Connecting Heat Transfer Macromodels for MEMS-Array Structures

by
Tamara Bechtold[1], Evgenii B. Rudnyi[1], Markus Graf[2], Andreas Hierlemann[2], Jan G. Korvink[1]

[1]IMTEK - Institute for Microsystem Technology
University of Freiburg, Georges-Köhler-Allee 103,
79110 Freiburg, Germany

[2]Physical Electronics Laboratory, ETH Zurich,
HTP H8, CH-8093 Zurich, Switzerland

Short title
Heat Transfer Macromodels for MEMS Arrays

Connecting Heat Transfer Macromodels for Array MEMS Structures

Tamara Bechtold, Evgenii B. Rudnyi, Jan G. Korvink

IMTEK, University of Freiburg, Georges-Köhler-Allee 103, 79110 Freiburg, Germany

Markus Graf, Andreas Hierlemann

Physical Electronics Laboratory, ETH Zurich, HTP H8, CH-8093 Zurich, Switzerland

1 Introduction

In microelectronic and micro-electro-mechanical systems (MEMS), system-level modeling and simulation have become an essential requirement for system optimization and for the reduction of design cycles. Unfortunately, analytical solutions for the governing partial differential equations (PDEs) of microelectronic and MEMS components are only available for simple device geometries, whereas for complex geometries, either approximations or numerical methods must be used. However, the numerical solution of the PDEs, via e. g. finite elements (FE), is often impractical or even prohibitive for the system simulation of a large number of interconnected devices. Note that the number of resulting ordinary differential equations (ODEs) for a single device easily exceeds 100 000. Even by using a domain decomposition technique on parallel computers, this huge number of unknowns demands high resources on CPU-time and memory. Hence, a reduction of the number of unknowns to a lower-dimensional system, known as compact- or macromodeling is about to become a standard for MEMS and microsystem simulation in various physical domains [1] - [14]. Concerning the all-important heat transfer macromodeling, in recent years several methodologies for the extraction of dynamic compact thermal models (DCTM) of microelectronic and MEMS devices have been proposed [4] - [14]. However, they all consider the DCTM of a single device only. Since microelectronic and MEMS are usually composed of subsystems that

interconnect to each other, array structures for example, it is desirable, especially for a large number of subsystems, to extract a heat-transfer macromodel of each subsystem on its own and then to couple them back together. Hence, we seek a kind of compact thermal multiport representation, which admits thermal fluxes to cross the boundaries and allows straightforward coupling to the next thermal multiport. The main problem hereby, is that the thermal flow is not lumped by nature as for example the electrical flow is along metallic wire interconnect. The ratio of electrical conductivity of metals and that of insulators is of the order of 10^8. Hence, the electrical current flow takes place almost solely in metal paths. This is not the case with heat flow, because the ratio of thermal conductivities in microtechnology is only of the order of 10^2 (see Figure 1). Therefore, it is unclear how to lump the thermal fluxes at shared surfaces between two finite element models in order to form the thermal ports (Figure 2) which would serve to couple together several compact models. As a matter of fact, there appear to be very few works on how to couple (dynamic) compact thermal models [15], [16].

At this point, we would like to emphasize the difference between our method and the domain decomposition technique [17]. The goal of domain decomposition technique is the solution of the original ODE system on parallel computers. To this end, the system is split into several subsystems, which is similar to our starting point. However, in the domain decomposition technique each subsystem is solved on a single processor within an iterative process, which converges to the exact solution over the whole domain. In our case, we target a compact model for each subsystem and try to couple those reduced models with each other afterwards.

The goal of this paper is to present and discuss the currently available solutions for model order reduction (MOR) of thermal micro-array structures. In section 2 our MEMS case study, a MOS-transistor-based microhotplate array model [18], is presented. It is a high dimensional model, con-

taining several hundreds of FE nodes at each shared interface.

Section 3 describes the possibility of reducing the entire array, i.e., without decoupling its parts, via the block Arnoldi procedure [25]. This projection algorithm is suitable for multiple-input-multiple-output (MIMO) linear systems. It offers a simple and effective solution for the reduction of array structures, but in a way "hides" the problem of coupling the dynamic compact thermal models.

Section 4 discusses the possibility of coupling reduced order models via substructuring based on a modified Guyan algorithm [29]. This method is usually used in structural mechanics, but is available for the thermal domain as well. As it demands the preservation of all shared (coupling) nodes, the resulting orders of compact models are much larger than by block Arnoldi.

In Section 5 we discuss the possibility to combine the two described approaches, i.e., to reduce each subsystem using projection and then to couple the reduced models. Finally, section 6 concludes the paper with a brief comparison of the available methods.

2 Microhotplate Array

The microhotplate array model is built upon a real gas sensor device fabricated in CMOS technology [18], [19]. It is composed of a $500\mu m \times 500\mu m$ CMOS multilayer fully suspended membrane and a ring PMOS transistor heater of $5\mu m$ gate-length and $720\mu m$ overall gate with, buried in a silicon island under the membrane, as shown in Figure 3. Figure 4 shows an example of a fabricated chip with digital circuitry in the lower part, analog circuitry in the middle and an array of three microhotplates in the upper section. The microhotplates are covered with drop-deposited noncrystalline SnO_2 as sensitive layers. Full advantage is taken of the features offered by applying CMOS-technology. All sensor values can be set and read out via the digital interface,

which drastically reduces the packaging complexity since the number of bond wires is the same as for a single microhotplate.

