Elastic PMLs for resonator anchor loss simulation

By

David S. Bindel

and

Sanjay Govindjee

February 2005

Department of Civil and Environmental Engineering
University of California, Berkeley
Elastic PMLs for resonator anchor loss simulation

David S. Bindel\textsuperscript{1} and Sanjay Govindjee\textsuperscript{2}\textsuperscript{*}

\textsuperscript{1} Department of Electrical Engineering and Computer Science; University of California Berkeley; Berkeley, CA 94720.
\textsuperscript{2} Structural Engineering, Mechanics, and Materials; Department of Civil Engineering; University of California Berkeley; Berkeley, CA 94720.

SUMMARY

Electromechanical resonators and filters, such as quartz, ceramic, and surface-acoustic wave devices, are important signal-processing elements in communication systems. Over the past decade, there has been substantial progress in developing new types of miniaturized electromechanical resonators using microfabrication processes. For these micro-resonators to be viable they must have high and predictable quality factors ($Q$). Depending on scale and geometry, the energy losses that lower $Q$ may come from material damping, thermoelastic damping, air damping, or radiation of elastic waves from an anchor. Of these factors, anchor losses are the least understood because such losses are due to a complex radiation phenomena in a semi-infinite elastic half space. Here, we describe how anchor losses can be accurately computed using an absorbing boundary based on a \textit{perfectly matched layer} (PML) which absorbs incoming waves over a wide frequency range for any non-zero angle of incidence. We show how to interpret the PML as a complex-valued change of coordinates, and illustrate how this interpretation leads to a simpler finite element implementation than was given in its original presentations. We also examine the convergence and accuracy of the method, and give guidelines for how to choose the parameters effectively. As an example application, we compute the anchor loss in a micro disk resonator and compare it to experimental data. Our analysis illustrates a surprising mode-mixing phenomenon which can substantially affect the quality of resonance.

KEY WORDS: Perfectly matched layer, Anchor loss, Resonator loss, High Q, Semi-infinite half space

1. Introduction

Modern communication systems rely on high-frequency electromechanical resonators to act as frequency references and filters. Though designers currently use quartz, ceramic, and surface-acoustic wave devices, surface-micromachined microelectromechanical system

\textsuperscript{*}Correspondence to: govindjee@ce.berkeley.edu

Contract/grant sponsor: National Science Foundation; contract/grant number: ECS-0426660

Contract/grant sponsor: University of California MICRO program

Contract/grant sponsor: Sun Microsystems
resonators (MEMS resonators) in development offer an attractive alternative. Because they can be integrated into standard complementary metal-oxide semiconductor (CMOS) technology, MEMS resonators have the potential to use less area and power, and cost less money than existing commercial devices [24]. But to be viable, energy losses in these MEMS resonators must be minimized. The usual measure of this energy loss is the quality factor $Q$ of a resonant peak, defined as

$$Q = 2\pi \left( \frac{\text{Stored energy}}{\text{Energy lost per period}} \right).$$

For an ideal linear single degree of freedom oscillator $Q = |\omega|/2 \text{Im}[\omega]$, where $\omega$ is the oscillator’s complex-valued eigenvalue [27, p. 158]. Resonators in cell phone filters, for example, require $Q$ values greater than 1000 for good performance, and higher values are preferable [24, 2].

Depending on scale, geometry, and materials, the energy losses that lower $Q$ may come from material damping, air damping, thermoelastic damping, or radiation of elastic waves from an anchor [12]. While losses in low-frequency resonators are dominated by air damping, for which increasingly accurate compact models are available [35, 7], high-frequency disk resonators have similar measured performance in vacuum or air [34]. Thermoelastic damping is a frequently-cited source of losses at high frequencies [17, 21, 1, 23, 29]. In most cases, the damping is estimated by fitting parameters in a model originally developed by Zener [36, 37, 38]; unfortunately, this parameter-fitting makes it difficult to tell what should be attributed to thermoelastic effects and what should be attributed to other sources of damping with similar functional form. Though anchor damping is a recognized source of losses [12], there are relatively few MEMS papers (see e.g. [28, 25, 26]) dealing with losses at the anchor.

Although it is not well studied, in several designs for high MHz or GHz frequency resonators, the dominant loss mechanism appears to be radiation of elastic energy through anchors. In these designs, the resonating device is much smaller than the silicon substrate on which it sits, and waves radiating from the anchor are so attenuated by the time they reflect from the sides of the microchip that the reflected waves are negligible. That is, the bulk of the chip can be modeled without loss as a semi-infinite half-space. To simulate the response of a semi-infinite domain, one usually employs boundary dampers, infinite elements, boundary integrals, or exact Dirichlet-to-Neumann (DtN) boundary conditions so that a domain of simulation can be finite and allow for the application of finite element or finite difference methods; see e.g. [40, Chapter 8], [18, 19, 5]. Each of these methods truncates the simulation domain with an artificial boundary at which outgoing waves are absorbed. For an elastic half space, Green’s function is not known in closed form, and so highly accurate global conditions, such as DtN conditions, cannot be used. Instead we model the semi-infinite domain using a perfectly matched layer (PML), which absorbs waves from any angle of incidence, but which does not require knowledge of Green’s function [8].

Basu and Chopra [8] demonstrated the superior performance of their PML method for problems related to earthquake engineering. Here, we examine the utility of a PML for anchor loss computations in MEMS resonators. We begin with a brief review of PMLs for time-harmonic motion. This is followed by a discussion of the relevant finite element expressions. Our presentation, while similar to that of [8], leads to a simpler implementation. We analyze the effects of discretization on the PML behavior, and give describe how to choose the PML parameters to obtain good accuracy. Because of the large computational scale of resonator problems we also investigate the use of reduced-order models that preserve the complex
symmetric structure of the PML equations.

To illustrate the PML technology, we analyze the behavior of a family of MEMS disk resonators. Several disk resonators have already been built [34, 33, 11], and in the laboratory they have shown quality factors as high as 55000 at frequencies of up to 1.14 GHz. Through numerical experiments, we explain in detail the mechanism of anchor loss in these devices; the predictions of our model are supported by recent experimental work [11], which we also partially report here.

