Example 1

| | $R_1$ | $R_2$ | $R_3$ | $R_4$ |
|---|---|---|---|---|
| Case 1 | 96Ω | 44Ω | 93Ω | 122Ω |
| Case 2 | 31Ω | 99Ω | 119Ω | 99Ω |
| Case 3 | 56Ω | 42Ω | 143Ω | 51Ω |

**Table 4.11** CPU Cost Comparison for Example 1

| | Case 1 | Case 2 | Case 3 | Ave. speed-up |
|---|---|---|---|---|
| Original System | 2906.4 | 2765.1 | 2712.6 | - |
| Traditional PMR | 383.4 | 336.8 | 381.6 | 7.6 |
| Sparse Multi-level PMR | 7.9 | 7.7 | 8.3 | 350.8 |

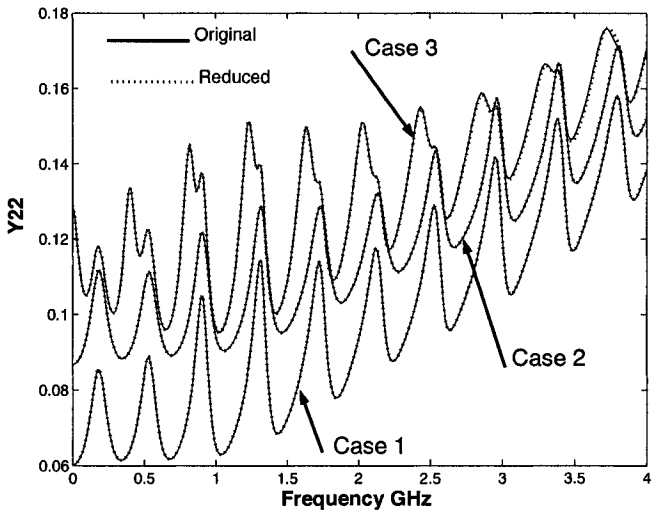


**Fig. 4.24** Frequency responses for example 2 (comparison between the original system and the proposed macromodel)

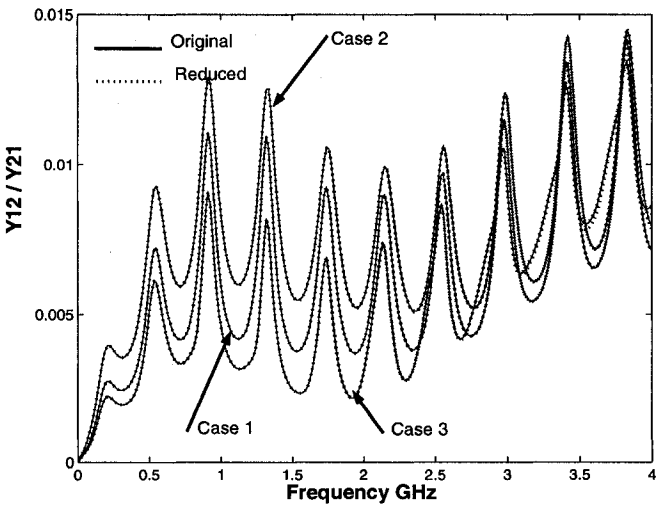


**Fig. 4.25** Frequency responses for example 2 (comparison between the original system and the proposed macromodel)

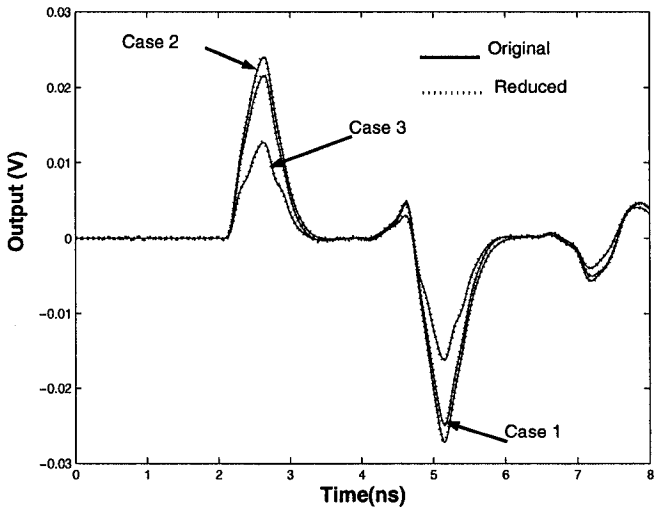


**Fig. 4.26** Time responses for example 2 (comparison between the original system and the proposed macromodel)

### Example 2

The second example is another interconnect network, this time consisting of 9 non-coupled transmission lines, 2 nine-coupled transmission line systems, 3 capacitors, and 29 resistors. The MNA size of this 2-port system is 6223. This system is parametric with respect to 2 resistor values and 2 capacitor values, $R_1, R_2, C_1$, and $C_2$.

**Table 4.12** Test Cases for Example 2

| | $R_1$ | $R_2$ | $C_1$ | $C_2$ |
|---|---|---|---|---|
| Case 1 | 211Ω | 168Ω | 1.70pF | 1.25pF |
| Case 2 | 131Ω | 209Ω | 0.94pF | 1.77pF |
| Case 3 | 50Ω | 4Ω | 2.07pF | 1.07pF |

Using the proposed sparse multi-level reduction method, the size of the reduced system after doing first level reduction is 628 while the macromodel is further reduced to 326 after doing the second level reduction. The size comparison is shown in Table 4.9. The values of the two

**Table 4.13** CPU Cost Comparison for Example 2

| | Case 1 | Case 2 | Case 3 | Ave. speed-up |
|---|---|---|---|---|
| Original System | 3272.1 | 3184.5 | 3187.5 | - |
| Traditional PMR | 1848.8 | 1817.8 | 1865.1 | 1.7 |
| Sparse Multi-level PMR | 50.3 | 50.8 | 47.7 | 64.8 |

parametric resistors are varied from $1 \sim 200\Omega$. The values of the two parametric capacitors are varied from $0.5pF \sim 5pF$ as shown in Table 4.12. The frequency responses for three sampling parameter cases are shown in Fig. 4.24 and Fig. 4.25. As can be seen, the results from the reduced macromodel match those from the original system up to 4GHz. The average relative error of Y parameters for all testing cases is $0.1\%$. Fig. 4.26 shows the transient responses for three sampling cases of parameters due to a $1V$ pulse with a rise time of $0.5ns$ and a pulse width of $2ns$. As expected, it is also accurate. A summary of CPU cost comparison between the original network, traditional parametric model order reduction (PMR), and the proposed method to obtain transient responses is shown in Table 4.13. In this example, the average CPU speed-up of 64.8 has been achieved.

In summary, a new parametric model order reduction method was presented that produced macromodels which are small yet also sparse. As shown through numerical examples, the sparse and small characteristics enable faster simulation, which is essential since parametric macromodels are often used repeatedly.

# Chapter 5

# Nonlinear Model Order Reduction

## 5.1 Introduction

Model order reduction methods based on congruence transformation were developed for obtaining passive [19] efficient macromodels for linear interconnect networks [5], [19]. Such congruence transformation based techniques have become the methods of choice for model order reduction of interconnect networks due to their accuracy, numerical stability and passivity [5]. These methods are, however, frequency domain methods and are thus inherently limited to linear subcircuits. In [64], the concept of congruence transformation based reduction was extended to nonlinear equations in the time domain and was shown to be stable and passive. However, while the approach in [64] leads to significant CPU cost savings, it is fundamentally a simulation method based circuit reduction and it therefore does not produce a nonlinear macromodel which can be reused under different input waveforms and load conditions. In terms of the nonlinear macromodeling techniques, nonlinear reduction methods based on Taylor series were proposed in [35]–[37], where a set of linearizations is obtained by dropping the higher order terms of the Taylor series. Each linearization is then reduced using Krylov projection methods. The main drawback of these methods is that the reduction can only be performed to weakly nonlinear systems. Furthermore, the exponentially increasing cost with the number of expansion terms in-

cluded makes these methods only limited in practice to cubic expansion. A trajectory piecewise linear approach [38], [80], [81] was proposed to handle strong nonlinearity. The key observation in trajectory model order reduction is that the nonlinearity is characterized by collections of linearized systems at different expansion points around state trajectory. Each individual linearized system is reduced using Krylov projection methods and the final macromodel is then obtained by taking a weighted combination of the resulting reduced order linearized models. However, finding a good weighting function is not an easy task for a broad class of nonlinear systems.

Given the problems associated with the existing nonlinear macromodeling techniques, a parametric sparse macromodeling technique for nonlinear networks is presented in this chapter. This method results in a sparse reduced order macromodel which is also valid over a range of parameter values. This nonlinear parametric macromodel improves the simulation time for parametric nonlinear networks since the macromodel only needs to be created once and can be used many times with different internal circuit parameters. In order to achieve this goal, a new formulation is proposed. The proposed formulation allows for the decoupling of both nonlinear and parametric equations by introducing constrained ports. The reduction and sparsification are performed on the linear portion of the macromodel in the space of the new formulation scheme, and the reduced macromodel is then brought back to the traditional representation by reincorporating the parametric and nonlinear equations without losing sparsity. It is to be noted that the newly introduced constrained ports have a negligible impact on the size of the reduced macromodel.

This chapter is organized into 6 sections. Following this introduction, Section 5.2 describes the system formulation for nonlinear networks. Nonlinear model order reduction based on congruence transformation is presented in Section 5.3. In Section 5.4, the proposed method for obtaining sparse nonlinear macromodels is presented. The parametric sparse nonlinear macromodel is proposed in Section 5.5. Finally, numerical examples are shown in Section 5.6.

## 5.2 Network Formulation

Consider a multi-port interconnect network consisting of many linear and nonlinear components. The nonlinear modified nodal analysis [63] formulation of such a $p$-port network as defined in equation (2.9), can be written as

$$\boldsymbol{G}\boldsymbol{x}(t) + \boldsymbol{C}\dot{\boldsymbol{x}}(t) + \boldsymbol{f}(\boldsymbol{x}(t)) = \boldsymbol{R}\boldsymbol{u}(t) + \boldsymbol{b}(t) \quad (5.1a)$$

$$\boldsymbol{i}(t) = \boldsymbol{L}^T\boldsymbol{x}(t) \quad (5.1b)$$

where $\boldsymbol{x}(t) \in \mathbb{R}^n$ is a vector of node voltages appended by independent voltage source currents, linear inductor currents, nonlinear capacitor charges and nonlinear inductor fluxes; $\boldsymbol{G} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{C} \in \mathbb{R}^{n\times n}$ contain the contributions of the memoryless and memory elements respectively as was outlined in Section 2.1.3.

The formulation in (5.1) can be generalized to take into account certain design parameters (interconnect geometries, resistors and capacitors etc). In such a case, the parametrized modified nodal analysis formulation [31], [63] including the nonlinear and parametric components can be expressed as:

$$\left(\boldsymbol{G} + \sum_{i=1}^{r} h_i(\lambda_i)\boldsymbol{D}_i\right)\boldsymbol{x}(t) + \boldsymbol{f}(\boldsymbol{x}(t)) + \left(\boldsymbol{C} + \sum_{j=1}^{c} g_j(\phi_j)\boldsymbol{E}_j\right)\dot{\boldsymbol{x}}(t) = \boldsymbol{R}\boldsymbol{u}(t) + \boldsymbol{b}(t) \quad (5.2a)$$

$$\boldsymbol{i} = \boldsymbol{L}^T\boldsymbol{x}(t) \quad (5.2b)$$

where $\boldsymbol{D}_1, ..., \boldsymbol{D}_r$ are matrices each containing the modified nodal analysis formulation stamp of a particular memoryless parameter, $\boldsymbol{E}_1, ..., \boldsymbol{E}_c$ are matrices each containing the modified nodal analysis formulation stamp of a particular memory parameter, $\lambda_1, ..., \lambda_r$ are input scalars corresponding to the variable parameters represented by $\boldsymbol{D}_1, ..., \boldsymbol{D}_r$ respectively, $\phi_1, ..., \phi_c$ are input scalars corresponding to the variable parameters represented by $\boldsymbol{E}_1, ..., \boldsymbol{E}_c$ respectively. $h_i$ is an algebraic function describing the contributions of the parameter $\lambda_i$ to the memoryless elements. $g_j$ is an algebraic function describing the contributions of the parameter $\phi_j$ to the memory

elements.

## 5.3 Nonlinear macromodeling based on congruence transformation

### 5.3.1 Reduced macromodel

In this section, model order reduction is applied in the time domain on the nonlinear MNA multi-port formulation in (5.1). The resulting reduced system will therefore be in macromodel form and can thus be connected to other circuits [50]. Furthermore, the congruence transformation matrix $\boldsymbol{Q}$ used for the reduction is chosen such that it is valid over the desired user defined range of load conditions and input waveforms.

By applying the change of variables $\boldsymbol{x}(t) = \boldsymbol{Q}\hat{\boldsymbol{x}}(t)$ to equation (5.1), and pre-multiplying by $\boldsymbol{Q}^T$, we obtain the reduced order macromodel as

$$\hat{\boldsymbol{G}}\hat{\boldsymbol{x}}(t) + \hat{\boldsymbol{C}}\dot{\hat{\boldsymbol{x}}}(t) + \hat{\boldsymbol{f}}(\hat{\boldsymbol{x}}(t)) = \hat{\boldsymbol{R}}\boldsymbol{u}(t) + \hat{\boldsymbol{b}}(t); \quad \boldsymbol{i}(t) = \hat{\boldsymbol{R}}^T\hat{\boldsymbol{x}}(t) \tag{5.3}$$

where

$$\hat{\boldsymbol{G}} = \boldsymbol{Q}^T\boldsymbol{G}\boldsymbol{Q}; \quad \hat{\boldsymbol{C}} = \boldsymbol{Q}^T\boldsymbol{C}\boldsymbol{Q}$$

$$\hat{\boldsymbol{R}} = \boldsymbol{Q}^T\boldsymbol{R}; \quad \hat{\boldsymbol{b}}(t) = \boldsymbol{Q}^T\boldsymbol{b}(t); \quad \hat{\boldsymbol{f}}(\hat{\boldsymbol{x}}(t)) = \boldsymbol{Q}^T\boldsymbol{f}(\boldsymbol{Q}\hat{\boldsymbol{x}}(t)) \tag{5.4}$$

and $\boldsymbol{Q} \in \mathbb{R}^{n\times q}$ is a real orthonormal matrix, $\hat{\boldsymbol{x}}(t) \in \mathbb{R}^q$ is the solution to the reduced system, and $q \ll n$. The process of obtaining $\boldsymbol{Q}$ will be discussed in the next section. The macromodel in (5.3) can be stamped in the MNA equations of a large network.

### 5.3.2 Reduction subspace

In order for the change of the variables $\boldsymbol{x}(t) = \boldsymbol{Q}\hat{\boldsymbol{x}}(t)$ to be valid, the congruence transformation matrix $\boldsymbol{Q}$ must span the subspace containing $\boldsymbol{x}(t)$ over the range of loading conditions and input waveforms of interest. The subspace is generated by performing transient analyses and sampling

$\boldsymbol{x}(t)$ from the initial time point ($t_0$) up to the terminal time point ($t_h$). This is done several times on different load conditions and using a range of input waveforms. The subspace containing $\boldsymbol{x}(t)$ is then defined as

$$\boldsymbol{K} = [\boldsymbol{x}^{(1)}(t_0), \ldots, \boldsymbol{x}^{(1)}(t_i), \ldots, \boldsymbol{x}^{(1)}(t_h), \ldots, \boldsymbol{x}^{(m)}(t_0), \ldots, \boldsymbol{x}^{(m)}(t_i), \ldots, \boldsymbol{x}^{(m)}(t_h)] \tag{5.5}$$

where $\boldsymbol{x}^{(m)}(t_i)$ is the $i^{th}$ time point for the $m^{th}$ transient response. Given the fact that there is typically a lot of similarities between various transient responses, and the fact that, for a large system, the elements of $\boldsymbol{x}(t)$ are not all linearly independent (which is the reason why linear model order reduction is possible), the subspace $\boldsymbol{K}$ typically contains a lot of redundant dimensions. In order to obtain an optimal orthonormal basis $\boldsymbol{Q}$ of this subspace, the singular value decomposition(SVD) [85] is used to identify the dominant directions. The subspace $\boldsymbol{K}$ is therefore decomposed using singular value decomposition process such that

$$\boldsymbol{K} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\boldsymbol{T}} \tag{5.6}$$

where $\boldsymbol{S}$ is a diagonal matrix containing the singular values in decreasing order and $\boldsymbol{U}, \boldsymbol{V}$ are orthogonal matrices. The matrix $\boldsymbol{Q}$ is constructed by taking only the leading $k$ columns of $\boldsymbol{U}$ which correspond to large values in $\boldsymbol{S}$, and thus contain the dominant dimensions of the subspace. The order $k$ is chosen using the following criteria:

$$\frac{s(k,k)}{s(1,1)} \leq err \tag{5.7}$$

$s(1,1)$ and $s(k,k)$ is the first and the $k^{th}$ elements of the diagonal elements of $\boldsymbol{S}$. $err$ is the error bound to choose the dominant subspace. Thus, the dominant subspace $\boldsymbol{Q}$ is given by:

$$\boldsymbol{Q} = [\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}] \tag{5.8}$$

where $u_1$, $u_2$, ..., $u_k$ are the first $k$ columns of $U$. It is to be noted that this subspace is valid over the range of inputs and loads used to construct it.

