Efficient Evaluation of Damping in Resonant MEMS

by

Tsuyoshi Koyama

B.Eng. (University of Tokyo, Japan) 2001
M.Eng. (University of Tokyo, Japan) 2003

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Engineering-Civil and Environmental Engineering

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Sanjay Govindjee, Chair
Professor James W. Demmel
Professor Robert L. Taylor

Spring 2008

The dissertation of Tsuyoshi Koyama is approved:

Chair Date

Date

Date

University of California, Berkeley

Spring 2008

Efficient Evaluation of Damping in Resonant MEMS

Tsuyoshi Koyama

Abstract

Efficient Evaluation of Damping in Resonant MEMS

by

Tsuyoshi Koyama

Doctor of Philosophy in Engineering-Civil and Environmental Engineering

University of California, Berkeley

Professor Sanjay Govindjee, Chair

This dissertation is about numerical methods for efficiently simulating damping behavior in Microelectromechanical Systems (MEMS). Within the class of MEMS devices, focus is put on the simulation of high-frequency mechanical resonators which have potential applications as on-chip, high-performance, low-power signal-processing elements such as filters or oscillators in radio-frequency (RF) wireless technology. The performance of these devices are defined by the quality factor Q, which is defined as the maximum stored energy divided by the energy dissipation per radian of oscillation. High Q values are desired, but can be limited by energy dissipation mechanisms such as anchor loss, thermoelastic damping, air damping, material losses, and ohmic loss. The design process of these MEMS resonators can be accelerated significantly by accurate and efficient numerical simulations which can predict the amount of damping in the system, and ultimately Q. In this work, numerical methods to efficiently evaluate Q for systems with anchor loss and thermoelastic damping are developed. Anchor loss contributions to Q are computed from the solution of a complex-symmetric eigenvalue problem through a Jacobi-Davidson QZ eigensolver in combination with a scalable (to millions of unknowns) geometric multigrid we have developed specifically for this type of problem. Thermoelastic contributions to Q are simulated by efficient evaluation of transfer functions by a

second-order Krylov subspace based structure preserving reduced order model. The MEMS resonators introduced here are components of an electrical circuit and actuated electrostatically or piezoelectrically, making them electromechanically coupled systems. Efficient transfer function evaluation through the extraction of equivalent circuit parameters based on a variational framework extendable to various resonator geometries is also presented. The numerical simulations of a class of disk resonators reveal contradictory results to experimental claims regarding decreases in Q with respect to post misalignment. The simulations only incorporate purely mechanical effects and ideal geometry, which implies that sources such as electromechanical coupling effects and geometrical variations are responsible for this Q sensitivity.

Professor Sanjay Govindjee
Dissertation Committee Chair

To my parents

and

Junko

Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank my adviser Prof. Sanjay Govindjee for the support that he has given me throughout my Ph.D. I am grateful for the countless advice he has given me and the patience he has taken upon in my research. Had it not been for the CE231 Mechanics of Solids lecture that I had taken with him, I do not think I would have selected the area of computational mechanics as my field of research, from which I have gained so much. I still remember the time I had visited his office in search for a Ph.D. research project, and his mentioning of the SUGAR project related to the simulation of RF-MEMS devices which I have taken upon. Through this project I have had the fortune of being able to interact with people from different areas as well as broaden my knowledge and understanding of various topics. I would also like to thank him for giving me the opportunity to conduct research alongside him overseas at ETH, Switzerland.

I would like to thank Prof. James Demmel for the advice and insight he has given me through the SUGAR group and Math 221 Advanced Matrix Computations lecture which has introduced me to the fascinating area of numerical linear algebra which has become an indispensable part of my research and interests. I would like to thank Prof. Robert Taylor for shaping the final pieces of my research with his deep insight in the area of finite element analysis. I would like to thank Prof. Alexander Givental and Prof. Fraydoun Rezakhanlou for teaching me the importance of mathematical analysis.

I would like to thank Dr. David Bindel for the countless hours of discussion, long detailed emails, spectacular code, and invaluable advice. My basic understanding of numerical computation and programming are all due to his guidance and teachings. Had it not been for his support I believe that my research would not have progressed as it has.

I would like to thank all of my friends who have supported me during my Ph.D, my friends in Berkeley, ETH, and in Japan. Having had the opportunity to study at 3 different places on 3 different continents, I am privileged to have met so many people who have influenced me and shaped me academically and as a person. I am forever in their debt.

Lastly I would like to thank my parents for all their loving support during my studies.

The work described here was in part supported by the National Science Foundation Grant ECS-0426660; the University of California MICRO program; and Sun Microsystems.

Chapter 1

Introduction

The work in this dissertation involves efficient numerical evaluation of damping phenomenon that occur in resonant Microelectromechanical Systems(MEMS) [161]. MEMS is the broad term used to describe micron-sized devices that interact not only with the electrical and mechanical domain, but also with the thermal and fluidic domain. Within MEMS, we focus our attention on devices called resonant MEMS, which have applications as filters and oscillators in integrated circuitry(IC) technology, and sensors.

1.1 Outline and contributions

- In the remaining portion of this introduction, an overview of resonant MEMS devices along with their applications, the damping mechanisms that affect their performance, and the numerical methods that exist to evaluate the quality factor Q, which is a measure of damping, are presented.
- Chapter 2 focuses on a methodology to evaluate damping that arises from the energy dissipation mechanism called anchor loss. This is modeled through a technology called Perfectly Matched Layers (PML). We have developed heuristics for selecting optimal parameters in the PML for

a 1D scalar wave problem and they are confirmed in higher dimensions. Here, optimality is defined in terms of the least computational effort required for a desired accuracy in the solution. In contrast to the work of Bindel, such a selection of optimal parameters is made possible by the expression for the interface reflection that we have constructed from close observation of results obtained from numerical experiments. The quality factor Q can be evaluated by solving a generalized eigenvalue problem obtained from a finite element discretization of the governing equations with PML applied. The application of PML results in complex-symmetric (non-Hermitian) system matrices, rendering the solution of the eigenvalue problem non-trivial when the size of the matrices required for physically meaningful solutions exceeds the order of millions. The generalized eigenvalue problem is solved with a Jacobi-Davidson QZ solver in combination with an efficient geometric multigrid preconditioner that we have developed for solving the complex-symmetric linear systems arising from the applications of PML. To the best of our knowledge, this is the first successful attempt in applying geometric multigrid to these linear systems. The details of the geometric multigrid are presented and the convergence behavior of the method with respect to PML parameters is presented. This is the also the first application of the Jacobi-Davidson QZ solver to problems on the order of millions of degrees of freedom.

- Chapter 3 focuses on a methodology to evaluate damping that arises from the mechanism called thermoelastic damping. This is modeled by solving the coupled mechanical balance of equilibrium and heat equation. Besides evaluating the eigenvalues of this system, the quality factor Q can be computed by transfer function evaluation. A reduced order modeling technique based on second-order Krylov subspaces is presented. The technique preserves the structure of the finite element discretized system of equations to produce a reduced order model which matches twice as many moments of the transfer function for systems with symmetric mechanical forcing and sensing. In order to show this property, a new theorem proving the moment matching properties based on a formulation through second-order Krylov subspaces is presented.

- Chapter 4 focuses on efficient damping evaluation of electromechanically coupled systems through equivalent circuit model parameter extraction and transfer function evaluation. In the design of a resonant MEMS device as a component in an electrical circuit, the engineer is interested in the behavior of the mechanical device at a specific frequency and mode of vibration, namely at resonance. The device at this frequency can be modeled by a single degree of freedom system through lumped mechanical parameters. From an analogy between a single degree of freedom mechanical and electrical system, these lumped mechanical parameters can be translated to equivalent circuit model parameters. The engineer can then use these few parameters to effectively model the complex mechanical resonator and subsequently simulate the entire response of the circuit including the mechanical system with ease. As opposed to the slightly ad hoc parameter extraction process presently used, a systematic parameter extraction process based on a variational framework for modeling electrostatic and piezoelectric electromechanically coupled systems is presented. The new framework allows one to treat systems with various geometry and to incorporate the electromechanical coupling effect into the mechanical modes of vibration.

- Chapter 5 presents numerical examples of MEMS structure which are used to exhibit the behavior of the numerical methods that we have presented for efficient modeling of systems with damping. The numerical simulations are conducted using the open-source software *HiQLab* [35]. *HiQLab* has been initiated by Bindel and members of the *SUGAR* [85] group, including myself, have made contributions in developing the software. A brief overview of *HiQLab* is presented along with details regarding the interface to the parallel iterative solver library *PETSc*, which is used to solve linear systems up to the size of millions. This is followed by the simulation of three MEMS devices. The first is the disk resonator with which the damping mechanism of anchor loss is modeled through the application of PML. The method that we have developed to compute the quality factor Q from the complex-valued frequencies through the computation of a complex-symmetric generalized eigenvalue problem is presented. The

simulations are conducted on a parallel processor machine to observe how the compute time scales as the number of processors and number of degrees of freedom of the problem increase. Fully 3D mechanical simulations are conducted on a family of disk resonators for which experimental results have claimed a large decrease in the quality factor Q with respect to increasing post misalignment. Our purely mechanical simulations reveal small decrease of Q with respect to post alignment. This implies that other sources such as electromechanical coupling effects and geometrical variations which we have not incorporated must be the cause for the sensitive behavior in Q. Next a dielectric transduced resonator is analyzed using the technology developed for electrostatic electromechanically coupled systems. The extracted equivalent circuit model parameters produce a transfer function with relatively accurate behavior near the resonance frequency with less computation time than the full model. Parametric studies of the equivalent circuit model parameters with respect to variations in the material and geometrical parameters reveal results which are not accessible by simple parallel plate assumptions of the electromechanical coupling effect. An application of a formula based on a simple parallel plate assumption can neglect the amplitude of vibration of the parallel plate, resulting in an false estimation of a 4th power dependence of the motional resistance on the gap size, which in reality should be a 2nd power dependence. Our simulations are able to predict this behavior. The last example is a thermoelastic beam resonator and ring resonator, with which the reduced order modeling technology for the thermoelastic problem is used to compute the transfer function. Analogously to the case of the equivalent circuit model parameter extraction, fast accurate evaluation of the transfer function is presented.

1.2 Resonant MEMS

Resonant MEMS are a class of devices which have applications in radio frequency (RF) wireless technology as potential replacements for off-chip bulky components, reducing the total size, cost, and

energy consumption of devices like wireless transceivers [155, 145, 146]. Figure 1.1 shows a schematic of a present-day wireless transceiver front-end architecture [147] and Figure 1.2 shows a schematic of the envisioned future multi-band receiver front-end architecture [147]. The components shaded in gray represent the mechanical components which are either filters (RF BPF: Radio Frequency Band-Pass Filter) or oscillators (Xstal Osc: Crystal Oscillator). One can clearly see from their abundance, the importance that mechanical components currently play and will continue to play in wireless technology. A filter is a signal processing unit used to select desired or remove unwanted components from a spectrum of frequencies, and an oscillator is used as a reference frequency generator for a mixer (depicted in the diagram as $\otimes$), which either upconverts or downconverts signals between very high-frequency (VHF) and intermediate frequency (IF) ranges [120].

One may wonder why mechanical components are used instead of electrical components. This is due to the performance requirements of the components in the RF electrical circuit. For a system under sinusoidal excitation at a given frequency ω, an index called the Quality Factor (Q) defined as

$$Q \quad := \quad \frac{\text{Maximum energy stored in system}}{\text{Energy dissipated per radian}}, \tag{1.1}$$

is used to measure its performance. As will be explained in more detail later in this chapter, Q is intimately related with the transfer function of the system and the sharpness of the resonance peak. Narrow bandwidth and low energy loss are desired for components in RF circuitry filters which requires high Q values on the order of 10,000 and higher. These are attainable only through mechanical components. With purely electrical components, Q values are limited to values on the order of 1 to 10. A high Q is also desired in RF oscillators since a low Q can increase the amount of phase noise in the system (phase noise floor) to the level of totally swamping the desired low-power signals.

Current state-of-the-art technology uses ceramic filters, surface acoustic wave (SAW) filters, quartz filters, film bulk acoustic resonator (FBAR) filters, and quartz oscillators, which are capable

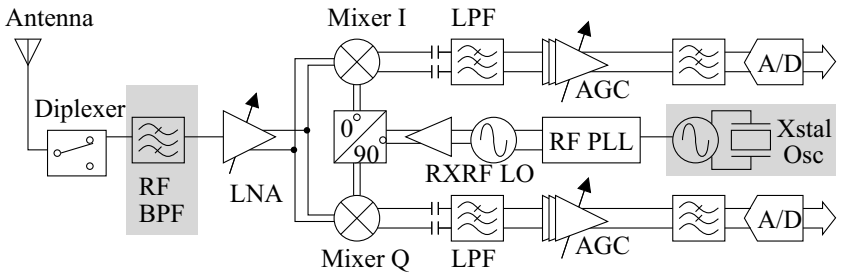


Figure 1.1: Present wireless receiver front-end architecture (Adapted from [147])

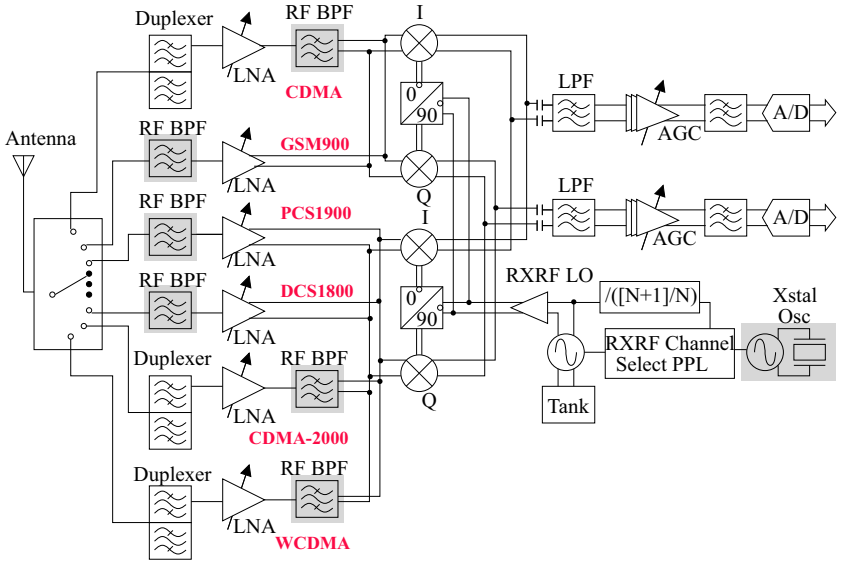


Figure 1.2: Future multiband wireles receiver front-end architecture (Adapted from [147])

of achieving Q's up to 10,000 needed for RF bandpass filtering and frequency generation. The only problem is that they are all off-chip components taking up space and comprising a sizable fraction of the parts and assembly cost. By replacing these off-chip components with on-chip integrated MEMS components, there is a large gain in terms of space, cost, and energy consumption [143].

The success in development and fabrication of resonant MEMS has introduced a new paradigm regarding the originally individual stand-alone MEMS devices as hierarchical building blocks for more complex and integrated structures. This approach is similar to the process in which integrated circuitry (IC) has developed. By combining the individual MEMS devices either mechanically,

electronically, or by both, in quantities ranging from 2 to 60, successful filter designs with low insertion and sufficient band-width have been constructed [126]. Our goal is to make the design of such devices, hitherto quite experimental, systematic and straightforward.

1.3 Damping mechanisms

As mentioned in the previous section, maximization of Q in resonant MEMS is crucial in obtaining superior performance as filter or oscillator components. One sees from the given definition of Q that energy dissipation within the system, i.e., damping in the system, must be minimized for this objective. This situation is contrary to applied areas such as earthquake engineering where damping is a considered a virtue rather than a vice. Numerous experiments and analysis have been conducted to determine the main causes of damping in high-frequency resonant MEMS, revealing mechanisms such as anchor loss [97], thermoelastic damping [66, 2, 195, 159, 97, 197, 46], material loss, air damping [97, 46], and ohmic loss. Within these mechanisms, in RF MEMS applications, anchor loss and thermoelastic damping have been experimentally acknowledged to be prominent sources of damping. Thus in this research, focus is centered on the efficient numerical evaluation of these mechanisms.

1.3.1 Anchor loss

Anchor loss is the mechanism in which energy is lost from a resonating system in the form of outgoing propagating waves that do not return. In resonant MEMS, the devices are fabricated on top of a Silicon substrate to which it is connected through anchors. Compared to the micron-sized device, the substrate is orders of magnitudes larger and can be considered almost a semi-infinite half domain. As the MEMS device resonates, this motion couples with the underlying substrate through its anchors, sending non-returning waves into the substrate. This damping mechanism has been experimentally observed to be prominent in resonant MEMS at high-frequencies larger than

the MHz range and depends strongly and sensitively on the design and mode of vibration of the device. There exist experimental data where breaking one of the anchors in the device has lead to order of magnitude difference in measured Q of the device. Contrary to this importance, not many attempts have been conducted to analyze and simulate this phenomenon.

There exist two approaches besides the one that we have made: semi-analytical and fully analytical. In the semi-analytical approach of Park and Park [151, 152], as the first step, a Fourier transform is applied in the plane of the substrate surface, which can be considered infinite. The direction orthogonal to this plane is left untransformed. This step is conducted to evaluate the impedance of the substrate. This is combined with the impedance of the structure under proper interface matching conditions to obtain the overall system response. The Q is the extracted from the transfer function which can be obtained from the impedance. This method is advantageous in the sense that the substrate is not modeled as semi-infinite and wave reflections on the backside of the substrate are modeled properly. This can be important when the substrate cannot be considered semi-infinite, such as the case when the propagating wave length in the substrate is large. The disadvantage is that a spatial inverse Fourier transform is required to evaluate the entries of the impedance matrix for the substrate which can be costly and tedious.

In the approach of Hao and Ayazi [151, 152, 90, 89], contrary to Park and Park [151, 152], a purely analytical expression for Q is sought based on simplifying assumptions on the structure and mode of vibration of the device. The Q here is computed by assuming that the work imposed on the substrate by the anchor is equal to the energy loss. The analytical closed form expressions obtained are particularly attractive to designers due to their simplicity, but the drawback is that they are only valid under the assumptions that are made and available for restricted geometries. Thus if the geometry is special such that the mechanism of damping is not known, this method is inapplicable. For example, the disk resonator fabricated by Wang et al. [188] has been analyzed by Hao and Ayazi [89] but due to these limitations they cannot simulate the effect of post misalignment on Q. Contrary to this we will present the results of simulations on such general geometries using

our method in Section 5.3. We will see that for the first time it is possible to rapidly simulate the effects of anchor loss for general geometries, especially the effect of post misalignment to the quality factor Q.

1.3.2 Thermoelastic damping

Thermoelastic damping is the mechanism in which energy is dissipated through the coupling between the mechanical and thermal domain. Mechanical oscillation couples into the thermal domain via volumetric expansion causing local temperature fluctuations. This in turn produces heat flow in the system leading to energy dissipation.

The first to extensively study this phenomenon was Clarence Zener, who published a series of papers starting in 1937 ranging from theoretical clarification of the mechanism and applications in polycrystals to experimental verifications in metals [200, 201, 202, 203, 204]. The formula he derived for evaluation of thermoelastic damping for beam structures in bending has been widely accepted due to its closed form, simplicity, and applicability. Though the formula was originally formulated for macro beams, the phenomenon has been experimentally observed in MEMS beam resonators down to the micron range. The approximations made in his formula were Euler-Bernoulli beam bending and use of only the first thermal eigenmode across the thickness in expressing the thermal field. Recently, Lifshitz and Roukes [130] have developed a slightly more refined model in absence of the approximation of the thermal field by the first eigenmode, but still restricted to beams. Slight variants have also been proposed by Srikar and Senturia [168] for polycrystalline beams (though the formula has not been experimentally verified), by Prabhakar and Vengallatore [157] for bilayered beams, and Wong et al. [191] for ring structures vibrating in in-plane flexural modes.

The formula proposed by Zener has been verified to be accurate, but is only applicable to beams or thin structures vibrating in their flexural mode. As designers explore different geometries which are no longer of this type, more general methods based on numerical simulations are required. The approaches that have been conducted are all based on the finite element discretization of the coupled

equations of thermoelasticity, i.e., the linearized balance of linear momentum and linearized heat equation [83, 60, 121, 199]. They have been shown to be valid and verified against experimental data for beam type geometries. Solution of the coupled equations has one drawback, it increases the degrees of freedom of the problem and increases the computational intensiveness. Thus what is now required are efficient methods to evaluate thermoelastic damping. In Chapter 3 we present a method based on transfer function evaluation of a reduced-order model which can be up to 60 times faster than a full finite element discretization, and as accurate.

1.4 Numerical evaluation of the quality factor

As mentioned in the previous section, the quality factor Q is a measure of the performance of a device and a large value is desired. The expression to evaluate Q was introduced in Equation (1.1). Unfortunately this expression is not suited for experimental or numerical evaluation, since difficulty can arise in computing or measuring energy dissipation and defining maximum stored energy. In experimental work, the quality factor is often evaluated from the transfer function. In numerical simulations, the quality factor can be evaluated from either the transfer function or an eigenvalue computation. These two methods are used to evaluate the quality factor Q in our simulations.

The equivalence between the three methods for evaluating the quality factor Q of a single degree of freedom system can be easily shown. The governing equation of motion for such a system is

$$m\ddot{x} + c\dot{x} + kx \quad = \quad f \tag{1.2}$$

where m is the mass, c is the damping coefficient, k is the stiffness, x is the displacement, and f is the forcing term. m, c, and k are all assumed real. Under a time-harmonic forcing term with

frequency ω, f, x, and velocity v can be expressed as,

$$\begin{aligned} f &= f_0 e^{i\omega t}, && (1.3)\\ x &= x_0 e^{i\omega t}, && (1.4)\\ v &= \dot{x} = v_0 e^{i\omega t}, && (1.5)\\ &= i\omega x_0 e^{i\omega t}. && (1.6) \end{aligned}$$

By inserting these terms into the equation of motion, one obtains the force-velocity transfer function,

$$H_v(\omega) := \frac{v_0}{f_0} = i\omega \frac{x_0}{f_0} = \frac{1}{\frac{k-m\omega^2}{i\omega} + c}. \quad (1.7)$$

The quality factor Q computed for this example with the three different methods are as follows.

1. **Energy computation(original definition)**

   Here we assume $\omega \in \mathcal{R}$. The maximum energy stored in the system is,

   $$W_{\text{Maximum stored energy}} = \frac{1}{2} k \left|x_0\right|^2 = \frac{1}{2} k \left|\frac{f_0}{\omega}\right|^2 \left|\frac{v_0}{f_0}\right|^2. \quad (1.8)$$

   The energy dissipated per radian in the system, which is the amount of work done per cycle is,

   $$\begin{aligned} W_{\text{Energy dissipated per radian}} &= \frac{1}{2\pi} \oint \mathbf{Re}(f)\mathbf{Re}(v)dt \\ &= \frac{1}{2\omega} \mathbf{Re}\left(\overline{f_0} v_0\right) \\ &= \frac{|f_0|^2}{2\omega} \mathbf{Re}\left(\frac{v_0}{f_0}\right). \quad (1.9) \end{aligned}$$

   Thus from Equation. (1.1),

   $$Q_{\text{energy}}(\omega) = \frac{k}{c\omega}. \quad (1.10)$$

2. **Eigenvalue evaluation**

   Given Equation (1.2) and the time-harmonic assumption on the displacement in Equation (1.4), the complex eigenfrequencies ω of the system satisfy the equation,

   $$-m\omega^2 + ic\omega + k\omega = 0. \quad (1.11)$$

The expressions for the complex eigenfrequencies are,

$$\omega = \frac{ic \pm \sqrt{4mk - c^2}}{2m} \ . \tag{1.12}$$

By defining the quality factor computed from the complex eigenfrequency as,

$$Q_{\text{eigen}} \quad = \quad \frac{|\omega|}{2\mathbf{Im}(\omega)}, \tag{1.13}$$

one obtains the expressions,

$$Q_{\text{eigen}} = \frac{\sqrt{mk}}{c}. \tag{1.14}$$

This expression is identical to the quality factor obtained from the energy computation definition $Q_{\text{energy}}(\omega)$, when $Q_{\text{energy}}(\omega)$ is evaluated at the forcing frequency $\omega_0 = \sqrt{\frac{k}{m}}$. ω_0 is the complex eigenfrequency of the undamped system with $c = 0$, as well as the value at which $H_v(\omega)$ takes a maximum for $\omega \in \mathbb{R}$.

3. **Transfer function evaluation**

As components in electrical circuits, mechanical resonators behave as transducers, transferring information between the electrical domain and mechanical domain [161] in the form of energy. The index of importance in measuring this energy transfer is the amount of energy transfered per unit time, which is power. In the electrical domain, the power conjugate variables are voltage V and current I. In the mechanical domain, the analogue of this power conjugate variable pair is the force f and velocity v. For this reason we would like to focus on the force-velocity transfer function $H_v(\omega)$ with $\omega \in \mathbb{R}$.

The force-velocity transfer function presented in Equation (1.7) has a magnitude of,

$$|H_v(\omega)| \quad = \quad \frac{1}{\sqrt{\left(\frac{k - m\omega^2}{\omega}\right)^2 + c^2}}, \tag{1.15}$$

with peak at frequency $\omega_{\text{peak}} = \omega_0 = \sqrt{\frac{k}{m}}$ and value $|H_v(\omega_0)| = \sqrt{\frac{1}{c^2}}$. The function $|H_v(\omega)|$ has a peak at ω_0 and decreases locally as ω differs from ω_0. Let us define the values of the

frequency where $|H_v(\omega)|$ drops to the value of $\sqrt{\frac{1}{2c^2}}$ in the neighborhood of ω_0 as ω_l and ω_u where,

$$\omega_l = \frac{-c+\sqrt{c^2+4mk}}{2m}, \tag{1.16}$$

$$\omega_u = \frac{c+\sqrt{c^2+4mk}}{2m}. \tag{1.17}$$

By defining the $1/\sqrt{2}$ bandwidth of the force-velocity transfer function in the neighborhood of ω_0, as the the difference between the frequencies that take the value $\frac{1}{\sqrt{2}}|H_v(\omega_0)|$, we obtain,

$$\Delta\omega_{\text{Bandwidth at } 1/\sqrt{2}} = \frac{c}{m}. \tag{1.18}$$

Finally we define the quality factor Q from this transfer function as,

$$Q_{\text{transfer}} := \frac{\omega_{\text{peak frequency}}}{\Delta\omega_{\text{Bandwidth at } 1/\sqrt{2}}}, \tag{1.19}$$

one obtains the expressions,

$$Q_{\text{transfer}} = \frac{\sqrt{mk}}{c}. \tag{1.20}$$

This is again identical to $Q_{\text{energy}}(\omega_0)$ and Q_{eig}.

Remark 1: The expression -3[dB] used in literature to find the bandwidth in Equation (1.19) is an approximation to the more proper expression,

$$20\log_{10}\frac{1}{\sqrt{2}} \approx -10 * 0.3010 = -3[\text{dB}].$$

For a multiple degree of freedom system, an exact equivalence is difficult to show. Consider the multiple degree of freedom system below,

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{C}\dot{\mathbf{x}} + \mathbf{K}\mathbf{x} = \mathbf{B}f, \tag{1.21}$$

$$y = \mathbf{L}^*\mathbf{x}, \tag{1.22}$$

where $\mathbf{M}$ is the mass matrix, $\mathbf{C}$ is the damping matrix, $\mathbf{K}$ is the stiffness matrix, $\mathbf{x}$ is the displacement, f is the scalar input, $\mathbf{B}$ is the vector which defines the input forcing pattern, y is the desired

scalar output, and $\mathbf{L}$ is the vector which defines the output sensing pattern. The input f is assumed time-harmonic $f := \hat{f}\exp(i\omega t)$, resulting in a time-harmonic displacement $\mathbf{x} := \hat{\mathbf{x}}\exp(i\omega t)$ and output $y := \hat{y}\exp(i\omega t)$. Here a single input single output (SISO) system has been selected for clarity in the discussion. Depending on the type of problem the matrices $\mathbf{M}, \mathbf{C}, \mathbf{K}$ have different properties. For example, as is shown in Chapter 3, the thermoelastic system of equations leads to a singular real-symmetric $\mathbf{M}$, singular real-unsymmetric $\mathbf{C}$ and $\mathbf{K}$. The anchor loss simulations with PML presented in Chapter 2, lead to complex-symmetric $\mathbf{M}$ and $\mathbf{K}$. In such cases, the definition of maximum energy stored in the system and energy dissipated per radian can become vague, which can make the applicability of Q evaluation method 1 troublesome. For the thermoelastic system of equations, should one consider just the mechanical energy? For the case of the anchor loss simulations with PML, should one consider the real part of $\mathbf{K}$ in evaluating the maximum stored energy? In this aspect the Q evaluation methods of 2 and 3 are more robust and can be applied in a straightforward manner.

Numerical evaluation of the quality factor Q from the eigenvalue (Q evaluation method 2), results in solving for the eigenvalues of the quadratic eigenvalue problem,

$$\left(-\omega^2\mathbf{M} + i\omega\mathbf{C} + \mathbf{K}\right)\hat{\mathbf{x}} \quad = \quad 0, \tag{1.23}$$

obtained from inserting $\mathbf{x} = \hat{\mathbf{x}}\exp(i\omega t)$ into Equation (1.21). With the complex-valued eigenvalue ω_{eig}, Q is evaluated from Equation (1.13). Thus the amount of computation required for evaluating Q is an eigensolve, which for large sparse systems may involve several linear system solves. Often one is interested in Q values for modes that have eigenvalues in the interior of the spectrum which can add difficulty to the linear solves.

Numerical evaluation of the quality factor Q from the transfer function (Q evaluation method 3) requires evaluating the function $H(\omega)$,

$$\hat{y}(\omega) \quad = \quad H(\omega)\hat{f}, \tag{1.24}$$

$$H(\omega) \quad := \quad \mathbf{L}^*\left(-\omega^2\mathbf{M} + i\omega\mathbf{C} + \mathbf{K}\right)^{-1}\mathbf{B}, \tag{1.25}$$

for discrete sampling points. Q is then extracted from the approximate graph. The computational effort required for evaluation of Q is the number of sampling points times the computational effort to apply $\left(-\omega^2\mathbf{M} + i\omega\mathbf{C} + \mathbf{K}\right)^{-1}$.

The equivalence between the quality factor obtained from the eigenvalue computation and transfer function evaluation for the multidimensional case can be argued under the assumptions of $Q >> 1$ and through Laurent series expansions of the transfer function $H(\omega)$ around the complex-valued eigenvalue ω_{eig} which is a pole of the transfer function [33].

The computation of Q from the eigenvalue problem presented in Equation (1.23), and the evaluation of Q from the transfer function in Equation (1.25) can become difficult when the size of the matrices involved in the computations become large.

- Larger systems simply require more computational effort, i.e., computational time, to solve.
- Algorithms available for smaller systems may not be applicable to larger systems due to their increase in compute time for a solution and required computational memory.

One can overcome these difficulties in two ways.

1. By taking advantage of parallelism in the solution method and using multiple processor machines. (Note that for parallel computing, one must produce algorithms which exploit the multiple processors).

   This approach is taken in Chapter 2 to simulate the quality factor Q for anchor loss simulations through eigenvalue evaluation.

2. By devising elegant reduced modeling schemes to circumvent repeated large scale intensive computations.

   This approach is taken in Chapter 3 to simulate the quality factor Q for thermoelastic damping through transfer function evaluation and in Chapter 4 to observe the behavior and compute the quality factor Q of an electronic circuit including an electromechanical resonator.

Chapter 2

Anchor loss simulations

2.1 Introduction

In the design of high-frequency micromechanical resonators, the energy dissipation mechanism called anchor loss has been recognized as one of the most prominent mechanisms in lowering the quality factor Q. Anchor loss is the mechanism in which energy is lost from a resonating system in the form of outgoing propagating waves that do not return. High-frequency micromechanical resonators are fabricated on top of a Silicon substrate to which it is connected through its anchors. Compared to the micron-sized device, the substrate is orders of magnitudes larger and can be considered almost a semi-infinite half domain. As the device resonates, this motion couples with the underlying substrate through its anchors, sending non-returning waves into the substrate. To model this phenomenon in a numerical simulation, one must truncate this semi-infinite domain finitely and apply appropriate boundary conditions to mimic the semi-infinite domain behavior. Application of a simple fixed or free boundary conditions results in unphysical standing waves, and thus a radiation boundary condition that enforces only outgoing wave solutions must be employed.

Situations similar to this problem are encountered in other areas such as computational acoustics and computational electromagnetics in the form of wave scattering problems. In the acoustics

literature the radiation boundary condition is referred to as the Sommerfeld radiation boundary condition. The numerical boundary conditions which approximate the exact radiation boundary condition can be classified into three groups: absorbing boundary conditions (e.g., local BGT and globalDtN), infinite elements(e.g., Burnett, Astley-Leis elements), and absorbing layers(e.g., perfectly matched layers). Each method has its advantages and disadvantages, which are summarized in the review articles by Harari [91] and Thompson [176]. Perfectly Matched Layers (PML), the radiation boundary condition that we use for modeling, can be interpreted as applying a sponge like layer at the computational domain boundary to damp and absorb outgoing waves. This method has been chosen for its ability to absorb all outgoing propagating waves at any angle of incidence with zero impedence mismatch at the interface in the ideal case. Additionally in a finite element formulation, the method preserves the sparsity of the linear system one must solve, compared to global DtN boundaries which couple all the degrees of freedom on the boundary of the domain. Such sparsity leads to faster solution time of the linear system of equations. The method also amends itself easily to Cartesian coordinate systems as opposed to cylindrical or spherical coordinate systems. This technology was originally developed by Berenger for solving Maxwell's equation in the time-harmonic case [29] and finite difference time-domain (FDTD) case [30]. The original formulation for the Maxwell's equation posed in first order form in terms of both electric and magnetic fields involved a splitting of the fields into orthogonal and tangential components. This mixed formulation was applied to the elasticity equations in a compressional and shear wave potential form based FDTD scheme by Hastings et al. [94] , in the velocity-stress formulated FDTD scheme by Chew and Liu [50] and Collino and Tsogka [56], and further adapted by Basu into a pure displacement based finite element formulation for both the time-harmonic [25] and transient case [26].

In the continuous case, an infinitely large PML domain is able to absorb all outgoing waves propagating at an angle of incidence into the medium with zero impedance mismatch at the PML interface. In the numerically discretized case, an infinitely large domain is not possible and discretization introduces error. The discretization reflection or error in the ability to mimic the radiation boundary

condition depends on many factors including the thickness of the PML, the absorbing function used in the PML, discretization of the mesh, and the dominant propagating wave length and speed. For the time-harmonic PML, dispersion analysis based on constant absorbing function [93], optimization of the values the absorbing function takes in the PML [55, 95], application of unbounded absorbing functions [31], and studies on dividing the numerical reflection into contributions from end and interface reflection [37] have been conducted to understand the behavior of the numerical reflection and determine optimal parameter selection in the discretized version of PML.

To compute Q of a mechanical system, one must either conduct a series of force computations under a time-harmonic load to evaluate the transfer function, or conduct an eigenvalue computation to evaluate the complex-valued frequencies corresponding to the mode of interest. In both cases one is faced with solving a linear system of equations. The application of PML produces a linear system of equations which are complex-symmetric in the forced computation [37]. For linear systems up to the order of 100,000, direct methods involving an LU factorization [57] of the linear system are the method of choice in terms of solution time and computational memory. When the size of the linear system exceeds this order, direct methods are no longer tractable and iterative methods [160] are the methods of choice. Such cases easily arise in 3D problems and cases where fine meshes are required for highly accurate solutions. The current technology of iterative methods is not as robust as direct methods, i.e., there do exist general iterative methods such as GMRES applicable for all linear systems but the time required for a solution may not be feasible even for relatively small size problems. One must take into account the properties of the linear system of equations to select the adequate combination of iterative method and preconditioner required to accelerate the iterative method [28].

For the class of symmetric positive definite linear systems arising from discretization of elliptic equations, the multigrid iterative method has proven to be optimal [182]. The multigrid method can be used alone as an iterative solution method, or as a preconditioner in combination with iterative methods such as Conjugate Gradients [160]. The equation governing linear elasticity in the

quasi-static case result in a symmetric positive definite linear system for which geometric as well as algebraic variants [69, 185, 135] of multigrid have been developed. The equations governing linear elasticity in the dynamic case are not symmetric positive definite; for the transient dynamic case, one obtains a vector-valued Helmholtz equation. The discretization results in a symmetric indefinite linear system. Application of the PML additionally renders the system complex-valued symmetric. Multigrid methods for the indefinite and complex-symmetric case are not as well understood, and convergence proofs exist only for the real-symmetric indefinite case when the coarse grid is assumed to be sufficiently fine enough [22, 41, 163]. The existence of several modes with small eigenvalues for which error is not sufficiently reduced, poses a problem for multigrid in the real-symmetric indefinite case. Wave-ray multigrid methods aim to treat these modes separately from the standard multigrid iteration [132], but such a method requires a priori knowledge of the problematic modes. Similar to the idea of explicitly including problematic modes, is an approach analogous to the smooth aggregation multigrid method [185] where the assumed problematic modes are included in the coarse grid in a two grid solve [186]. Multigrid applicability to the complex-symmetric case, includes all problematic issues existing for the real-symmetric indefinite case, with the addition of a complex-valued spectrum.

The choice of the optimal iterative method for solving Helmholtz's equation which results in either a real-symmetric indefinite or complex-symmetric system is not well established. Methods such as GMRES preconditioned with multigrid with GMRES smoothers [67, 5] have been proposed but still lack theoretical background and robustness in their application. Additionally, most applications focus on the discretization of scalar-valued Helmholtz equations whose difficulty differs from the vector-valued elasticity problem. For the scalar-valued Helmholtz equations, a multigrid preconditioner constructed from a complex-valued shifted Laplacian operator has been presented [68] with adequate performance, but applicability to the vector-valued elasticity equations is not clear.

Non-multigrid methods that have been applied to the real-symmetric indefinite or complex-symmetric system arising from the discretization of the Helmholtz equation are GMRES precondi-

tioned with incomplete LU factorization [109] and domain-decomposition methods such as FETI-H [71] and FETI-DPH [70]. There also exists a vast literature on other solution methods for the Helmholtz problem and the interested reader is referred to the review articles of current technology by Harari [91] and Thompson [176].

In order to compute the quality factor Q from the complex-valued frequencies of a system with PML applied, one must solve a complex-symmetric generalized eigenvalue problem. Analogous to the solution of linear systems, a problem arises when the matrices involved are larger or equal to millions of degrees of freedom. Eigenvalue methods for sparse systems must be employed. Currently one of the most widely used and robust methods to compute eigenvalues of large systems is the Arnoldi method based on Krylov subspace construction and projection onto this subspace [18]. For high-frequency MEMS resonators design, the mode of interest may have an eigenfrequency in the interior of the spectrum. In this case a combination of the Arnoldi method with a shift-and-invert spectral transformation yields fast results. The crucial part of the algorithm is the ability to apply the shift-and-invert operator with high accuracy, involving the solution of a linear system. If high accuracy cannot be attained, the generated projection subspace no longer suffices as a Krylov subspace and Ritz-values and Ritz-vectors obtained may not be of sufficient accuracy. For systems of increasing size, direct methods become unfeasible in terms of memory and time and iterative methods must be employed. Attaining high accuracy for solves with iterative methods can be expensive due to the large number of iterations required to attain the desired accuracy. The Jacobi-Davidson method [167] is capable of circumventing this problem by tolerating moderately accurate solves. The Jacobi-Davidson method also belongs to the class of projection methods, just like Arnoldi, but constructs the projection method in a way different from a Krylov subspace. A correction equation is solved to find a good direction to "correct" the current eigenvector approximation, and this new correction direction is used to expand the projection subspace. This correction equation can be solved to any desired accuracy, with the consequence that low accuracy can result in a requirement of a larger projection subspace. For the generalized eigenvalue problem a variant using the QZ

algorithm called Jacobi-Davidson QZ [74] can be employed.

In this chapter, a method to evaluate the optimal PML parameters based on the presentation of the numerical reflection coefficient by Bindel [37] is presented. Optimality is defined in terms of obtaining a desired accuracy in the solution under the least amount of time. In the case of applying direct methods to solve the systems with PML, the time required for solution is proportional to the size of the computational domain. Thus the shortest PML layer possible yielding the desired accuracy leads to optimal solution time. In the case of applying iterative methods such as multigrid to solve the systems with PML, a limit is enforced on the possible PML parameters that can be selected for the method to work. Also in this case, the computational time is proportional to the size of the computational domain, but the determination of the shortest PML has a constraint such that the corresponding PML parameters are amenable to the multigrid method. Thus a clear understanding of the relationship between the desired accuracy, PML length, and PML parameters is required. A method to determine the shortest PML thickness and corresponding PML profile parameter for a desired reflection given the type of elements and discretization is developed. This method is formulated for the 1D scalar wave equation, and the claims are confirmed with 2D scalar wave and 2D elastodynamics equations through an index equivalent to the numerical reflection coefficient called the energy dissipation error. This is followed by the presentation of a geometric multigrid method applicable for solving complex-symmetric systems with PML, including the presentation of smoothers and interpolation operators. The effectiveness and behavior of the method is illustrated and confirmed by computation of two grid convergence factors for 2D scalar wave and 2D elastodynamics problems. Finally, an eigenvalue computation method for Q computation with the Jacobi-Davidson method is presented. Through the analysis of a 1D scalar wave problem, heuristics for selecting PML parameters to attain a desired accuracy in Q are also presented.

2.2 Optimal perfectly matched layers parameter estimation

To apply the method of Perfectly Matched Layers (PML) for numerical approximation of the radiation boundary conditions, one must select parameters for the PML such as the length of the PML and absorbing function. Unless these parameters are selected properly, the performance of the PML can degrade, and desired accuracy cannot be obtained. The main problem in the proper selection of PML parameters arised from the numerical discretizaion of the problem. In the application of PML to a continuous problem defined on a finitely truncated domian, the selection of PML parameters is not too difficult, since one only has to treat the wave reflections arising from the finite termination of the PML. When the PML is discretized, additional wave reflection can occur at the PML interface, leading to less accurate results. Various polynomial absorbing function profiles and PML layer lengths have been experimented with to search for an optimal combination. The selection of PML parameters also differs between frequency-domain and time-domain applications.

For the finite difference time-domain (FDTD) electromagnetic case, the search for optimal parameters initiates with the work of Bérenger in which it is identified that absorbing functions with imaginary parts rising from zero at the PML interface produce small numerical reflections [29]. Complying to this claim, polynomial profiles have been suggested [30]. From the numerical experiments with various profiles, Bérenger states that the largest source of error of the PML arises from reflection at the PML interface [30]. Through extensive numerical experiments, Gedney [78] presents an optimal choice of PML parameter for 5 and 10 node spacings in the PML. The work of Collino and Monk [55] take an optimization approach to determining the PML parameters for 5 and 10 node spacings in the PML. In the time-domain problem, the PML layer must absorb waves of various frequencies, and the magnitude of each contribution is not clear. Thus it is unclear if these derived heuristics are applicable for the frequency-domain problem.

For the frequency-domain case, several studies have been conducted by Harari and Albocher [92] and an optimization for the polynomial absoribing function profile has been presented by Heikkola

et al. [95]. In both cases the presented results are numerical experiments than a presentation of heuristics for optimal PML parameter selection. A more precise analysis of the reflection was presented by Bindel [37], where the total reflection due to PML is seperated into the contributions arising from the end termination reflection and the interface reflection. The presented framework enables one to select optimal PML parameters in a rational way.

In this section, a method based on the work of Bindel for estimating optimal parameters for the PML with desired accuracy in the reflection is presented. The method is developed for the 1D scalar wave equation, where the models to evaluate the contribution of the end termination reflection and interface reflection is presented. Based on these two models the PML parameter which optimizes the computational time required for the solution, through estimation of the shortest PML possible for a desired accuracy, is presented. To confirm the applicability of the obtained heuristics in the multidimensional and scalar-valued case, the energy dissipation error index, which is shown to be equivalent to the amount of wave reflection, is introduced. Through this index, the 2D scalar wave and 2D elastodynamic equations are investigated. The nomenclature introduced in this section is summarized in Table 2.1.

Table 2.1: Nomenclature

Reflection and error

r_{end}	:	End termination reflection. The reflection introduced by a finite termination of the PML.
$r_{\text{interface}}$	:	Interface reflection. The reflection introduced by the discretized PML at the PML interface.
r_{computed}	:	Computed reflection. The reflection computed from the numerical simulation. This includes the contribution of both the end termination reflection and interface reflection.
r_{model}	:	The reflection obtained from the model, as a sum of r_{end} and $r_{\text{interface}}$.
k_{relerr}	:	The relative error in the approximation of the wave number.
E_{relerr}	:	The energy dissipation error. The relative error in the amount of energy dissipated in the system.

PML parameters

β	:	End value of PML absorbing function.
$\gamma(s)$	:	Absorbing function profile defined on the unit interval. Normalized to equal 1 at $s = 1$.
p	:	The polynomial order of a polynomial absorbing function profile $\gamma(s) = s^p$.
l_{pml}	:	Length of the PML.
h	:	Distance between nodes in the PML.
n_{wpml}	:	Number of wave lengths in the PML.
n_{npml}	:	Number of nodes in the PML. $\left(\frac{l_{\text{pml}}}{h}\right)$.
n_{npw}	:	Number of nodes per wave length in the discretization.
μ	:	$\frac{\beta}{n_{\text{npml}}^p}$.

2.2.1 1D scalar wave equation

The 1D scalar wave equation with the application of PML depicted in Figure 2.1 is considered. The entire computational domain $\Omega := [0, L_p]$ is defined as the union of the bounded elastic domain $\Omega_{\text{bd}} := [0, L]$ and the wave absorbing PML domain $\Omega_{\text{pml}} := [L, L_p]$. The governing equations for the

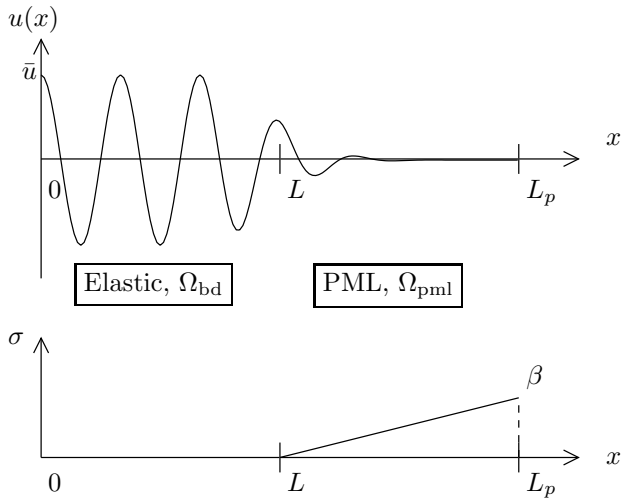


Figure 2.1: 1D scalar wave configuration with linear absorbing function $\sigma(s)$

system are,

$$\rho\frac{\partial^2 u}{\partial t^2} - E\frac{\partial^2 u}{\partial \tilde{x}^2} = 0, \quad x \in [0, L_p], \tag{2.1}$$

$$\tilde{x} = \int_0^x \lambda(s)ds, \tag{2.2}$$

$$\lambda(s) = 1 - \sigma(s)i, \tag{2.3}$$

$$\sigma(s) = \begin{cases} 0 & 0 \le s < L \\ \text{some real value} & L \le s \le L_p \end{cases}, \tag{2.4}$$

where ρ is the material density, E is the Young's modulus, t is the time, x is the original coordinate system, $\tilde{x}$ is the complex-stretched coordinate system, and i is $\sqrt{-1}$. We also define $c := \sqrt{\frac{E}{\rho}}$ as the wave speed. By definition, x and $\tilde{x}$ are differentially related as,

$$\frac{d\tilde{x}}{dx} = \lambda(x), \qquad \frac{d}{d\tilde{x}} = \frac{1}{\lambda(x)}\frac{d}{dx}\ . \tag{2.5}$$

The function $\lambda(x)$ is called the absorbing function and determines the wave absorbing behavior of the PML. For the given specific form of $\lambda(x)$, which is often used in practice, the relation between the coordinates become,

$$\tilde{x}(x) = x - i\int_0^x \sigma(s)ds\ . \tag{2.6}$$

Under time-harmonic assumptions $u(x,t) = \hat{u}(x)\exp(i\omega t)$, where ω is the forcing frequency, Equation (2.1) takes the form of the Helmholtz equation,

$$\frac{d^2\hat{u}}{d\tilde{x}^2} + k^2\tilde{x} = 0, \tag{2.7}$$

where $k := \dfrac{\omega}{c}$ is the wave number. Solutions of this equation take the form,

$$\hat{u}(x) = c_{\text{out}}\exp(-ikx)\exp\left(-k\int_0^x \sigma(s)ds\right) + c_{\text{in}}\exp(ikx)\exp\left(k\int_0^x \sigma(s)ds\right). \tag{2.8}$$

c_{out} and c_{in} are constants determined by the boundary condition. One observes in these expressions, the exponential damping behavior of outgoing waves in the PML domain. By applying the fixed boundary condition, often used in practice to terminate the PML,

$$\hat{u}(0) = \bar{u}, \tag{2.9}$$

$$\hat{u}(L_p) = 0, \tag{2.10}$$

the constants are,

$$\alpha := \int_L^{L_p} \sigma(s)ds, \tag{2.11}$$

$$c_{\text{out}} = \frac{1}{1-e^{-2k\alpha-2ikL_p}}\bar{u}, \tag{2.12}$$

$$c_{\text{in}} = \frac{-e^{-2k\alpha-2ikL_p}}{1-e^{-2k\alpha-2ikL_p}}. \tag{2.13}$$

2.2.2 End termination reflection

To mimic the infinite domain boundary condition, one desires $c_{\text{in}} = 0$, such that only outgoing waves are permitted as solutions. The ratio,

$$r_{\text{end}} := \left|\frac{c_{\text{in}}}{c_{\text{out}}}\right| = e^{-2k\alpha}. \tag{2.14}$$

can be considered a normalized measure of the quality of the boundary condition. This quantity will be given the name "end termination reflection", since it is the reflection arising from a finite end

termination of the PML. For a fixed wave number k, r_{end} tends to zero as the length of the PML $l_{\text{pml}} := L_p - L$ is increased or if the size of the function $\sigma(s)$ is increased.

For monotonically damped wave propagation behavior in the PML, $\sigma(s)$ is assumed positive and monotonically increasing. By defining $\beta := \sigma(L_p)$, $\sigma(s)$ is represented in terms of a normalized function $\gamma(s)$ defined on the unit interval $[0, 1]$ with $\gamma(1) = 1$,

$$\sigma(s) = \beta\gamma\left(\frac{s-L}{l_{\text{pml}}}\right), \quad s \in [L, L_p], \tag{2.15}$$

$$l_{\text{pml}} := L_p - L\ , \tag{2.16}$$

and therefore,

$$\alpha = \beta l_{\text{pml}} \int_0^1 \gamma(s) ds. \tag{2.17}$$

In literature it is often stated that in order to attain the perfectly matched property between the bounded domain and PML domain, the absorbing function $\sigma(s)$ must equal zero at the PML interface, i.e., $\gamma(0) = 0$ must be enforced. It should be stressed that this is not a requirement for the continuous PML, and is a heurstic that has been developed for time-domain computations. The original formulation developed by Bérenger states that the impedance between the bounded domain and PML domain must be matched. For this 1D scalar wave problem, the impedance of the bounded domain Z_{bd} is defined as $Z_{\text{bd}} := \sqrt{\rho E}$ The PML can be interpreted as an anisotropic complex-valued material [37], which yields the complex-valued density $\rho_{\text{pml}} = \lambda(s)\rho$ and complex-valued Young's modulus $E_{\text{pml}} = \frac{E}{\lambda(s)}$ in the PML. The impedance of the PML is $Z_{\text{pml}} := \sqrt{\rho_{\text{pml}} E_{\text{pml}}} = \sqrt{\rho E}$, which is equal to the impedance of the bounded domain for any absorbing function profile $\lambda(s)$. Thus any absorbing function is perfectly matched at the PML interface in the continuous case. The problem arises under discretization, such that abrupt discontinuities across the PML with coarse mesh discretization lead to large reflection at the PML interface and degradation in performance. This claim is verified with the numerical simulations presented in Section 2.2.3. where a constant PML absorbing function is selected.

