in Proceedings of the 32^{st} Design Automation Conference, 12–15 June 1995, San Francisco, California, USA

Efficient Reduced-Order Modeling of Frequency-Dependent Coupling Inductances associated with 3-D Interconnect Structures

L. Miguel Silveira   Mattan Kamon   Jacob White
Research Laboratory of Electronics
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract

Since the first papers on asymptotic waveform evaluation (AWE), reduced order models have become standard for improving interconnect simulation efficiency, and very recent work has demonstrated that bi-orthogonalization algorithms can be used to robustly generate AWE-style macromodels. In this paper we describe using block Arnoldi-based orthogonalization methods to generate reduced order models from FASTHENRY, a multipole-accelerated three dimensional inductance extraction program. Examples are analyzed to demonstrate the efficiency and accuracy of the block Arnoldi algorithm.

1 Introduction

The dense three-dimensional packaging used in compact electronic systems may produce magnetic interactions which interfere with system performance. Such effects are difficult to simulate because they occur only as a result of an interaction between the field distribution in a complicated geometry of conductors, and the circuitry connected to those conductors. Recent work on reduced-order modeling techniques have made it possible to efficiently simulate circuits combined with interconnect [1], but generating the reduced-order models from realistic 3-D structures has received less attention. Reduced-order models can be generated very efficiently by exploiting the fact that 3-D field solvers typically use Krylov-subspace based iterative methods. These iterative methods can provide more than just a solution at a particular frequency; they can be used to directly construct reduced-order models [2].

In this paper, we present a numerically robust and accurate approach for computing reduced-order models of magnetoquasistatic coupling in complicated 3-D structures. The approach is based on using the multipole-accelerated program FASTHENRY [3], combined with the block Krylov-subspace algorithm Arnoldi [4]. We begin, in section 2, by describing the mesh-formulation approach of FASTHENRY. In section 3, the standard Padé approximation approach as well as an Arnoldi-based approach are derived. In section 4 results are presented comparing the accuracy of the two model-order reduction methods on a package example. Finally, in section 5, we present conclusions and acknowledgments.

2 The Mesh Formulation Approach

The frequency dependent resistance and inductance matrices describing the terminal behavior of a set of conductors can be rapidly computed with the multipole-accelerated mesh-formulation approach as implemented in FASTHENRY [3]. To describe the approach, consider that each conductor is approximated as piecewise-straight sections. The volume of each straight section is then discretized into a collection of parallel thin filaments through which current is assumed to flow uniformly.

To derive a system of equations for the filament currents, we start by assuming the system is in sinusoidal steady-state and follow the partial inductance approach in [5]. The branch current phasors can be related to branch voltage phasors by

$$\boldsymbol{V_b} = (\boldsymbol{R} + \boldsymbol{j\omega L})\boldsymbol{I_b} = \boldsymbol{Z I_b}, \tag{1}$$

where $\boldsymbol{V_b}$, $\boldsymbol{I_b} \in \mathbb{C}^b$, b is the number of branches (number of current filaments), and ω is excitation frequency. The entries of the diagonal matrix $\boldsymbol{R} \in \mathbb{R}^{b\times b}$ represent the DC resistance of each current filament, and $\boldsymbol{L} \in \mathbb{R}^{b\times b}$ is the dense matrix of partial inductances.

Kirchhoff's voltage law, which implies that the sum of branch voltages around each mesh in the network is represented by

$$\boldsymbol{M V_b} = \boldsymbol{V_s} \qquad \boldsymbol{M}^T \boldsymbol{I_m} = \boldsymbol{I_b}, \tag{2}$$

where $\boldsymbol{V_s} \in \mathbb{C}^m$ is the mostly zero vector of source branch voltages, $\boldsymbol{I_m} \in \mathbb{C}^m$ is the vector of mesh currents, $\boldsymbol{M} \in \mathbb{R}^{m\times b}$ is the mesh matrix. Combining (2) and (1) yields

$$\boldsymbol{M Z M}^T \boldsymbol{I_m} = \boldsymbol{V_s}. \tag{3}$$

The complex admittance matrix which describes the external terminal behavior of a t-conductor system, denoted $\boldsymbol{Y_t} = \boldsymbol{Z_t}^{-1}$, can by derived from (3) by noting that

$$\boldsymbol{I_t} = \boldsymbol{Y_t V_t}$$

where $\boldsymbol{I_t}$ and $\boldsymbol{V_t}$ are the terminal source currents and voltages of the t-conductor system. These values are related to the mesh quantities by $\boldsymbol{I_t} = \boldsymbol{N}^T \boldsymbol{I_m}, \boldsymbol{V_s} = \boldsymbol{N V_t}$, where $\boldsymbol{N} \in \mathbb{R}^{m\times t}$ is an easily constructed terminal incidence matrix.