2. Perfectly matched layers

Except for scale, a microresonator atop a silicon chip is much like a structure on the earth’s surface during an earthquake. While we are concerned with waves radiating away from a structure and the earthquake engineer is concerned with waves radiating toward a structure, in both cases the substrate is much larger than the structure, and it can be modeled as an elastic half-space (possibly heterogeneous). This infinite-domain approximation occurs in many physical models: acoustic waves radiating from a musical instrument, electromagnetic waves reflecting from aircraft, elastic waves scattering from a crack in a solid, and water waves in an open harbor are only a few additional examples [31], [40, Chapter 8]. The essential characteristic of the infinite-domain solution is that only outgoing waves are allowed. To model infinite-domain problems on a computer, we need finite-size discretizations which enforce this radiation condition.

One way to enforce the radiation condition is to discretize an exact boundary equation satisfied by outgoing waves. For example, outside of a sphere containing any radiators and scatterers, waves can be written as a multipole expansion; in this expansion, the radiation condition just says that certain coefficients corresponding to incoming waves should be zero. A related global condition is the DtN map, which specifies how Dirichlet conditions and Neumann conditions must be related at a surface [31]. These boundary conditions are rigorously derived and highly accurate, but they usually require that the artificial boundary have a particular shape. They are also nonlocal in space: every boundary unknown is directly related to every other boundary unknown, and consequently the matrix of boundary terms is dense, and expensive to form and to solve. Furthermore, exact boundary conditions may be unavailable for problems in which no analytically tractable Green’s function is known, as in our case.

A second approach is to build approximate boundary conditions based on the asymptotic behavior of outgoing waves. These approximate conditions are local and inexpensive, but only absorb waves over a small range of angles of incidence [31, 8]. Consequently, a large computational domain may be needed for accurate results. Further, they often have difficulty with surface waves and interface waves. Yet another approach is to add a nonphysical “sponge layer” to dissipate waves before they reach the artificial boundary. Waves passing through the sponge layer are damped on the way to the artificial boundary, and are further damped when they are reflected back, so that most of the signal entering the layer is absorbed. To be effective, though, the layer must be designed so that there is no impedance mismatch to reflect waves back from the interface between the layer and the rest of the domain.

A perfectly matched layer (PML) is a refinement of a sponge layer. Bérenger invented the perfectly matched layer for problems in electromagnetic wave propagation [9], and it was later re-interpreted as a complex-valued change of coordinates which could be applied to any
linear wave equation [13, 32, 30]. Not only do these layers rapidly attenuate waves, they also “perfectly match” the rest of the domain; that is, there are no spurious reflections at the interface due to perfect impedance matching. In [8], a perfectly matched layer for time-harmonic elastodynamics is described which – unlike previous elastodynamic PMLs such as those in [14] – can be implemented with finite elements in a standard displacement framework, with no non-standard global unknowns. We describe an alternate interpretation of the PML described in [8], and show how our interpretation further simplifies implementation in a finite element code.

2.1. A motivating example

A PML model of an infinite domain problem is composed of two parts: a sub-domain where the actual equation of interest is dealt with explicitly and, a sub-domain that produces the desired effect of a far-field radiation boundary condition. To set terminology and provide insight into the workings of PMLs, we review a simple 1-D example.

2.1.1. 1-D elastic wave 

Consider a longitudinal wave propagating in a homogeneous, semi-infinite rod with axial coordinate \( x \in [0, \infty) \). If waves travel with speed \( c \), the one-dimensional wave equation that describes this system is

\[
\frac{\partial^2 u}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2} = 0
\]

(2)

where \( u(x, t) \) is the displacement. Time-harmonic solutions \( u(x, t) = \hat{u}(x)e^{i\omega t} \) are governed by a Helmholtz equation

\[
\frac{d^2 \hat{u}}{dx^2} + k^2 \hat{u} = 0,
\]

(3)

where \( k = \omega/c \) is the wave number and \( i = \sqrt{-1} \). Solutions to this problem have the form

\[
\hat{u} = c_{\text{out}} e^{-ikx} + c_{\text{in}} e^{ikx}
\]

(4)

where \( c_{\text{out}} \) is the magnitude of the outgoing wave traveling from the origin toward infinity, and \( c_{\text{in}} \) is the magnitude of the incoming wave traveling from infinity toward the origin. In general, we assume there is no source at infinity, so physically meaningful solutions to such problems have \( c_{\text{in}} = 0 \).

2.1.2. 1-D elastic wave in a perfectly matched medium

We now consider the Helmholtz equation (3) under a change of coordinates. Let \( \lambda : \mathbb{R} \to \mathbb{C} \) be a continuous function which is nowhere zero, and define a new coordinate

\[
\tilde{x} = \int_0^x \lambda(s) \, ds.
\]

(5)

By definition, \( \tilde{x} \) and \( x \) are differentially related

\[
\frac{d\tilde{x}}{dx} = \lambda(x) \quad \frac{d}{d\tilde{x}} = \frac{1}{\lambda(x)} \frac{d}{dx}.
\]

(6)

Now suppose that the stretched coordinate \( \tilde{x} \) is used as the independent variable in equation (3). Then in terms of \( \tilde{x} \), the equation is

\[
\frac{1}{\lambda} \frac{d}{dx} \left( \frac{1}{\lambda} \frac{d\hat{u}}{dx} \right) + k^2 \hat{u} = 0.
\]

(7)
where the derivative may be taken in a weak sense, since $\lambda$ need not be $C^1$. Equation (7) describes wave propagation in a \textit{perfectly matched medium} (PMM).