Combining Equations (2.14) and (2.17) yield the result,

$$\log(r_{\text{end}}) \quad = \quad -4\pi\beta n_{\text{wpml}} \int_0^1 \gamma(s)ds, \tag{2.18}$$

$$n_{\text{wpml}} \quad := \quad \frac{l_{\text{pml}}}{\lambda} \ , \tag{2.19}$$

where $\lambda = \frac{2\pi}{k}$ is the wave length, and n_{wpml} is the number of wave lengths in the PML. Given an absorbing function profile $\gamma(s)$, the contours of constant r_{end} take the form of hyperbolas with respect to β and n_{wpml}. Taking a polynomial profile for $\gamma(s)$, as is often done in practice,

$$\gamma(s) = s^p, \tag{2.20}$$

results in the expression,

$$-\frac{p+1}{4\pi}\log(r_{\text{end}}) \quad = \quad \beta \ n_{\text{wpml}}. \tag{2.21}$$

The contours for constant r_{end} for $p = \{0, 1\}$ are shown in Figure 2.2. One observes that lower order polynomials display smaller end termination reflection for any combination of (n_{wpml}, β).

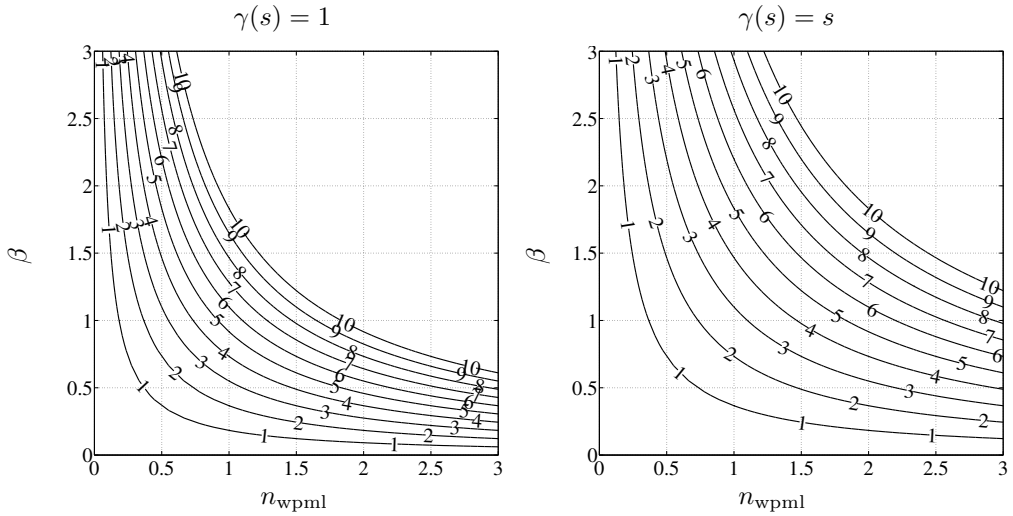


Figure 2.2: End termination reflection $-\log_{10}(r_{\text{end}})$ for the absorbing function profiles $\gamma(s) = \{1, s\}$.

2.2.3 Interface reflection

Compared to the continuous problem where r_{end} is the only source of wave reflection, numerical discretization introduces another source of reflection. The behavior of this reflection, assigned the name "interface reflection" $r_{\text{interface}}$ by Bindel [33], is demonstrated through 1D computations of the reflection r_{computed}. The computed reflection r_{computed} contains the contributions of both the end termination reflection r_{end} and the interface reflection $r_{\text{interface}}$. When the length of the PML l_{pml} is long enough, such that the end termination reflection r_{end} is sufficiently smaller than interface reflection $r_{\text{interface}}$, one can assume $r_{\text{interface}} \approx r_{\text{computed}}$. The reflections are computed through dispersion analysis [33, 55, 142]. The bounded domain and PML domain are both discretized by 1D finite elements of varying order n_{eorder} with constant distance h between nodes.

The infinite domain problem, that this model mimics, can be considered an infinite chain of 1D elements with size $(n_{\text{eorder}} + 1) \times h$. Similar to the evaluation of dispersion relationships on lattices in solid state physics [112], one can consider a single element of size $(n_{\text{eorder}} + 1) \times h$ as the unit cell, and compute the eigenmodes.

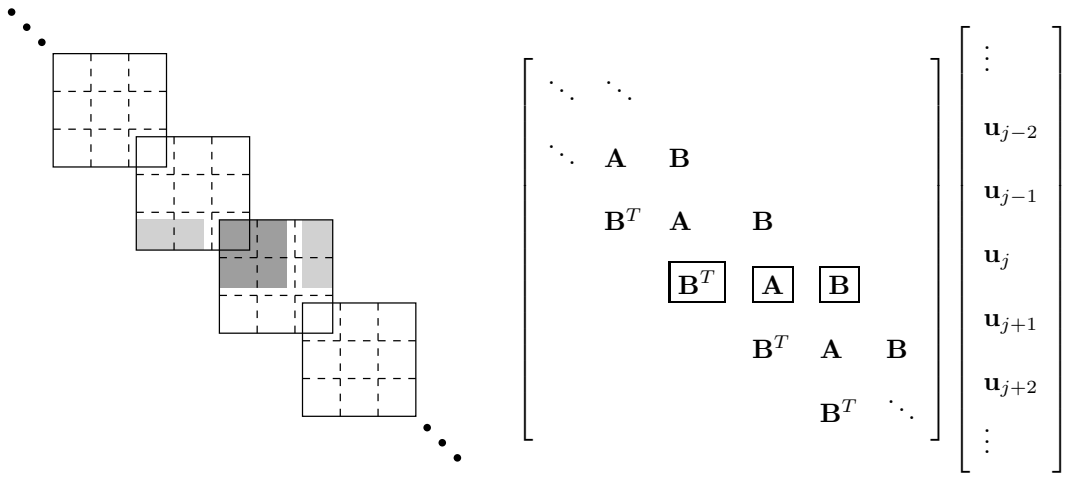


Figure 2.3: Matrix schematic of a 1D chain of quadratic elements. Left diagram denotes the assembled structure and overlap of the individual element matrices, where large boxes denote the individual element matrices. Right diagram denotes the matrix structure. The shaded boxes on the left correspond the boxed matrices on the right

The example for quadratic elements is shown in Figure 2.3. The boxes in the left figure denote

the structured assembly of the individual element matrices with nine entries per element. Each box denotes,

$$\square \quad := \quad \mathbf{k}_{\text{element}} - \omega^2 \mathbf{m}_{\text{element}}, \tag{2.22}$$

where the $\mathbf{k}_{\text{element}}$ and $\mathbf{m}_{\text{element}}$ are the 3-by-3 element stiffness and mass matrices for the quadratic element. Since the elements are connected to its neighbhors one each side by a single node, there is an overlap of one.

The figure on the right shows the repeating structure of the total system with labeled matrices. One has the relation,

$$\mathbf{B}^T \mathbf{u}_{j-1} + \mathbf{A}\mathbf{u}_j + \mathbf{B}\mathbf{u}_{j+1} \quad = \quad 0, \tag{2.23}$$

and assuming a traveling wave solution,

$$\mathbf{u}_j \quad := \quad \hat{\mathbf{u}} \exp(-ikx_j), \tag{2.24}$$

one obtains a quadratic eigenvalue problem for the eigenmodes and wave numbers k representable by the infinite chain,

$$\left(\mathbf{B}^T + ik\mathbf{A} - k^2\mathbf{B}\right) \hat{\mathbf{u}} \quad = \quad 0. \tag{2.25}$$

For this quadratic case, one obtains 4 eigenmodes, 2 traveling wave (acoustic) modes $[\mathbf{v}_{a+}, \mathbf{v}_{a-}]$ with wave numbers $[k_{a+}, k_{a-}]$ moving in opposite directions ($+$ denotes right moving, $-$ denotes left moving), and 2 optical modes $[\mathbf{v}_{o1}, \mathbf{v}_{o2}]$ with wave numbers $[k_{o1}, k_{o2}]$.

To estimate the quality of the PML in approximating the radiation boundary condition at the forced frequency of ω, the finite domain problem terminated by a PML domain is solved,

$$\left(\mathbf{K} - \omega^2 \mathbf{M}\right) \mathbf{x} \quad = \quad \mathbf{F}. \tag{2.26}$$

From the full solution $\mathbf{x}$, the solution corresponding to two cells in the bounded domain are extracted,

$$\mathbf{x} \quad = \quad \begin{bmatrix} \vdots \\ \mathbf{x}_j \\ \mathbf{x}_{j+1} \\ \vdots \end{bmatrix}. \tag{2.27}$$

In this case, cells j and $j+1$ have been selected. The contributions in the direction of the traveling waves are computed by,

$$\begin{pmatrix} c_{a+} \\ c_{a-} \end{pmatrix} \quad := \quad [\mathbf{v}_{a+}, \mathbf{v}_{a+}]^{+} \begin{pmatrix} \mathbf{x}_j \\ \mathbf{x}_{j+1} \end{pmatrix}, \tag{2.28}$$

where $^{+}$ denotes the pseudo-inverse. The reflection is computed as the ratio of the left moving and right moving components,

$$r_{\text{computed}} \quad = \quad \frac{c_{a-}}{c_{a+}}, \tag{2.29}$$

similar to the method in which the end termination reflection r_{end} was defined in Equation (2.14). For the continuous problem, a mode oscillating at the frequency ω must have the wave vector $k = c\omega$. This relation does not hold exactly for a numerically discretized mesh and an error exists between the computed wave number $k_{\text{a+}}$ and the exact wave number k. This relative error is defined as,

$$k_{\text{relerr}} \quad := \quad \frac{|k_{\text{a+}} - k|}{|k|}, \tag{2.30}$$

and represents the discretization error.

Remark: Since this computation involves projection of the solution not onto the continuous modes, but onto the actual discrete modes of the system, the contribution of discretization error is removed from the evaluation of the computed reflection r_{computed}. This is seen in the following computations, where computed reflections several orders smaller in magnitude compared to the discretization error are obtained.

The computed reflection r_{computed} is evaluated for three sets of examples with varying $\beta \in [0, 10]$ and $n_{\text{wpml}} \in [0, 10]$. n_{npw} defines the number of nodes per wave length used in the discretization.

1. Figure 2.4: Vary the discretization $n_{\text{npw}} = \{12, 24, 48, 96\}$. Fix the the order of the element=$\{$linear$\}$ and absorbing function profile $\gamma(s) = s$.

2. Figure 2.5: Vary the order of the element=$\{$linear, quadratic, cubic$\}$. Fix the discretization $n_{\text{npw}} = 12$ and absorbing function profile $\gamma(s) = s$.

3. Figure 2.6: Vary the absorbing function profile $\gamma(s) = \{1, s, s^2, s^3\}$. Fix the order of the element=$\{$linear$\}$ and discretization $n_{\text{npw}} = 12$.

The following observations can be made from the figures.

1. The contours of constant computed reflection seem to consist of two parts, a hyperbola and a curve emanating from the origin $(0, 0)$. Since the hyperbolas arise from the end termination reflection r_{end}, the additional curves emanating from the origin must be occur from the discretization.

2. Figure 2.4: The contours are composed of a hyperbola and a straight line emanating from the origin. Increasing the discretization n_{npw} rotates the error contour curves emanating from the origin counter-clockwise.

3. Figure 2.5: The contours are composed of a hyperbola and a straight line emanating from the origin. Increasing the element order, i.e., the polynomial interpolation, rotates the error contour curves emanating from the origin counter-clockwise.

4. Figure 2.6: The contours are composed of a hyperbola and a curve that seem to have the same shape as the absorbing function profile $\gamma(s)$.

The computed reflection generating the constant contour curves emanating from the origin will be called "interface reflection", $r_{\text{interface}}$. From these observations, one can conclude the following.

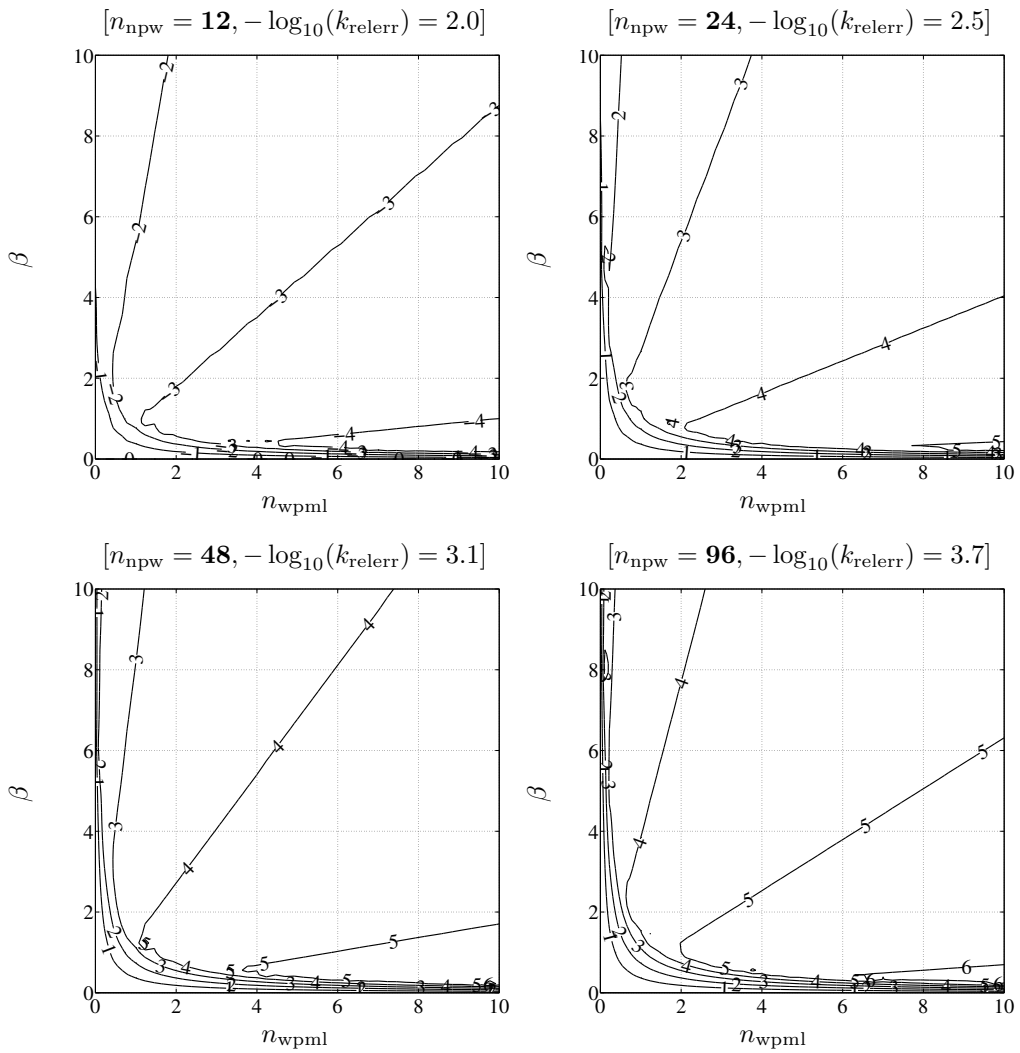


Figure 2.4: Computed reflection $-\log_{10}(r_{\mathrm{computed}})$ for varying discretizations, $n_{\mathrm{npw}} = \{12, 24, 48, 96\}$, fixing the parameters, element=$\{$linear$\}$, $\gamma(s) = \{s\}$.

- Figure 2.4: Fine discretization is required for small computed reflection. As the discretization is made finer, the hyperbolas representing smaller end termination reflection r_{end} appear. The hyperbolas initiate their appearance in the region of small β and large n_{wpml}. Thus one can state that small computed reflection is easier to obtain in the region of small β and large n_{wpml}, i.e., it is better to make the PML longer than making the end absorbing function value large.

  When the discretization is made finer the interface reflection is reduced, resulting in the counter clockwise shift of the constant computed reflection contours. The hyperbolas of smaller end

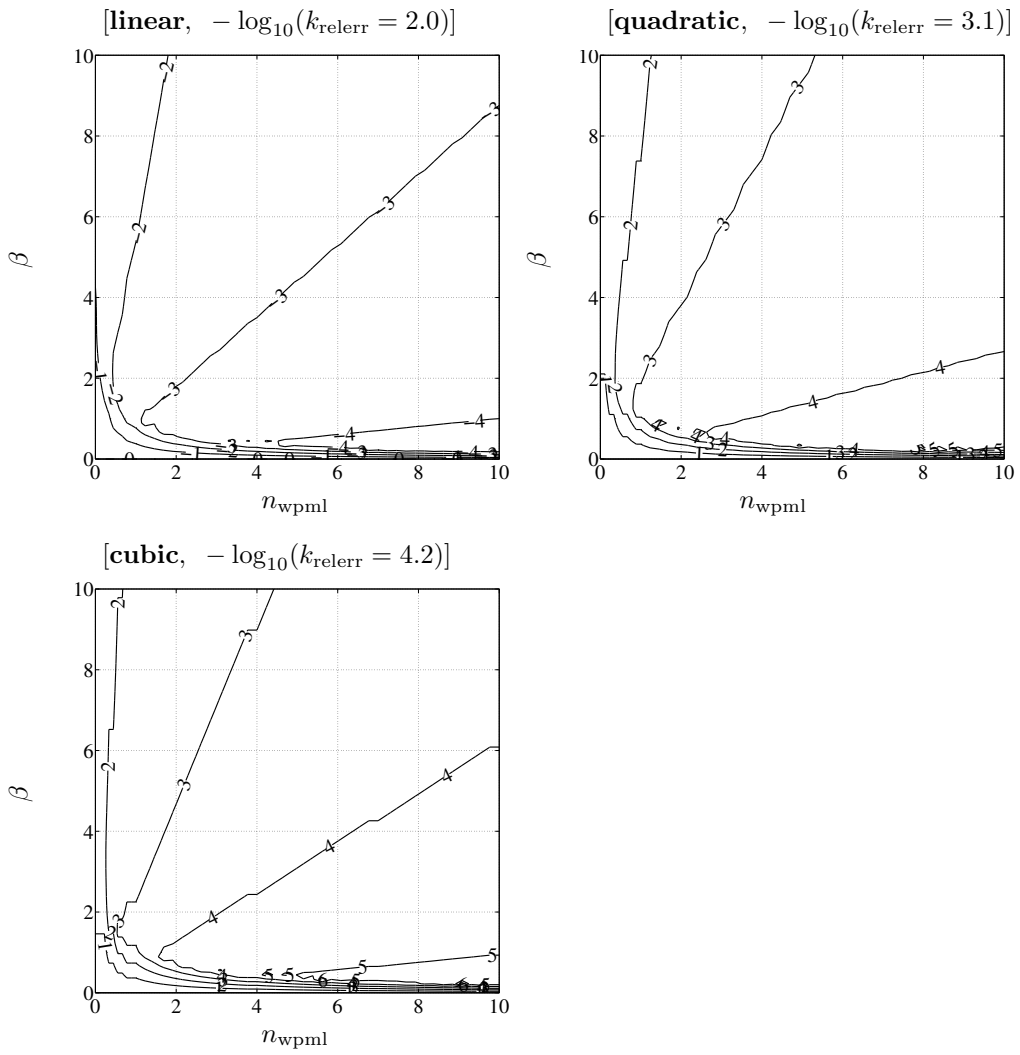


Figure 2.5: Computed reflection $-\log_{10}(r_{\text{computed}})$ for varying element= {linear, quadratic, cubic}, fixing the parameters, $\gamma(s) = \{s\}, n_{\text{npw}} = \{12\}$.

termination reflection appear as soon as $r_{\text{interface}} < r_{\text{end}}$.

- Figure 2.5: Increasing the order of the elements leads to better interpolation of the wave for the same number of nodes per wave and better transition at the bounded domain-PML domain interface, reducing the interface reflection.

  One must also take into account the decrease in discretization error of the wave, represented in k_{relerr}. The computed reflections r_{computed} in these simulations do not include the discretization

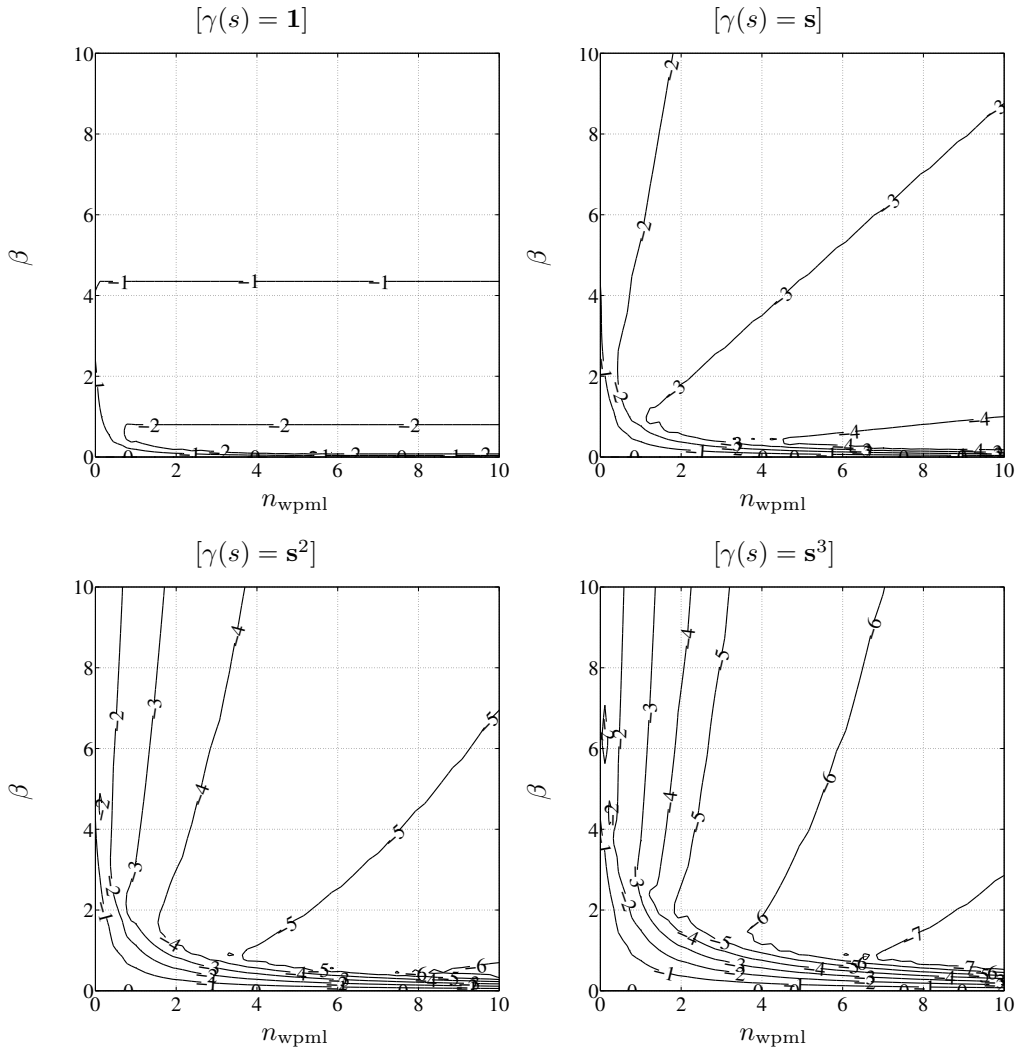


Figure 2.6: Computed reflection $-\log_{10}(r_{\text{computed}})$ for varying absorbing function profiles, $\gamma(s) = \{1, s, s^2, s^3\}$, fixing the parameters, element=$\{$linear$\}$, $n_{\text{npw}} = \{12\}$. Here, $-\log_{10}(k_{\text{relerr}}) = 2.0$.

error, which ultimately defines the accuracy of the simulation. Any computed reflection smaller or equal to the discretization error can be considered the same. Thus when one desires small discretization error for the smallest number of nodes representing a wave, higher order elements are the elements of choice.

- Figure 2.6: Increasing the order of the polynomial in the absorbing function profile leads to smaller computed reflections in the small β and large n_{wpml} regime. But there is a increase in

computed reflection in the small β and small n_{wpml} regime. This is due to a decrease in the value of the integral of $\gamma(s) = 1$ over $[0, 1]$, which defines the distance of the hyperbolas from the origin.

The curve representing the constant contours of interface reflection have exactly the same functional form as the absorbing function $\gamma(x)$, such that $\beta = A\gamma(n_{\text{wpml}})$ for a constant A. This implies that the interface reflection is the same for all pairs (n_{wpml}, β) that satisfy this relation. Should one assume the relation holds even for the case when $n_{\text{wpml}} \to 0$, it is clear that what we had named the interface reflection $r_{\text{interface}}$, truly arises from local behavior at the PML interface. The behavior of the interface reflection near the origin is unfortunately not observable since the end termination reflection dominates in this regime.

As mentioned in the previous item, the computed reflection must be considered in accordance with the discretization error. Numerical reflection on the order of 10^{-7} is shown to be be attainable with the absorbing profile $\gamma(s) = s^3$ for large n_{wpml} and small β, but for this discretization of $n_{\text{npw}} = 12$ with discretization error on the order of 10^{-2}, such values are unnecessary and indistinguishable in reality. Under this claim, one can see that constant PML with $\gamma(s) = 1$ is a valid choice for the absorbing function profile as long as β is selected sufficiently small.

These observations suggest a method to evaluate the interface reflection. Since the contours of constant interface reflection take the same form as the absorbing function profile, let us assume they can be represented as,

$$\beta = c(r_{\text{interface}}, n_{\text{npw}}, \gamma, \text{element})\gamma(n_{\text{wpml}}), \tag{2.31}$$

where $c(r_{\text{interface}}, n_{\text{npw}}, \gamma, \text{element})$ is the scaling coefficient depending on the interface reflection, discretization, choice of absorbing function, and element type. Further assuming the distance between the nodes as h, and a polynomial absorbing function profile $\gamma(s) = s^p$, Equation (2.31) can

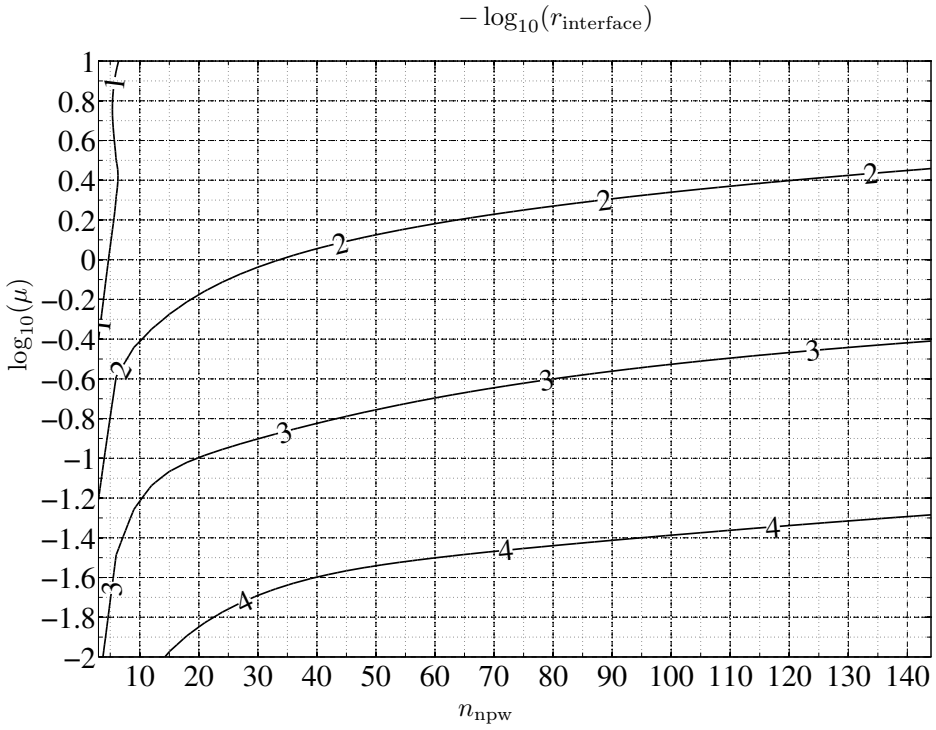


Figure 2.7: Interface reflection $-\log_{10}(r_{\text{interface}})$ due to discretization varying μ and n_{npw} for linear elements with $\gamma(s) = s$

be manipulated into the form,

$$\beta \quad = \quad c(r_{\text{interface}}, n_{\text{npw}}, p, \text{element}) \left(\frac{l_{\text{pml}}}{h \; n_{\text{npw}}} \right)^p. \tag{2.32}$$

Let us now define the quantity μ,

$$\mu \quad := \quad \frac{\beta}{n_{\text{npml}}^p} = \frac{\beta h^p}{l_{\text{pml}}^p} = \frac{c(r_{\text{interface}}, n_{\text{npw}}, p, \text{element})}{n_{\text{npw}}^p}, \tag{2.33}$$

where $n_{\text{npml}} := \frac{l_{\text{pml}}}{h}$ is the number of nodes in the PML. For the case $p = 1$, since β is the value of the absorbing function at the end, μ is the increase in the absorbing function value per node. This is equivalent to the index, damping per element used by Bindel [33].

In Figure 2.7, the contours of constant interface reflection $r_{\text{interface}}$ are plotted with respect to μ and n_{npw}, for linear elements with absorbing function profile $\gamma(s) = s$. The figure is generated with a large n_{wpml} such that end termination reflection r_{end} is negligible, resulting in $r_{\text{interface}} \approx r_{\text{computed}}$.

Table 2.2: Corresponding μ and c for different discretizations n_{npw} for $r_{\text{interface}} = 1 \times 10^{-4}$

n_{npw}	12	24	48	96
$\log_{10}(\mu)$	-2.10	-1.77	-1.56	-1.39
μ	0.00794	0.0170	0.0275	0.0407
$c(r = 1 \times 10^{-4}, n_{\text{npw}}, p = 1, \text{linear})$	0.0958	0.408	1.32	3.91

One can use this figure to reconstruct Figure 2.4 in the following steps.

1. Given the discretization n_{npw} and a specified reflection r, find the μ corresponding to $r_{\text{interface}} = r$. For the selection $r = 1 \times 10^{-4}$, the μ corresponding to each n_{npw} is summarized in Table 2.2.

2. Compute the corresponding $c(r_{\text{interface}}, n_{\text{npw}}, p, \text{element})$ from Equation (2.33) as,

$$c(1 \times 10^{-4}, n_{\text{npw}}, p = 1, \text{linear}) \quad = \quad \mu \times n_{\text{npw}}^{p} \ . \tag{2.34}$$

   The obtained values are also shown in Table 2.2.

3. Plot the interface reflection contour curve from Equation (2.31) along with the end termination reflection hyperbola from Equation (2.21) with respect to $n_{\text{npw}} - \beta$. These are shown in Figure 2.8

By comparing Figure 2.8 with Figure 2.4, one sees that the obtained curves lie right on top of those obtained from the simulations. This method of reconstruction is possible for any combination of element order and absorbing function profile, as long as one can construct the corresponding interface reflection $n_{\text{npw}} - \mu$ diagram.

2.2.4 Optimal parameter estimation

Given the interface reflection $n_{\text{npw}} - \mu$ diagram (e.g., Figure 2.7) for a combination of element type and absorbing function profile, one can obtain an expression for the constant interface reflection contour curve and constant end termination reflection contour hyperbola (e.g., Figure 2.8). From these two curves, one can compute the intersection, which corresponds to the combination of shortest PML $n_{\text{wpml,optimal}}$ and β_{optimal} possible for a desired reflection. Any PML shorter or β larger only

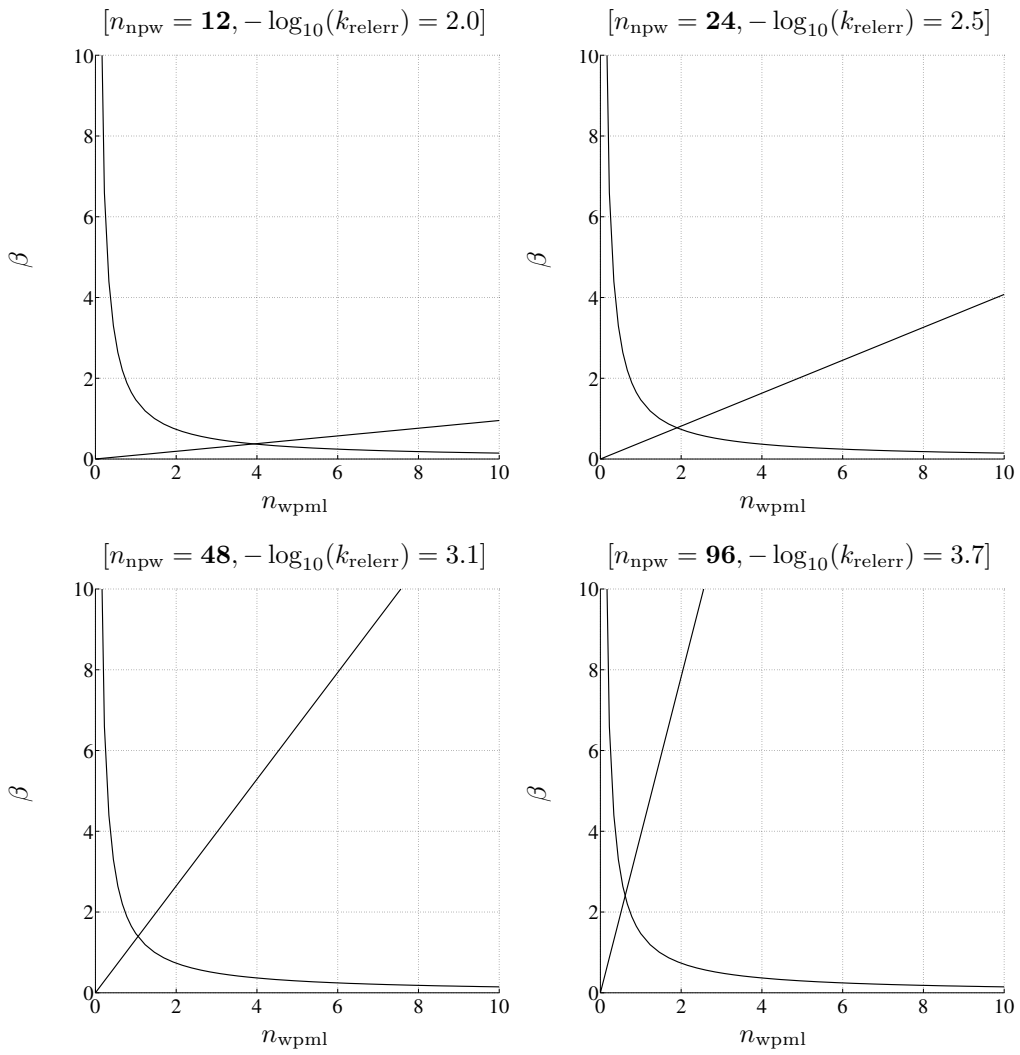


Figure 2.8: Simulated constant contour of reflection equal to $-\log_{10}(r) = 4$, for discretizations $n_{\text{npw}} = \{12, 24, 48, 96\}$, fixing the parameters element=$\{\text{linear}\}$, $\gamma(s) = \{s\}$.

results in an increase in the reflection. What this essentially means is that given a desired reflection, there is a limit to how high one should make the β, and a limit to how short one can make the length of the PML n_{wpml}. From Equation (2.21) and Equation (2.31), the "optimal" parameters can be expressed as,

$$n_{\text{wpml,optimal}} = c(r, n_{\text{npw}}, p, \text{element})^{-\frac{1}{p+1}} \left\{ \frac{-\log(r)}{4\pi}(p+1) \right\}^{\frac{1}{p+1}}, \tag{2.35}$$

$$\beta_{\text{optimal}} = c(r, n_{\text{npw}}, p, \text{element})^{\frac{1}{p+1}} \left\{ \frac{-\log(r)}{4\pi}(p+1) \right\}^{\frac{p}{p+1}}. \tag{2.36}$$

The optimal n_{wpml} and β for linear elements with absorbing function profile $\gamma(s) = s$ is computed and shown in Figure 2.9. The dashed lines represent the contours of constant interface reflection, and the solid lines represent the optimal parameters. The method of reading off the optimal parameters for the absorbing profile $\gamma(s) = s$, using Figure 2.9 is presented in the following example:

1. Assume linear elements, a discretization of $n_{\text{npw}} = 12$, and a desired reflection $r = 1 \times 10^{-4}$.

2. From Figure 2.9, find the corresponding point $(n_{\text{npw}}, r) = (12, 10^{-4})$ assuming that the contours of constant reflection and the grid lines for n_{npw} form a grid.

3. For the obtained point, read off the corresponding optimal parameter, assuming the contours of constant optimal parameter $\{n_{\text{wpml,optimal}}, \beta_{\text{optimal}}\}$ and grid lines for n_{npw} form a grid. The values obtained are, $(n_{\text{wpml,optimal}}, \beta_{\text{optimal}}) = (3.9, 0.4)$. The case of other discretizations is shown in the following table.

n_{npw}	12	24	48
$n_{\text{wpml,optimal}}$	3.9	1.9	1.0
β_{optimal}	0.4	0.8	1.4

From a practical viewpoint, $n_{\text{wpml,optimal}}$ may not be so useful since this does not give an idea of exactly how large the PML layer has to be in terms of numbers of degrees of freedom. The more practical index is,

$$n_{\text{npml,optimal}} := n_{\text{wpml,optimal}} \times n_{\text{npw}}, \tag{2.37}$$

which is the number of nodes in the PML. This index relates directly to the number of unknowns and computational effort required to solve the linear system of equations. The parameter β_{optimal} may also not be so useful for some cases since it is discretization n_{npw} dependent, which may not be ideal for one who is interested in simulating the behavior at several frequencies. Recall from Section 2.2.1, that the absorbing function was defined as,

$$\lambda(s) = 1` - \sigma(s)i \tag{2.38}$$

$$= 1 - \beta\gamma(s)i. \tag{2.39}$$

This expression can be rewritten as,

$$\lambda(s) = 1 - \frac{\beta\omega}{\omega}\gamma(s)i \tag{2.40}$$

$$= 1 - \frac{\bar{\beta}}{\omega}\gamma(s)i, \tag{2.41}$$

$$\bar{\beta} := \beta\omega \tag{2.42}$$

where the parameter $\bar{\beta}$ has been defined. This parameter $\bar{\beta}$ can be shown to be farely insensitive to the discretization n_{npw}.

The values for $n_{\text{npml,optimal}}$ and $\bar{\beta}_{\text{optimal}}$ are computed for linear elements and linear absorbing function profile $\gamma(s) = s$ and presented in Figure 2.10. One observes that the contours of constant optimal parameter run fairly parallel with the contours of constant desired accuracy. This implies that one selection of optimal parameters is valid for a wide range of wave discretizations n_{npw}. It is surprising to see that reflection of magnitude 10^{-2} can be obtained with only 5 nodes (5 linear elements) and $\bar{\beta} = 1$ for the range of discretizations.

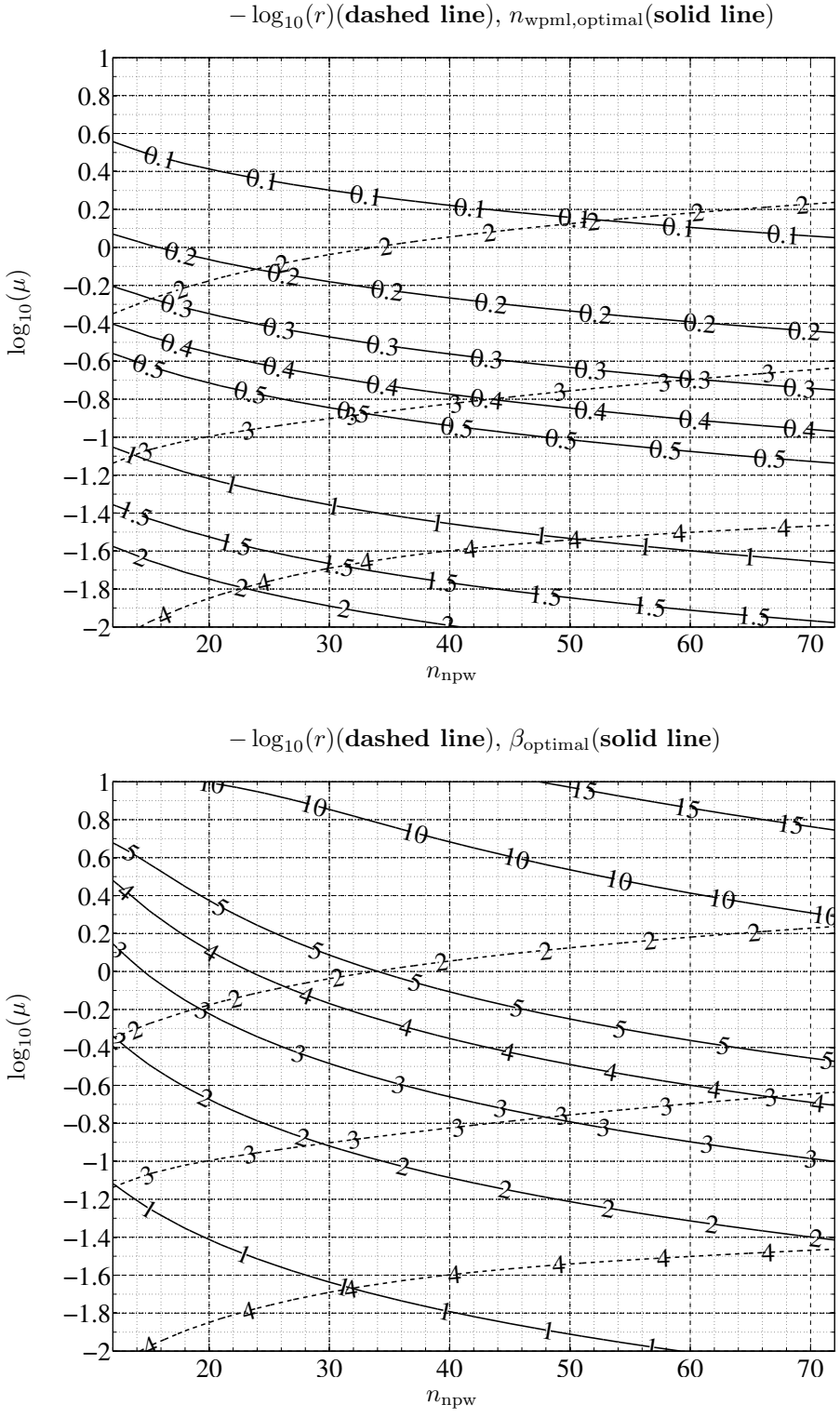


Figure 2.9: Optimal n_{wpml} and β for **linear** elements with $\boldsymbol{\gamma(s) = s}$

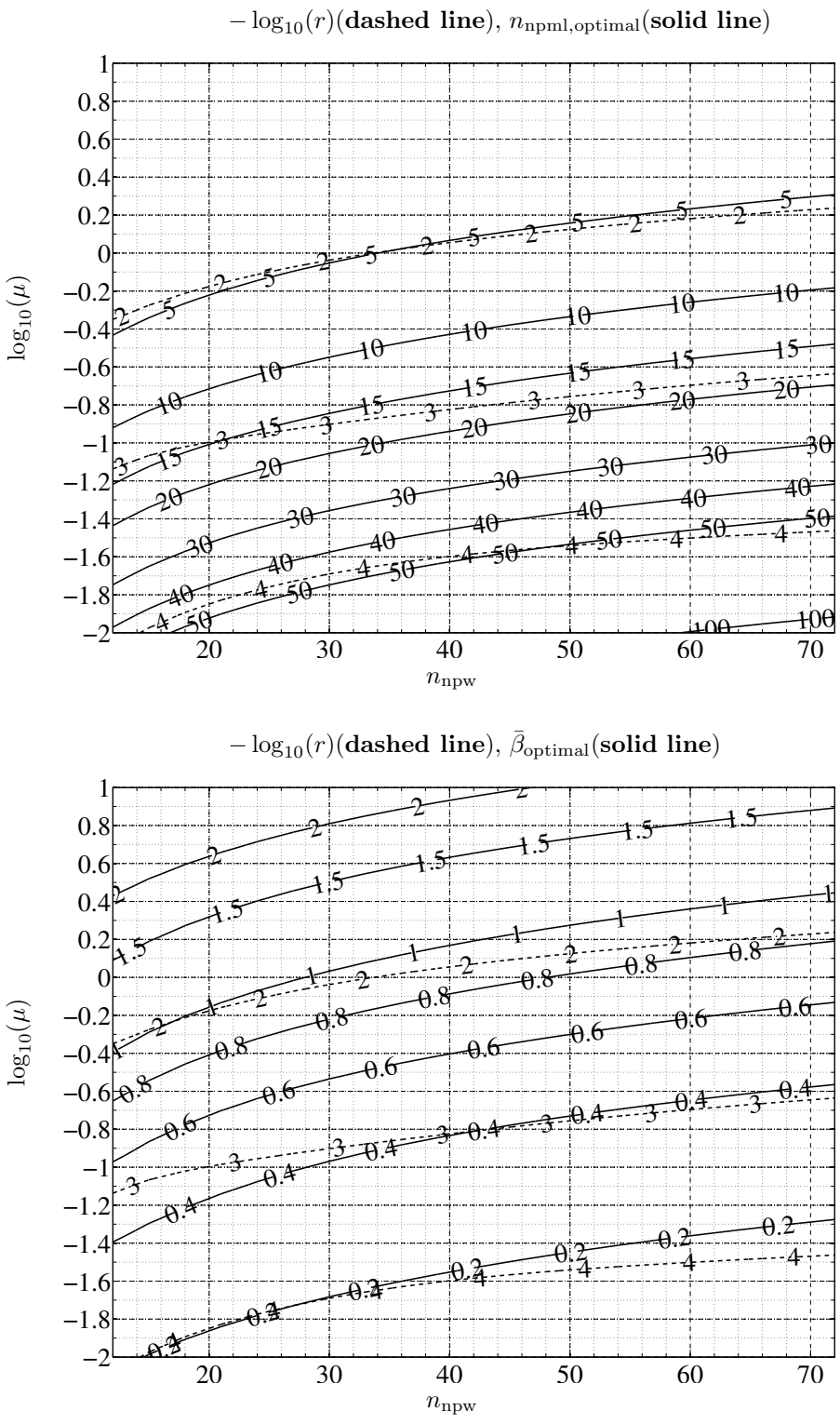


Figure 2.10: Optimal n_{npml} and $\bar{\beta}$ for **linear** elements with $\boldsymbol{\gamma(s) = s}$

2.2.5 Attainable numerical reflection

In the simulations presented in previous sections, the computed reflection r_{computed} was obtained from a method which was insensitive to the discretization error, such that even for a wave discretization error of 10^{-2}, a computed reflection of 10^{-4} is attainable. (See Figure 2.4). This is because the computed reflection is obtained from a projection of the solution onto the discrete modes represented by the discretization.

The discretization error can be measured by the relative error in the wave vector k_{relerr}, defined in Equation (2.30). Computed reflection r_{computed} smaller than k_{relerr} does not lead to better approximations, and thus limits the reflection accuracy one should desire for a given discretization, i.e., number of nodes per wave n_{npw}.

To verify the claim that computed reflection smaller than the discretization error is artificial, the reflection $r_{\text{radiation}}$ is computed with the exact radiation boundary condition enforced at the end instead of the PML. This is possible only in 1D, and the boundary condition applied is,

$$\hat{u}(0) = \bar{u}, \tag{2.43}$$

$$\frac{d\hat{u}}{d\tilde{x}}(L_p) = -ik\hat{u}. \tag{2.44}$$

k_{relerr} and the reflection $r_{\text{radiation}}$ obtained by varying the number of points per wave n_{npw} is shown in in Figures 2.11 and 2.12 in log and linear scale respectively. The horizontal axis of Figure 2.12 coincides with the figures for optimal parameter estimation for ease of comparison.

Note that even though we use an exact boundary condition, there still are incoming parts to the numerical solution which result in a reflection on the same order as the discretization error k_{relerr}, and no smaller. The rate of convergence for the various element orders follow the asymptotic convergence rates of the finite elements, namely order 2 for linear, order 4 for quadratic, and order 6 for cubic. From this it is clear that in order to obtain highly accurate infinite domain solutions with only outgoing solution components, besides applying correct absorbing boundary conditions, the discretization must be made fine enough.

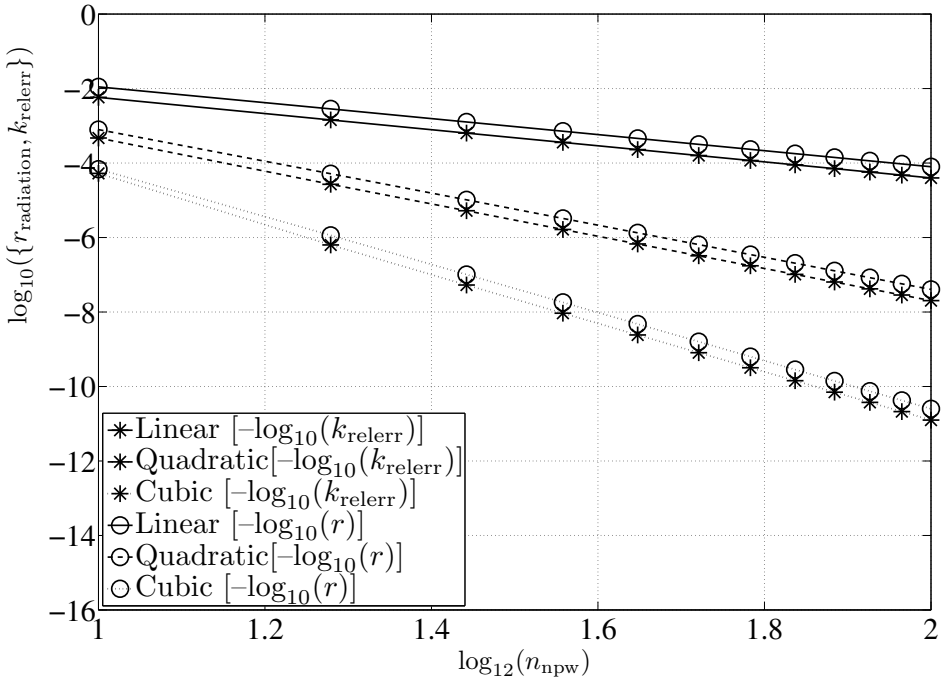


Figure 2.11: Relative error in k and reflection $r_{\text{radiation}}$ under the exact radiation boundary condition with respect to number of nodes per wave length

Thus an added heuristic in selecting the appropriate PML parameters is the following.

- Given n_{npw}, first determine the accuracy of the wave discretization by Figure 2.11. Then select the optimal parameters for a reflection r on the same order or slightly smaller.

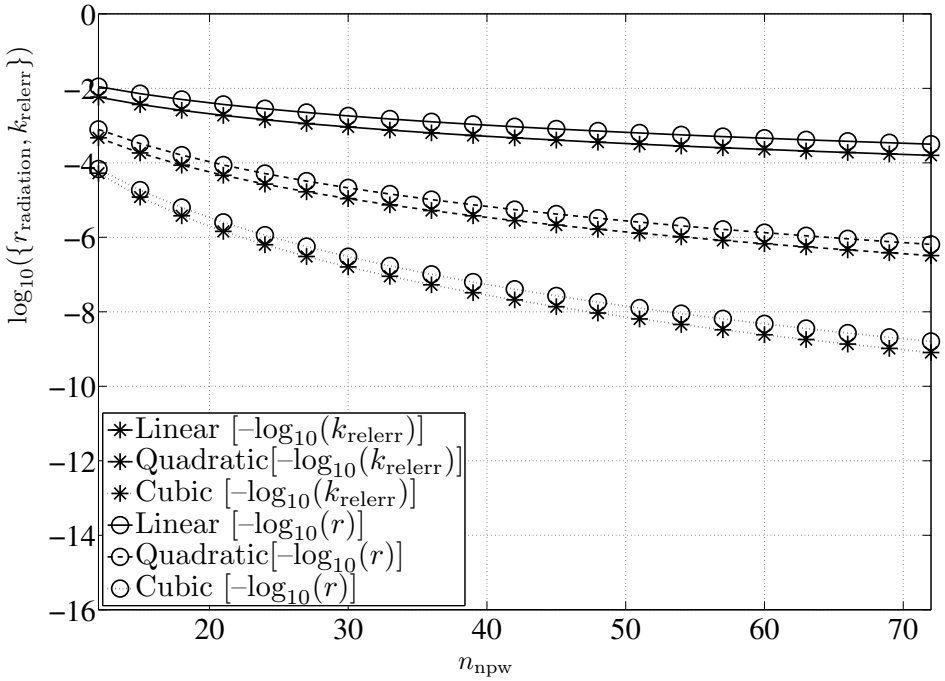


Figure 2.12: Relative error in k and reflection $r_{\mathrm{radiation}}$ under the exact radiation boundary condition with respect to number of nodes per wave length

2.2.6 Energy dissipation error

The discrete reflection r_{computed} is a convenient measure to evaluate the effectiveness of the PML in 1D, but difficult to generalize in multi-dimensions. Another measure based on an energy approach is introduced to evaluate the effectiveness of absorbing boundary conditions in multi-dimensions. The approach is similar to the definition of Q, in which the energy dissipated per radian under a time-harmonic excitation is considered.

