Krylov-Subspace-Based Order Reduction Methods Applied to Generate Compact-Electro-Thermal Models for MEMS

T. Bechtold*, B. Salimbahrami**, E. B. Rudnyi*, B. Lohmann** and J. G. Korvink*

*IMTEK, University of Freiburg, 79110 Freiburg, Germany
Tel:+49-761-2037389, Fax:+49-761-2037382,bechtold@imtek.de
**IAT, University of Bremen, 28359 Bremen, Germany
Tel:+49-421-218-3906, Fax:+49-421-218-4707,salimbahrami@iat.de

ABSTRACT

The high power dissipation density in today's miniature electronic/mechanical systems makes on-chip thermal management crucial. In order to achieve quick-to-evaluate, yet accurate electro-thermal models, needed for the thermal management of microsystems, model order reduction is necessary. In this paper, we use Krylov-subspace methods for the order reduction of a electro-thermal MEMS model, illustrated by a novel type of micropropulsion device.
Comparison between different moment-matching algorithms including a new two-sided Arnoldi algorithm, is performed.

Keywords: compact thermo-electric model, order reduction, Krylov-subspace, Arnoldi process, Lanczos algorithm.

1 INTRODUCTION

Modeling of thermo-electric processes becomes increasingly important for a variety of applications, including power transistors, thick-film circuits, prediction of electrostatic discharge, hotplate sensors, oxide-confined vertical-cavity lasers, and so on. Therefore, it is necessary to develop a electro-thermal model which computes the dependence between power dissipation and temperature distribution over the device. Moreover, such a heat transfer analysis needs to be done quickly in response to every design alteration. The model must also provide good accuracy in order to return precise temperature values.

The main problem of electro-thermal modeling is that ordinary differential equation (ODE) systems resulting from finite element, finite difference or some other spatial discretization method can easily reach an order of 100 000 or more. Even with the increasing speed of modern computers it is not possible to perform effective system-level simulation without simplification or model reduction.

Conventionally, the reduction of thermo-electric models for micro-electronic and MEMS devices is performed through a lumped-element decomposition of the model followed by parameter fitting [1]. Such a non-automatic approach requires the designer to choose the corect reduced model structure without strict guidelines, and to perform a time-consuming parametrization including one or more simulations of the full-scale model. Moreover, sufficient accuracy is provided only for a limited parametric domain.

In order to achieve both efficiency and accuracy in thermal management of microsystems, we propose a different, automatic order reduction approach, suitable for linear electro-thermal models, and based on Krylov-subspace methods. In adittion to the Arnoldi [2] and Lanczos [3] algorithm reported previously, we introduce a new method called *two-sided Arnoldi* to find two basis necessary for projection and calculating the reduced order model. All three algorithms were tested and compared with regard to their computational complexity, accuracy of approximation, numerical stability, preservation of the stability and passivity of the original system, invariance properties and approximation of the complete output.

As a MEMS test case a new class of high energy actuator, which integrates solid fuel with three silicon micromachined wafers [4], was used. This microthruster is ignited by passing an electric current through a polysilicon resistor embedded in the membrane, as shown in Fig. 1.

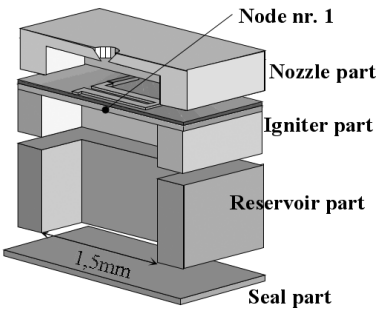

Fig. 1 Microthruster Structure.

The present work considers the initial heating phase of the fuel, right up to the onset of ignition, described through the following equations:

$$\nabla\bullet(\kappa\nabla T) + Q - \rho C_p\frac{\partial T}{\partial t} = 0,\ Q = \frac{\boldsymbol{j}^2}{\sigma} \tag{1}$$

where κ is the thermal conductivity, C_p is the specific heat capacity, ρ is the mass density, T is the temperature distribution, Q is the heat generation, $\boldsymbol{j}$ is the spartially varying electric current density vector and σ is the specific electric conductivity.