Suppose
\[ \lambda(s) = 1 - i\sigma(s)/k; \tag{8} \]
then the solutions to the PMM equation (7) are
\[ \hat{u} = c_{\text{out}} \exp \left( - \int_0^x \sigma(s) \, ds \right) \exp(-ikx) + c_{\text{in}} \exp \left( \int_0^x \sigma(s) \, ds \right) \exp(ikx). \tag{9} \]

So long as $\sigma = 0$, both the incoming and outgoing solutions to the PMM equation (7) agree with the solutions to the original Helmholtz equation (3). Where $\sigma > 0$, the wave decays in the direction of travel. Since the outgoing wave and the incoming wave travel in opposite directions, the outgoing wave amplitude decays with increasing $x$, while the incoming wave amplitude decays with decreasing $x$. For example, assume $\sigma$ is defined to be zero on $[0, L]$ and $\sigma = \beta(s-L)$ on $[L, \infty)$. Then for $x > L$, the outgoing wave amplitude is $c_{\text{out}} \exp(-\beta(x-L)^2/2)$, and the incoming wave amplitude is $c_{\text{in}} \exp(\beta(x-L)^2/2)$.

Because waves decay so rapidly as they travel through the PMM region, we obtain a good approximation to the infinite-domain problem even if we force $\hat{u}(L_p) = 0$ for some finite $L_p > L$. This generates the concept of a \textit{perfectly matched layer} (PML); i.e. a PML is a finite PMM attached to a region with regular wave behavior. For example, suppose we prescribe $\hat{u}(0) = 1$ and $\hat{u}(L_p) = 0$. For convenience, define $\gamma = \beta(L_p - L)^2$; then the boundary conditions become
\[ \begin{bmatrix} \hat{u}(0) \\ \hat{u}(L_p) \end{bmatrix} = \begin{bmatrix} 1 \\ e^{-\gamma/2 + ikL_p} \end{bmatrix} \begin{bmatrix} c_{\text{out}} \\ c_{\text{in}} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{10} \]
and therefore
\[ c_{\text{out}} = \frac{1}{1 - e^{-\gamma - 2ikL_p}} = 1 + O(e^{-\gamma}) \quad c_{\text{in}} = \frac{-e^{-\gamma - 2ikL_p}}{1 - e^{-\gamma - 2ikL_p}} = -O(e^{-\gamma}). \tag{11} \]

Even for modest $\gamma$, the bounded-domain solution is a good approximation to the infinite domain solution. For $\gamma \approx 4.6$, only 1% of the outgoing wave is reflected. Increasing $\gamma$ decreases the reflection in the continuous case; however, in the discrete equations obtained from finite difference or finite element approximations, we must be careful about how we increase $\gamma$. If $\beta$ is too large, the waves entering the PML will decay rapidly, effectively creating a boundary layer; if the discretization is too coarse to resolve this decay, the numerical solution will be polluted by spurious reflections. We discuss this phenomenon and its implications further in Section 2.5.

UCB/SEMM-2005/01
2.2. Elastic perfectly matched layers

The multi-dimensional equations of motion for a time-harmonic elastodynamic medium with no body forces are

\[
\omega^2 \rho u + \nabla \cdot \sigma = 0 \\
\sigma = C : \epsilon \\
\epsilon(u) = \left( \frac{\partial u}{\partial x} \right)^s
\]

where \( u \) is the displacement field, \( \epsilon \) is the infinitesimal strain tensor, \( \sigma \) is the stress tensor, \( C \) is the material stiffness tensor, and \( \rho \) is the density. A simple isotropic elastic medium admits propagating disturbances moving at two characteristic velocities – compression waves (\( P \) waves) and shear waves (\( S \) waves). An anisotropic medium admits further characteristic wave speeds, and inhomogeneities and interfaces add yet more wave types. However, as in the one-dimensional case, a complex-valued coordinate transformation can be used to attenuate each of these waves in the direction of travel without spurious reflections from artificial interfaces.

2.2.1. Multi-dimension PMM equations

Though it is possible to introduce the coordinate transformation into the local form of the equations [8], it is simpler to first recast the equations in weak form and then transform. The weak form of the time-harmonic elastodynamic equation is

\[
\int \epsilon(w) : \sigma(u) \, d\Omega - \omega^2 \int \rho w \cdot u \, d\Omega = \int_{\Gamma} w \cdot t \, d\Gamma
\]

where the domain is \( \Omega \), part of the boundary \( \Gamma \subset \partial \Omega \) is subject to tractions \( t \), and \( w \) is a weight function. As before, suppose \( \tilde{x} \) is a transformed coordinate such that the Jacobian \( \frac{\partial \tilde{x}}{\partial x} = \Lambda \) is continuously defined and everywhere nonsingular. Replacing \( x \) with \( \tilde{x} \) everywhere in (15), we have

\[
\int_{\tilde{\Omega}} \tilde{\epsilon}(w) : \tilde{\sigma}(u) \, d\tilde{\Omega} - \omega^2 \int \rho w \cdot u \, d\tilde{\Omega} = \int_{\tilde{\Gamma}} w \cdot \tilde{\sigma}(u) \cdot \tilde{n} \, d\tilde{\Gamma}
\]

We now map back to the \( x \) coordinate system:

\[
\int_{\Omega} \epsilon(w) : \sigma(u) \, d\Omega - \omega^2 \int_{\Omega} \rho w \cdot u \, d\Omega = \int_{\Gamma} w \cdot \sigma(u) \cdot (\Lambda^{-T} n) \, d\Gamma.
\]

In the \( x \) coordinate system, the transformed strain and stress tensors are

\[
\tilde{\epsilon}(u) = \left( \frac{\partial u}{\partial \tilde{x}} \right)^s = \left( \frac{\partial u}{\partial x} \Lambda^{-1} \right)^s \\
\tilde{\sigma}(u) = C : \tilde{\epsilon}(u).
\]

The local form of (17), which can be derived either from (17) or directly from transforming (12), is

\[
\text{trace} \left( \frac{\partial \tilde{\sigma}(u)}{\partial x} \Lambda^{-1} \right) + \omega^2 \rho u = 0
\]

or, in indicial form,

\[
\frac{\partial \tilde{\sigma}_{ij}}{\partial x_k} (\Lambda^{-1})_{kj} + \omega^2 \rho u_i = 0.
\]

UCB/SEMM-2005/01
2.3. Anisotropic medium interpretation

We now present a different way to look at the PML equations, in which the original form of the elasticity equations is maintained, but with different material coefficients. This leads to a succinct and intrinsically symmetric implementation for multi-dimensional elastic PMLs.