Before the microhotplate devices are fabricated, the designs undergo an extensive simulation process such as thermal modelling using finite-element simulations. Modeling the transducer and hotplate behavior is an extremely important step towards the formulation of a compact sensor model for monolithic system realizations. The parameters of interest are, for example, the hotplate thermal resistance and thermal time constant, whose prediction facilitates and accelerates the design of the monolithic array systems for a given microhotplate design. The heat conduction within the array is described through the nonstationary heat transfer equation:

$$\rho C_p \frac{\partial T}{\partial t} - \nabla \bullet (\kappa \nabla T) = Q \tag{1}$$

where κ is the thermal conductivity, C_p is the specific heat capacity, ρ is the mass density, T is the temperature distribution and Q is the heat generation rate. In the following a linearity assumption will be made, which implies that all relevant material properties are considered to be temperature independent. This assumption is common for dynamic compact thermal models [10]. Under further assumption, that Q is uniformly distributed within the heating area, the finite element based spatial discretization converts (1) into a large linear ODE system of the form:

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

where $C,K \in R^{nxn}$ are the global heat capacity and heat conductivity matrices, $T(t) \in R^n$ is the temperature vector, $F \in R^{n \times m}$ and $E \in R^{n \times p}$ are the input and the output matrices, $Q(t) = \left[Q_1 \; Q_2 \; \ldots \; Q_m \right]^T$ is the vector of heat sources, n is the dimension of the system and m

and p denote the number of inputs (heat sources) and outputs.

For test purposes, we have built a finite element model of a 2x2 microhotplate array structure, which contains 100934 DOF (meshed and discretized with ANSYS8.0 using SOLID90 elements). We have modeled the membrane as a single layer with material properties as proposed in [20]. The transistor heater is modeled as a lumped element circular heater buried in the Si island under the membrane, which allows for a purely thermal simulation. Our model captures the thermal efficiency of the real device, which is $6°C/mW$ [21], to within 70%.

For the reduction of interconnected thermal systems it is necessary to describe the thermal crosstalk within a reduced order model. To accomplish this, we have modeled a 2x2 array as a part of still larger array. At all four side walls (Γ_q) of the chip, the convection boundary conditions were applied:

$$\forall \boldsymbol{r} \in \Gamma_q,\ q(r) = h(T - T_{ambient}) \tag{3}$$

where q is a heat flux through the boundary and h is the convection coefficient, which characterizes the thermal contact between the conducting solid and its adjacent gaseous material with the temperature $T_{ambient}$. The convection coefficient was set to $h = 10^4$ and the ambient temperature was set to $T_{ambient} = 0°C$. The initial temperatures was also set to zero. The heat source power for each circular heater was set to $40mW$. In sections 3 and 4 it will be shown that it is possible to turn the heat sources on and off also after the model order reduction step. Temperature contour plots in Figure 5 through Figure 7 demonstrate that the chosen BCs result in a temperature increase of about $10°C$ per added heat source, everywhere on the chip. This effect must be taken into account if, e. g., no automatic temperature control is employed. Note, however, that the model only resembles the real device and hence, the true crosstalk level may be different.

3 Block Arnoldi

The ideas of mathematical model order reduction have been developed in control theory [22], [23], and are applicable to the first order linear ODE systems, such as (2). They target a reduction of the number of state variables, i. e., equations within a system, while trying to preserve the input-output behavior. Mathematically, this is equivalent to projecting the system (2) onto some lower-dimensional subspace, using two orthonormal basis matrices $U \in R^{n \times r}$ and $V \in R^{n \times r}$, where n and r are the dimensions of the full and reduced systems, respectively. If such matrices are found, a reduced model is defined. Different MOR methods construct U and V in different ways. Methods suggested in [22] are mathematically optimal, i. e., offer a global error estimate, but require a computational complexity of $O(n^3)$, and hence are only applicable for systems with less than a few thousand unknowns. On the other hand, the Krylov-subspace based methods (Arnoldi process and Lanczos algorithm) [25] have no global error estimate, but can easily be applied to high-dimensional ODE systems.

In simple terms, the Arnoldi process [24] is a mathematical algorithm which can be used to find U and V with which to reduce the number of state variables of the first order linear ODE system, while preserving the all-important transfer function properties. Although it was originally applied to single-input-single-output systems, its modification to block Arnoldi [25] can be applied to approximate matrix-valued transfer functions as well. It offers a straightforward approach to reduce a model of an array structure with multiple heat sources, provided that the number of devices within the array remains moderate. By block Arnoldi the entire discretized array model (2) is reduced to a system with the same form:

$$C_r \cdot \dot{z} + K_r \cdot z = \boldsymbol{F}_r \cdot Q(t)$$
$$y_r = E_r^{\,T} \cdot z \tag{4}$$

but with much smaller dimension $r \ll n$. A generalized coordinate z can be seen as a projection of the n-dimensional temperature vector onto an r-dimensional subspace, subject to some (hopefully small) error ε:

$$\boldsymbol{T} = V \cdot z + \varepsilon, \quad z \in R^r, \quad r \ll n \tag{5}$$

and the entries of the vector $y_r \in R^p$ are those linear combinations of the reduced states z, which correspond to the required outputs y in (2). The matrix V in (5) is composed from r n-dimensional vector-columns that form a basis for the reduced subspace. When the subspace is found, the equation system (2) is projected onto it by applying (5) (ε is neglected) and then multiplying (2) from the left side by V^T:

$$V^T C V \cdot \dot{T}_r + V^T K V \cdot T_r = V^T F \cdot u(t)$$
$$y_r(t) = E^T V \cdot T_r \tag{6}$$

