EXTRACTION OF FREQUENCY DEPENDENT MACROMODELS FOR MEMS SQUEEZED-FILM DAMPING EFFECTS

Yao-Joe Joseph Yang*   Chih-Ming Chien**

Department of Mechanical Engineering
National Taiwan University
Taipei, Taiwan 10617, R.O.C.

Mattan Kamon***   Vladimir L. Rabinovich***   John R. Gilbert****

Coventor, Inc.,
Cambridge, MA, U.S.A.

ABSTRACT

In this paper, an efficient macromodel extraction technique for dynamical MEMS gas damping effects is presented. The technique applies an Arnoldi-based model-order-reduction algorithm to generate low-order macromodels from a FEM approximation of the governing equation of the squeeze-film fluidic damping effect, the Reynolds equation. We demonstrate that this approach is more than 100 times efficient than previous approaches, which solve the Reynolds equation using transient finite-element/finite-difference methods. The generated gas-damping macromodels can be easily inserted into system-level modeling packages, such as SPICE, Saber and Simulink, for transient and frequency coupled-domain analysis. We also demonstrated that the simulated results are in good agreement with experimental results for various MEMS devices.

Keywords: Macromodel, Squeeze-film damping, MEMS, Model order reduction.

1. INTRODUCTION

A large class of MEMS devices must operate at significant gas pressures. To understand the dynamical behaviors of these devices, the effects of the gas surrounding the movable component can be critical. In terms of operation modes, the MEMS gas damping effects can be classified as two types [1]. The first type corresponds to the devices operating in in-plane motions. This type of damping is also known as the lateral damping. Comb-drive-type devices, such as gyro, in-plane actuators, are the most typical examples that are under lateral damping effect [2]. The second type of damping effect corresponds to the devices operating in out-of-plane motion. The squeeze-film damping, which involves two parallel plates moving close or away from each other, is one of common damping effect for the device in out-of-plane motion. Parallel-plate-type devices, such as accelerometers, microrelay, switches, optical modulators and micromirrors, are the typical examples whose dynamics are significantly affected by the squeeze-film damping effect [3].

Most previous works in the MEMS squeeze-film damping were focused on getting more accurate simulations of the small or large-signal 3D gas damping and spring effects [3~9] using the finite-element or the finite-difference method (FEM/FDM) solvers. The typical results of such analyses are similar to Fig. 1, which shows the gas damping and spring components vs. oscillation frequencies for the microrelay [10] shown in Fig. 2(a). It is well-known that dynamical coupled-domain simulations are very important for the device performance prediction and optimization, which are especially critical during the design phase of MEMS product commercialization. Although FEM/FDM solvers can perform explicit transient simulations or spectral analyses, the computational costs are very intensive, making the simulations extremely difficult or infeasible for coupled-domain dynamical analyses. In order to perform efficient coupled-domain simulations, creating accurate macromodels (i.e., compact models or reduced-order models) for different physical domains is a key step, since almost all the MEMS devices are coupled-domain systems. The step of creating macromodels is analogous to the process of extracting SPICE circuit models from the full-3D physical models of CMOS or BJT transistors. This paper is about the
development of efficient techniques to automatically extract compact damping models from the full-3D fluidic mechanics damping models.

The extraction procedure is shown in Fig. 3. A 3D model of a MEMS device is built using the Coventorware [11]. The finite-element (FE) matrix system for solving transient small amplitude motion of the Reynolds equation, the governing equation for the squeeze-film damping effects, is constructed and then reduced to a low order system using the Arnoldi-based algorithm [12–15]. After the reduced system is scaled with proper gap-dependent coefficients, it is then converted to a time domain representation written in a hardware description language (HDL), in this case MAST for use in SABER [16,17]. Because the model is generated directly from the FEM system, the transient calculations of the full FEM model for a wide frequency range, as reported in [7], is no longer necessary to be performed. This brings the total computation time to solve a frequency-dependent gas-damping model to a few minutes vs. a few hours/days in previous approaches [5,7,9].

2. THEORY

The models are extracted using the Arnoldi-based model reduction algorithm as PRIMA [12–15], which is commonly used for model reduction of electrical interconnects for integrated circuits. Here, we apply this algorithm for the squeeze-film damping effect exclusively and build models in the mechanical vibrating mode shapes. Such an approach would allow these models to be readily combined with the mode-shape-based models (the normal mode method) in the low order model of the entire system [18,19].