In indicial form, we write the strain associated with a displacement field \( u \) as

\[
\varepsilon_{ij}(u) = \frac{1}{2} \left( \delta_{ip} \delta_{jq} + \delta_{iq} \delta_{jp} \right) \frac{\partial u_p}{\partial x_q}.
\]  

(22)

The PML-transformed strain has the same form, except with one of the Kronecker \( \delta \) functions replaced by \(~\delta\):

\[
\tilde{\varepsilon}_{ij}(u) = \frac{1}{2} \left( \delta_{ip} \delta_{jr} + \delta_{ir} \delta_{jp} \right) \frac{\partial u_p}{\partial x_q} \left( \Lambda^{-1} \right)_{qr} \]

(23)

\[
\tilde{\varepsilon}_{ij}(w) \tilde{C}_{ijkl} \tilde{\varepsilon}_{kl}(u) = \frac{\partial w_p}{\partial x_q} \tilde{\varepsilon}_{ijpq} \tilde{C}_{ijkl} \tilde{\varepsilon}_{klrs} \frac{\partial u_r}{\partial x_s} \]

(25)

\[
\frac{\partial w_p}{\partial x_q} \tilde{\varepsilon}_{pqrs} \frac{\partial u_r}{\partial x_s} \]

(26)

where we define

\[
\tilde{\varepsilon}_{pqrs} := \tilde{\varepsilon}_{ijpq} \tilde{C}_{ijkl} \tilde{\varepsilon}_{klrs}.
\]

(27)

Note that \( \tilde{\varepsilon}_{pqrs} \) inherits the major and minor symmetries of \( C_{ijkl} \). That is,

\[
C_{ijkl} = C_{klij} \quad \Rightarrow \quad \tilde{\varepsilon}_{pqrs} = \tilde{\varepsilon}_{rspq}
\]

(28)

\[
C_{ijkl} = C_{jikl} \quad \Rightarrow \quad \tilde{\varepsilon}_{pqrs} = \tilde{\varepsilon}_{qprs}.
\]

(29)

Because of the minor symmetries (\( \tilde{\varepsilon}_{pqrs} = \tilde{\varepsilon}_{qprs} \) and \( \tilde{\varepsilon}_{pqrs} = \tilde{\varepsilon}_{rpqs} \)), we can rewrite (26) as

\[
\tilde{\varepsilon}_{pq}(w) C_{pqrs} \tilde{\varepsilon}_{rs}(u) = \frac{\partial w_p}{\partial x_q} \tilde{\varepsilon}_{pqrs} \frac{\partial u_r}{\partial x_s} = \varepsilon_{pq}(w) \tilde{\varepsilon}_{pqrs} \varepsilon_{rs}(u).
\]

(30)

If we substitute (30) into the weak form of the PML equation (17), and assume that there is no loading on the transformed part of the boundary, we have

\[
\int_{\Omega} \varepsilon(w) : \tilde{\varepsilon}(u) - w^2 \int_{\Omega} \rho w \cdot u \cdot \det(\Lambda) d\Omega = \int_{\Gamma} w \cdot t d\Gamma.
\]

(31)

Now define

\[
C_{PML} = \tilde{\varepsilon} \det(\Lambda)
\]

(32)

\[
\rho_{PML} = \rho \det(\Lambda).
\]

(33)

so that (31) becomes

\[
\int_{\Omega} \varepsilon(w) : C_{PML} \varepsilon(u) d\Omega - w^2 \int_{\Omega} \rho_{PML} w \cdot u d\Omega = \int_{\Gamma} w \cdot t d\Gamma.
\]

(34)

The form of (34) is identical to the form of the standard elasticity equation (15), but with inhomogeneous, anisotropic, complex-valued material properties.
2.4. Finite element implementation

To derive the weak form of the PML equations in the $x$ coordinate system, Equation (17), we performed a change of variables in the integrals of Equation (16). Because isoparametric finite elements already use mapped integration, we can combine the change of variables associated with the PML mapping with the change of variables associated with the isoparametric coordinate transformation.

Consider the element in Figure 2. Suppose we choose shape functions $N_I$, so that we have interpolations within elements of the form $u = \sum_I N_I u_I$ and $w = \sum_I N_I w_I$. Then the nodal submatrices for the element stiffness and mass are given by

$$k_{IJ}^e = \int_{\Omega^e} \tilde{B}_I^T D \tilde{B}_J \tilde{J} d\Omega^\Omega$$

$$m_{IJ}^e = \left( \int_{\Omega^e} \rho \tilde{N}_I^T N_J \tilde{J} d\Omega^\Omega \right) 1$$

where $1$ is the second order identity tensor and the nodal matrices $\tilde{B}_I$ come from transforming coordinates in the standard $B$-matrix formulation [39, Chapter 4], $D$ is the standard matrix of material parameters, and $\tilde{J}$ is the Jacobian of the composition of the PML mapping with the isoparametric mapping:

$$\tilde{J} = \det \left( \frac{\partial x}{\partial \xi} \right) \det (\Lambda).$$

In practice, we evaluate the integrals numerically by Gaussian quadrature in the parent domain. Whether the quadrature is done analytically or numerically, the form of the integrands in (35) and (36) guarantees that the mass and stiffness matrices will be complex symmetric.

Remarks:

1. This interpretation of the PML in terms of an additional coordinate transformation works with plane stress, plane strain, axisymmetric, or three-dimensional problems. In the axisymmetric case, however, the factor of $r$ that appears in the integrands should not be transformed into the PML coordinate systems, since that factor of $r$ comes from the Jacobian of the mapping to the $(r, z)$ coordinates, and not from the mapping to the $(\tilde{r}, \tilde{z})$ coordinates.