We use a two dimensional axi-symmetric model, which after the finite element (FE) based spatial discretization of the governing equations (1) (assuming that the heat generation Q is uniformly distributed within the heating area) results in a linear system of about 1000 ordinary differential equations (ODEs) of the form:

$$\begin{aligned}[C]\dot{\boldsymbol{T}} + [K]\boldsymbol{T} &= \boldsymbol{F}I(t)^2R\\ y &= E^T\cdot\boldsymbol{T}\end{aligned} \tag{2}$$

where $[K],[C] \in R^{n\times n}$ are the global heat conductivity and heat capacity matrix, $\boldsymbol{T}(t),\boldsymbol{F}, E \in R^n$ are the temperature (state), the load and the output vector respectively and n is the dimension of the system. The electric current $I(t)$ through the heater with electric resistivity R is the input to the system.The equation (2) represent a Single-Input-Single-Output (SISO) system. The present work also considers a special case when E is an identity matrix, that is, $y = \boldsymbol{T}$, which we call a Single-Input-*Complete* Output (SICO) system.

2 MODEL ORDER REDUCTION

Most of the practical work in model reduction of large linear dynamic systems has been tied with moment matching of the transfer function via Krylov subspaces by means of either the Arnoldi or Lanczos process. They define a projection from the high dimensional space of the original model (2) to a lower dimensional space and thereby define the reduced order model. By applying the transformation $\boldsymbol{T} = [V]\cdot\boldsymbol{T_r}$ where $V \in R^{n\times r}$ and $r < n$, to system (2) and then multiplying the state equation by transpose of some matrix $W \in R^{n\times r}$, a model with reduced order r can be found:

$$\begin{aligned}W^T[C]V\dot{\boldsymbol{T}}_r + W^T[K]V\boldsymbol{T}_r &= W^T\boldsymbol{F}I(t)^2R\\ y_r &= E^T\cdot V\boldsymbol{T}_r\end{aligned} \tag{3}$$

The reduced system matrices and load vector are computed by:

$$[C]_r = W^T[C]V; [K]_r = W^T[K]V; \boldsymbol{F}_r = W^T\boldsymbol{F} \tag{4}$$

The key question is: how to find V and W?

Let us rewrite the system (2) as:

$$\begin{aligned}[A]\dot{\boldsymbol{T}} &= T + \boldsymbol{b}u(t)\\ y &= \boldsymbol{c}^T\cdot\boldsymbol{T}\end{aligned} \tag{5}$$

with $A = -[K]^{-1}[C]$, $\boldsymbol{b} = -[K]^{-1}\boldsymbol{F}$ and $\boldsymbol{c} = E$. As already mentioned, the basic idea behind the Krylov-subspace-based algorithms is to write down the transfer function of (5) in the frequency domain using a Taylor series in the Laplace variable s around $s_0 = 0$:

$$\{G(s)\} = -\sum_{i=0}^{\infty}\{m\}_i s^i \tag{6}$$

where $\{m\}_i = \boldsymbol{c}^T(A)^i\boldsymbol{b}$ is called the i*th* moment, and then to find a much lower order system (of the same form as (5)) whose transfer function $\{G_r(s)\}$ has the same moments as $\{G(s)\}$ up to some degree. Due to the numerical instability of (6), the moments are not computed explicitly. Instead, a Krylov subspace (of the dimension *r*) defined as:

$$K_r\{A, \boldsymbol{b}\} = span\{\boldsymbol{b}, A\boldsymbol{b}, \ldots, A^{r-1}\boldsymbol{b}\} \tag{7}$$

is used. The vectors that span the subspace are called the basic vectors.