The governing equation of the squeeze-film damping is the isothermal Reynolds Eq. [3]. The isothermal Reynolds equation is actually a simplified equation by combining of the Navier-Stokes equation, the continuity equation and the equation of state. The detailed derivation of the isothermal Reynolds equation is tedious, and can be found in [3]. However, it is important to stress the assumptions for deriving the Reynolds equation these assumptions are also the conditions for appropriately using the Reynolds equation. These assumptions are: (1) the pressure distribution across the gap is uniform (2) the fluidic velocity component perpendicular to plate surfaces is negligible (3) the gap between the plates is much smaller than the length (width) of the movable plate. For the device with small amplitude motion, the Reynolds equation can be further simplified (linearized) as:

\[
\frac{h_0}{P_0} \frac{\partial p}{\partial t} = \frac{k_0^3}{12\mu} \nabla^2 p - \frac{\partial e}{\partial t}
\]

where the variation in plate spacing \( h \) is assumed to be small compared to the mean spacing, \( h_0 \), given by

\[
h = h_0 + e(x, t)
\]

with \( x \in \mathbb{R}^2 \) and \( e \ll h_0 \). \( \mu \) denotes the air viscosity which can be represented as a function of Knudsen number if the gas rarefaction effect becomes significant. In this work, \( \mu \) is set to be a constant since the initial gaps of our simulation cases are large enough (about 3 \( \mu m \) or larger) that neglecting gas rarefaction effect will not introduce noticeable modeling errors. \( p \) is the pressure variation around the ambient pressure \( p_0 \), and \( |p| \) is much less than \( p_0 \). Therefore, the absolute pressure can be written as:

\[
P = P_0 + p
\]

Equation (1) can be solved by the finite element method for a given \( e(x, t) \). Let \( f(x) \) be the mode shape of the displacement so that \( e(x, t) = f(x) \cdot u(t) \).

The dynamic system obtained by discretizing Eq. (1) using the finite element method can be written in state space form as

\[
\frac{h_0}{P_0} B \frac{d\mathbf{p}}{dt} = \frac{k_0^3}{12\mu} A\mathbf{p} + B\mathbf{f}(t)
\]

where \( A, B \in \mathbb{R}^{n\times n} \), where \( n \) is the number of nodal degrees of freedom, \( \mathbf{p} \) is a column vector which contains the pressure variation at each node, and \( \mathbf{f} \) is \( f(x) \) evaluated at the node points. \( y \) is then the net force projected into the shape defined by \( f(x) \) in a finite element sense.

Equation (4) can be rewritten as:

\[
\frac{d\mathbf{p}}{dt} = c_1 \mathbf{p} + c_2 \mathbf{S}(t)
\]

\[
y = Q^T \mathbf{p}
\]

Fig. 1 Results of 3D physics analysis of gas spring and damping for beam switch of Fig. 2(a), operating in its first vibration mode. The solid lines are obtained using the full Reynolds equation solver. The points show the overlap of the macromodel obtained from the Arnoldi-based model-order-reduction technique.
where \( R = B^{-1}A \), \( S = \mathbf{f} \), \( c_1 = \frac{P_k h_k^2}{12\mu} \), \( c_2 = \frac{P_k}{h_0} \) and \( Q = BT \).

During the finite-element discretization process, \( B \) is a diagonal matrix if the lumped-mass formulation is used. Therefore, the computation of matrix \( R \) is trivial (a matrix inversion computation is not required).

If the moving structure is rigid and is in a uniform-displacement motion (\( f(x) = 1 \)), the input excitation is \( e(x, t) = e(t) = u(t) \). Therefore, the output \( y \) can be considered as the total damping force against the plate. In this case, the system dynamic equation can be written as

\[
M \ddot{z} + K z = F_D
\]

where \( F_D \) is the total damping force and is equal to the output \( y = \mathbf{f}^T B^T \mathbf{p} \). \( M \) and \( K \) are the corresponding lumped mass and lumped spring constant for this uniform displacement motion. \( z \) is the displacement and \( z \) is equal to the input \( u(t) \) in Eq. (4).

If the structure is vibrating in a specific non-uniform mode shape, the input excitation becomes \( e(x, t) = f(x) \cdot u(t) \). In this case, the output \( y \) is the total damping force for this mode. In order to appropriately use the damping force corresponding to this mode, the dynamic equation of the system should be constructed using the normal mode method [19].

\[
M_g \ddot{z}_g + K_g z_g = F_{Dg}
\]

where \( M_g \) and \( K_g \) are the corresponding generalized mass and spring constant, respectively. \( F_{Dg} \) is the total damping force projected on to the mode shape and is equal to the output \( y \). \( z_g \) is the generalized displacement for this oscillating mode, and \( z_g \) is equal to the input \( u(t) \) in Eq. (4).