Hence, to compute the i^{th} column of $\boldsymbol{Y_t}$, set $\boldsymbol{V_{t_i}} = 1$ and the rest of $\boldsymbol{V_{t_i}}$ to zero. Solve (3) with a $\boldsymbol{V_s} = \boldsymbol{N V_t}$ and then extract the entries of $\boldsymbol{I_m}$ associated with the source branches via $\boldsymbol{I_t} = \boldsymbol{N}^T \boldsymbol{I_m}$.

To solve (3) by Gaussian Elimination would require $\mathcal{O}(m^3)$ operations since $\boldsymbol{Z}$ is dense. Instead, programs like FASTHENRY solve (3) using a multipole-accelerated GMRES iterative algorithm [4, 6], which requires $\mathcal{O}(b)$ operations.

3 Reduced-Order Modeling

Consider forming the state-space representation of (3). Expanding $\boldsymbol{Z}$ into $\boldsymbol{R} + s\boldsymbol{L}$ gives

$$s(\boldsymbol{M L M}^T)\boldsymbol{I_m} = -(\boldsymbol{M R M}^T)\boldsymbol{I_m} + \boldsymbol{V_s}. \tag{4}$$

From (4) we can obtain

$$\begin{aligned} \tfrac{d}{dt}\boldsymbol{I_m} &= \boldsymbol{A}^{-1}\boldsymbol{I_m} + \boldsymbol{A}^{-1}\boldsymbol{B V_t} \\ \boldsymbol{I_t} &= \boldsymbol{C}^T \boldsymbol{I_m}. \end{aligned} \tag{5}$$

or equivalently the matrix transfer function

$$\frac{\boldsymbol{I_t}}{\boldsymbol{V_t}} = \boldsymbol{Y_t}(s) = \boldsymbol{C}^T (\boldsymbol{I} - s\boldsymbol{A})^{-1} \boldsymbol{B} \tag{6}$$

where $\boldsymbol{A} = -(\boldsymbol{M R M}^T)^{-1}(\boldsymbol{M L M}^T)$, $\boldsymbol{B} = (\boldsymbol{M R M}^T)^{-1}\boldsymbol{N}$, $\boldsymbol{I}$ is the identity matrix in $\mathbb{R}^{m\times m}$ and $\boldsymbol{C} = \boldsymbol{N}$.

It is possible to use (5) directly in a circuit simulator as a model for the interconnect, but such an approach can be computationally expensive. For example, in a complicated package, the dense matrix $\boldsymbol{A}$ can easily be larger than $10,000 \times 10,000$.

3.1 Order Reduction using Padé Approximations

A more efficient approach to including a system described by (5) in a circuit simulator is to compute a reduced-order model. Consider a given entry, $\boldsymbol{Y_{ij}}$, of the admittance matrix, $\boldsymbol{Y_t}(s)$, given by

$$\frac{\boldsymbol{I_{t_i}}}{\boldsymbol{V_{t_j}}} = \boldsymbol{Y_{ij}}(s) = \boldsymbol{c}^T (\boldsymbol{I} - s\boldsymbol{A})^{-1} \boldsymbol{b} \tag{7}$$

where $\boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^m$ are appropriately chosen columns of the original matrices $\boldsymbol{B}$ and $\boldsymbol{C}$ respectively. Expanding (7) into a MacLaurin series we get

$$\boldsymbol{Y_{ij}}(s) = \boldsymbol{c}^T (\boldsymbol{I} - s\boldsymbol{A})^{-1} \boldsymbol{b} = \sum_{k=0}^{\infty} m_k s^k. \tag{8}$$

where

$$m_k = \boldsymbol{c}^T \boldsymbol{A}^k \boldsymbol{b} \tag{9}$$

is known as the k^{th} moment of the transfer function. A diagonal Padé approximation of q^{th} order for $\boldsymbol{Y_{ij}}$ is then defined as the rational function

$$\boldsymbol{G}_q^P(s) = \frac{u_{q-1}s^{q-1} + \cdots + u_1 s + u_0}{w_q s^q + w_{q-1}s^{q-1} + \cdots + w_1 s + 1} \tag{10}$$

in the coefficients u_i and w_i such that [7]

$$\lim_{s\to 0} \frac{1}{k!}\frac{d^k}{ds^k}\boldsymbol{G}_q^P(s) = m_k. \tag{11}$$