Obtaining the subspace defined in (5.5) requires samples of the transient responses under different input waveforms and loading conditions. However, it is not possible to perform a transient analysis directly on the multi-port formulation in (5.1). For that purpose, the load information is integrated into the original differential equations in (5.1) to obtain a new formulation called the nonlinear parametric port formulation. It is to be noted that the unknown variables in the new formulation are the same as the unknown variables $x(t)$ in the original differential equations. The detailed process of obtaining such a nonlinear parametric port formulation can be found in Section 5.4.4.

## 5.4 Sparse Nonlinear Macromodel

If nonlinear time domain model order reduction was applied directly to the original system in (5.1) as was outlined in the previous section, the resulting nonlinear macromodel in (5.3) would be dense. This significantly reduces the efficiency of the simulation. In this section, we propose a new approach which allows for the application of time domain nonlinear macromodeling as well as for the sparsification of the macromodel without any significant impact on the size of the reduced order macromodel [52].

### 5.4.1 Macromodel formulation suitable for sparsification

In the first step of the proposed approach, the system in (5.1) is reformulated such that the nonlinear equations in $f(x(t))$ are decoupled from the nonlinear equations through the introduction

of constrained ports. The resulting formulation is in the form of:

$$\boldsymbol{G}_p\boldsymbol{x}_p(t) + \boldsymbol{C}_p\dot{\boldsymbol{x}}_p(t) = \boldsymbol{R}_{p_1}\boldsymbol{u}(t) + \boldsymbol{b}_p(t) + \boldsymbol{R}_{p_2}\boldsymbol{u}_2 \tag{5.9a}$$

$$\boldsymbol{i}(t) = \boldsymbol{L}_{p_1}^T\boldsymbol{x}_p(t) \tag{5.9b}$$

$$\boldsymbol{f}_p(\boldsymbol{u}_2) = \boldsymbol{L}_{p_2}^T\boldsymbol{x}_p(t) \tag{5.9c}$$

where $\boldsymbol{G}_p$, $\boldsymbol{C}_p$ are obtained by adding the constrained ports to the original system in (5.1). The variable $\boldsymbol{x}_p(t)$ is obtained by appending the new constrained port currents to $\boldsymbol{x}(t)$. $\boldsymbol{R}_{p_1}$ is essentially the same as $\boldsymbol{R}$ in (5.1a) except for adding new rows with all zeros corresponding to the new constrained ports. $\boldsymbol{R}_{p_2}$ is a selector matrix mapping the new constrained port voltages into the node space. $\boldsymbol{L}_{p_1}$ is essentially the same as $\boldsymbol{L}$ in (5.1b) except for adding new rows with all zeros corresponding to the new constrained ports. Equation (5.9c) represents the nonlinear equations of the circuit which are expressed as nonlinear constraints on the newly introduced ports. It is to be noted that these constraints are utilized in the computation of the reduction matrix in order to ensure that the size of the reduced macromodel is not affected by the addition of the constrained ports.

Having the modified formulation in (5.9), the reduced system is then obtained by using congruence transformation, resulting in

$$\hat{\boldsymbol{G}}_p\hat{\boldsymbol{x}}_p(t) + \hat{\boldsymbol{C}}_p\dot{\hat{\boldsymbol{x}}}_p(t) = \hat{\boldsymbol{R}}_{p_1}\boldsymbol{u}(t) + \hat{\boldsymbol{b}}_p(t) + \hat{\boldsymbol{R}}_{p_2}\hat{\boldsymbol{u}}_2 \tag{5.10a}$$

$$\boldsymbol{i}(t) = \hat{\boldsymbol{L}}_{p_1}^T\hat{\boldsymbol{x}}_p(t) \tag{5.10b}$$

$$\boldsymbol{f}_p(\hat{\boldsymbol{u}}_2) = \hat{\boldsymbol{L}}_{p_2}^T\hat{\boldsymbol{x}}_p(t) \tag{5.10c}$$

where:

$$\hat{G}_p = Q_p^T G_p Q_p; \quad \hat{C}_p = Q_p^T C_p Q_p \tag{5.11a}$$

$$\hat{R}_{p_1} = Q_p^T R_{p_1}; \quad \hat{R}_{p_2} = Q_p^T R_{p_2} \tag{5.11b}$$

$$\hat{L}_{p_1} = Q_p^T L_{p_1}; \quad \hat{L}_{p_2} = Q_p^T L_{p_2}; \quad \hat{b}_p(t) = Q_p^T b_p(t) \tag{5.11c}$$

The computation of the congruence transformation $Q_p$ in the time domain is described in Section 5.4.4. A simple example of this representation is illustrated in Section 5.4.5.

### 5.4.2 Sparsification of the reduced macromodel

The reduced macromodel in (5.10) is generally a dense nonlinear macromodel. However, we note that the formulation in (5.10a) is in the form of general linear multi-port network. It is therefore possible to apply diagonalization techniques such as the one described in [75] to equation (5.10a). Equation (5.10a) can be reformulated by premultiplying $\hat{G}_p^{-1}$

$$\hat{x}_p(t) + \hat{G}_p^{-1}\hat{C}_p\dot{\hat{x}}_p(t) = \hat{G}_p^{-1}\hat{R}_{p_1}u(t) + \hat{G}_p^{-1}\hat{b}_p(t) + \hat{G}_p^{-1}\hat{R}_{p_2}\hat{u}_2 \tag{5.12}$$

By applying eigen-decomposition to $\hat{G}_p^{-1}\hat{C}_p$, we obtain

$$\hat{G}_p^{-1}\hat{C}_p = VDV^{-1} \tag{5.13}$$

$V$ and $D$ are in general complex; however a real diagonalization can be derived from (5.13) in the form

$$\hat{G}_p^{-1}\hat{C}_p = S_r D_r {S_r}^{-1} \tag{5.14}$$

where $\boldsymbol{D}_r$ is a real block diagonal matrix and

$$\boldsymbol{S}_r = \boldsymbol{V}\boldsymbol{P}^{-1} \tag{5.15}$$

$$\boldsymbol{D}_r = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1} \tag{5.16}$$

$\boldsymbol{P}$ is defined as:

1. when the corresponding element of the diagonal matrix $\boldsymbol{D}$, $d_{i,i}$, is real, the element of $\boldsymbol{P}$ matrix is $p_{i,i} = 1$,
2. when the corresponding elements of the diagonal matrix $\boldsymbol{D}$, $d_{i,i}$ and $d_{i+1,i+1}$ are complex conjugate, the elements of $\boldsymbol{P}$ matrix are: $p_{i:i+1,i:i+1} = \begin{bmatrix} 1 & 1 \\ j & -j \end{bmatrix}$

By substituting (5.14) into (5.12), the sparse reduced system becomes

$$\breve{\boldsymbol{G}}_p\breve{\boldsymbol{x}}_p(t) + \breve{\boldsymbol{C}}_p\dot{\breve{\boldsymbol{x}}}_p(t) = \breve{\boldsymbol{R}}_{p_1}\boldsymbol{u}(t) + \breve{\boldsymbol{b}}_p(t) + \breve{\boldsymbol{R}}_{p_2}\breve{\boldsymbol{u}}_2 \tag{5.17a}$$

$$\boldsymbol{i}(t) = \breve{\boldsymbol{L}}_{p_1}^T\breve{\boldsymbol{x}}_p(t) \tag{5.17b}$$

$$\boldsymbol{f}_p(\breve{\boldsymbol{u}}_2) = \breve{\boldsymbol{L}}_{p_2}^T\breve{\boldsymbol{x}}_p(t) \tag{5.17c}$$

where $\breve{\boldsymbol{G}}_p$ is an identity matrix,

$$\breve{\boldsymbol{C}}_p = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1} \tag{5.18a}$$

$$\breve{\boldsymbol{R}}_{p_1} = (\boldsymbol{V}\boldsymbol{P}^{-1})^{-1}\hat{\boldsymbol{G}}_p^{-1}\hat{\boldsymbol{R}}_{p_1} \tag{5.18b}$$

$$\breve{\boldsymbol{b}}_p = (\boldsymbol{V}\boldsymbol{P}^{-1})^{-1}\hat{\boldsymbol{G}}_p^{-1}\hat{\boldsymbol{b}}_p \tag{5.18c}$$

$$\breve{\boldsymbol{R}}_{p_2} = (\boldsymbol{V}\boldsymbol{P}^{-1})^{-1}\hat{\boldsymbol{G}}_p^{-1}\hat{\boldsymbol{R}}_{p_2} \tag{5.18d}$$

$$\breve{\boldsymbol{L}}_{p_1} = (\boldsymbol{V}\boldsymbol{P}^{-1})^T\hat{\boldsymbol{L}}_{p_1} \tag{5.18e}$$

$$\breve{\boldsymbol{L}}_{p_2} = (\boldsymbol{V}\boldsymbol{P}^{-1})^T\hat{\boldsymbol{L}}_{p_2} \tag{5.18f}$$

$$\breve{\boldsymbol{x}}_p(t) = (\boldsymbol{V}\boldsymbol{P}^{-1})^{-1}\hat{\boldsymbol{x}}_p(t) \tag{5.18g}$$

Note that $\breve{G}_p$ and $\breve{C}_p$ are sparse real reduced matrices.

### 5.4.3 Sparse reduced order nonlinear macromodel

The final step is to reincorporate the nonlinear constraints in (5.17c) into the overall sparse macromodel equations in (5.17a). In order to achieve this, the reduced macromodel is treated as a subsection, which is represented by (5.17a) and (5.17b). Then the nonlinear elements, which were removed before are connected back to the constrained ports of the subsection to form the sparse reduced order nonlinear macromodel.

$$\begin{bmatrix} \boldsymbol{G}_\phi & \boldsymbol{D}\breve{\boldsymbol{L}}_p^T \\ -\breve{\boldsymbol{R}}_p\boldsymbol{D}^T & \breve{\boldsymbol{G}}_p \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_\phi(t) \\ \breve{\boldsymbol{x}}_p(t) \end{bmatrix} + \begin{bmatrix} \boldsymbol{C}_\phi & 0 \\ 0 & \breve{\boldsymbol{C}}_p \end{bmatrix} \begin{bmatrix} \dot{\boldsymbol{x}}_\phi(t) \\ \dot{\breve{\boldsymbol{x}}}_p(t) \end{bmatrix}$$

$$+ \begin{bmatrix} \boldsymbol{f}_\phi(\boldsymbol{x}_\phi(t)) \\ 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{R}_\phi \\ 0 \end{bmatrix} \boldsymbol{u}(t) + \begin{bmatrix} 0 \\ \breve{\boldsymbol{b}}_p(t) \end{bmatrix} \tag{5.19}$$

where

$$\breve{\boldsymbol{L}}_p^T = \begin{bmatrix} \breve{\boldsymbol{L}}_{p_1}^T \\ \breve{\boldsymbol{L}}_{p_2}^T \end{bmatrix} \tag{5.20}$$

$$\breve{\boldsymbol{R}}_p = \begin{bmatrix} \breve{\boldsymbol{R}}_{p_1} & \breve{\boldsymbol{R}}_{p_2} \end{bmatrix} \tag{5.21}$$

$\boldsymbol{G}_\phi$ and $\boldsymbol{C}_\phi$ represent the general port formulation of input and output ports. $\boldsymbol{R}_\phi$ is a selector matrix. $\boldsymbol{f}_\phi(\boldsymbol{x}_\phi(t))$ represents the nonlinear elements for constrained ports. $\boldsymbol{D}$ is a selector matrix with elements $d_{i,j} \in \{0, 1\}$ and a maximum of one non-zero in each row or column that maps the vectors of port voltages and port currents entering subsection into the node space of the network. $\breve{\boldsymbol{x}}_p(t)$ is the unknowns of the reduced order macromodel in (5.17a). The above formulation is illustrated by a simple example in Section 5.4.5. In summary, the final reduced order macromodel

based on (5.19) is in the form

$$\tilde{G}\tilde{x}(t) + \tilde{C}\dot{\tilde{x}}(t) + \tilde{f}(\tilde{x}(t)) = \tilde{R}u(t) + \tilde{b}(t) \quad (5.22a)$$

$$i(t) = \tilde{L}^T\tilde{x}(t) \quad (5.22b)$$

where $\tilde{G}$ and $\tilde{C}$ are small and sparse matrices.

### 5.4.4 Reduction subspace

#### Nonlinear parametric port formulation

The congruence transformation matrix $Q_p$ used in the reduced order macromodel must span the subspace containing $x_p(t)$ over the desired range of loading conditions and input waveforms. However it is not possible to perform a transient analysis directly on the multi-port formulation in (5.9a). It is in fact necessary to take into account the conditions imposed by equations (5.9b) and (5.9c) and obtain nonlinear parametric port formulation. For clarify of presentation, we will first consider a simple two-port network with one nonlinear element, shown in Fig 5.1, then the results will be extended into general multi-port nonlinear networks. Since there is one nonlinear

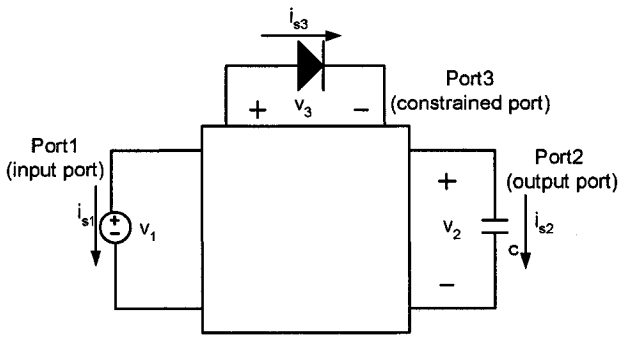


**Fig. 5.1** An example two-port network with one nonlinear element

element, a constrained port (port3) is added to the network. Here, the two ports are divided into input port (port1) and output port (port2). The voltages across the input port, output port and the constrained port are $v_1$, $v_2$ and $v_3$ respectively and the currents are $i_{s1}$, $i_{s2}$ and $i_{s3}$ respectively. If

we examine the modified nodal analysis formulation equation of such a system defined in (5.9), we note that the last three rows of the $G_p$ matrix, corresponding to the port1, port2 and port3 equations, in (5.9a) are:

$$\boldsymbol{G}_p = \begin{bmatrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 \end{bmatrix} \tag{5.23}$$

and

$$\boldsymbol{x}_p = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \cdots \\ i_{s1} \\ i_{s2} \\ i_{s3} \end{bmatrix} \boldsymbol{R}_{p1} = \begin{bmatrix} 0 & 0 \\ \cdots & \cdots \\ -1 & 0 \\ 0 & -1 \\ 0 & 0 \end{bmatrix} \boldsymbol{u}(t) = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \tag{5.24}$$

$$\boldsymbol{R}_{p2} = \begin{bmatrix} 0 \\ \cdots \\ 0 \\ -1 \end{bmatrix} \boldsymbol{u}_2 = \begin{bmatrix} v_3 \end{bmatrix} \tag{5.25}$$