The most typical way to study the transient dynamics as well as frequency responses of a coupled-physical-domain MEMS system is to use a system-level simulator, such as the Matlab, the SPICE or the SABER. A system-level simulator is suitable to solve a coupled-domain system in which only a relatively small number of state variables (e.g., less than 100) are allocated for each physical domain. Therefore, it is impractical (and in most cases, impossible) to plug a model of a big physical domain (e.g. the model with more than 1,000 states) into a system-level simulator to perform coupled-domain transient analyses.

Furthermore, the rank of the system (Eq. (5)) generated by the finite-element or the finite-different schemes is equal to the total number of the nodes with unknown pressure. For a device with complicated geometry, the total number of nodes might be as high as 10,000 – 100,000 (e.g., a plate with many perforations). Therefore, we apply a model order reduction (MOR) algorithm to generate a low order representation of Eq. (5) which still accurately captures the dynamic behavior. In Laplace domain, the transfer function of the system (Eq. (5)) can be written:

\[
T(s) = Q^T (I - s(c_i R)^{-1})^{-1}(c_i / c_i) b
\]

where

\[
b = -R^{-1} s
\]

After expanding the transfer function in Taylor series about \( s = 0 \), we obtain:

\[
T(s) = Q^T (I + s(c_i R)^{-1} + s^2(c_i R)^{-2} + ...) (c_i / c_i) b = \sum_{k=0} m_k s^k
\]

where \( m_k \) are the coefficients of the Taylor series:

\[
m_k = Q^T ((c_i R)^{-1}) (c_i / c_i) b = (c_i / c_i e_0) Q^T R^{-k} b
\]

It is possible to obtain a reduced-order model by truncating the Taylor expansion to approximate the original transfer function \( T(s) \). However, since \( R^{-k} b \) (see Eq. (8)) quickly line up with a single eigenvector as \( k \) increases, the procedure of computing the coefficients of the Taylor series is usually numerically unstable. Therefore, the Arnoldi-based algorithm [12] is applied to stably compute the first \( k \) orthogonal bases \( \tilde{v}_i \) that span the Krylov subspace:

\[
K_q(R^{-i} b) = \text{span}\{b, R^{-1} b, R^{-2} b, \ldots, R^{-(i-1)} b\}
\]

The procedure of generating the orthogonal bases \( \{\tilde{v}_i\} \) is in fact the Gram-Schmidt process with the initial basis \( \{b, R^{-1} b, R^{-2} b, \ldots, R^{-(k-1)} b\} \). Given the matrix \( V_q \) whose columns are \( \tilde{v}_i \), this algorithm can be used to generate a \( k \)-th order compact model (macromodel) by reducing the matrices \( R, S, \) and \( Q \) into smaller matrices \( R_q, S_q, \) and \( C_q \):

\[
V_q^T R V_q = R_q, \quad V_q^T S = S_q, \quad V_q^T Q = Q_q
\]

Finally, Eq. (5) can also be rewritten in the reduced form:

\[
\dot{\tilde{p}}_q = c_i R_q \tilde{p}_q + c_2 S_q u(t)
\]

\[
\tilde{y} = Q_q^T \tilde{p}_q
\]

where \( \tilde{p}_q \) is the reduced state vector. Note that the reduced system, as shown in Eq. (9), has the same input \( u(t) \) and output \( \tilde{y} \) as those in Eq. (5). Since the typical rank of the reduced system is much smaller than that of the original system, its computational efficiency for simulating transient responses and frequency responses is much higher. This reduced system is the so-called macromodel which can be used for system-level transient or frequency analyses. The attractive property of such an approach is that the first \( k \) Taylor series coefficients of the reduced model (Eq. (9)) match those of the original model (Eq. (5)) [0]. Therefore, potentially Eq. (9) is a good candidate of reduced order...
model of the original system. As will be seen in the next section, \( k = 5 \) is generally adequate for an accurate damping model. Finally, the method easily extends to a single model with multiple inputs \( \{ u_1(x), u_2(x), \ldots \} \) corresponding to multiple mode shapes, \( \{ f_1(x), f_2(x), \ldots \} \).