Given the exact energy dissipation per cycle of the system with an infinite domain under a time-harmonic motion E_{infinite}, the effectiveness of the PML can be determined by computing the energy dissipated per cycle E_{pml}, and evaluating the relative error,

$$E_{\text{relerr}} := \frac{|E_{\text{infinite}} - E_{\text{pml}}|}{|E_{\text{pml}}|}. \tag{2.45}$$

This index will be called the "energy dissipation error" index. In the 1D scalar wave case, E_{relerr} can be shown to be equivalent to the discrete reflection r_{computed} defined in Equation 2.29. The equivalence can also be shown by the resemblance of the plots obtained from the energy dissipation error varying β and n_{wpml}, with those of the discrete reflection r_{computed}.

First, we will provide a proof of this claim and then look at heuristics based upon it. In particular, we will take this understanding and apply it to the 2D scalar wave and 2D elastodynamic cases.

1D scalar wave

Consider the 1D scalar wave problem introduced in Section 2.2.1. To sustain harmonic motion, finite energy must be constantly pumped into the system per cycle at the forced displacement boundary $x = 0$. For a fixed boundary condition at $x = L_p$ without PML, there is no energy dissipation, and the energy dissipated per cycle is zero. With application of PML, the energy dissipated in the continuous system is computed as [111],

$$E_{\text{pml}} = -\oint \text{Re}(\sigma(t,0))\text{Re}(v(t,0))dt, \tag{2.46}$$

where,

$$\begin{aligned}
u(t,\tilde{x}) &= \hat{u}(\tilde{x})e^{i\omega t}, \\
\hat{u}(\tilde{x}) &= c_{\text{out}}e^{-ik\tilde{x}} + c_{\text{in}}e^{ik\tilde{x}}, \\
v(t,\tilde{x}) &= \frac{\partial u}{\partial t}, \\
&= \hat{v}(\tilde{x})e^{i\omega t}, \\
\hat{v}(\tilde{x}) &= i\omega\hat{u}(\tilde{x}), \\
\sigma(t,\tilde{x}) &:= c\frac{\partial u}{\partial \tilde{x}}, \\
&= \hat{\sigma}(\tilde{x})e^{i\omega t}, \\
\hat{\sigma}(\tilde{x}) &:= ikc\left(-c_{\text{out}}e^{-ik\tilde{x}} + c_{\text{in}}e^{ik\tilde{x}}\right),
\end{aligned}$$

and σ is a stress or force like quantity. By defining the complex-valued reflection as,

$$R := \frac{c_{\text{in}}}{c_{\text{out}}}, \tag{2.47}$$

and using the boundary condition $\tilde{u}(x=0) = \bar{u} \in \mathbb{R}$, one obtains,

$$\hat{\sigma}(x=0) = -ikc\bar{u}\frac{1-R}{1+R}. \tag{2.48}$$

Computation of E_{pml} yields,

$$\begin{aligned}
E_{\text{pml}} &= \frac{T}{2}\text{Re}\left\{\text{Conj}(\hat{\sigma}(0))\hat{v}(0)\right\} \\
&= \pi kc\bar{u}^2\text{Re}\left(\frac{1-R}{1+R}\right), \qquad (2.49)
\end{aligned}$$

where $T := \frac{2\pi}{\omega}$ is the period and Conj denotes complex conjugation. E_{infinite} is obtained by setting $R = 0$ in E_{pml},

$$E_{\text{infinite}} = \pi kc\bar{u}^2, \tag{2.50}$$

and thus,

$$\begin{aligned}
E_{\text{relerr}} &= \text{Re}\left(\frac{2R}{1+R}\right) \\
&\approx 2\text{Re}(R) \quad (\text{for } |R| << 1). \qquad (2.51)
\end{aligned}$$

It is clear that when the reflection R is small, E_{relerr} is equivalent to r_{computed}.

For the discrete case, E_{pml} is computed as,

$$E_{\text{pml,computed}} := \frac{T}{2}\text{Re}\{\text{Conj}(\mathbf{F}(1))i\omega\mathbf{U}(1)\}, \tag{2.52}$$

where $\mathbf{F}$ and $\mathbf{U}$ are the discrete vector of forces and nodal displacements obtained from the computation, and 1 denotes the index of the forced end.

The energy dissipation error is computed in Figure 2.13 for the same case shown in Figure 2.4. One observes the smae trends as the discrete reflection. Note that only contours of constant error up to the order of the discretization are depicted. This is because the computation of $E_{\text{pml,computed}}$ includes contribution from discretization error.

This 1D scalar example motivates using the energy dissipation error as an alternative index to the discrete reflection r_{computed}.

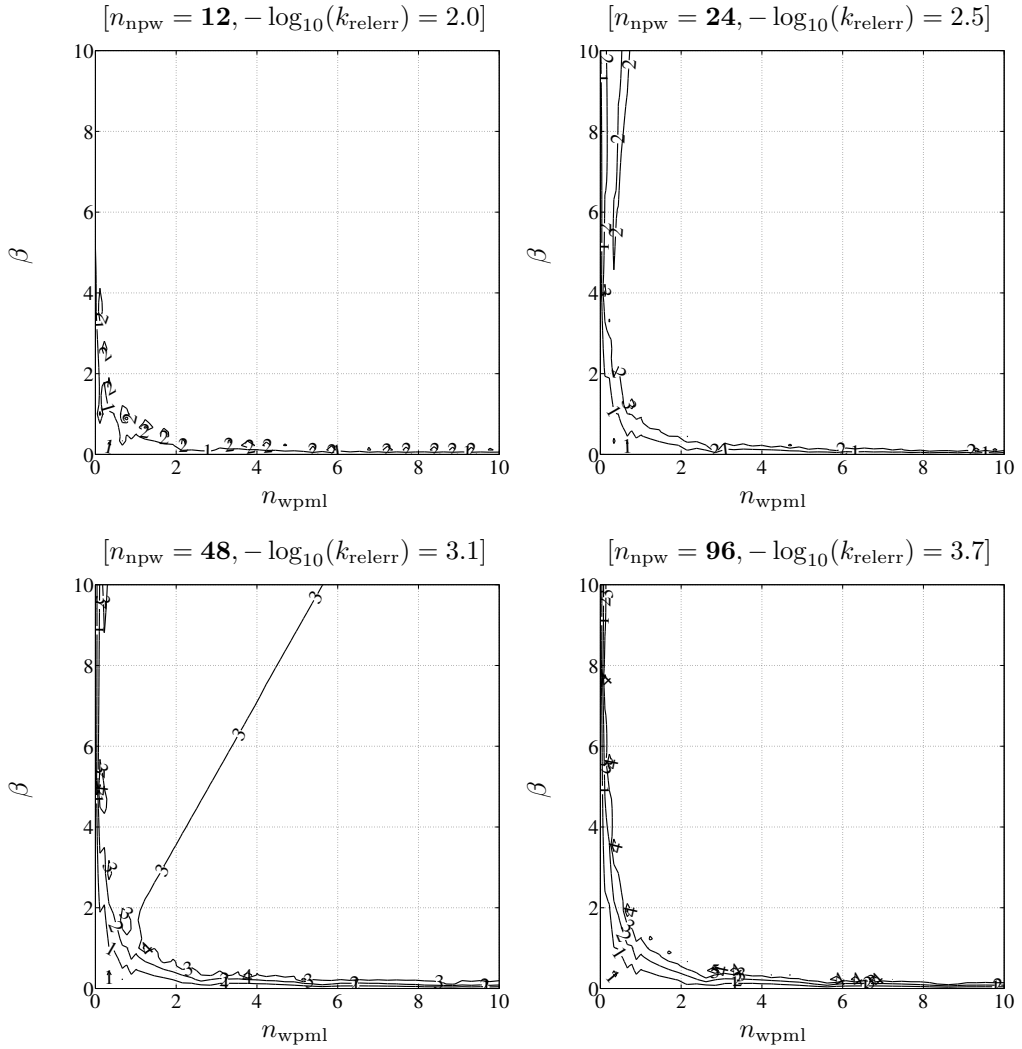


Figure 2.13: Energy dissipation error $-\log_{10}(E_{\mathrm{relerr}})$ for varying discretizations $n_{\mathrm{npw}} = \{12, 24, 48, 96\}$ fixing the parameters element=$\{$linear elements$\}$, $\gamma(s) = \{s\}$.

2D scalar wave

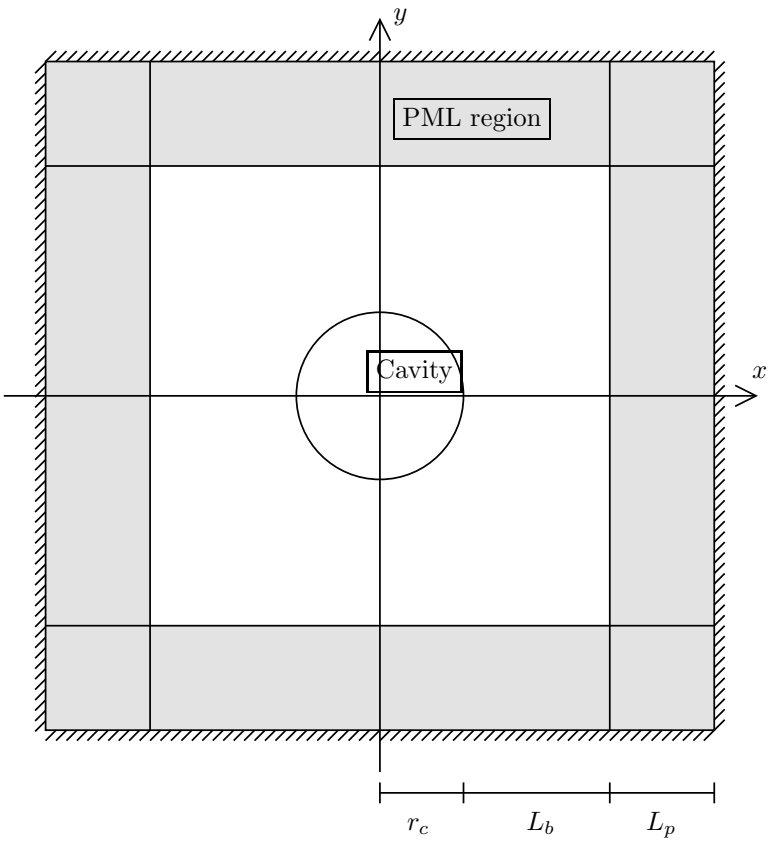


Figure 2.14: 2D scalar wave configuration

To verify the observations made in the 1D scalar problem and to check the validity of the optimal PML parameter selection for multi-dimensional scalar problems, the 2D scalar wave equation is

investigated. The model problem is the 2D membrane [84],

$$\rho\frac{\partial^2 u}{\partial t^2} - \mathrm{S}\left(\frac{\partial^2 u}{\partial \tilde{x}^2} + \frac{\partial^2 u}{\partial \tilde{y}^2}\right) = 0, \quad (x,y) \in [-L_d, L_d] \times [-L_d, L_d] \setminus \Omega_0, \tag{2.53}$$

$$\Omega_0 := \left\{(x,y) | x^2 + y^2 < r_c^2\right\}, \tag{2.54}$$

$$L_d := r_c + L_b + L_p, \tag{2.55}$$

$$\tilde{x} = \int_0^x \lambda(s) ds, \tag{2.56}$$

$$\tilde{y} = \int_0^y \lambda(s) ds, \tag{2.57}$$

$$\lambda(s) = 1 - \sigma(s) i, \tag{2.58}$$

$$\sigma(s) = \begin{cases} 0 & 0 \le s < r_c + L_b \\ \text{some value} & r_c + L_b \le s \le L_d \end{cases}. \tag{2.59}$$

where ρ is the density, S is the given tension, and again $\tilde{x}$ and $\tilde{y}$ are the stretched coordinates. Under time harmonic assumptions, one obtains the 2D Helmholtz equation,

$$\frac{d^2\hat{u}}{d\tilde{x}^2} + \frac{d^2\hat{u}}{d\tilde{y}^2} + k^2\hat{u} = 0, \tag{2.60}$$

where $c := \sqrt{\dfrac{\mathrm{T}}{\rho}}$ is the wave speed and $k := \dfrac{\omega}{c}$ is the wave number. Solutions of this equation in the non-PML part of the domain take the form [84],

$$\hat{u}(r) = c_{\text{out}}\mathrm{H}_0^{(2)}(kr) + c_{\text{in}}\mathrm{H}_0^{(1)}(kr), \tag{2.61}$$

$$r := \sqrt{x^2 + y^2}, \tag{2.62}$$

where H_0^1 and H_0^2 are the Hankel functions, and c_{out} and c_{in} are constants determined by the boundary condition. $\mathrm{H}_0^{(2)}$ represents the outgoing solution from the origin, and $\mathrm{H}_0^{(1)}$ represents the incoming solution from infinity. For infinite domain outgoing wave propagation behavior, one desires $c_{\text{in}} = 0$.

The boundary conditions applied are the following,

$$\hat{u}(x,y) = \bar{u} \qquad \text{for } (x^2 + y^2 = r_c^2), \tag{2.63}$$

and additionally,

$$\hat{u}(x,y) = 0 \qquad \text{for } (|x| = L_p \text{ or } |y| = L_p). \tag{2.64}$$

for the PML problem.

E_{infinite}, the energy dissipated per cycle for the continuous infinite domain problem, is computed from assuming only outgoing wave solutions of the form,

$$\hat{u}(x,y) = A\mathrm{H}_0^{(2)}(kr), \tag{2.65}$$

with A determined by the boundary condition as,

$$A = \frac{\bar{u}}{\mathrm{H}_0^{(2)}(kr_c)}. \tag{2.66}$$

E_{infinite} is computed as,

$$E_{\text{infinite}} = -\oint \mathrm{Re}(q(t,r_c))\mathrm{Re}(v(t,r_c))dt, \tag{2.67}$$

where,

$$\begin{aligned}
v(t,r) &= \frac{\partial u}{\partial t}, \\
&= \hat{v}(r)e^{i\omega t}, \\
\hat{v}(r) &= i\omega\hat{u}(r), \\
q(t,r) &:= \mathrm{S}\frac{\partial u}{\partial r} \times 2\pi r, \\
&= \hat{q}(r)e^{i\omega t}, \\
\hat{q}(r) &:= -k\mathrm{S}A\mathrm{H}_1^{(2)}(kr) \times 2\pi r.
\end{aligned}$$

$q(t,r)$ is the vertical force acting on the membrane at the circle of radius r. Given these expressions,

$$\begin{aligned}
E_{\text{infinite}} &= \frac{T}{2}\mathrm{Re}\left\{\mathrm{Conj}(\hat{q}(r_c))\hat{v}(r_c)\right\} \\
&= 2T\omega\mathrm{S}|A|^2 \ ,
\end{aligned} \tag{2.68}$$

where $T := \frac{2\pi}{\omega}$ is the period. For the discrete case, $E_{\text{pml,computed}}$ is computed as,

$$E_{\text{pml,computed}} := \frac{T}{2}\mathrm{Re}\{\mathbf{F}(\text{nodes on } r_c)^* i\omega\mathbf{U}(\text{nodes on } r_c)\}, \tag{2.69}$$

where $\mathbf{F}$ and $\mathbf{U}$ are the discrete vector of forces and nodal displacements obtained from the computation, (nodes on r_c) denotes the set of indices corresponding to degrees of freedom on the circle of radius r_c, and * denotes conjugate transposition of the vector.

In computing the energy dissipation error E_{relerr}, one must adequately discretize the forced displacement boundary condition and mesh to obtain small energy dissipation error. This is different from the 1D case, in which boundary conditions can be enforced exactly.

A sample mesh used in the computation and sample solution of wave propagation is shown in Figure 2.15. The red denotes positive displacement and blue denotes negative displacement. One can visually see the effectiveness of the PML in approximating the infinite domain boundary condition.

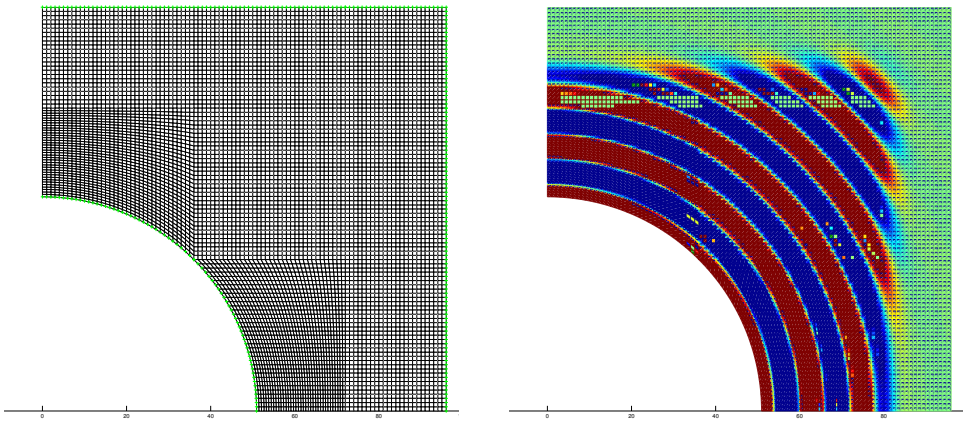


Figure 2.15: 2D scalar wave sample discretized mesh and wave motion

The energy dissipation error is computed for 4 cases:

1. Figure 2.16 left: Linear elements. Absorbing function profile fixed $\gamma(s) = s$. The number of nodes per wave is different for each plot $n_{\text{npw}} = \{12, 24, 48\}$. Within each plot the number of waves in the pml $n_{\text{wpml}} \in [0, 3]$ and $\beta \in [0, 3]$ is varied.

2. Figure 2.16 right: Linear elements. Absorbing function profile fixed $\gamma(s) = 1$. The number of nodes per wave is different for each plot $n_{\text{npw}} = \{12, 24, 48\}$. Within each plot the number of

waves in the pml $n_{\text{wpml}} \in [0, 3]$ and $\beta \in [0, 3]$ is varied.

3. Figure 2.17 left: Cubic elements. Absorbing function profile fixed $\gamma(s) = s$. The number of nodes per wave is different for each plot $n_{\text{npw}} = \{12, 24, 48\}$. Within each plot the number of waves in the pml $n_{\text{wpml}} \in [0, 3]$ and $\beta \in [0, 3]$ is varied.

4. Figure 2.17 right: Cubic elements. Absorbing function profile fixed $\gamma(s) = 1$. The number of nodes per wave is different for each plot $n_{\text{npw}} = \{12, 24, 48\}$. Within each plot the number of waves in the pml $n_{\text{wpml}} \in [0, 3]$ and $\beta \in [0, 3]$ is varied.

The following observations can be made.

- The contours of constant energy dissipation error have exactly the same behavior as observed for the 1D scalar energy dissipation error (Figure 2.13) and the 1D scalar computed reflection (Figure 2.4).

- The constant PML with absorbing function profile $\gamma(s) = 1$ works as well as the linear profile, with smaller energy dissipation error for all combinations of (n_{wpml}, β).

- Given the same discretization, cubic elements are capable of obtaining energy dissipation error two orders of magnitude smaller than linear elements.

Given these results, one can have some confidence in applying the heuristics developed for the 1D scalar wave to the 2D scalar valued case.

Remark: The results show surprisingly good results for the constant PML. This behavior follows from the explanation given in Section 2.2.2. One should not, however, assume that any complex-valued material will behave this well. The success of a constant PML lies in the anisotropic property and the addition of complex-valuedness in the mass matrix. Making the PML stretch the same in all of the PML domain leads to unphysical results.

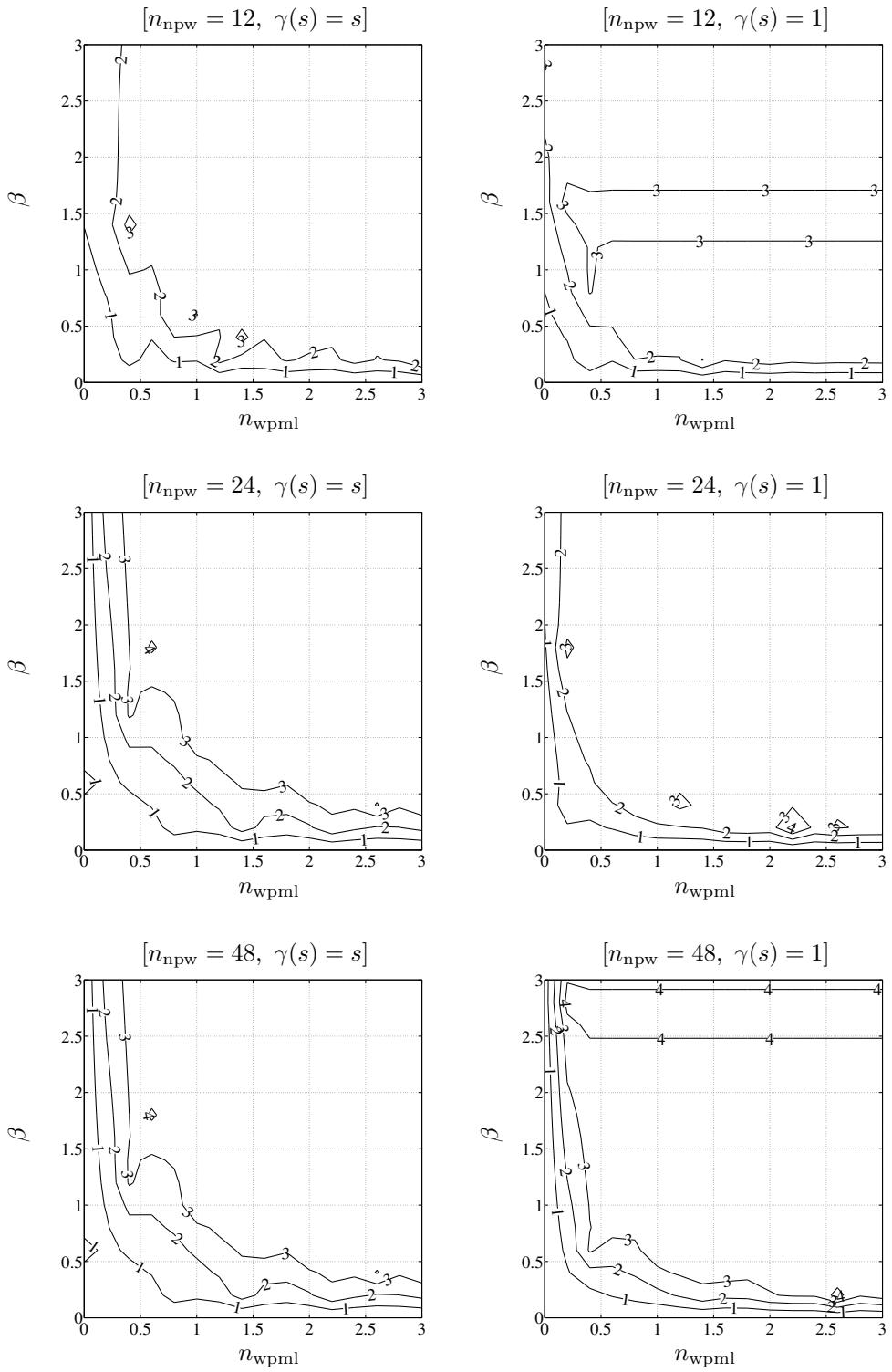


Figure 2.16: Energy dissipation error $-\log_{10}(E_{\text{relerr}})$ for the 2D scalar case with varying discretizations $n_{\text{npw}} = \{12, 24, 48\}$, varying absorbing function profiles $\gamma(s) = \{1, s\}$, fixing the parameters element=$\{\textbf{linear}\}$.

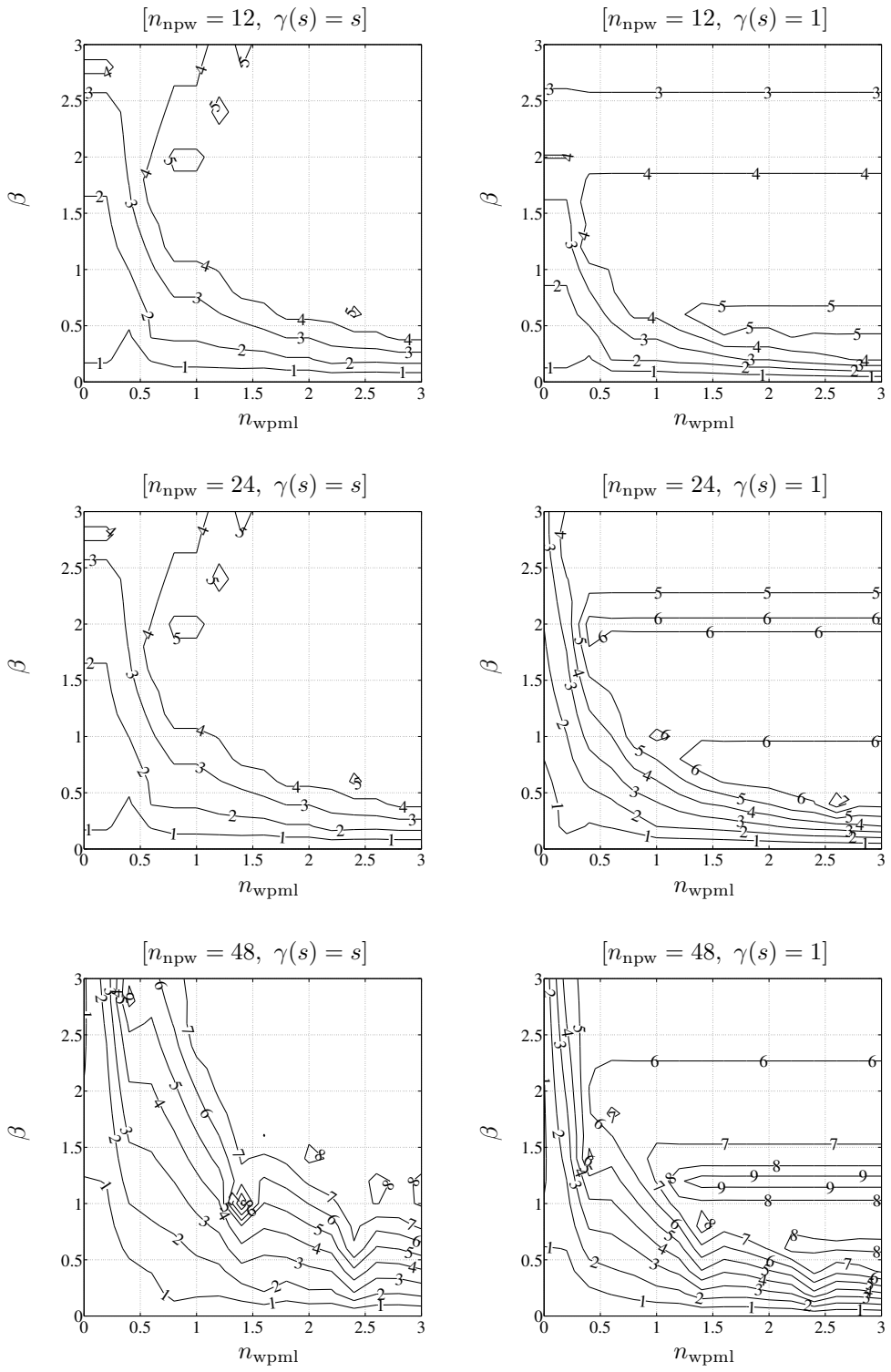


Figure 2.17: Energy dissipation error $-\log_{10}(E_{\text{relerr}})$ for the 2D scalar case with varying discretizations $n_{\text{npw}} = \{12, 24, 48\}$, varying absorbing function profiles $\gamma(s) = \{1, s\}$, fixing the parameters element=$\{$**cubic**$\}$.

2D elastodynamics

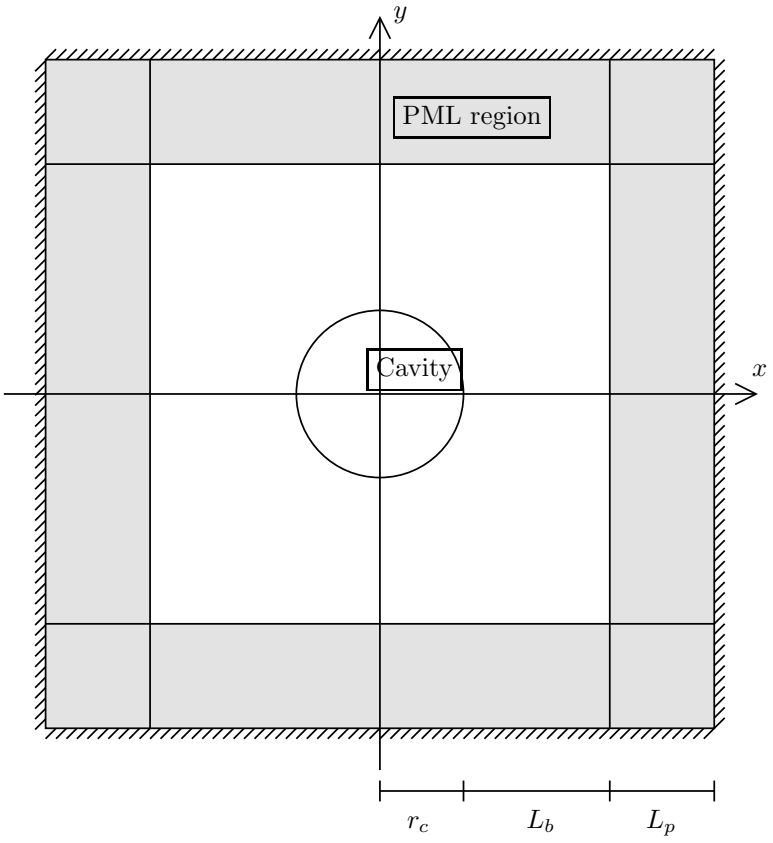


Figure 2.18: 2D elastic wave configuration

To verify the observations made in the 1D scalar problem, and check the validity of the optimal PML parameter selection for multi-dimensional vector-valued problems, the 2D elastodynamic problem is considered. Harmonic volumetric and shear waves propagating from an infinitely extending cylindrical cavity, can be treated as a 2D plane strain problem [84]. The governing equations for elastodynamics in the absence of body forces can be given in Navier's form as,

$$(\lambda + \mu)\boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \mathbf{u}) + \mu\boldsymbol{\nabla}^2\mathbf{u} \quad = \quad \rho\ddot{\mathbf{u}}, \tag{2.70}$$

where λ, μ are the Lame constants, ρ is the density, and $\mathbf{u}$ is the displacement vector. This equation

is solved on the domain,

$$(x,y) \quad \in \quad [-L_d, L_d] \times [-L_d, L_d] \setminus \Omega_0, \tag{2.71}$$

$$\Omega_0 \quad := \quad \left\{(x,y) | x^2 + y^2 < r_c^2\right\}, \tag{2.72}$$

$$L_d \quad := \quad r_c + L_b + L_p. \tag{2.73}$$

Introducing scalar and vector potentials $\Psi, \mathbf{H}$,

$$\mathbf{u} \quad = \quad \boldsymbol{\nabla}\Phi + \boldsymbol{\nabla} \times \mathbf{H}, \quad \boldsymbol{\nabla} \cdot \mathbf{H} = 0, \tag{2.74}$$

and assuming plane strain axisymmetry,

$$u_z = \frac{\partial}{\partial \theta} = \frac{\partial}{\partial z} = 0, \tag{2.75}$$

one obtains,

$$\mathbf{u} \quad = \quad u_r \mathbf{e}_r + u_\theta \mathbf{e}_\theta, \tag{2.76}$$

$$u_r \quad = \quad \frac{\partial \Phi}{\partial r}, \tag{2.77}$$

$$u_\theta \quad = \quad -\frac{\partial H_z}{\partial r}, \tag{2.78}$$

$$\sigma_{rr} \quad = \quad (\lambda + 2\mu)\left(\frac{\partial^2 \Phi}{\partial r^2} + \frac{1}{r}\frac{\partial \Phi}{\partial r}\right) - \frac{2\mu}{r}\frac{\partial \Phi}{\partial r}, \tag{2.79}$$

$$\sigma_{r\theta} \quad = \quad -\mu\left(\frac{\partial^2 H_z}{\partial r^2} - \frac{1}{r}\frac{\partial H_z}{\partial r}\right), \tag{2.80}$$

where σ is the stress. The equations governing the volumetric and shear waves are,

$$\boldsymbol{\nabla}^2 \Phi \quad = \quad \frac{1}{c_v^2}\frac{\partial^2 \Phi}{\partial t^2}, \tag{2.81}$$

$$\boldsymbol{\nabla}^2 H_z \quad = \quad \frac{1}{c_s^2}\frac{\partial^2 H_z}{\partial t^2}, \tag{2.82}$$

where $c_v := \dfrac{\lambda + 2\mu}{\rho}$ and $c_s := \dfrac{\mu}{\rho}$ are the volumetric and shear wave speed. Under time-harmonic assumptions, Equations (2.81) and (2.82) both take the form of Helmholtz equations,

$$\boldsymbol{\nabla}^2 \hat{\Phi} \quad = \quad k_v^2 \hat{\Phi}, \tag{2.83}$$

$$\boldsymbol{\nabla}^2 \hat{H}_z \quad = \quad k_s^2 \hat{H}_z, \tag{2.84}$$

where $k_v := \dfrac{\omega}{c_v}$ and $k_s := \dfrac{\omega}{c_s}$ are wave number like quantities. The solution to both equations take the form of Hankel functions with respect to r, leading to,

$$\hat{\Phi}(r) = A_v \mathrm{H}_0^{(2)}(k_v r), \tag{2.85}$$

$$\hat{H}_z(r) = A_s \mathrm{H}_0^{(2)}(k_s r), \tag{2.86}$$

$$\tag{2.87}$$

for the outgoing wave solution, where $A_v,\ A_s$ are constants determined by boundary conditions. Applying displacement boundary conditions at the cavity $r = r_c$,

$$u_r(r_c) = \bar{u}_r \tag{2.88}$$

$$u_\theta(r_c) = \bar{u}_\theta \tag{2.89}$$

$$\tag{2.90}$$

yields,

$$A_v = -\frac{\bar{u}_v}{k_v \mathrm{H}_1^2(k_v r_c)}, \tag{2.91}$$

$$A_s = \frac{\bar{u}_\theta}{k_v \mathrm{H}_1^2(k_s r_c)}. \tag{2.92}$$

Given these expressions, the stress can be computed and the energy dissipation for the infinite domain under time-harmonic excitation can be computed similar to the previous sections as,

$$E_{v,\text{infinite}} = T\omega \frac{2(\lambda + 2\mu)|\bar{u}_r|^2}{|H_1^{(2)}(k_v r_c)|^2}, \tag{2.93}$$

$$E_{s,\text{infinite}} = T\omega \frac{2\mu|\bar{u}_\theta|^2}{|H_1^{(2)}(k_s r_c)|^2}, \tag{2.94}$$

where $T := \frac{2\pi}{\omega}$ is the period. For the discrete case, $E_{\text{pml,computed}}$ is computed as,

$$E_{\text{pml,computed}} := \frac{T}{2}\mathrm{Re}\{\mathbf{F}(\text{nodes on } r_c)^* i\omega \mathbf{U}(\text{nodes on } r_c)\}, \tag{2.95}$$

where $\mathbf{F}$ and $\mathbf{U}$ are the discrete vector of forces and nodal displacements obtained from the computation, (nodes on r_c) denotes the set of indices corresponding to degrees of freedom on the circle of radius r_c, and * denotes conjugate transposition of the vector.

In computing the energy dissipation error E_{relerr}, one must adequately discretize the forced boundary condition and mesh to obtain small energy dissipation error. This is different from the 1D case, in which boundary conditions can be enforced exactly.

For the elastodynamic time-harmonic problem, two different waves propagate through the medium compared to one in the scalar problem. These two waves are the volumetric wave and shear wave. Under time-harmonic forcing at a frequency of ω, they have wave length of,

$$\lambda_v = \frac{2\pi}{k_v} = \frac{2\pi c_v}{\omega}, \tag{2.96}$$

$$\lambda_s = \frac{2\pi}{k_s} = \frac{2\pi c_s}{\omega}. \tag{2.97}$$

The ratio between the volumetric and shear speed in terms of the Poisson ratio ν is,

$$\kappa_{vs} := \frac{c_v}{c_s} = \sqrt{\frac{2-2\nu}{1-2\nu}} \geq \sqrt{2}, \quad (0.5 \geq \nu \geq 0). \tag{2.98}$$

Because the wave lengths differ, given a mesh discretization, the number of nodes per wave length and the number of waves in the PML for the two differ. Since,

$$n_{\text{npw,v}} = \kappa_{vs} n_{\text{npw,s}}, \tag{2.99}$$

$$n_{\text{wpml,v}} = \frac{1}{\kappa_{vs}} n_{\text{wpml,s}}, \tag{2.100}$$

for sufficient discretization of waves, one must place enough nodes per shear wave length λ_s, and for sufficient PML length, one must adjust to the volumetric wave length λ_v.

The energy dissipation error is computed for 8 cases. In the first 4 cases, volumetric waves are excited and in the latter 4 cases, shear waves are excited.

1. Figure 2.19 left: Volumetric waves. Linear elements. Absorbing function profile fixed $\gamma(s) = s$. The number of nodes per wave is different for each plot $n_{\text{npw,v}} = \{12, 24, 48\}$. Within each plot the number of waves in the pml $n_{\text{wpml,v}} \in [0, 3]$ and $\beta \in [0, 3]$ is varied.

2. Figure 2.19 right: Volumetric waves. Linear elements. Absorbing function profile fixed $\gamma(s) = 1$. The number of nodes per wave is different for each plot $n_{\text{npw,v}} = \{12, 24, 48\}$. Within each plot the number of waves in the pml $n_{\text{wpml,v}} \in [0, 3]$ and $\beta \in [0, 3]$ is varied.

3. Figure 2.20 left: Volumetric waves. Cubic elements. Absorbing function profile fixed $\gamma(s) = s$. The number of nodes per wave is different for each plot $n_{\text{npw,v}} = \{12, 24, 48\}$. Within each plot the number of waves in the pml $n_{\text{wpml,v}} \in [0, 3]$ and $\beta \in [0, 3]$ is varied.

4. Figure 2.20 right: Volumetric waves. Cubic elements. Absorbing function profile fixed $\gamma(s) = 1$. The number of nodes per wave is different for each plot $n_{\text{npw,v}} = \{12, 24, 48\}$. Within each plot the number of waves in the pml $n_{\text{wpml,v}} \in [0, 3]$ and $\beta \in [0, 3]$ is varied.

5. Figure 2.21 left: Shear waves. Linear elements. Absorbing function profile fixed $\gamma(s) = s$. The number of nodes per wave is different for each plot $n_{\text{npw,s}} = \{12, 24, 48\}$. Within each plot the number of waves in the pml $n_{\text{wpml,s}} \in [0, 3]$ and $\beta \in [0, 3]$ is varied.

6. Figure 2.21 right: Shear waves. Linear elements. Absorbing function profile fixed $\gamma(s) = 1$. The number of nodes per wave is different for each plot $n_{\text{npw,s}} = \{12, 24, 48\}$. Within each plot the number of waves in the pml $n_{\text{wpml,s}} \in [0, 3]$ and $\beta \in [0, 3]$ is varied.

7. Figure 2.22 left: Shear waves. Cubic elements. Absorbing function profile fixed $\gamma(s) = s$. The number of nodes per wave is different for each plot $n_{\text{npw.s}} = \{12, 24, 48\}$. Within each plot the number of waves in the pml $n_{\text{wpml,s}} \in [0, 3]$ and $\beta \in [0, 3]$ is varied.

8. Figure 2.22 right: Shear waves. Cubic elements. Absorbing function profile fixed $\gamma(s) = 1$. The number of nodes per wave is different for each plot $n_{\text{npw,s}} = \{12, 24, 48\}$. Within each plot the number of waves in the pml $n_{\text{wpml,s}} \in [0, 3]$ and $\beta \in [0, 3]$ is varied.

The following observations can be made.

- The contours of constant energy dissipation error for the volumetric wave and shear wave have almost identical contours. They also resemble the contours obtained from the 2D scalar wave case.

- The constant PML with absorbing function profile $\gamma(s) = 1$ works as well as the linear profile, with smaller energy dissipation error for all combinations of (n_{wpml}, β).

- Given the same discretization, cubic elements are capable of obtaining energy dissipation error two orders of magnitude smaller than linear elements.

Given these results, one can have some confidence in applying the heuristics developed for the 1D scalar wave to the 2D vector-valued case.

Remark: For the elastodynamic semi-infinite half space problem, or wave propagation in layered media, one must take into account the surface Rayleigh waves which have wave speed approximately equal to the shear wave, and interface Love waves, in the selection of appropriate discretization of the problem and PML parameter selection.

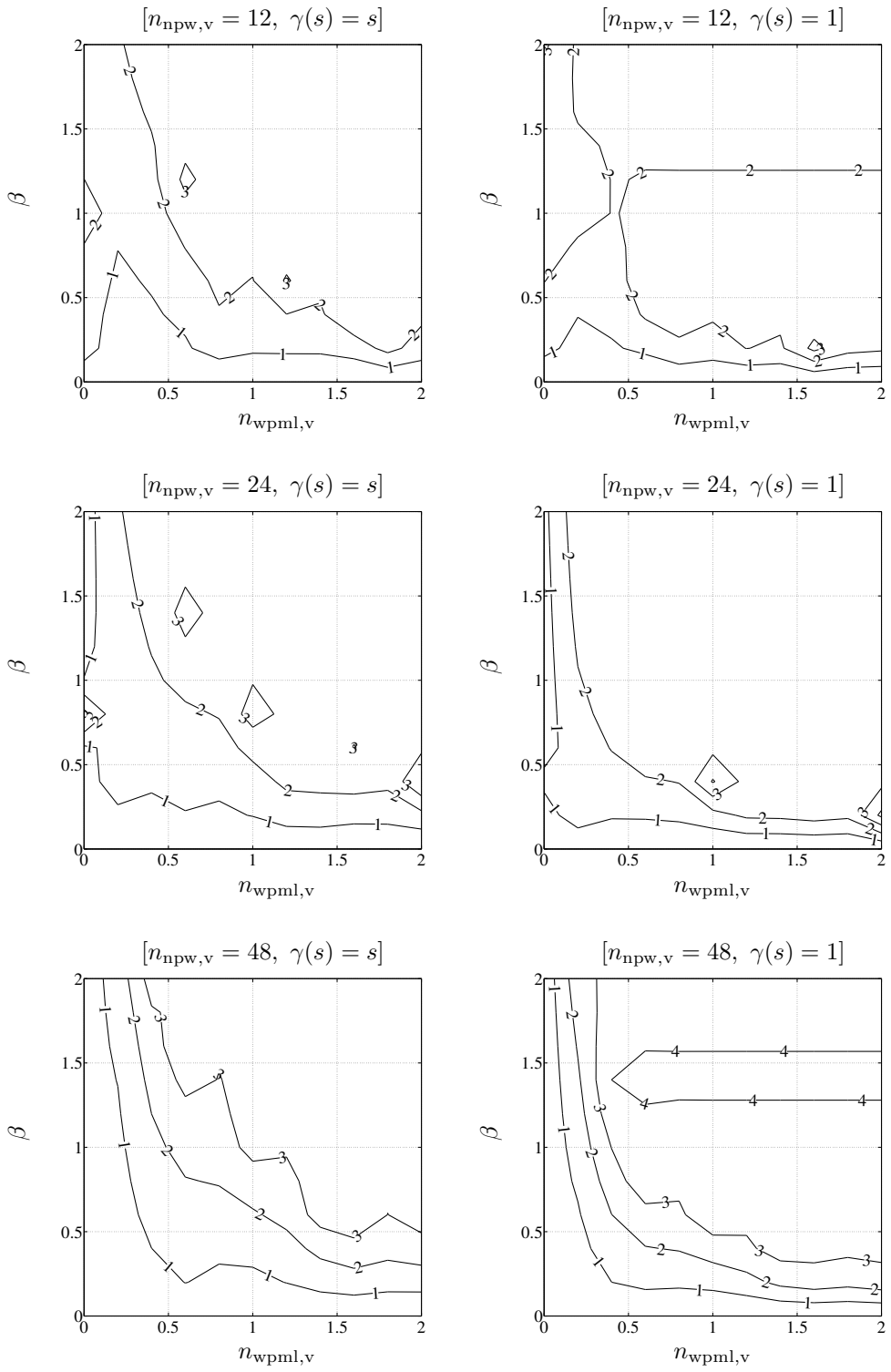


Figure 2.19: Energy dissipation error $-\log_{10}(E_{\mathrm{relerr}})$ for the 2D elastodynamic **volumetric** wave propationg with varying discretizations $n_{\mathrm{npw,v}} = \{12, 24, 48\}$, varying absorbing function profiles $\gamma(s) = \{1, s\}$, fixing the parameters element=$\{$**linear**$\}$.

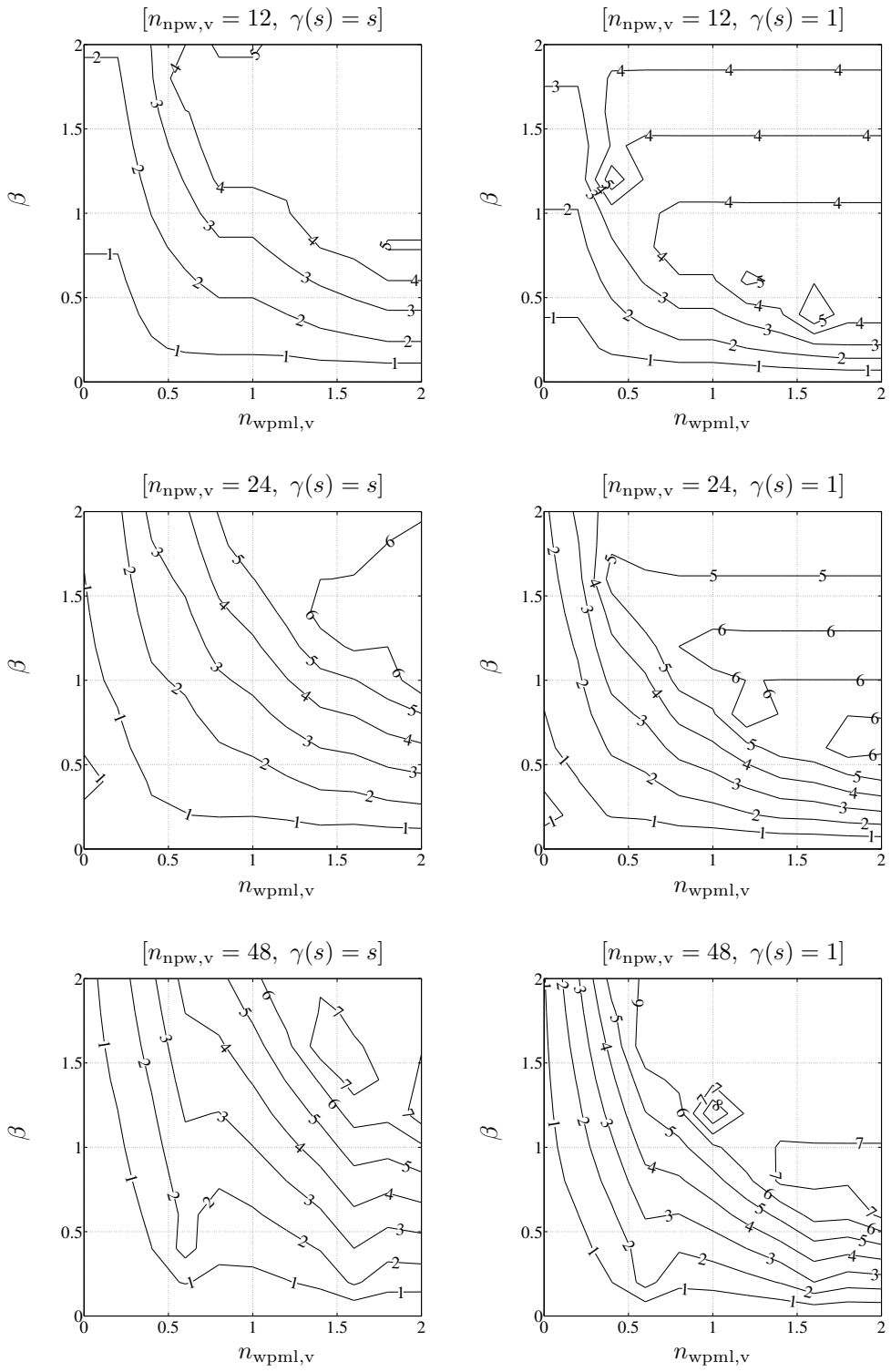


Figure 2.20: Energy dissipation error $-\log_{10}(E_{\mathrm{relerr}})$ for the 2D elastodynamic **volumetric** wave propationg with varying discretizations $n_{\mathrm{npw,v}} = \{12, 24, 48\}$, varying absorbing function profiles $\gamma(s) = \{1, s\}$, fixing the parameters element=$\{$**cubic**$\}$.

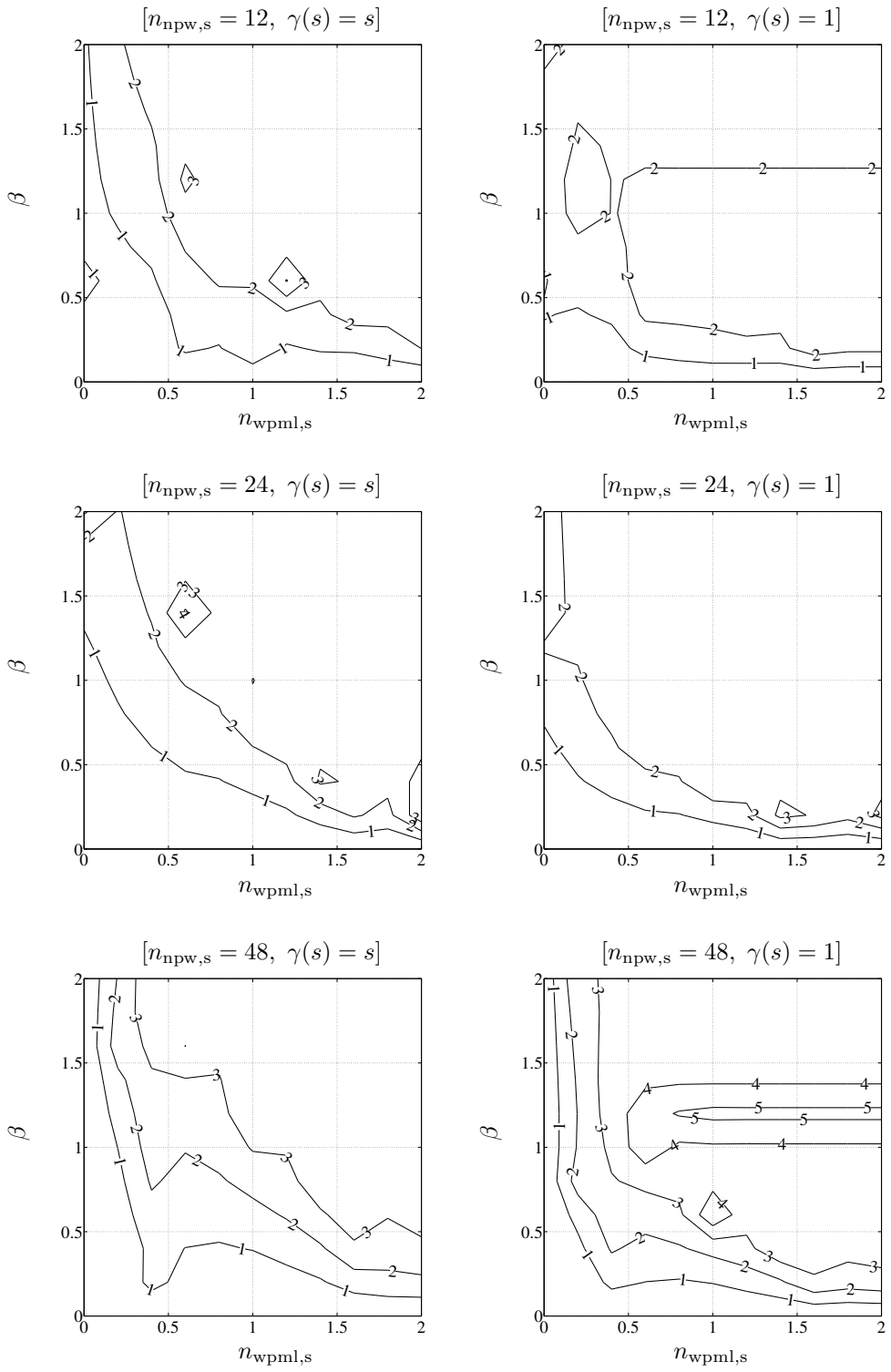


Figure 2.21: Energy dissipation error $-\log_{10}(E_{\mathrm{relerr}})$ for the 2D elastodynamic **shear** wave propationg with varying discretizations $n_{\mathrm{npw,s}} = \{12, 24, 48\}$, varying absorbing function profiles $\gamma(s) = \{1, s\}$, fixing the parameters element=$\{$**linear**$\}$.