Using this general port representation, the port equations in the last three rows are simply $v_1 = v_1$, $v_2 = v_2$, and $v_3 = v_3$. In other words, the port voltage can be arbitrarily set by the boundary conditions. Note that port1 is the input port and port2 is designated as output port with capacitive loads and port3 is constrained by the nonlinear element. In this case, the equation at port1 remains

unchanged but the equation at port2 becomes:

$$i_{s2} - c\dot{v}_2 = 0 \tag{5.26}$$

and the equation at port3 becomes:

$$i_{s3} - I_s\left(e^{\frac{v_3}{v_T}} - 1\right) = 0 \tag{5.27}$$

Equation (5.27) represents a nonlinear constraint. In this case, a simple diode model was used, where $I_s$ is the reverse bias saturation current of the diode and $v_T$ is thermal voltage. Incorporating (5.26) and (5.27) into modified nodal analysis formulation stamp has the effect of converting the multi-port network in (5.9a) into parametric "single-port" network while keeping the vector of unknowns $\boldsymbol{x}_p$ unchanged. This is done by modifying the rows corresponding to port2, port3 equations in the modified nodal analysis formulation in (5.9a) as follows:

$$\bar{\boldsymbol{G}}_p\boldsymbol{x}_p(t) + (\boldsymbol{C}_p + \boldsymbol{C}_1)\dot{\boldsymbol{x}}_p(t) + \bar{\boldsymbol{f}}(\boldsymbol{x}_p(t)) = \boldsymbol{R}_1\bar{\boldsymbol{u}}_1(t) + \boldsymbol{b}_p(t), \tag{5.28}$$

where

$$\bar{\boldsymbol{G}}_p = \begin{bmatrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -1 \end{bmatrix} \tag{5.29}$$

$$\boldsymbol{C}_1 = \begin{bmatrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & c & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \tag{5.30}$$

$$\boldsymbol{R}_1 = \begin{bmatrix} 0 \\ \cdots \\ -1 \\ 0 \\ 0 \end{bmatrix} \quad \bar{\boldsymbol{u}}_1(t) = \begin{bmatrix} v_1 \end{bmatrix} \tag{5.31}$$

$$\bar{\boldsymbol{f}}(\boldsymbol{x}_p(t)) = \begin{bmatrix} 0 \\ \cdots \\ 0 \\ 0 \\ I_s\left(e^{\frac{v_3}{v_T}} - 1\right) \end{bmatrix} \tag{5.32}$$

It is to be noted that there is one parameter in (5.28), $c$, corresponding to the capacitive load. To extend the above method into $M$ input ports, $N$ output ports network ($M + N = p$), and $E$ constrained ports, the nonlinear parametric port formulation with capacitive loads becomes

$$\bar{\boldsymbol{G}}_p \boldsymbol{x}_p(t) + (\boldsymbol{C}_p + \boldsymbol{C}_N)\dot{\boldsymbol{x}}_p(t) + \bar{\boldsymbol{f}}(\boldsymbol{x}_p(t)) = \boldsymbol{R}_M \bar{\boldsymbol{u}}_M(t) + \boldsymbol{b}_p(t), \tag{5.33}$$

where $\bar{\boldsymbol{G}}_p$ is obtained by modifying $(N + E)$ rows, corresponding to output port and constrained port equations, from port voltages into port currents, $\boldsymbol{C}_N$ is a matrix with elements containing capacitive load parameters, $c_1, c_2, \cdots, c_N$ and $\boldsymbol{R}_M$ is a selector matrix that maps input port voltages into the node space of the circuit.

If a resistor is connected to the port2 in Fig. 5.1, then the port2 equation becomes:

$$i_{s2} - g v_2 = 0 \tag{5.34}$$

Here we use conductance $g$ to represent the value of the resistive load.

Incorporating (5.34) and (5.27) into (5.9a) results in

$$(\bar{\boldsymbol{G}}_p + \boldsymbol{G}_1)\boldsymbol{x}_p(t) + \boldsymbol{C}_p\dot{\boldsymbol{x}}_p(t) + \bar{\boldsymbol{f}}(\boldsymbol{x}_p(t)) = \boldsymbol{R}_1\bar{\boldsymbol{u}}_1(t) + \boldsymbol{b}_p(t), \tag{5.35}$$

The above equation is similar to that in (5.28) except for the matrix $\boldsymbol{G}_1$, containing one nonzero elements, $g$. To extend the above method into $M$ input ports, $N$ output ports network ($M + N = p$), and $E$ constrained ports, the nonlinear parametric port formulation with resistive loads becomes:

$$(\bar{\boldsymbol{G}}_p + \boldsymbol{G}_N)\boldsymbol{x}_p(t) + \boldsymbol{C}_p\dot{\boldsymbol{x}}_p(t) + \bar{\boldsymbol{f}}(\boldsymbol{x}_p(t)) = \boldsymbol{R}_M\bar{\boldsymbol{u}}_M(t) + \boldsymbol{b}_p(t), \tag{5.36}$$

where $\boldsymbol{G}_N$ is a matrix with elements containing resistive load parameters, $g_1, g_2, \cdots, g_N$.

If the loads are the parallel combination of a resistor and a capacitor, then the nonlinear parametric port formulation with combination of capacitive and resistive loads becomes

$$(\bar{\boldsymbol{G}}_p + \boldsymbol{G}_N)\boldsymbol{x}_p(t) + (\boldsymbol{C}_p + \boldsymbol{C}_N)\dot{\boldsymbol{x}}_p(t) + \bar{\boldsymbol{f}}(\boldsymbol{x}_p(t)) = \boldsymbol{R}_M\bar{\boldsymbol{u}}_M(t) + \boldsymbol{b}_p(t) \tag{5.37}$$

where $\boldsymbol{G}_N$ and $\boldsymbol{C}_N$ are matrices with elements containing resistive load parameters, $g_1, g_2, \cdots, g_N$ and capacitive parameters,$c_1, c_2, \cdots, c_N$ respectively [40].

### Congruence Transformation Matrix

Using (5.37), the subspace is defined by performing transient responses sampling $\boldsymbol{x}_p(t)$ from the initial time point ($t_0$) up to the terminal time point ($t_h$). This is done several times on different load conditions (for capacitive loads, $c_1^j, \cdots, c_N^j$, ; for resistive loads, $g_1^j, \cdots, g_N^j$), and using a range of input waveforms $u_1^i(t), \cdots, u_M^i(t)$ as shown in Fig. 5.2. Four different input waveforms are used to generate the subspace data. They are 5V step input with fall time of $50ps$ and $500ps$ and 5V step input with rise time of $50ps$ and $500ps$ as shown in Fig. 5.3. The subspace containing

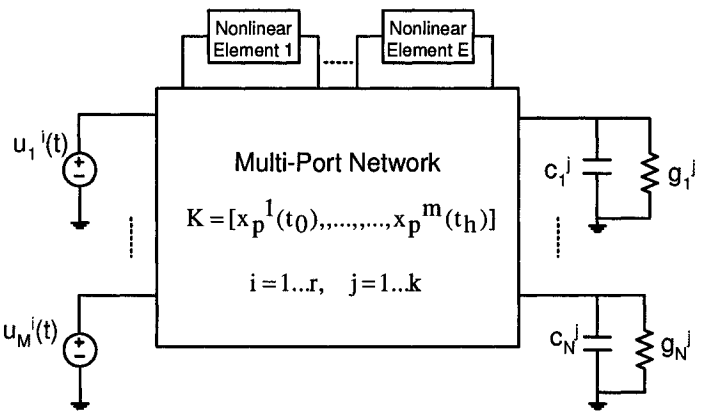


**Fig. 5.2** Network used for generation of subspace data

$\boldsymbol{x}_p(t)$ is then defined as

$$\boldsymbol{K} = [\boldsymbol{x}_p^1(t_0), \ldots, \boldsymbol{x}_p^1(t_h), \ldots, \boldsymbol{x}_p^m(t_0), \ldots, \boldsymbol{x}_p^m(t_h)] \tag{5.38}$$

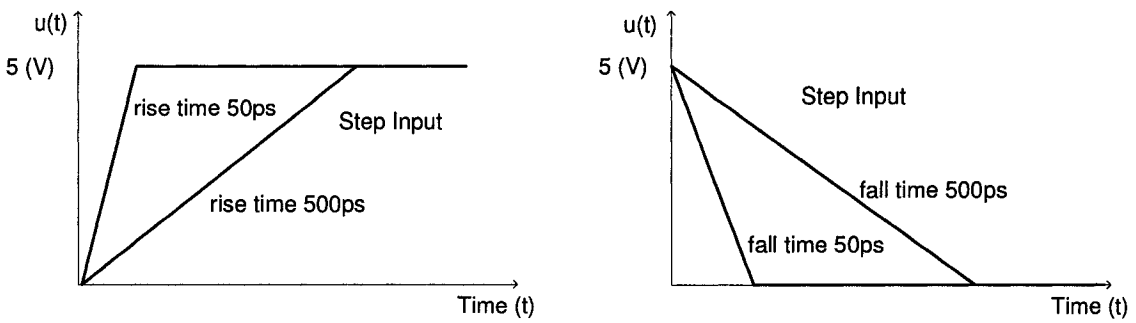


**Fig. 5.3** Input waveforms for calculation of congruence transformation

Given the fact that there is typically a lot of similarities between various transient responses, the subspace $\boldsymbol{K}$ typically contains a lot of redundant dimensions. In order to obtain an optimal orthonormal basis $\boldsymbol{Q}_p$ of this subspace, singular value decomposition [45], [85] is used to identify the dominant directions, which results in

$$\boldsymbol{K} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T \tag{5.39}$$

where $\boldsymbol{S}$ is a diagonal matrix with the singular values and in decreasing order. $\boldsymbol{U}$ and $\boldsymbol{V}$ are unitary matrices. The singular values along the diagonal of $\boldsymbol{S}$ measure the relative importance of the corresponding columns of $\boldsymbol{U}$. This provides a convenient way to filter out the redundant directions. Taking only the first $k$ columns of $\boldsymbol{U}$, those corresponding to the highest singular values, produces the congruence transformation matrix, $\boldsymbol{Q}_p$. Note that although the congruence matrix $\boldsymbol{Q}_p$ is obtained from (5.37) containing the information on the output ports and constrained ports, the order reduction procedure is applied to the general port formulation in (5.9) so that the reduced system is a macromodel expressed in (5.22) that can be connected to the predefined range of loads.

In summary, the proposed sparse nonlinear macromodeling algorithm consists of eight main steps as follows

1. The original system in (5.1) is reformulated such that the nonlinear elements are decoupled through the introduction of constrained ports. The new multi-port formulation is in (5.9).

2. In order to construct the subspace, a new nonlinear parametric port formulation in (5.37) is obtained by embedding the nonlinear constraints in (5.9c) and the load information into the multi-port formulation in (5.9a).

3. Generate the subspace $K$ by performing transient responses on (5.37) over the range of input waveforms and loads ($K = [\boldsymbol{x}^{(1)}(t_0), \cdots, \boldsymbol{x}^{(1)}(t_h), \cdots, \cdots, \boldsymbol{x}^{(m)}(t_h)]$).

4. Decompose the subspace $K$ into three matrices using SVD decomposition , $[U, S, V] \leftarrow$SVD($K$).

5. Extract the dominant subspace by taking the leading $k$ columns of $U$, $Q_p \leftarrow$first $k$ columns of $U$.

6. Performing the congruence transformation to (5.9) results in a dense reduced macromodel in (5.10).

7. The dense reduced macromodel in (5.10) is sparsified into a sparse reduced macromodel in (5.17).

8. The final reduced macromodel in (5.22) is obtained by reincorporating nonlinear constraints in (5.17c) into the sparse macromodel in (5.17a).

### 5.4.5 Illustration example

In this subsection, a simple example is given to illustrate the various mutation of the previous subsection. Consider a two-port network including an nonlinear component, a diode, shown in Fig. 5.4, the modified nodal analysis formulation of the original network is as in the form of (5.1):

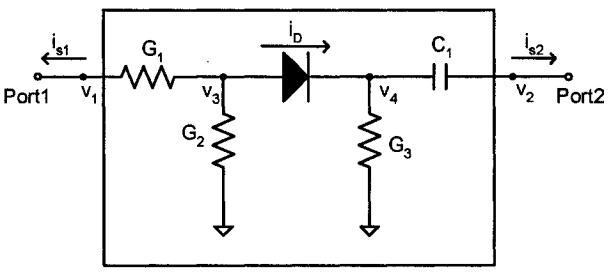


**Fig. 5.4** An example circuit for illustration

$$\begin{bmatrix} G_1 & 0 & -G_1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -G_1 & 0 & G_1+G_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & G_3 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ i_{s1} \\ i_{s2} \end{bmatrix}$$

$$+\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & C_1 & 0 & -C_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -C_1 & 0 & C_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{v}_1 \\ \dot{v}_2 \\ \dot{v}_3 \\ \dot{v}_4 \\ \dot{i}_{s1} \\ \dot{i}_{s2} \end{bmatrix}$$

$$+\begin{bmatrix} 0 \\ 0 \\ I_s\left(e^{\frac{v_3-v_4}{v_T}}-1\right) \\ -I_s\left(e^{\frac{v_3-v_4}{v_T}}-1\right) \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \tag{5.40}$$

where $I_s$ is the reverse bias saturation current of the diode, $v_T$ is thermal voltage. As the nonlinear element in the above equation hinder the sparsification of the reduced system, it is decoupled from the original modified nodal analysis equation by adding a nonlinear constrained port, port3, as shown in Fig. 5.5. Therefore the resulting formulation is obtained as shown in (5.9a):

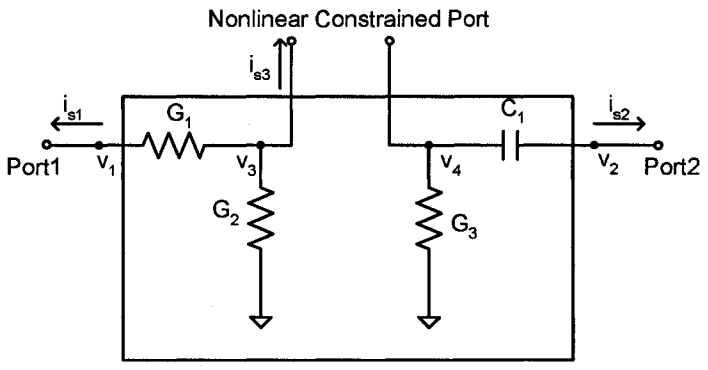


**Fig. 5.5** Nonlinear constrained port

$$
\begin{bmatrix}
G_1 & 0 & -G_1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
-G_1 & 0 & G_1+G_2 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & G_3 & 0 & 0 & -1 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 1 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ i_{s1} \\ i_{s2} \\ i_{s3} \end{bmatrix}
$$

$$
+
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & C_1 & 0 & -C_1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -C_1 & 0 & C_1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix} \dot{v}_1 \\ \dot{v}_2 \\ \dot{v}_3 \\ \dot{v}_4 \\ \dot{i}_{s1} \\ \dot{i}_{s2} \\ \dot{i}_{s3} \end{bmatrix}
$$

$$
=
\begin{bmatrix}
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
-1 & 0 \\
0 & -1 \\
0 & 0
\end{bmatrix}
\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}
+
\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ -1 \end{bmatrix}
\begin{bmatrix} v_{34} \end{bmatrix}
\tag{5.41}
$$

where

$$
v_{34} = v_3 - v_4 \tag{5.42}
$$

The variable vector in (5.41) differs from the vector in (5.40) by adding a new variable, the introduced nonlinear constrained port current, $i_{s3}$. Furthermore, the new current is bounded by

the following nonlinear equation

$$I_s\left(e^{\frac{v_{34}}{v_T}}-1\right)=\begin{bmatrix}0&0&0&0&1\end{bmatrix}\begin{bmatrix}v_1\\v_2\\v_3\\v_4\\i_{s1}\\i_{s2}\\i_{s3}\end{bmatrix} \tag{5.43}$$

The above equation is the nonlinear constraint on the newly introduced constrained port as in (5.9c). The port3 is different from the general port in the way that the load connected to the port3 is completely known. In the illustration circuit, it is a diode. This load information can be exploited in order to obtain a congruence transformation which is not affected by the new added ports. Note that the last two rows in (5.41), corresponding to the port2, port3 equations, are simply

$$-v_2 = -v_2 \tag{5.44a}$$

$$-v_3+v_4 = -v_{34} \tag{5.44b}$$

In other words, the port2, port3 can be arbitrarily set by the boundary condition. Consider that the port3 is connected to the diode, the port2 is designated as the output port with capacitive load, embedding (5.43) and (5.26) into (5.41) without changing the unknown variables would result in

the nonlinear parametric port formulation in (5.33)

$$\begin{bmatrix} G_1 & 0 & -G_1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ -G_1 & 0 & G_1+G_2 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & G_3 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ i_{s1} \\ i_{s2} \\ i_{s3} \end{bmatrix}$$

$$+ \left( \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & C_1 & 0 & -C_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -C_1 & 0 & C_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & c & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \right) \begin{bmatrix} \dot{v}_1 \\ \dot{v}_2 \\ \dot{v}_3 \\ \dot{v}_4 \\ \dot{i}_{s1} \\ \dot{i}_{s2} \\ \dot{i}_{s3} \end{bmatrix}$$

$$+ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ I_s\left(e^{\frac{v_{34}}{v_T}} - 1\right) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} v_1 \end{bmatrix} \tag{5.45}$$