3.2 Order Reduction using Arnoldi Iterations

As pointed out in [8] and applied in [2, 8, 9], a numerically robust approach for computing $\boldsymbol{G}_q^P(s)$ of very high order is to use a Lanczos-based algorithm.

An alternative approach, which robustly generates a somewhat different approximation, can be derived using an Arnoldi process as in the GMRES algorithm used in FASTHENRY. The idea behind this approach is similar to that of [4], and is that of selecting an orthonormal basis for the Krylov subspace $\mathcal{K}_k(\boldsymbol{A}, \boldsymbol{b}) = span\{\boldsymbol{b}, \boldsymbol{A b}, \boldsymbol{A}^2\boldsymbol{b}, \cdots, \boldsymbol{A}^{k-1}\boldsymbol{b}\}$. The Arnoldi algorithm is a better conditioned process than direct evaluation of the moments because it generates an orthogonal set of vectors which span $\boldsymbol{A}^k\boldsymbol{b}, k = 0, \ldots, 2q-1$.

After q steps, the Arnoldi algorithm returns a set of $q+1$ orthonormal vectors, as the columns of the matrix $\boldsymbol{V_q} \in \mathbb{R}^{m\times q}$ and the vector $\boldsymbol{v_{q+1}}$ where the first column of $\boldsymbol{V_q}$ is $\boldsymbol{v_1} = \boldsymbol{b}/||\boldsymbol{b}||_2$. As a result of the orthogonalization process, the Arnoldi process also produces the $q \times q$ upper Hessenberg matrix $\boldsymbol{H}_q$ and scalar $h_{q+1,q}$ which satisfy

$$\boldsymbol{A}\boldsymbol{V}_q = \boldsymbol{V}_q\boldsymbol{H}_q + h_{q+1,q}\boldsymbol{v}_{q+1}\boldsymbol{e}_q^T \tag{12}$$

where $\boldsymbol{e}_q$ is the q^{th} unit vector in $\mathbb{R}^{m\times m}$.

From (12) and $\boldsymbol{v}_1 = \boldsymbol{b}/||\boldsymbol{b}||_2$, it can easily be seen that after q steps of an Arnoldi process, for $k < q-1$,

$$\boldsymbol{A}^k\boldsymbol{b} = ||\boldsymbol{b}||_2\boldsymbol{A}^k\boldsymbol{V}_q\boldsymbol{e}_1 = ||\boldsymbol{b}||_2\boldsymbol{V}_q\boldsymbol{H}_q^k\boldsymbol{e}_1. \tag{13}$$

With this relation, the moments (9) can be related to $\boldsymbol{H}_q$ by

$$m_k = \boldsymbol{c}^T\boldsymbol{A}^k\boldsymbol{b} = ||\boldsymbol{b}||_2\boldsymbol{c}^T\boldsymbol{V}_q\boldsymbol{H}_q^k\boldsymbol{e}_1 \tag{14}$$

and so the q^{th} order Arnoldi-based approximation to $\boldsymbol{Y}_{ij}$ can be written as

$$\boldsymbol{G}_q^A(s) = ||\boldsymbol{b}||_2\boldsymbol{c}^T\boldsymbol{V}_q\left(\boldsymbol{I} - s\boldsymbol{H}_q\right)^{-1}\boldsymbol{e}_1 \tag{15}$$

corresponding to the state-space realization using the triplet $[\boldsymbol{A}_k, \boldsymbol{b}_k, \boldsymbol{c}_k] = [\boldsymbol{H}_q, \boldsymbol{e}_1, ||\boldsymbol{b}||_2\boldsymbol{V}_q^T\boldsymbol{c}]$.

Note that the rational function $\boldsymbol{G}_q^A(s)$ is *not* a Padé approximation. Equation (13) is only valid for $k < q-1$, thus $\boldsymbol{G}_q^A(s)$ has q poles but only matches $q-2$. Conversely, $\boldsymbol{G}_q^P(s)$ matches $2q-1$ moments.