2. For many problems, a reasonable choice of coordinate transformations is to independently stretch each coordinate $x_i$, so that $\Lambda$ is a diagonal matrix; i.e. $\Lambda = \text{diag}(\lambda_i)$. If we further choose stretching functions so that $\Lambda$ can be described by low-order polynomials, then
it makes sense to also use isoparametric interpolation to compute the values of the stretching function. That is, given values for $\lambda_i$ at each node, we compute $\Lambda = \text{diag}(\lambda_i)$ by interpolation at the Gauss points where it is evaluated.

3. By writing the PML equations in this form we can easily institute an economy of programming where every element in a mesh is a “PML element.” Regular elements are formed using the coordinate transformation $\Lambda = 1$ and true PML elements by $\Lambda = \text{diag}(\lambda_i)$. Thus the creation of PML elements only requires a minor modification of the traditional element mapped integration routines.

2.5. Effects of discretization

In Section 2.1.1, we analyzed a one-dimensional continuous model problem with a PML. We now consider the effect of discretization on this model problem. We will show that the reflection in the discrete solution is the result of two effects: reflection due to the finite termination of the PML, as discussed in the continuum case; and spurious reflection from the PML interface due to a mismatch in the discretized wave behaviors in the bounded domain and in the PML.

Recall that if we discretize a standard wave equation on $(-\infty, \infty)$ with a uniform mesh, using any standard finite element or finite difference method, we obtain left-traveling (incoming) and right-traveling (outgoing) discrete wave solutions at node $j$ of the form $\xi_j$ and $\xi_j^\prime$. The value $\xi$ is an algebraic function of the dimensionless parameter $kh$, and $\xi(kh) \approx e^{kh}$ when $kh$ is small. Because $\xi$ is algebraic, it cannot match $e^{kh}$ exactly, and consequently there is always numerical dispersion; that is, for a fixed $h$ the discrete wave speed depends on the wave number $k$. Similarly, if the mesh size $h$ is not uniform, there will be numerical reflections at interfaces where the mesh density—and consequently the discrete wave speed—changes. These spurious effects come from the error implicit in the discretization, and they vanish in the limit as $kh \to 0$.

Now consider the discretization of the model problem using the domain and the damping profile illustrated in Figure 1 with a uniform mesh size in $x$. We will assume throughout our analysis that the PML begins at an element boundary. While the mesh size is uniform in the original coordinate system $x$, the elements are not uniformly sized with respect to the transformed coordinate $\tilde{x}$; it is unsurprising, then, that there is a mismatch in the discrete wave behavior at the PML interface, and that this mismatch grows more severe as $\beta h$ increases.

To be more precise, we will consider the discrete reflection coefficient $r$ (the ratio of the magnitude of the incoming to the outgoing component of the solution in the bounded domain) for various values of the PML parameters. For fixed values of $\beta h$ and $kh$, the reflection coefficient rapidly approaches a limiting value $r_{\text{interface}}$ as the length of the PML increases. That is, reflections due to the finite length of the PML become negligible in comparison to reflections at the discrete interface between the bounded domain and the PML. In Figure 3, we show the effects of interface reflections in the PML for linear and quadratic elements. We see from these plots that fine discretizations (either from smaller values of $kh$ or from higher-order elements) and low values of $\beta$ lead to decreased interface reflections. We also note that quadratic elements perform substantially better than linear elements.

The overall reflection in the discrete system can be estimated effectively as $r = r_{\text{interface}} + r_{\text{nominal}}$, where $r_{\text{nominal}}$ is the far-end reflection in the continuum model. When both effects are of the same order of magnitude, cancellation can occur so that the estimated reflection coefficient is slightly too high; otherwise, this prediction closely matches the observed behavior.

UCB/SEMM-2005/01
Figure 3. Amount of spurious reflection from a discrete PML interface for varying mesh densities and PML parameters. Both linear (left) and quadratic (right) elements are shown.

Figure 4. Actual (left) and estimated (right) numerical reflection for varying PML parameters at ten quadratic elements per wave.

Figure 4 shows the estimated and actual reflection coefficients for a fixed value of $kh$ as the PML parameter $\beta$ and the PML length are varied. The observed decomposition of the PML reflections into interface and far-end effects suggests that the parameters should be chosen so that the two sources of error are comparable.

This model problem suggests a strategy for choosing PML parameters in the more complicated case of two-dimensional or three-dimensional elasticity. Because the discrete elastic wave equation is dispersive, and because the mesh density observed by a discrete wave is a function of the wave direction, there will be a range of possible numerical wave numbers $kh$. There is less numerical interface reflection for long wavelengths than for short wavelengths, but long waves also do not decay as rapidly as they pass through the PML. Therefore, we recommend that $\beta h$ be chosen so that the shortest expected wavelengths have sufficiently low reflection from the discrete PML interface; and then that the PML length be chosen so that the longest expected wavelengths can be expected to decay adequately according to the continuum
model.

3. Quality factors and forced motion computations

In the MEMS problems of interest to us, we wish to compute quality factors and to compute forced motion responses. The governing equations after spatial discretization of the weak form are given by:

$$K_{\text{dyn}}(\omega)u = F,$$

where $K_{\text{dyn}}(\omega) := K(\omega) - \omega^2 M(\omega)$. As we noted at the end of Section 2.2, the PML coordinate transformations suggested in [8] are dependent on the frequency, so that the attenuation through the PML layer will be independent of the forcing frequency. This makes the system matrices dependent upon the drive frequency. Also note that the system matrices are complex symmetric. The points taken together create a somewhat involved problem. However, a number of basic observations can be used to greatly simplify the situation.

3.1. Quality factors via an eigencomputation

In a properly designed high quality MEMS resonator the drive pattern and frequency are always chosen to excite some resonant mode. Because of this, it suffices for many cases to simply compute the complex-valued eigenvalue of the system closest to the real-valued drive frequency. From this complex-valued eigenvalue, $Q$ is formally defined according to Equation (1) as the ratio of the stored energy to the energy loss per radian, which in terms of a single mode damped oscillator can be expressed as

$$Q = \frac{\omega}{2 \text{Im}(\omega)}$$

where $\omega$ is the computed eigenvalue [27].