2.1 Arnoldi Algorithm

In case of Arnoldi algorithm $V = W$. This algorithm generates a set of orthonormal vectors (with length one and orthogonal to each other) which simultaneously represent a basis for the given Krylov-subspace (7). They are saved as columns of the matrix *V*. This means that $V^T\cdot V = I$. The algorithm further generates an upper Hessenberg matrix $(a_{ij} = 0, \forall(i,j), |j-i| > 1)$ $H_A \in R^{r\times r}$, related to the system matrix A as follows:

$$V^T\cdot A\cdot V = H_A \tag{8}$$

The matrix H_A can be considered as an orthogonal projection of the matrix A onto the Krylov-subspace (7), and it is equal to the system matrix of the reduced system A_r. It can be proved that the first r moments of $\{G_r(s)\}$ and $\{G(s)\}$ match [5].

2.2 Lanczos algorithm

For the Lanczos algorithm $V \neq W$. The columns of the matrix V form a basis (but not an orthonormal one) of the Krylov-subspace (7), which is also called an input Krylov-subspace. The columns of the matrix W form a non-orthonormal basis of the output Krylov-subspace defined as:

$$K_r\{A^*, \boldsymbol{c}\} = span\left\{\boldsymbol{c}, A^*\boldsymbol{c}, \ldots, (A^*)^{r-1}\boldsymbol{c}\right\} \tag{9}$$

where A^* is the conjugate transpose of the matrix A. The basis V and W of the subspaces (7) and (9) are biorthogonal, which means that $W^T \cdot V = I$. The algorithm generates further a tridiagonal matrix T_L related to the original system matrix as:

$$W^T \cdot A \cdot V = T_L \tag{10}$$

The matrix T_L can be considered as an oblique projection of the matrix A onto the input Krylov-subspace (7), and it is equal the system matrix of the reduced system A_r. In this case the first $2r$ moments of $\{G_r(s)\}$ and $\{G(s)\}$ match [5].

2.3 Two-Sided Arnoldi Algorithm

The two-sided Arnoldi algorithm computes matrix V as an orthonormal basis for the input Krylov-subspace (7), and matrix W as an orthonormal basis for the output Krylov-subspace (9), by twice using the one-sided Arnoldi algorithm (section 2.1). This means that $V^T \cdot V = I$ and $W^T \cdot W = I$. The reduced system is computed as in (3), and the first $2r$ moments of $\{G_r(s)\}$ and $\{G(s)\}$ match as in the Lanczos algorithm. Moreover, the transfer function of the reduced model gained from the two-sided Arnoldi method equals the one gained from the Lanczos algorithm [5]. Further details can be found in [6].

3 RESULTS

In Fig. 2 a relative error between the full and the 5th order reduced model, using both the one-sided Arnoldi and two-sided Arnoldi algorithm, is shown. The results for Lanczos algorithm are identical as for two-sided Arnoldi.

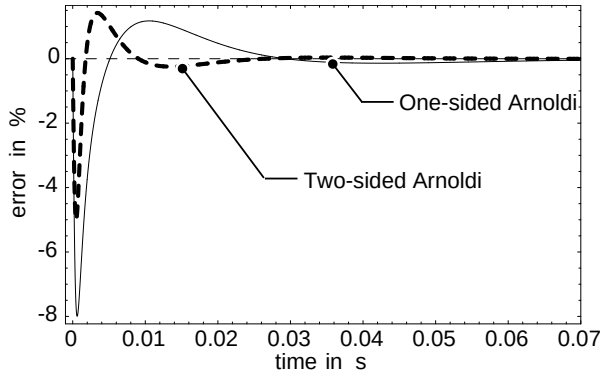

Fig. 2 Relative error between the full and the 5th order reduced model for a single node (node 1 in Fig 1).