3.3 Systems with Multiple Inputs and Outputs

Consider the case of a multiple input, multiple output system with r inputs and p outputs (MIMO system). Obtaining a reduced-order model of order q for such a system using Padé-based techniques requires at best $\mathcal{O}(2rq)$ computations where it is assumed $r \le p$ without loss of generality. On the other hand, using the technique described in section 3.2 requires running the Arnoldi algorithm r times, once for each column of $\boldsymbol{B}$. The total cost of computing the matrix transfer function using this algorithm is thus $\mathcal{O}(rq)$ for a q^{th} order approximation.

An alternative approach to that described in section 3.3 for computing a reduced-order model of (6) is to use a block algorithm. Block versions of the Arnoldi algorithm exist for handling multiple right-hand side vectors.

After $\hat{q}$ steps, the block Arnoldi algorithm returns a set of $\hat{q}$ orthonormal blocks as the block columns of the matrix $\boldsymbol{V}_{\hat{q}}^b \in \mathbb{R}^{m\times r\hat{q}}$, and a $r\hat{q} \times r\hat{q}$ upper band Hessenberg matrix $\boldsymbol{H}_{\hat{q}}^b$ whose entries are $r \times r$ blocks $\boldsymbol{H}_{i,j}$. These two matrices satisfy a relationship similar to that in (12), namely

$$\boldsymbol{A}\boldsymbol{V}_{\hat{q}}^b = \boldsymbol{V}_{\hat{q}}^b\boldsymbol{H}_{\hat{q}}^b + \boldsymbol{V}_{\hat{q}+1}\boldsymbol{H}_{q+1,q}\boldsymbol{E}_{\hat{q}}^T \tag{16}$$

Method	# mv prods	# nonzeros in $\boldsymbol{A}_k$
Padé	$2pq$	$\approx\ 3p^2q$
Arnoldi	pq	$\approx\ \frac{3}{2}pq^2$
Block Arnoldi	q	$\approx\ \frac{1}{2}q^2$

Table 1: Costs for a q^{th} order approximation of a p-input, p-output system (number of matrix vector products and number of nonzeros in the system matrix).

where $\boldsymbol{E}_{\hat{q}}$ is the matrix of the last r columns of the identity in $\mathbb{R}^{m\times m}$.

From (16), the $\hat{q}^{th}$ order block Arnoldi-based approximation to $\boldsymbol{Y}_{ij}$ can be written as

$$\boldsymbol{G}_{\hat{q}}^{bA}(s) = \boldsymbol{R}_1\boldsymbol{C}^T\boldsymbol{V}_{\hat{q}}^b\left(\boldsymbol{I} - s\boldsymbol{H}_{\hat{q}}^b\right)^{-1}\boldsymbol{E}_1 \tag{17}$$

corresponding to the state-space realization using the triplet $[\boldsymbol{A}_k, \boldsymbol{B}_k, \boldsymbol{C}_k] = [\boldsymbol{H}_{\hat{q}}^b, \boldsymbol{E}_1, \boldsymbol{R}_1\boldsymbol{V}_{\hat{q}}^{b^T}\boldsymbol{C}]$ where $\boldsymbol{R}_1$ results from the QR factorization of $\boldsymbol{B}$.

The total cost of computing the matrix transfer function using the block Arnoldi algorithm is $\mathcal{O}(r\hat{q})$ resulting in an approximation of order $r\hat{q}$.

3.4 Complexity Comparisons

In all of these methods, the computation of $\boldsymbol{b}$ is inexpensive since $\boldsymbol{M}\boldsymbol{R}\boldsymbol{M}^T$ is sparse. However, because L is dense, the dominant cost of each iteration or moment computation is a matrix-vector product, $\boldsymbol{A}\boldsymbol{x} = -(\boldsymbol{M}\boldsymbol{R}\boldsymbol{M}^T)^{-1}(\boldsymbol{M}\boldsymbol{L}\boldsymbol{M}^T)\boldsymbol{x}$. In practice, the matrix-vector cost dominates even when the dense part, $(\boldsymbol{M}\boldsymbol{L}\boldsymbol{M}^T)\boldsymbol{x}$, is rapidly computed with a hierarchical multipole-algorithm as in FastHenry.