The block Arnoldi algorithm computes the transformation matrix V as an orthogonal basis for the so-called right block Krylov subspace of dimension *r*, induced by $A = K^{-1}C$ and $B = K^{-1}F$:

$$K_r(A, B) = colspan\{B, AB, \ldots, A^{j-1}B\} \tag{7}$$

The r column vectors of (7) should be linearly independent and hence a so-called deflation step may be needed (see [25] for more details). The algorithm takes as inputs the matrices K, C and the target reduced order r. The transformation matrix V is a direct output. Several variations of this algorithm are described in [26]. A crucial property of (7) is that the transfer functions of the

systems (2) and (4) defined as $G(s) = E^T(sC + K)^{-1}F$ and $G_r(s) = E_r{}^T(sC_r + K_r)^{-1}F_r$, when developed into Taylor series expansions around $s_0 = 0$:

$$G(s) = \sum_{i=0}^{\infty} m_i s^i \quad with \quad m_i = E^T(-K^{-1}C)^i K^{-1}F$$
$$G_r(s) = \sum_{i=0}^{\infty} m_{ri} s^i \; with \; m_{ri} = E_r{}^T(-K_r{}^{-1}C_r)^i K_r{}^{-1}F_r \tag{8}$$

where m_i and m_{ri} are called the i-*th* moments, match in the first r moments:

$$m_i = m_{ri}, i = 0, \ldots, r-1 \tag{9}$$

Furthermore, the passivity and stability of the original system are preserved in the reduced system [27]. The computation of $V = \left[v_1 \; v_2 \; \ldots \; v_r\right]$ within block Arnoldi is an iterative process and requires a new matrix-vector product in each iteration, namely:

$$v_{new} = A \cdot v_{previous} = K^{-1}C \cdot v_{previous} \tag{10}$$

with the starting vectors $v_1 \; v_2 \; \ldots \; v_m$ computed as m columns of the matrix $B = \left[b_1 \; b_2 \; \ldots \; b_m\right]$. Note that one never needs to form A and B explicitly. Instead, with a direct solver, one uses the sparse factorization of K (LU or Cholesky decomposition), and a fast back substitution in each iteration. In this way, model order reduction by Arnoldi can be performed approximately at the cost of a stationary system solution.

Figure 8 and Figure 9 compare the step responses of the full-scale and of the reduced order 50 model of the 2x2 microhotplate array, computed by block Arnoldi, for the case when two heat sources out of four (each $40mW$) are switched on. The four observed outputs are located in the center of each hotplate. Figure 10 shows further that the reduced model accurately reproduces the

crosstalk effect.

This results show, that block Arnoldi offers a simple and effective solution for the model order reduction of the 2x2 microhotplate array. As the multiple input vector $Q(t) = \left[Q_1\ Q_2\ Q_3\ Q_4\right]^T$ does not explicitly partake in the model order reduction process, it is possible to separately switch each heat source on and off after the order reduction has been performed. Furthermore, the input terms $Q_1 \ldots Q_4$ can also be nonlinear, i. e., the functions of time and temperature and can still be transferred to the reduced system (see [28] for more details). A further advantage of block Arnoldi is that the output matrix also does not explicitly participate in the model order reduction procedure and hence the approximation of the complete output (E is an identity matrix of dimension n) is warranted as well. The disadvantage, however, may become significant with the growing number of devices within an array. At some point, the system matrices C and K, although sparse, grow so large, that a LU- or Cholesky decomposition and matrix vector multiplications, as needed for (10), may become difficult or even prohibitive. In such cases, it is clearly better to search for an alternative to reduce each device on its own and then to couple these compact models back into an array model. How this can be done is discussed in section 5.

4 Coupling of the reduced order models via substructuring

Substructuring is a commercially available method [28], based on the modification of the Guyan algorithm [29], [30]. It has been for many years in common use in mechanical engineering. The Guyan algorithm is another model order reduction method, which projects (similar as equations (5) and (6) in the block Arnoldi) a high dimensional ODE system to a lower-dimensional one. A projection subspace is, however, chosen based on engineering intuition, rather than on mathematical rigour. Practically, this means that the designer can choose the “important” FE nodes which

are to be physically preserved within a reduced model. Below, we briefly explain the Guyan-based model order reduction process and show afterwards how it is possible to couple the reduced models in a process called substructuring. The presented results are computed using ANSYS 8.0.

The system matrices *C* and *K* obtained by spatial discretization of the heat transfer PDE, contain terminal nodes, which connect to external circuitry as well as to internal nodes. The terminal nodes can be chosen by the designer at will. Hence, the system matrices can be split into four blocks:

$$\begin{bmatrix} C_{ee} & C_{ei} \\ C_{ie} & C_{ii} \end{bmatrix} \cdot \begin{bmatrix} \dot{T}_e \\ \dot{T}_i \end{bmatrix} + \begin{bmatrix} K_{ee} & K_{ei} \\ K_{ie} & K_{ii} \end{bmatrix} \cdot \begin{bmatrix} T_e \\ T_i \end{bmatrix} = \begin{bmatrix} F_e \\ F_i \end{bmatrix} \cdot u(t) \tag{11}$$

with the index sets *e* and *i* ranging over all external and internal nodes respectively. In the stationary case ($C = 0$) it is possible to eliminate the equations for the non-terminal nodes by means of linear algebra operations (e. g. the Schur complement) [31] and to obtain a reduced system:

$$K_r \cdot T_e = F_r \cdot u(t) \tag{12}$$

with

$$\begin{aligned} K_r &= K_{ee} - K_{ei} \cdot K_{ii}^{-1} \cdot K_{ie} \\ F_r &= F_e - K_{ei} \cdot K_{ii}^{-1} \cdot F_i \end{aligned} \tag{13}$$