The transient responses for (5.45) are sampled from the initial time point to the terminal time point to form the subspace containing the system responses over the desired input waveforms and output loads. Then the congruence transformation matrix is found by taking the dominant

direction of this subspace by using singular value decomposition. Having the transformation matrix, the congruence transformation is applied to (5.41) to find the reduced system. It is to be noted that the reduced system is a three-port network with an input port, an output port connected to a capacitive load and a constrained port connected to the nonlinear element, diode. After doing the sparsification as explained in Section 5.4.2, the reduced sparsified macromodel is in the form of (5.17). The final step is to connect the diode back to the nonlinear constrained port, which results in the macromodel expressed in (5.19) as shown in Fig. 5.6, where:

$$\boldsymbol{G}_\phi = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}; \quad \boldsymbol{x}_\phi(t) = \begin{bmatrix} v_1(t) \\ v_2(t) \\ v_{34}(t) \\ i_{s1}(t) \\ i_{s2}(t) \\ i_{s3}(t) \end{bmatrix} \tag{5.46}$$

$$\boldsymbol{R}_\phi = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}; \quad \boldsymbol{f}_\phi(\boldsymbol{x}_\phi(t)) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ I_s\left(e^{\frac{v_{34}}{v_T}} - 1\right) \end{bmatrix}; \quad \boldsymbol{u}(t) = \begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix} \tag{5.47}$$

## 5.5 Nonlinear Parametric Macromodel

The nonlinear macromodel presented in the previous section is valid for a specific nonlinear circuit. This macromodel has to be regenerated each time as the internal parameters change. These parameters could be the length and width of transmission lines or the values of resistances

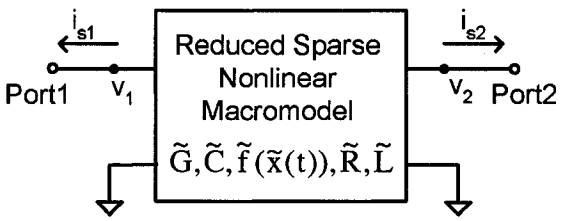


**Fig. 5.6** Reduced nonlinear sparse macromodel

and capacitances of circuits such as the parametric circuit shown in Fig. 5.10. Given the fact that the system equations for interconnect networks are very large due to the discretization, it is computational time consuming for performing simulation. In order to improve the efficiency, a nonlinear parametric macromodel technique is presented in this section.

Consider the parametric nonlinear networks in (5.2) with $M$ input ports, $N$ output ports, $E$ nonlinear elements and $D$ parametric elements, because of the parametric terms $(\sum_{i=1}^{r} h_i(\lambda_i)\boldsymbol{D}_i)$, $\left(\sum_{j=1}^{c} g_j(\phi_j)\boldsymbol{E}_j\right)$ and the nonlinear terms $\boldsymbol{f}(\boldsymbol{x}(t))$, the standard sparsification process cannot be operated. They are decoupled from the original equation by adding constrained ports for the parametric terms and nonlinear terms. The resulting formula suitable for the sparsification is as follows:

$$\boldsymbol{G}_w\boldsymbol{x}_w(t) + \boldsymbol{C}_w\dot{\boldsymbol{x}}_w(t) = \boldsymbol{R}_{w_1}\boldsymbol{u}(t) + \boldsymbol{b}_w(t) + \boldsymbol{R}_{w_2}\boldsymbol{u}_2 + \boldsymbol{R}_{w_3}\boldsymbol{u}_3 \tag{5.48a}$$

$$\boldsymbol{i}(t) = \boldsymbol{L}_{w_1}^T\boldsymbol{x}_w(t) \tag{5.48b}$$

$$\boldsymbol{f}_w(\boldsymbol{u}_2) = \boldsymbol{L}_{w_2}^T\boldsymbol{x}_w(t) \tag{5.48c}$$

$$\boldsymbol{g}_w(\boldsymbol{u}_3) = \boldsymbol{L}_{w_3}^T\boldsymbol{x}_w(t) \tag{5.48d}$$

Equation (5.48c) and (5.48d) represent the nonlinear elements and parametric elements respectively expressed as the nonlinear or parametric conditions on the newly introduced constrained ports. Therefore, in the parametric nonlinear system, we have two kinds of constrained ports, parametric and nonlinear ports. The nonlinear constrained ports are connected to nonlinear elements, while parametric constrained ports are connected to parametric elements. Having the

formulation, the real congruence transformation is applied to the system in (5.48) to obtain:

$$\hat{\boldsymbol{G}}_w\hat{\boldsymbol{x}}_w(t) + \hat{\boldsymbol{C}}_w\dot{\hat{\boldsymbol{x}}}_w(t) = \hat{\boldsymbol{R}}_{w_1}\boldsymbol{u}(t) + \hat{\boldsymbol{b}}_w(t) + \hat{\boldsymbol{R}}_{w_2}\hat{\boldsymbol{u}}_2 + \hat{\boldsymbol{R}}_{w_3}\hat{\boldsymbol{u}}_3 \tag{5.49a}$$

$$\hat{\boldsymbol{i}}(t) = \hat{\boldsymbol{L}}_{w_1}^T\hat{\boldsymbol{x}}_w(t) \tag{5.49b}$$

$$\hat{\boldsymbol{f}}_w(\hat{\boldsymbol{u}}_2) = \hat{\boldsymbol{L}}_{w_2}^T\hat{\boldsymbol{x}}_w(t) \tag{5.49c}$$

$$\hat{\boldsymbol{g}}_w(\hat{\boldsymbol{u}}_3) = \hat{\boldsymbol{L}}_{w_3}^T\hat{\boldsymbol{x}}_w(t) \tag{5.49d}$$

where:

$$\hat{\boldsymbol{G}}_w = \boldsymbol{Q}_w^T\boldsymbol{G}_w\boldsymbol{Q}_w; \quad \hat{\boldsymbol{C}}_w = \boldsymbol{Q}_w^T\boldsymbol{C}_w\boldsymbol{Q}_w \tag{5.50a}$$

$$\hat{\boldsymbol{R}}_{w_1} = \boldsymbol{Q}_w^T\boldsymbol{R}_{w_1}; \quad \hat{\boldsymbol{R}}_{w_2} = \boldsymbol{Q}_w^T\boldsymbol{R}_{w_2} \tag{5.50b}$$

$$\hat{\boldsymbol{L}}_{w_1} = \boldsymbol{Q}_w^T\boldsymbol{L}_{w_1}; \quad \hat{\boldsymbol{b}}_w(t) = \boldsymbol{Q}_w^T\boldsymbol{b}_w(t) \tag{5.50c}$$

$$\hat{\boldsymbol{L}}_{w_2} = \boldsymbol{Q}_w^T\boldsymbol{L}_{w_2}; \quad \hat{\boldsymbol{L}}_{w_3} = \boldsymbol{Q}_w^T\boldsymbol{L}_{w_3} \tag{5.50d}$$

To sparsify the reduced order model in (5.49a), the diagonalizing process explained in the Section 5.4.2 is performed, which results in

$$\breve{\boldsymbol{G}}_w\breve{\boldsymbol{x}}_w(t) + \breve{\boldsymbol{C}}_w\dot{\breve{\boldsymbol{x}}}_w(t) = \breve{\boldsymbol{R}}_{w_1}\boldsymbol{u}(t) + \breve{\boldsymbol{b}}_w(t) + \breve{\boldsymbol{R}}_{w_2}\breve{\boldsymbol{u}}_2 + \breve{\boldsymbol{R}}_{w_3}\breve{\boldsymbol{u}}_3 \tag{5.51a}$$

$$\breve{\boldsymbol{i}}(t) = \breve{\boldsymbol{L}}_{w_1}^T\breve{\boldsymbol{x}}_w(t) \tag{5.51b}$$

$$\breve{\boldsymbol{f}}_w(\breve{\boldsymbol{u}}_2) = \breve{\boldsymbol{L}}_{w_2}^T\breve{\boldsymbol{x}}_w(t) \tag{5.51c}$$

$$\breve{\boldsymbol{g}}_w(\breve{\boldsymbol{u}}_3) = \breve{\boldsymbol{L}}_{w_3}^T\breve{\boldsymbol{x}}_w(t) \tag{5.51d}$$

Finally the nonlinear constraints in (5.51c) and parametric constraints in (5.51d) are reincorporated into the overall macromodel equation by taking (5.51a) and (5.51b) as a subsection. The

following sparse reduced order macromodel is obtained

$$\left(\tilde{\boldsymbol{G}} + \sum_{i=1}^{r} h_i(\lambda_i)\tilde{\boldsymbol{D}}_i\right)\tilde{\boldsymbol{x}}(t) + \tilde{\boldsymbol{f}}(\tilde{\boldsymbol{x}}(t)) + \left(\tilde{\boldsymbol{C}} + \sum_{j=1}^{c} g_j(\phi_j)\tilde{\boldsymbol{E}}_j\right)\dot{\tilde{\boldsymbol{x}}}(t) = \tilde{\boldsymbol{R}}\boldsymbol{u}(t) + \tilde{\boldsymbol{b}}(t) \quad (5.52a)$$

$$\boldsymbol{i} = \tilde{\boldsymbol{L}}^T\tilde{\boldsymbol{x}}(t) \quad (5.52b)$$

where $\tilde{\boldsymbol{G}}$, $\tilde{\boldsymbol{C}}$, $\tilde{\boldsymbol{D}}_i$ and $\tilde{\boldsymbol{E}}_j$ are small and sparse matrices.

The congruence transformation $\boldsymbol{Q}_w$ used in model order reduction must contain the information about the parametric elements, the nonlinear elements and the loads. In order to do that, the load information, the nonlinear elements information as well as the parametric information are incorporated into the formulation in (5.48a) and result in

$$\left(\bar{\boldsymbol{G}}_w + \boldsymbol{G}_N + \sum_{i=1}^{r} h_i(\lambda_i)\bar{\boldsymbol{D}}_i\right)\boldsymbol{x}_w(t) + \left(\bar{\boldsymbol{C}}_w + \sum_{j=1}^{c} g_j(\phi_j)\bar{\boldsymbol{E}}_j\right)\dot{\boldsymbol{x}}_w(t) + \bar{\boldsymbol{f}}(\boldsymbol{x}_w(t)) = \boldsymbol{R}_M\boldsymbol{u}_M(t) + \boldsymbol{b}_w(t) \quad (5.53)$$

$$\left(\bar{\boldsymbol{G}}_w + \sum_{i=1}^{r} h_i(\lambda_i)\bar{\boldsymbol{D}}_i\right)\boldsymbol{x}_w(t) + \left(\bar{\boldsymbol{C}}_w + \boldsymbol{C}_N + \sum_{j=1}^{c} g_j(\phi_j)\bar{\boldsymbol{E}}_j\right)\dot{\boldsymbol{x}}_w(t) + \bar{\boldsymbol{f}}(\boldsymbol{x}_w(t)) = \boldsymbol{R}_M\boldsymbol{u}_M(t) + \boldsymbol{b}_w(t) \quad (5.54)$$

$$\left(\bar{\boldsymbol{G}}_w + \boldsymbol{G}_N + \sum_{i=1}^{r} h_i(\lambda_i)\bar{\boldsymbol{D}}_i\right)\boldsymbol{x}_w(t) + \left(\bar{\boldsymbol{C}}_w + \boldsymbol{C}_N + \sum_{j=1}^{c} g_j(\phi_j)\bar{\boldsymbol{E}}_j\right)\dot{\boldsymbol{x}}_w(t) + \bar{\boldsymbol{f}}(\boldsymbol{x}_w(t)) = \boldsymbol{R}_M\boldsymbol{u}_M(t) + \boldsymbol{b}_w(t) \quad (5.55)$$

Equation (5.53), (5.54), and (5.55) are nonlinear parametric formulation with resistive loads, capacitive loads and combination of resistive and capacitive loads respectively. Where $\boldsymbol{G}_N$ contains resistive load parameters, $g_1, g_2, \cdots, g_N$. $\boldsymbol{C}_N$ contains capacitive load parameters, $c_1, c_2, \cdots, c_N$. $\bar{\boldsymbol{G}}_w$ is obtained by modifying $(N + E + D)$ rows corresponding to output ports, nonlinear con-

strained ports and parametric constrained ports equations, from port voltages into port currents. $\bar{\boldsymbol{D}}_i$ and $\bar{\boldsymbol{E}}_j$ contains parametric information. $\boldsymbol{R}_M$ is a selector matrix that maps input port voltages into the node space of the circuit.

In order to obtain the required subspace data, transient analyses over the predefined input waveforms, load conditions and parametric conditions using formulation in (5.53), (5.54) or (5.55) are performed. The congruence transformation matrix $\boldsymbol{Q}_w$ is found by extracting the dominant subspace using singular value decomposition.

## 5.6 Numerical Results

In this section, we present two examples. The first example considers an interconnect system containing nonlinear elements. The second example considers a nonlinear parametric interconnect system. The transient responses using the proposed method are compared to those obtained from original system. As was expected, the results match very well. Also CPU comparisons for two examples demonstrate the efficiency of the proposed method. The proposed algorithms for all examples were implemented in MATLAB.

### 5.6.1 Example 1

The first example is a nonlinear network containing 9 coupled transmission lines, nine single transmission lines and two inverters. The length for the coupled transmission lines is $0.1m$ and the length for the single transmission lines is $0.05m$. The per unit length parameters of the 9 coupled transmission lines are given in [88]. The per unit length parameters of the single transmission lines are $R = 3.74\Omega/m, L = 284nH/m, C = 84.6pF/m$. Inverter1 is connected between two transmission lines, while inverter2 is connected at the output of one transmission line. The output of inverter2 is considered as the output of the network as shown in Fig. 5.7. After discretization of network, the resulting size of the modified nodal analysis matrices was 3533. Four different input waveforms are used for computing the subspace data. They are 5V

step inputs with rise time of 50ps and 500ps and 5V step inputs with fall time of 50ps and 500ps as shown in Fig. 5.3. The load condition was set to be a capacitor with value ranging from 0.1pF to 10pF. Using the proposed approach, the size of the reduced macromodel was 320 as shown in Table 5.2.

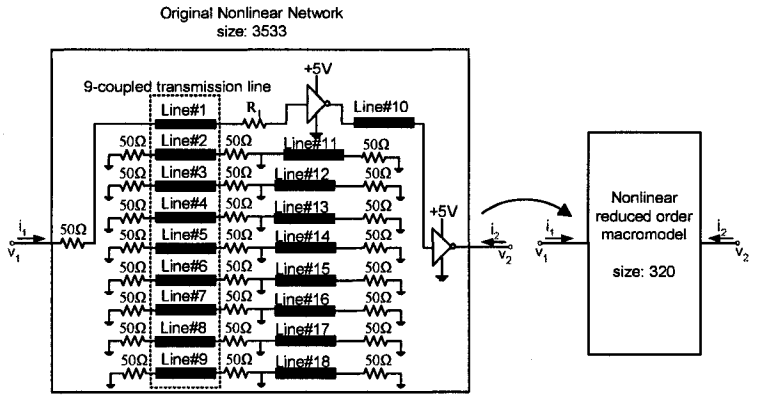


**Fig. 5.7** Interconnect network with two inverters as example 1

**Table 5.1** Three Cases for Example 1

| | C | Input Waveform |
|---|---|---|
| Case 1 | 3.2pF | step input with 488ps fall time |
| Case 2 | 0.71pF | step input with 79ps rise time |
| Case 3 | 0.49pF | sinusoidal input at frequency 2GHz |

In order to test the proposed method, a capacitor was connected at the output of the reduced macromodel. Different cases of load capacitor values and input waveforms were considered. The transient responses for three of these cases are shown in Fig. 5.8 and Fig. 5.9. The values of the loads and input waveform for each case are given in Table 5.1. As can be seen the results match very well with the transient responses of the original system. The CPU cost of the reduced system for three cases in Table 5.1 range from 38s to 78s, while the CPU cost for the original system range from 770s to 938s. The average speed-up of 15.8 was therefore achieved. A summary of CPU cost comparisons for example 1 to obtain the transient responses is shown in Table 5.2.