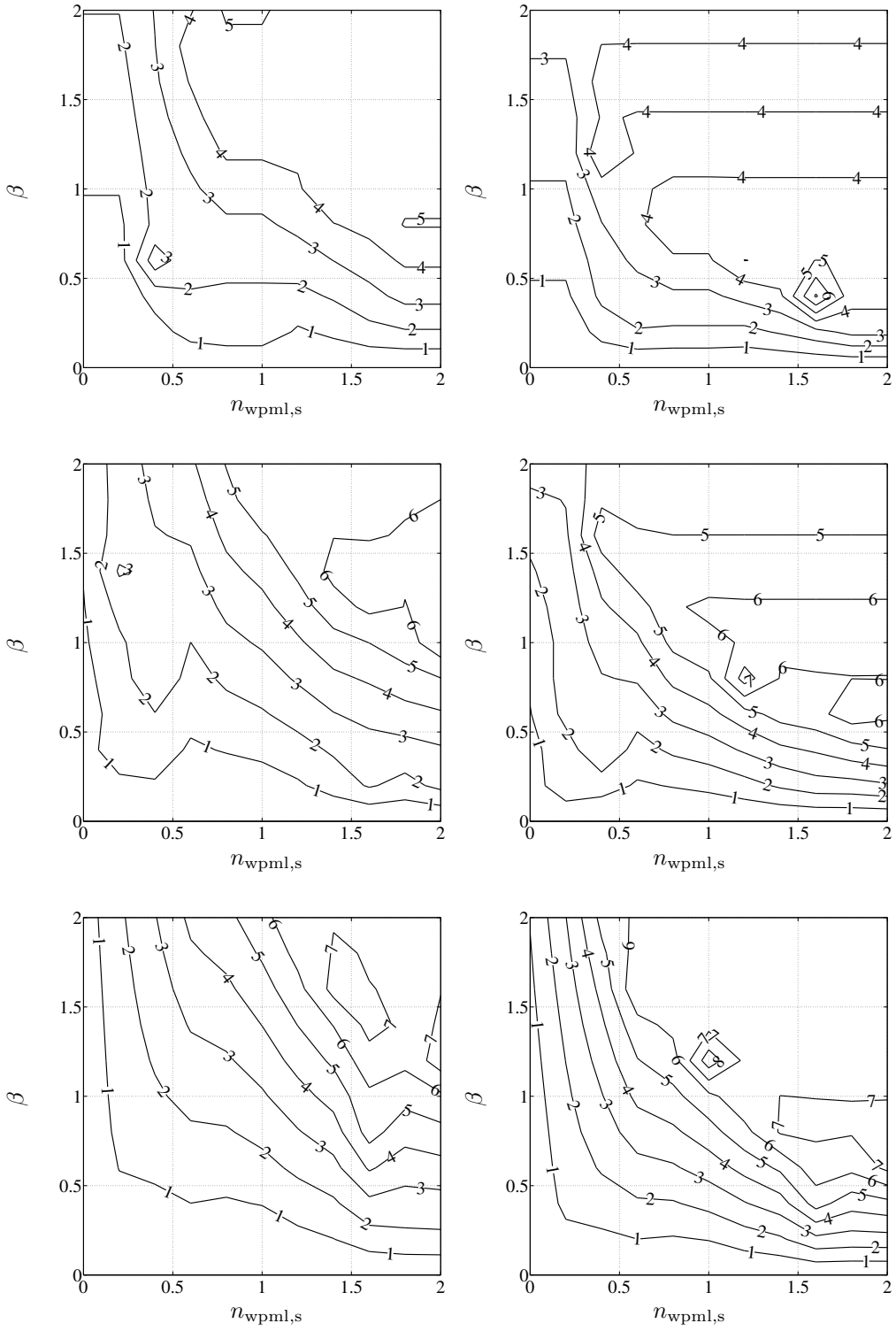


Figure 2.22: Energy dissipation error $-\log_{10}(E_{\text{relerr}})$ for the 2D elastodynamic **shear** wave propationg with varying discretizations $n_{\text{npw,s}} = \{12, 24, 48\}$, varying absorbing function profiles $\gamma(s) = \{1, s\}$, fixing the parameters element=$\{$**cubic**$\}$.

2.3 Geometric multigrid for forced motion computation

The quality factor Q, which one would like to obtain from the system, can be computed by evaluating the transfer function of the sytem or computing the complex-valued frequencies of the system. For transfer function evaluation, one must solve a series of linear systems of the form,

$$\left(\mathbf{K} - \omega^2 \mathbf{M}\right) \mathbf{u} \quad = \quad \mathbf{F}, \tag{2.101}$$

with varying forcing frequency ω, where $\mathbf{K}, \mathbf{M}$ are the stiffness and mass matrices, $\mathbf{F}$ is the forcing vector, and $\mathbf{u}$ is the displacement vector. For the evaluation of the complex-valued eigenfrequency, one must also solve linear systems of this form, as will be explained in Section 2.4. Thus it is crucial that one be able to solve linear systems of the form given in Equation (2.101) efficiently. For linear systems of size up to the order of hundreds of thousands, directive methods are employed due to their robustness [57]. For systems exceeding this size, such as those obtained from 3D discretizations of physical systems, direct methods are not optimal in the time required for the solution and memory requirements. This forces one to turn to iterative methods. Though the applicability of iterative methods are not as general as direct methods, they have less memory requirements, and if they are combined with a good preconditioner to accelerate the speed of convergence to the solution, i.e., decrease the number of iterations required, they can be very powerful [160, 24].

The systems which must be solved with the application of PML are complex-symmetric. Iterative Krylov subspace based methods such as GMRES and BICGSTAB for general linear systems, methods implemented in QMRPACK [75], and COCG [184] tailored for complex-symmetric linear systems do exist, but without the help of an efficient preconditioner, they have poor convergence rates. A preconditioners that is applicable for general complex-symmetric systems is incomplete LU factorizations [160], but for robustness for various problem types, one must tolerate large fill-ins which again result in large memory requirements.

The selection of an efficient preconditioner must also take into account the application. For the system of Equation (2.101), in the case of no PML with zero shift $\omega = 0$, one has a symmetric

positive definite linear system obtained from a discretization of an elliptic system of equations. For this class of problems, multigrid methods have been shown to perform extremely well [182]. A multigrid method to solve,

$$\mathbf{A}\mathbf{x} \quad = \quad \mathbf{b} \tag{2.102}$$

constructs multilevels of coarse discretizations of the original matrix $\mathbf{A}(= \mathbf{A}_0)$, which are defined as $\mathbf{A}_k(k = 1, \ldots, n)$, each matrix having size n_k where $n_k > n_{k+1}$. Approximate solutions and residuals are mapped between these grids (discretizations) by mapping operators. The idea behind the method is error smoothing. The solution $\mathbf{x}$ to the problem above can be decomposed into the eigenmodes of the operator $\mathbf{A}$. An operation called smoothing is applied to the approximate solution at each grid level, which eliminates (smooths) the error components of high frequency, i.e., eigenvalues with large magnitude, expressed on that grid. As smoothing is applied to the approximate solution on each grid, the components corresponding to the high frequency on the corresponding grid is eliminated. This sequence of operations is terminated at the coarsest grid with a direct solve eliminating all components representable on that grid, which includes the lowest frequency components of the total problem.

The key components in the multigrid method are the selection of the smoother $\mathbf{S}$, coarse grids (coarse grid operators) $\mathbf{A}_k$, and the mapping operators between the grids. The mapping operator used to map the solution from the fine grid to coarse grid, i.e., map degrees of freedom on the kth level to the $k+1$th level, is called the restriction operator $\mathbf{R}_k^{k+1}$, and the operator used to map the solution from the coarse grid to the fine grid, i.e., map degrees of freedom on the $k+1$th level to the kth level, is called the prolongation operator $\mathbf{P}_{k+1}^{k}$. Using these components, the multgrid method can be applied with various cycles [182]. The most common cycle is the multigrid V-cycle, which is presented in Figure 2.23. One cycle initiates on the finest grid and restricts down through the multilevels with smoothing steps interleaved. Once the coarsest level is reached, a direct solve is conducted, and the solution is interpolated up through the multilevels with smoothing steps

interleaved. The cycle ends with a smoothing step on the finest grid.

For smoothing, typically, point smoothers such as damped Jacobi and Gauss-Seidel [182] or polynomial smoothers such as Chebyshev polynomials [160, 6] are employed. Krylov iterative methods can also be used as smoothers. In the development of multigrid, much emphasis has been placed on the construction of efficient coarse grids and corresponding prolongation and restriction operators, due to the non-triviality in their construction and their importance in the overall convergence of the method. Depending on the method of coarse grid construction, multigrid methods are divided into two groups, geometric multigrid (GMG) and algebraic multigrid (AMG).

GMG [182] constructs the coarse grid and corresponding prolongation and restriction operators using the geometric information of the underlying physical problem On the other hand AMG constructs these purely algebraically from the system matrix $\mathbf{A}$ [69], or with very little information [4, 185, 135]. When one has direct access to the geometry of the underlying physical problem, GMG is the method of choice, since one can take full advantage of the problem. One still must face the effort of constructing the coarser grids, prolongation, and restriction operators, which may become difficult for complex geometry and special algorithms must be developed for unstructured mesh geometry [3]. When one only has access to the system matrix $\mathbf{A}$, AMG is the method of choice, since the method is able to construct the coarse grids purely algebraically from the matrix. Information such as the connectivity of the system is extracted from the weighted graph interpretation of the matrix [182]. Smooth aggregation AMG methods incorporate vectors that belong to the near null space of the system matrix in the coarse grids [185, 135]. When the system matrix is derived from Poisson's equation, these are the constant vectors, and in the case of elasticity, these are the rigid body modes represented by the translations and rotations [4]. The drawback of AMG methods is the ability of the method to produce adequate coarse grids for non-symmetric indefinite complex-valued linear systems, due to the assumptions made in developing the theory, such as symmetry, real-valuedness, and positive definiteness. The requirement of positive definiteness is also required in the convergence proof of GMG. The coarse grid operators $\mathbf{A}_k$ can be obtained

independently from the prolongation and restriction operators or from a Petrov-Galerkin projection using the prolongation and restriction operators as,

$$\mathbf{A}_k = \mathbf{R}_{k-1}^k \mathbf{A}_{k-1} \mathbf{P}_k^{k-1} (k = 1, \ldots, n), \tag{2.103}$$

where to simplify notation,

$$\mathbf{P}_k = \mathbf{P}_k^{k-1}, \tag{2.104}$$

$$\mathbf{R}_k = \mathbf{R}_{k-1}^k \ . \tag{2.105}$$

One can also select $\mathbf{R}_{k-1}^k = \left(\mathbf{P}_k^{k-1}\right)^T$ to obtain the standard Galerkin projection.

The application of PML and the non zero shift $\omega = 0$, does make the matrix $\mathbf{A}$ complex-valued and indefinite, but if the complex-valued nature and indefiniteness is small, one can hope that the multigrid method designed for symmetric positive definite systems is still applicable. Based on such an idea, multigrid for the real-symmetric indefinite case has been investigated, and convergence has been proven under the assumption that the coarsest grid in the multilevel be sufficiently fine [22, 41]. This restriction decreases the optimality of the multigrid method, but produces a solution method. Incorporation of Krylov iterations as smoothers in the multigrid method has also been proposed to treat the indefiniteness of the operator [67, 5], but this method can be costly when the number of smoothing iterations required for convergence increases. Several attempts have been made in the application of Ruge-Stüben classical AMG to complex symmetric matrices arising from the discretization of the time-harmonic Maxwell equations in 2D scalar-valued potential form, which take the form of Equation (2.101) with real-valued $\mathbf{K}, \mathbf{M}$ and a complex-valued ω [158, 116]. It must be pointed out that the complex-valued symmetric structure obtained from the application of PML equations is fundamentally different from those arising from real-valued matrices with complex-valued shifts ω. As is explained later in the section, a complex-valued shift ω can improve the convergence of multigrid compared to a real-valued shift.

In our approach, GMG is adapted to solve the complex-symmetric indefinite linear system of Equation (2.101) arising from the discretization of the time-harmonic elastodynamic problem

with PML applied. This choice has been made due to the lack of existence of an AMG approach for constructing the coarse grids, the prolongation operators, and the restriction operators, for this type of problem. A GMG approach is possible in our case, since full access to the geometric information of the physical problem is available, including the mesh generation process. The systems of interest have regular geometry, allowing block generation of the finite element mesh. This allows a fully geometric approach in construction of the coarse grids, prolongation operators, and restriction operators. In this section the components of GMG that are used in our method are introduced. First, the smoothers that we consider for our method are introduced. The method of Local Fourier Analysis (LFA), which can be used to analyze the the two stationary smoothers that we employ, the Gauss-Seidel smoother and Kaczmarz smoother, is presented. This is followed by the presentation of a Lemma for complex-symmetric matrices that is employed in combination with the Chebyshev smoother that we also use. Next, the method by which the coarse grids, prolongation operators, and restriction operators are constructed is explained. The coarse grids are constructed by a block generation, and the prolongation operators are constructed based on evaluating coarse grid shape functions at the fine grid points [3, 72]. The section is closed by presenting the 2-grid convergence factors of the proposed GMG method on 2D scalar wave and 2D elastodynamic problems. For the 3D scalar wave and 3D elastodynamic problem, only the results obtained from the LFA is presented.

2.3.1 Smoothers

The types of smoothers that are available in multigrid can be grouped into the following two categories [182],

1. Stationary methods: Jacobi, Weighted Jacobi, Gauss-Seidel, Symmetric Gauss-Seidel, SOR, Polynomial smoothers,

2. Nonstationary methods: Krylov methods (CG, GMRES, BICGSTAB,.etc.).

function MGV($\mathbf{A}_i$,$\mathbf{b}_i$)

if there is a coarser grid $i+1$

$\mathbf{x}_i \leftarrow \mathbf{S}^{\nu 1}(\mathbf{A}_i, \mathbf{b}_i)$ -$\nu 1$ iterations of the pre-smoother
$\mathbf{r}_i \leftarrow \mathbf{b}_i - \mathbf{A}_i\mathbf{x}_i$ -residual calculation
$\mathbf{r}_{i+1} \leftarrow \mathbf{R}_{i+1}\mathbf{r}_i$ -restriction of residual to coarse grid
$\mathbf{e}_{i+1} \leftarrow \text{MGV}(\mathbf{R}_{i+1}\mathbf{A}_i\mathbf{P}_{i+1})$ -the recursive application of MG
$\mathbf{x}_i \leftarrow \mathbf{x}_i + \mathbf{P}_{i+1}\mathbf{e}_{i+1}$ -prolongation of coarse grid correction
$\mathbf{r}_i \leftarrow \mathbf{b}_i - \mathbf{A}_i\mathbf{x}_i$
$\mathbf{x}_i \leftarrow \mathbf{x}_i + \mathbf{S}^{\nu 2}(\mathbf{A}_i, \mathbf{r}_i)$ -$\nu 2$ iterations of the post-smoother

else

$\mathbf{x}_i \leftarrow \mathbf{A}_i^{-1}\mathbf{r}_i$ -direct solve on coarsest grid

return $\mathbf{x}_i$

Figure 2.23: Multigrid V-cycle algorithm

A smoother is called stationary when it is representable by a linear operator, i.e., a matrix. The behavior of nonstationary smoothers depend on the vector it is applied to. For the symmetric positive definite case where all eigenvalues are real and positive, the ideal smoother eliminates error modes corresponding to large eigenvalues, unaffecting modes corresponding to small eigenvalues which are to be eliminated by the direct solve on the coarsest grid. When the matrix is non-symmetric and indefinite, the action of the smoother is not as clearly defined, but still one desires the similar behavior, such that error modes corresponding to eigenvalues with large magnitude are eliminated.

For application of multigrid to the linear system with PML, the performance of three types of stationary smoothers are investigated: Gauss-Seidel, Kaczmarz, and Chebyshev. The investigation of nonstationary smoothers are excluded here, since they have difficulty as components in multigrid preconditioners in combination with the Jacobi-Davidson method used for obtaining eigenvalues in Section 2.4. First the method of Local Fourier Analysis (LFA) is introduced to analyze the behavior of the two stationary smoothers, Gauss-Seidel and Kaczmarz. This is followed by a brief explanation of the Gauss-Seidel and Kaczmarz smoother. Next the details of the Chebyshev polynomial

smoothers are presented along with a Lemma that enables one to obtain a bound on the real part and imaginary part of the eigenvalues of a complex-symmetric matrix, by computing the eigenvalues of the real and imaginary part of the complex-symmetric matrix.

Remark: Though the behavior of nonstationary methods depend on the acting vector, one can quantitatively argue their behavior. In the case of CG applied as a smoother to real symmetric positive definite systems, it will try to minimize the error in the operator norm $||\mathbf{e}||_{\mathbf{A}} := ||\mathbf{e}^*\mathbf{A}\mathbf{e}||$. This implies that error modes corresponding to large eigenvalues are weighted more, and tend to be eliminated more than those corresponding to small eigenvalues. In the case of GMRES applied as a smoother to general systems, it will try to minimize the residual $||\mathbf{r}|| := ||\mathbf{e}^*\mathbf{A}^*\mathbf{A}\mathbf{e}||$, i.e., the error in the $\mathbf{A}^*\mathbf{A}$ operator norm. This implies that error modes corresponding to large eigenvalues of the operator $\mathbf{A}^*\mathbf{A}$ are weighted more, and tend to be eliminated more than those corresponding to small eigenvalues.

Local Fourier analysis

The linear system that one desires to apply the smoother on,

$$\mathbf{A}\mathbf{u} = \mathbf{b}, \tag{2.106}$$

in most cases arises from the discretization of some linear partial differential equation,

$$\mathcal{L}u = b, \tag{2.107}$$

where $\mathcal{L}$ is the linear operator to which $\mathbf{A}$ is a discretization of, and u, b are the continuous versions of $\mathbf{u}$ and $\mathbf{b}$. When the smoothing step for Equation (2.106) can be written in the form,

$$\mathbf{E}\mathbf{u}_{\text{new}} = \mathbf{F}\mathbf{u}_{\text{old}} + \mathbf{b}, \tag{2.108}$$

$$\mathbf{A} = \mathbf{E} - \mathbf{F}, \tag{2.109}$$

where $\mathbf{u}_{\text{old}}$ is the solution before the smoothing step and $\mathbf{u}_{\text{new}}$ is the solution after the smoothing step, one can employ the method of local Fourier analysis (LFA) [182]. LFA measures the effectiveness of

the smoother on Equation (2.107) on an infinite grid, neglecting the influence and effect of boundary conditions. For the 2D infinite grid,

$$\mathcal{G}_h \quad := \quad \left\{ \mathbf{x} = h\mathbf{k} | \mathbf{k} \in \mathbb{Z}^2 \right\}, \tag{2.110}$$

with uniform grid size of h in both x, y directions, the eigenfunctions are the Fourier modes, expressed as,

$$\begin{aligned} \varphi(\boldsymbol{\theta}, \mathbf{x}) \quad &= \quad \exp(i\boldsymbol{\theta} \cdot \mathbf{x}/h) \\ &= \quad \exp(i\theta_x x/h) \exp(i\theta_y y/h). \end{aligned} \tag{2.111}$$

Without loss of generality, it is assumed that,

$$(\theta_x, \theta_y) \in [-\pi, \pi)^2 \ . \tag{2.112}$$

The linear operator $\mathcal{L}$ discretized on the grid $\mathcal{G}_h$ is denoted as $\mathbf{L}_h$, and is considered to have a difference stencil representation,

$$\mathbf{L}_h w(\mathbf{x}) \quad = \quad \sum_{\mathbf{k} \in \text{Finite Set}} l_{\mathbf{k}} w(\mathbf{x} + h\mathbf{k}) \ , \tag{2.113}$$

with constant coefficients $l_{\mathbf{k}}$. $\mathbf{A}$ is essentially $\mathbf{L}_h$ for some h with appropriate boundary conditions. For nodes sufficiently away from the boundary, $\mathbf{A}$ is assumed to have this structure.

$$(\mathbf{A}\mathbf{x})_i \quad = \quad \sum_{j \in \text{Finite Set}} A_{ij} x_j \ . \tag{2.114}$$

The symbol or eigenvalue λ_θ corresponding to the Fourier mode φ is defined as,

$$\mathbf{L}_h \varphi(\boldsymbol{\theta}, \mathbf{x}) \quad = \quad \lambda_\theta \varphi(\boldsymbol{\theta}, \mathbf{x}), \tag{2.115}$$

with,

$$\lambda_\theta \quad := \quad \sum_{\mathbf{k} \in \text{Finite Set}} l_{\mathbf{k}} \exp(i\boldsymbol{\theta} \cdot \mathbf{k}) \ . \tag{2.116}$$

For smoothers applied to the problem,

$$\mathbf{L}_h \mathbf{u} \quad = \quad \mathbf{b}, \tag{2.117}$$

which can be written locally as,

$$\mathbf{L}_h^+ \mathbf{u}_{\text{new}} + \mathbf{L}_h^- \mathbf{u}_{\text{old}} \quad = \quad \mathbf{b}, \tag{2.118}$$

subtraction of the exact solution $\mathbf{u}$ results in the expression,

$$\mathbf{L}_h^+ \mathbf{e}_{\text{new}} \quad = \quad -\mathbf{L}_h^- \mathbf{e}_{\text{old}}, \tag{2.119}$$

$$\mathbf{e}_{\text{new}} \quad = \quad \mathbf{u}_{\text{new}} - \mathbf{u}, \tag{2.120}$$

$$\mathbf{e}_{\text{old}} \quad = \quad \mathbf{u}_{\text{old}} - \mathbf{u} \ . \tag{2.121}$$

Since φ are also eigenfunctions of both $\mathbf{L}_h^+$ and $\mathbf{L}_h^-$, by denoting their eigenvalues as λ_θ^+ and λ_θ^-, the action of the smoother on an error component corresponding to an eigenfunction,

$$\mathbf{e}_{\text{old}} \quad = \quad \varphi(\boldsymbol{\theta}, \mathbf{x}), \tag{2.122}$$

is,

$$\mathbf{e}_{\text{new}} \quad = \quad -\frac{\lambda_\theta^-}{\lambda_\theta^+} \varphi(\boldsymbol{\theta}, \mathbf{x}). \tag{2.123}$$

If one considers an infinite grid coarser than $\mathcal{G}_h$ with grid size $2h$,

$$\mathcal{G}_{2h} \quad := \quad \left\{ \mathbf{x} = 2h\mathbf{k} | \mathbf{k} \in \mathbb{Z}^2 \right\}, \tag{2.124}$$

one observes that eigenfunctions $\varphi(\boldsymbol{\theta}, \mathbf{x})$ on $\mathcal{G}_h$ with,

$$\boldsymbol{\theta} \in [-\pi/2, \pi/2)^2, \tag{2.125}$$

are representable on $\mathcal{G}_{2h}$. From this property, one can divide the eigenfunctions on $\mathcal{G}_h$ into two groups,

$$T_{\text{low}} \quad := \quad \left\{ \boldsymbol{\theta} | \boldsymbol{\theta} \in [-\pi/2, \pi/2)^2 \right\}, \tag{2.126}$$

$$T_{\text{high}} \quad := \quad \left\{ \boldsymbol{\theta} | \boldsymbol{\theta} \in [-\pi, \pi)^2 \setminus T_{\text{low}} \right\}, \tag{2.127}$$

those modes that are representable on coarse grids T_{low} and those modes that are highly oscillating and not representable on coarse grids T_{high}. In multigrid, the direct solve on the coarse grid is

designed to eliminate the low frequency errors, and the smoother is designed to remove the high frequency errors. Thus one can define the reduction factor of a smoother $\mathbf{S}$ whose action is representable as Equation (2.118) as,

$$\mu_{\text{loc}}(\mathbf{S}) \quad := \quad \sup\left\{\left|\frac{\lambda^-(\boldsymbol{\theta})}{\lambda^+(\boldsymbol{\theta})}\right| : \boldsymbol{\theta} \in T_{\text{high}}\right\} . \tag{2.128}$$

One can also define the smoothing factor in each direction as,

$$\mu_{\text{x,loc}}(\mathbf{S}) \quad := \quad \sup\left\{\left|\frac{\lambda^-(\boldsymbol{\theta})}{\lambda^+(\boldsymbol{\theta})}\right| : \boldsymbol{\theta} \in T_{\text{x,high}}\right\}, \tag{2.129}$$

$$\mu_{\text{y,loc}}(\mathbf{S}) \quad := \quad \sup\left\{\left|\frac{\lambda^-(\boldsymbol{\theta})}{\lambda^+(\boldsymbol{\theta})}\right| : \boldsymbol{\theta} \in T_{\text{y,high}}\right\} , \tag{2.130}$$

where,

$$T_{\text{x,high}} \quad := \quad \{\boldsymbol{\theta} | \theta_x \in [-\pi, -\pi/2] \cup [\pi/2, \pi]\}, \tag{2.131}$$

$$T_{\text{y,high}} \quad := \quad \{\boldsymbol{\theta} | \theta_y \in [-\pi, -\pi/2] \cup [\pi/2, \pi]\}. \tag{2.132}$$

It is clear that $\mu_{\text{loc}} = \max(\mu_{\text{x,loc}}, \mu_{\text{y,loc}})$. In the case of vector valued functions, this definition can be generalized to,

$$\mu_{\text{loc}}(\mathbf{S}) \quad := \quad \sup\left\{\left|\rho\left(\mathbf{L}^+(\boldsymbol{\theta})^{-1}\mathbf{L}^-(\boldsymbol{\theta})\right)\right| : \boldsymbol{\theta} \in T_{\text{high}}\right\} , \tag{2.133}$$

where ρ denotes the spectral radius of the operator. Smoothing factors in each direction can be defined analogously to the scalar case.

$$\mu_{\text{x,loc}}(\mathbf{S}) \quad := \quad \sup\left\{\left|\rho\left(\mathbf{L}^+(\boldsymbol{\theta})^{-1}\mathbf{L}^-(\boldsymbol{\theta})\right)\right| : \boldsymbol{\theta} \in T_{\text{x,high}}\right\} , \tag{2.134}$$

$$\mu_{\text{y,loc}}(\mathbf{S}) \quad := \quad \sup\left\{\left|\rho\left(\mathbf{L}^+(\boldsymbol{\theta})^{-1}\mathbf{L}^-(\boldsymbol{\theta})\right)\right| : \boldsymbol{\theta} \in T_{\text{y,high}}\right\} . \tag{2.135}$$

This tool is used to analyze the smoothing properties of Gauss-Seidel and Kaczmarz for discretizations of the 2D scalar, 2D elastodynamic, 3D scalar, and 3D elastodynamic problem in Section 2.3.3.

Gauss-Seidel

The Gauss-Seidel smoother is defined as the following iteration,

$$(\mathbf{L}+\mathbf{D})\,\mathbf{x}_{\text{new}} \quad = \quad -\mathbf{U}\mathbf{x}_{\text{old}}+\mathbf{b}, \tag{2.136}$$

where $\mathbf{L},\mathbf{U},\mathbf{D}$ are the strictly lower triangular, strictly upper triangular, and diagonal parts of the matrix $\mathbf{A}$. One knows that for Hermitian positive definite or diagonally dominant $\mathbf{A}$, the method is convergent [80]. A criterion for complex-valued symmetric matrix does not exist. To check the convergence of the iteration for general matrices, one must look at the spectral radius of the iteration matrix $-(\mathbf{L}+\mathbf{D})^{-1}\mathbf{U}$. For structured meshes, one can also employ LFA, as is done in Section 2.3.3 to investigate the convergence of this method for problems employing PML.

Kaczmarz

The Kaczmarz smoother is defined as the following iteration,

$$\mathbf{x}_{\text{new}} \quad = \quad \left[\mathbf{I}-\mathbf{A}^{*}\left(\mathbf{L}_{\mathbf{AA}^{*}}+\mathbf{D}_{\mathbf{AA}^{*}}\right)\mathbf{A}\right]\mathbf{x}_{\text{old}}+\mathbf{A}^{*}\left(\mathbf{L}_{\mathbf{AA}^{*}}+\mathbf{D}_{\mathbf{AA}^{*}}\right)\mathbf{b}, \tag{2.137}$$

where $\mathbf{L}_{\mathbf{AA}^{*}},\mathbf{D}_{\mathbf{AA}^{*}}$ are the strictly lower triangular, and diagonal parts of the matrix $\mathbf{AA}^{*}$. These equations can be obtained from application of Gauss-Seidel to,

$$\mathbf{AA}^{*}\mathbf{y} \quad = \quad \mathbf{b}, \tag{2.138}$$

$$\mathbf{y} \quad := \quad \mathbf{A}^{-*}\mathbf{x}. \tag{2.139}$$

Here * denotes conjugate transposition. This method is also referred to as a row-projection method [160], and the method has been proven to converge for non-singular $\mathbf{A}$ [174]. The convergence can also be seen from the Hermitian positive definite property of $\mathbf{AA}^{*}$ and Gauss-Seidel for non-singular $\mathbf{A}$.

The advantage of this smoothing method, is its applicability as a smoother to any non-singular system. The disadvantage is its weak smoothing properties [182]. In Section 2.3.3, it is observed that this method is still effective as a smoother for the scalar-valued problem but not for elasticity.

Chebyshev polynomials

Polynomial smoothers [160] are a class of smoothers which have error smoothing operators of the form,

$$\mathbf{e}_{\text{new}} = Q_n(\mathbf{A})\mathbf{e}_{\text{old}}, \tag{2.140}$$

$$\mathbf{e}_{\text{new}} := \mathbf{x}_{\text{new}} - \mathbf{x}, \tag{2.141}$$

$$\mathbf{e}_{\text{old}} := \mathbf{x}_{\text{old}} - \mathbf{x}, \tag{2.142}$$

where $\mathbf{x}_{\text{old}}, \mathbf{x}_{\text{new}}, \mathbf{x}$ are the old, new, and exact solutions of the problem. $Q_n(x)$ is a n-th order polynomial with the restriction $Q_n(0) = 1$, since the error should not change if the operator is zero. As a smoother, one desires,

$$||\mathbf{e}_{\text{new}}||_2 < ||Q_n(\mathbf{A})\mathbf{e}_{\text{old}}||_2, \tag{2.143}$$

where the 2-norm has been selected without loss of generality. Assuming $\mathbf{A}$ is diagonalizable $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$, where $\mathbf{V}$ are the eignvectors and $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues, the error in the eigenvectors coordinates $\hat{\mathbf{e}} := \mathbf{V}^{-1}\mathbf{e}$ is,

$$||\hat{\mathbf{e}}_{\text{new}}||_2 < ||Q_n(\mathbf{\Lambda})\hat{\mathbf{e}}_{\text{old}}||_2. \tag{2.144}$$

When $\mathbf{A}$ is Hermitian, $||\hat{\mathbf{e}}||_2 = ||\mathbf{e}||_2$, and Equations (2.143) and (2.144) are equivalent. Further manipulation of Equation (2.144), yields,

$$||\hat{\mathbf{e}}_{\text{new}}||_2 < \max_{\lambda_i \in \mathbf{\Lambda}} |Q_n(\lambda_i)| \;\; ||\hat{\mathbf{e}}_{\text{old}}||_2. \tag{2.145}$$

Thus one would like to find an optimal polynomial $Q_n(x)$ such that,

- $Q_n(0) = 1$
- $\max_{x \in \Omega} |Q_n(x)|$ is small as possible for some domain $\Omega \supset \mathbf{\Lambda}$.

When $\mathbf{A}$ is Hermitian positive definite, $\mathbf{\Lambda}$ is a subset of the positive real line, and the optimal polynomials are the Chebyshev polynomials [160]. For the case of complex-symmetric $\mathbf{A}$, the eigenvalues

lie on the whole complex plane, in which case the Chebyshev polynomials are not optimal [73] but still have an effective smoothing property as long as the eigenvalues of $\mathbf{A}$ are not too close to zero, and lie on the positive real half of the complex plane. The Chebyshev polynomials are defined recursively as,

$$T_0(z) = 1, \tag{2.146}$$

$$T_1(z) = z, \tag{2.147}$$

$$T_{n+1}(z) = 2zT_n(z) - T_{n-1}(z), \tag{2.148}$$

or,

$$T_n(z) = \cosh(n \cosh^{-1}(z)) \ . \tag{2.149}$$

For the purposes of the polynomial preconditioner, one considers the scaled and translated version,

$$P_n(z) = \frac{T_n\left(\frac{d-z}{c}\right)}{T_n\left(\frac{d}{c}\right)}, \tag{2.150}$$

where $P_n(0) = 1$. In the complex plane, asymptotically, the contours of constant magnitude of the polynomial form ellipses centered at d with foci at $d + c$ and $d - c$. An example of the case $d = 2, c = \{1, 1i\}$ and $n = 3$ is shown in Figure 2.24. One sees that selecting a complex valued c can rotate the ellipse.

Thus even if the eigenvalues lie off the real axis, as long as they are sufficiently confined within these ellipses, one can obtain sufficient smoothing. As is observed in the two grid convergence factors of Section 2.3.3, for PML parameters with β not too large, the eigenvalues do not part too much from the real axis. For a complex-symmetric matrix, the magnitude of the imaginary part of the eigenvalue can be easily estimated by the imaginary part of the matrix as is shown in the following Lemma.

Lemma 2.3.1 *Given a complex symmetric matrix* $\mathbf{A}$*,*

$$\mathbf{A} := \mathbf{A}_r + i\mathbf{A}_i, \tag{2.151}$$

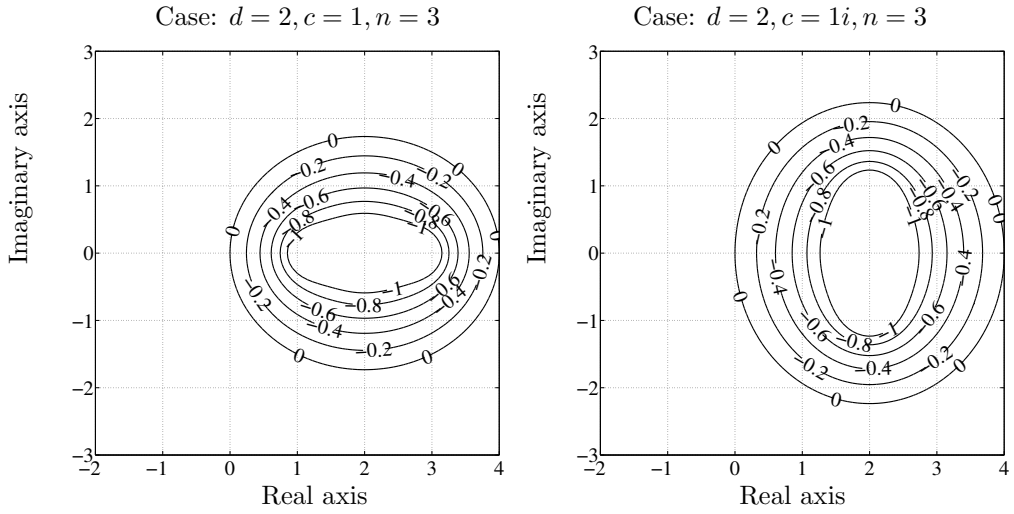


Figure 2.24: The convergence factor of Chebyshev smoothers

where $\mathbf{A}_r^T = \mathbf{A}_r$ *is the real part and* $\mathbf{A}_i^T = \mathbf{A}_i$ *is the imaginary part, the real and imaginary part of the eigenvalues are bounded by the eigenvalues of* $\mathbf{A}_r$ *and* $\mathbf{A}_i$ *respectively,*

$$\lambda_{r,min} \leq \mathrm{Re}(\lambda) \leq \lambda_{r,max}, \tag{2.152}$$

$$\lambda_{i,min} \leq \mathrm{Im}(\lambda) \leq \lambda_{i,max}. \tag{2.153}$$

Proof The eigenvalue λ and corresponding eigenvector $\mathbf{x}$ are related as:

$$\lambda = \frac{\mathbf{x}^*\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{x}} \tag{2.154}$$

$$= \frac{\mathbf{x}^*\mathbf{A}_r\mathbf{x}}{\mathbf{x}^*\mathbf{x}} + i\frac{\mathbf{x}^*\mathbf{A}_i\mathbf{x}}{\mathbf{x}^*\mathbf{x}} \ . \tag{2.155}$$

Since for a real symmetric matrix, the Rayleigh quotient is always real,

$$\mathrm{Re}(\lambda) = \frac{\mathbf{x}^*\mathbf{A}_r\mathbf{x}}{\mathbf{x}^*\mathbf{x}} \in \mathbb{R}, \tag{2.156}$$

$$\mathrm{Im}(\lambda) = \frac{\mathbf{x}^*\mathbf{A}_i\mathbf{x}}{\mathbf{x}^*\mathbf{x}} \in \mathbb{R} \ . \tag{2.157}$$

q.e.d.□.

With this Lemma, one can bound the eigenvalues within a bounding box. Thus if one has a situation such as is shown in Figure 2.25, the Chebyshev smoother can be applied.

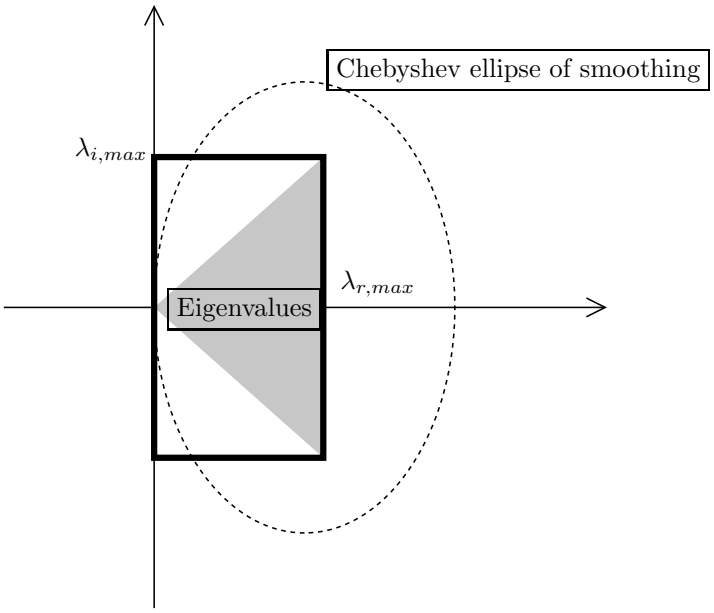


Figure 2.25: Chebyshev ellipse enclosing the region of eigenvalues(gray) using estimates obtained from the bounding box defined through the Lemma

Though adaptive methods exist to estimate the convex hull of the eigenvalues of the matrix [44, 137, 136], and to select optimal c and d, here the Lemma presented is used to estimate c and d. $\lambda_{r,max}$ and $\lambda_{i,max}$ are computed from the real and imaginary part of the matrix by a power method or Lanczos method. Then c and d are set as,

$$d = \lambda_{r,max}, \tag{2.158}$$

$$c = \lambda_{i,max} i. \tag{2.159}$$

This choice is able to enclose all eigenvalues for mild PML parameters. It is observed in Section 2.3.3 that this choice does lead to convergent results.

Though Cheybshev smoothers require evaluation of several parameters, they only involve matrix-vector operations which makes it easier to efficiently parallelize them compared to methods such as Gauss-Seidel and Kaczmarz.

2.3.2 Prolongation operators

The geometric multigrid approach is taken to construct the coarse grids and corresponding prolongation and restriction operators. The restriction operator is defined as the transpose of the prolongation operator $\mathbf{R} = \mathbf{P}^T$, and the coarse grid operators are defined by a Galerkin projection. The geometries of the resonators analyzed have fairly regular structure, which allows one to generate the mesh of the structure by mapped structured Cartesian block meshes.

Without loss of generality, the method of constructing the coarse grids and corresponding prolongation operators are presented for the two grid case. The computational domain is first divided into superblocks, which are each mapped squares for the 2D case, and mapped cubes for the 3D case. Both the fine grid $\mathcal{G}_{\text{fine}}$ and coarse grid $\mathcal{G}_{\text{fine}}$ share the same domain constructed from these superblocks. The difference between the two is the number of nodes placed along the edges and faces. It is assumed that for any super block, the number of nodes across an edge or face on the fine grid is larger than the coarse grid. To clarify this setup, the two grid case for the scalar 2D problem is employed to explain this method of grid construction. The problem is defined on the V-shaped domain depicted in Figure 2.26. The meshes are constructed from bilinear finite elements. The left mesh shows standard mesh construction, where the coarse and fine mesh defined by the coarse and fine grid nodes are not particularly related. The right mesh shows the block generated mesh construction, where two superblocks are defined, one in dark gray, and the other in light gray. For the coarse mesh, each region is meshed with zero nodes per side. The fine mesh is generated from the underlying super blocks, such that there are two nodes on each side of the super blocks.

The prolongation operator $\mathbf{P}$ from the coarse grid to fine grid is constructed by the evalution of fine grid nodes at the coarse grid shape functions. Let us denote the coarse grid nodes as $\mathbf{x}_i^c$ and the fine grid nodes as $\mathbf{x}_i^f$. Using the finite element shape functions on the coarse grid $N_j^c(\mathbf{x})$, the field on the coarse grid is represented as,

$$u(\mathbf{x}) \quad = \quad \sum_{j \in \mathcal{G}_{\text{coarse}}} N_j^c(\mathbf{x}) u_j^c \ . \tag{2.160}$$

Using the coarse grid values u_i^c, one can obtain an approximation of the nodal values on the fine grid u_i^f, by the interpolated values from the coarse grid shape functions [72],

$$u_i^f \quad := \quad u(\mathbf{x}_i^f) = \sum_{j \in \mathcal{G}_{\text{coarse}}} N_j^c(\mathbf{x}_i^f) u_j^c \ . \tag{2.161}$$

This is a linear mapping from the coarse grid values u_i^c to the fine grid values u_i^f, which is defined as the prolongation operator,

$$\mathbf{u}^f \quad = \quad \mathbf{P}\mathbf{u}^c \ , \tag{2.162}$$

whose components are,

$$P_{ij} \quad = \quad N_j^c(\mathbf{x}_i^f), \qquad i \in \mathcal{G}_{\text{fine}}, \quad j \in \mathcal{G}_{\text{coarse}}. \tag{2.163}$$

For the vector-valued case, each field is interpolated separately by the method introduced above. Since the finite element nodal shape functions have small support and little overlap, the prolongation operator is sparse.

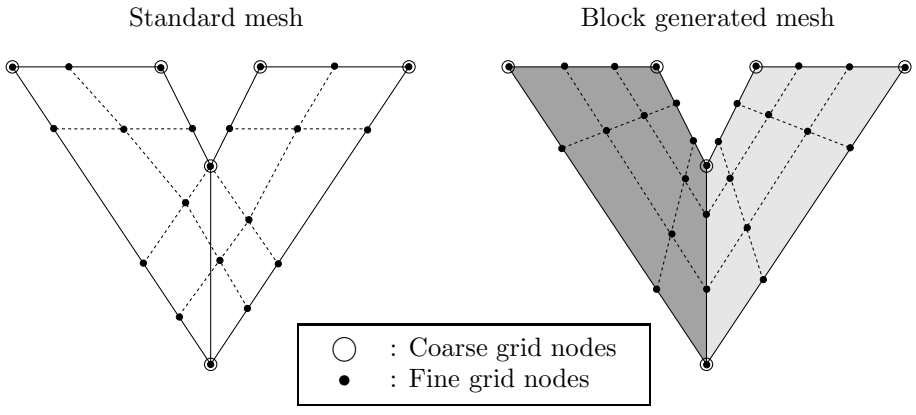


Figure 2.26: Standard mesh and block generated mesh

In this method of prolongation operator construction, one computational bottleneck can arise in the evaluation of the entries of $\mathbf{P}$. If one evaluates every entry, the complexity is the product of the number of elements of the coarse grid and number of nodes of the fine grid. For a general mesh,

determining the location of the nonzeros is non-trivial, and an efficient search algorithm must be employed. In our case, such a search can be avoided by utilizing the block generated structure of the mesh. Between any coarse and fine grid, within each super block, one can easily identify which fine node belongs to which coarse element. In the mesh generation step, by constructing a data structure with the information of which node belongs to which element, the prolongation operators can be formed efficiently through a table look up. The data structure is not large with size equal to twice the number of fine grid nodes, since it only consists of the fine grid node number and the corresponding coarse grid element number it belongs to.

For discretization of the fine grid with linear finite elements, it is natural to assume discretization of the coarse grid with linear finite elements leading to linear shape functions $N_i^c(\mathbf{x})$ that are evaluated for construction of the prolongation operator. For the case of discretization of the fine grid with quadratic finite elements, one has two choices, quadratic discretization of the coarse mesh leading to quadratic shape functions evaluated for prolongation operator construction, or linear discretization of the coarse mesh leading to linear shape functions evaluated for prolongation operator construction. In the case of nested meshes, where the coarse grid nodes are a subset of the fine grid nodes, selection of linear finite elements lead to a coarse grid operator which is a linear discretization of the fine grid mesh with linear finite elements with corners matching those of the original quadratic finite elements. Essentially the lower order finite element space is being used for the coarse grid correction, which is similar to the idea of using lower order finite element spaces as preconditioners for higher order elements [134].

The use of higher order finite elements only on the finest grid, and linear finite elements on the coarser grids, lead to less sparser coarse grid operators, due to the smaller support of linear shape functions. For this reason, the prolongation operators are constructed from linear discretizations on the coarser grids. With regards to sparsity it is also preferable to have nested meshes, since non-nested meshes can increase the support.

The numerical results presented in Section 5.3.1 show that the proposed method of coarse grid

and prolongator construction is efficient and scalable.

2.3.3 Two-grid convergence factors

In order to exhibit the convergent behavior of the geometric multigrid method that we propose, two grid convergence factors are computed. The two grid convergence factor is defined as the norm of the iteration matrix for a two level multigrid method, and is a measure of the reduction in the error in one iteration. It is a necessary condition that the two grid convergence factor is smaller than one for the multigrid method to work. The convergence of the multigrid method for arbitrary number of grids, often called h independent convergence in literature where h is the characteristic mesh size of the finest grid h, is based on a convergent two grid multigrid method. The method in which multigrid convergence is generally proven, is outlined in the following.

Using the notation introduced in Section 2.3, a two multigrid algorithm is written as,

$$\mathbf{x}_{\text{new}} = \mathbf{M}_h \mathbf{x}_{\text{old}} + \mathbf{F}_h(\nu_1, \nu_2)\mathbf{b}, \tag{2.164}$$

$$\mathbf{M}_h := \mathbf{S}^{\nu_1}\left(\mathbf{I} - \mathbf{P}\mathbf{A}_c^{-1}\mathbf{R}\mathbf{A}\right)\mathbf{S}^{\nu_2}, \tag{2.165}$$

where $\mathbf{M}_h$ is the two grid multigrid iteration matrix, $\mathbf{F}(\nu_1, , \nu_2)$ is a matrix depending on the number of pre- and post smoothing iterations ν_1, ν_2 and $\mathbf{A}_c := \mathbf{RAP}$. The error in the solution between $\mathbf{x}_{\text{old}}$ and $\mathbf{x}_{\text{new}}$ is related by,

$$\mathbf{e}_{\text{new}} = \mathbf{S}^{\nu_1}\left(\mathbf{I} - \mathbf{P}\mathbf{A}_c^{-1}\mathbf{R}\mathbf{A}\right)\mathbf{S}^{\nu_2}\mathbf{e}_{\text{old}}. \tag{2.166}$$

By taking norms one obtains,

$$||\mathbf{e}_{\text{new}}|| \leq ||\mathbf{M}_h|| \cdot ||\mathbf{e}_{\text{old}}|| \leq ||\mathbf{S}^{\nu_1}|| \cdot ||\left(\mathbf{A}^{-1} - \mathbf{P}\mathbf{A}_c^{-1}\mathbf{R}\right)|| \cdot ||\mathbf{A}\mathbf{S}^{\nu_2}|| \cdot ||\mathbf{e}_{\text{old}}||. \tag{2.167}$$

The h independent convergence of multigrid is proven in two steps [87].

1. An h independent two grid convergence factor ρ_{2grid} is derived from a combination of the bound on the smoothing properties of the smoother,

$$||\mathbf{S}^{\nu_1}|| \leq 1, \qquad ||\mathbf{A}\mathbf{S}^{\nu_2}|| \leq \eta(\nu_2)h^{-m} \text{ with } \eta(\nu_2) \to 0(\nu_2 \to \infty), \tag{2.168}$$

and approximation properties of the prolongation and restriction operators,

$$|| \left(\mathbf{A}^{-1} - \mathbf{P}\mathbf{A}_c^{-1}\mathbf{R} \right) || \leq Ch^m \tag{2.169}$$

where $m > 0$ and C is a constant. These yield the relation,

$$||\mathbf{M}_h|| \leq C\eta(\nu_2) =: \rho_{2grid}, \tag{2.170}$$

which is h independent, and smaller than 1 for sufficiently large ν_2.

2. A recursion formula for the convergence factor for n grids ρ_n is derived as follows.

Assume there are L grids, and that on grid l, the multigrid iteration matrix is denoted by $\mathbf{M}_l$ and the operator is denoted by $\mathbf{A}_l$. Then $\mathbf{M}_l$ is defined recursively by,

$$\begin{aligned}
\mathbf{M}_l &= \mathbf{S}^{\nu_1} \left(\mathbf{I}_l - \mathbf{P}_{l-1}^l \left[\mathbf{I}_{l-1} - \mathbf{M}_{l-1}^\gamma \right] \mathbf{A}_{l-1}^{-1} \mathbf{R}_l^{l-1} \mathbf{A}_l \right) \mathbf{S}^{\nu_2}, \\
&= \mathbf{S}^{\nu_1} \left(\mathbf{I}_l - \mathbf{P}_{l-1}^l \mathbf{A}_{l-1}^{-1} \mathbf{R}_l^{l-1} \mathbf{A}_l \right) \mathbf{S}^{\nu_2} + \mathbf{S}^{\nu_1} \left(\mathbf{P}_{l-1}^l \mathbf{M}_{l-1}^\gamma \mathbf{A}_{l-1}^{-1} \mathbf{R}_l^{l-1} \mathbf{A}_l \right) \mathbf{S}^{\nu_2}, \\
&= \mathbf{M}_{2grid} + \left(\mathbf{S}^{\nu_1} \mathbf{P}_{l-1}^l \right) \mathbf{M}_{l-1}^\gamma \left(\mathbf{A}_{l-1}^{-1} \mathbf{R}_l^{l-1} \mathbf{A}_l \mathbf{S}^{\nu_2} \right).
\end{aligned} \tag{2.171}$$

Here γ denotes the number of times the iteration matrix $\mathbf{M}_{l-1}^\gamma$ is applied at level $l-1$. $\gamma = 1$ corresponds to a multigrid V-cycle (Figure 2.23), and $\gamma = 2$ corresponds to a multigrid W-cycle. In the W-cycle, at each level the multigrid iteration matrix is applied twice. By defining,

$$\rho_l := ||\mathbf{M}_l||, \tag{2.172}$$

$$C := ||\mathbf{S}^{\nu_1} \mathbf{P}_{l-1}^l|| \cdot ||\mathbf{A}_{l-1}^{-1} \mathbf{R}_l^{l-1} \mathbf{A}_l \mathbf{S}^{\nu_2}||, \tag{2.173}$$

and taking norms of Equation (2.171), one obtains,

$$\rho_l \leq \rho_{2grid} + C\rho_{l-1}^\gamma. \tag{2.174}$$

For the case of $4C\rho_{2grid} < 1$ and $\gamma = 2$ (Multigrid W-Cycle), as the limit $n \to \infty$ is taken a convergence rate $\rho^* < 1$ independent of the number of grids can be obtained. Two prove multigrid convergence for the V-cycle, a slight refinement of the proof must be made [40].

The discretization of elliptic partial differential equations such as Poisson's equation and standard linear elasticity lead to positive definite systems for which the process denoted above can be used to prove multigrid h independent convergence. For cases when the linear system is not of this type, difficulty can arise. When the system is indefinite, standard stationary smoothers such as Gauss-Seidel lose their norm decreasing property $||\mathbf{S}|| \leq 1$, and the error in modes which have negative eigenvalues are enhanced. When the number of these modes are large, one must use other smoothers such as the Kaczmarz smoother which smooths unconditionally but has a very slow convergence rate with $||\mathbf{S}||$ close to unity. For the case of Helmholtz's equation, when the shift ω is large, the approximation property of Equation (2.169) is lost, which restricts the size of the coarse grid possible in such applications [22, 41].