Assuming that no loads have been applied to the internal nodes ($F_i = 0$), the Schur complement amounts to a coordinate transformation of the form:

$$T = V_G \cdot T_e = \begin{bmatrix} I \\ -K_{ii}^{-1} K_{ie} \end{bmatrix} \cdot T_e \tag{14}$$

where I is the unity matrix, whose dimension corresponds to the number of terminal nodes. Note that, different than in equation (5) by block Arnoldi, a reduced state vector T_e is more than a generalized coordinate. It is a vector of physical temperatures at terminal nodes. Furthermore, the reduced heat conductivity matrix and the load vector (13) are seen to be built (equivalent as in (6)) by projection:

$$\begin{aligned} K_r &= V_G{}^T \cdot K \cdot V_G \\ F_r &= V_G{}^T \cdot F \end{aligned} \tag{15}$$

Guyan was first to generalize (via the energy functions) the projection (14) to order reduction of nondamped mechanical systems, i. e., to the mass matrix. A further generalization (without mathematical background) to the model order reduction of dynamical thermal systems (we call this method the modified Guyan algorithm) is done by projecting the heat capacity matrix C in the similar manner as in (15):

$$C_r = V_G{}^T \cdot C \cdot V_G \quad = \quad C_{ee} - C_{ei}K_{ii}^{-1}K_{ie} - K_{ei}K_{ii}^{-1}C_{ie} + K_{ei}K_{ii}^{-1}C_{ii}K_{ii}^{-1}K_{ie} \tag{16}$$

The main disadvantage of modified Guyan's algorithm versus Arnoldi, when applied to thermal models, is the much bigger dimension of the reduced system by the same accuracy [13]. It's big advantage however, is that it allows the coupling of several reduced models into an array structure, because it is possible to physically preserve the models surface nodes during reduction. Let us briefly explain this coupling.

An equation system which describes two thermal models, RM_1 with order r_1 and RM_2 with order r_2, reduced by the modified Guyan algorithm, using (15) and (16) is given as:

$$
\begin{aligned}
C_{r1} \cdot \dot{T}_{e1} + K_{r1} \cdot T_{e1} &= F_{r1} \cdot u_1(t) \\
C_{r1} \cdot \dot{T}_{e2} + K_{r2} \cdot T_{e2} &= F_{r2} \cdot u_2(t)
\end{aligned}
\tag{17}
$$

where $T_{e1} \in R^{r_1 \times 1}$ and $T_{e2} \in R^{r_2 \times 1}$ are terminal nodes vectors of RM_1 and RM_2 (see Figure 11). Of course, by performing a Guyan reduction of each submodel, the chosen terminal nodes T_{e1} and T_{e2} must contain the surface nodes which are to be coupled to another model. Depending on the desired accuracy and efficiency, additional nodes can be chosen as terminals. In Figure 11 the shared nodes are denoted as $T_{e1,\,i}$ and $T_{e2,\,i}$.

The coupling is done through the set of constraint equations for the temperatures and for the fluxes on the shared interface:

$$
\begin{aligned}
T_{e1,\,i} - T_{e2,\,i} &= 0,\ 1 \le i \le m \\
\dot{T}_{e1,\,i} - \dot{T}_{e2,\,i} &= 0,\ 1 \le i \le m
\end{aligned}
\tag{18}
$$

which actually reduce the number of variables in (17) by m. We now transform the equation system (17) into a block-matrix form:

$$
\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \cdot \begin{bmatrix} \dot{T}_1 \\ \dot{T}_2 \end{bmatrix} + \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} \cdot \begin{bmatrix} T_1 \\ T_2 \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}
\tag{19}
$$

where $T_1 = T_{e1,\,i}$ are the coupling nodes belonging to RM_1, and T_2 are all other nodes belonging to both models $T_2 = T_{e1,\,bulk} \bigcup T_{e2}$. For simplicity, the input terms $u_1(t)$ and $u_2(t)$ in (19) have been incorporated in the load vector. We further rewrite the constraint equations (18) as $T_1 = X \cdot T_2$ and $\dot{T}_1 = X \cdot \dot{T}_2$ or:

$$\begin{bmatrix} I & -X \end{bmatrix} \cdot T = 0$$
$$\begin{bmatrix} I & -X \end{bmatrix} \cdot \dot{T} = 0 \tag{20}$$

where I stands for the unity matrix of dimension $m \times m$ and $T = \begin{bmatrix} T_1 & T_2 \end{bmatrix}^T$.

The problem of solving (19) under the constraints (20) can be carried out by several standard techniques, like the Lagrange multipliers adjunction method, penalty function method, perturbed Lagrangian method etc. [32] - [34]. In the Lagrange multipliers adjunction method, the constraints are imposed by adding new equations containing the Lagrange multipliers, which leads to the multiplier-augmented form of (19) (see equation (6.3.41) and its generalization to transient problem (6.3.45) in [33]). In our special case of linear system with linear constraints, it is possible to eliminate both Lagrange multipliers and m temperatures T_1 of shared nodes. This turns the equation system (19) into a coupled reduced system of the form:

$$\tilde{C} \cdot T_2 + \tilde{K} \cdot T_2 = \tilde{F} \tag{21}$$

where

$$\begin{aligned} \tilde{C} &= X^T \cdot C_{11} \cdot X + X^T \cdot C_{12} + C_{21} \cdot X + C_{22} \\ \tilde{K} &= X^T \cdot K_{11} \cdot X + X^T \cdot K_{12} + K_{21} \cdot X + K_{22} \\ \tilde{F} &= X^T \cdot F_1 + F_2 \end{aligned} \tag{22}$$

Please note, that at this stage the heat sources are still independent, i. e. it is possible to scale the load vectors by scaling F_1 and F_2.