Because the system matrices depend upon $\omega$, the eigenvalues $\omega$ will correspond to solutions of the nonlinear eigenvalue problem $\det(K_{\text{dyn}}(\omega)) = 0$. However, when the frequency range of interest is not too wide, the parameters of the coordinate transformation may be chosen once to give acceptable attenuation over the desired range, so that the approximate dynamic stiffness for $\omega$ near a fixed reference frequency $\omega_0$ is

$$K_{\text{dyn}}^0(\omega) := K(\omega_0) - \omega^2 M(\omega_0).$$

Finding roots such that $\det(K_{\text{dyn}}^0(\omega)) = 0$ is a linear (generalized) eigenvalue problem, which we can approach using standard tools.

To find the damped eigenvalues near some specified (real-valued) reference frequency $\omega_0$, we use a shift-and-invert Arnoldi procedure, described in standard references on numerical linear algebra [20, Chapter 9], [16, Chapter 7]. This procedure computes an orthonormal basis $V$ for the Krylov subspace

$$V_n(K_{\text{dyn}}(\omega_0), u_0) = \text{span}\{u_0, K_{\text{dyn}}^0(\omega_0)^{-1}u_0, \ldots, K_{\text{dyn}}^0(\omega_0)^{-(n-1)}u_0\};$$

the eigenvalues are then approximated by the eigenvalues of the much smaller problem $V^H K_{\text{dyn}}^0(\omega)V$, where $(\cdot)^H$ indicates complex conjugate transpose. For computing a few
isolated eigenvalues near $\omega_0$, the main cost of the shift-and-invert Arnoldi procedure is to compute the factorization needed to apply $K_{\text{dyn}}^0(\omega_0)^{-1}$. In our numerical experiments, we use UMFPACK [15] to factor the shifted matrix, and we use eigs, MATLAB’s interface to the implicitly-restarted Arnoldi code ARPACK [22], to compute the desired eigenvalues.

3.2. Efficient forced motion computations

Though damped mode eigencomputations are illuminating, they do not give a complete picture of the frequency response behavior. In practice, we are interested in systems in which there is a single periodically-forced input and a single output determined by some sensed displacement. Suppose $F$ is a time-harmonic load pattern vector, and the output is a linear function of displacement $P^T u$. Then we are really interested in computing the transfer function

$$H(\omega) = P^T K_{\text{dyn}}(\omega)^{-1} F$$

which we approximate for a range of frequencies near $\omega_0$ by

$$H^0(\omega) = P^T K_{\text{dyn}}^0(\omega)^{-1} F.$$ (43)

Even if $K_{\text{dyn}}^0(\omega)$ is nearly singular at some frequency, the response amplitude $|H^0(\omega)|$ may not peak, since $F$ may be nearly orthogonal to the forcing that drives the mode, or $P$ may be orthogonal to the modal displacement pattern. Therefore, one normally examines $H^0(\omega)$ directly using a Bode plot.

Since the dimension $N$ of $K_{\text{dyn}}(\omega)$ will be large, it is expensive to evaluate $H^0(\omega)$ directly. Instead, we construct a reduced-order model of dimension $n \ll N$, which we use to approximate $H^0(\omega)$. Krylov-subspace projections are often used to build reduced models of large systems [6, 3]. If we build an orthogonal basis $V$ for a small Krylov subspace using shift-and-invert Arnoldi, then we can approximate $H^0(\omega)$ near the shift $\omega_0$ by

$$\hat{H}^0(\omega) := (V^H P)^H (V^H K_{\text{dyn}}^0(\omega) V)^{-1} (V^H F) \approx H^0(\omega).$$ (44)

Though $\hat{H}^0(\omega)$ is often a good approximation to $H^0(\omega)$, the projected system matrix $V^H K_{\text{dyn}}^0(\omega) V$ does not preserve the complex symmetric structure of the original discretization. We can construct a symmetry-preserving reduced-order model by choosing an orthonormal projection basis $W$ such that

$$\text{span}(W) = \text{span}([\text{Re}(V), \text{Im}(V)]).$$ (45)

Because the span of $W$ contains the span of $V$, a reduced model based on $W$ will be at least as accurate as the standard Arnoldi-based reduced model. Also, because $W$ is a real-valued basis, projection onto the space spanned by $W$ corresponds to a Bubnov-Galerkin discretization of the PML equation with shape functions $N_{\text{red}}^J \sum J W_{IJ} N_{IJ}$. While these facts alone might induce us to use $W$ rather than $V$ as a projection basis [10], we expect projection onto $W$ to yield much better accuracy than standard Arnoldi projection, as we now describe.

If $(K, M)$ is a Hermitian pencil and $M$ is positive definite, then the pencil will have an orthonormal eigensystem, so that if $v$ is a column eigenvector, $v^H$ will be a corresponding row eigenvector. In the study of such eigenproblems, the Rayleigh quotient

$$\rho(v) = \frac{v^H K v}{v^H M v}$$ (46)
plays a special role. When \( v \) is an eigenvector, \( \rho(v) \) is a corresponding eigenvalue; and further, since \( \rho \) is stationary when and only when the argument is an eigenvector, the Rayleigh quotient produces second-order accurate eigenvalue estimates from eigenvector estimates which are only accurate to first order. For general pencils, a column eigenvector \( v \) and the corresponding row eigenvector \( w^H \) need have no such simple relationship, and so the Rayleigh quotient only provides first-order accurate eigenvalue estimates. The appropriate generalization of the Rayleigh quotient to the non-Hermitian case is \( (w^H K v) / (w^H M v) \), a ratio which again yields second order accuracy (so long as the degenerate case \( w^H M v \neq 0 \) is avoided). When \( K \) and \( M \) are complex symmetric, we know the left and right eigenvectors are simply (non-conjugated) transposes of each other, and so we re-write the second-order accurate quotient estimate as

\[
\theta(v) = \frac{v^T K v}{v^T M v}.
\]

This modified Rayleigh quotient was used in [4] as part of a Jacobi-Davidson strategy for solving complex symmetric eigenvalue problems from PML discretizations of problems in electromagnetics.