Though full theoretical proof cannot be established for two grid convergence and h independent convergence for the elastodynamic equation with PML, verification of the two grid convergence factor through numerical computation suffices to give a good idea on whether the multigrid method is convergent and can extend to multilevels. In this section, the two grid convergence factor for the 2D scalar wave equation and 2D elastodynamic equation with PML are computed for different PML parameters. The smoothing factor is also computed through LFA for the 2D scalar wave, 2D elastodynamic, 3D scalar wave, and 3D elastodynamic problem for different PML parameters. These simulations present the range of PML parameters that can be chosen for convergent behavior of the multigrid method that we propose.

2D scalar wave

In this section the 2D scalar wave equation under time-harmonic point excitation, i.e., Helmholtz's equation, is considered on a square domain.

$$\frac{d^2\hat{u}}{d\tilde{x}^2} + \frac{d^2\hat{u}}{d\tilde{y}^2} + k^2\hat{u} = \delta(0,0), \quad (x,y) \in [-L_d, L_d] \times [-L_d, L_d], \tag{2.175}$$

$$L_d := L_b + L_p, \tag{2.176}$$

$$\tilde{x} = \int_0^x \lambda(s)ds, \tag{2.177}$$

$$\tilde{y} = \int_0^y \lambda(s)ds, \tag{2.178}$$

$$\lambda(s) = 1 - \sigma(s)i, \tag{2.179}$$

$$\sigma(s) = \begin{cases} 0 & 0 \le s < L_b \\ \text{some value} & L_b \le s \le L_d \end{cases} \tag{2.180}$$

$$\hat{u}(x,y) = 0 \qquad \text{for } (|x| = L_d \text{ or } |y| = L_d). \tag{2.181}$$

Local Fourier Analysis

The effectiveness of the Gauss-Seidel and Kaczmarz smoother for bilinear finite elements are investigated through the method of LFA. Since LFA requires constant coefficients, only the constant absorbing function profile $\gamma(s) = 1$ is considered. For smoothing, the problematic case arises when PML is applied in one direction only, and thus only the case of constant PML in the x direction is considered. To focus on the effect of β on the smoothing factor, ω is set to zero in $\mathbf{K} - \omega\mathbf{M}$. The smoothing factors $\mu_{\text{x,loc}}, \mu_{\text{y,loc}}$ defined in Equation (2.129) and (2.130) with respect to varying β are shown in Figure 2.27. The smoothing factor μ for Gauss-Seidel exceeds 1 at $\beta = 0.7$. This predicts failure of the multigrid for such PML parameters, which is confirmed in the actual two grid convergence factor computations. Contrary to this, the smoothing factor for Kaczmarz has $\mu \le 1$ for all β, confirming the unconditional convergence of the method. One does see though that the smoothing factor is unacceptable for large β.

Figure 2.28 shows the smoothing factor in the (θ_x, θ_y) plane for $\beta = \{0, 0.7, 5\}$. $\beta = 0$ corresponds to the symmetric positive definite case. For this value of β, the smoothing property of

Kaczmarz is weaker than Gauss-Seidel for all modes, since most of the modes near the origin have smoothing factors close to 0.8 or larger. For $\beta = 0.7$, the point of failure of the Gauss-Seidel iteration, μ becomes large in the region of high oscillation in both x, y directions. In the extreme case of $\beta = 5$, one observes anisotropic behavior in smoothing, such that smoothing exists only in the y direction. This behavior can be explained from the governing Helmholtz's equation. Under the application of PML, Equation (2.175) can be rewritten as,

$$\frac{1}{(1-i\beta)^2}\frac{d^2\hat{u}}{dx^2} + \frac{d^2\hat{u}}{dy^2} + k^2\hat{u} \quad = \quad \delta(0,0), \tag{2.182}$$

in the unstretched coordinates. As β is increased, the coefficient in front of the derivatives in the x direction increases, leading to anisotropy and weak coupling in the x direction for which smoothing is difficult. For such a case, an anisotropic multigrid with semi-coarsening only in the y direction can be attempted, but as presented in Section 2.2, such high β values are unnecessary for small reflection.

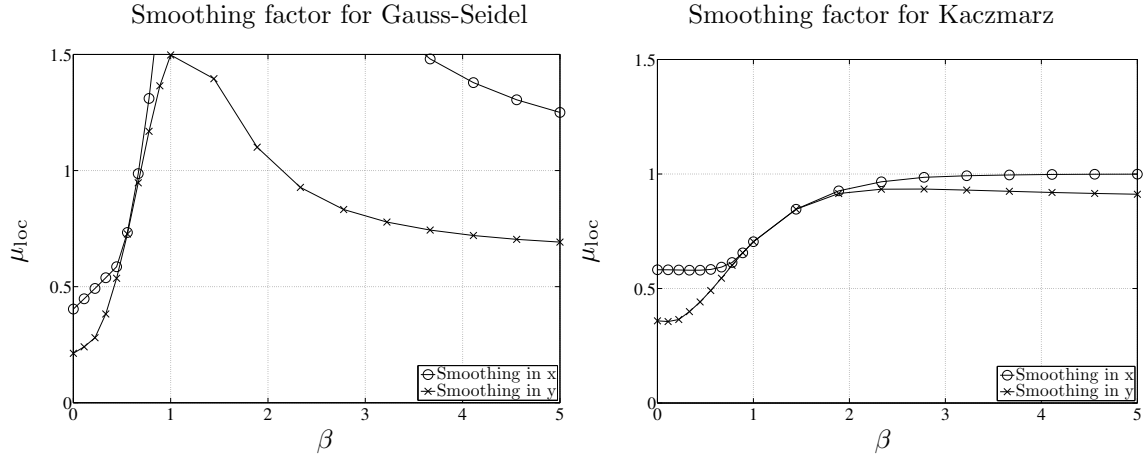


Figure 2.27: Smoothing factor for 2D scalar wave equation with PML in the x direction for Gauss-Seidel and Kaczmarz smoothers

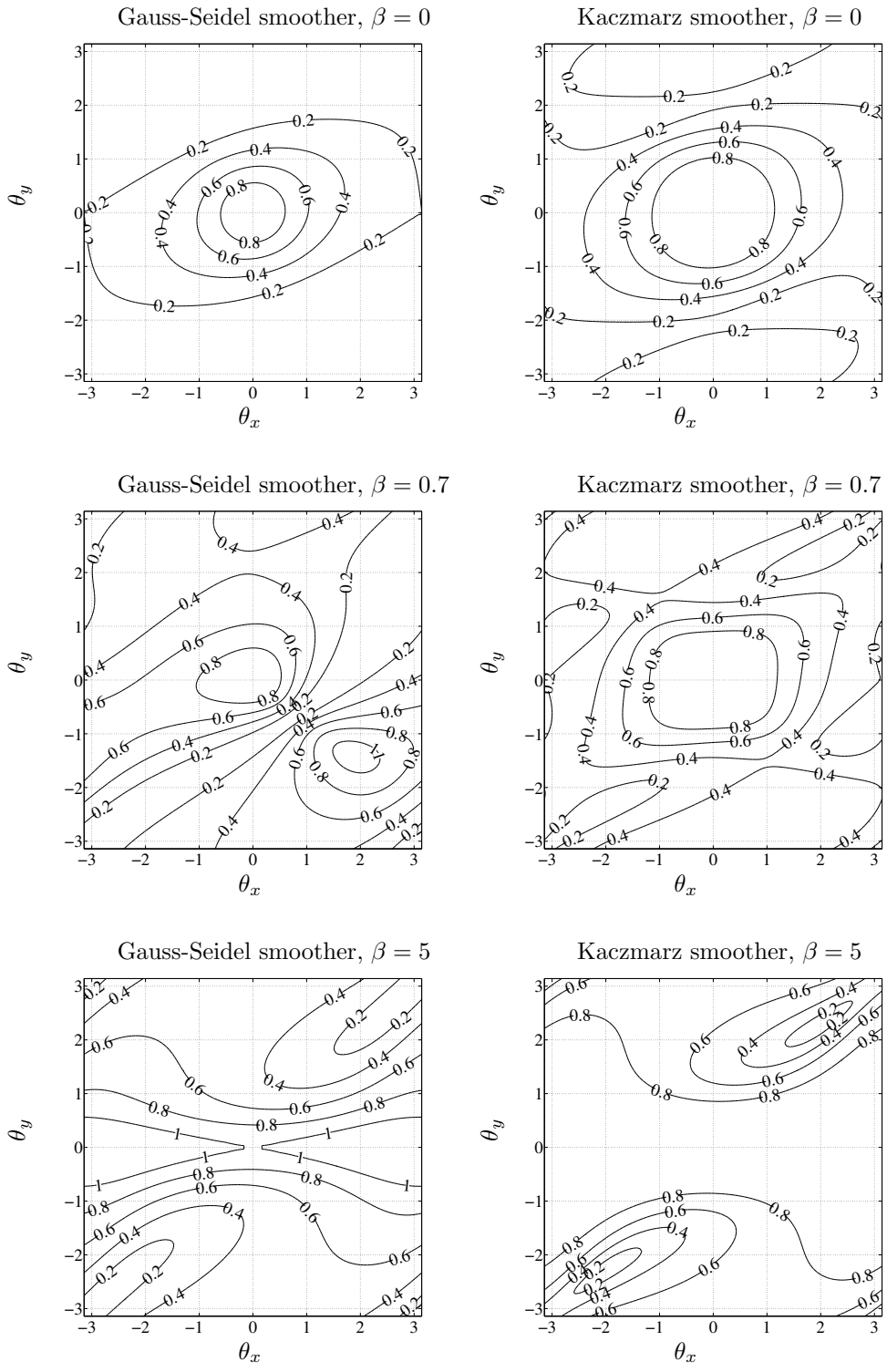


Figure 2.28: Smoothing factors for Gauss-Seidel and Kaczmarz at $\beta = \{0, 0.7, 5\}$

Two grid convergence factor with constant PML absorbing profile $\gamma(s) = 1$

Two grid convergence factors for bilinear elements with an absorbing function profile of $\gamma(s) = 1$ are computed. This absorbing function profile is selected for comparison with the LFA results. The parameters varied are the number of nodes per wave $n_{\text{npw}} \in [6, 24]$ and $\beta \in [0, 1]$. For this two grid method, the number of nodes per wave on the coarse grid is half of n_{npw}. The number of waves in the bounded domain n_{wbd} and the number of waves in the PML n_{wpml} are set to the following 4 cases.

1. Figure 2.29 left : $(n_{\text{wbd}}, n_{\text{wpml}}) = (1, 1)$
2. Figure 2.29 right: $(n_{\text{wbd}}, n_{\text{wpml}}) = (2, 1)$
3. Figure 2.30 left : $(n_{\text{wbd}}, n_{\text{wpml}}) = (1, 2)$
4. Figure 2.30 right: $(n_{\text{wbd}}, n_{\text{wpml}}) = (2, 2)$

The shift ω in the linear system of equations solved,

$$\left(\mathbf{K} - \omega^2 \mathbf{M}\right) \quad = \quad \mathbf{F} \tag{2.183}$$

is set to,

$$\omega \quad = \quad c\frac{2\pi}{h\ n_{\text{npw}}}, \tag{2.184}$$

where c is the wave speed and h is the distance between the nodes, so that propagating waves with n_{npw} per wave length are excited.

From these figures, one can make the following comments on the effectiveness of each smoother.

- Gauss-Seidel: Multigrid divergence occurs at $\beta = 0.7$, as LFA predicts, and a representation of approximately 6 nodes per wave is required on the coarse grid for convergence.
- Kaczmarz: Convergence is obtained for values of β up to 1 as LFA predicts. Convergence is obtained for fairly small n_{npw} on the coarse grid, since the smoother is able to smooth error

modes that are not appropriately reduced by the coarse grid correction. For fast convergence, a representation of approximately 6 nodes per wave is required on the coarse grid for convergence.

- Chebyshev: Convergence factors are quite insensitive to β up to 1. For fast convergence, a representation of approximately 6 nodes per wave is required on the coarse grid for convergence.

In general the following comments can be made.

- Increase in the size of the boundary n_{wbd} leads to increase in convergence factor. (Slow convergence).
- Increase in the size of the PML n_{wpml} leads to decrease in convergence factor. (Fast convergence).
- Divergence occurs for $\beta \leq 0.1$. In the case of $\beta = 0$, the real-symmetric indefinite system is solved.

These observations are all related to the behavior of multigrid with nonzero shift ω. The cause and further investigation as well as the explanation of this behavior is postponed to the next example of 2D elastodynamics.

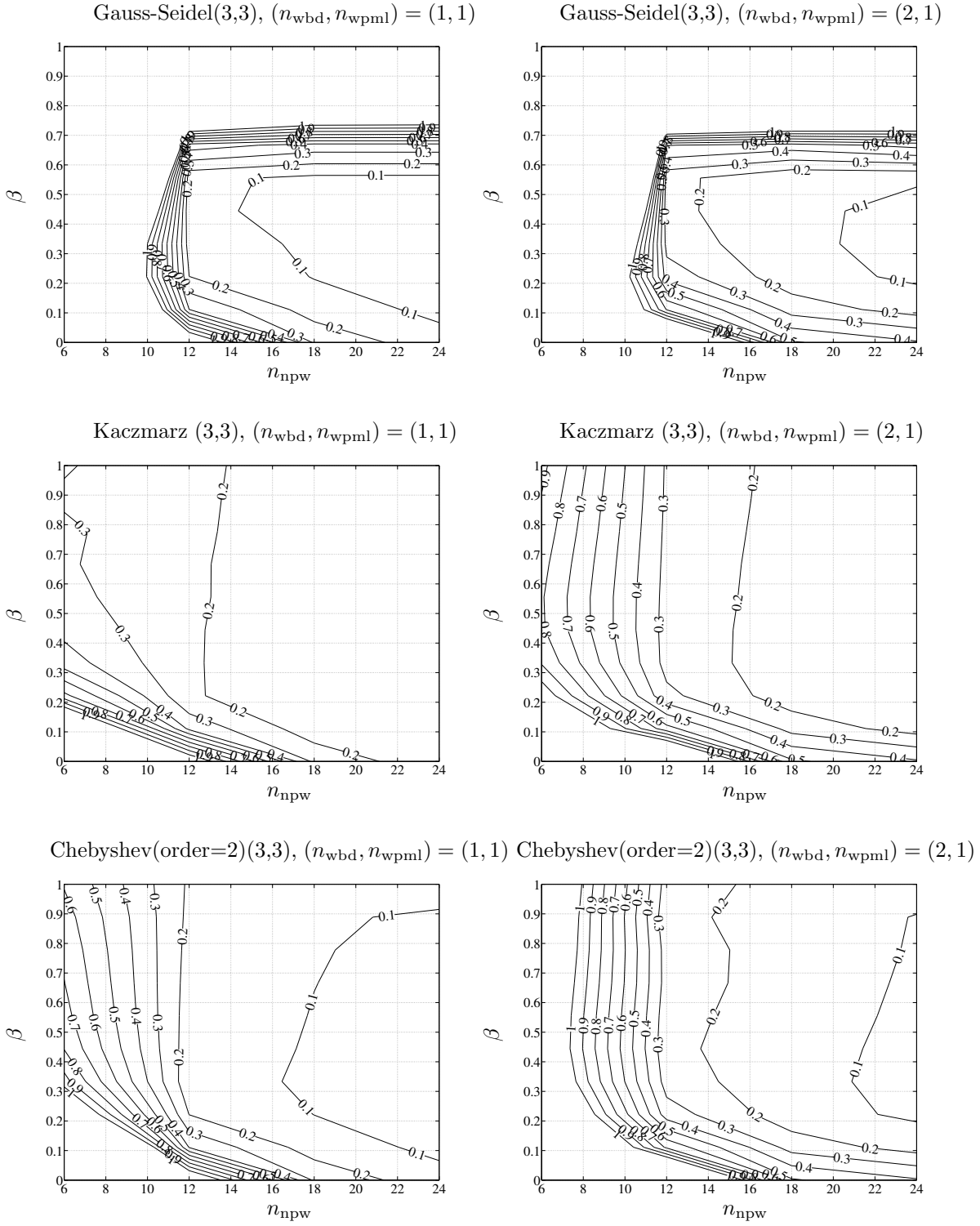


Figure 2.29: 2D scalar wave equation: two grid convergence factor for linear elements, $\boldsymbol{\gamma(s) = 1}$, $n_{\text{wbd}} = \{1, 2\}$, $\boldsymbol{n}_{\text{wpml}} \boldsymbol{= 1}$

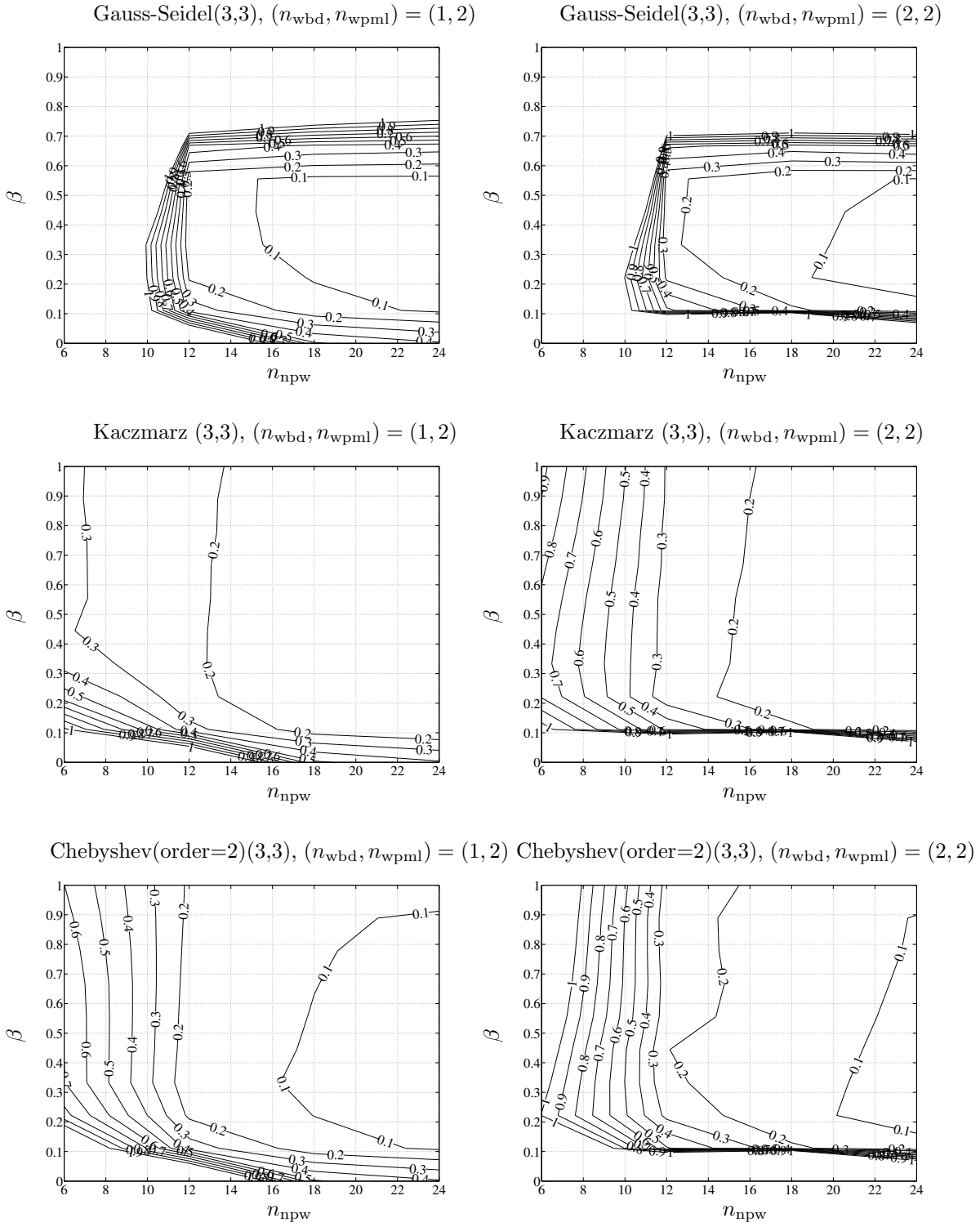


Figure 2.30: 2D scalar wave equation: two grid convergence factor for linear elements, $\boldsymbol{\gamma(s) = 1}$, $n_{\text{wbd}} = \{1, 2\}$, $\boldsymbol{n_{\text{wpml}} = 2}$

Two grid convergence factor with linear PML absorbing profile $\gamma(s) = s$

The two grid convergence factors for bilinear elements with a linear absorbing function profile of $\gamma(s) = s$ is also computed. The parameters varied are number of nodes per wave $n_{\text{npw}} \in [6, 24]$ and $\beta \in [0, 1]$. For this two grid method, the number of nodes per wave on the coarse grid is half of n_{npw}. The number of waves in the bounded domain n_{wbd} and the number of waves in the PML n_{wpml} are set to the following 2 cases.

1. Figure 2.31 left : $(n_{\text{wbd}}, n_{\text{wpml}}) = (1, 1)$

2. Figure 2.31 right: $(n_{\text{wbd}}, n_{\text{wpml}}) = (2, 1)$

One obtains similar results to the constant absorbing function profile case, with the exception that the behavior for Gauss-Seidel is slightly better for larger values of β.

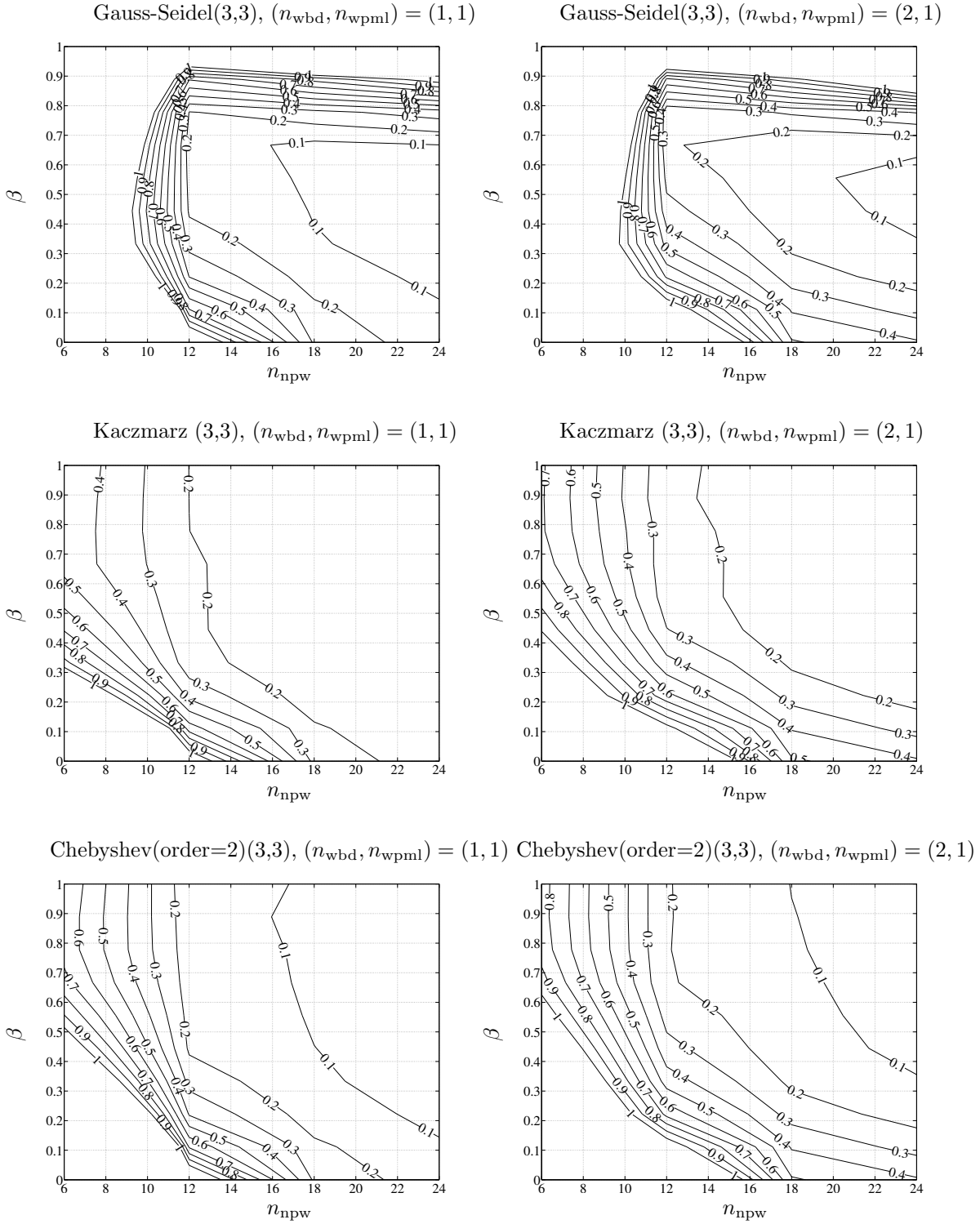


Figure 2.31: 2D scalar wave equation: two grid convergence factor for linear elements, $\boldsymbol{\gamma(s) = s}$, $n_{\text{wbd}} = \{1, 2\}$, $\boldsymbol{n}_{\text{wpml}} = \mathbf{1}$

2D elastodynamics

In this section the 2D elastodynamic equation under time-harmonic excitation, is considered on a rectangular domain. This model is supposed to mimic behavior of a semi-infinite half domain.

$$(\lambda+\mu)\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\hat{\mathbf{u}})+\mu\boldsymbol{\nabla}^2\hat{\mathbf{u}} = -\omega^2\rho\hat{\mathbf{u}}, \quad (x,y)\in[-L_d,L_d]\times[-L_d,0], \tag{2.185}$$

$$L_d := L_b+L_p, \tag{2.186}$$

$$\tilde{x} = \int_0^x \lambda_x(s)ds, \tag{2.187}$$

$$\tilde{y} = \int_0^y \lambda_y(s)ds, \tag{2.188}$$

$$\lambda(s) = 1-\sigma(s)i, \tag{2.189}$$

$$\sigma(s) = \begin{cases} 0 & 0\le|s|<L_b \\ \text{some value} & L_b\le|s|\le L_d \end{cases} \tag{2.190}$$

$$\hat{u}(x,y) = 0 \qquad \text{for } (|x|=L_d \text{ or } |y|=L_d), \tag{2.191}$$

$$\mathbf{t}\cdot\mathbf{n}(0,0) = (0,1)^T \tag{2.192}$$

Local Fourier Analysis

The effectiveness of the Gauss-Seidel and Kaczmarz smoother for bilinear finite elements are investigated through the method of LFA. Since LFA requires constant coefficients, only the constant absorbing function profile $\gamma(s)=1$ is considered. For smoothing, the problematic case arises when PML is applied in one direction only, and thus only the case of constant PML in the x direction is considered. To focus on the effect of β on the smoothing factor, ω is set to zero in $\mathbf{K}-\omega\mathbf{M}$. The smoothing factors $\mu_{\mathrm{x,loc}}, \mu_{\mathrm{y,loc}}$ defined in Equation (2.134) and (2.135) for each direction with respect to varying β are shown in Figure 2.32. The smoothing factor μ for Gauss-Seidel exceeds 1 at $\beta=0.7$. This predicts failure of the multigrid for such PML parameters, which is confirmed in the actual two grid convergence factor computations. Contrary to this, the smoothing factor for Kaczmarz has $\mu\le 1$ for all β, confirming the unconditional convergence of the method. One does see though that the smoothing factor is close to unity even for small β, i.e., the smoothing and consequently convergence will be quite slow

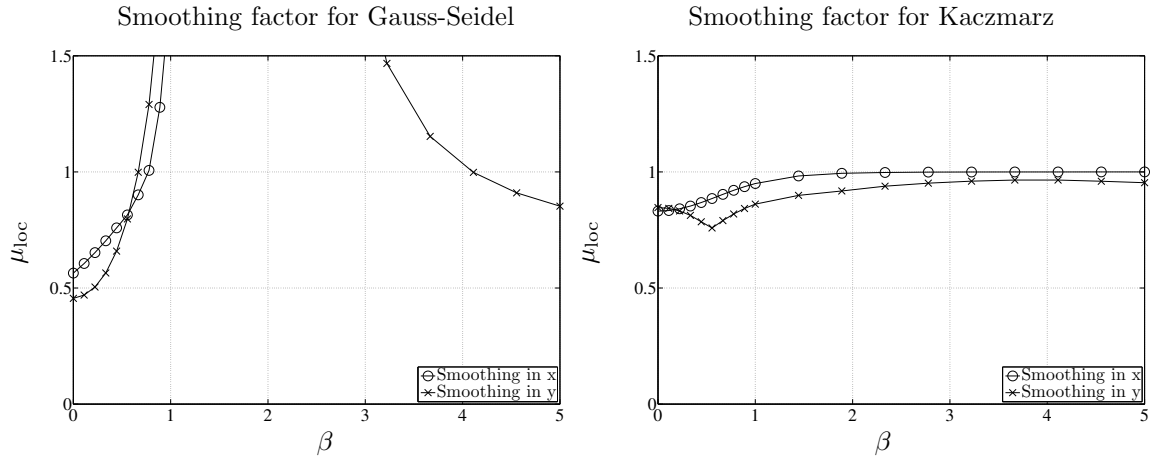


Figure 2.32: Smoothing factor for 2D elastodynamic equation with PML in the x direction for Gauss-Seidel and Kaczmarz smoothers

Two grid convergence factor with constant PML absorbing profile $\gamma(s) = 1$

Two grid convergence factors for bilinear elements with an absorbing function profile of $\gamma(s) = 1$ are computed. This absorbing function profile is selected for comparison with the LFA results. The parameters varied are the number of nodes per shear wave $n_{\text{npw,s}} \in [6, 24]$ and $\beta \in [0, 1]$. For this two grid method, the number of nodes per shear wave on the coarse grid is half of $n_{\text{npw,s}}$. The number of shear waves in the bounded domain is set to $n_{\text{wbd,s}} = 2$ and the number of shear waves in the PML is set to $n_{\text{wpml,s}} = 2$. For this 2D elasticity problem, the shear wave length λ_s is selected as the reference wave length, since $\lambda_v \approx 1.8\lambda_s(\nu = 0.3)$ and the wave length of the Rayleigh waves $\lambda_{\text{rayleigh}} \approx \lambda_s$. Adequate discretization of the shear wave will imply adequate discretization of the other types of waves. For appropriate attenuation of waves propagating in the PML, one must select the length of the PML in terms of λ_v.

The shift ω in the linear system of equations solved,

$$\left(\mathbf{K} - \omega^2\mathbf{M}\right) \quad = \quad \mathbf{F} \tag{2.193}$$

is set to,

$$\omega \quad = \quad \omega_s := c_s \frac{2\pi}{h\ n_{\text{npw,s}}}, \tag{2.194}$$

in the left side of Figure 2.33, and

$$\omega \quad = \quad 0, \tag{2.195}$$

in the right side of Figure 2.33. Here c_s is the shear wave speed and h is the distance between the nodes. The forcing frequency of ω_s is selected to excite shear waves with $n_{\text{npw,s}}$ nodes per shear wave. Since the excitation mode is not purely shear, volumetric and Rayleigh wave are also excited, but since these waves are of the same wave length as the shear wave or longer, they have wave discretizations of $n_{\text{npw,s}}$ or more. The case $\omega = 0$, gives the ideal convergence in the absence of added indefiniteness in the linear system that multigrid must treat. By comparing the two figures, it

is clear that lack of convergence for small β is due to a nonzero ω. With $\beta = 0$ one obtains the real-symmetric indefinite problem, for which it is clear that with $\omega = \omega_s$, convergence is unattainable. The explanation for convergence in the case of mild β parameters follows from the argument made by Elman [67] and Brandt [42] in their explanation of the difficulty in applicability of multigrid to the real-symmetric indefinite equation.

Let us solve the trivial problem,

$$\mathbf{A}_f \mathbf{x}_f = \mathbf{0}, \tag{2.196}$$

with non-singular fine grid matrix $\mathbf{A}_f$ by application of two-grid multigrid. The initial guess on the fine grid $\mathbf{x}_f^0$ is assumed $\mathbf{v}_f$, an eigenvector of $\mathbf{A}_f$ with eigenvalue λ_f. This mode $\mathbf{v}_f$ is assumed smooth enough such that application of smoothing does not change the mode $\mathbf{S}\mathbf{v}_f \approx \mathbf{v}_f$. The residual on the fine grid is,

$$\mathbf{r}_f = -\mathbf{A}_f \mathbf{v}_f = -\lambda_f \mathbf{v}_f. \tag{2.197}$$

If the coarse grid is fine enough such that $\mathbf{v}_f$ is reasonably well represented, the restriction of $\mathbf{v}_f$ onto the coarse grid $\mathbf{R}\mathbf{v}_f$ will be close to an eigenvector $\mathbf{v}_c$ of the coarse grid operator $\mathbf{A}_c$ with eigenvalue λ_c. The correction on the coarse grid is,

$$\mathbf{A}_c \mathbf{e}_c = \mathbf{R}\mathbf{r}_f = -\lambda_f \mathbf{v}_c \tag{2.198}$$

and thus,

$$\mathbf{e}_c = -\lambda_f \mathbf{A}_c^{-1} \mathbf{v}_c = -\frac{\lambda_f}{\lambda_c} \mathbf{v}_c, \tag{2.199}$$

leading to the expression.

$$\begin{aligned} \mathbf{x}_f^{new} &= \mathbf{x}_f^0 + \mathbf{P}\mathbf{e}_c, \\ &= \left(1 - \frac{\lambda_f}{\lambda_c}\right) \mathbf{v}_f. \end{aligned} \tag{2.200}$$

For error reduction, one requires,

$$\left| 1 - \frac{\lambda_f}{\lambda_c} \right| \leq 1. \tag{2.201}$$

When the eigenvalues are identical on the two grids, an error of this mode is completely nulled. When they differ largely, an error of this mode can be enhanced. For the low modes in a symmetric positive definite systems such as Poisson's equation and quasistatic linear elasticity, $\lambda_c > 0$, $\lambda_f > 0$, and $\lambda_f < \lambda_c$, which imply that all low error modes are reduced or at least not enhanced. For the real symmetric indefinite case, the sign of λ_c and λ_f are not determined and can be arbitrarily close to zero. When the eigenvalues are of opposite sign or if $\lambda_f > 2\lambda_c$, the error mode is enhanced.

With the application of PML, the linear system is no longer real-symmetric indefinite, and eigenvalues are no longer real. Qualitatively, for PML with β not too large, the eigenvalues for low modes on the coarse grid and fine grid can be expressed as,

$$\lambda_{f,pml} = \lambda_{f,nopml} + i\delta_f, \tag{2.202}$$

$$\lambda_{c,pml} = \lambda_{c,nopml} + i\delta_c, \tag{2.203}$$

where $\lambda_{f,nopml} \in \mathbb{R}$, $\lambda_{c,nopml} \in \mathbb{R}$ are the eigenvalues of the non-PML system and $\delta_f, \delta_c \in \mathbb{R}$. With the addition of the imaginary term, the reduction factor for problematic modes of the real-symmetric indefinite problem with $\lambda_{c,nopml} \approx 0$ and $\lambda_{f,nopml} << 1$ now become,

$$\left|1 - \frac{\lambda_f}{\lambda_c}\right| \approx \left|1 - \frac{\delta_f}{\delta_c}\right|, \tag{2.204}$$

which is smaller than one since for low modes $\delta_c > \delta_f$ and for properly damped modes $\delta_c > 0, \delta_f > 0$. Damping can be introduced in other forms other than PML, but have the same effect. Such advantages of added damping to the original equations for efficient solutions of the Helmholtz's equation have been conducted by Erlangga [68] in the modified Laird preconditioner for the scalar Helmholtz's equation. This analysis also shows why multigrid methods to solve the scalar Helmholtz equation obtained from potential discretization of 2D Maxwells equation [158, 116],

$$(\mathbf{K} + i\omega\mathbf{M})\,\mathbf{x} = \mathbf{F}, \tag{2.205}$$

with $\mathbf{K}, \mathbf{M}$ real-symmetric positive definite and $\omega \in \mathbb{R}$, work well. The complex shift acts as a shift of the real-valued spectrum away from the real axis, adding imaginary components to the real-valued

eigenvalues.

This analysis implies that for multigrid methods, solving the actual physical problem with damping and radiation type boundary conditions is in fact easier to solve than the system without any damping.

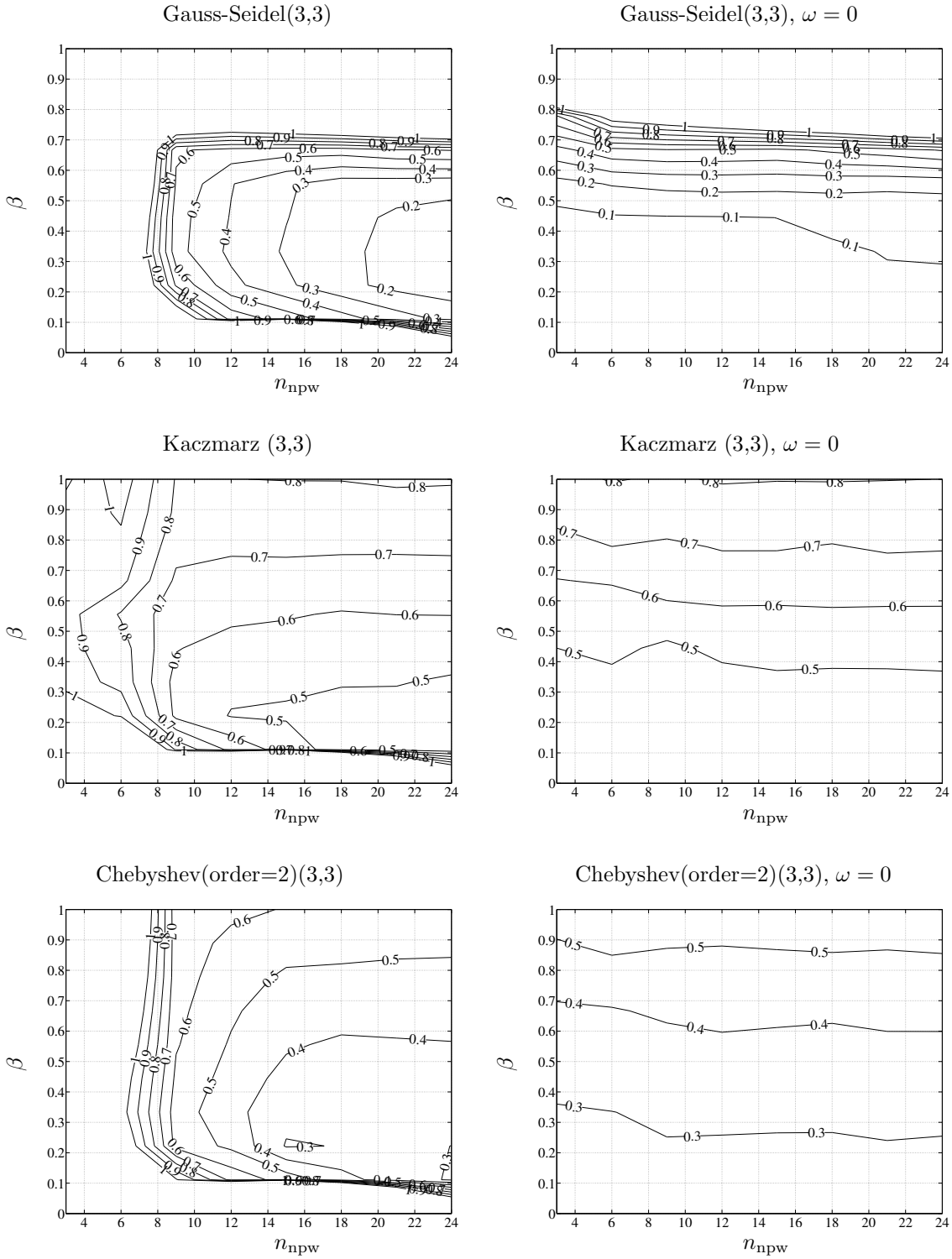


Figure 2.33: Elasticity 2D multigrid convergence factor for linear elements, with shift and with no shift $\omega = 0$, $\gamma(s) = 1$, $n_{\mathrm{wbd,s}} = 2$, $n_{\mathrm{wpml,s}} = 2$

Two grid convergence factor with linear PML absorbing profile $\gamma(s) = s$

The two grid convergence factors for bilinear elements with linear absorbing function profile of $\gamma(s) = s$ are also computed for the same parameters as the constant case. One obtains similar results to the constant absorbing function profile case, with the exception that the behavior for Gauss-Seidel is slightly better for larger values of β.

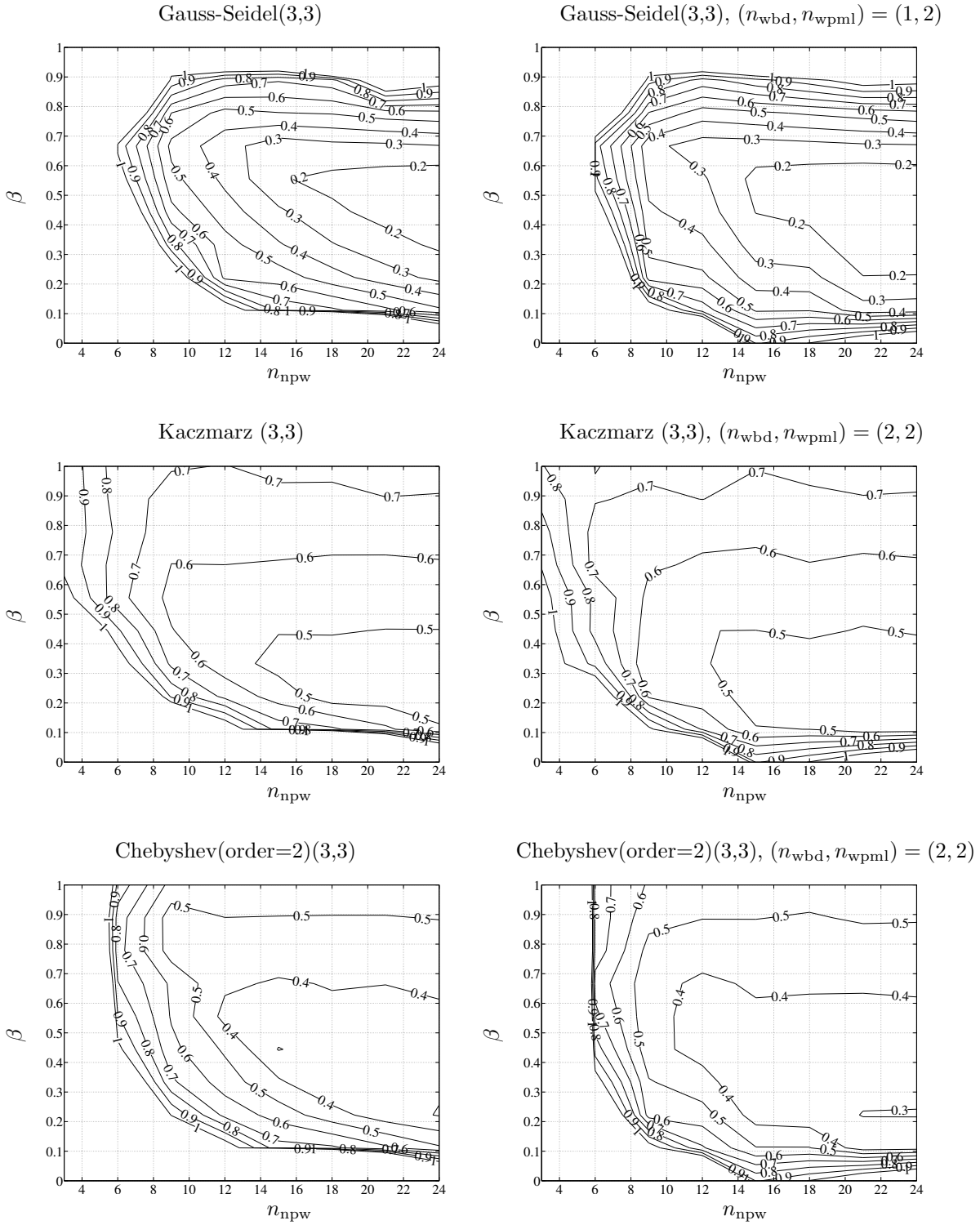


Figure 2.34: Elasticity 2D multigrid convergence factor for linear elements, $\gamma(s) = s$, $n_{\mathrm{wbd,s}} = 2$, $n_{\mathrm{wpml,s}} = 2$

Two grid convergence factor: effect of n_{wpml} and n_{wbd}

Here the effect of the convergence factor with respect to the size of the domain $n_{\text{wbd,s}}$ and the size of the PML $n_{\text{wpml,s}}$ is investigated.

The two grid convergence factors for bilinear elements with absorbing function profile of $\gamma(s) = 1$ are computed. For this two grid method, the number of nodes per shear wave on the coarse grid is half of $n_{\text{npw,s}}$. For this 2D elasticity problem, the shear wave length λ_s is selected as the reference wave length, since $\lambda_v \approx 1.8\lambda_s(\nu = 0.3)$ and the wave length of the Rayleigh waves $\lambda_{\text{rayleigh}} \approx \lambda_s$. Adequate discretization of the shear wave will imply adequate discretization of the other types of waves. On the other hand, for appropriate attenuation of waves propagating in the PML, one must select the length of the PML in terms of λ_v.

4 cases are considered.

1. Figure 2.35 left : The number of nodes per shear wave $n_{\text{npw,s}} = 24$ and the number of shear waves in the PML $n_{\text{wpml,s}} = 2$ is fixed. The number of shear waves in the bounded domain $n_{\text{wbd,s}} \in [0.5, 4]$ and $\beta \in [0, 1]$ are varied.

2. Figure 2.35 right: The number of nodes per shear wave $n_{\text{npw,s}} = 24$ and the number of shear waves in the bounded domain is fixed $n_{\text{wbd,s}} = 1$. The number of shear waves in the PML domain $n_{\text{wpml,s}} \in [0.5, 4]$ and $\beta \in [0, 1]$ are varied.

3. Figure 2.36 left : $\beta = 0.5$ is fixed and the number of shear waves in the PML $n_{\text{wpml,s}} = 2$ is fixed. The number of shear waves in the bounded domain $n_{\text{wbd,s}} \in [0.5, 4]$ and the number of nodes per shear wave $n_{\text{npw,s}}$ is varied.

4. Figure 2.36 right : $\beta = 0.5$ is fixed and the number of shear waves in the bounded domain $n_{\text{wbd,s}} = 2$ is fixed. The number of shear waves in the PML $n_{\text{wpml,s}} \in [0.5, 4]$ and the number of nodes per shear wave $n_{\text{npw,s}}$ is varied.

The following observations can be made.

1. Figure 2.35: For this discretizaion of $n_{\text{npw,s}}$ in this range of parameters, the convergence factor is stable with respect to increase in the size of the bounded domain and the PML domain.

2. Figure 2.36 right: Given any discretization, the convergence factor increases as the bounded domain is enlarged. (Slower convergence).

3. Figure 2.36 left: Given any discretization, the convergence factor is stable as the PML domain is enlarged.

It is clear that the selection of $n_{\text{npw,s}} = 24$ was fine enough so that an increase in the size of the domain was not affecting its convergence rate.

One can see from these results that for fast convergence, the bounded domain should be made as small as possible and elongation of the PML domain for moderate β does not cause a problem in terms of convergence. This can be explained qualitatively from an eigenvalue viewpoint. As the size of the bounded domain is enlarged, the given shift ω moves more and more into the interior of the spectrum, increasing the indefiniteness of the operator. The imaginary part of the eigenvalues close to the shift ω also decreases relatively, which can cause problems with the reduction factor introduced in Equation (2.204). As the size of the PML domain is increased, the ω may move into the interior of the spectrum but the imaginary part of the eigenvalue does not decrease in magnitude, which does not cause a problem with the reduction factor. Additionally, any new eigenvalues introduced that are smaller than ω have large imaginary parts which again does not degrade the reduction factor.

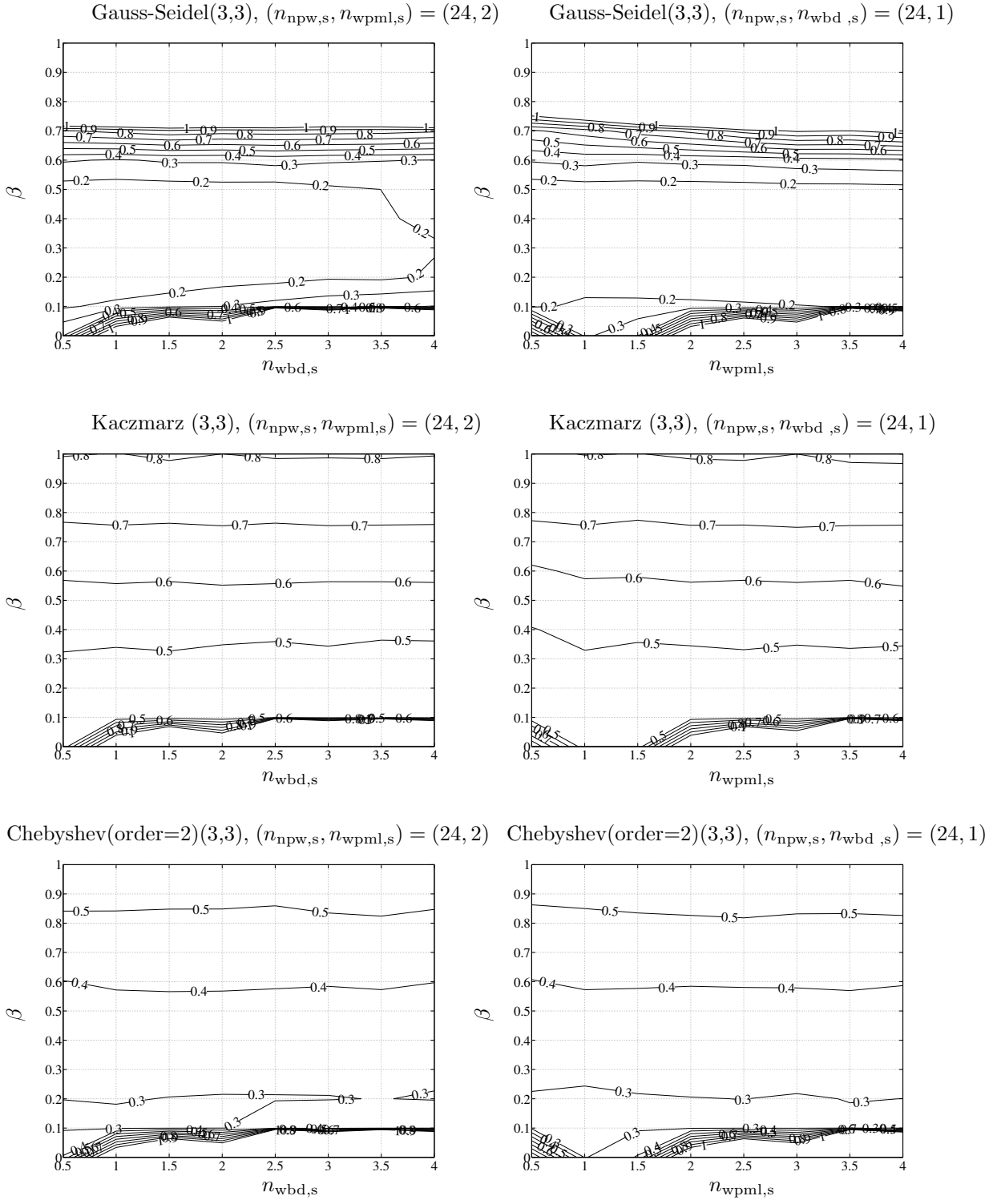


Figure 2.35: Elasticity 2D multigrid convergence factor for linear elements, $\gamma(s) = 1$, varying $n_{\text{wbd,s}} = \{1, 2\}$, fixing $n_{\text{wpml,s}} = 2$, $n_{\text{npw,s}} = 24$,

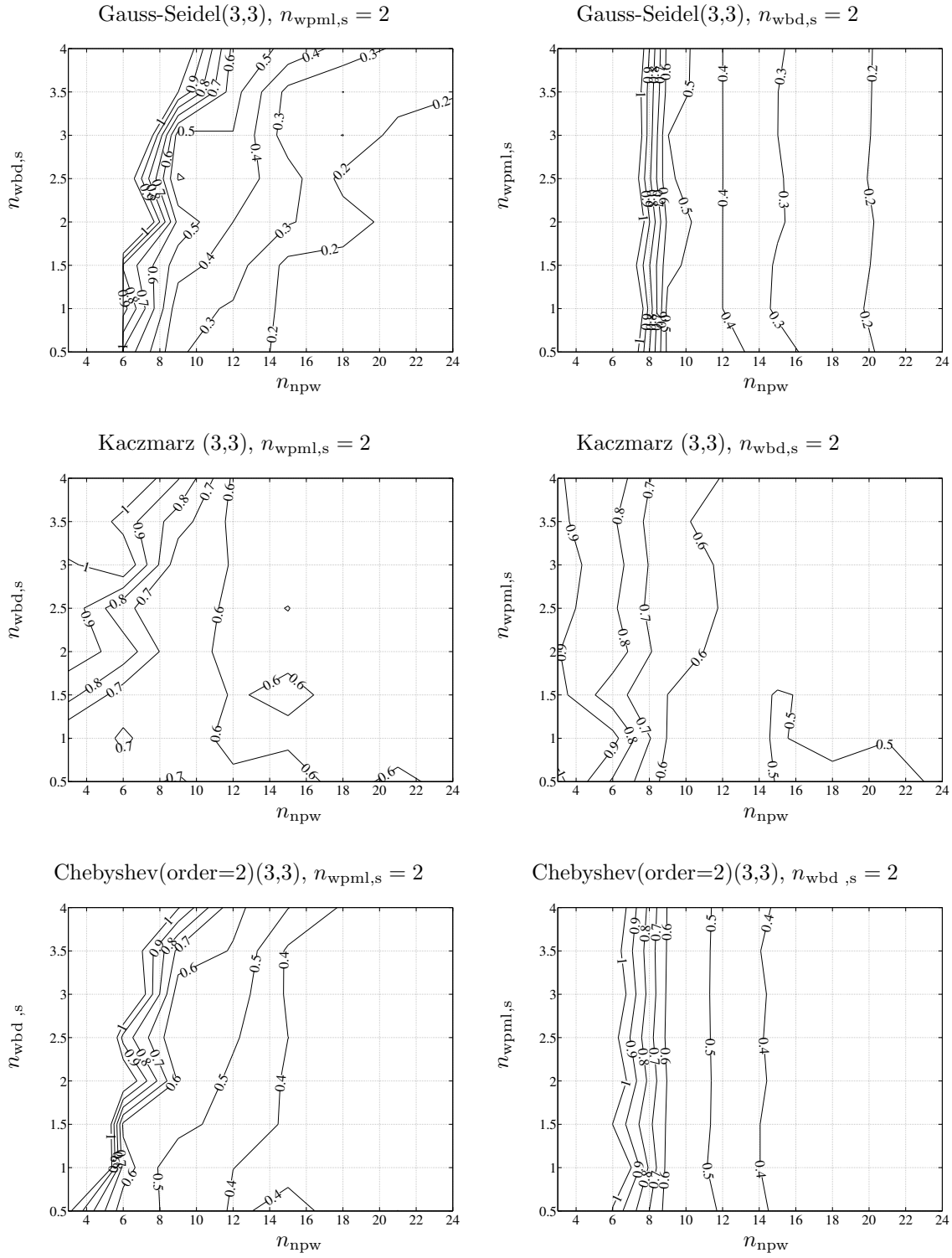


Figure 2.36: Elasticity 2D multigrid convergence factor for linear elements, $\gamma(s) = s$, $n_{\text{wbd,s}} = 2$, $n_{\text{wpml,s}} = 2$

Two grid convergence factor: higher order elements

In Section 2.2, it was observed that higher order elements lead to smaller reflection. Here the two grid convergence factors of biquadratic and bicubic elements with an absorbing function profile of $\gamma(s) = 1$ are computed. This absorbing function profile is selected for comparison with results from bilinear elements presented in Figure 2.33. The parameters varied are the number of nodes per shear wave $n_{\text{npw,s}} \in [6, 24]$ and $\beta \in [0, 1]$. For this two grid method, the number of nodes per shear wave on the coarse grid is half of $n_{\text{npw,s}}$. The number of shear waves in the bounded domain is set to $n_{\text{wbd,s}} = 2$ and the number of shear waves in the PML is set to $n_{\text{wpml,s}} = 2$. For this 2D elasticity problem, the shear wave length λ_s is selected as the reference wave length, since $\lambda_v \approx 1.8\lambda_s(\nu = 0.3)$ and the wave length of Rayleigh waves $\lambda_{\text{rayleigh}} \approx \lambda_s$. Adequate discretization of the shear wave will imply adequate discretization of the other types of waves. On the other hand, for appropriate attenuation of waves propagating in the PML, one must select the length of the PML in terms of λ_v.