In order to accurately describe the heating area of a single hotplate, during the reduction we have preserved the 948 heater nodes in addition to the surface nodes. As a result, the reduced model of

a single microhotplate contains 2140 ordinary differential equations. The lower limit of the reduced model, when applying the modified Guyan method, would have been the description of the interfaces, which requires 1192 ODEs. The coupling of four reduced models into an array structure according to equations (21) and (22) results in an equation system of 7351 ODEs. Note once more that the full-scale FE model contains 100934 equations. An important computational issue here is that the reduced system (21) is dense and hence it's computation may not necessarily represent a large decrease in CPU time versus the computation of the full-scale system, which is sparse. Figure 12 shows the assembled contour plots of four substructured and coupled microhotplates when two heat sources are switched on. The approximation of the thermal crosstalk effect, when compared with the full-scale FE model (Figure 6), is excellent. Figure 13 further compares the step response of the substructured array with the step responses of the full-scale model and of the block Arnoldi reduced model from Figure 8.

Let us emphasize ones more, that the convenience of modified Guyan reduction algorithm is that it preserves the chosen terminal nodes in a physical sense, while using the projection (14). Hence, it is possible to decouple an array, to reduce each device on it's own, while preserving the surface nodes as terminals and to couple several reduced models back into an array structure. The coupling is done via the terminals, by using any of the methods for treating linear multifreedom constraints in finite elements systems [32] - [34]. The main disadvantage, is that the lower size limit of a single reduced model is defined by the description of the surfaces, which implies, when using accurate meshes, relatively large number of equations. For the microhotplate array, several hundred of nodes per single surface were present. It is of course possible to reduce the number of coupling nodes by coarsening the mesh on the interface, but this results in a more difficult mesh generation step and lower mesh quality. Besides, we may need to choose more internal (non ter-

minal) nodes in order to reach the desired accuracy (see discussion in [13]).

5 Discussion

Both presented methods have their advantages and disadvantages. Block Arnoldi is straightforward, for it reduces the entire array model and offers high accuracy. Its crucial disadvantage is that the computation can become prohibitive as the number of devices within an array increases. Substructuring requires the decoupling of the array model and physical preservation of all shared nodes during reduction of a single device using a modified Guyan algorithm. This allows easy back coupling of the reduced models, but results in an unnecessary large reduced array model. For the presented case study, the block Arnoldi method could reduce an array model of the order 100934 down to order 50, whereas substructuring required an order of 7351 for the reduced model. Hence, the best way would be to reduce each subsystem by (Krylov-subspace-based) projection, as it guarantees smaller dimension for the same accuracy than Guyan-based substructuring [13], and then to try to couple several reduced models back into the array structure. Let us explain how this could be done.

If the MOR of a single device model is done by Krylov-subspace projection (Arnoldi process), the surface nodes are not preserved. Instead, we have generalized coordinates z. In this case the Lagrange multipliers method cannot be applied any more and hence, the only way to couple the reduced models back into an array structure is over additional “flux inputs”, as follows.

The two, still uncoupled models (M1 and M2 in Figure 14), are described through the equation system:

$$\begin{bmatrix} C_{11} & 0 \\ 0 & C_{22} \end{bmatrix} \cdot \begin{bmatrix} \dot{T}_1 \\ \dot{T}_2 \end{bmatrix} + \begin{bmatrix} K_{11} & 0 \\ 0 & K_{22} \end{bmatrix} \cdot \begin{bmatrix} T_1 \\ T_2 \end{bmatrix} = \begin{bmatrix} F_1 \cdot u_1 \\ F_2 \cdot u_2 \end{bmatrix} \tag{23}$$

Let us for the beginning suppose that M1 and M2 are the full-scale models with surface nodes vectors T_α and T_β. The heat fluxes between the models surfaces are introduced as:

$$\begin{aligned} \Phi_\alpha &= h \cdot (T_\alpha - T_\beta) \\ \Phi_\beta &= h \cdot (T_\beta - T_\alpha) \end{aligned} \tag{24}$$

where the heat transfer coefficient *h* can be chosen on the basis of e. g., experimental measurements. The coupling of M1 and M2 can be done by simply adding the right hand side of (24) to (23) while first rewriting (24) into a matrix form:

$$\begin{bmatrix} \Phi_\alpha \\ \Phi_\beta \end{bmatrix} = H \cdot \begin{bmatrix} T_1 \\ T_2 \end{bmatrix} \tag{25}$$

Note that *H* is of low rank, i. e., only the rows and columns corresponding to the coupling nodes T_α and T_β are non-zero. These rows contain *h* and -*h* entries at the α_i and β_i locations:

$$H = \begin{array}{c} \\ \alpha_i \\ \\ \beta_i \\ \end{array} \begin{array}{c} \begin{array}{ccccc} \alpha_i & & & & \beta_i \end{array} \\ \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 0 & h & \dots & -h & 0 \\ & & & & \\ 0 & -h & \dots & h & 0 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix} \end{array} \tag{26}$$