Table 1 compares the number of matrix-vector products and also the number of nonzeros in the reduced-order system matrix for each method where, for simplicity, we assume $r = p$. Based on the number of matrix-vector products, block Arnoldi is the most efficient at forming the reduced-order model. Also, since block Arnoldi has the fewest nonzeros, its system matrix would be the most efficient during circuit simulation.

The authors would like to note that the researchers in [2] have developed an a block Lanczos algorithm which we understand will produce an $\boldsymbol{A}$ matrix with fewer nonzero entries.

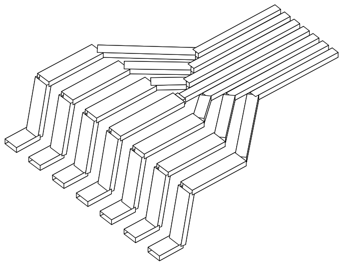

Figure 1: Seven pins of a cerquad pin package.

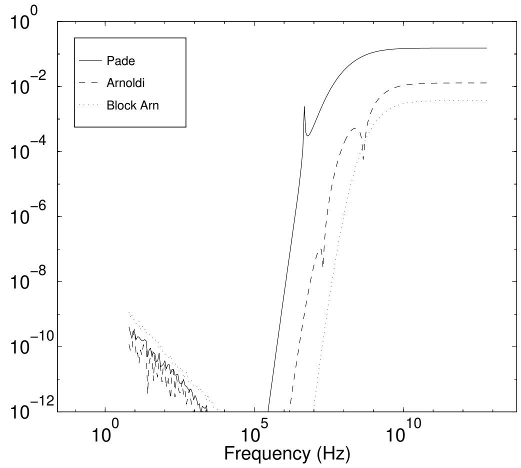

Figure 2: Relative error for $G_4^P(s)$, $G_8^A(s)$ and $G_{56}^{bA}(s)$. Each method required 56 matrix-vector products.

4 Experimental Results

Consider the small set of package pins. as shown in Figure 1. To compute the resistance and inductance matrices with FASTHENRY, the pins were discretized into three filaments along their height and four along their width producing a system of size $m = 887$. This discretization allows the modeling of changes in resistance and inductance due to skin and proximity effects. Since there are seven pins, the model to be produced has only seven inputs and seven outputs.

For the admittance between pins 1 and 2, Figure 2 shows the relative error for Padé, Arnoldi, and block Arnoldi admittance modes with model order 4, 8, and 56, respectively. The model order was chosen such that the computation would require 56 matrix-vector products for each method. It is worth noting that Padé, Arnoldi, and block Arnoldi match 7, 6, and 6 moments, respectively, yet block Arnoldi clearly gives the best approximation.

Table 2 compares the computational cost and the complexity of the reduced order models for a desired fixed accuracy of 5% pointwise error. The table shows that block Arnoldi requires the fewest matrix-vector products, and thus requires the least computation. The table also shows that with only 42 nonzeros in the reduced-order system matrix, it is also the most efficient for subsequent simulation using its model.

Method	Order	# mv prods	# nonz in A_k
Padé	8	112	1764
Arnoldi	8	56	252
Block Arn.	14	14	42

Table 2: Order of approximation, number of matrix vector products and number of nonzeros in the reduced-order system matrix for approximation yielding an accuracy of 5%.

4.1 Coupled Simulation Results

Consider the crosstalk between the pins in Fig. 1. Assume the five middle lines carry output signals from the chip and the two outer pins carry power and ground. The signals are driven and received with CMOS inverters. The drivers are capable of driving a large current to compensate for the impedance of the package pins. The inductance of the pins is modeled as described in the previous sections, and the capacitance is assumed to be 8pF. The interconnect from the end of pin to the receiver is modeled with a capacitance of 5pF. A $0.1\mu F$ decoupling capacitor is connected between the driver's power and ground to minimize supply fluctuations.

The reduced-order model for each entry in the admittance matrix is incorporated into SPICE3 as a frequency-dependent voltage-controlled current source VCCS. As a sample time domain simulation, imagine that at time $t_0 = 4$ns the signal on the middle pin of Fig.1 is to switch from high to low and the other four signal pins switch from low to high but due to delay on chip, the other four pins switch at $t_1 = 5$ns. In this case, significant current will suddenly pass through the late pins while the middle pin is in transition. Due to crosstalk, this large transient of current has significant effects on the input of the receiver of the middle pin, as shown in Fig. 3. Note that the input does not rise monotonically. Fig. 3 also shows that the bump in the waveform is carried through to the output of receiver, as a large glitch.