For the microthruster model, the simple SISO setup for both the one-sided and the two-sided Arnoldi algorithm was sufficient to approximate not only a single output response but also the transient thermal response in all the finite element nodes of the microthruster (SICO setup). Fig. 3 shows

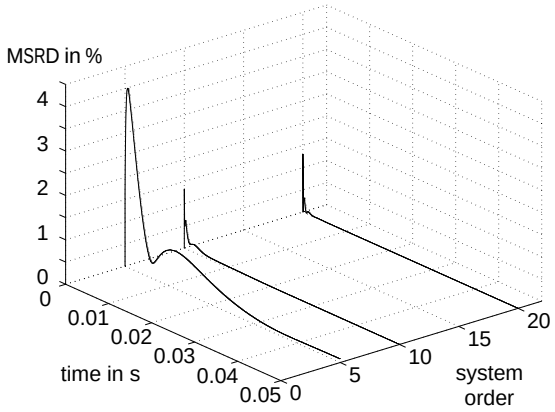

Fig. 3 Mean square relative difference (MSRD) for all the nodes during the initial 0.05s, for a two-sided Arnoldi reduction from order 1071 to 20, 10 and 5.

the mean relative difference for all the nodes between the full-scale and the reduced different order models. Hence it was possible, after the simulation of the reduced model, to recover the full solution (for all the 1071 nodes) by applying the projection $\boldsymbol{T} = [V] \cdot \boldsymbol{T_r}$.

In Table 1 the maximal relative error for a single output node (node 1 in Fig 1) e_1 and a maximal mean square relative differences for approximating the complete output e_2 for different orders of the reduced model are given. The results for the Lanczos algorithm are identical as for the two-sided Arnoldi.

Method	order	*e1*[%]	*e2*[%]
One-sided Arnoldi	20	0.964576	0.339707
	10	0.937031	1.44194
	5	7.99637	3.01407
Two-sided Arnoldi	20	0.0087617	1.30918
	10	1.67593	1.33582
	5	5.00608	4.03799

Table 1: Results for one-sided and two-sided Arnoldi

4 DISCUSSION AND CONCLUSION

Accuracy of approximation: As mentioned in chapter 2, the one-sided Arnoldi algorithm matches only *r* moments of $\{G_r(s)\}$ and $\{G(s)\}$, whereas the two-sided Arnoldi and Lanczos algorithm match *2r* moments. Hence, the two-sided Arnoldi and Lanczos algorithm are optimal in a sense that they match as many moments as there are free coefficients in the reduced order transfer function (*r* poles and *r* zeroes). This provides a unique reduced order model, i. e., invariance regarding the representation and the realization of the original system, but does not guarantee the preservation

of stability and passivity. In many applications it is better to trade some of the „optimality“ of two-sided methods to gain guaranteed stable and passive reduced order models, for example by using the one-sided Arnoldi reduction algorithm**.**

Preservation of the stability and passivity of the original system: In circuit simulation, reduced-order modeling is mostly applied to large passive linear subcircuits (such as RLC networks), and preservation of passivity is crucial for the stability of the simulation of the whole circuit. Unfortunately, the reduced models gained by two-sided Arnoldi and Lanczos algorithm are not passive and stable in general [3]. In the microthruster example, the reduced model of order 9 is unstable. It has been shown that the coordinate transformed one-sided Arnoldi algorithm on the other side, can generate guaranteed stable reduced-order models [7].

Computational complexity: One disadvantage of the Arnoldi method is that each new Arnoldi vector (new column of matrix *V*) should be orthogonal to all previously generated vectors. This means that the computational cost for orthogonalisation over the *r* steps of algorithm grows as $O(2r^2n)$ with the dimensions *n* of the full space and *r* of the Krylov-subspace. Additionally *r* steps of the Arnoldi procedure require *r* matrix-vector products at the cost of *2rNz(A)*, where *Nz(A)* is a number of nonzero elements of *A* ($Nz(A)=n^2$ for a dense matrix). Thus, on average the computational costs for (one-sided) Arnoldi algorithm grows as $O(2r^2n + 2rNz(A))$. For the two-sided Arnoldi algorithm the costs are double. The Lanczos algorithm has smallest computational costs for orthogonalisation. In each step it is necessary to deal with just two previously generated vectors (matrix T_L is tridiagonal), which makes the orthogonalisation costs over the *r* steps of the algorithm grow only as *O(16rn)*. Together with matrix-vector products, the Lanczos algorithm´s costs are of *O(16rn+4rNz(A))*. Hence, for large *r* ´the Lanczos algorithm is the fastest of three. However, it is numerically less stable than both the Arnoldi and the two-sided Arnoldi process.