The usefulness of having both left and right eigenvectors explains why model reduction methods based on nonsymmetric Lanczos iteration often approximate better than Arnoldi methods: a nonsymmetric Lanczos iteration simultaneously builds a basis for a right Krylov subspace, which typically contains good approximations for column eigenvectors; and a left Krylov subspace, which typically contains good approximations for row eigenvectors. For complex symmetric matrices, however, left and right subspaces are simply conjugates of each other, and the definition of \( \text{span}(W) \) given above is equivalent to

\[
\text{span}(W) = \text{span}([V; \text{conj } V]).
\]

That is, if \( V \) is an Arnoldi basis for a right Krylov subspace, then both the right Krylov subspace and the corresponding left Krylov subspace are subspaces of \( \text{span}(W) \). As explained in the previous paragraph, then, the position of any eigenvalues (poles) which are estimated by vectors in \( V \) will be determined to second-order accuracy by projection onto \( W \), where projection onto \( V \) would typically attain only first-order accuracy. For similar reasons, if \( P \) and \( F \) are proportional to each other, the estimated transfer function obtained from projecting onto \( W \) will match \( H^0 \) in \( 2n \) moments, rather than the \( n \) moments typical of a standard Arnoldi projection [6].

4. Study of disk resonators

To exercise and test our proposed methods we examine in detail the response of several radial disk resonators which have been fabricated recently [34, 33, 11]. A schematic of these devices is shown in Figure 5. A thin disk is supported on a post above a substrate. The disk is surrounded by drive electrodes, and the potential difference between the disk and the drive electrodes pulls the disk radially outward at the rim. The disk is driven near the frequency of the first or second axisymmetric, bulk radial, in-plane mode. In [34], the disk is made of polysilicon; [33] describes both polysilicon and polydiamond disks; [11] is concerned with poly-SiGe disks. The choice of these disk resonators allows us to idealize the systems as axisymmetric.

UCB/SEMM-2005/01
A major source of energy loss in these devices is the propagation of elastic waves down the supporting post and into the wafer below, where they are largely dissipated. Relative to the size of the resonating device, the wafer is large. We assume that the wafer is effectively infinite in extent, so that none of the waves that radiate into the substrate will be reflected. We use a perfectly matched layer to model the wafer as a semi-infinite half space.

In the actual devices, there are nitride and oxide films between the device and the wafer. For our simulations we have ignored these geometric features as they typically only have a minor effect on resonator performance. Note that this is a simplification in our model, not an inherent limitation of the PML technology; indeed, one of the attractions of PMLs is the ability to handle layered media and other embedded scatterers.

In the examples to follow, we examine issues of

2. Elucidation of the physical mechanism of anchor loss.
3. Design sensitivity in such resonators.

4.1. Convergence of $Q$

We first consider the polysilicon disk resonator described in [34], which has a disk 20 $\mu$m in diameter and 2 $\mu$m thick, supported 0.5 $\mu$m above the substrate by a post 2 $\mu$m in diameter. When the disk was driven in the second radial mode, the measured $Q$ was 7330 in vacuum and 6100 in air, and the center frequency was 733 MHz. In our best-resolved simulation, we computed a $Q$ value of 6250 at a center frequency of 715.6 MHz. Perhaps surprisingly, relatively fine resolution was required to obtain convergence. When the mesh was under-resolved, the $Q$ factor was drastically underestimated, possibly because the small flux reaching the anchor base could not be resolved by the mesh, and was therefore overestimated. We computed the value of $Q$ using both linear and higher-order elements at several mesh densities; see Fig. 6. For a given mesh density parameter $m$, we chose elements so that nodes were as near as possible to $1/m$ $\mu$m apart. We computed $Q$ and $\omega_{\text{center}}$ by using shift-and-invert Arnoldi with an initial shift of 715 MHz to find the closest complex-valued eigenvalue; most of the time in these computations was spent in the LU factorization of the shifted matrix. The plot shows a clear advantage to $p$-refinement for this class of problems.

UCB/SEMM-2005/01
Figure 6. Convergence of $Q$ and $\omega_{\text{center}}$ for the second radial mode of the polysilicon disk in [34].
4.2. Observed energy loss mechanism

Figure 7 shows the behavior of the same disk when driven at 715 MHz, just slightly below the resonant frequency. The amplitudes of both the radial and vertical displacements are shown at the point when the forcing is maximal. Even though the design intent was to excite a pure radial mode, it is clear that the mode contains a bending component. This is a reflection of
the fact that pure radial modes are not possible in supported structures. The majority of the displacement occurs in the disk itself, but there is some motion in the post as well. Though the forcing is in the radial direction, the Poisson effect leads to motion in the vertical direction at the center of the disk. This vertical “pump” motion results in displacement waves that travel down the post and into the substrate. In an animation, it is possible to see low-amplitude waves radiating away from the post to be absorbed into the perfectly matched layer.

To better understand the behavior shown in Figure 7, we compute the energy density $F$
\begin{equation}
F(t) = -\text{Re}(\sigma e^{i\omega t}) \text{Re}(ve^{i\omega t})
\end{equation}
where $\sigma$ is the stress tensor and $v$ is the velocity vector. Since the energy flux changes over time, we time-average over a single period to obtain the mean energy density flux:
\begin{equation}
\bar{F} = -\frac{1}{2} \text{Re}(\sigma v^*)
\end{equation}
where $v^*$ is the complex conjugate of the velocity. For a standing wave, the displacement and stress are pure real, and $v = iku$ is pure imaginary, so there is no mean energy flux. Therefore, the mean energy flux field tells us something about the departure from the lossless standing-wave behavior. Figure 8 shows the mean energy flux for a region of the resonator near the edge of the post. The flux vectors in the body of the disk form cycles which carry energy around inside the disk, but do not let it escape into the substrate. Near the post, however, the cycle pattern is broken, and the energy flux plot shows a “spray” of energy that travels down the post and into the substrate.