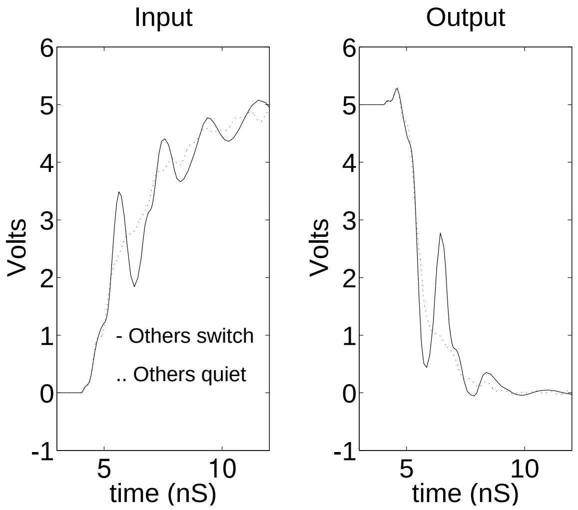

Figure 3: The middle pin's receiver when four adjacent pins switch 1ns after the middle pin.

5 Conclusions

In this paper we described an accurate approach using block Arnoldi based algorithms to compute reduced-order models of linear systems, such as the frequency-dependent inductance matrices associated with complicated 3-D structures. One key advantage of this method is that it is no more expensive than computing the inductance matrix at a single frequency. We also compared three approaches to model order reduction, the Padé-based approach, an Arnoldi-based approach, and a block Arnoldi method. We showed that both the Arnoldi and block Arnoldi algorithms can have advantages over Padé in certain applications.

We showed that although the original Padé algorithm requires order $2pq$ work to compute a q^{th} order reduced model for a p-terminal package, block Arnoldi algorithms can reduce the work to produce a reduced order model to order q. Examples are analyzed to demonstrate the efficiency and accuracy of the Arnoldi-based algorithms.

Finally it should be noted that an extension of the Lanczos algorithm has been developed in [10] for computing multi-point Padé approximations.

Acknowledgments

The authors wish to acknowledge the very helpful discussions with Dr. Peter Feldmann and Dr. Roland Freund of the A.T. & T research center.

This work was supported by the Defense Advanced Research Projects Agency under Contract N00014-91-J-1698, the Semiconductor Research Corporation under Contract SJ-558, the National Science Foundation contract (9117724-MIP), an NSF Graduate Research Fellowship, and grants from IBM and DEC.

References

[1] Lawrence T. Pillage and Ronald A. Rohrer. Asymptotic Waveform Evaluation for Timing Analysis. *IEEE Trans. CAD*, 9(4):352–366, April 1990.

[2] Peter Feldmann and Roland W. Freund. Efficient linear circuit analysis by Padé approximation via the Lanczos process. In *Proceeding of the Euro-DAC*, September 1994.

[3] M. Kamon, M. J. Tsuk, and J. White. Fasthenry: A multipole-accelerated 3-d inductance extraction program. *IEEE Transactions on Microwave Theory and Techniques*, 42(9):1750–1758, September 1994.

[4] Y. Saad and M. H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing*, 7:856–869, July 1986.

[5] A. E. Ruehli. Inductance calculations in a complex integrated circuit environment. *IBM J. Res. Develop.*, 16:470–481, September 1972.

[6] L. Greengard. *The Rapid Evaluation of Potential Fields in Particle Systems.* M.I.T. Press, Cambridge, Massachusetts, 1988.

[7] G. A. Baker Jr. The theory and application of the Padé approximant methods. In K. A. Brueckner, editor, *Advances in Theoretical Physics, vol. 1*, pages 1–58. Academic Press, New York, 1965.

[8] W. B. Gragg and A. Lindquist. On the partial realization problem. *Linear Algebra and its Applications*, 50:277–319, 1983.

[9] K. Gallivan, E. Grimme, and P. Van Dooren. Asymptotic Waveform Evaluation via a Lanczos Method. *Applied Mathematics Letters*, 7(5):75–80, 1994.

[10] K. Gallivan, E. Grimme, and P. Van Dooren. Multi-point padé approximants of large-scale systems via a two-sided rational krylov algorithm. In 33^{rd} *IEEE Conference on Decision and Control*, December 1994.