This low order model can be inserted directly into a system simulator such as SPICE or SABER. Note that since the vector space spanned by \( \{ v_i \} \) does not depend on the mean gap, ambient pressure or viscosity, the above model is valid for any choice of those parameters. Going one step further, we can model large signal behavior by letting the mean gap vary with time, \( h_0 = h_0(t) \). This approach would be valid if \( h_0(t) \) varies slowly compared to \( u(t) \) because the piecewise-linear approximation is applicable for this condition [21,22]. In fact, from numerical experiment, we find that replacing \( h_0 \) in Eq. (5) with \( h_0(t) = h_0 + u(t) \) for even large \( u(t) \) gives good results.

3. RESULTS AND DISCUSSIONS

Once we obtain a macromodel by the MOR algorithm, an HDL template of the gas model written in MAST is automatically generated. MAST is a hardware description language which is used by the system-level simulator Saber [17]. A system-level model of a coupled-domain MEMS device with this frequency-dependant gas-damping macromodel can be easily built using Saber, and small signal as well as transient analyses can be easily simulated.

3.1 Small Signal Oscillations

Three experimental data sets are used to verify the gas-damping models generated by the Arnoldi-based algorithm: a microrelay from IMT [10], an optical modulator from Lucent Technology [23] (shown in Fig. 2(b)), and a low-frequency accelerometer from Motorola (shown in Fig. 4) [9]. For small signal analysis, the system-level is constructed using the Saber (a system-level simulator). The mode-shape, the generalized spring constant and the generalized mass of the resonance mode of the device are obtained using the modal analysis module of a typical solid mechanics simulator (e.g., ABAQUS). Note that in this work, only the first-resonance-mode damping is simulated and measured in the small signal analyses. As described in previous section, the effect of the mode shape (non-uniform displacement) will be considered by multiplying the mode shape function \( f(x) \) with the external input excitation \( e(t) \) when a reduced damping model is generated. In the Saber, the generalized mass of the first mode is modeled as a lumped mass, and the spring constant of the first mode is modeled as a lumped spring constant. A typical Saber schematic of a system-level model for small signal analysis is shown in Fig. 5.

The dimensions of the microrelay and can be found in [7,10]. However, the detailed designs of the optical modulator and the low-frequency accelerometer are proprietary so that their dimensions cannot be released.

Figure 5 shows the frequency responses (small AC analysis by Saber) of an IMT microrelay for different ambient pressures. The IMT fixed-fixed-beam microrelay operates in its first oscillation mode. The fixed-fixed beam is 300\( \mu \)m in length, 30\( \mu \)m in width and 1.2\( \mu \)m in thickness. The gap between the beam and the substrate is 3\( \mu \)m. In the simulations, the effective Young’s modulus is 155GPa, and the compressive residual stress is 130MPa. The changes in quality factors as well as resonance shifting effects can be easily observed in Fig. 6 as the ambient pressure varies.

Figure 7 shows the gas damping coefficients vs. frequency for these three devices. The gas spring constant vs. frequency is shown in Fig. 8. The damping coefficients are obtained by considering the pressure force component that is in phase with the velocity of the structure (the damping force component). The spring constant, on the other hand, is obtained by considering the component that is in phase with the displacement of the motion (the spring force component). The simulated results in these figures are generated by macromodels of order 5. Note that the experimental and simulated results are in good agreement spanning about 6 orders of magnitude in frequency.

The comparison of computational times and the discrepancy for calculating the frequency-dependent damping the spring components (e.g., Fig. 1) is shown in Table 1. In the previous approaches [0, 0, 0], at least 20 transient simulations with different small-amplitude-oscillation frequencies were needed to generate the damping and spring components spanning across a desired frequency range. Our results show this new approach is at least 100 times more efficient than previous approaches. Note that the discrepancy is the average error between the total damping forces calculated by previous approach and by the reduced-order damping models of this work for 20 different frequencies.

This new approach also automatically builds time-domain macromodels for system-level analysis. Note that the computational time for the macromodels is the required time to generate each macromodel (the model generation time). It is in fact a “one time cost”. Once a macromodel is generated from the full FEM/FDM
3D Solid Model Creation

Fig. 3 Procedure of the efficient and accurate air damping modeling described in this paper

Fig. 4 A SEM of the Motorola accelerometer [9]

Fig. 5 The Saber schematic of a system-level model for small signal analysis

Fig. 6 Frequency response of an IMT microrelay for different pressures of 0.1, 1, 5, 10, and 15 mbar (curves from top to bottom). The quality factors and resonance shifting effects can be easily extracted from the curves.