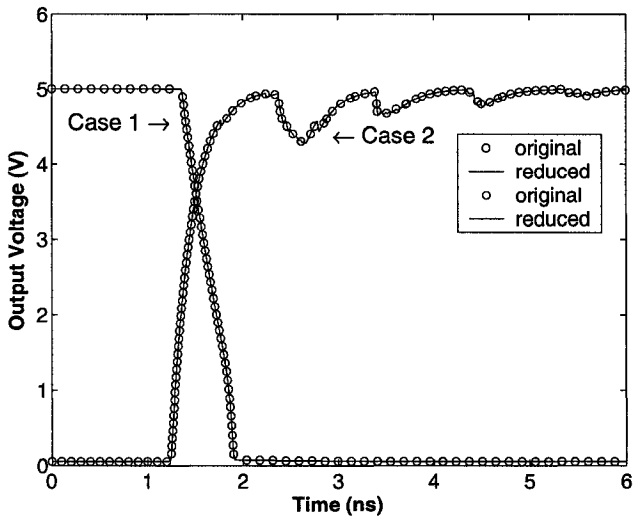


**Fig. 5.8** Transient response comparison between the original system and the proposed macromodel for example 1 (cases 1 and 2)

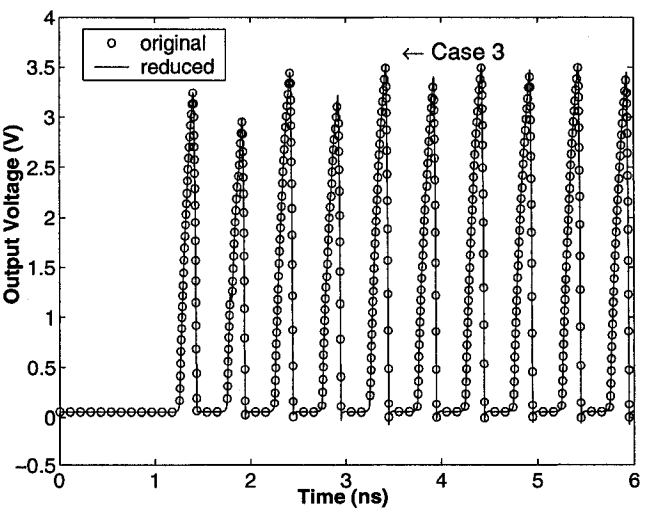


**Fig. 5.9** Transient response comparison between the original system and the proposed macromodel for example 1 (case 3)

**Table 5.2** Size and CPU Cost Comparisons

| | Example 1 |
|---|---|
| Size of Original System | 3533 |
| Size of Proposed Macromodel | 320 |
| CPU Cost for Original System | 770s $\sim$ 938s |
| CPU Cost for Reduced System | 38s $\sim$ 78s |
| Average CPU Cost Speed-up | 15.8 |

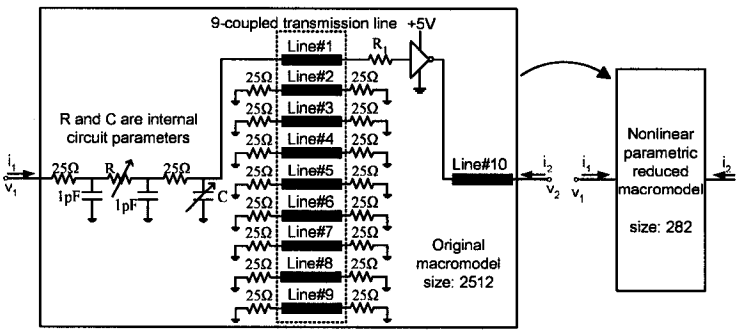


**Fig. 5.10** Interconnect network with two internal circuit parameters as example 2

**Table 5.3** Size Comparisons for Example 2

| | |
|---|---|
| Size of Original System | 2512 |
| Size of Proposed Macromodel | 282 |

**Table 5.4** Three sample cases from 100 test cases for example 2

| | Parameter#1 $R$ | Parameter#2 $C$ | Load $C_L$ | Input Waveform |
|---|---|---|---|---|
| Case 1 | 76Ω | 6.2pF | 0.24pF | step input with 150ps rise time |
| Case 2 | 11Ω | 2.1pF | 3.7pF | 4GHz sine waveform |
| Case 3 | 55Ω | 0.12pF | 0.86pF | step input with 590ps fall time |

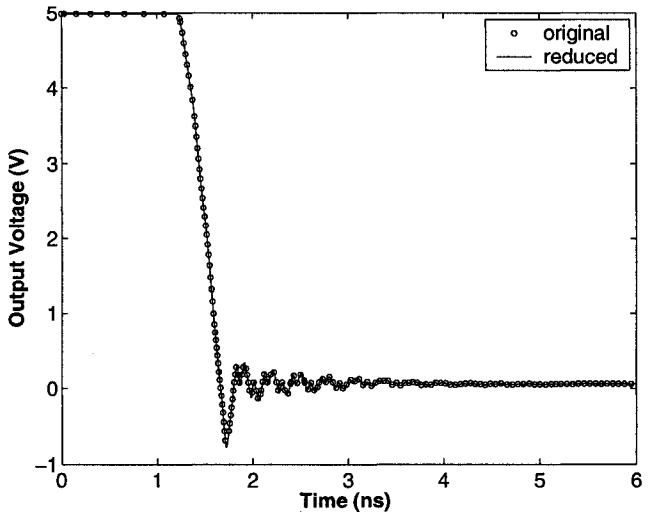


**Fig. 5.11** Transient response comparison between the original system and the proposed macromodel for example 2 (cases 1)

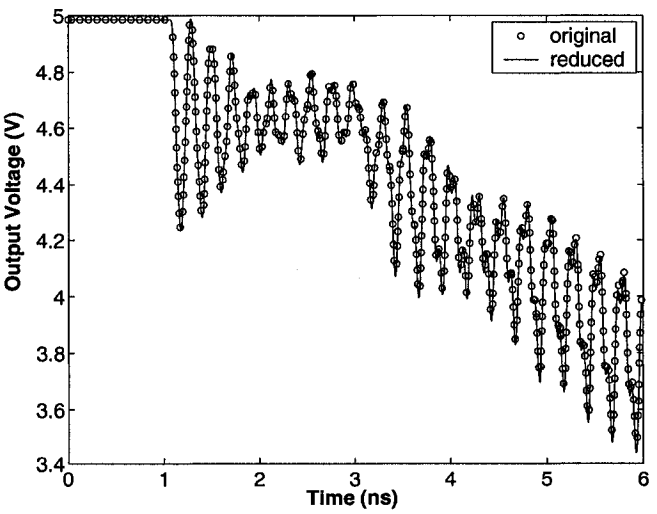


**Fig. 5.12** Transient response comparison between the original system and the proposed macromodel for example 2 (case 2)

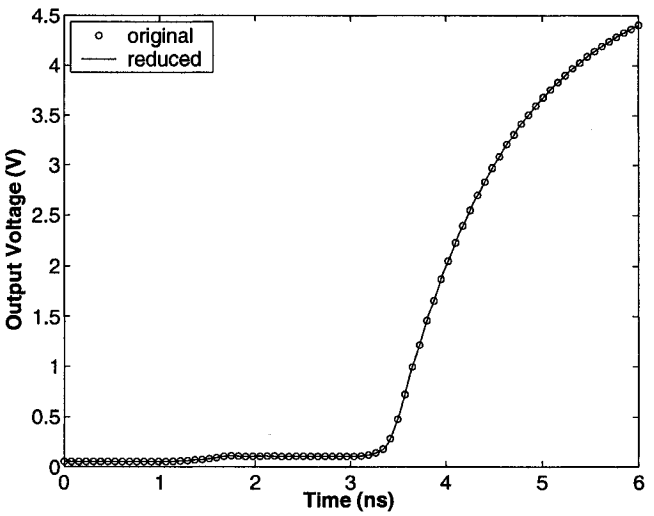


**Fig. 5.13** Transient response comparison between the original system and the proposed macromodel for example 2 (case 3)

**Table 5.5** CPU Comparisons for Example 2 Based on 100 Test Cases

| | Reduced System | Original System | Average speed-up |
|---|---|---|---|
| Reduction overhead | 2191.7s | - | - |
| Simulation for 100 test cases | 1883.8s | 46349.9s | 24.6 |
| Overall for 100 test cases | 4075.5s | 46349.9s | 11.4 |

### 5.6.2 Example 2

The second example is an interconnect network with one single transmission line, nine coupled transmission lines system and one inverter. The length for the coupled transmission lines is $0.1m$ and the length for the single transmission line is $0.05m$. The per unit length parameters of the 9 coupled transmission lines are given in [88]. The per unit length parameters of the single transmission line are $R = 3.74\Omega/m, L = 28.4nH/m, C = 64.6pF/m$. This system is a parametric 2-port network with 1 resistor $R$ and 1 capacitor $C$ as internal circuit parameters as shown in Fig. 5.10. For this network, the original modified nodal analysis matrix size is 2512. When generating the subspace data, the desired range for the parametric resistor was set to $1 \sim 100\Omega$, and the desired range for the parametric capacitor was set to $0.1 \sim 10pF$. The load at the output port was set to be capacitive with $C_L$ ranging from $0.1 \sim 10pF$. The input waveforms used for generating the subspace data were 5V rising edge and falling edge steps with 50ps and 500ps rise/fall time as shown in Fig. 5.3.

Using the parametric nonlinear macromodel described in this chapter, the size of the reduced model is 282 as shown in Table 5.3. In order to test the accuracy and efficiency of the macromodel, one hundred randomly chosen sample cases are tested. The parameter values as well as the load values were chosen randomly within the acceptable range defined above. The input waveforms for testing include the step input with rise time randomly chosen between 50ps and 500ps; step input with fall time randomly chosen between 50ps and 500ps; pulse waveforms with different rise and fall times and different pulse widths; piece wise linear waveforms; sinusoidal waveform at frequencies ranging from 1GHz to 4GHz. The transient responses for three cases from one hundred test cases are shown in Fig. 5.11, Fig. 5.12 and Fig. 5.13. The values of the loads and input waveforms for the three test cases are given in Table 5.4.

In order to calculate the CPU speed-up for the reduced order macromodel over the original system, we include the CPU time to obtain the macromodel, which is referred to the reduction overhead time. The overhead time consists of the time for generating the subspace data (1776.6s) and the time for doing singular value decomposition (415.1s). The time to do the simulation

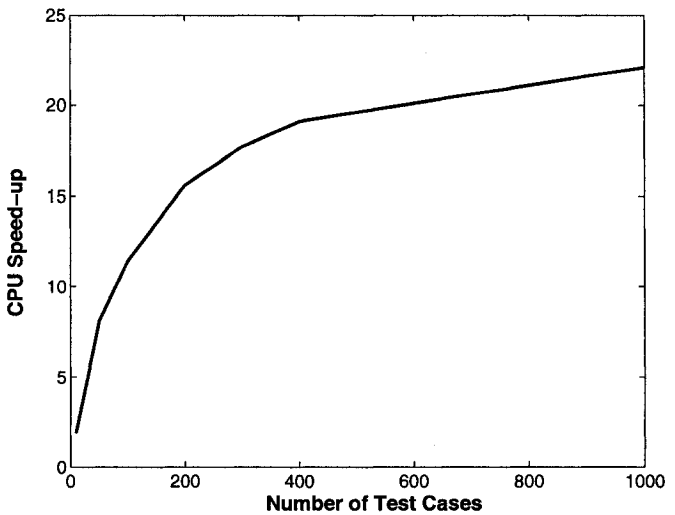


**Fig. 5.14** Average CPU speed-up versus number of test cases

for 100 test cases using the reduced macromodel is 1883.8s, the overall time of the 100 test cases for the proposed approach is 4075.5s when the reduction overhead is included. On the other hand, the simulation time of the original system for 100 test cases is 46349.9s. Therefore a CPU speed-up of 11.4 is achieved as shown in Table 5.5. It is to be noted that the time for generating the macromodel is a one-time cost. It is therefore leveraged over many simulation runs of the reduced macromodel. As can be seen from Fig. 5.14, the overall speed-up approaches the simulation speed-up of 24.6 when the number of test cases is large.

# Chapter 6

# Summary and Future work

## 6.1 Summary

This thesis presents a number of methods to enhance the efficiency and accuracy for modeling and simulation of interconnect networks. First, a new method for the computation of the congruence transformation matrix to handle systems with a large number of ports has been proposed. The new approach is based on taking advantage of prior information regarding to the port conditions. Specifically, resistive, capacitive, transmission line loads as well as nonlinear loads such as diodes and inverters are considered. For that purpose, a new parametric port formulation has been developed which allows us to embed the load parameters without any modification to the vector of unknowns in the modified nodal analysis (MNA) formulation. Parametric model order reduction technique is then used to obtain the reduction subspace. It has been shown that the block moments with respect to frequency as well as the block moments with respect to the load parameters are conserved. As can be seen in the examples, the size of the proposed reduced macromodels is much less sensitive to the number of ports than those obtained from traditional methods. This demonstrates that the proposed method significantly extends the range of applicability of model order reduction to systems with a large number of ports when the information about the types of loads is available.

Secondly, new multi-level reduction methods based on singular value decomposition (SVD) for regular as well as parametric systems have been developed. The new SVD based reduction can reduce by more than one-third the size of the macromodels compared to those obtained from the traditional techniques. In addition, a new parametric formulation has been developed to allow the sparsification of parametric networks. This new formulation replaces the stamps of the parametric elements with constrained port formulation. The reduction procedure and sparsification are then performed on the macromodel in the space of the new formulation. CPU speed-ups from 1 to 2 orders of magnitude have been achieved due to the sparsification. Furthermore, a new sparse multi-level parametric model order reduction method based on the multi-level reduction and sparsification has been developed. The proposed macromodel can achieve up to 350 times faster than the original system and therefore significantly improve the simulation efficiency.

Finally, a time domain nonlinear macromodeling technique, which is able to deal with arbitrary nonlinearity, has been developed as well as a nonlinear formulation suitable for sparsification has been proposed. This results in an efficient reduced order nonlinear macromodel which is sparse, and is valid over a predefined range of input waveforms and load conditions. Furthermore, a sparse parametric nonlinear model order reduction approach has been developed. Using this method, the macromodel only needs to be created once, and can be reused many times over different input waveforms, different load conditions, as well as different internal circuit parameters. As demonstrated in the examples, the proposed nonlinear macromodel can achieve up to 40 speed-up over the original nonlinear system to obtain the transient responses. This has many applications for interconnect networks containing nonlinear elements and internal circuit parameters, and also has the potential to provide an automatic and systematic approach to develop dynamic macromodels for nonlinear drivers.

## 6.2 Future work

1. Passive macromodeling for interconnect networks characterized by tabulated/measured data: The model order reduction methods proposed in this thesis are aimed for intercon-

nect networks which can be characterized by mathematically partial differential equations. However, in some cases, such analytical equations are not available and one must rely on experimental measurements or results of full wave simulation in order to characterize the interconnect networks. A number of methods have been proposed in the literature to address the macromodeling problem for interconnect networks characterized by tabulated/measured data. However obtaining an efficient macromodel for interconnect networks based on tabulated data while preserving important properties such as passivity is still an open problem facing the VLSI design and simulation community.

2. Efficient nonlinear macromodels for analog/RF circuits: In this thesis, projection based model order reduction was proposed for nonlinear macromodeling in the time domain. This method is general for any nonlinear network but is more suitable for large interconnect networks with a few nonlinear elements, and the sparsity of the resulting macromodel would deteriorate as the number of nonlinear elements becomes large. It is therefore worthwhile to explore methods for obtaining efficient nonlinear macromodels which are targeted towards specific classes of analog/RF circuits (e.g. amplifiers, and mixers) and using this information and insight into the circuit operation to develop systematic approaches for accurate nonlinear macromodels for such circuits.

3. Nonlinear macromodeling for Microelectromechanical System (MEMS): Integrating electronics with miniature MEMS devices at the microscopic level has caused the complexity of design to increase dramatically. The rise in complexity has made the CAD tools a more essential part of the design cycle of MEMS. A promising solution to the rising complexity problem is the use of accurate and efficient reduced model. Current macromodeling techniques for MEMS have generally focused on linearized models, however, in many cases, a nonlinear macromodel is needed. The nonlinear macromodeling technique proposed in this thesis can be extended to model nonlinear MEMS.

# Appendix A

# Proof of Conservation of Moments

In this appendix, the proof of conservation of moments is presented. This proof that the reduced system conserves $q_{g_h}$ moments ($\boldsymbol{M}_{g_h} \in \mathbb{R}^{n \times q_{g_h}}$) with respect to conductance $g_h$ is based on mathematical induction. First we prove that the $zero$th moment obtained from the reduced system is conserved with respect to the conductance. Next, we demonstrate that the $k$th moment is conserved if the previous $(k-1)$th moment is conserved.