A coupled system is then given through:

$$\begin{bmatrix} C_{11} & 0 \\ 0 & C_{22} \end{bmatrix} \cdot \begin{bmatrix} \dot{T}_1 \\ \dot{T}_2 \end{bmatrix} + \begin{bmatrix} K_{11} & 0 \\ 0 & K_{22} \end{bmatrix} \cdot \begin{bmatrix} T_1 \\ T_2 \end{bmatrix} = \begin{bmatrix} F_1 \cdot u_1 \\ F_2 \cdot u_2 \end{bmatrix} + H \cdot \begin{bmatrix} T_1 \\ T_2 \end{bmatrix} \tag{27}$$

Notice that we need a strategy on how to choose an optimal *h*, and that in this way the temperatures in the coupling nodes T_α and T_β, differently than in substructuring method, are not necessarily equal to each other during time evolution. It is now possible to treat the coupling flux terms as additional inputs by rewriting (27) as:

$$\begin{aligned} C_{11} \cdot \dot{T}_1 + (K_{11} - h_{11}) \cdot T_1 &= F_1 \cdot u_1 + h_{12} \cdot T_2 \\ C_{22} \cdot \dot{T}_2 + (K_{22} - h_{22}) \cdot T_2 &= F_2 \cdot u_2 + h_{21} \cdot T_1 \end{aligned} \tag{28}$$

where $h_{11}, \ldots, h_{22}$ are the four matrix blocks of H. For convenience we will transform (28) into the right hand side formulation:

$$\begin{aligned} \dot{T}_1 &= A_1 \cdot T_1 + B_1 \cdot u_1 + h_1 \cdot T_2 \\ \dot{T}_2 &= A_2 \cdot T_2 + B_2 \cdot u_2 + h_2 \cdot T_2 \end{aligned} \tag{29}$$

with $A_1 = -C_{11}^{-1} \cdot (K_{11} - h_{11})$, $A_2 = -C_{22}^{-1} \cdot (K_{22} - h_{22})$, $B_1 = C_{11}^{-1} \cdot F_1, B_2 = C_{22}^{-1} \cdot F_2$, $h_1 = C_{11}^{-1} \cdot h_{12}$ and $h_2 = C_{22}^{-1} \cdot h_{21}$. Equation (29) stands for a system of two coupled models, when no MOR has been performed. Let us emphasize once more, that what we really want, is to couple two reduced models for case when no surface nodes T_α and T_β have been preserved. In order to be able to follow the reverse path, we project (29) using:

$$\begin{aligned} T_1 &= V_1 \cdot z_1 \\ T_2 &= V_2 \cdot z_2 \end{aligned} \tag{30}$$

to get a coupled reduced system of the form:

$$\begin{aligned} \dot{z}_1 &= V_1{}^T \cdot A_1 \cdot V_1 \cdot z_1 + V_1{}^T \cdot B_1 \cdot u_1 + V_1{}^T \cdot h_1 \cdot V_2 \cdot z_2 \\ \dot{z}_2 &= V_2{}^T \cdot A_2 \cdot V_2 \cdot z_2 + V_2{}^T \cdot B_2 \cdot u_2 + V_2{}^T \cdot h_2 \cdot V_1 \cdot z_1 \end{aligned} \tag{31}$$

where z_1 and z_2 are the generalized coordinates. Note, that (31) is formally correct, regardless the rank of the matrices h_1 h_2 and the choice of V_1 and V_2. Hence, if we know the projection matrices and the heat transfer coefficients at the shared surfaces, equation (31) couples the reduced models. The key question becomes, how to find a good projection subspace, especially if the number of coupling nodes is large, as it is usually the case after the spatial discretization of thermal domain.

The authors in [15] suggest a modal approach, which is to construct the projection matrices V_1 and V_2 by using the eigenvectors of the original system matrices $C_{11}^{-1} \cdot K_{11}$ and $C_{22}^{-1} \cdot K_{22}$ of (23). Such foregoing offers a facility of not having to recreate the reduced model corresponding to the new coupling conditions, but provides no guideline on how to choose the important eigenvectors.

It is also possible to reduce a number of shared surface nodes already at the meshing level, by coarsening the mesh on the interface. In such a case, the rank of the matrices h_1 and h_2 is low. Hence, it is convenient to apply a block Arnoldi projection to each subsystem in (31) with the starting vectors defined as (non zero) columns of $\left[B_1\ h_1\right]$ and $\left[B_2\ h_2\right]$. V_1 and V_2 are computed as basis for the block Krylov subspaces:

$$K_{r1}\left\{A_1, \left[B_1\ h_1\right]\right\} \text{ and } K_{r2}\left\{A_2, \left[B_2\ h_2\right]\right\} \tag{32}$$

However, the reduction of shared node numbers at meshing level, requires additional knowledge of ANSYS meshing tool and makes a process non automatic.

It is further possible to consider only a sum of the columns of h_1 and h_2 by constructing (32) and then to compute (31) as written above. This is equivalent to forcing the entire heat flow between two devices through a single point. Alternatively, the average heat flow could be taken as well.

Both last approaches require a sacrifice of precision at the model's surfaces. In the case when there is only a small temperature gradient over the surface nodes (by analogy this is the case we have for electrical interconnects), this might be an acceptable approximation. It is also worthwhile mentioning that, the solution sensitivity with respect to the number of nodes on the interface have been studied for the domain decomposition technique [35], [36] and hence, some of the ideas may be "borrowed" in the future. An implementation of these issues certainly deserves further research.