1. Figure 2.37 left : biquadratic elements, $\omega = \omega_s$, $(n_{\text{wbd,s}}, n_{\text{wpml,s}}) = (2, 2)$
2. Figure 2.37 right: biquadratic elements $\omega = 0$, $(n_{\text{wbd,s}}, n_{\text{wpml,s}}) = (2, 2)$
3. Figure 2.38 left : bicubic elements $\omega = \omega_s$, $(n_{\text{wbd,s}}, n_{\text{wpml,s}}) = (2, 2)$
4. Figure 2.38 right: bicubic elements $\omega = 0$, $(n_{\text{wbd,s}}, n_{\text{wpml,s}}) = (2, 2)$

One observes that higher order elements lead to larger convergence factors, i.e. slower convergence. This effect is not just due to indefiniteness or the application, since one can see degradation in the convergence rate even for the symmetric positive definite case which corresponds to the $\beta = 0$ in the right side of Figures 2.37 and 2.38. The degradation in the Gauss-Seidel is slightly weaker than the other smoothers. The advantage of the Chebyshev smoother is the ability to increase the order for faster convergence rates.

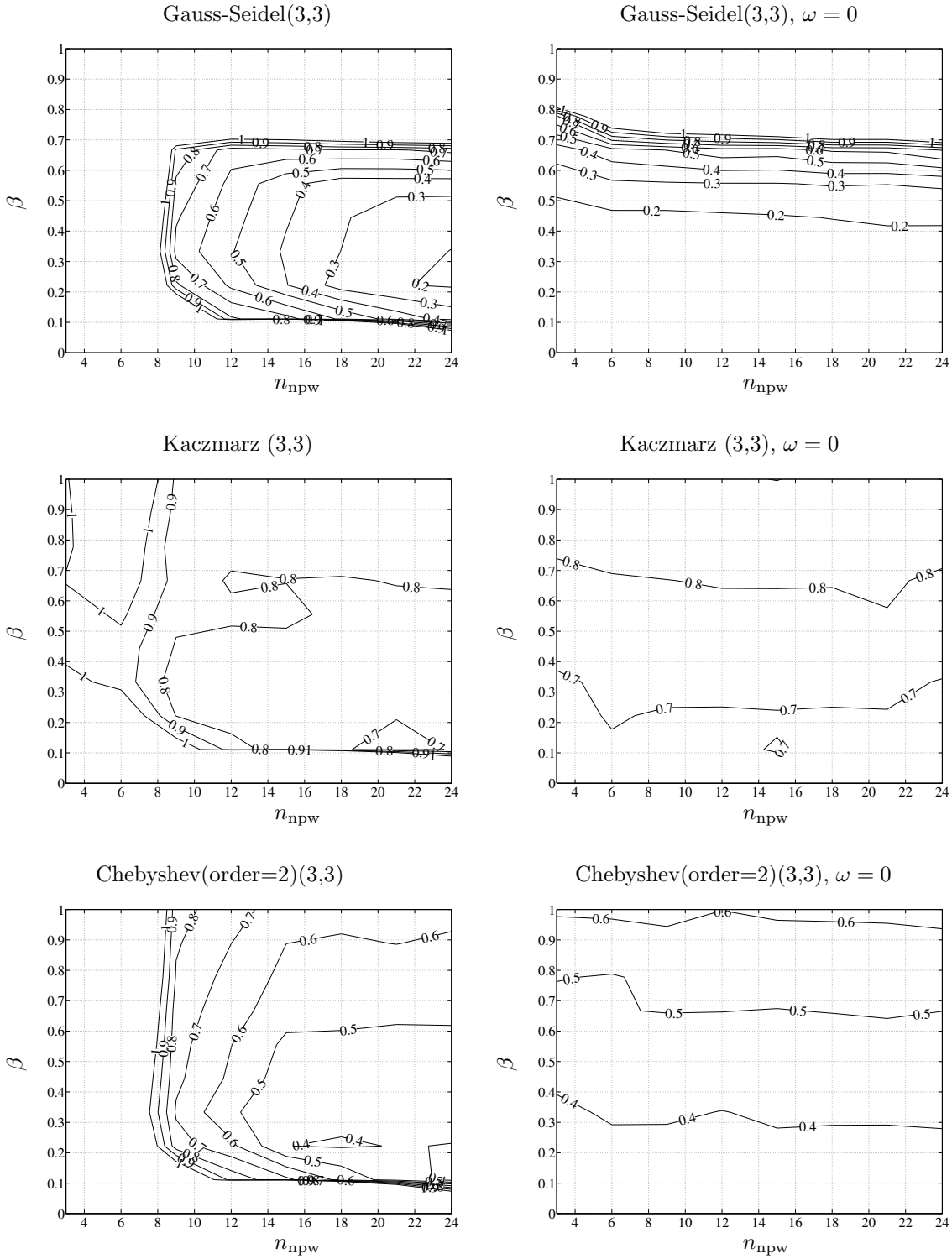


Figure 2.37: Elasticity 2D multigrid convergence factor for quadratic elements, with shift and with no shift $\omega = 0$, $\gamma(s) = 1$, $n_{\mathrm{wbd,s}} = 2$, $n_{\mathrm{wpml,s}} = 2$

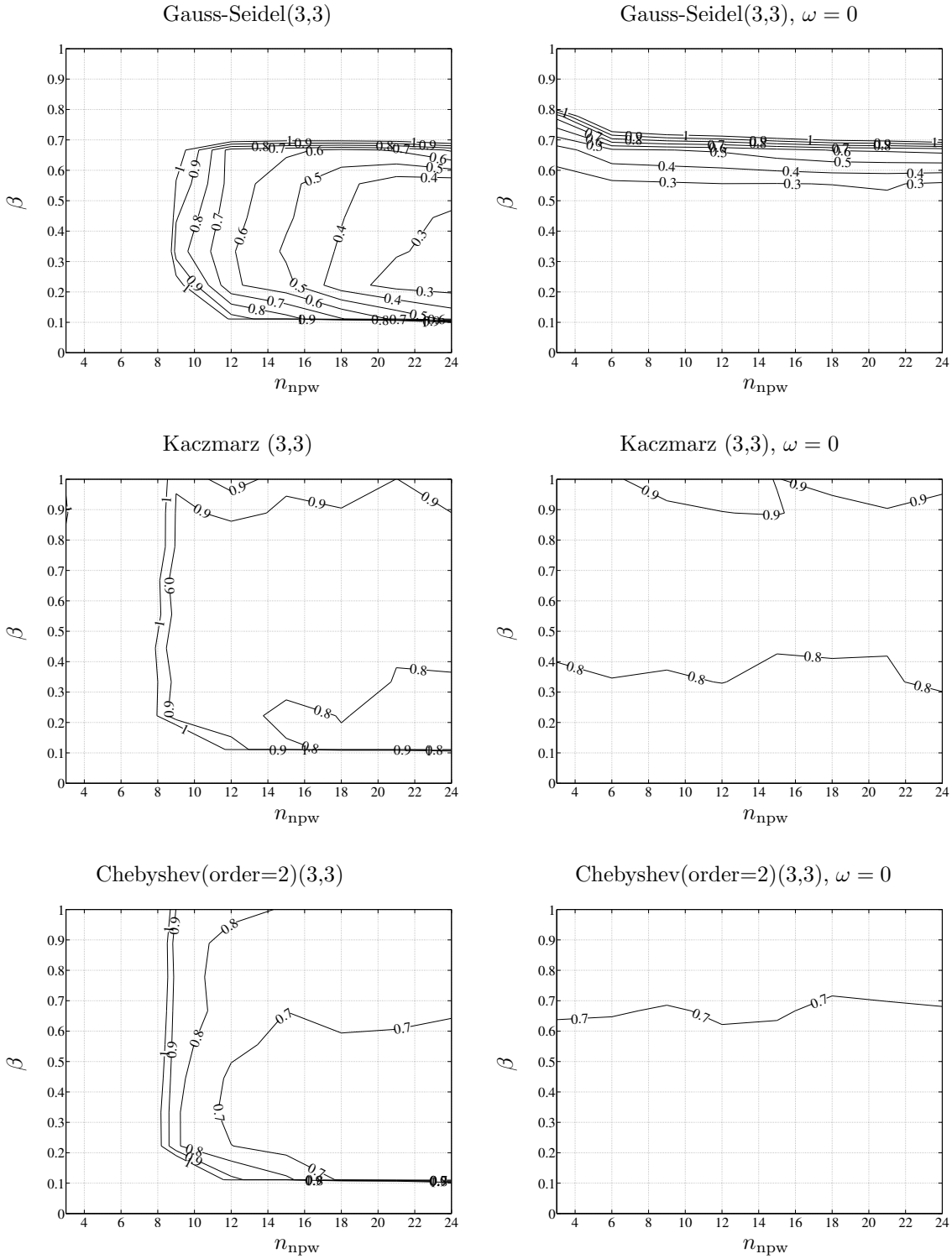


Figure 2.38: Elasticity 2D multigrid convergence factor for cubic elements, with shift and with no shift $\omega = 0$, $\gamma(s) = 1$, $n_{\mathrm{wbd,s}} = 2$, $n_{\mathrm{wpml,s}} = 2$

3D scalar wave

Local Fourier Analysis

In this section the LFA of the 3D scalar wave equation is investigated. The effectiveness of the Gauss-Seidel and Kaczmarz smoother on trilinear finite elements are investigated through the method of LFA. Since LFA requires constant coefficients, only the constant absorbing function profile $\gamma(s) = 1$ is considered. For smoothing, the problematic case arises when PML is applied in one direction only, and thus only the case of constant PML in the x direction is considered. To focus on the effect of β on the smoothing factor, ω is set to zero in $\mathbf{K} - \omega\mathbf{M}$. The smoothing factors $\mu_{\mathrm{x,loc}}, \mu_{\mathrm{y,loc}}$ defined in Equation (2.129) and (2.130) with respect to varying β are shown in Figure 2.39. The smoothing factor μ for Gauss-Seidel exceeds 1 at $\beta = 0.7$. This predicts failure of the multigrid for such PML parameters. Contrary to this, the smoothing factor for Kaczmarz has $\mu \leq 1$ for all β, confirming the unconditional convergence of the method. One does see though that the smoothing factor is unacceptable for large β.

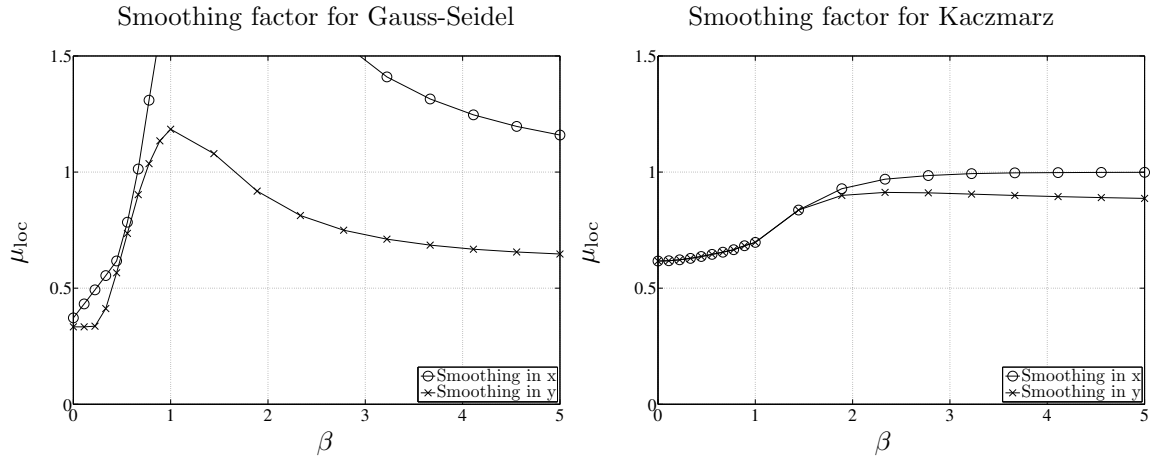


Figure 2.39: Smoothing factor for 3D scalar wave equation with PML in the x direction for Gauss-Seidel and Kaczmarz smoothers

3D elastodynamic

Local Fourier Analysis

In this section the LFA of the 3D elastodynamic equation is investigated. The effectiveness of the Gauss-Seidel and Kaczmarz smoother on trilinear finite elements are investigated through the method of LFA. Since LFA requires constant coefficients, only the constant absorbing function profile $\gamma(s) = 1$ is considered. For smoothing, the problematic case arises when PML is applied in one direction only, and thus only the case of constant PML in the x direction is considered. To focus on the effect of β on the smoothing factor, ω is set to zero in $\mathbf{K} - \omega\mathbf{M}$. The smoothing factors $\mu_{\mathrm{x,loc}}, \mu_{\mathrm{y,loc}}$ defined in Equation (2.134) and (2.135) for x and y with respect to varying β of the two smoothers are shown in Figure 2.40. The smoothing factor μ for Gauss-Seidel exceeds 1 at $\beta = 0.7$. This predicts failure of the multigrid for such PML parameters. Contrary to this, the smoothing factor for Kaczmarz has $\mu \leq 1$ for all β, confirming the unconditional convergence of the method. One does see though that the smoothing factor is close to unity even for small β, i.e., the smoothing will be quite slow.

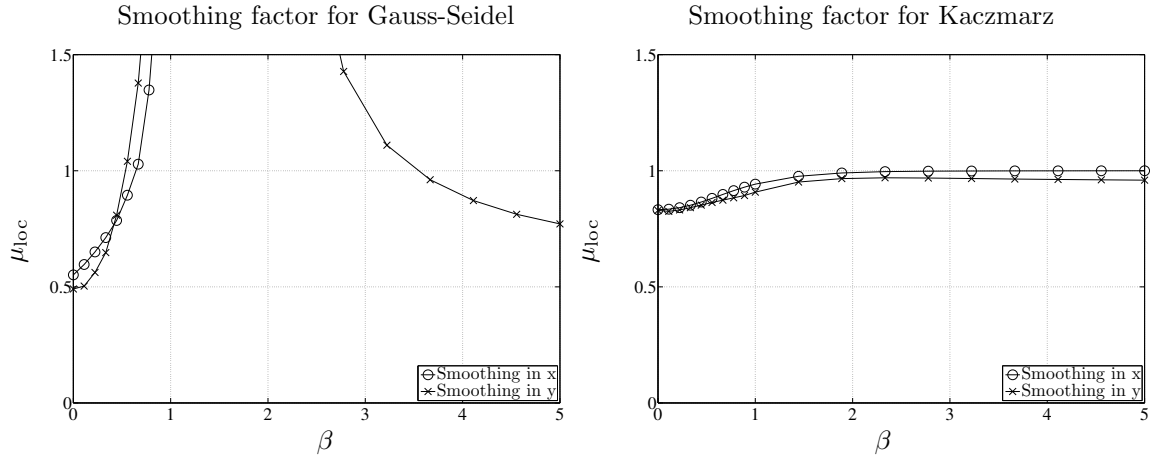


Figure 2.40: Smoothing factor for 3D elasticity wave equation with PML in the x direction for Gauss-Seidel and Kaczmarz smoothers

2.4 Quality factors via eigenvalue computation

In this section the quality factor Q is evaluated by computing the complex-valued frequencies ω of the system. Q is obtained from the expression,

$$Q = \frac{|\omega|}{2\mathrm{Im}(\omega)}. \tag{2.206}$$

The complex-valued frequencies in turn are computed as eigenvalues of the generalized eigenvalue problem,

$$\mathbf{Kx} = \omega^2\mathbf{Mx}, \tag{2.207}$$

obtained from the numerical discretization of the mechanical system. Here, $\mathbf{K}$ is the stiffness matrix, $\mathbf{M}$ is the mass matrix, and $\mathbf{x}$ is the corresponding eigenvector. Discretization by numerical methods such as the finite element produce sparse matrices $\mathbf{K}, \mathbf{M}$ for which sparse eigenvalue methods are employed. In the application of computing Q, one is interested in only a few eigenvalues close to a particular value, for which projection methods prove to be very effective. For clarity the eigenvalue problem is restated as,

$$\mathbf{Ax} = \lambda\mathbf{Bx}\ . \tag{2.208}$$

$\mathbf{A}$ and $\mathbf{B}$ are assumed nonsingular.

In projection methods [18], an approximation subspace $\mathcal{V} = \mathrm{span}\{\mathbf{V}\}$ spanned by the columns of the matrix $\mathbf{V}$ is constructed from which an eigenvector approximate with corresponding eigenvalue close to a desired value is extracted. Simultaneously, a test space $\mathcal{W} = \mathrm{span}\{\mathbf{W}\}$ spanned by the columns of the matrix $\mathbf{W}$ is constructed. The eigenvector approximate $\mathbf{x} \in \mathcal{V}$ and eigenvalue approximate θ are enforced to obey the condition,

$$(\mathbf{A} - \theta\mathbf{B})\,\mathbf{x} \perp \mathcal{W}. \tag{2.209}$$

Since vectors in $\mathcal{V}$ can be represented as $\mathbf{Vy}$, where $\mathbf{y}$ is a vector of the same size as the number of

columns in $\mathbf{V}$, one can alternatively state the problem as computing the pair $(\mathbf{y}, \theta)$ that satisfy,

$$\begin{aligned} \mathbf{W}^* \left(\mathbf{A} - \theta \mathbf{B}\right) \mathbf{V}\mathbf{y} &= 0 \\ \mathbf{W}^* \mathbf{A} \mathbf{V} \mathbf{y} &= \theta \mathbf{W}^* \mathbf{B} \mathbf{V} \mathbf{y} \ . \end{aligned} \tag{2.210}$$

When $\mathcal{V}$ and $\mathcal{W}$ are equal this is called a Galerkin projection, and when they are different, a Petrov-Galerkin projection.

The most widely used and developed projection subspace for eigenvalue extraction is the Krylov subspace. Given an operator $\mathbf{A}$ and initial vector $\mathbf{v}$, the k-th Krylov subspace is defined as,

$$\mathcal{K}^k(\mathbf{A}, \mathbf{v}) = \text{span}\{\mathbf{v}, \mathbf{A}\mathbf{v}, \mathbf{A}^2\mathbf{v}, \ldots, \mathbf{A}^{k-1}\mathbf{v}\} \ . \tag{2.211}$$

This Krylov subspace has the character of approximating eigenvectors that correspond to eigenvalues on the exterior of the spectrum of the operator $\mathbf{A}$. For the generalized eigenvalue problem, several different Krylov subspaces can be constructed, each approximating different parts of the spectrum better.

1. $\mathcal{K}^k(\mathbf{B}^{-1}\mathbf{A}, \mathbf{v})$: For exterior eigenvalues of the pencil $(\mathbf{A}, \mathbf{B})$,

2. $\mathcal{K}^k(\mathbf{A}^{-1}\mathbf{B}, \mathbf{v})$: For eigenvalues of the pencil $(\mathbf{A}, \mathbf{B})$ close to 0,

3. $\mathcal{K}^k((\mathbf{A} - \sigma\mathbf{B})^{-1}\mathbf{B}, \mathbf{v})$: For interior eigenvalues of the pencil $(\mathbf{A}, \mathbf{B})$ close to σ.

The 3rd constructed subspace referred to as "shift-and-invert", in which a spectral transformation has been applied to the pencil $(\mathbf{A}, \mathbf{B})$,

$$(\mathbf{A} - \sigma\mathbf{B})^{-1}\mathbf{B}\mathbf{x} = \frac{1}{\lambda - \sigma}\mathbf{x} \ , \tag{2.212}$$

is most effective for computing interior eigenvalues of the spectrum. The only drawback of such a transformation is that the inverse of the operator $\mathbf{A} - \sigma\mathbf{B}$ must be computed and a linear system of equations,

$$(\mathbf{A} - \sigma\mathbf{B})\mathbf{v}_{k+1} = \mathbf{v}_k \tag{2.213}$$

must be solved for the new Krylov vector $\mathbf{v}_{k+1}$. In the case that the matrices are not too large (order of up to a million), a direct solve through an LU factorization is feasible in terms of time and memory, and thus accurate applications of the inverse operator are possible. When the matrices are very large (more than millions), LU factorizations are no longer tractable. An alternative to a direct solve is to iteratively solve the linear system. This takes considerably less memory than direct solves, but the drawback is the accuracy attainable. The termination of an iterative method for solving $\mathbf{Ax} = \mathbf{b}$ is often measured by the size of the relative residual define as $||\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}||/||\mathbf{b}||$, where $\hat{\mathbf{x}}$ is the current approximate solution. The number of iterations required for high accuracy is usually very large making it unfeasible. Thus one must be satisfied with a relative residual of approximately $10^{-6} - 10^{-10}$. This can cause problems for Krylov methods, since the space constructed through such iterations is no longer an exact Krylov subspace. In such case, one must work with a modified definition of the Krylov subspace to include the inexactness [18].

Ideally one would like to obtain a good approximation for interior eigenvalues that monotonically converges to the exact eigenvalues as the Krylov subspace is expanded, without conducting a "shift-and-invert" since this is costly. Extraction of interior eigenvalues from the 1st subspace is possible, but one does not obtain monotonic convergence for these values. A better approximation for interior eigenvalues can be obtained from the Harmonic-Ritz values [149] which arise from a Petrov-Galerkin projection. The test space $\mathcal{W}$ is defined as $\mathbf{A}\mathcal{V}$, and the eigenvector approximate $\mathbf{x} \in \mathcal{V}$ and eigenvalue $\hat{\theta}$ are sought under the Petrov-Galerkin condition,

$$\left(\mathbf{Ax} - \hat{\theta}\mathbf{x}\right) \perp \mathcal{W} \tag{2.214}$$

which can be restated as computing the pair $(\mathbf{y}, \theta)$ that satisfy,

$$\begin{aligned} \mathbf{W}^* \left(\mathbf{A} - \hat{\theta}\mathbf{I}\right) \mathbf{Vy} &= 0 \\ \mathbf{W}^*\mathbf{AVy} &= \hat{\theta}\mathbf{W}^*\mathbf{Vy} \\ \mathbf{W}^*\mathbf{Wy} &= \hat{\theta}\mathbf{W}^*\mathbf{A}^{-1}\mathbf{Wy} \ . \end{aligned} \tag{2.215}$$

The eigenvector approximate $\mathbf{x} = \mathbf{Vy}$ is called the Harmonic-Ritz vector and $\hat{\theta}$ the Harmonic-Ritz

value. By constructing $\mathbf{W}$ to be orthonormal, the relationship,

$$\mathbf{W}^*\mathbf{A}^{-1}\mathbf{W}\mathbf{y} = \frac{1}{\hat{\theta}}\mathbf{y}, \tag{2.216}$$

is obtained. Thus one observes that the Petrov-Galerkin projection can be interpreted as a Galerkin projection of $\mathbf{A}^{-1}$ under the subspace $\mathbf{W}$. Thus the Harmonic-Ritz values are interpreted as the reciprocal of the eigenvalue approximates of $\mathbf{A}^{-1}$. Since the Galerkin projection approximates eigenvalues on the exterior of the spectrum well, eigenvectors with small $\hat{\theta}$ are represented more accurately. This method can also be applied to the generalized eigenvalue problem with a shift. Construct the test space as $\mathbf{W} = (\mathbf{A} - \sigma\mathbf{B})\mathbf{V}$. Then project such that,

$$\begin{aligned} \mathbf{W}^* \left(\mathbf{A} - \sigma\mathbf{B}\right) \mathbf{V}\mathbf{y} &= (\lambda - \sigma)\mathbf{W}^*\mathbf{V}\mathbf{y} \\ \mathbf{W}^*\mathbf{W}\mathbf{y} &= (\lambda - \sigma)\mathbf{W}^* \left(\mathbf{A} - \sigma\mathbf{B}\right)^{-1} \mathbf{W}\mathbf{y} \\ \frac{1}{(\lambda - \sigma)}\mathbf{W}^*\mathbf{W}\mathbf{y} &= \mathbf{W}^* \left(\mathbf{A} - \sigma\mathbf{B}\right)^{-1} \mathbf{W}\mathbf{y} \ . \end{aligned} \tag{2.217}$$

By computing the eigenvalues of this reduced generalized eigenvalue problem, one can obtain better estimates for interior eigenvalues and corresponding eigenvectors. For the symmetric pencil, one can observe monotonic convergence of the interior eigenvalues. It must be noted that since the projection space is not as good as the shift-and-invert case, fast convergence equivalent to the shift-and-invert case cannot be expected.

Another type of projection subspace other than the Krylov subspace, which does not have the requirement of highly accurate solves, can be constructed based on the Jacobi-Davidson method [167]. This method can be considered a pseudo-Newton type of method for obtaining eigenvalues and eigenvectors. A correction is computed at each step to update the current eigenvalue and eigenvector estimate. This method can be combined with the Harmonic-Ritz value eigenvalue extraction to compute the eigenvalues of the generalized eigenvalue problem in the Jacobi-Davidson QZ method [74].

In order to compute the quality factor Q of the system from the generalized eigenvalue problem, one must first understand what PML parameters are appropriate to obtain a physically accurate approximation. In Section 2.2, the relation between desired reflection and PML parameter selection

was identified. Here the relation between the desired accuracy in Q and PML parameters is presented for a 1D problem, which lead to heuristics for parameter selection in the general case. With such heuristics at hand, an overview of the Jacobi-Davidson QZ algorithm is given, with details regarding our implementation.

2.4.1 Effect of perfectly matched layers on eigenvalue accuracy

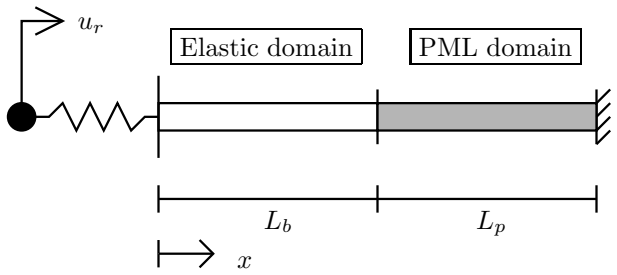


Figure 2.41: Mass-spring system attached to elastic and PML domain

The following 1D model is used to study the effect of the PML parameters on the accuracy of the quality factor Q computed for the system. A schematic of the model is shown in Figure 2.41. The model consists of a mass-spring system attached to a continuum 1D elastic bounded domain terminated with a PML. This system can be interpreted as a resonator (mass-spring system) situated on top of the infinite substrate (1D elastic domain with PML termination). The governing equations of the system are,

$$\rho \frac{\partial^2 u}{\partial t^2} - E \frac{\partial^2 u}{\partial x^2} = 0, \tag{2.218}$$

$$m_r \frac{\partial^2 u_r}{\partial t^2} + k_r \{u_r - u(t,0)\} = 0, \tag{2.219}$$

with the interface condition for force balance,

$$k_r \{u_r - u(t,0)\} = -E \frac{\partial u}{\partial x}(t,0), \tag{2.220}$$

where ρ is the density per length with dimensions [M/L], E is the stiffness with dimensions [ML/T^2],

m_r is the resonator mass, and k_r is the resonator stiffness. Non-dimensionalization of these equations with the dimensional parameters,

$$[T] := \frac{2\pi}{\sqrt{k_r/m_r}}, \tag{2.221}$$

$$[L] := 2\pi\frac{\sqrt{E/\rho}}{\sqrt{k_r/m_r}}, \tag{2.222}$$

$$[M] := \rho\,[L]\,, \tag{2.223}$$

lead to,

$$\frac{\partial^2\tilde{u}}{\partial\tilde{t}^2} - \frac{\partial^2\tilde{u}}{\partial\tilde{x}^2} = 0, \tag{2.224}$$

$$\frac{\partial^2\tilde{u}_r}{\partial\tilde{t}^2} + (2\pi)^2\left\{\tilde{u}_r - \tilde{u}(t,0)\right\} = 0, \tag{2.225}$$

with the interface condition,

$$2\pi\alpha\left\{\tilde{u}_r - \tilde{u}(\tilde{t},0)\right\} = -\frac{\partial\tilde{u}}{\partial\tilde{x}}(t,0), \tag{2.226}$$

where,

$$\alpha := \frac{\sqrt{m_r k_r}}{\sqrt{\rho E}}, \tag{2.227}$$

is the impedance which defines the degree of coupling between the mass-spring system and the substrate. The tilde denotes the non-dimensionalized quantities. With an abuse of notation, the tildes defining the non-dimensionalization are dropped. Under time-harmonic assumptions,

$$u_r(t,x) = \hat{u}_r\exp(i\Omega t), \tag{2.228}$$

$$u(t,x) = \hat{u}\exp(i\Omega t), \tag{2.229}$$

where,

$$\Omega := 2\pi\frac{\omega}{\sqrt{k_r/m_r}} \tag{2.230}$$

is the non-dimensionalized frequency, the equations become,

$$\Omega^2 \hat{u} + \frac{d\hat{u}}{dx^2} = 0, \tag{2.231}$$

$$-\Omega^2 \hat{u}_r + (2\pi)^2 \{\hat{u}_r - \hat{u}(t,0)\} = 0, \tag{2.232}$$

$$\tag{2.233}$$

with the interface condition,

$$2\pi\alpha \{\hat{u}_r - \hat{u}(0)\} = -\frac{\partial \hat{u}}{\partial x}(0). \tag{2.234}$$

$$\tag{2.235}$$

Solving Equation (2.231), and restricting the solution to only those that are outgoing yields,

$$\hat{u} = c_{\text{out}} \exp(-i\Omega x). \tag{2.236}$$

$$\tag{2.237}$$

Expressions for the eigenvalue and Q

By combining Equations (2.232), (2.234), and (2.236), one obtains a 1D eigenvalue problem for Ω,

$$\Omega \left[\left(\frac{\Omega}{2\pi} \right)^2 - (2\pi\alpha i) \left(\frac{\Omega}{2\pi} \right) - 1 \right] c_{\text{out}} = 0, \tag{2.238}$$

which yields,

$$\frac{\Omega}{2\pi} = 0, \quad \pi\alpha i \pm \sqrt{1 - (\pi\alpha)^2}. \tag{2.239}$$

$\pi\alpha i + \sqrt{1 - (\pi\alpha)^2}$ denotes the eigenvalue corresponding to the right traveling wave. When the resonator is vibrating, the mode of vibration which leads to energy loss from the resonator is this right traveling wave mode, and thus the quality factor Q corresponding to this eigenvalue is of interest,

$$Q = \frac{1}{2\pi\alpha}. \tag{2.240}$$

A closed form analytical expression for the complex-eigenvalue and corresponding Q has been obtained for the system. With this expression, the behavior of the eigenvalue and Q convergence with respect to PML parameters can be studied by comparing the computed numerical values with the exact analytical values.

Error bound for the quality factor Q

Before moving onto the numerical computation of the quality factor Q for different PML parameters, an error bound for Q, is derived in terms of error in the eigenvalue. Assume,

$$\begin{aligned} \omega_e &= \omega_r + i\omega_i, && (2.241)\\ Q_e &= \frac{|\omega_e|}{2\mathrm{Im}(\omega_e)}, && (2.242)\\ \omega &= \omega_r(1+\epsilon_r) + i\omega_i(1+\epsilon_i), && (2.243)\\ Q &= \frac{|\omega|}{2\mathrm{Im}(\omega)}, && (2.244) \end{aligned}$$

where ω_e is the exact eigenvalue and ω is the computed eigenvalue with relative error of ϵ_r in the real and ϵ_i in the imaginary part. By defining $\nu := \frac{\omega_i}{\omega_r} = \frac{1}{\sqrt{4Q_e^2-1}}$, one has,

$$\begin{aligned} \frac{Q}{Q_e} &= \frac{1+\epsilon_r}{1+\epsilon_i}\sqrt{1+\nu^2\left(\frac{1+\epsilon_i}{1+\epsilon_r}\right)^2} && (2.245)\\ &\leq \frac{1+\epsilon_r}{1+\epsilon_i}\left[1+\frac{1}{2}\nu^2\left(\frac{1+\epsilon_i}{1+\epsilon_r}\right)^2\right]. && (2.246) \end{aligned}$$

With the additional assumption that $|\epsilon_i| > |\epsilon_r|$, which is valid when one has $\nu < 1$, one obtains

$$\begin{aligned} \left|\frac{Q-Q_e}{Q_e}\right| &\leq 2|\epsilon_i| + \frac{1}{2}\nu^2(1+2|\epsilon_i|) && (2.247)\\ &\leq 2|\epsilon_i| + \frac{1}{2}\frac{1}{4Q_e^2-1}(1+2|\epsilon_i|). && (2.248) \end{aligned}$$

When ν is small, i.e., a large Q_e, the accuracy of Q is determined by the accuracy of the imaginary part. On the other hand when high accuracy is obtained in the imaginary part, the quality factor error is governed by Q_e. This implies that the numerically accuracy to which the quality factor can be computed is bounded by its own magnitude. These observations are confirmed in the following numerical simulations.

Numerical simulations

The relative error of Q, as well as the energy dissipation error E_{relerr}, relative error in the wave number k_{relerr}, and relative error in the eigenvalue is computed for 4 different cases. The absorbing function profile $\gamma(s) = s$, $\beta = 1$, and $\alpha = 1 \times 10^{-3}$, are the same for all cases. The first two cases correspond to the case where one fixes the size of the domain of computation and successively increase the discretization of the mesh to decrease the interface reflection and obtain smaller total reflection. The latter two cases correspond to the case where one fixes the size of the bounded domain and discretization, and increasingly extends the size of the PML domain to decrease both end termination reflection and interface reflection, for smaller total reflection.

1. Figure 2.42: Linear elements. Vary the number of nodes per wave $n_{\text{npw}} \in [6, 96]$ for two cases of number of waves in the PML $n_{\text{wpml}} = \{0., 6.5\}$.

2. Figure 2.43 Cubic elements. Vary the number of nodes per wave $n_{\text{npw}} \in [6, 96]$ for two cases of number of waves in the PML $n_{\text{wpml}} = \{0., 6.5\}$.

3. Figure 2.44 Linear elements. $n_{\text{npw}} = 96$. Vary the number of waves in the PML $n_{\text{wpml}} \in [0, 4]$.

4. Figure 2.45 Cubic elements. $n_{\text{npw}} = 96$. Vary the number of waves in the PML $n_{\text{wpml}} \in [0, 4]$.

The following observations can be made for all cases.

- The relative error in Q, the energy dissipation error E_{relerr}, and the relative error in the imaginary part of the eigenvalue $\omega_{i,\text{relerr}}$ is bounded below by the discretization error k_{relerr}.

- The error in Q is always bounded from below by $\frac{1}{2}\nu^2$.

- The error in Q is always bounded from below by the energy dissipation error E_{relerr}.

- The relative error in Q coincides with the energy dissipation error E_{relerr} and the relative error in the imaginary part of the eigenvalue $\omega_{i,\text{relerr}}$, until it is bounded below by $\frac{1}{2}\nu^2$. This is understandable since $Q \approx E_{\text{dissipated}} \approx \omega_i$.

- The rate at which the relative error of the eigenvalue ω and its real and imaginary parts decrease is the same.

For each case the following comments can be made.

1. Figure 2.42: For the case $n_{\text{wpml}} = 0.5$, the relative error in Q is bounded by the energy dissipation error E_{relerr}. The energy dissipation error cannot decrease with finer discretization due to the end termination reflection for $\beta = 1$ which is approximately 10^{-1}.

   For the case $n_{\text{wpml}} = 6.5$, the end termination reflection for $\beta = 1$ is fairly small, such that the energy dissipation error is bounded by the discretization error. Thus the relative error in Q is bounded by the discretization error.

2. Figure 2.43 For the case $n_{\text{wpml}} = 0.5$, the situation is analogous to the linear element case.

   For the case $n_{\text{wpml}} = 6.5$, the end termination reflection is sufficiently small. The discretization error can be made sufficiently small, such that the relative error in Q is bounded by $\frac{1}{2}\nu^2$.

3. Figure 2.44 As the PML domain is extended, the relative error in Q decreases to the point that it is bounded by the discretization error.

4. Figure 2.45 In this case the discretization error is smaller than $\frac{1}{2}\nu^2$, such that the relative error in Q decreases to the point that it is bounded by this value.

These observations and comments lead to the following heuristics for PML parameter selection and Q evaluation.

- Given a desired relative accuracy in Q, select the PML parameters which give the same relative error in the energy or reflection using the heuristics developed in Section 2.2. Compute the eigenvalue for finer discretization until convergence in the desired relative accuracy in Q is obtained.

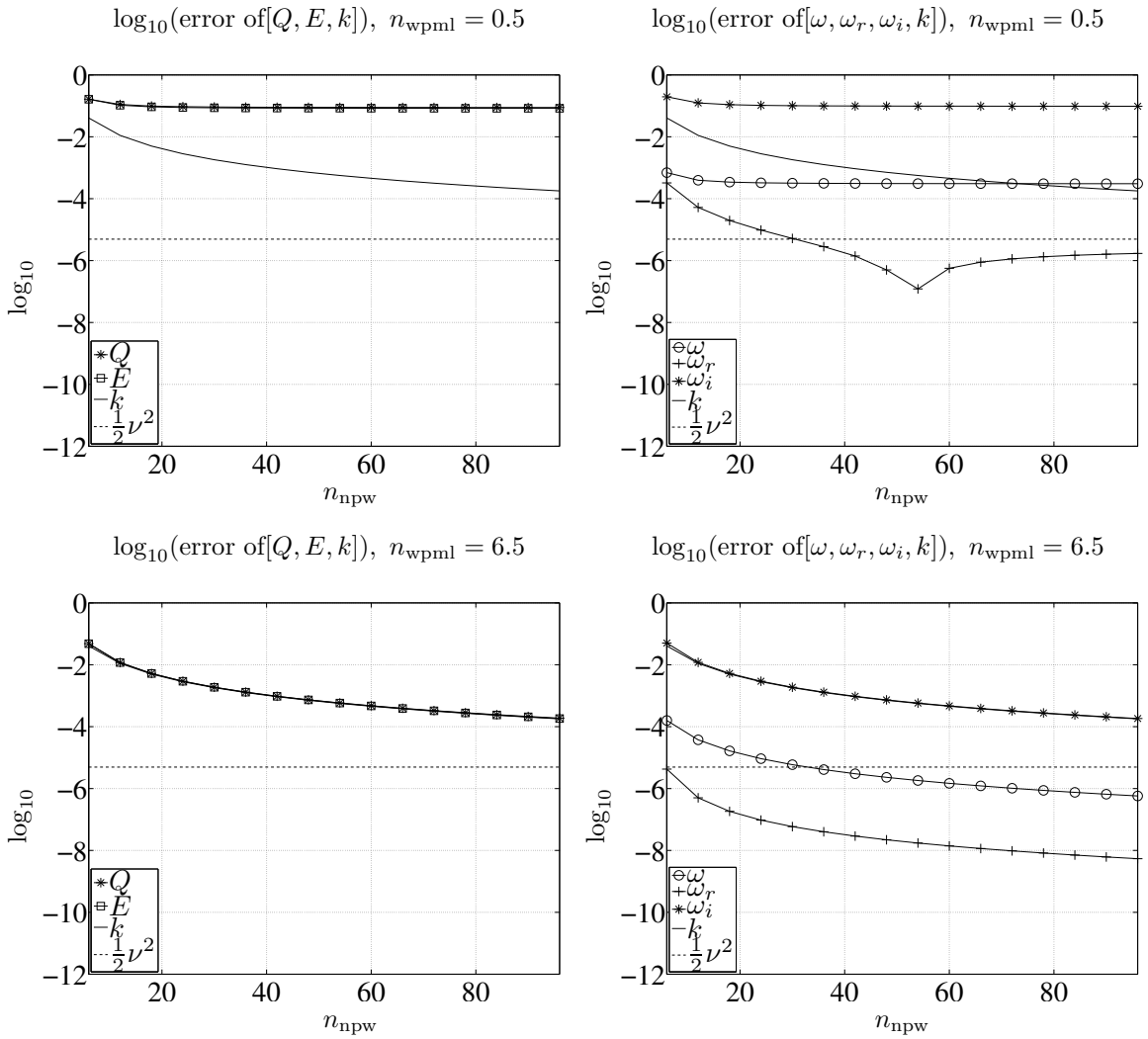


Figure 2.42: Relative error of $Q, E, k, \omega, \omega_r, \omega_i$ with respect to varying number of nodes per wave n_{npw}, keeping parameters [linear elements, $\gamma(s) = s, \beta = 1, \alpha = 1 \times 10^{-3}$] constant.

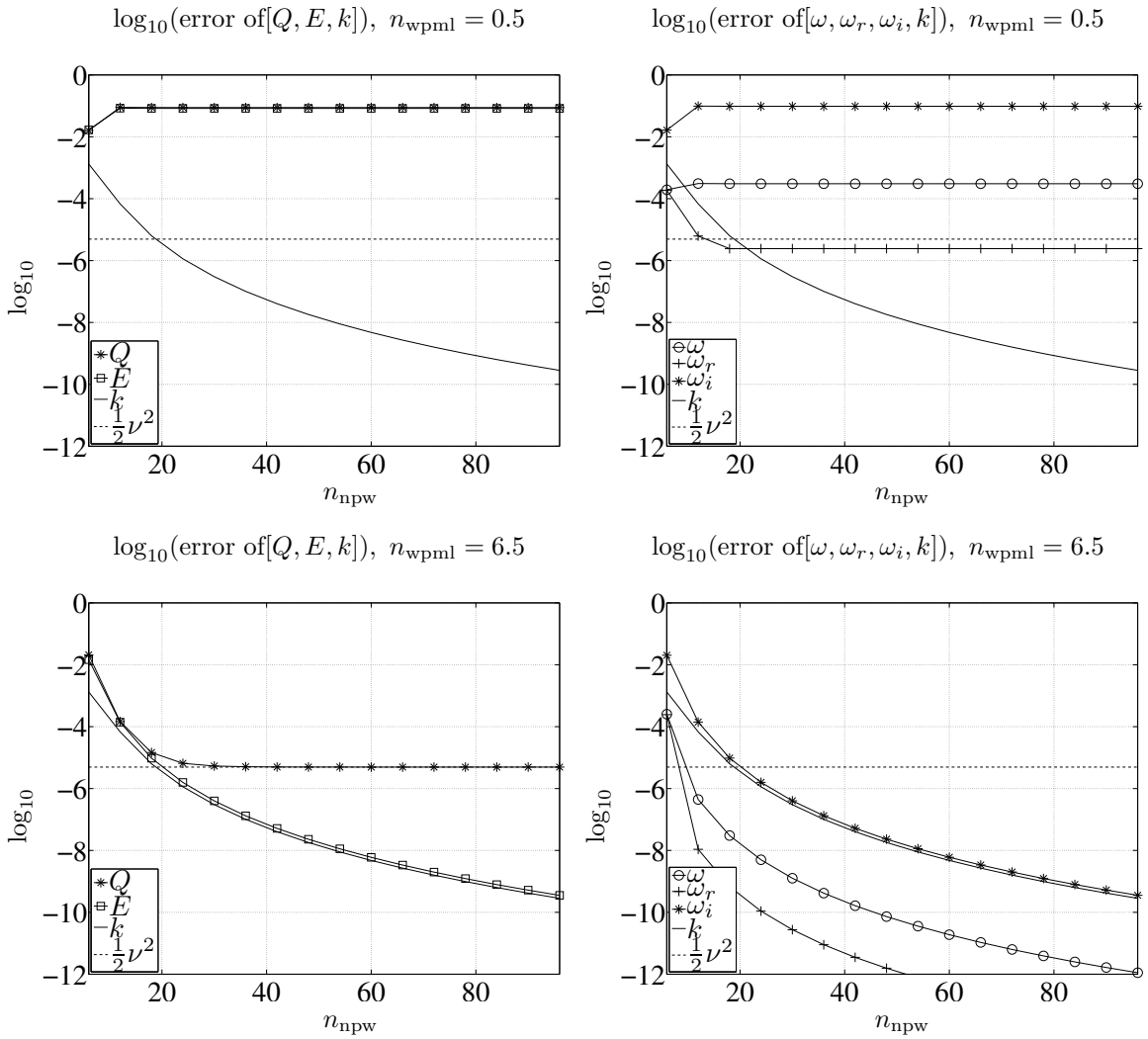


Figure 2.43: Relative error of $Q, E, k, \omega, \omega_r, \omega_i$ with respect to varying number of nodes per wave n_{npw}, keeping parameters [cubic elements, $\gamma(s) = s, \beta = 1, \alpha = 1 \times 10^{-3}$] constant.

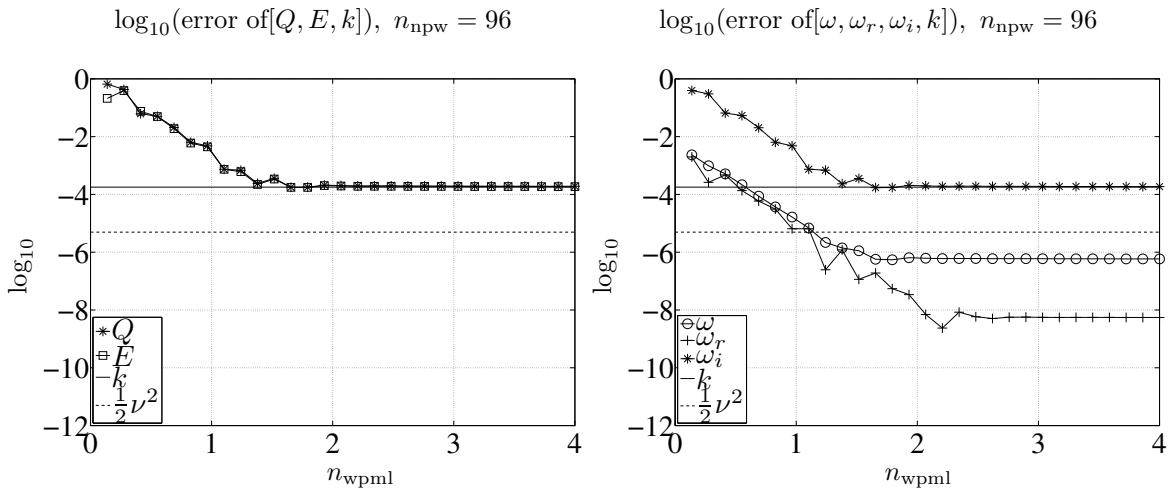


Figure 2.44: Relative error of $Q, E, k, \omega, \omega_r, \omega_i$ with respect to varying length of the pml n_{wpml}, keeping parameters [linear elements, $\gamma(s) = s, \beta = 1, \alpha = 1 \times 10^{-3}$] constant.

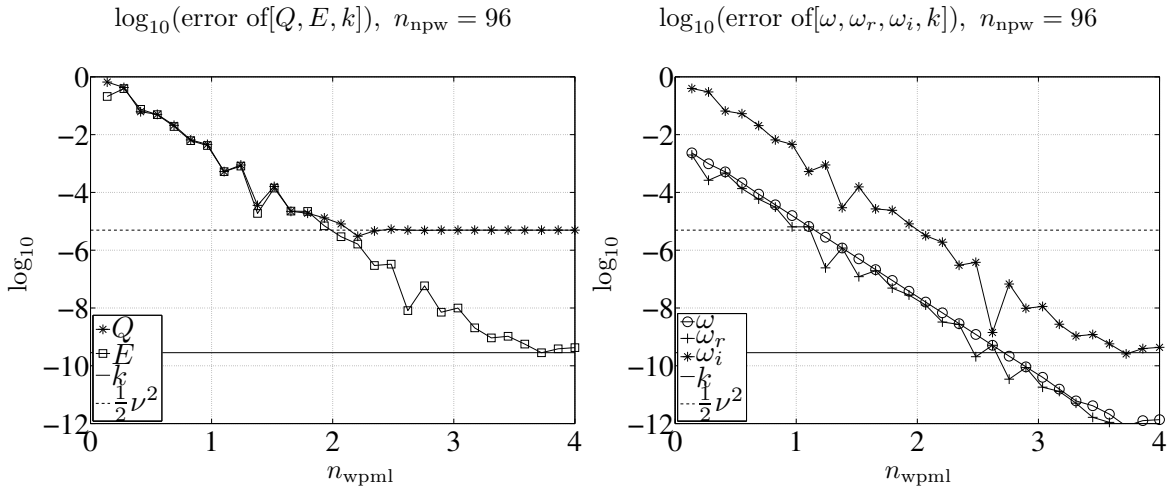


Figure 2.45: Relative error of $Q, E, k, \omega, \omega_r, \omega_i$ with respect to varying length of the pml n_{wpml}, keeping parameters [cubic elements, $\gamma(s) = s, \beta = 1, \alpha = 1 \times 10^{-3}$] constant.