The $zero$th block moment $\boldsymbol{M}^0$ at the expansion point $s = 0, g_1 = g_1^a, \cdots, g_N = g_N^a, c_1 = 0, \cdots, c_N = 0$ is given by:

$$\boldsymbol{G}_\phi \boldsymbol{M}^0 + \boldsymbol{D}_n^a \boldsymbol{R}_N \boldsymbol{L}_N^T \boldsymbol{M}^0 + \boldsymbol{R}_N \boldsymbol{R}_N^T \boldsymbol{M}^0 = \boldsymbol{R}_M \tag{A.1}$$

where $\boldsymbol{D}_n^a \in \mathbb{R}^{n \times n}$ is a diagonal matrix. The first $n - N$ diagonal elements are zero, while the last $N$ diagonal elements are $g_j^a$, $j = 1, \cdots, N$.

The $k$th block moment ($k \neq 0$) with respect to frequency, $\boldsymbol{M}_s^k$, at the expansion points $s = 0, g_1 = g_1^a, \cdots, g_N = g_N^a, c_1 = 0, \cdots, c_N = 0$ can be evaluated from (3.23) and is given by:

$$\boldsymbol{G}_\phi \boldsymbol{M}_s^k + \boldsymbol{C} \boldsymbol{M}_s^{(k-1)} + \boldsymbol{D}_n^a \boldsymbol{R}_N \boldsymbol{L}_N^T \boldsymbol{M}_s^k + \boldsymbol{R}_N \boldsymbol{R}_N^T \boldsymbol{M}_s^k = 0 \tag{A.2}$$

note that $\boldsymbol{M}_s^0 = \boldsymbol{M}^0$ was calculated in (A.1).

The $k$th block moment ($k \neq 0$) with respect to $g_h$, $\boldsymbol{M}_{g_h}^k$, at the expansion points $s = 0, g_1 = g_1^a, \cdots, g_N = g_N^a, c_1 = 0, \cdots, c_N = 0$ can also be calculated from (3.23) and is given by:

$$\boldsymbol{G}_\phi \boldsymbol{M}_{g_h}^k + \boldsymbol{r}_{M+h} \boldsymbol{l}_{M+h}^T \boldsymbol{M}_{g_h}^{(k-1)} + \boldsymbol{D}_n^a \boldsymbol{R}_N \boldsymbol{L}_N^T \boldsymbol{M}_{g_h}^k + \boldsymbol{R}_N \boldsymbol{R}_N^T \boldsymbol{M}_{g_h}^k = 0 \tag{A.3}$$

note that $\boldsymbol{M}_{g_h}^0 = \boldsymbol{M}^0$ was calculated in (A.1).

As shown in (A.1), at an expansion point, $s = 0$, $g_1 = g_1^a, \cdots, g_N = g_N^a$. The *zero*th block moment $\boldsymbol{M}_{g_h}^0$ with respect to $g_h$ is the solution of

$$\boldsymbol{G}_\phi \boldsymbol{M}_{g_h}^0 + \boldsymbol{D}_n^a \boldsymbol{R}_N \boldsymbol{L}_N^T \boldsymbol{M}_{g_h}^0 + \boldsymbol{R}_N \boldsymbol{R}_N^T \boldsymbol{M}_{g_h}^0 = \boldsymbol{R}_M \tag{A.4}$$

It is to be noted that the first $(n - N)$ rows of $\boldsymbol{D}_n^a \boldsymbol{R}_N \boldsymbol{L}_N^T \boldsymbol{M}_{g_h}^0$ and $\boldsymbol{R}_N \boldsymbol{R}_N^T \boldsymbol{M}_{g_h}^0$ are zero, while the last $N$ rows of $\boldsymbol{G}_\phi \boldsymbol{M}_{g_h}^0$ and $\boldsymbol{R}_M$ are zero. The above equation can be separated into two parts: the first $(n - N)$ equations and the last $N$ equations.

$$\boldsymbol{G}_\phi \boldsymbol{M}_{g_h}^0 = \boldsymbol{R}_M \tag{A.5}$$

$$\boldsymbol{D}_N^a \boldsymbol{L}_N^T \boldsymbol{M}_{g_h}^0 + \boldsymbol{R}_N^T \boldsymbol{M}_{g_h}^0 = 0 \tag{A.6}$$

where $\boldsymbol{D}_N^a \in \mathbb{R}^{N \times N}$ is a diagonal matrix with diagonal elements $g_j^a$, $j = 1, \cdots, N$. Note that $\boldsymbol{L}_N^T \boldsymbol{M}_{g_h}^0$ is nothing more than the *zero*th block moment of the constrained port voltages, denoted by $\boldsymbol{M}_{g_h(V_N)}^0 \in \mathbb{R}^{N \times M}$, while $\boldsymbol{R}_N^T \boldsymbol{M}_{g_h}^0$ is nothing more than the *zero*th block moment of the constrained port currents, denoted by $\boldsymbol{M}_{g_h(I_N)}^0 \in \mathbb{R}^{N \times M}$, therefore (A.6) is rewritten as follows:

$$\boldsymbol{D}_N^a \boldsymbol{M}_{g_h(V_N)}^0 + \boldsymbol{M}_{g_h(I_N)}^0 = 0 \tag{A.7}$$

To facilitate the presentation, the original system (3.1) is expressed in a similar form as the

parametric port model in (3.21). The original system can thus be recast as:

$$\boldsymbol{G}_{\phi}\boldsymbol{x} + s\boldsymbol{C}\boldsymbol{x} + \boldsymbol{R}_N\boldsymbol{L}_N^T\boldsymbol{x} = \boldsymbol{R}_M\boldsymbol{u}_M + \boldsymbol{R}_N\boldsymbol{V}_N \tag{A.8}$$

$$\boldsymbol{I}_N = \boldsymbol{R}_N^T\boldsymbol{x} = -\boldsymbol{D}_N\boldsymbol{V}_N \tag{A.9}$$

Where

$$\boldsymbol{G}_{\phi} + \boldsymbol{R}_N\boldsymbol{L}_N^T = \boldsymbol{G} \tag{A.10}$$

$\boldsymbol{I}_N \in \mathbb{R}^{N\times 1}$ is a vector containing constrained port currents, $\boldsymbol{V}_N \in \mathbb{R}^{N\times 1}$ is a vector containing constrained port voltages. $\boldsymbol{D}_N \in \mathbb{R}^{N\times N}$ diagonal matrix with diagonal elements $g_j$, $j = 1, \cdots, N$. On the other hand, the reduced system can be expressed in form of :

$$\hat{\boldsymbol{G}}_{\phi}\hat{\boldsymbol{x}} + s\hat{\boldsymbol{C}}\hat{\boldsymbol{x}} + \hat{\boldsymbol{R}}_N\hat{\boldsymbol{L}}_N^T\hat{\boldsymbol{x}} = \hat{\boldsymbol{R}}_M\boldsymbol{V}_M + \hat{\boldsymbol{R}}_N\hat{\boldsymbol{V}}_N \tag{A.11}$$

$$\hat{\boldsymbol{I}}_N = \hat{\boldsymbol{R}}_N^T\hat{\boldsymbol{x}} = -\boldsymbol{D}_N\hat{\boldsymbol{V}}_N \tag{A.12}$$

where $\hat{\boldsymbol{I}}_N$ and $\hat{\boldsymbol{V}}_N \in \mathbb{R}^{N\times 1}$ are vectors containing respectively the constrained port currents and constrained port voltages in the reduced system. It follows from (A.11) and (A.12) that at an expansion point, $s = 0$, $g_1 = g_1^a, \cdots, g_N = g_N^a$, the *zero*th block moment for the reduced system is determined by:

$$\hat{\boldsymbol{G}}_{\phi}\hat{\boldsymbol{M}}_{g_h}^0 + \hat{\boldsymbol{R}}_N\hat{\boldsymbol{L}}_N^T\hat{\boldsymbol{M}}_{g_h}^0 = \hat{\boldsymbol{R}}_M + \hat{\boldsymbol{R}}_N\hat{\boldsymbol{M}}_{g_h(V_N)}^0 \tag{A.13}$$

$$\hat{\boldsymbol{M}}_{g_h(I_N)}^0 = \hat{\boldsymbol{R}}_N^T\hat{\boldsymbol{M}}_{g_h}^0 = -\boldsymbol{D}_N^a\hat{\boldsymbol{M}}_{g_h(V_N)}^0 \tag{A.14}$$

where $\hat{\boldsymbol{M}}_{g_h}^0$ is the *zero*th block moment for the reduced system. $\hat{\boldsymbol{M}}_{g_h(V_N)}^0 \in \mathbb{R}^{N\times M}$ and $\hat{\boldsymbol{M}}_{g_h(I_N)}^0 \in \mathbb{R}^{N\times M}$ are respectively the *zero*th block moment of constrained port voltages and constrained port currents for the reduced system. Substituting (3.26) into (A.13), we get:

$$\boldsymbol{Q}^T\boldsymbol{G}_{\phi}\boldsymbol{Q}\hat{\boldsymbol{M}}_{g_h}^0 + \boldsymbol{Q}^T\boldsymbol{R}_N\boldsymbol{L}_N^T\boldsymbol{Q}\hat{\boldsymbol{M}}_{g_h}^0 = \boldsymbol{Q}^T\boldsymbol{R}_M + \boldsymbol{Q}^T\boldsymbol{R}_N\hat{\boldsymbol{M}}_{g_h(V_N)}^0 \tag{A.15}$$

Next we substitute $\boldsymbol{K} = \boldsymbol{QR}$ in the above equation and use (A.10), which results in:

$$\boldsymbol{K}^T\boldsymbol{GKR}^{-1}\hat{\boldsymbol{M}}^0_{g_h} - \boldsymbol{K}^T\boldsymbol{R}_N\hat{\boldsymbol{M}}^0_{g_h(V_N)} = \boldsymbol{K}^T\boldsymbol{R}_M \tag{A.16}$$

Substitute $\boldsymbol{R}_M$ from (A.5) into (A.16)

$$\boldsymbol{K}^T\boldsymbol{GKR}^{-1}\hat{\boldsymbol{M}}^0_{g_h} - \boldsymbol{K}^T\boldsymbol{R}_N\hat{\boldsymbol{M}}^0_{g_h(V_N)} = \boldsymbol{K}^T\boldsymbol{G}_\phi\boldsymbol{M}^0_{g_h} \tag{A.17}$$

Referring to (3.13), we have

$$\boldsymbol{M}^0_{g_h(V_N)} = \boldsymbol{L}^T_N\boldsymbol{M}^0_{g_h} \tag{A.18}$$

which results in

$$\boldsymbol{K}^T\boldsymbol{R}_N\boldsymbol{M}^0_{g_h(V_N)} = \boldsymbol{K}^T\boldsymbol{R}_N\boldsymbol{L}^T_N\boldsymbol{M}^0_{g_h} \tag{A.19}$$

Adding (A.17) and (A.19) and using (A.10)

$$\boldsymbol{K}^T\boldsymbol{GKR}^{-1}\hat{\boldsymbol{M}}^0_{g_h} + \boldsymbol{K}^T\boldsymbol{R}_N(\boldsymbol{M}^0_{g_h(V_N)} - \hat{\boldsymbol{M}}^0_{g_h(V_N)})$$

$$= \boldsymbol{K}^T\boldsymbol{GM}^0_{g_h} \tag{A.20}$$

Also substitute (3.26) in (A.14)

$$\boldsymbol{R}^T_N\boldsymbol{Q}\hat{\boldsymbol{M}}^0_{g_h} = -\boldsymbol{D}^a_N\hat{\boldsymbol{M}}^0_{g_h(V_N)} \tag{A.21}$$

Subtracting (A.21) from (A.7), we obtain

$$\boldsymbol{R}^T_N\boldsymbol{Q}\hat{\boldsymbol{M}}^0_{g_h} + \boldsymbol{D}^a_N\hat{\boldsymbol{M}}^0_{g_h(V_N)} - \boldsymbol{D}^a_N\boldsymbol{M}^0_{g_h(V_N)} - \boldsymbol{M}^0_{g_h(I_N)} = 0 \tag{A.22}$$

Now we substitute $\boldsymbol{K} = \boldsymbol{QR}$ into (A.22) and combine the resulting equation with (A.20) in

matrix form to obtain

$$\begin{bmatrix} \boldsymbol{K}^T\boldsymbol{G}\boldsymbol{K} & \boldsymbol{K}^T\boldsymbol{R}_N \\ \boldsymbol{R}_N^T\boldsymbol{K} & -\boldsymbol{D}_N^a \end{bmatrix} \begin{bmatrix} \boldsymbol{R}^{-1}\hat{\boldsymbol{M}}_{g_h}^0 \\ \boldsymbol{M}_{g_h(V_N)}^0 - \hat{\boldsymbol{M}}_{g_h(V_N)}^0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{K}^T\boldsymbol{G}\boldsymbol{M}_{g_h}^0 \\ \boldsymbol{M}_{g_h(I_N)}^0 \end{bmatrix} \tag{A.23}$$

The above equation can be recast as follows:

$$\begin{bmatrix} \boldsymbol{K}^T & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{E(N)} \end{bmatrix} \begin{bmatrix} \boldsymbol{G} & \boldsymbol{R}_N \\ \boldsymbol{R}_N^T & -\boldsymbol{D}_N^a \end{bmatrix} \begin{bmatrix} \boldsymbol{K} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{E(N)} \end{bmatrix} \begin{bmatrix} \boldsymbol{R}^{-1}\hat{\boldsymbol{M}}_{g_h}^0 \\ \boldsymbol{M}_{g_h(V_N)}^0 - \hat{\boldsymbol{M}}_{g_h(V_N)}^0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{K}^T\boldsymbol{G}\boldsymbol{M}_{g_h}^0 \\ \boldsymbol{M}_{g_h(I_N)}^0 \end{bmatrix} \tag{A.24}$$

where $\boldsymbol{I}_{E(N)} \in \mathbb{R}^{N\times N}$ identity matrix. Since (A.24) is nonsingular and therefore has a unique solution, which is the first $M$ columns of the identity matrix $\boldsymbol{e}_{M_1}$.

$$\boldsymbol{R}^{-1}\hat{\boldsymbol{M}}_{g_h}^0 = \boldsymbol{e}_{M_1} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \tag{A.25}$$

It follows that:

$$\boldsymbol{Q}\hat{\boldsymbol{M}}_{g_h}^0 = \boldsymbol{K}\boldsymbol{R}^{-1}\hat{\boldsymbol{M}}_{g_h}^0 = \boldsymbol{K}\boldsymbol{e}_{M_1} = \boldsymbol{M}_{g_h}^0 \tag{A.26}$$

Also, the following equality can be obtained from (A.24):

$$\boldsymbol{M}_{g_h(V_N)}^0 = \hat{\boldsymbol{M}}_{g_h(V_N)}^0 \tag{A.27}$$

Therefore the *zero*th moment is conserved. Next, we demonstrate that if the $(k-1)$th moment

is conserved, then the $k$th moment is also conserved. In other words, if

$$\boldsymbol{Q}\hat{\boldsymbol{M}}_{g_h}^{(k-1)} = \boldsymbol{M}_{g_h}^{(k-1)} \tag{A.28}$$

$$\boldsymbol{M}_{g_h(V_N)}^{k-1} = \hat{\boldsymbol{M}}_{g_h(V_N)}^{k-1} \tag{A.29}$$

then

$$\boldsymbol{Q}\hat{\boldsymbol{M}}_{g_h}^{(k)} = \boldsymbol{M}_{g_h}^{(k)} \tag{A.30}$$

$$\boldsymbol{M}_{g_h(V_N)}^{k} = \hat{\boldsymbol{M}}_{g_h(V_N)}^{k} \tag{A.31}$$

The $k$th block moment is given by (A.3) and has the form of:

$$\boldsymbol{G}_\phi \boldsymbol{M}_{g_h}^k + \boldsymbol{r}_{M+h}\boldsymbol{l}_{M+h}^T \boldsymbol{M}_{g_h}^{(k-1)} + \boldsymbol{D}_n^a \boldsymbol{R}_N \boldsymbol{L}_N^T \boldsymbol{M}_{g_h}^k + \boldsymbol{R}_N \boldsymbol{R}_N^T \boldsymbol{M}_{g_h}^k = 0 \tag{A.32}$$

The above matrix is separated into two parts, the first $n - N$ equations and the last $N$ equations.