6 Conclusion

In this paper we have presented and applied two methods for the MOR of interconnected thermal systems, block Arnoldi and the Guyan-based substructuring. The application of the block Arnoldi algorithm is straightforward. It reduces the entire array model and results in a much smaller reduced model sizes than by using substructuring. Its main disadvantage is that it is not well scalable to a large number of devices within an array. Substructuring decouples the array model and physically preserves the shared nodes. This allows easy back-coupling of the reduced models, but results in unnecessary large array model sizes. Hence, both methods require alternatives for use in engineering practice.

We have discussed the possibility to find a “mixture” of both methods, which could work in the general case, i. e., to reduce each subsystem by some kind of projection and then to couple the reduced models via the surface fluxes into an array. Equation (31) couples the reduced models regardless of the number of coupling nodes, and regardless of the choice of the projection subspaces. The question which remains open is, how to find good projection matrices. Due to the large number of shared nodes (several hundreds), it appears that the Krylov subspace methods have met their limits. However, by observing the numerical simulation results for the microhotplate array, we find almost no temperature gradient over the models surface nodes. This can be expected for other thermal MEMS models, which by designed allow a heat transfer mostly in horizontal direction, as well. In such cases, it may be acceptable to focus the heat flux to one or several points, which allow us to consider the coupling flux terms as additional inputs. Hence, the subspaces (32) could be easily computed and block Arnoldi could be applied for the reduction of each subsystem. The back coupling can be done afterwards using equation (31). Figure 15 schematically summarizes the main characteristics of the discussed methods.

Let us mention at the end that currently, a new method, called structure preserving MOR, is being researched by mathematicians [37], [38]. It seems to offer the possibility to reduce a subsystem without even decoupling an array. This method, as well as the reduction of shared node numbers certainly deserves further research.

7 References

[1] X. Zhao, E. M. Abdel-Rahmann, A. H. Nayfeh, “A reduced order model for electrically actuated microplates“, *J. Micromech. Microeng.* **14**, pp. 900-906, (2004).

[2] J. Lienemann, D. Billger, E. B. Rudnyi, A. Greiner, J. G. Korvink, “MEMS Compact Modeling Meets Model Order Reduction: Examples of the Application of Arnoldi Methods for Microsystem Devices”, *Proc. NANOTECH2004*, pp. 303-306, (2004).

[3] E. B. Rudnyi, J. Lienemann, A. Greiner, J. G. Korvink, “mor4ansys: Generating Compact Models Directly from ANSYS Models”, *Proc. NANOTECH2004*, pp. 279-282, (2004).

[4] Y-J. Yang, K-Y. Shen, “Non-linear heat-transfer macromodeling for MEMS thermal devices“, *J. Micromech. Microeng.* **15**, pp. 408-418, (2005).

[5] T. Bechtold, E. B. Rudnyi, J. G. Korvink, “Error indicators for fully automatic extraction of heat-transfer macromodels for MEMS”, *J. Micromech. Microeng.* **15**, pp. 430-440, (2005).

[6] J. Palacin, M Salleras, M Puig, J Samitier and S Marco, “Evolutionary algorithms for compact thermal modelling of microsystems: application to a micro-pyrotechnic actuator“, *J. Micromech. Microeng.* **14**, pp. 1074-1082, (2004).

[7] T. Bechtold, J. Hildenbrandt, J. Wöllenstein, J. G. Korvink, “Model Order Reduction of 3D Electro-Thermal Model for a Novel Micromachined Hotplate Gas Sensor“, *Proc. EUROSIME2004*, pp. 263-267,(2004).

[8] V. Szekely, M. Rencz, “Thermal dynamics and the time constant domain”, *IEEE Trans. Comp. Pack. Tech.,* **23**(3), pp. 587-594, (2000).

[9] Y. C. Gerstenmaier, G. Wachutka "Time dependent temperature fields calculated using eigenfunctions and eigenvalues of the heat conduction equation", *Microelectronics Journal* **32**(10-11), pp. 801-808, (2001).

[10] M. N. Sabry, “Dynamic Compact Thermal Models Used for Electronic Design: A Review of Reacent Progress“, *Proc. IPACK03*, (2003).

[11] Y. Yang, C.Yu, “Extraction of heat-transfer macromodels for MEMS device“, *J. Micromech. Microeng.* **14**, pp. 587-596, (2004).

[12] L. Codecasa, D. D’Amore, P. Maffezzoni, “An Arnoldi Based Thermal Network Reduction Method for electro-Thermal analysis“, *IEEE Trans. Comp. Pack. Tech.*, **26**(1), pp. 186 - 192, (2003).

[13] T. Bechtold, E. B. Rudnyi, J. G. Korvink, “Automatic Generation of Compact Electro-Thermal Models for Semiconductor Devices”, *IEICE Trans. Electron.*, **3**, pp. 459 - 465, (2003).

[14] B. Shapiro, “Creating Compact Models of Complex Electronic Systems: An Overview and Suggested Use of Existing Model Reduction and Experimental System Identification Tools“, *IEEE Trans. Comp. Pack. Tech.*, **26**, pp. 165-172, (2003).

[15] D. Petit, R. Hachette, “Model Reduction in Linear Heat Conduction: Use of Interface Fluxes for the Numerical Coupling“, *Int. J. Heat Mass Transfer*, **41**, pp. 3177-3189, (1998).

[16] X. Guo, D. Celo, D. J. Walkey, T. Smy, “The Use of Constant Heat Flow Ports for Thermal Macro-Models“, *Proc. THERMICNIC04*, pp.323-328, (2004).