2.4.2 Jacobi-Davidson QZ combined with geometric multigrid

Constructing a Krylov subspace by shift-and-invert with a specific σ rich in eigenvector approximates with corresponding eigenvalues close to σ can be costly. This is due to the requirement for highly accurate linear solves for a large system of equations. An alternative type of projection subspace can be constructed by the Jacobi-Davidson method. This method can be interpreted as a pseudo-Newton type of method for obtaining eigenvalues and eigenvectors. A correction is computed at each step to update the current eigenvalue and eigenvector approximate. The Jacobi-Davidson QZ method, a variant of the Jacobi-Davidson tailored for the generalized eigenvalue proceeds as follows.

For the generalized eigenvalue problem presented in Equation (2.208), assume that one has an approximation subspace $\mathbf{V}$, and test subspace $\mathbf{W}$. Let the triplet $(\theta, \mathbf{q}, \mathbf{p})$ be the Petrov value, right Petrov vector, and the left Petrov vector, extracted from $\mathbf{V}$ and $\mathbf{W}$, with θ close to the desired eigenvalue of σ. The relationship between the triplet is,

$$(\mathbf{A} - \theta\mathbf{B})\,\mathbf{q} \perp \mathbf{W}, \qquad (\mathbf{A} - \theta\mathbf{B})^*\,\mathbf{p} \perp \mathbf{V}, \qquad \mathbf{p}^*\mathbf{A}\mathbf{q} = \theta\mathbf{p}^*\mathbf{B}\mathbf{q}. \tag{2.249}$$

Essentially $(\theta, \mathbf{q})$ is the eigenvalue, eigenvector approximate with residual,

$$\mathbf{r} \;:=\; (\mathbf{A} - \theta\mathbf{B})\,\mathbf{q}. \tag{2.250}$$

With these approximates one would like to find an update $\mathbf{t} \perp \mathbf{q}$, such that $\mathbf{q} + \mathbf{t}$ is the eigenvector corresponding to σ,

$$(\mathbf{A} - \sigma\mathbf{B})\,(\mathbf{q} + \mathbf{t}) = 0. \tag{2.251}$$

This equation can be written equivalently in terms of projections orthogonal to $\mathbf{p}$,

$$\Leftrightarrow \quad \begin{cases} \mathbf{p}\mathbf{p}^*\,(\mathbf{A} - \sigma\mathbf{B})\,(\mathbf{q} + \mathbf{t}) &= 0 \\ (\mathbf{I} - \mathbf{p}\mathbf{p}^*)\,(\mathbf{A} - \sigma\mathbf{B})\,(\mathbf{q} + \mathbf{t}) &= 0 \end{cases}, \tag{2.252}$$

$$\Leftrightarrow \quad \begin{cases} \sigma = \dfrac{\mathbf{p}^*\mathbf{A}(\mathbf{q} + \mathbf{t})}{\mathbf{p}^*\mathbf{B}(\mathbf{q} + \mathbf{t})} \\ (\mathbf{I} - \mathbf{p}\mathbf{p}^*)\,(\mathbf{A} - \sigma\mathbf{B})\,\mathbf{t} = -\,(\mathbf{I} - \mathbf{p}\mathbf{p}^*)\,(\mathbf{A} - \sigma\mathbf{B})\,\mathbf{q} \end{cases}. \tag{2.253}$$

These equations are nonlinear when the exact eigenvalue σ is unknown, and σ is approximated by the Petrov value θ, yielding the "correction equations",

$$(\mathbf{I} - \mathbf{p}\mathbf{p}^*)(\mathbf{A} - \theta\mathbf{B})(\mathbf{I} - \mathbf{q}\mathbf{q}^*)\,\mathbf{t} = -\mathbf{r}. \tag{2.254}$$

The projection in front of $\mathbf{t}$ is inserted from the condition $\mathbf{t} = (\mathbf{I} - \mathbf{q}\mathbf{q}^*)\mathbf{t}$. If this correction equation is solved exactly, locally superlinear convergence is attainable [165]. The method can progress even in the case of moderate accuracy, which allows the use of iterative methods for large systems.

Once this correction equation is solved for $\mathbf{t}$, the approximation subspace is expanded by appending $\mathbf{t}$, and the test subspace is expanded by appending a linear combination of $\mathbf{At}$ and $\mathbf{Bt}$. To incorporate Harmonic-Ritz value approximation for interior eigenvalues close to σ_0, the test subspace is constructed as $\mathbf{W} = (\mathbf{A} - \sigma_0\mathbf{B})\mathbf{V}$, and the appended vector is $(\mathbf{A} - \sigma_0\mathbf{B})\mathbf{t}$.

To summarize, at each step an approximate Petrov triplet is computed by solving the projected generalized eigenvalue problem for the pencil $(\mathbf{W}^*\mathbf{AV}, \mathbf{W}^*\mathbf{BV})$ with approximation subspace $\mathbf{V}$ and test subspace $\mathbf{W}$. Then the correction equation is solved for the update $\mathbf{t}$. This correction $\mathbf{t}$ is appended to $\mathbf{V}$, and $(\mathbf{A} - \sigma_0\mathbf{B})\mathbf{t}$ is appended to $\mathbf{W}$, for Harmonic-Ritz approximation close to σ_0.

Preconditioning and the correction equation

In one step of the Jacobi-Davidson QZ method, a correction equation must be solved. The application of iterative methods is ideal, since one can select the accuracy of the computed solution, trading time for less accuracy and more efficiency.

For the application of PML, the linear system is complex-symmetric and requires iterative solvers for general linear systems. BICGSTAB and other solvers may also be applicable for complex-symmetric systems, but here we restrict our study to GMRES. Iterative methods must be combined with efficient preconditioners for fast convergence. The application of a preconditioner $\mathbf{P}$ for $\hat{\mathbf{A}} := (\mathbf{I} - \mathbf{p}\mathbf{p}^*)(\mathbf{A} - \theta\mathbf{B})(\mathbf{I} - \mathbf{q}\mathbf{q}^*)$, in solving the correction equations require special treatment, such

that the projected version,

$$\hat{\mathbf{P}} := (\mathbf{I} - \mathbf{p}\mathbf{p}^*)\,\mathbf{P}\,(\mathbf{I} - \mathbf{q}\mathbf{q}^*)\,, \tag{2.255}$$

must be used. In the Krylov iterative method for the correction method, the operation,

$$\hat{\mathbf{P}}^{-1}\hat{\mathbf{A}}\mathbf{v}_k = \mathbf{v}_{k+1} \tag{2.256}$$

$$\Leftrightarrow \quad \mathbf{v}_{k+1} = \left[\mathbf{I} - \mathbf{P}^{-1}\mathbf{p}\left(\mathbf{q}^*\mathbf{P}^{-1}\mathbf{p}\right)^{-1}\mathbf{q}^*\right]\mathbf{P}^{-1}\mathbf{A}\mathbf{v}_k, \tag{2.257}$$

is required to obtain the new vector.

Among the many types of preconditioners available, the geometric multigrid method presented in Section 2.3 is employed. It has been stated in the section on multigrid, that non-stationary smoothers are avoided in connection with the Jacobi-Davidson eigenvalue method. The reason for such a choice lies in the application of the preconditioner as is presented in Equation (2.257). Non-stationary methods have different behavior depending on the application vector, implying that the effect in $\mathbf{P}^{-1}\mathbf{p}$ and $\mathbf{P}^{-1}\mathbf{A}\mathbf{v}_k$ is different. For solution of linear systems, such non-stationary behavior of the preconditioner can be dealt with by methods such as FGMRES. The usage of FGMRES in combination with multigrid non-stationary smoothers for solution of the correction equations in the JDQZ method has been observed to lead to delayed or non-convergent behavior in the outer JDQZ iteration.

Remark: The non-convergent behavior of non-stationary smoothers for multigrid preconditioning of the Jacobi-Davidson correction equations is due to the non-trivial interaction between the inner and outer iterations of Krylov and the Jacobi-Davidson eigenvalue iteration.

Geometric multigrid for initial vector approximations

The computation of eigenvalues is a nonlinear process, with convergence to the eigenvalue depending strongly on the quality of the initial approximation subspace $\mathbf{V}$ and test subspace $\mathbf{W}$. If one lacks information of a good initial direction, the random vector is used to initiate the construction of $\mathbf{V}$ and corresponding $\mathbf{W}$.

For the use of the geometric multigrid preconditioner, a hierarchy of meshes, i.e., approximations, are constructed for projection. One can utilize these grids to compute eigenvalue and eigenvector approximations on coarse grids $(\lambda_c, \mathbf{v}_c)$, and use the eigenvalue λ_c as σ_0, and the eigenvector combined with the prolongation operators to compute the initial vector approximates, $\mathbf{v}_{f,init} = \mathbf{P}\mathbf{v}_c$. This multilevel method can be helpful, since one can identify the desired modes on a coarser mesh which requires less time for computation of modes, and use these modes to guide the finer more expensive and time-consuming computations. The acceleration of this initial vector construction is presented in the numerical examples in Section 5.3.2.

2.5 Conclusion

In this chapter, the components necessary to evaluate the quality factor Q of high-frequency MEMS resonators with anchor loss through an eigenvalue computation are presented.

Anchor loss is simulated by terminating the computational domain boundary with Perfectly Matched Layers (PML), which numerically models the radiation boundary condition. All waves outgoing from the computational domain and propagating into the PML are absorbed with zero impedance mismatch at the interface. Unfortunately, this zero impedance property holds true only for the continuous case. Under numerical discretization, reflection can occur with magnitude depending on the discretization of the propagating wave (n_{npw}, number of points per wave), and selection of the PML absorbing function. The PML absorbing function can be characterized by two components. The value of the function at the end of the PML defined as β, and the length of the PML in units of wave length defined as n_{nwpml}. Unless the correct combination of these parameters are selected, the results obtained from the numerical simulation may not be accurate. This is due to insufficient wave absorption in the PML leading to end termination reflection or excessive discretized wave interface reflection at the PML interface. Additionally, the size of n_{nwpml} can directly affect the size of the linear system of equations that one must solve for, i.e., computational expense. Thus one would like to find an optimal combination of parameters that balances these contributions to yield results of desired accuracy with least computational time. In Section 2.2 of this chapter, we develop a method by which one can attain this combination of parameters for a 1D scalar wave problem. An explicit functional form for contours of constant discretized wave interface reflection is formulated based on observations from numerical simulations. The intersection between this curve and the contour of constant end termination reflection yield the combination $(\beta, n_{\text{wpml}}$ which gives the shortest PML possible, given a wave discretization n_{npw} and desired accuracy. Since this is the shortest PML possible for the desired accuracy, it minimizes the size of the linear system that must be solved, and thus minimizes time. Since the optimal parameters are presented in the non-dimensional form of number

of waves in the PML n_{wpml} and the end value of the absorbing function profile β, it is amenable to any length scale. The heuristics developed for the 1D problem are confirmed in higher-dimensional and vector-valued problems.

The quality factor Q of the mechanical system with PML applied can be computed through an eigenvalue computation. As it will be shown in the numerical simulations of disk resonators in Section 5.3, fine discretization can be required to fully resolve the energy dissipation mechanism arising from anchor loss modeled by PML. Thus in the case of 3D simulations, one may be forced to solve eigenvalue problems of size on the order of millions. This is not a trivial task due to the shear size, and with the addition of PML, which makes the matrices involved complex-valued symmetric. In order to compute the eigenvalues, we have developed a method based on the combination of the Jacobi-Davidson QZ eigensolver with a geometric multigrid preconditioned GMRES to solve the correction equation. Since the geometric multigrid method has been developed for solving real symmetric positive definite systems arising from the discretization of elliptic problems, its application to complex-valued symmetric problems is not clear. In Section 2.3 of this chapter, we have developed a geometric multigrid with smoother and prolongation operator components which are scalable and applicable to the complex-valued symmetric system arising from the application of PML. The Gauss-Seidel and Chebyshev smoothers have been shown to be work well under a restriction on the selectable β value. In order to apply the Chebyshev smoother, a Lemma has been proven to estimate the parameters of the smoother based on the complex-valued symmetric property of the system matrices. The prolongation operators are constructed purely geometrically with superblocks that allow scalability in their construction. The geometric multigrid method that has been developed has a restriction on the selectable β value, but as it is shown in the disk resonators examples in Section 5.3, this does not prohibit the analysis for the class of high-frequency MEMS resonators vibrating at frequencies above hundreds of MHz.

In order to understand how this restriction on β can affect the accuracy of the computed quality factor, a 1D scalar wave example is studied in Section 2.4 under this context. Numerical results

reveal that β only has to be large enough to obtain a discretized wave reflection on the same order of accuracy as is desired in the quality factor Q, i.e., for a desired Q accuracy of n digits, PML parameters for obtaining reflection of 10^{-n} should be chosen.

The numerical examples presented in Chapter 5 will confirm the applicability and scalability of our method on 3D discretization of disk resonators with millions of degrees of freedom. Sensitivity with respect to post misalignment for these structures which can only be observed with 3D simulations are presented using our method.

Chapter 3

Thermoelastic damping

3.1 Introduction

Thermoelastic damping is an intrinsic energy loss mechanism which arises from the coupling of the mechanical domain with the thermal domain. The phenomenon, if treated at the atomic scale, involves the treatment of thermal phonons, which are quantized lattice vibrations of the crystal, and their equilibration process [129, 148]. When the mean free path of the phonons l_T is much smaller than the characteristic length scale L of the system, i.e., the diffusive regime ($l_T \ll L$), the thermal phonons scatter often and can be treated as a gas. This allows us to define a local temperature T and to treat the phenomenon at the continuum level, where the complicated interactions between the mechanical and thermal domains are captured by a single parameter, the material's linear coefficient of thermal expansion α_T. Local stress fluctuations induce local strain variations, which create local temperature gradients. These are compensated by irreversible heat flow, leading to energy dissipation. This mechanism is called *thermoelastic damping*.

The first widely accepted theoretical work on *thermoelastic damping* was conducted by Zener in 1937 [200]. In a series of papers, he analyzes and presents a closed form algebraic expression of Q for an isotropic homogeneous beam vibrating in a flexural mode using the coupled thermomechanical

equations and Euler-Bernoulli beam theory. In 2000, Lifshitz and Roukes re-examined Zener's theory to obtain an exact solution to Zener's problem, different from Zener who assumed a series expansion of the solution. [130]. Zener's formula has been verified by multiple experiments (see for example [159]) to yield good results. Because of this success, extensions of Zener's theory have been attempted in order to account for polycrystalline material [168] and beam-like geometries [2], though in each case the theory has not been fully verified. What is consistent in both theoretical and experimental work is that only beam-like geometries have been examined. This restriction clearly arises from the inability to obtain closed form solutions for complex geometrical and material devices.

To increase the class of devices that can be analyzed, a numerical method based on finite element analysis (FEA) which directly treats the governing coupled partial differential equations can be employed to evaluate *thermoelastic damping* and its effect on Q [83, 121, 199, 66]. The FEA approach is often computationally expensive due to the fine discretization that is required for accurate values. To circumvent this expense, we propose a reduced order model technique which reduces the size of the system, but maintains accuracy in evaluating the transfer function of the system. The method is based on Krylov subspace projection methods [63, 19, 20, 17], with structure preservation [124] for additional accuracy. By exploiting the structure of the equations, for symmetric mechanical forcing and sensing we obtain a doubling of matched moments in the transfer function with the same number of second-order Krylov subspace iterations [20]. This leads to a more efficient method for evaluating Q from the transfer function.

We begin with a review of the coupled infinitesimal thermoelastic equations, and their non-dimensionalized form. This will reveal the underlying structure of the equations, which is directly inherited by the finite element discretized system of equations. This structure is exploited in the structure preserving reduced order model. A theorem related with the accuracy of the model reduction method follows.

3.2 Linear thermoelasticity

Thermoelastic damping at the continuum level can be modeled by the coupled thermoelastic equations, which consist of the balance of linear and angular momentum for the mechanical domain, and the two thermodynamics principles. Body forces will be neglected, since they scale as length cubed and are negligible at the MEMS scale. Heat sources will also be neglected, since they do not arise in resonators. For reasonable generality, we assume that the material model has cubic symmetry, which results in a scalar linear thermal expansion coefficient and thermal conductivity. The cubic crystal assumption incorporates modeling both single crystal silicon a cubic crystal and polysilicon an isotropic material, which are two typical MEMS materials. The linear elastic behavior of the two materials and small deformations that occur in the resonators justify the use of a linear constitutive model and infinitesimal kinematics. Due to the weak coupling between the mechanical and thermal domain, the temperature fluctuations resulting from the deformation will also be small, which justifies linearizing the temperature around the ambient operating temperature. Additionally, since we assume small temperature fluctuations, the heat capacity at constant volume will be assumed constant.

3.2.1 Review of the balance equations

The equations governing linear thermoelasticity are well known, and are derived from the balance of linear and angular momentum,

$$\text{div}\ \boldsymbol{\sigma} = \rho\ \ddot{\mathbf{u}} \tag{3.1}$$

$$\boldsymbol{\sigma}^T = \boldsymbol{\sigma} \tag{3.2}$$

and the energy equality and entropy inequality,

$$\rho\ \dot{e} = -\text{div}\ \mathbf{h} + \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} \tag{3.3}$$

$$\rho\ \dot{\eta} \geq -\frac{\text{div}\ \mathbf{h}}{T} + \frac{1}{T^2}\text{grad}T \cdot \mathbf{h}\ \ . \tag{3.4}$$

Here, $\mathbf{u}$ is the displacement field, $\boldsymbol{\sigma}$ is the stress tensor, $\boldsymbol{\varepsilon}$ is the infinitesimal strain tensor, ρ is the density, e is the internal energy per unit mass, η is the entropy per unit mass, $\mathbf{h}$ is the heat flux, and T is the absolute temperature. From the Clausius-Duhem inequality and Legendre transform of the internal energy e to Helmholtz's free energy ψ, the two principles of thermodynamics are combined into the form,

$$\rho c_v \dot{T} \quad = \quad -\text{div } \mathbf{h} + \rho T \frac{\partial^2 \psi}{\partial T \partial \boldsymbol{\varepsilon}} : \dot{\boldsymbol{\varepsilon}} \quad , \tag{3.5}$$

where the heat capacity at constant volume has been defined as $c_v = -\frac{\partial^2 \psi}{\partial T^2} T$. From the last term in Equation (3.5), it is apparent that the equation is nonlinear. Because of our assumption of small temperature fluctuations, we decompose the temperature as:

$$T \quad = \quad T_0 + \theta \tag{3.6}$$

where T_0 is a reference temperature and θ is the temperature fluctuation. The energy balance (3.5) is then simultaneously linearized with respect to θ and the displacement field. Throughout, we assume linearized kinematics,

$$\boldsymbol{\varepsilon}(\mathbf{u}) \quad = \quad \left(\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \right)^s \quad , \tag{3.7}$$

and Fourier's law for the heat flux,

$$\mathbf{h} \quad = \quad -\kappa_T \nabla \theta \ , \tag{3.8}$$

where κ_T is the thermal conductivity. This results in two coupled governing equations for cubic linear thermoelasticity:

$$\begin{aligned} \rho \ddot{\mathbf{u}} \quad &= \quad \boldsymbol{\nabla} \cdot [\mathbb{C} : \boldsymbol{\varepsilon}] \quad -3\kappa\alpha_T \boldsymbol{\nabla}\theta \\ \rho c_v \dot{\theta} \quad &= \quad \kappa_T \boldsymbol{\nabla}^2 \theta \qquad -3\kappa\alpha_T T_0 \text{tr}\,(\dot{\boldsymbol{\varepsilon}}) \ . \end{aligned} \tag{3.9}$$

Here, we have assumed a Helmholtz Free energy of the form,

$$\psi \quad = \quad \frac{1}{2\rho} (\boldsymbol{\varepsilon} - \alpha_T \mathbf{1}\theta) : \mathbb{C} : (\boldsymbol{\varepsilon} - \alpha_T \mathbf{1}\theta) \quad , \tag{3.10}$$

which implies,

$$\begin{aligned} \boldsymbol{\sigma} &= \rho\frac{\partial\psi}{\partial\boldsymbol{\varepsilon}} = \mathbb{C} : (\boldsymbol{\varepsilon} - \alpha_T \mathbf{1}\theta) \\ &= \mathbb{C} : \boldsymbol{\varepsilon} - 3\kappa\alpha_T \mathbf{1}\theta \quad , \end{aligned} \tag{3.11}$$

where $\mathbb{C}$ is the elasticity tensor, α_T is the linear coefficient of thermal expansion, κ is the bulk modulus. The second term on the right hand side of the two equations in (3.9) represent the coupling between the equations. The term coupling the thermal variable into the mechanical domain results from the usual assumption that thermal strains are linearly related to temperature fluctuations. The other term coupling the mechanical variable into the thermal domain can easily be interpreted in analogy to the thermodynamics of gases, where the volumetric strain rate $\mathrm{tr}\,(\dot{\boldsymbol{\varepsilon}})$ corresponds to the volumetric rate of expansion of a gas. We know that when a gas expands adiabatically the temperature decreases, and vice versa, which is observed in the equation above.

3.2.2 Dimensional analysis of the governing equations

A dimensional analysis of the governing equations is useful in evaluating the effect of material parameters on the equations. The parameters governing the problem and values used to evaluate the non-dimensionalized coefficients are summarized in Table 3.1, where values appropriate for polysilicon have been selected. The characteristic values for non-dimensionalization of length, time,

Table 3.1: Governing parameters and polysilicon material parameters [168]

Domain	**Parameter**		**Value**	
Mechanical parameter	Length	L	1×10^{-6}	[m]
	Young's Modulus	E	150.0	[GPa]
	Poisson's Ratio	ν	0.226	[-]
	Density	ρ	2330.0	[kg/m^3]
Thermal parameter	Thermal Expansion Coeff.	a_T	2.6×10^{-6}	[1/K]
	Thermal Capacity at Const. Pressure	c_p	712.0	[J/kg/K]
	Thermal Conductivity	κ_T	30.0	[W/m/K]
	Referential Temperature	T_0	293.15	[K]

and mass can be chosen intuitively as follows.

$$\begin{aligned}
&\text{Characteristic length} &:L_D &= L \\
&\text{Characteristic time} &:T_D &= L_D/c \qquad (c = \sqrt{E/\rho}) \\
&\text{Characteristic mass} &:M_D &= \rho L_D^3
\end{aligned}$$

The characteristic value for temperature Θ_D is left to be determined from the analysis. The non-dimensionalized form of the governing Equations (3.9) becomes,

$$\begin{aligned}
\ddot{\widetilde{\mathbf{u}}} &= \widetilde{\boldsymbol{\nabla}} \cdot \left[\widetilde{\mathbb{C}} : \widetilde{\boldsymbol{\varepsilon}}\right] \quad -\xi_1 \widetilde{\boldsymbol{\nabla}} \widetilde{\theta} \\
\dot{\widetilde{\theta}} &= \xi_2 \widetilde{\boldsymbol{\nabla}}^2 \widetilde{\theta} \qquad\quad -\xi_3 \dot{\widetilde{\boldsymbol{\varepsilon}}}
\end{aligned} \tag{3.12}$$

where the tildes represent the non-dimensionalized quantities. The dimensionless parameters are,

$$\xi_1 = \alpha_T \Theta_D \ , \quad \xi_2 = \frac{\kappa_T}{\rho c_v} \frac{L_D}{T_D} \ , \quad \xi_3 = \frac{3\alpha_T \kappa T_D}{c_v} \frac{T_0}{\Theta_D} \ , \quad \widetilde{\mathbb{C}} = \frac{1}{E}\mathbb{C} \ . \tag{3.13}$$

The choice of $\xi_3 = 1$ presents us with a convenient choice for Θ_D.

$$\Theta_D = \frac{3\alpha_T \kappa T_D}{c_v} T_0 \ . \tag{3.14}$$

Using the material properties in Table 3.1, the coupling coefficients become:

$$\xi_1 = 4.6 \times 10^{-7} \ , \quad \xi_2 = 1.1 \times 10^{-8} \ , \quad \xi_3 = 1 \ . \tag{3.15}$$

For typical MEMS problems, it is clear that both ξ_1 and ξ_2 are small. $\xi_1 << 1$ implies weak coupling of the thermal variable into the mechanical domain, that the response is almost purely mechanical. The situation for the thermal domain is significantly different. The combination of $\xi_2 << 1$ and $\xi_3 = 1$ implies a strong coupling of the mechanical variable into the thermal domain, that thermal response is mainly driven by the mechanical motion.

3.2.3 Finite Element Discretization

The finite element method is employed to discretize and numerically solve the linear thermoelastic Equations (3.12). The non-dimensionalized form of the governing equations has been selected, for their simplicity and usefulness in observing the weak coupling between the domains. With an

abuse of notation, the tilde's which represent the normalization are dropped for clarity. ξ_3 is also omitted, since we have chosen the characteristic temperature to render this quantity unity. The governing equations in weak form are,

$$\begin{aligned} \int_\Omega \ddot{\mathbf{u}} \cdot \mathbf{w} d\Omega + \int_\Omega \boldsymbol{\sigma} : \boldsymbol{\varepsilon}(\mathbf{w}) d\Omega &= \int_\Gamma (\boldsymbol{\sigma} \cdot \mathbf{n}) \cdot \mathbf{w} d\Gamma \\ \int_\Omega \dot{\theta}\delta\theta \ d\Omega + \xi_2 \int_\Omega \boldsymbol{\nabla}\delta\theta \cdot \boldsymbol{\nabla}\theta d\Omega + \int_\Omega \dot{\boldsymbol{\varepsilon}} : (\mathbb{C} : \mathbf{1}) \, \delta\theta \ d\Omega &= \int_\Gamma (\boldsymbol{\nabla}\theta) \cdot \mathbf{n} \ \delta\theta \ d\Omega \end{aligned} \quad , (3.16)$$

where we have multiplied each equation by admissible test functions $\mathbf{w}$ and $\delta\theta$ respectively, integrated over the whole body Ω, and conducted an integration by parts. Here we are assuming a constitutive relation of the form,

$$\boldsymbol{\sigma} = \mathbb{C} : \boldsymbol{\varepsilon} - \xi_1 \theta \mathbf{1} \ . \tag{3.17}$$

The element matrices take the form,

$$\begin{aligned} \mathbf{m}_e &= \begin{bmatrix} \int_\Omega \mathbf{N}_2^T \mathbf{N}_2 \ d\Omega & 0 \\ 0 & 0 \end{bmatrix} \\ \mathbf{d}_e &= \begin{bmatrix} 0 & 0 \\ \int_\Omega \mathbf{N}_1^T \ \mathbf{q}^T \mathbf{B}_2 \ d\Omega & \int_\Omega \mathbf{N}_1^T \mathbf{N}_1 \ d\Omega \end{bmatrix} \\ \mathbf{k}_e &= \begin{bmatrix} \int_\Omega \mathbf{B}_2^T \mathbf{C} \mathbf{B}_2 \ d\Omega & -\xi_1 \int_\Omega \ \mathbf{B}_2^T \ \mathbf{q} \mathbf{N}_1 \ d\Omega \\ 0 & \xi_2 \int_\Omega \mathbf{B_1}^T \mathbf{B}_1 \ d\Omega \end{bmatrix} , \end{aligned} \tag{3.18}$$

where $\mathbf{N}_1, \mathbf{N}_2$ are the interpolation functions for the scalar and vector valued problem, $\mathbf{B}_1, \mathbf{B}_2$ are the discrete gradient operator for the scalar and vector valued problem, $\mathbf{C}$ is the matrix of material parameters, and $\mathbf{q}$ is the thermal stiffness matrix. The globally assembled system of equations inherits the matrix structure of each element, giving rise to the following global system:

$$\begin{bmatrix} \mathbf{M}_{uu} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{pmatrix} \ddot{\mathbf{u}} \\ \ddot{\boldsymbol{\theta}} \end{pmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{D}_{tu} & \mathbf{D}_{tt} \end{bmatrix} \begin{pmatrix} \dot{\mathbf{u}} \\ \dot{\boldsymbol{\theta}} \end{pmatrix} + \begin{bmatrix} \mathbf{K}_{uu} & \xi_1 \mathbf{K}_{ut} \\ \mathbf{0} & \xi_2 \mathbf{K}_{tt} \end{bmatrix} \begin{pmatrix} \mathbf{u} \\ \boldsymbol{\theta} \end{pmatrix} = \begin{pmatrix} \mathbf{F}_u \\ \mathbf{F}_t \end{pmatrix} \ . \tag{3.19}$$

In more compact form, we have

$$\mathbf{M}\ddot{\mathbf{z}} + \mathbf{D}\dot{\mathbf{z}} + \mathbf{K}\mathbf{z} = \mathbf{F} \ . \tag{3.20}$$

The matrices have the following unique structure:

1. The mass matrix $\mathbf{M}$ and damping matrix $\mathbf{D}$ are rank deficient, and damping and stiffness matrices $\mathbf{D}, \mathbf{K}$ are unsymmetric. This is completely different from the purely mechanical problem, where $\mathbf{M}, \mathbf{K}$ are symmetric positive definite.

2. The submatrices $\mathbf{M}_{uu}, \mathbf{D}_{tt}, \mathbf{K}_{uu}, \mathbf{K}_{tt}$ are symmetric positive definite, since they correspond to system matrices obtained from the purely mechanical or thermal problem.

3. The submatrices representing the thermoelastic coupling are related to one another by

$$\mathbf{D}_{tu} = -\mathbf{K}_{ut}^T \ . \tag{3.21}$$

4. The weak coupling of the thermal variable into the mechanical domain is inherited in the discretized system by the coefficient $\xi_1 << 1$.

5. In what is to follow, these equations are utilized by proportionally driving $\mathbf{F}$ in steady oscillation and the proportionally sensing $\mathbf{z}$. This naturally leads to the notion of evaluating a transfer function. Section 3.3 discusses an efficient means to do this by exploiting the structure of the equations.

3.3 Reduced order modeling

Finite element discretization of partial differential equations often results in large size N linear systems of equations, when an accurate solution is sought. For a linear static problem, this may not be of utmost importance, since only one solve is required. But in a steady state linear dynamic problem for computing a transfer function, this does make a difference, because the system of equations must be solved repeatedly for each data point in the transfer function. To reduce the time and computational effort of each solve, model reduction techniques have been developed to project the large size N solution space of the original system onto a size n subspace of solutions, where $n << N$, producing a substantially smaller system of equations to solve.

Here, we extend a technique based on second order Krylov subspaces, proposed by Bai and Su [20, 19], to match the moments of the second order transfer function for the reduced system with that of the original full system. It is highly advantageous to directly treat the second order form compared to converting to first order form, since there is no need to double the number of variables, which is computationally expensive.

Several lemmas and theorems considering the moment matching properties for the transfer function in second order form are presented. These differ from the versions presented by Li and Bai [124], which are given for the transfer function in first order form. The conversion to first order form has arbitrariness in choosing the auxiliary variable, making it difficult to show the inclusion of subspaces necessary in proving moment matching properties. Application to the thermoelastic problem makes clear the advantage of structure preservation, which leads to twice as many matched moments as one would normally expect. This doubling of the moment matching is due to the inclusion of both left and right second order Krylov subspaces in our projection subspace.

3.3.1 Moments of the transfer function in second order form

Let us assume our forcing of the true system can be represented by a forcing pattern $\mathbf{b}$ with time varying part $u(t)$. Further, let us assume that our system output $y(t)$ can be represented by a fixed sense pattern $\mathbf{l}$. This gives us the following continuous time-invariant second order system:

$$\mathbf{M}\ddot{\mathbf{z}}(t) + \mathbf{D}\dot{\mathbf{z}}(t) + \mathbf{K}\mathbf{z}(t) = \mathbf{b}u(t) \tag{3.22}$$

$$y(t) = \mathbf{l}^*\mathbf{z}(t) \ , \tag{3.23}$$

where a superposed $*$ denotes conjugate transpose. By taking the Laplace transform of the system, we have

$$s^2\mathbf{M}\widetilde{\mathbf{z}}(s) + s\mathbf{D}\widetilde{\mathbf{z}}(s) + \mathbf{K}\widetilde{\mathbf{z}}(s) = \mathbf{b}\widetilde{u}(s) \tag{3.24}$$

$$\widetilde{y}(s) = \mathbf{l}^*\widetilde{\mathbf{z}}(s) \ , \tag{3.25}$$

where the tilde's represent the Laplace transform of each variable. The relevant system transfer function in second order form is written as,

$$H(s) \quad = \quad \mathbf{l}^* \left(s^2\mathbf{M} + s\mathbf{D} + \mathbf{K}\right)^{-1} \mathbf{b} \ . \tag{3.26}$$

Evaluating the transfer function for a range of frequencies can be computationally expensive, due to the large size N of the linear system of equations. Here, we conduct model reduction on this system based on a subspace projection method proposed by Li and Bai [124]. Matrices $\mathbf{X}$ and $\mathbf{Y}$, whose columns span selected subspaces, are used as the projection space for model reduction. By defining,

$$\mathbf{l}_R = \mathbf{X}^*\mathbf{l} \ , \quad \mathbf{b}_R = \mathbf{Y}^*\mathbf{b} \ , \quad \mathbf{M}_R = \mathbf{Y}^*\mathbf{M}\mathbf{X} \ , \quad \mathbf{D}_R = \mathbf{Y}^*\mathbf{D}\mathbf{X} \ , \quad \mathbf{K}_R = \mathbf{Y}^*\mathbf{K}\mathbf{X} \ , \tag{3.27}$$

the transfer function in second order form for the reduced order model is,

$$H_R(s) \quad = \quad \mathbf{l}_R^* \left(s^2\mathbf{M}_R + s\mathbf{D}_R + \mathbf{K}_R\right)^{-1} \mathbf{b}_R \tag{3.28}$$

If $\mathbf{X}$ and $\mathbf{Y}$ are of dimension N-by-n where $n << N$, then $H_R(s)$ can be evaluated quickly. However, for $H_R(s)$ to be useful, we must choose $\mathbf{X}$ and $\mathbf{Y}$ carefully so that it is an accurate approximation to $H(s)$, in some range of frequencies.

The type of accuracy that we aim for is moment matching between the reduced order model and the original full model. The moments of a function are defined as the coefficients of the power series expansion around a given point. The transfer function expanded at $s = 0$ for each model is,

$$H(s) \quad = \quad \sum_{i=0}^{\infty} M_i s^i \tag{3.29}$$

$$H_R(s) \quad = \quad \sum_{i=0}^{\infty} M_{Ri} s^i \ , \tag{3.30}$$

where M_i and M_{Ri} are the moments. We present the following lemma to compute the moments of the second order system.

Lemma 3.3.1 *An equation of the form,*

$$f(s) \quad = \quad \mathbf{p}^* \left(\mathbf{I} - s\mathbf{A} - s^2\mathbf{B}\right)^{-1} \mathbf{q} \tag{3.31}$$

has the representation around $s = 0$*,*

$$f(s) \;=\; \sum_{i=0}^{\infty} \left(\mathbf{p}^* \mathbf{E}^i \mathbf{q}\right) s^i \;, \tag{3.32}$$

where $\mathbf{A}, \mathbf{B}, \left\{\mathbf{E}^i\right\}_{i=0}^{\infty}$ *are* N*-by-*N *matrices and* $\mathbf{p}, \mathbf{q}$ *are size* N *vectors. The matrices* $\mathbf{E}^i (0 \leq i)$ *are given by the recursion,*

$$\mathbf{E}^0 \;=\; \mathbf{I} \tag{3.33}$$

$$\mathbf{E}^1 \;=\; \mathbf{A} \tag{3.34}$$

$$\mathbf{E}^i \;=\; \mathbf{A}\mathbf{E}^{i-1} + \mathbf{B}\mathbf{E}^{i-2} \quad (i \geq 2) \tag{3.35}$$

$$\;=\; \mathbf{E}^{i-1}\mathbf{A} + \mathbf{E}^{i-2}\mathbf{B} \quad (i \geq 2) \;. \tag{3.36}$$

Proof The proof of this Lemma follows as a consequence of A.0.1 in the Appendix. □

Using this lemma the transfer function can be written in the following two equivalent forms,

$$\begin{aligned} H(s) \;&=\; \mathbf{l}^* \left(\mathbf{I} + s\mathbf{K}^{-1}\mathbf{D} + s^2\mathbf{K}^{-1}\mathbf{M}\right)^{-1} \mathbf{K}^{-1}\mathbf{b} \\ &=\; \sum_{i=0}^{\infty} \left[\mathbf{l}^* \mathbf{E}_r^i \left(\mathbf{K}^{-1}\mathbf{b}\right)\right] s^i \end{aligned} \tag{3.37}$$

where $\{\mathbf{E}_r^i\}_{i=0}^{\infty}$ is defined by the recursion in Lemma 3.3.1 with the substitution, $\mathbf{A} \rightarrow -\mathbf{K}^{-1}\mathbf{D}$, $\mathbf{B} \rightarrow -\mathbf{K}^{-1}\mathbf{M}$ and,

$$\begin{aligned} H(s) \;&=\; \mathbf{l}^*\mathbf{K}^{-1} \left(\mathbf{I} + s\mathbf{D}\mathbf{K}^{-1} + s^2\mathbf{M}\mathbf{K}^{-1}\right)^{-1} \mathbf{b} \;\;, \\ &=\; \left[\mathbf{b}^* \left(\mathbf{I} + s^*\mathbf{K}^{*,-1}\mathbf{D}^* + s^{*2}\mathbf{K}^{*,-1}\mathbf{M}^*\right)^{-1} \mathbf{K}^{*,-1}\mathbf{l}\right]^* \\ &=\; \left[\sum_{i=0}^{\infty} \left[\mathbf{b}^* \mathbf{E}_l^{i^*} \left(\mathbf{K}^{*,-1}\mathbf{l}\right)\right] s^{*i}\right]^* \\ &=\; \sum_{i=0}^{\infty} \left[\mathbf{b}^* \mathbf{E}_l^{i^*} \left(\mathbf{K}^{*,-1}\mathbf{l}\right)\right]^* s^i \;\;, \end{aligned} \tag{3.38}$$

where, $\{\mathbf{E}_l^{i^*}\}_{i=0}^{\infty}$ is defined by the recursion with substitution $\mathbf{A} \rightarrow -\mathbf{K}^{*,-1}\mathbf{D}^*$, $\mathbf{B} \rightarrow -\mathbf{K}^{*,-1}\mathbf{M}^*$. The second form is presented here to motivate the notion and the equivalence of the left second order Krylov subspace to the right second order Krylov subspace, which are both presented in the

following section. The subscripts for the sequences of matrices $\{\mathbf{E}^i\}_{i=0}^{\infty}$ adhere to the convention, that r denotes the sequence generated by the right second order Krylov subspace and l denotes that generated by the left second order Krylov subspace. The expressions for the moments become,

$$M_n = \mathbf{l}^*\mathbf{E}_r^i\mathbf{K}^{-1}\mathbf{b} = \mathbf{l}^*\mathbf{K}^{-1}\mathbf{E}_l^i\mathbf{b} \ , \tag{3.39}$$

and the moments for the reduced transfer functions are obtained by substituting $\mathbf{M}, \mathbf{D}, \mathbf{K}$ with their corresponding versions subscripted by R,

$$M_{Ri} = \mathbf{l}_R^*\mathbf{E}_{r,R}^i\mathbf{K}_R^{-1}\mathbf{b}_R = \mathbf{l}_R^*\mathbf{K}_R^{-1}\mathbf{E}_{l,R}^i\mathbf{b}_R \ . \tag{3.40}$$

Here $\{\mathbf{E}_{r,R}^i\}_{i=0}^{\infty}$ is defined by the recursion with substitution, $\mathbf{A} \rightarrow -\mathbf{K}_R^{-1}\mathbf{D}_R,\ \mathbf{B} \rightarrow -\mathbf{K}_R^{-1}\mathbf{M}_R$ and $\{\mathbf{E}_{l,R}^{i\ *}\}_{i=0}^{\infty}$ by the recursion with substitution $\mathbf{A} \rightarrow -\mathbf{K}_R^{*,-1}\mathbf{D}_R^*,\ \mathbf{B} \rightarrow -\mathbf{K}_R^{*,-1}\mathbf{M}_R^*$.

The form of the recursion relationship given in Lemma 3.3.1 and its relation to the moments give rise to the following definition of the second order Krylov subspaces.

3.3.2 Second order Krylov subspaces

The number of matching moments k differ depending on the selection of $\mathbf{X}$ and $\mathbf{Y}$. In the method that we select, $\mathbf{X}$ and $\mathbf{Y}$ are constructed to contain second order Krylov subspaces. A kth second order Krylov subspace $\mathcal{G}_k(\mathbf{A}, \mathbf{B}, \mathbf{r})$, generated by the matrices $\mathbf{A}, \mathbf{B}$ and initial vector $\mathbf{r}$, is defined as the subspace spanned by the sequence of vectors $\{\mathbf{r}_i\}_{i=0}^{k-1}$ generated by the following recursion.

$$\mathbf{r}_0 = \mathbf{r} \tag{3.41}$$

$$\mathbf{r}_1 = \mathbf{A}\mathbf{r}_0 \tag{3.42}$$

$$\mathbf{r}_i = \mathbf{A}\mathbf{r}_{i-1} + \mathbf{B}\mathbf{r}_{i-2} \quad (2 \leq i \leq k-1) \ . \tag{3.43}$$

An orthogonal basis spanning this second order Krylov subspace can be numerically generated by the Second Order Arnoldi (SOAR) method [20]. Comparison between the definition of the second order Krylov subspace and the recursive nature of the sequence of matrices $\{\mathbf{E}^i\}_{n=0}^{\infty}$ in Lemma 3.3.1

reveals the following correspondence between this matrix sequence and sequence of vectors spanning the subspace.

Lemma 3.3.2 *The sequence of vectors* $\{\mathbf{r}_i\}_{i=0}^{k-1}$ *spanning the second order Krylov subspace* $\mathcal{G}_k(\mathbf{A}, \mathbf{B}, \mathbf{r})$ *is defined by,*

$$\mathbf{r}_i = \mathbf{E}^i \mathbf{r} , \tag{3.44}$$

where $\{\mathbf{E}^i\}_{i=0}^{k-1}$ *is the sequence of matrices defined in Lemma 3.3.1.*

Proof This is easily seen by comparing the expressions for $\mathbf{E}^i$ in Lemma 3.3.1 with that obtained from inserting Equation (3.44) into Equations (3.41,3.42, 3.43). □

The second order Krylov subspace for the transfer function in Equation (3.26) is defined as $\mathcal{G}_k\left(\mathbf{K}^{-1}\mathbf{D}, \mathbf{K}^{-1}\mathbf{M}, \mathbf{K}^{-1}\mathbf{b}\right)$. From Lemma 3.3.2, we obtain the expression for the sequence of vectors $\{\mathbf{r}_{r,i}\}_{i=0}^{k-1}$ spanning this subspace.

$$\mathbf{r}_{r,i} = \mathbf{E}_r^i \mathbf{K}^{-1}\mathbf{b} . \tag{3.45}$$

Analogous to the Standard Krylov subspaces, we can define a left second order Krylov subspace $\mathcal{G}_k\left(\mathbf{K}^{*,-1}\mathbf{D}^*, \mathbf{K}^{*,-1}\mathbf{M}, \mathbf{K}^{*,-1}\mathbf{l}\right)$ for the transfer function. Again from the same Lemma, we obtain the expression for the sequence of vectors $\{\mathbf{r}_{l,i}\}_{i=0}^{k-1}$ spanning this subspace.

$$\mathbf{r}_{l,i} = \mathbf{E}_l^{i^*} \mathbf{K}^{*,-1}\mathbf{l} . \tag{3.46}$$

Again, similar to the case for the sequence $\{\mathbf{E}^i\}_{i=0}^{\infty}$, subscript r denotes relation to the right and subscript l denotes relation to the left second order Krylov subspace.

3.3.3 Moment matching theorems

The technique used to prove the moment matching property between the reduced and full model by selection of $\mathbf{X}$ as $\mathbf{Y}$ is based on projection techniques. Here, we extend the technique originally applied to the transfer function in first order form [124] to the transfer function in second order

form. The details of the proof of the theorem are presented in the Appendix. The extension of the moment matching theorem to second order form is essential for showing that we can obtain double moment matching in the thermoelastic problem if we perform the moment matching in second order form.

Theorem 3.3.1 *Let matrices be defined as in Equations (3.27). Let integers $k, r \geq 0$. If,*

$$\mathcal{G}_k\left(\mathbf{K}^{-1}\mathbf{D}, \mathbf{K}^{-1}\mathbf{M}; \mathbf{K}^{-1}\mathbf{b}\right) \subset \mathsf{span}\left(\mathbf{X}\right) \tag{3.47}$$

$$\mathcal{G}_k\left(\mathbf{K}^{*,-1}\mathbf{D}^*, \mathbf{K}^{*,-1}\mathbf{M}^*; \mathbf{K}^{*,-1}\mathbf{l}\right) \subset \mathsf{span}\left(\mathbf{Y}\right) \tag{3.48}$$

then

$$\mathbf{E}_r^i\mathbf{K}^{-1}\mathbf{b} = \mathbf{X}\mathbf{E}_{R,r}^i\mathbf{K}_R^{-1}\mathbf{b}_R \qquad (0 \leq i \leq k-1) \tag{3.49}$$

$$\mathbf{E}_l^{i\,*}\mathbf{K}^{*,-1}\mathbf{l} = \mathbf{Y}{\mathbf{E}_{R,l}^i}^*\mathbf{K}_R^{*,-1}\mathbf{l}_R \qquad (0 \leq j \leq r-1)\ . \tag{3.50}$$

As a result,

$$M_i = M_{Ri} \qquad (0 \leq i \leq k+r-1)\ . \tag{3.51}$$

Proof See Appendix. □

In resonator applications, the frequency range of interest for the transfer function is centered around some non-zero value s_0. The second order transfer function can be rewritten incorporating this shift as,

$$H(s) = \mathbf{l}^*\left((s-s_0)^2\mathbf{M} + (s-s_0)\mathbf{D}_{s_0} + \mathbf{K}_{s_0}\right)^{-1}\mathbf{b}\ , \tag{3.52}$$

where the subscript s_0 denotes the shifted versions of $\mathbf{D}$ and $\mathbf{K}$.

$$\mathbf{D}_{s_0} = 2s_0\mathbf{M} + \mathbf{D} \tag{3.53}$$

$$\mathbf{K}_{s_0} = s_0^2\mathbf{M} + s_0\mathbf{D} + \mathbf{K}\ . \tag{3.54}$$

Application of Theorem 3.3.1 to the case incorporating a shift s_0 results in the following corollary.

Corollary 3.3.1 *Let integers $k, r \geq 0$. Let matrices be defined as in Equations (3.27, 3.53, 3.54). Additionally define,*

$$\mathbf{D}_{s_0,R} = \mathbf{Y}^* \mathbf{D}_{s_0} \mathbf{X} \ , \quad \mathbf{K}_{s_0,R} = \mathbf{Y}^* \mathbf{K}_{s_0} \mathbf{X} \ , \tag{3.55}$$

If,

$$\mathcal{G}_k \left(\mathbf{K}_{s_0}^{-1} \mathbf{D}_{s_0}, \mathbf{K}_{s_0}^{-1} \mathbf{M}; \mathbf{K}_{s_0}^{-1} \mathbf{b} \right) \subset \mathsf{span} \left(\mathbf{X} \right) \tag{3.56}$$

$$\mathcal{G}_k \left(\mathbf{K}_{s_0}^{*,-1} \mathbf{D}_{s_0}, \mathbf{K}_{s_0}^{*,-1} \mathbf{M}^*; \mathbf{K}_{s_0}^{*,-1} \mathbf{l} \right) \subset \mathsf{span} \left(\mathbf{Y} \right) \tag{3.57}$$

then

$$M_{s_0,i} = M_{s_0,Ri} \qquad (0 \leq i \leq k + r - 1) \ , \tag{3.58}$$

where,

$$M_{s_0,i} \ = \ \mathbf{l}^* \mathbf{E}_{r,s_0}^i \mathbf{K}_{s_0}^{-1} \mathbf{b} \qquad = \mathbf{l}^* \mathbf{K}_{s_0}^{-1} {\mathbf{E}_{l,s_0}^i}^* \mathbf{b} \tag{3.59}$$

$$M_{s_0,Ri} \ = \ \mathbf{l}_R^* \mathbf{E}_{r,s_0,R}^i \mathbf{K}_{s_0,R}^{-1} \mathbf{b}_R = \mathbf{l}_R^* \mathbf{K}_{s_0,R}^{-1} {\mathbf{E}_{l,s_0,R}^i}^* \mathbf{b}_R \ . \tag{3.60}$$

This implies,

$$H(s) \ = \ H_R(s) + \mathcal{O} \left((s - s_0)^{k+r} \right) \ . \tag{3.61}$$

3.3.4 Structure preservation for the thermoelastic problem

Here, we focus on model reduction of the linear thermoelastic problem in Equation (3.19). Preserving matrix structure of an original transfer function in second order form has been shown to be advantageous in obtaining highly accurate reduced order models [124, 37]. This also holds true in the thermoelastic problem, where we merge the two approaches mentioned above to apply model reduction to a system incorporating both thermoelastic damping and anchor loss.

Since the method of actuation and sensing is purely mechanical in resonator applications, we

assume that the thermal parts of $\mathbf{b}$ and $\mathbf{l}$ are zero.