Numerical stability: Because of rounding errors in all three algorithms, the column vectors of basis *V* and *W* for the Krylov-subspaces (7) and (9) may become non-orthogonal. How quickly this happens depends on the chosen dimension *r* of the Krylov-subspaces. For the one- and two-sided Arnoldi algorithm each new vector should be orthogonal to all previously generated vectors so that the rounding errors accumulate slower than by Lanczos algorithm, where each new vector is orthogonalised only with respect to the last two generated vectors.

Invariance properties: Changing the representation (multiplying the state equation with some matrix) or the realization (change of state vector) of the original system does not change the input-output behavior of the reduced order models generated by two-sided methods [5]. In case of one-sided Arnoldi algorithm these invariance properties are not given, due to matching only *r* moments which is half of the number of unknowns in $\{G_r(s)\}$.

Approximation of the complete output: In general the two-sided Arnoldi and Lanczos algorithms produce reduced-order models which are „optimized“ for particular output(s). This is due to their taking into account the output vector $\boldsymbol{c}$ by using a basis *W* of the output Krylov-subspace (9). It means that, even if in the case of the microthruster model two-sided Arnoldi properly describe the complete state vector (see e_2 in Table 1), in general this may not work. The one-sided Arnoldi algorithm on the other hand, doesn´t take into account the output vector at all, so we can expect the approximation of the complete output also in the general case.

Conclusion: Our results show that, for moderate values of *r* the two-sided Arnoldi is definitely better than Lanczos algorithm. Stability and passivity properties of two-sided methods however, need to be further researched. Meanwhile, we consider the one-sided Arnoldi algorithm as an acceptable tool for compacting linear electro-thermal models.

5 ACKNOWLEDGMENTS

This work is partially funded by the EU through the project MICROPYROS (IST-1999-29047), partially by the DFG project MST-Compact (KO-1883/6) and partially by an operating grant of the University of Freiburg.

6 REFERENCES

[1] V. d‘Alessandro, N.Rinaldi, “A Critical Review of Thermal Models for Electro-thermal Simulation”, *Solid-State Electronics*, **46**, p. 487-496 (2002).

[2] D. Ramaswami, J. White, “Automatic Generation of Small-Signal Dynamic Macromodels from 3D Simulation”, *Proc. 4th MSM*, p. 27-30 (2001).

[3] R. W. Freund, „Krylov-subspace Methods for Reduced Order Modeling in Circuit Simulation“, *Journal of Comutational and Applied Mathematics*, **123**, p. 395-421 (2000).

[4] E. B. Rudnyi, T. Bechtold, J. G. Korvink, and C. Rossi, „Solid Propellant Microthruster: Theory of Operation and Modelling Strategy“, *Proc. Nanotech’02, AIAA paper 2002-5755*, 2002.

[5] B. Salimbahrami, B. Lohmann, „Invariance Properties of Krylov Subspace Methods in Linear Model Order Reduction“, *Systems and Control Letters paper,* (2002).

[6] B. Salimbahrami, B. Lohmann, T. Bectold, J. G. Korvink, „Two-sided Arnoldi Algorithm and Its Application in Order Reduction of MEMS“, *Proc. MATHMOD*, (2003).

[7] L. M. Silveira, M. Kamon, I. Elfadel, J. White,“A Coordinate-transformed Arnoldi Algorithm for Generating Guaranteed Stable Reduced-Order Models of RL C Circuits”, *Comp. Methods. Appl. Mech. Eng.*, **169**, p. 377-389 (1999).