[17] B. Smith, P. Bjorstad, W. Gropp, “Domain Decomposition Parallel Multilevel Methods for Eliptic Partial Differential Equations“, *Cambridge University Press*, (1996).

[18] M. Graf, S. Taschini, P. Käser, C. Hagleitner, A. Hierlemann, H. Baltes, “Digital MOS-Transistor-Based Microhotplate Array for Simultaneous Detection of Enviromentally Relevant Gases", *Proc. IEEE MEMS 2004*, pp.351-354, (2004).

[19] M. Graf, R. Juriscka, D. Barrettino, A. Hierelemann, “3D nonlinear modeling of microhot-

plates in CMOS technology for use as metal-oxide-based sensors“, *J. Micromech. microeng.*, **15**, pp. 190-200, (2005).
[20] Martin von Arx “Thermal Properties of CMOS Thin Films“, *PhD thesis submitted to ETH Zürich No. 12743*, (1998).
[21] M. Graf, D. Barrettino, P. Käser, J. Cerda, A. Hierlemann, and H. Baltes, “Smart Single-Chip CMOS Microhotplate Array for Metal-Oxide-Based Gas Sensors", *Proc. IEEE Transducers'03*, pp.123-126, (2003).
[22] G. Obinata, B. D. O. Anderson, “Model Reduction for Control System Design”, *Springer,* (2004).
[23] B. N. Datta, “Numerical methods for Linear Control Systems”, *Elsevier Inc.*, (2004).
[24] W. E. Arnoldi, “The principle of minimized iterations in the solution of the matrix eigenvalue problem“, *Quart. Appl. Math.*, **9**, pp. 17-29, (1951).
[25] R. Freund, “Krylov-subspace methods for reduced order modeling in circuit simulation“, *J. Comp. App. Math.*, **123**(1-2), pp. 395-421, (2000).
[26] Y. Saad, “Iterative Methods for Sparse Linear Systems”,*PWS Publishing Company*, (1996).
[27] R. Freund, “Passive reduced-order modeling via Krylov-subspace methods “, *Proc. Inter. Conf. Computer Aided Contr. Sys. Design*, (2000).
[28] ANSYS Release 5.6, *User Manuals-Theory Manual*, (1999).
[29] R. J. Guyan, “Reduction of stiffness and mass matrices“, *AIAA Journal*, **3**, 2, p. 138, (1965).
[30] Z. Q. Qu,“Model Order Reduction Techniques with applications in finite element analysis”, *Springer*, (2005).
[31] C. Meier, M. Emmenegger, S. Taschini, H. Baltes, J. G. Korvink, “Equivalent Circuit Model of Resistive IC Sensors Derived with the Box Integration Method”, *IEEE Trans. on Comp.-Aided Design of Int. Circ. Syst.*, **18**(7), (1999).
[32] Zienkiewicz, “The Finite Elemente Method“, *McGraw-Hill Book Company (UK)*, (1977).
[33] T. Belytschko, W. K. Liu, B. Moran, “Nonlinear Finite Elements for Continua and Structures“, John Wiley & Sons Ltd., (2000).
[34] Introduction to Finite Element Methods (ASEN 5007) - Fall 2004, Department of Aerospace Engineering Sciences, University of Colorado at Boulder, *http://titan.colorado.edu/courses.d/IFEM.d/*
[35] P. L. Tallec, T. Sassi, M. Vidrascu, “Three-dimensional Domain Decomposition Methods with Nonmatching Grids and Unstructured Coarse Solvers“, *Contemporary Mathematics*, **180**, pp. 61-71, (1993).
[36] J. Pousin, T. Sassi, “Domain decomposition with non matching grids and adaptive finite element techniques“, *East-West J. Numer. Math.*, **8**(3), pp. 243-256, (2000).
[37] R. C. Li, Z. Bai, “Structure-Preserving Model Reduction“, *Proc. PARA'04,* (2004).
[38] A, Vanderdorpe, P. Van Dooren, “Model Reduction of Interconnected Systems”, Proc. *MTNS2004, (2004)*.

Captions for figures

Figure 1 Comparison of Distributive effects: electric flow is lumped and heat flow not.

Figure 2 Continuous thermal flux through the shared interface of two FE models. The goal is to model the “FE cubes” as thermal multiports i. e. to lumpe a flux.

Figure 3 Cross-section of a single microhotplate.

Figure 4 Micrograph of the chip with microhotplate array and circuitry.

Figure 5 Steady-state solution of a 2x2 array with a single heat source of 40mW.

Figure 6 Steady-state solution of a 2x2 array with two heat sources of $40mW$ each.

Figure 7 Steady-state solution of a 2x2 array with three heat sources of $40mW$ each.

Figure 8 Step response of the full-scale and reduced order models in two output points for two heat sources of $40mW$ each.

Figure 9 Step response errors corresponding to the Figure 8.

Figure 10 Step response of the reduced order 50 models in a single output point, when one, two or three heat sources of $40mW$ each are switched on.

Figure 11 Two Guyan-based reduced models with terminal nodes T_{e1} and T_{e2} which can be coupled over the surface nodes $T_{e1,i}$ and $T_{e2,i}$.

Figure 12 Expanded solution of four coupled reduced models.

Figure 13 Single output step responses of the full-scale and reduced order models created by block Arnoldi and Guyan-based substructuring. Each heat source has $40mW$.

Figure 14 Coupling via surface fluxes.

Figure 15 Possibilities for MOR of interconnected systems.