4.3. Mode mixing and design sensitivity in the disk resonator

To further test our simulation technology on anchor loss, we now consider a series of five poly-Si$_{0.4}$Ge$_{0.6}$ disk resonators with 41.5 $\mu$m radii disks with different thicknesses [11]. Figure 9 shows one such 41.5 $\mu$m radius resonator. The disk itself is supported on a conical post with upper radius 1.49 $\mu$m, lower radius 1.61 $\mu$m, and nominal height 1 $\mu$m. The drive is clearly not fully axisymmetric but we model it as such for simplicity. For the material we use a density of 4127 kg/m$^3$ computed by linear interpolation and assume a Poisson ratio of 0.28; Young’s modulus was estimated from an acoustic measurement as 139 GPa.

Figure 10 shows the computed in-phase radial and vertical displacements in one of the disks when it is driven with a radial forcing on its outer edge; the computed $Q$ of the dominant mode is 140000. The radial motion is coupled to a small bending motion due to the stem as mentioned earlier. Thus, as in the prior example, the dominant mode for this disk is not a pure radial motion. The bending motion of the mode along with the Poisson effect induces a vertical motion in the stem which pumps displacement waves into the substrate, where they carry away the energy of the resonance.

Figure 11 shows measured $Q$ values from the five 41.5 $\mu$m disks. The error bars on thickness are indicative of the limits of the SEM geometry measurement method. Also shown in Figure 11 is simulation data using the measured geometry from the resonators. The two curves correspond to $Q$ values for the two eigenvalues nearest the shift (45 MHz). The high $Q$ curve is associated with the radial extension mode we wish to drive; the low $Q$ mode is dominated by a bending motion. The agreement between the measured and computed $Q$ values is good, and the trend with respect to changing disk thickness is captured well.

UCB/SEMM-2005/01
Figure 9. SEM of 41.5 μm radius poly-SiGe disk resonator.

Figure 10. Radial and vertical displacement fields illustrating the mode mixing. Disk radius is 41.5 μm and film thickness is 1.6 μm. The drive frequency is 45 MHz and $Q = 140000$. Displacement contours are in units of μm.
The presence of the nearby second mode has a large influence on the high $Q$ mode’s quality factor. This is seen in the large swings in the curves of Fig. 11 which are computed solely from the system eigenvalues. A good method of visualizing the pole interaction is to examine a root-locus diagram for the two interacting poles (eigenvalues) parameterized by film thickness. Figure 12 was computed for the 41.5 μm radius disks. As the thickness changes the two poles approach each other. The first mode’s frequency first moves away from the real axis increasing damping, then back toward the real axis decreasing damping. The speed of the first pole increases the closer it is to the second pole in the complex plane. The dip in the resonator’s $Q$ correlates well with the thickness at which the poles are closest.

4.4. Performance of model reduction method

As seen above, the bending-dominated mode significantly affects the resonant peaks associated with the radial dominated modes but not in a way that is immediately obvious. For this reason, Bode plots are helpful in understanding such systems. As an example of our model reduction technology, we start with a shift drawn from a hand estimate for the disk frequency:

$$\omega_{\text{shift}} = 2.405c/R,$$

where $R$ is the disk radius and $c$ is the compression-wave velocity in the disk material. With this shift, we require only two steps of shift-and-invert Arnoldi to resolve the two-dimensional
Figure 12. Root-locus plot parameterized by varying film thickness. Lower curve corresponds to the first mode and the upper curve to the next nearest mode.

invariant subspace for both relevant eigenvalues. We perform the projections in two ways: first, using a standard Arnoldi projection (with three complex vectors); and second, using a symmetry-preserving projection based on splitting the real and imaginary parts of the Arnoldi vectors (with five real vectors). Both reduced models produce Bode plots which closely match the original 24265 degree of freedom system (Figure 13). In Figure 14, we compare the accuracy of the two models; clearly, the structure-preserving algorithm is more accurate.

4.5. Conclusions

In this paper, we have described, developed, and enhanced tools suitable for the simulation of quality factors in very high frequency MEMS resonators. The simulation of this regime of physical behavior, to date, has been largely ignored due to the analytical and numerical difficulty of estimating anchor losses which dominate such systems. The primary numerical advances made are as follows:

1. We have described an alternate interpretation of a PML for time-harmonic elasticity which was introduced in [8] and in doing so have shown how to construct a particularly simple finite element implementation for all relevant classes of analysis.

2. Through the use of one-dimensional analysis, we have elucidated the effect of discretization on the perfect matching property in order to derive heuristics for choosing the PML parameters.

UCB/SEMM-2005/01
3. We have performed convergence studies which show the care which must be taken to resolve eigenvalues with small imaginary parts compared to their magnitudes, and we have highlighted the advantage of $p$-refinement in resolving such eigenvalues.

4. By exploiting the complex symmetry inherent in the PML equations, we have also described how to improve the accuracy of standard methods for computing free vibrations and for building reduced models for forced frequency-response analysis.

On the engineering side, our motivation for studying PMLs was to investigate anchor loss in MEMS resonators. We have demonstrated the utility of our method by analyzing the behavior of a family of disk resonators. Our developed tools allow us, in detail, to describe the physical mechanism by which these resonators lose energy by radiation of elastic waves from the anchor. Because MEMS fabrication processes are not exact, we have analyzed the effect of variations in the thickness of the resonating disks. Our analysis showed that the quality factor of these resonators is highly sensitive to the film thickness, and experimental results confirmed this analysis. Our new tools allow us to explain the variations in the quality factor in terms of interactions between the desired radial extension mode and a parasitic bending mode which resonates at nearly the same frequency. The sensitivity to film thickness was first discovered numerically, and later was verified experimentally.

ACKNOWLEDGEMENTS

UCB/SEMM-2005/01
Figure 14. Errors in a structure-preserving reduced model (+) and a standard Arnoldi reduced model (*). Both reduced models are generated from the same number of Krylov vectors.

The authors wish to thank Prof. Roger Howe, Dr. Emmanuel Quévy, Dr. Jong Woo Shin, Dr. Sunil Bhave, Prof. James Demmel, and Dr. Ushnish Basu for their interest and assistance in this work.

REFERENCES