$$\boldsymbol{G}_\phi \boldsymbol{M}_{g_h}^k = 0 \tag{A.33}$$

$$\boldsymbol{M}_{g_{h(V_h)}}^{(k-1)} + \boldsymbol{D}_N^a \boldsymbol{M}_{g_h(V_N)}^k + \boldsymbol{M}_{g_h(I_N)}^k = 0 \tag{A.34}$$

where $\boldsymbol{M}_{g_{h(V_h)}}^{(k-1)} \in \mathbb{R}^{N\times M}$ is a matrix with each column containing only one nonzero, the $(k-1)th$ block moment of the $h$th constrained port voltage. $\boldsymbol{M}_{g_h(V_N)}^k$ and $\boldsymbol{M}_{g_h(I_N)}^k \in \mathbb{R}^{N\times M}$ are the $k$th block moment of the constrained port voltages and constrained port currents for the original network. On the other hand, the $k$th block moment with respect to $g_h$ for the reduced system can be calculated from (A.11) and (A.12). Substituting (3.26) into the resulting equations results in

$$\boldsymbol{Q}^T \boldsymbol{G}_\phi \boldsymbol{Q} \hat{\boldsymbol{M}}_{g_h}^k + \boldsymbol{Q}^T \boldsymbol{R}_N \boldsymbol{L}_N^T \boldsymbol{Q} \hat{\boldsymbol{M}}_{g_h}^k = \boldsymbol{Q}^T \boldsymbol{R}_N \hat{\boldsymbol{M}}_{g_h(V_N)}^k \tag{A.35}$$

$$\hat{\boldsymbol{M}}_{g_h(I_N)}^k = \boldsymbol{R}_N^T \boldsymbol{Q} \hat{\boldsymbol{M}}_{g_h}^k = -\boldsymbol{D}_N^a \hat{\boldsymbol{M}}_{g_h(V_N)}^k - \hat{\boldsymbol{M}}_{g_{h(V_h)}}^{(k-1)} \tag{A.36}$$

where $\hat{\boldsymbol{M}}^{(k-1)}_{g_{h(V_h)}} \in \mathbb{R}^{N\times M}$ is a matrix with each column containing only one nonzero, the $(k-1)$th block moment of the $h$th constrained port voltage for the reduced system. $\hat{\boldsymbol{M}}^{k}_{g_{h(V_N)}}$ and $\hat{\boldsymbol{M}}^{k}_{g_{h(I_N)}} \in \mathbb{R}^{N\times M}$ are the $k$th block moment of the constrained port voltages and currents for the reduced system. Inserting $\boldsymbol{K} = \boldsymbol{QR}$ into (A.35) and using (A.10) result in

$$\boldsymbol{K}^T\boldsymbol{GKR}^{-1}\hat{\boldsymbol{M}}^{k}_{g_h} - \boldsymbol{K}^T\boldsymbol{R}_N\hat{\boldsymbol{M}}^{k}_{g_{h(V_N)}} = 0 \tag{A.37}$$

Substitute (A.33) in (A.37)

$$\boldsymbol{K}^T\boldsymbol{GKR}^{-1}\hat{\boldsymbol{M}}^{k}_{g_h} - \boldsymbol{K}^T\boldsymbol{R}_N\hat{\boldsymbol{M}}^{k}_{g_{h(V_N)}} = \boldsymbol{K}^T\boldsymbol{G}_\phi\boldsymbol{M}^{k}_{g_h} \tag{A.38}$$

Referring to (3.13), we have:

$$\boldsymbol{K}^T\boldsymbol{R}_N\boldsymbol{M}^{k}_{g_{h(V_N)}} = \boldsymbol{K}^T\boldsymbol{R}_N\boldsymbol{L}_N^T\boldsymbol{M}^{k}_{g_h} \tag{A.39}$$

Adding (A.38) and (A.39) and using (A.10), we obtain

$$\boldsymbol{K}^T\boldsymbol{GKR}^{-1}\hat{\boldsymbol{M}}^{k}_{g_h} + \boldsymbol{K}^T\boldsymbol{R}_N(\boldsymbol{M}^{k}_{g_{h(V_N)}} - \hat{\boldsymbol{M}}^{k}_{g_{h(V_N)}}) = \boldsymbol{K}^T\boldsymbol{G}\boldsymbol{M}^{k}_{g_h} \tag{A.40}$$

Note that the $(k-1)$th block moment is conserved in (A.29), which results in

$$\boldsymbol{M}^{(k-1)}_{g_{h(V_h)}} = \hat{\boldsymbol{M}}^{(k-1)}_{g_{h(V_h)}} \tag{A.41}$$

Also subtracting (A.36) from (A.34) and using (A.41)

$$\boldsymbol{R}_N^T\boldsymbol{Q}\hat{\boldsymbol{M}}^{k}_{g_h} - \boldsymbol{D}^a_N(\boldsymbol{M}^{k}_{g_{h(V_N)}} - \hat{\boldsymbol{M}}^{k}_{g_{h(V_N)}}) = \boldsymbol{M}^{k}_{g_{h(I_N)}} \tag{A.42}$$

Now we substitute $\boldsymbol{K} = \boldsymbol{QR}$ in (A.42) and combine the resulting equation with (A.40) in matrix

form to obtain

$$\begin{bmatrix} \boldsymbol{K}^T\boldsymbol{G}\boldsymbol{K} & \boldsymbol{K}^T\boldsymbol{R}_N \\ \boldsymbol{R}_N^T\boldsymbol{K} & -\boldsymbol{D}_N^a \end{bmatrix} \begin{bmatrix} \boldsymbol{R}^{-1}\hat{\boldsymbol{M}}_{g_h}^k \\ \boldsymbol{M}_{g_h(V_N)}^k - \hat{\boldsymbol{M}}_{g_h(V_N)}^k \end{bmatrix} = \begin{bmatrix} \boldsymbol{K}^T\boldsymbol{G}\boldsymbol{M}_{g_h}^k \\ \boldsymbol{M}_{g_h(I_N)}^k \end{bmatrix} \tag{A.43}$$

Again, since (A.43) is nonsingular and therefore has only one unique solution, which is the $R$th $M$ columns of the identity matrix $\boldsymbol{e}_{\boldsymbol{M_R}}$, where $R$ corresponds to the location of $\boldsymbol{M}_{g_h}^k$ in $\boldsymbol{K}$. It follows that:

$$\boldsymbol{Q}\hat{\boldsymbol{M}}_{g_h}^k = \boldsymbol{K}\boldsymbol{R}^{-1}\hat{\boldsymbol{M}}_{g_h}^k = \boldsymbol{K}\boldsymbol{e}_{\boldsymbol{M_R}} = \boldsymbol{M}_{g_h}^k \tag{A.44}$$

Also, the following equality can be obtained:

$$\boldsymbol{M}_{g_h(V_N)}^k = \hat{\boldsymbol{M}}_{g_h(V_N)}^k \tag{A.45}$$

Therefore the $k$th moment of the system with respect to $g_h$ is conserved if the previous moments are conserved. Thus, by mathematical induction, we can conclude that the first $q_{g_h}$ moments are conserved.

Similarly, we proceed with the proof that the reduced system conserves $q_s$ moments with respect to frequency. Note that the zeroth moment with respect to conductance $\boldsymbol{M}_{g_h}^0 = \boldsymbol{M}^0$ is the same as the zeroth moment with respect to frequency $\boldsymbol{M}_s^0 = \boldsymbol{M}^0$ as defined in (A.1), the proof for the conservation of the zeroth moment with respect to frequency is equivalent to that of the zeroth moment with respect to conductance as given before.

Next we show that if the $(k-1)$th moment is conserved, then the $k$th moment is also conserved. In other words, if

$$\boldsymbol{M}_s^{(k-1)} = \boldsymbol{Q}\hat{\boldsymbol{M}}_s^{(k-1)} \tag{A.46}$$

$$\boldsymbol{M}_{s(V_N)}^{(k-1)} = \hat{\boldsymbol{M}}_{s(V_N)}^{(k-1)} \tag{A.47}$$

then

$$\boldsymbol{M}_s^k = \boldsymbol{Q}\hat{\boldsymbol{M}}_s^k \tag{A.48}$$

$$M^k_{s(V_N)} = \hat{M}^k_{s(V_N)} \tag{A.49}$$

The $k$th block moment ($k \neq 0$) with respect to frequency, $M^k_s$, at the expansion points $s = 0, g_1 = g^a_1, \cdots, g_N = g^a_N, c_1 = 0, \cdots, c_N = 0$ is given by (A.2) and has the form of:

$$\boldsymbol{G_\phi M^k_s + C M^{(k-1)}_s + D^a_n R_N L^T_N M^k_s + R_N R^T_N M^k_s = 0} \tag{A.50}$$

The above matrix is separated into two parts, the first $n - N$ equations and the last $N$ equations.

$$\boldsymbol{G_\phi M^k_s + C M^{(k-1)}_s = 0} \tag{A.51}$$

$$\boldsymbol{D^a_N M^k_{s(V_N)} + M^k_{s(I_N)} = 0} \tag{A.52}$$

where $\boldsymbol{M^k_{s(V_N)}}$ and $\boldsymbol{M^k_{s(I_N)}} \in \mathbb{R}^{N \times M}$ are the $k$th block moment of the constrained port voltages and constrained port currents for the original network. On the other hand, the $k$th block moment with respect to frequency for the reduced system can be calculated from (A.11) and (A.12). Substituting the congruence transformation in (3.26) into the resulting equations results in

$$\boldsymbol{Q^T G_\phi Q \hat{M}^k_s + Q^T C Q \hat{M}^{(k-1)}_s + Q^T R_N L^T_N Q \hat{M}^k_s = Q^T R_N \hat{M}^k_{s(V_N)}} \tag{A.53}$$

$$\boldsymbol{\hat{M}^k_{s(I_N)} = R^T_N Q \hat{M}^k_s = -D^a_N \hat{M}^k_{s(V_N)}} \tag{A.54}$$

where $\boldsymbol{\hat{M}^k_{s(V_N)}}$ and $\boldsymbol{\hat{M}^k_{s(I_N)}} \in \mathbb{R}^{N \times M}$ are the $k$th block moment of the constrained port voltages and currents for the reduced system. Inserting $\boldsymbol{K = QR}$ into (A.53) and using (A.10) result in

$$\boldsymbol{K^T G K R^{-1} \hat{M}^k_s + K^T C K R^{-1} \hat{M}^{(k-1)}_s - K^T R_N \hat{M}^k_{s(V_N)} = 0} \tag{A.55}$$

Substitute (A.51) in (A.55)

$$\boldsymbol{K}^T\boldsymbol{G}\boldsymbol{K}\boldsymbol{R}^{-1}\hat{\boldsymbol{M}}_s^k + \boldsymbol{K}^T\boldsymbol{C}\boldsymbol{K}\boldsymbol{R}^{-1}\hat{\boldsymbol{M}}_s^{(k-1)} - \boldsymbol{K}^T\boldsymbol{R}_N\hat{\boldsymbol{M}}_{s(V_N)}^k = \boldsymbol{K}^T\boldsymbol{G}_\phi\boldsymbol{M}_s^k + \boldsymbol{K}^T\boldsymbol{C}\boldsymbol{M}_s^{(k-1)} \tag{A.56}$$

note that the $(k-1)$th block moment is conserved in (A.46), it follows that

$$\boldsymbol{K}^T\boldsymbol{G}\boldsymbol{K}\boldsymbol{R}^{-1}\hat{\boldsymbol{M}}_s^k - \boldsymbol{K}^T\boldsymbol{R}_N\hat{\boldsymbol{M}}_{s(V_N)}^k = \boldsymbol{K}^T\boldsymbol{G}_\phi\boldsymbol{M}_s^k \tag{A.57}$$

Referring to (3.13), we have:

$$\boldsymbol{K}^T\boldsymbol{R}_N\boldsymbol{M}_{s(V_N)}^k = \boldsymbol{K}^T\boldsymbol{R}_N\boldsymbol{L}_N^T\boldsymbol{M}_s^k \tag{A.58}$$

Adding (A.57) and (A.58) and using (A.10), we obtain

$$\boldsymbol{K}^T\boldsymbol{G}\boldsymbol{K}\boldsymbol{R}^{-1}\hat{\boldsymbol{M}}_s^k + \boldsymbol{K}^T\boldsymbol{R}_N(\boldsymbol{M}_{s(V_N)}^k - \hat{\boldsymbol{M}}_{s(V_N)}^k) = \boldsymbol{K}^T\boldsymbol{G}\boldsymbol{M}_s^k \tag{A.59}$$

Also subtracting (A.54) from (A.52)

$$\boldsymbol{R}_N^T\boldsymbol{Q}\hat{\boldsymbol{M}}_s^k - \boldsymbol{D}_N^a(\boldsymbol{M}_{s(V_N)}^k - \hat{\boldsymbol{M}}_{s(V_N)}^k) = \boldsymbol{M}_{s(I_N)}^k \tag{A.60}$$

Now we substitute $\boldsymbol{K} = \boldsymbol{Q}\boldsymbol{R}$ in (A.60) and combine the resulting equation with (A.59) in matrix form to obtain

$$\begin{bmatrix} \boldsymbol{K}^T\boldsymbol{G}\boldsymbol{K} & \boldsymbol{K}^T\boldsymbol{R}_N \\ \boldsymbol{R}_N^T\boldsymbol{K} & -\boldsymbol{D}_N^a \end{bmatrix} \begin{bmatrix} \boldsymbol{R}^{-1}\hat{\boldsymbol{M}}_s^k \\ \boldsymbol{M}_{s(V_N)}^k - \hat{\boldsymbol{M}}_{s(V_N)}^k \end{bmatrix} = \begin{bmatrix} \boldsymbol{K}^T\boldsymbol{G}\boldsymbol{M}_s^k \\ \boldsymbol{M}_{s(I_N)}^k \end{bmatrix} \tag{A.61}$$

Again, since (A.61) is nonsingular and therefore has only one unique solution, which is the $r$th $M$ columns of the identity matrix $\boldsymbol{e}_{M_r}$, where $r$ corresponds to the location of $\boldsymbol{M}_s^k$ in $\boldsymbol{K}$. It follows that:

$$\boldsymbol{Q}\hat{\boldsymbol{M}}_s^k = \boldsymbol{K}\boldsymbol{R}^{-1}\hat{\boldsymbol{M}}_s^k = \boldsymbol{K}\boldsymbol{e}_{M_R} = \boldsymbol{M}_s^k \tag{A.62}$$

Also, the following equality can be obtained:

$$M^k_{s(V_N)} = \hat{M}^k_{s(V_N)} \tag{A.63}$$

# References


[1] H. B. Bakoglu, *Circuits, interconnects and packaging for VLSI.* Reading, MA: Addison-Wesley, 1990.

[2] C. R. Paul, *Analysis of Multiconductor Transmission Lines.* New York: Wiley, 1994.

[3] A. Deutsch, "Electrical charactersitics of interconnections for high performance systems," *Proceedings IEEE*, vol. 86, no. 2, pp. 315–355, Feb. 1998.

[4] M. A. B. Jackson and E. S. Kuh, "Estimating and optimizing rc interconnect delay during physical design," in *Proc. Intl. Symp. Circuits Syst.*, 1990, pp. 869–871.

[5] R. Achar and M. Nakhla, "Simulation of high-speed interconnects," *Proceedings IEEE*, vol. 89, no. 5, pp. 693–728, May 2001.

[6] E. Bogatin, *Signal Integrity -simplified.* Upper Saddle River, NJ: Prentice Hall, 2004.

[7] T. Dhane and D. Zutter, "Selection of lumped element models for coupled lossy transmission lines," *IEEE Trans. Computer-Aided Design*, vol. 11, no. 7, pp. 805–815, Jul. 1992.

[8] M. Celik, A. C. Cangellaris, and A. Yaghmour, "An all purpose transmission line model for interconnect simulation in SPICE," *IEEE Trans. Microwave Theory Tech.*, vol. 45, no. 10, pp. 1857–1867, Oct. 1997.

[9] A. Cangellaris, S. Pasha, J. L. Prince, and M. Celik, "A new discrete transmission line model for passive model order reduction and macromodeling of high-speed interconnections," *IEEE Trans. Adv. Packag.*, vol. 22, no. 3, pp. 356–364, Aug. 1999.

[10] W. T. Beyene and J. Schutt-Aine, "Accurate frequency-domain modeling and efficient circuit simulation of high-speed packaging interconnects," *IEEE Trans. Microwave Theory Tech.*, vol. 45, pp. 1941–1947, Oct. 1997.