Table 1 Comparison of computation times and the discrepancy between previous and new approaches

<table>
<thead>
<tr>
<th>Device</th>
<th>Previous Approach [9]</th>
<th>Reduce Order Models</th>
<th>Discrepancy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Computational time</td>
<td>Computation time</td>
<td>Order3: 3.1%</td>
</tr>
<tr>
<td>Lucent</td>
<td>948min</td>
<td>1.5min</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Order3: 2.3%</td>
</tr>
<tr>
<td>Motorola</td>
<td>431min</td>
<td>1.1min</td>
<td></td>
</tr>
<tr>
<td>IMT</td>
<td>200min</td>
<td>0.4min</td>
<td></td>
</tr>
</tbody>
</table>

Note that the discrepancy is the average error between the total damping forces calculated by previous approach and by the reduced-order damping models of this work for 20 different frequencies.
model, it can be used for further system-level simulations without incurring any model generation time. Since the order of generated macromodels are usually very small (e.g., 5–10), the typical computational time for simulating the macromodels is almost negligible (about a tenth of a second).

3.2 Transient Analysis

Figure 9 is an example of a Saber coupled-domain system-level model of a microswitch [24], which includes a lumped inertia, two mechanical springs, two automatically-generated frequency-dependent gas-damping models for two gaps, a electrostatic force actuator with a voltage source, and two mechanical stoppers. Figure 10 shows the experimental and simulated transient responses of the microswitch. The corresponding device parameters can be found in [24]. A step voltage of 10 volts is applied between the plate and substrate for 1ms. The plate collapses on the stopper, then is released from the substrate (after turning off the applied voltage) and oscillates around its neutral position. The simulated result follows the measured result reasonably well.

The large amplitude macromodel transient results are also compared with the full Reynolds simulations [5] that use the finite-difference method (FDM). Figures 11 and 12 show the transient results for a square plate suspended by tethers over a counter electrode with applying step voltages of 2 volts (below pull-in voltage) and 3 volts (above pull-in voltage). The ambient pressure is 100N/m², the gap is 2μm, and the plate is $500 \times 500 \mu m^2$. Below pull-in voltage, the macromodel and the FDM results match very well, as shown in Fig. 11. However, the results of the model with fixed $h_0$ [4] (see Eq. (2)) underestimates the damping as the plate moves away from its original gap.

When the applied voltage is larger than pull-in voltage, the plate velocity increases significantly when the plate is close to touch down, which in turn significantly increases the pressure back force against the plate. It is due to the fact that the total pressure force against the plate is approximately inverse-proportional to the gap between the plate and the substrate [3]. This large pressure increase breaks the small-amplitude Reynolds equation assumption that the pressure variation is much less than absolute ambient pressure. The macromodel result diverges from the FDM result after the plate travels across half gap, as shown in Fig. 12, due to this highly non-linear pressure force. The result of the model with fixed $h_0$ [4] starts to diverge from the FDM result after the plate travels about a quarter of the gap. Our preliminary simulated results show that the macromodel predicts the damping effect accurately as long as the external excitation force never drives the device beyond pull-in.

4. CONCLUSION

A new approach to extract frequency-dependent gas-damping dynamical models for MEMS devices is...
demonstrated. The Arnoldi-based algorithm is applied for creating a low-order model from the transient FE system matrices. The frequency-dependant gas damping and spring effects can be obtained using the low-order models without any computationally intensive transient simulation for a wide frequency range. A reduction of more than two orders of magnitude in computational time has been demonstrated. The numerical results of small signal gas damping and spring effects have shown in good agreement with experimental results for a wide range of operating frequencies. Examples of large-amplitude transient analysis and comparison to experimental results are also provided. Preliminary study shows that the generated macromodels can also provide accurate results for large-amplitude motions if pull-in does not occur.

ACKNOWLEDGEMENTS

Valuable discussions with Dr. Romanowicz and Dr. da Silva regarding to HDL are greatly appreciated.

NOMENCLATURE

\[ A \] system matrix formulated by FEM discretization
\[ \tilde{A} \] reduced system matrix \[ A \]
\[ B \] system matrix after FEM discretization
\[ \tilde{B} \] reduced system matrix \[ B \]
\[ e \] displacement of out-of-plane motion
\[ f \] normalized modeshape of out-of-plane oscillation
\[ F_D \] total damping force for a uniform displacement motion
\[ F_{Dq} \] total damping force for a specific oscillating mode
\[ h_0 \] initial gap between movable structure and the substrate
\[ h \] gap between movable structure and the substrate
\[ K_k \] Krylov subspace

REFERENCES


(Manuscript received September 29, 2004, accepted for publication January 21, 2005.)