[11] A. Dounavis, X. Li, M. Nakhla, and R. Achar, "Passive closed-loop transmission line model for general purpose circuit simulators," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 12, pp. 2450–2459, Dec. 1999.

[12] A. Dounavis, R. Achar, and M. Nakhla, "A general class of passive macromodels for lossy multiconductor transmission lines," *IEEE Trans. Microwave Theory Tech.*, vol. 49, no. 10, pp. 1686–1696, Oct. 2001.

[13] A. Dounavis, R. Achar, and M. S. Nakhla, "On passive time-domain macromodels of distributed transmission line networks," in *IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 2002, pp. 975–978.

[14] M. Nakhla and R. Achar, *The VLSI Handbook.* Boca Raton, FL: CRC Press, 2000, ch. Interconnect modeling and simulation, pp. 17.1–17.29.

[15] T. Tank and M. Nakhla, "Analysis of high-speed VLSI interconnects using asymptotic waveform evaluation," *IEEE Trans. Computer-Aided Design*, pp. 2107–2116, Mar. 1992.

[16] E. Chiprout and M. Nakhla, "Analysis of interconnect networks using complex frequency hopping (CFH)," *IEEE Trans. Computer-Aided Design*, vol. 14, pp. 186–200, Feb. 1995.

[17] ——, *Asymptotic Waveform Evaluation and Moment Matching for Interconnect Analysis.* Boston: Kluwer, 1993.

[18] L. T. Pillage and R. A. Rohrer, "Asymptotic waveform evaluation for timing analysis," *IEEE Trans. Computer-Aided Design*, vol. 9, pp. 352–366, Apr. 1990.

[19] A. Odabasioglu, M. Celik, and L. T. Pileggi, "PRIMA: passive reduced-order interconnect macromodeling algorithm," *IEEE Trans. Computer-Aided Design*, vol. 17, no. 8, pp. 645–654, Aug. 1998.

[20] P. Feldmann and R. W. Freund, "Efficient linear circuit analysis by Padé via Lanczos process," *IEEE Trans. Computer-Aided Design*, vol. 14, pp. 639–649, May 1995.

[21] J. E. Bracken, V. Raghavan, and R. A. Rohrer, "Interconnect simulation with asymptotic waveform evaluation," *IEEE Trans. Circuits Syst.*, vol. 39, pp. 869–878, Nov. 1990.

[22] Y. Liu, P. L.T., and S. A.J., "Model order-reduction of rc(l) interconnect including variational analysis," in *Proc. Design Automation Conf.*, New Orleans, LA, 1999, pp. 201–206.

[23] I. M. Elfadel and D. D. Ling, "A block rational arnoldi algorithm for multipoint passive model-order reduction of multiport RLC networks," in *Proc. ACM Int. Conference on Computer Aided Design*, San Jose, California, Nov. 1997, pp. 66–71.

[24] J. R. Phillips, L. Daniel, and L. M. Silveira, "Guaranteed passive balancing transformations for model order reduction," *IEEE Trans. Computer-Aided Design*, vol. 22, pp. 1027–1041, Aug. 2003.

[25] J. R. Phillips and L. M. Silveira, "Poor man's TBR: a simple model reduction scheme," *IEEE Trans. Computer-Aided Design*, vol. 24, no. 1, pp. 43–55, Jan. 2005.

[26] K. J. Kerns and A. T. Yang, "Preservation of passivity during RLC network reduction via split congruence transformations," *IEEE Trans. Computer-Aided Design*, vol. 17, pp. 582–591, Jul. 1998.

[27] A. Dounavis, E. Gad, and M. Nakhla, "Passive model-reduction of multiport distributed networks including frequency dependent parameters," *IEEE Trans. Microwave Theory Tech.*, pp. 2325–2334, Dec. 2000.

[28] Y. Saad, *Iterative Methods for Sparse Linear Systems*. Boston: PWD Publishing Company, 1996.

[29] R. Achar, P. Gunupudi, M. Nakhla, and E. Chiprout, "Passive interconnect reduction algorithm for distributed/measured networks," *IEEE Trans. Circuits Syst. II*, vol. 47, no. 4, pp. 287–301, Apr. 2000.

[30] M. Kamon, F. Wang, and J. White, "Generating nearly optimally compact models from Krylov subspace based reduced-order models," *IEEE Trans. Circuits Syst. II*, vol. 47, no. 4, pp. 239–248, Apr. 2000.

[31] P. Gunupudi, R. Khazaka, and M. S. Nakhla, "Analysis of transmission line circuits using multi-dimensional model reduction techniques," *IEEE Trans. on Packaging*, vol. 25, pp. 174–180, May 2002.

[32] L. Daniel, O. Siong, K. Lee, and J. White, "A multiparameter moment-matching model-reduction approach for generating geometrically parameterized interconnect performance models," *IEEE Trans. Computer-Aided Design*, vol. 23, pp. 678–693, May 2004.

[33] Y. Chen, "Model order reduction for nonlienar systems," Master's thesis, Massachusettes Institute of Technology, 1999.

[34] J. Roychowdhury, "Reduced-order modeling of time-varying systems," *IEEE Trans. Circuits Syst. II*, vol. 46, no. 10, pp. 1273–1288, Oct. 1999.

[35] J. Phillips, "Projection framework for model reduction of weakly nonlinear systems," in *Proc. Design Automation Conf.*, Los Angeles, CA, 2000, pp. 184–189.

[36] ——, "Automated extraction of nonlinear circuit macromodels," in *Proc. IEEE Custom Integrated Circuits Conf.*, Orlando, Florida, May 2000, pp. 451–454.

[37] P. Li and L. T. Pileggi, "Compact reduced-order modeling of weakly nonlinear analog and rf circuits," *IEEE Trans. Computer-Aided Design*, vol. 23, no. 2, pp. 184–203, Feb. 2005.

[38] M. Rewienski and J. White, "A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices," *IEEE Trans. Computer-Aided Design*, vol. 22, no. 2, pp. 155–203, Feb. 2003.

[39] N. Dong and J. Roychowdhury, "Automated nonlinear macromodelling of output buffers for high-speed digital applications," in *Proc. Design Automation Conf.*, Anaheim, CA, 2005, pp. 51–56.

[40] M. Ma and R. Khazaka, "Passive model order reduction of interconnect networks with large number of ports," in *IEEE MTT-S Int. Microwave Symp. Dig.*, Long Beach, CA, Jun. 2005, pp. 1779–1782.

[41] ——, "Passive SVD based model order reduction with parametric port formulation," in *Proc. IEEE workshop on Signal Propagation on Interconnects SPI*, Berlin, Germany, May 2006, pp. 101–104.

[42] ——, "Model order reduction with load constraints," in *Proc. IEEE conf. on Electrical Performance of Electronic Packaging*, Scottsdale, AZ, Oct. 2006, pp. 193–196.

[43] ——, "Multi-level reduction with nonlinear port constraints," in *Proc. Intl. Symp. Circuits Syst.*, New Orleans, LA, May 2007.

[44] ——, "Model order reduction with parametric port formulation," *IEEE Trans. Adv. Packag.*, accepted with minor revision for publication as a regular paper.

[45] ——, "Efficient projection based macromodel for interconnect networks," in *Proc. IEEE workshop on Signal Propagation on Interconnects SPI*, Garmisch, Germany, May 2005, pp. 181–184.

[46] B. Haddadin, M. Ma, T. S. Roseanu, and R. Khazaka, "Efficient macromodel for interconnects excited by incident fields," in *Proc. Canadian Conference on Computer and Electrical Engineering*, Ottawa, Canada, May 2006, pp. 90–93.

[47] T. S. Roseanu, M. Ma, and R. Khazaka, "Time domain reduced order macromodel for interconnects excited by incident fields," in *Proc. IEEE workshop on Signal Propagation on Interconnects SPI*, Garmisch, Germany, May 2005, pp. 197–200.

[48] M. Ma, A. Leung, and R. Khazaka, "Multi-level parametric model order reduction for interconnect networks," in *Proc. IEEE conf. on Electrical Performance of Electronic Packaging*, Austin, TX, Oct. 2005, pp. 235–238.

[49] ——, "Sparse macromodels for parametric networks," in *Proc. Intl. Symp. Circuits Syst.*, kos, Greece, Jun. 2006.

[50] M. Ma and R. Khazaka, "Nonlinear macromodeling using model order reduction," in *Proc. IEEE conf. on Electrical Performance of Electronic Packaging*, Austin, TX, Oct. 2005, pp. 131–134.

[51] ——, "Sparse macromodeling for nonlinear network," in *IEEE MTT-S Int. Microwave Symp. Dig.*, San Francisco, CA, Jun. 2006.

[52] ——, "Sparse macromodeling for parametric nonlinear network," *IEEE Trans. Microwave Theory Tech.*, vol. 54, no. 12, pp. 4305–4315, Dec. 2006.

[53] R. W. Freund and P. Feldmann, "Small-signal circuit analysis and sensitivity computations with the PVL algorithm," *IEEE Trans. Circuits Syst.*, vol. 43, no. 8, pp. 577–585, Aug. 1996.

[54] P. Gunupudi, M. Nakhla, and R. Achar, "Simulation of high-speed distributed interconnects using krylov-space techniques," *IEEE Trans. Computer-Aided Design*, vol. 19, no. 7, pp. 799–808, Jul. 2000.

[55] Q. Yu, J. M. L. Wang, and E. S. Kuh, "Passive multipoint moment matching model order reduction algorithm on multiport distributed interconnect networks," *IEEE Trans. Circuits Syst.*, vol. 46, no. 1, pp. 140–160, Jan. 1999.

[56] R. W. Freund, "Krylov-subspace methods for reduced-order modeling in circuit simulation," *Journal of Computational and Applied Mathematics*, vol. 123, pp. 395–421, 2000.

[57] E. J. Grimme, "Krylov projection methods for model reduction," Ph.D. dissertation, University of Illinos at Urbana-Champaign, 1997.

[58] R. Khazaka and M. S. Nakhla, "Analysis of high-speed interconnects in the presence of electromagnetic interference," *IEEE Trans. Microwave Theory Tech.*, vol. 46, pp. 940–947, Jul. 1998.

[59] K. J. Kerns and A. T. Yang, "Stable and efficient reduction of large multiport RC networks by pole analysis via congruence transformations," in *Proc. Design Automation Conf.*, 1996, pp. 280–285.

[60] A. C. Cangellaris and M. Igarashi, "Rules for robust generation of accurate reduced-order models for high-speed coupled interconnections," *IEEE Trans. Adv. Packag.*, vol. 24, no. 2, pp. 120–125, May 2001.

[61] B. Moore, "Principal component analysis in linear system: Controllability, observability and model reduction," *IEEE Trans. Automat. Contr.*, no. 1, pp. 17–32, Feb. 1981.

[62] M. Rabiei, P.; Pedram, "Model order reduction of large circuits using balanced truncation," in *Proc. Asia and South Pacific Design Automation Conference*, Hong Kong, Jan. 1999, pp. 237–240.

[63] C. W. Ho, A. E. Ruehli, and P. A. Brennan, "The modified nodal approach to network analysis," *IEEE Trans. Circuits Syst.*, vol. 22, pp. 504–509, Jun. 1975.

[64] P. Gunupudi and M. S. Nakhla, "Nonlinear circuit-reduction of high-speed interconnect networks using congruent transformation techniques," *IEEE Trans. on Packaging*, vol. 24, pp. 317–325, 2001.

[65] J. Vlach and K. Singhal, *Computer Methods for Circuit Analysis and Design.* New York: Van Norstrand Reinolds, 1983.

[66] I. Erdin, R. Khazaka, and M. S. Nakhla, "Simulation of high-speed interconnects in a multilayered medium in the presence of incident field," *IEEE Trans. Microwave Theory Tech.*, vol. 46, no. 12, pp. 2251–2257, Dec. 1998.

[67] R. Khazaka, J. Poltz, M. S. Nakhla, and Q. Zhang, "A fast method for the simulation of lossy interconnects with frequency dependent parameters," in *Proc. IEEE Multi-Chip Module conference*, Feb. 1996, pp. 95–98.

[68] J. Y. Lee, X. Huang, and R. A. Roher, "Pole and zero sensitivity calculation in asymptotic waveform evaluation," *IEEE Trans. Computer-Aided Design*, vol. 11, pp. 586–597, May 1992.

[69] M. S. Nakhla, E. Chiprout, R. Achar, and R. Khazaka, "Recent progress in the simulation of high speed VLSI interconnects," in *Proc. IEEE conf. on Electrical Performance of Electronic Packaging*, Oct. 1995, pp. 193–196.

[70] R. Achar, R. Khazaka, R. Griffith, M. S. Nakhla, and Q. J. Zhang, "Simulation of delay and crosstalk in high speed vlsi interconnects," in *Proc. Canadian Conference on Computer and Electrical Engineering*, Sep. 1995, pp. 385–388.

[71] R. Khazaka and M. S. Nakhla, "Analysis of high-speed interconnects in the presence of electromagnetic interference," in *IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 1999, pp. 1811–1814.

[72] R. Sanaie, E. Chiprout, M. Nakhla, and Q. J. Zhang, "A fast method for frequency and time domain simulation of high speed VLSI interconnects," *IEEE Trans. Microwave Theory Tech.*, vol. 42, pp. 2562–2571, Dec. 1994.

[73] R. Achar, M. Nakhla, and E. Chiprout, "Block CFH: A model-reduction technique for distributed interconnect networks," *IEEE Trans. Microwave Theory Tech.*, vol. 42, pp. 2562–2571, Dec. 1997.

[74] J. W. Demmel, *Applied Numerical Linear Algebra.* Philadelphia, PA: SIAM Publishers, 1997.

[75] A. Odabasioglu, M. Celik, and L. Pileggi, "Practical considerations for passive reduction of RLC circuits," in *Proc. ACM Int. Conference on Computer Aided Design*, San Jose, California, Nov. 1999, pp. 214–219.

[76] J. Phillips, "Projection-based approaches for model reduction of weakly nonlinear, time varying systems," *IEEE Trans. Computer-Aided Design*, vol. 22, no. 2, pp. 171–187, Feb. 2003.

[77] P. Gunupudi, R. Khazaka, M. Nakhla, T. Smy, and D. Celo, "Passive parametrized time-domain macromodels for high-speed transmission-line networks," *IEEE Trans. Microwave Theory Tech.*, vol. 51, pp. 2347–2354, Dec. 2003.

[78] M. P. C. W. R. Laub, A.J.; Heath, "Computation of system balancing transformations and other applications of simultaneous diagonalization algorithms," *IEEE Trans. Automat. Contr.*, no. 2, pp. 115–122, Feb. 1987.

[79] W. J. Rugh, *Nonlinear System Theory.* Baltimore: Johns Hopkins University Press, 1981.

[80] M. Rewienski and J. White, "A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices," in *Proc. ACM Int. Conference on Computer Aided Design*, San Jose, CA, Nov. 2001, pp. 252–257.

[81] M. J. Rewienski, "A trajectory piecewise-linear approach to model order reduction of nonlinear dynamical systems," Ph.D. dissertation, Massachusettes Institute of Technology, 2003.

[82] P. Gunupudi and M. S. Nakhla, "Mutli-dimensional model reduction of vlsi interconnects," in *Proc. IEEE Custom Integrated Circuits Conf.*, Orlando, Florida, May 2000, pp. 499–502.

[83] A. Leung and R. Khazaka, "Parametric model order reduction technique for design optimization," in *Proc. Intl. Symp. Circuits Syst.*, May 2005, pp. 1290–1293.

[84] R. Khazaka and M. Nakhla, "Multi-level passive order reduction of interconnect networks," in *IEEE MTT-S Int. Microwave Symp. Dig.*, May 2001, pp. 1155–1158.

[85] G. H. Golub and C. Van Loan, *Matrix Computations.* Baltimore, MD: John Hopkins Univ. Press, 1989.

[86] P. Gunupudi, R. Khazaka, and M. S. Nakhla, "Analysis of transmission line circuits using multi-dimensional model reduction techniques," in *Proc. IEEE conf. on Electrical Performance of Electronic Packaging*, Oct. 2001, pp. 43–46.

[87] E. S. Kuh and R. A. Rohrer, *Theory of Linear Active Networks.* San Francisco: Holden-Day, 1967.

[88] A. C. Cangellaris and A. E. Ruehli, "Model order reduction techniques applied to electromagnetic problems," in *Proc. IEEE conf. on Electrical Performance of Electronic Packaging*, Oct. 2000, pp. 239–242.