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_u \\ \mathbf{F}_t \end{bmatrix} = \mathbf{b}u(t) \tag{3.62}$$

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}_u \\ \mathbf{b}_t \end{bmatrix} = \begin{bmatrix} \mathbf{b}_u \\ \mathbf{0} \end{bmatrix} \tag{3.63}$$

$$\mathbf{l} = \begin{bmatrix} \mathbf{l}_u \\ \mathbf{l}_t \end{bmatrix} = \begin{bmatrix} \mathbf{l}_u \\ \mathbf{0} \end{bmatrix} . \tag{3.64}$$

For this problem, we assume the matrices $\mathbf{M}, \mathbf{D}, \mathbf{K}, \mathbf{z}$ have the substructure presented in Equation (3.19), $\mathbf{b} = [\mathbf{b}_u^T, \mathbf{b}_t^T]^T$, $\mathbf{l} = [\mathbf{l}_u^T, \mathbf{l}_t^T]^T$, and $\mathbf{M}, \mathbf{D}, \mathbf{K} \in \mathcal{C}^{N_u+N_t \times N_u+N_t}$, $\mathbf{z} \in \mathcal{C}^{N_u+N_t}$, $\mathbf{b}_u, \mathbf{l}_u \in \mathcal{C}^{N_u}$, $\mathbf{b}_t, \mathbf{l}_t \in \mathcal{C}^{N_t}$, where N_u, N_t are the mechanical and thermal degrees of freedom respectively. We define new matrices $\mathbf{K}^{new}, \mathbf{D}^{new}$ so that,

$$\mathbf{K}^{new} = \begin{bmatrix} \mathbf{K}_{uu}^{new} & \mathbf{K}_{ut}^{new} \\ \mathbf{0} & \mathbf{K}_{tt}^{new} \end{bmatrix} := \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\xi_1\mathbf{I} \end{bmatrix} \mathbf{K} = \begin{bmatrix} \mathbf{K}_{uu} & \xi_1\mathbf{K}_{ut} \\ \mathbf{0} & -\xi_1\xi_2\mathbf{K}_{tt} \end{bmatrix} \tag{3.65}$$

$$\mathbf{D}^{new} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{D}_{tu}^{new} & \mathbf{D}_{tt}^{new} \end{bmatrix} := \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\xi_1\mathbf{I} \end{bmatrix} \mathbf{D} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ -\xi_1\mathbf{D}_{tu} & -\xi_1\mathbf{D}_{tt} \end{bmatrix} , \tag{3.66}$$

and substitute these for their previous versions,

$$\mathbf{K} := \mathbf{K}^{new} , \quad \mathbf{D} := \mathbf{D}^{new} . \tag{3.67}$$

Now, $\mathbf{K}, \mathbf{D}$ become,

$$\mathbf{D} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{D}_{tu} & \mathbf{D}_{tt} \end{bmatrix} , \ \mathbf{K} = \begin{bmatrix} \mathbf{K}_{uu} & \mathbf{K}_{ut} \\ \mathbf{0} & \mathbf{K}_{tt} \end{bmatrix} , \tag{3.68}$$

and the transfer function in second order form is,

$$H(s) = \begin{bmatrix} \mathbf{l}_u \\ \mathbf{0} \end{bmatrix}^* \left(s^2 \begin{bmatrix} \mathbf{M}_{uu} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + s \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{D}_{tu} & \mathbf{D}_{tt} \end{bmatrix} + \begin{bmatrix} \mathbf{K}_{uu} & \mathbf{K}_{ut} \\ \mathbf{0} & \mathbf{K}_{tt} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{b}_u \\ \mathbf{0} \end{bmatrix} \tag{3.69}$$

$$= \mathbf{l}^* \left(s^2\mathbf{M} + s\mathbf{D} + \mathbf{K} \right) \mathbf{b} . \tag{3.70}$$

From the relation between the coupling submatrices in Equation (3.21), we now have $\mathbf{D}_{tu}^T = \mathbf{K}_{ut}$. To apply Corollary 3.3.1 for model reduction, $\mathbf{X}, \mathbf{Y}$ must contain the subspaces spanned by the right second order Krylov subspace $\mathcal{G}_k(\mathbf{K}_{s_0}^{-1}\mathbf{D}_{s_0}, \mathbf{K}_{s_0}^{-1}\mathbf{M}; \mathbf{K}_{s_0}^{-1}\mathbf{b})$, and left second order Krylov subspace $\mathcal{G}_k(\mathbf{K}_{s_0}^{*,-1}\mathbf{D}_{s_0}^{*}, \mathbf{K}_{s_0}^{*,-1}\mathbf{M}^{*}; \mathbf{K}_{s_0}^{*,-1}\mathbf{l})$, respectively. The sequence of vectors $\{\mathbf{r}_{r,i}\}_{i=0}^{k-1}$ which span the right second order Krylov subspace and sequence of vectors $\{\mathbf{r}_{l,i}\}_{i=0}^{k-1}$ which span the left second order Krylov subspace can be partitioned into their mechanical and thermal degrees of freedom as follows:

$$\mathbf{r}_{r,i} = \begin{bmatrix} \mathbf{r}_{r,i}^u \\ \mathbf{r}_{r,i}^t \end{bmatrix} \quad (0 \leq i \leq k-1) \tag{3.71}$$

$$\mathbf{r}_{l,i} = \begin{bmatrix} \mathbf{r}_{l,i}^u \\ \mathbf{r}_{l,i}^t \end{bmatrix} \quad (0 \leq i \leq k-1)\ . \tag{3.72}$$

For the special case when we have identical mechanical forcing and sensing vectors,

$$\mathbf{b}_u = \mathbf{l}_u\ , \quad \mathbf{b}_t = \mathbf{l}_t = \mathbf{0}\ , \tag{3.73}$$

we obtain the following relationship between $\{\mathbf{r}_{r,i}\}_{i=0}^{k-1}$ and $\{\mathbf{r}_{l,i}\}_{i=0}^{k-1}$,

$$\begin{bmatrix} \mathbf{r}_{l,i}^u \\ \mathbf{r}_{l,i}^t \end{bmatrix} = \begin{bmatrix} \mathbf{r}_{r,i}^u \\ \sum_{j=0}^{i} \left(-\frac{1}{s_0}\right)^i \mathbf{r}_{r,i-j}^t \end{bmatrix}\ . \tag{3.74}$$

When $s_0 = 0$, this relation reduces to,

$$\begin{bmatrix} \mathbf{r}_{l,i}^u \\ \mathbf{r}_{l,i}^t \end{bmatrix} = \begin{bmatrix} \mathbf{r}_{r,i}^u \\ \mathbf{r}_{r,i+1}^t \end{bmatrix}\ . \tag{3.75}$$

By defining the following matrices,

$$\mathbf{R}_r^u = \begin{bmatrix} \mathbf{r}_{r,0}^u & \mathbf{r}_{r,1}^u & \cdots & \mathbf{r}_{r,k-1}^u \end{bmatrix} \tag{3.76}$$

$$\mathbf{R}_r^t = \begin{bmatrix} \mathbf{r}_{r,0}^t & \mathbf{r}_{r,1}^t & \cdots & \mathbf{r}_{r,k-1}^t \end{bmatrix} \tag{3.77}$$

we have the following inclusion of subspaces,

$$\mathsf{span}\left(\begin{bmatrix} \mathbf{r}_{r,0}^u & \mathbf{r}_{r,1}^u & \cdots & \mathbf{r}_{r,k-1}^u \\ \mathbf{r}_{r,0}^t & \mathbf{r}_{r,1}^t & \cdots & \mathbf{r}_{r,k-1}^t \end{bmatrix}\right) \subset \mathsf{span}\left(\begin{bmatrix} \mathbf{R}_r^u & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_r^t \end{bmatrix}\right) \tag{3.78}$$

$$\mathsf{span}\left(\begin{bmatrix} \mathbf{r}_{l,0}^u & \mathbf{r}_{l,1}^u & \cdots & \mathbf{r}_{l,k-1}^u \\ \mathbf{r}_{l,0}^t & \mathbf{r}_{l,1}^t & \cdots & \mathbf{r}_{l,k-1}^t \end{bmatrix}\right) \subset \mathsf{span}\left(\begin{bmatrix} \mathbf{R}_r^u & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_r^t \end{bmatrix}\right) . \tag{3.79}$$

Thus, if we define $\mathbf{X}_s$ so that

$$\mathsf{span}\left(\mathbf{X}_s^u\right) = \mathsf{span}\left(\mathbf{R}_r^u\right) \tag{3.80}$$

$$\mathsf{span}\left(\mathbf{X}_s^t\right) = \mathsf{span}\left(\mathbf{R}_r^t\right) \tag{3.81}$$

$$\mathbf{X}_s = \begin{bmatrix} \mathbf{X}_s^u & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_s^t \end{bmatrix} \tag{3.82}$$

and $\mathbf{X}_s^* \mathbf{X}_s = \mathbf{I}$, then

$$\mathcal{G}_k(\mathbf{K}_{s_0}^{-1}\mathbf{D}_{s_0}, \mathbf{K}_{s_0}^{-1}\mathbf{M}; \mathbf{K}_{s_0}^{-1}\mathbf{b}) \subset \mathsf{span}(\mathbf{X}_s) \tag{3.83}$$

$$\mathcal{G}_k(\mathbf{K}_{s_0}^{*,-1}\mathbf{D}_{s_0}^*, \mathbf{K}_{s_0}^{*,-1}\mathbf{M}^*; \mathbf{K}_{s_0}^{*,-1}\mathbf{l}) \subset \mathsf{span}(\mathbf{X}_s) \ . \tag{3.84}$$

Here, the subscript s in $\mathbf{X}_s$ denotes structure preservation, which is apparent from the expressions for the reduced matrices presented below. Corollary 3.3.1 implies that, by producing the right second order Krylov subspace and selecting $\mathbf{X} = \mathbf{X}_s$ and $\mathbf{Y} = \mathbf{X}_s$, we can in fact obtain $2k$ matching moments. Under this projection, the reduced transfer function in second order form is,

$$H_R(s) = \mathbf{l}_R^* \left(s^2\mathbf{M}_R + s\mathbf{D}_R + \mathbf{K}_R\right)^{-1} \mathbf{b}_R \ , \tag{3.85}$$

with,

$$\mathbf{l}_R = \mathbf{b}_R = \mathbf{X}_s^* \mathbf{b} = \left[\left(\mathbf{X}_s^{u,*} \mathbf{b}_u \right)^* , \mathbf{0} \right]^* , \tag{3.86}$$

$$\mathbf{M}_R = \mathbf{X}_s^* \mathbf{M} \mathbf{X}_s = \begin{bmatrix} \mathbf{X}_s^{u,*} \mathbf{M}_{uu} \mathbf{X}_s^u & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{3.87}$$

$$\mathbf{D}_R = \mathbf{X}_s^* \mathbf{D} \mathbf{X}_s = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{X}_s^{t,*} \mathbf{D}_{tu} \mathbf{X}_s^u & \mathbf{X}_s^{t,*} \mathbf{D}_{tt} \mathbf{X}_s^t \end{bmatrix} \tag{3.88}$$

$$\mathbf{K}_R = \mathbf{X}_s^* \mathbf{K} \mathbf{X}_s = \begin{bmatrix} \mathbf{X}_s^{u,*} \mathbf{K}_{uu} \mathbf{X}_s^u & \mathbf{X}_s^{u,*} \mathbf{K}_{ut} \mathbf{X}_s^t \\ \mathbf{0} & \mathbf{X}_s^{t,*} \mathbf{K}_{tt} \mathbf{X}_s^t \end{bmatrix} . \tag{3.89}$$

Additionally, we see structure preservation of the original system of equations.

In applying Perfectly Matched Layers [37] to model anchor loss, the matrices that we obtain lose their Hermitian symmetry in favor of complex symmetry.

$$\mathbf{M}_{uu}^T = \mathbf{M}_{uu} \tag{3.90}$$

$$\mathbf{D}_{tt}^T = \mathbf{D}_{tt} \tag{3.91}$$

$$\mathbf{K}_{uu}^T = \mathbf{K}_{uu} \tag{3.92}$$

$$\mathbf{K}_{tt}^T = \mathbf{K}_{tt} \tag{3.93}$$

$$\mathbf{K}_{ut}^T = \mathbf{D}_{tu} \tag{3.94}$$

Assuming that $\mathbf{b}_u, \mathbf{l}_u \in \mathcal{R}^{N_u}$, the relationship between the sequence of vectors spanning the right and left second order Krylov subspace $\{\mathbf{r}_{r,i}\}_{i=0}^{k-1}$ and $\{\mathbf{r}_{l,i}\}_{i=0}^{k-1}$ becomes,

$$\begin{bmatrix} \mathbf{r}_{l,i}^u \\ \mathbf{r}_{l,i}^t \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{r}_{r,i}^u} \\ \sum_{j=0}^{i} \left(-\frac{1}{s_0} \right)^i \overline{\mathbf{r}_{r,i-j}^t} \end{bmatrix} . \tag{3.95}$$

When $s_0 = 0$, this relation reduces to,

$$\begin{bmatrix} \mathbf{r}_{l,i}^u \\ \mathbf{r}_{l,i}^t \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{r}_{r,i}^u} \\ \overline{\mathbf{r}_{r,i+1}^t} \end{bmatrix} . \tag{3.96}$$

If we define $\mathbf{X}_{rs}$ real so that,

$$\begin{aligned}
\mathsf{span}\,(\mathbf{X}^u_{sR}) &= \mathsf{span}\left(\frac{\mathbf{R}^u_r + \overline{\mathbf{R}^u_r}}{2}\right) && (3.97)\\
\mathsf{span}\,(\mathbf{X}^u_{sI}) &= \mathsf{span}\left(\frac{\mathbf{R}^u_r - \overline{\mathbf{R}^u_r}}{2i}\right) && (3.98)\\
\mathsf{span}\,(\mathbf{X}^t_{sR}) &= \mathsf{span}\left(\frac{\mathbf{R}^t_r + \overline{\mathbf{R}^t_r}}{2}\right) && (3.99)\\
\mathsf{span}\,(\mathbf{X}^t_{sI}) &= \mathsf{span}\left(\frac{\mathbf{R}^t_r - \overline{\mathbf{R}^t_r}}{2i}\right) && (3.100)\\
\mathbf{X}_{rs} &= \begin{bmatrix} \mathbf{X}^u_{sR} & \mathbf{X}^u_{sI} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}^t_{sR} & \mathbf{X}^t_{sI} \end{bmatrix} && (3.101)
\end{aligned}$$

and $\mathbf{X}^*_{rs}\mathbf{X}_{rs} = \mathbf{I}$ where subscript R denotes real, I denotes imaginary, and r denotes the real basis in the expression $\mathbf{X}$, then

$$\begin{aligned}
\mathcal{G}_k(\mathbf{K}^{-1}_{s_0}\mathbf{D}_{s_0}, \mathbf{K}^{-1}_{s_0}\mathbf{M}; \mathbf{K}^{-1}_{s_0}\mathbf{b}) &\subset \mathsf{span}(\mathbf{X}_{rs}) && (3.102)\\
\mathcal{G}_k(\mathbf{K}^{*,-1}_{s_0}\mathbf{D}^*_{s_0}, \mathbf{K}^{*,-1}_{s_0}\mathbf{M}^*; \mathbf{K}^{*,-1}_{s_0}\mathbf{l}) &\subset \mathsf{span}(\mathbf{X}_{rs})\ . && (3.103)
\end{aligned}$$

Thus, under this projection, we again obtain $2k$ matching moments by generating the right second order Krylov subspace. This split between the real part and imaginary part of the sequence of vectors is equivalent to the split in [37] for increased moment matching.

Remark 1. Additionally, we define the matrices $\mathbf{X}_{nosplit}$ and $\mathbf{X}_r$ so that,

$$\begin{aligned}
\mathsf{span}\,(\mathbf{X}_{nosplit}) &= \mathrm{span}\left(\begin{bmatrix} \mathbf{R}^u_r \\ \mathbf{R}^t_l \end{bmatrix}\right) && (3.104)\\
\mathsf{span}\,(\mathbf{X}_r) &= \mathrm{span}\left(\begin{bmatrix} \frac{\mathbf{R}^u_r + \overline{\mathbf{R}^u_r}}{2} & \frac{\mathbf{R}^u_r - \overline{\mathbf{R}^u_r}}{2i} \\ \frac{\mathbf{R}^t_r + \overline{\mathbf{R}^t_r}}{2} & \frac{\mathbf{R}^t_r - \overline{\mathbf{R}^t_r}}{2i} \end{bmatrix}\right), && (3.105)
\end{aligned}$$

where we orthogonalize the columns so that $\mathbf{X}^*_{nosplit}\mathbf{X}_{nosplit} = \mathbf{I}$ and $\mathbf{X}^*_r\mathbf{X}_r = \mathbf{I}$.

The possibilities we have in selecting our projection matrix $\mathbf{X}$ are summarized in Table 3.2 and 3.3. The name SOAR has been applied, since we generate the matrices using the Second Order Arnoldi (SOAR) method proposed by Su and Bai [19]. These projection matrices are used to

compare the moment matching accuracy claims for each case in the numerical examples presented in Section 5.5.

Table 3.2: ROMs generated by k iterations of SOAR (TED)

Name	**Projection matrix**	**Size of ROM**	**Matched Moments**
SOAR	$\mathbf{X}_{nosplit}$	k	k
SOAR-S	$\mathbf{X}_s$	2k	2k

Table 3.3: ROMs generated by k iterations of SOAR (TED/PML)

Name	**Projection matrix**	**Size of ROM**	**Matched Moments**
SOAR	$\mathbf{X}_{nosplit}$	k	k
SOAR-S	$\mathbf{X}_s$	2k	k
SOAR-R	$\mathbf{X}_r$	2k	k
SOAR-RS	$\mathbf{X}_{rs}$	4k	2k

Remark 2. Here, we would like to mention the difficulty of applying the theorems presented by Li and Bai [124] to prove the $2k$ moment matching property for the thermoelastic problem with purely mechanical forcing and sensing vectors, and the arbitrariness in converting the second order transfer function to first order form.

Our second order system can be converted into the following first order form,

$$s\mathbf{C}\widetilde{\mathbf{Z}}(s) + \mathbf{G}\widetilde{\mathbf{Z}}(s) = \widehat{\mathbf{b}}\widetilde{u}(s) \tag{3.106}$$

$$\widetilde{y}(s) = \widehat{\mathbf{l}}^{*}\widetilde{\mathbf{Z}}(s)\ . \tag{3.107}$$

Here, the matrices are defined as,

$$\widetilde{\mathbf{Z}}(s) = \begin{bmatrix} \mathbf{u}(s) \\ s\mathbf{u}(s) \\ \boldsymbol{\theta}(s) \end{bmatrix}, \ \widehat{\mathbf{b}} = \begin{bmatrix} \mathbf{0} \\ \mathbf{b}_u \\ \mathbf{0} \end{bmatrix}, \ \widehat{\mathbf{l}} = \begin{bmatrix} \mathbf{l}_u \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \tag{3.108}$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{W} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{uu} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_{tt} \end{bmatrix}, \ \mathbf{G} = \begin{bmatrix} \mathbf{0} & -\mathbf{W} & \mathbf{0} \\ \mathbf{K}_{uu} & \mathbf{0} & \mathbf{K}_{ut} \\ \mathbf{0} & \mathbf{D}_{tu} & \mathbf{K}_{tt} \end{bmatrix}, \tag{3.109}$$

and $\mathbf{W} \in \mathcal{C}^{N_u \times N_u}$ is an arbitrary nonsingular matrix. The transfer function can be written,

$$\begin{aligned} H(s) &= \widehat{\mathbf{l}}^* \left(\mathbf{G} + s\mathbf{C}\right)^{-1} \widehat{\mathbf{b}} \\ &= \widehat{\mathbf{l}}^* \left(\mathbf{I} + s\mathbf{G}^{-1}\mathbf{C}\right)^{-1} \mathbf{G}^{-1}\widehat{\mathbf{b}} \ . \end{aligned} \tag{3.110}$$

This expression is equivalent to Equation (3.69). To apply Theorem 3.3 in [124] to prove the $2k$ moment matching property, one requires symmetric $\mathbf{G}$, symmetric $\mathbf{C}$, and $\hat{\mathbf{b}} = \hat{\mathbf{l}}$. We can clearly see that even when $\mathbf{b}_u = \mathbf{l}_u$, the force and sense vectors for the first order form $\widehat{\mathbf{b}}, \widehat{\mathbf{l}}$ are not equal. This means that even by selection of $\mathbf{W} = -\mathbf{K}_{uu}$, which leads to symmetric $\mathbf{G}, \mathbf{C}$, the cannot be applied. By changing the order of the equation, it is possible to make $\widehat{\mathbf{b}} = \widehat{\mathbf{l}}$, but then $\mathbf{G}, \mathbf{C}$ will no longer be symmetric.

Additionally, if one has a nonzero shift s_0, there is an arbitrariness in selecting the first order form. The second order shifted transfer function is presented in Equation (3.52). Using the definition $\widetilde{\mathbf{Z}}(s)$ in Equation 3.108 in combination with Equations 3.106 and 3.107, one form for the first order form is given by,

$$\begin{aligned} H(s) &= \widehat{\mathbf{l}}^* \left(\mathbf{G}_{s_0} + (s - s_0)\mathbf{C}\right)^{-1} \widehat{\mathbf{b}} \\ &= \widehat{\mathbf{l}}^* \left(\mathbf{I} + (s - s_0)\mathbf{G}_{s_0}^{-1}\mathbf{C}\right)^{-1} \mathbf{G}_{s_0}^{-1}\widehat{\mathbf{b}} \ . \end{aligned} \tag{3.111}$$

where,

$$\mathbf{G}_{s_0} = \mathbf{G} + s_0\mathbf{C} = \begin{bmatrix} s_0\mathbf{W} & -\mathbf{W} & \mathbf{0} \\ \mathbf{K}_{uu} & s_0\mathbf{M}_{uu} & \mathbf{K}_{ut} \\ \mathbf{0} & \mathbf{D}_{tu} & \mathbf{K}_{tt} + s_0\mathbf{D}_{tt} \end{bmatrix} . \tag{3.112}$$

Another form can be obtained by defining,

$$\overline{\mathbf{Z}}(s) = \begin{bmatrix} \mathbf{u}(s) \\ (s - s_0)\mathbf{u}(s) \\ \boldsymbol{\theta}(s) \end{bmatrix} . \tag{3.113}$$

Equations (3.106,3.107) can be rewritten as,

$$(s - s_0)\mathbf{C}\overline{\mathbf{Z}}(s) + \overline{\mathbf{G}}_{s_0}\overline{\mathbf{Z}}(s) = \widehat{\mathbf{b}}\widetilde{u}(s) \tag{3.114}$$

$$\widetilde{y}(s) = \widehat{\mathbf{l}}^*\overline{\mathbf{Z}}(s) \ , \tag{3.115}$$

where,

$$\overline{\mathbf{G}}_{s_0} = \begin{bmatrix} \mathbf{0} & -\mathbf{W} & \mathbf{0} \\ \mathbf{K}_{uu} + s_0^2\mathbf{M}_{uu} & 2s_0\mathbf{M}_{uu} & \mathbf{K}_{ut} \\ s_0\mathbf{D}_{tu} & \mathbf{D}_{tu} & \mathbf{K}_{tt} + s_0\mathbf{D}_{tt} \end{bmatrix} . \tag{3.116}$$

The transfer function for this system is,

$$H(s) = \widehat{\mathbf{l}}^* \left(\mathbf{I} + (s - s_0)\overline{\mathbf{G}}_{s_0}{}^{-1}\mathbf{C} \right)^{-1} \overline{\mathbf{G}}_{s_0}{}^{-1}\widehat{\mathbf{b}}. \tag{3.117}$$

Clearly this expression differs from Equation (3.111), and the Standard Krylov subspaces that are generated by the two different expressions are also different, making it again difficult to apply Theorem 3.3 in [124].

3.4 Conclusions

In this section, we presented a finite element based numerical method to efficiently evaluate the transfer function for the coupled equations of linear thermoelasticity. From this transfer function,

the quality factor (Q) of MEMS resonators including the effect of thermoelastic damping (TED) can be computed. Finite element analysis allows Q evaluation of devices irrespective of fabrication material or geometry, freeing designers from the contemporary beam like structures.

For efficiency, a Krylov subspace approach is taken to project the large finite element discretized system of equations onto a smaller subspace of solutions. The vectors spanning this smaller subspace of solutions is efficiently generated by the Second-Order Arnoldi (SOAR) method. The reduced order model (ROM) produced by this projection, is substantially smaller than the original system enabling fast transfer function evaluation. The characteristic structure inherited from the governing continuous partial differential equations is taken advantage of to construct a highly accurate structure preserving ROM for modeling both the TED case and TED/anchor loss case. The increased accuracy in the reduced model is proven through a new moment matching theorem based on second-order Krylov subspaces. For symmetric mechanical forcing and sensing, the theorem is able to prove a doubling of matched moments in the transfer function with the same number of second-order Krylov subspace iterations.

The theoretical claims made are verified by two numerical examples presented in Section 5.5.

Chapter 4

Electromechanically coupled systems

4.1 Introduction

High-frequency MEMS resonators used as RF filters or oscillators are mainly mechanical in their vibrational characteristics, but couple with the electrical domain through their form of actuation. At the micron-scale, electrostatic forces as well as piezoelectric induced forces become prominent due to scaling effects [181]. Such form of forces also easily integrate with the electronic character of integrated circuitry. In order to model the full behavior of the resonator, one must treat an electromechanically coupled system.

Electrostatic forces are applied to the resonator through voltage potential gaps between the resonator and electrode, which are filled with some sort of dielectric material. Piezoelectric forces are applied by induced electrical fields within the resonator. In both cases, a time-harmonic varying voltage applied at the resonance frequency of the device is capable of creating resonance in the device. This induced motion in the device is capacitively sensed in the form of a time-harmonic

current. In this sense, from the viewpoint of the electronic circuitry or an electrical engineer, the behavior of a mechanical resonator can be fully characterized by its voltage-current characteristics.

The approach of modeling the mechanical resonator as an electronic component is widely used in the design of electromechanical transducers as acoustical filter [111] and electrical signal processing filter applications [105]. Such a method is possible from the simple behavior of the mechanical resonator close to its fundamental frequencies. The complex motion of the resonator can be modeled as a single degree of freedom system oscillating in the fundamental mode corresponding to its fundamental frequency. The mechanical mass-spring-damper single degree of freedom system has a direct analogue in the electrical domain as a inductor-capacitor-resistor system, enabling a component by component equivalent replacement. This method of lumped modeling of a mechanical resonator is often referred to as the method of "equivalent circuits" [105, 177, 178], and the extracted parameters are called equivalent circuit parameters. Equivalent circuits allow designers to extract the necessary properties of the mechanical resonator required in designing the surrounding components without having to deal with the intricacies of the mechanical behavior. Models ranging from simple equivalent circuits capturing the behavior of one mode, to complex models capturing multimodes exist.

Equivalent circuit modeling has been introduced in the MEMS resonator field and has quickly gained popularity for its affinity with the electrical engineers designing the devices. The analogy is being pushed even further by replacing not only a single resonator but multiple resonators and their coupling elements by electrical equivalents [125]. In many papers one encounters formulas to evaluate equivalent circuit parameters, but in most cases they are either based on parallel plate assumptions [194, 39], derived for special geometries [189, 61, 169, 156]. As the geometry of the devices become more complex such assumptions may not apply, and numerical evaluation is necessary. In the area of piezoelectric bulk acoustic wave resonators, finite element methods have been applied to analyze their behavior [122, 133], and in the process equivalent circuit parameter expressions have been presented for this problem. Unfortunately this type of approach has not been extended into

the MEMS computational domain.

In this section, the idea of equivalent circuits is first introduced through a simple electromechanical parallel plate example. This example incorporates all the ideas behind lumped circuit modeling of mechanical resonators. This is followed by a direct extension of the idea to general nonlinear 3D electromechanical continuum systems through a variational approach. By evaluating the contributions of electrostatic and piezoelectric forces, the nonlinear system that one must solve is reduced to a linear system for the piezoelectric electromechanical system, and a combination of a linear and nonlinear system for the electrostatic electromechanical system. The section is closed with the details in the actual evaluation of the equivalent circuit parameters as well as extensions. Example applications of the developed theory are presented in Chapter 5.

4.2 Electromechanical 1D parallel plate capacitor

To introduce the idea of equivalent circuit parameters for fast simulation of mechanical resonators integrated with electrical systems, the simple 1D electromechanical parallel plate capacitor is presented.

Consider a parallel plate capacitor with one end fixed, and the other end attached to a mass-spring system; as is shown in Figure 4.1, m is the mass, k is the spring stiffness, b is the velocity dependent damping coefficient, g is the gap distance between the plates, g_0 is the initial gap distance at rest, $V_{\rm ac}$ is an AC voltage term, and $V_{\rm dc}$ is a DC voltage term. Gravitational forces are not included here.

The static energy of the system is defined in terms of the displacement u of the top plate and surface charge Q as,

$$W'(u,Q) = \frac{1}{2}ku^2 + \frac{1}{2}\frac{Q^2}{C(u)}, \tag{4.1}$$

$$C(u) := \frac{\epsilon_0 A}{g_0+u}, \tag{4.2}$$

where ϵ_0 is the permittivity of free space, A is the area of the plate, and $C(u)$ is the variable

capacitance. The force F and voltage V are obtained as the conjugate variables,

$$F = \frac{\partial W'}{\partial u} = ku - \frac{1}{2}\frac{\partial C}{\partial u}\frac{Q^2}{C^2(u)}, \tag{4.3}$$

$$V = \frac{\partial W'}{\partial Q} = \frac{Q}{C(u)}. \tag{4.4}$$

By application of a Legendre transform for the (Q, V) pair, an expression for the energy in terms of the control variables (u, V) can be obtained,

$$W(u,V) := \min_{Q}\{W'(u,Q) - QV\} \tag{4.5}$$

$$= \frac{1}{2}ku^2 - \frac{1}{2}C(u)V^2, \tag{4.6}$$

where,

$$F(u,V) = \frac{\partial W}{\partial u} \tag{4.7}$$

$$= ku - \frac{1}{2}\frac{\partial C}{\partial u}V^2, \tag{4.8}$$

$$Q(u,V) = -\frac{\partial W}{\partial V} \tag{4.9}$$

$$= C(u)V. \tag{4.10}$$

Under the assumption of electrostatics, which is valid when the electrical field time scale is small compared to the mechanical field time scale, the kinetic energy of the system only has a mechanical contribution,

$$T(u,V) := \frac{1}{2}m\dot{u}^2. \tag{4.11}$$

Given an external load term,

$$W_{\text{ext}}(u,V) := Q_{\text{ext}}V, \tag{4.12}$$

the total static energy of the system Π is defined as,

$$\Pi(u,V) := W(u,V) + W_{\text{ext}}(u,V). \tag{4.13}$$

The velocity dependent damper leads to the dissipation function,

$$\mathcal{F}(\dot{u},\dot{Q}) := \frac{1}{2}b\dot{u}^2. \tag{4.14}$$

From the Lagrangian equations of motion [79],

$$L := T - \Pi, \tag{4.15}$$

$$\frac{d}{dt}\left[\frac{\partial L}{\partial(\dot{u},\dot{Q})}\right] - \left[\frac{\partial L}{\partial(u,Q)}\right] + \frac{\partial \mathcal{F}}{\partial(\dot{u},\dot{Q})} = 0, \tag{4.16}$$

one obtains the governing equations of the system,

$$m\ddot{u} + b\dot{u} + F(u,V) = 0, \tag{4.17}$$

$$-Q(u,V) = -Q_{\text{ext}}. \tag{4.18}$$

$$\tag{4.19}$$

Under the application of a time-varying voltage of the form,

$$V(t) := V_{\text{dc}} + V_{\text{ac}}(t), \tag{4.20}$$

with $||V_{\text{ac}}|| << ||V_{\text{dc}}||$, one can assume a decomposition of the displacement field into two contributions,

$$u(t) := u_{\text{dc}} + u_{\text{ac}}(t), \tag{4.21}$$

a static term u_{dc} and a time-varying term u_{ac} with $||u_{\text{ac}}|| << ||u_{\text{dc}}||$. Linearization of Equations (4.17) and (4.18) at the point $(u_{\text{dc}}, V_{\text{dc}})$ yields,

$$m\ddot{u}_{\text{ac}} + b\dot{u}_{\text{ac}} + F(u_{\text{dc}},V_{\text{dc}}) + \frac{\partial F}{\partial u}\Big|_{\text{dc}} u_{\text{ac}} + \frac{\partial F}{\partial V}\Big|_{\text{dc}} V_{\text{ac}} = 0 \tag{4.22}$$

$$-Q(u_{\text{dc}},V_{\text{dc}}) - \frac{\partial Q}{\partial u}\Big|_{\text{dc}} u_{\text{ac}} - \frac{\partial Q}{\partial V}\Big|_{\text{dc}} V_{\text{ac}} = -Q_{\text{ext}}. \tag{4.23}$$

By defining u_{dc} as the displacement for the static voltage V_{dc}, $(u_{\text{dc}}, V_{\text{dc}})$ solves the equation,

$$F(u_{\text{dc}},V_{\text{dc}}) = 0, \tag{4.24}$$

which implies static DC equilibrium for the top plate. Defining,

$$Q_{\text{dc}} := Q(u_{\text{dc}},V_{\text{dc}}), \tag{4.25}$$

$$Q_{\text{ac}} := Q_{\text{ext}} - Q_{\text{dc}}, \tag{4.26}$$

Equations (4.22) and (4.23) can be rewritten as,

$$\mathbf{M}\begin{pmatrix} \ddot{u}_{\rm ac} \\ \ddot{V}_{\rm ac} \end{pmatrix} + \mathbf{B}\begin{pmatrix} \dot{u}_{\rm ac} \\ \dot{V}_{\rm ac} \end{pmatrix} + \mathbf{K}_{\rm dc}\begin{pmatrix} u_{\rm ac} \\ V_{\rm ac} \end{pmatrix} = \begin{pmatrix} 0 \\ -Q_{\rm ac} \end{pmatrix}, \tag{4.27}$$

where,

$$\mathbf{M} := \begin{bmatrix} m & 0 \\ 0 & 0 \end{bmatrix}, \tag{4.28}$$

$$\mathbf{B} := \begin{bmatrix} b & 0 \\ 0 & 0 \end{bmatrix}, \tag{4.29}$$

$$\mathbf{K}_{\rm dc} := \begin{bmatrix} \frac{\partial^2 W}{\partial u^2} & \frac{\partial^2 W}{\partial u \partial V} \\ \frac{\partial^2 W}{\partial V \partial u} & \frac{\partial^2 W}{\partial V^2} \end{bmatrix}_{\rm dc} = \begin{bmatrix} k - \frac{1}{2}\frac{\partial^2 C}{\partial u^2}V_{\rm dc}^2 & -\frac{\partial C}{\partial u}V_{\rm dc} \\ -\frac{\partial C}{\partial u}V_{\rm dc} & -C_{\rm dc} \end{bmatrix}. \tag{4.30}$$

In the expression above, one observes the negative electrical stiffness contribution leading to softening,

$$k_{\rm dc} := k - \frac{1}{2}\frac{\partial^2 C}{\partial u^2}\Big|_{\rm dc} V_{\rm dc}^2. \tag{4.31}$$

The coupling between the mechanical and electrical domains is represented by the symmetric electromechanical coupling term,

$$\eta_{\rm dc} := -\frac{\partial C}{\partial u}\Big|_{\rm dc} V_{\rm dc}\ . \tag{4.32}$$

Under time-harmonic assumptions,

$$V_{\rm ac} := \hat{V}_{\rm ac}\exp(i\omega t), \tag{4.33}$$

$$u_{\rm ac} := \hat{u}_{\rm ac}\exp(i\omega t), \tag{4.34}$$

$$Q_{\rm ac} := \hat{Q}_{\rm ac}\exp(i\omega t), \tag{4.35}$$

$$I_{\rm ac} := \frac{dQ_{\rm ac}}{dt}, \tag{4.36}$$

$$= \hat{I}_{\rm ac}\exp(i\omega t), \tag{4.37}$$

$$\hat{I}_{\rm ac} := i\omega\hat{Q}_{\rm ac}, \tag{4.38}$$

one obtains the admittance $Y(\omega)$,

$$Y(\omega) := \frac{\hat{I}_{\rm ac}}{\hat{V}_{\rm ac}} \tag{4.39}$$

$$= \left[i\omega C_{\rm dc} + \frac{i\omega\eta_{\rm dc}^2}{k_{\rm dc} - m\omega^2 + ib\omega} \right] \tag{4.40}$$

$$= \left[i\omega C_{\rm dc} + \frac{1}{\frac{1}{i\omega}\frac{k_{\rm dc}}{\eta_{\rm dc}^2} + i\omega\frac{m}{\eta_{\rm dc}^2} + \frac{b}{\eta_{\rm dc}^2}} \right]. \tag{4.41}$$

An analogy of the form of this equation to the admittance of the LRCC circuit shown in Figure 4.2, leads to the representation of the mechanical resonator as an equivalent circuit with parameters [162],

$$L_{\rm eq} := \frac{\eta_{\rm dc}^2}{k_{\rm dc}}, \tag{4.42}$$

$$R_{\rm eq} := \frac{b}{\eta_{\rm dc}^2}, \tag{4.43}$$

$$C_{\rm eq} := \frac{m}{\eta_{\rm dc}^2}, \tag{4.44}$$

$$C_{0,\rm eq} := C_{\rm dc}. \tag{4.45}$$

These terms $L_{\rm eq}$, $R_{\rm eq}$, and $C_{\rm eq}$, are called the motional inductance, the motional resistance, and the motional capacitance. The motional resistance rewritten in terms of the quality factor Q ($:= \frac{\sqrt{mk}}{b}$) is,

$$R_{\rm eq} = \frac{\sqrt{mk}}{\eta_{\rm dc}^2 Q}. \tag{4.46}$$

The motional resistance is the resistance the circuit experiences at the resonance of the mechanical resonator, which defines the amount of energy lost in the system. For efficient low power consuming resonators, a small motional resistance is ideal. One way of lowering the motional resistance is maximizing the quality of resonance, Q, of the system. Thus, not only does Q minimization lead to better signal processing capabilities, but it also leads to lower energy consumption. Such observations have pushed the need for CAD tools to simulate Q in mechanical resonators. The other way of lowering the motional resistance is maximizing the electromechanical coupling coefficient, $\eta_{\rm dc}$. For the parallel plate electromechanical capacitor, the expression for the electromechanical coupling

coefficient yields,

$$\eta_{\text{eq}} = \frac{\epsilon_0 A V_{\text{dc}}}{(g_0 + u_{\text{dc}})^2} \approx \frac{\epsilon_0 A V_{\text{dc}}}{g_0^2}, \quad (\text{when } |u_{\text{dc}}| << |g_0|). \tag{4.47}$$

Minimizing the gap size g_0 leads to large reduction in the motional resistance due to the 4th power. This has lead to the fabrication of resonators with nano-airgaps between the resonator and transduction electrodes [173]. But this has its limitations due to the fabrication process and existence of a breakdown voltage [162]. The DC bias voltage V_{dc} can also be increased but again this also has a limitation due to breakdown. Increase in surface area is another method which has been implemented in the design of ring type resonators, for which A increases with the ring diameter [39]. For our parallel plate example, vacuum has been assumed between the plates. There is however no restriction on the type of material inserted between the plates, other than being insulators. A relatively new and interesting approach is to insert a high relative permittivity dielectric in the place of vacuum [32, 49]. Bulk materials such as Hafnium Oxide and Silicon Nitride, and fluidic materials such as water have been attempted [48]. The emergence of such designs strengthen the need for tools to evaluate the behavior of mechanical resonators, since the insertion of these materials can greatly affect the mechanical mode of vibration. The use of these layers of dielectric as actuation mechanisms in resonators to excite motion is called internal transduction, compared to the term external transduction used for the standard air gap type of actuation.

Remark: In the computation of transfer functions for evaluation of Q in mechanical systems, the force-displacement relation is often computed. For applications as transducers, it should be stressed, however, that the force-velocity transfer function is of importance. This is because the velocity is what induces a change in the current. The becomes obvious by rewriting the expression for current

in Equation (4.39) in terms of the voltage and velocity $\hat{v}_{\mathrm{ac}}$,

$$\begin{aligned}
\hat{I}_{\mathrm{ac}} &= i\omega C_{\mathrm{dc}}\hat{V}_{\mathrm{dc}} + \eta_{\mathrm{dc}} i\omega \hat{u}_{\mathrm{ac}} \\
&= i\omega C_{\mathrm{dc}}\hat{V}_{\mathrm{dc}} + \eta_{\mathrm{dc}}\hat{v}_{\mathrm{ac}}, && (4.48) \\
\hat{v}_{\mathrm{ac}} &:= i\omega \hat{u}_{\mathrm{ac}} && (4.49) \\
&= \frac{i\omega}{k_{\mathrm{dc}} - m\omega^2 + ib\omega}\hat{F}_{\mathrm{ac}}, && (4.50) \\
\hat{F}_{\mathrm{ac}} &:= -\eta_{\mathrm{dc}}\hat{V}_{\mathrm{ac}} && (4.51)
\end{aligned}$$

where $\hat{F}_{\mathrm{ac}}$ is the electrostatic force arising from the electromechanical coupling. The pairing of velocity and force (v, F) arises from its power duality between the current and voltage pair (I, V).

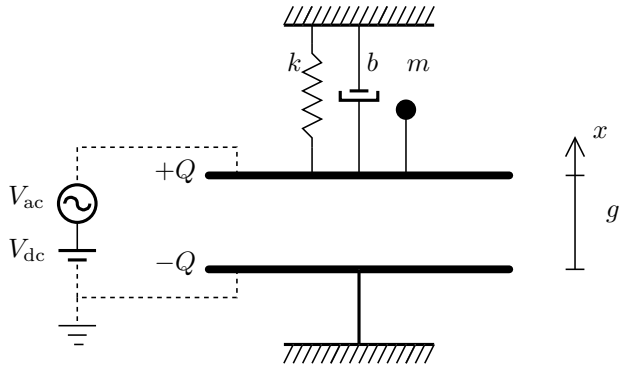


Figure 4.1: Schematic of electromechanical 1D parallel plate capacitor

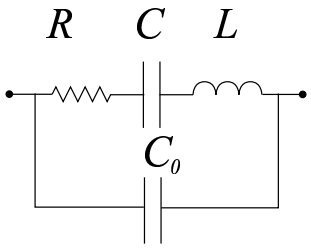


Figure 4.2: Schematic of the equivalent LRCC circuit

4.3 A variational approach to the electromechanical problem

In the previous section, equivalent circuit parameters were obtained for an idealized 1D electromechanical parallel plate configuration. In this section the method of evaluating equivalent circuit parameters for general 3D nonlinear continuum electromechanical systems is presented. Both electrostatic and piezoelectric contributions are included in the formulation, The derivation of the equations follows a similar approach to the 1D example, where the static energy for the system is defined, and derivatives of this energy give the equations of motion. This energy approach leads to a variational formulation, easily implementable by a finite element method for numerical evaluation.

For a general nonlinear electromechanical system, one can derive the following expression for the total static energy on the undeformed domain Ω with boundary Γ,

$$\Pi_{\text{total}}(\boldsymbol{\varphi}, \phi) := \Pi(\boldsymbol{\varphi}, \phi) + \Pi_{\text{ext}}(\boldsymbol{\varphi}, \phi), \tag{4.52}$$

$$\Pi(\boldsymbol{\varphi}, \phi) := \frac{1}{2} \int_{\Omega} W(\mathbf{F}(\boldsymbol{\varphi}), \mathbf{E}(\phi))\, d\Omega, \tag{4.53}$$

$$\Pi_{\text{ext}}(\boldsymbol{\varphi}, \phi) := \int_{\boldsymbol{\varphi}(\Gamma_{\mathbf{q}})} \bar{\sigma}_q \phi\, d\Gamma, \tag{4.54}$$

with,

$$\Gamma := \Gamma_{\boldsymbol{\varphi}} \sqcup \Gamma_{\mathbf{t}}, \tag{4.55}$$

$$:= \Gamma_{\phi} \sqcup \Gamma_{\mathbf{q}}, \tag{4.56}$$

$$\mathbf{x}(\mathbf{X}, t) := \boldsymbol{\varphi}(\mathbf{X}, t), \tag{4.57}$$

$$\mathbf{F} := \frac{\partial \boldsymbol{\varphi}}{\partial \mathbf{X}}, \tag{4.58}$$

$$\mathbf{E} := -\frac{\partial \phi}{\partial \mathbf{x}} \tag{4.59}$$

$$\mathbf{t} := \boldsymbol{\sigma} \cdot \mathbf{n} = 0, \qquad (\mathbf{x} \in \boldsymbol{\varphi}(\Gamma_{\mathbf{t}})) \tag{4.60}$$

$$\boldsymbol{\varphi} := \bar{\boldsymbol{\varphi}}, \qquad (\mathbf{X} \in \Gamma_{\boldsymbol{\varphi}}) \tag{4.61}$$

$$\sigma_q := \mathbf{D} \cdot \mathbf{n} = \bar{\sigma}_q, \quad (\mathbf{x} \in \boldsymbol{\varphi}(\Gamma_{\mathbf{q}})), \tag{4.62}$$

$$\phi := \bar{\phi}, \qquad (\mathbf{X} \in \Gamma_{\phi}). \tag{4.63}$$

Here, $\Gamma_{\boldsymbol{\varphi}}$ is the part of the boundary with displacement defined, $\Gamma_{\mathbf{t}}$ is the part of the boundary with traction defined, Γ_{ϕ} is the part of the boundary with potential defined, $\Gamma_{\mathbf{q}}$ is the part of the boundary with charge defined, $\boldsymbol{\varphi}$ is the deformation mapping, $\boldsymbol{\sigma}$ is the Cauchy stress tensor, $\mathbf{F}$ is the deformation gradient, $\mathbf{D}$ is the electric displacement vector, $\mathbf{E}$ is the electric field vector, ϕ is the potential, $\mathbf{t}$ is the traction vector, σ_q is the surface charge density, and $\mathbf{n}$ is the surface normal vector. The body force $\mathbf{b}$ and body charge density ρ_q are both assumed zero for simplicity.

An example configuration of an electromechanical problem is shown in Figure 4.3. An elastic dielectric material is sandwiched between two electrodes. The bottom plate is fixed enforcing a mechanical fixed displacement boundary condition $\Gamma_{\boldsymbol{\varphi}}$ at the bottom. The rest of the boundary is free to move; a mechanical free traction boundary condition $\Gamma_{\mathbf{t}}$. The bottom plate is grounded to 0 potential, enforcing an electrical fixed potential boundary condition Γ_{ϕ} at the bottom. The top plate can also be treated as an electrical fixed boundary condition since the potential is fixed to the sum of the DC bias voltage and time-varying AC voltage, but here it is interpreted as an electrical forced charge boundary condition $\Gamma_{q,1}$ to retain the charge variables on the top surface. This is formulation is also natural since our objective is to obtain an admittance like quantity, obtained from a forced current. The left and right sides of the material has zero surface charge; an electrical free charge boundary condition $\Gamma_{q,0}$. The total electrical forced charge boundary is $\Gamma_q = \Gamma_{q,0} \sqcup \Gamma_{q,1}$.

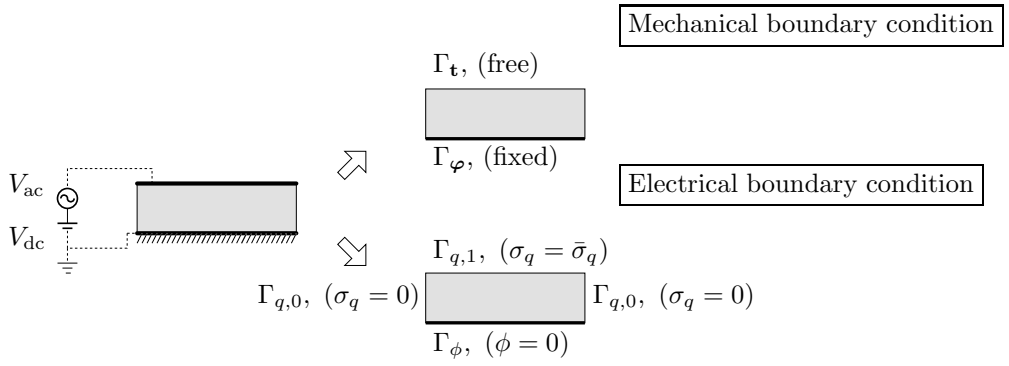


Figure 4.3: Example configuration of an electromechanical problem: Elastic dielectric sandwiched between two electrodes

The boundary conditions enforce the functions $\boldsymbol{\varphi}$ and ϕ to belong to the following space of functions,

$$\boldsymbol{\varphi} \in \mathcal{S}_{\boldsymbol{\varphi}} \quad := \quad \left\{ \boldsymbol{\varphi} : H^1(\Omega; \mathbb{C}^3) : \boldsymbol{\varphi} = \bar{\boldsymbol{\varphi}} \text{ on } \Gamma_{\boldsymbol{\varphi}} \right\}, \tag{4.64}$$

$$\phi \in \mathcal{S}_{\phi} \quad := \quad \left\{ \phi : H^1(\Omega; \mathbb{C}) : \phi = \bar{\phi} \text{ on } \Gamma_{\phi} \right\}. \tag{4.65}$$

The constitutive equations are defined by the stored energy function $W(\mathbf{F}, \mathbf{E})$ [113]. For a frame invariant energy function, one requires,

$$W(\mathbf{QF}, \mathbf{QE}) \quad = \quad W(\mathbf{F}, \mathbf{E}), \qquad \text{for any } \mathbf{Q} \in \mathrm{SO}(3). \tag{4.66}$$

For this to hold, the energy function must be of the form,

$$W(\mathbf{F}, \mathbf{E}) \quad = \quad \Psi(\mathbf{C}, \mathbf{E}_r), \tag{4.67}$$

$$\mathbf{C} \quad := \quad \mathbf{F}^T \mathbf{F}, \tag{4.68}$$

$$\mathbf{E}_r \quad := \quad \mathbf{F}^T \mathbf{E}, \tag{4.69}$$

where $\mathbf{C}$ is the right Cauchy-Green tensor and $\mathbf{E}_r$ is the electrical field in the reference configuration. We also define the electrical displacement in the reference configuration as,

$$\mathbf{D}_r \quad := \quad J \mathbf{F}^{-1} \mathbf{D}, \tag{4.70}$$

where $J := \det(\mathbf{F}) = \sqrt{\det(\mathbf{C})}$ is the Jacobian of the deformation gradient. The constitutive equations become,

$$\boldsymbol{\sigma} \quad = \quad 2 \frac{1}{J} \mathbf{F} \frac{\partial \Psi}{\partial \mathbf{C}} \mathbf{F}^T, \tag{4.71}$$

$$\mathbf{D} \quad = \quad \frac{1}{J} \mathbf{F} \frac{\partial \Psi}{\partial \mathbf{E}_r}. \tag{4.72}$$

Under the assumption of electrostatics, which is valid when the electrical field time scale is small compared to the mechanical field time scale, the kinetic energy of the system only has a mechanical contribution,

$$T(\boldsymbol{\varphi}, \phi) \quad := \quad \int_{\Omega} \frac{1}{2} \rho \dot{\boldsymbol{\varphi}}^2 \, d\Omega, \tag{4.73}$$

where ρ is the material density. From the Lagrangian equations of motion, one obtains,

$$L := T - \Pi_{\text{total}} \tag{4.74}$$

$$\frac{d}{dt}\left[\frac{\partial L}{\partial(\dot{\boldsymbol{\varphi}},\dot{\phi})}\right] = \frac{\partial L}{\partial(\boldsymbol{\varphi},\phi)} \ . \tag{4.75}$$

Given a time independent variation $[\delta\boldsymbol{\varphi}, \delta\phi]$ belonging to the space of functions,

$$\delta\boldsymbol{\varphi} \in \mathcal{V}_{\boldsymbol{\varphi}} := \left\{\delta\boldsymbol{\varphi} : H^1(\Omega;\mathbb{C}^3) : \delta\boldsymbol{\varphi} = \mathbf{0} \text{ on } \Gamma_{\boldsymbol{\varphi}}\right\}, \tag{4.76}$$

$$\delta\phi \in \mathcal{V}_{\phi} := \left\{\delta\phi : H^1(\Omega : \mathbb{C}) : \delta\phi = 0 \text{ on } \Gamma_{\boldsymbol{\varphi}}\right\}, \tag{4.77}$$

the equations of motions yield [138],

$$\int_\Omega \rho\ddot{\boldsymbol{\varphi}} \cdot \delta\boldsymbol{\varphi} d\Omega + \Pi_{\boldsymbol{\varphi}}(\boldsymbol{\varphi},\phi)[\delta\boldsymbol{\varphi}] = 0, \tag{4.78}$$

$$\Pi_{\phi}(\boldsymbol{\varphi},\phi)[\delta\phi] = -\int_{\boldsymbol{\varphi}(\Gamma_{\mathbf{q}})} \sigma_{\text{q}}\delta\phi \ d\Gamma \ . \tag{4.79}$$

The notation introduced here, $\Pi_{\boldsymbol{\varphi}}(\boldsymbol{\varphi},\phi)[\delta\boldsymbol{\varphi}]$, denotes the first variation of Π with respect to $\boldsymbol{\varphi}$ which is a linear form. The linear form has the argument $\delta\boldsymbol{\varphi}$. Let us a assume a decomposition of the solution $(\boldsymbol{\varphi},\phi)$,

$$\boldsymbol{\varphi}(\mathbf{X},t) := \boldsymbol{\varphi}_{\text{dc}}(\mathbf{X}) + \boldsymbol{\varphi}_{\text{ac}}(\mathbf{X},t), \tag{4.80}$$

$$\phi(\mathbf{X},t) := \phi_{\text{dc}}(\mathbf{X}) + \phi_{\text{ac}}(\mathbf{X},t), \tag{4.81}$$

where $\boldsymbol{\varphi}_{\text{ac}}, \boldsymbol{\varphi}_{\text{dc}} \in \mathcal{V}_{\boldsymbol{\varphi}}$ and $\phi_{\text{ac}}, \phi_{\text{dc}} \in \mathcal{V}_{\phi}$. Linearization of Equations (4.78) and (4.79) at $(\boldsymbol{\varphi}_{\text{dc}}, \phi_{\text{dc}})$ yields,

$$\int_\Omega \rho\ddot{\boldsymbol{\varphi}}_{\text{ac}} \cdot \delta\boldsymbol{\varphi} d\Omega + \Pi_{\boldsymbol{\varphi}}(\boldsymbol{\varphi}_{\text{dc}},\phi_{\text{dc}})[\delta\boldsymbol{\varphi}] +$$

$$\Pi_{\boldsymbol{\varphi}\boldsymbol{\varphi}}(\boldsymbol{\varphi}_{\text{dc}},\phi_{\text{dc}})[\delta\boldsymbol{\varphi},\boldsymbol{\varphi}_{ac}] + \Pi_{\boldsymbol{\varphi}\phi}(\boldsymbol{\varphi}_{\text{dc}},\phi_{\text{dc}})[\delta\boldsymbol{\varphi},\phi_{ac}] = 0, \tag{4.82}$$

$$\Pi_{\phi}(\boldsymbol{\varphi}_{\text{dc}},\phi_{\text{dc}})[\delta\phi] +$$

$$\Pi_{\phi\boldsymbol{\varphi}}(\boldsymbol{\varphi}_{\text{dc}},\phi_{\text{dc}})[\delta\phi,\boldsymbol{\varphi}_{ac}] + \Pi_{\phi\phi}(\boldsymbol{\varphi}_{\text{dc}},\phi_{\text{dc}})[\delta\phi,\phi_{ac}] = -\int_{\boldsymbol{\varphi}(\Gamma_q)} \sigma_{\text{q}}\delta\phi \ d\Gamma \ . \tag{4.83}$$

Let the pair $(\boldsymbol{\varphi}_{\text{dc}}, \phi_{\text{dc}})$ be the static solution to the following boundary value problem,

- Find the pair $(\boldsymbol{\varphi}, \phi)$ which is a stationary point of,

$$\Pi(\boldsymbol{\varphi}, \phi) + \Pi_{\text{ext,dc}}(\boldsymbol{\varphi}, \phi), \tag{4.84}$$

with,

$$\Pi_{\text{ext,dc}}(\boldsymbol{\varphi}, \phi) \quad := \quad \int_{\boldsymbol{\varphi}(\Gamma_{\mathbf{q}} \setminus \Gamma_{\text{dc}})} \bar{\sigma}_q \phi \; d\Gamma, \tag{4.85}$$

subject to the boundary conditions,

$$\mathbf{t} := \boldsymbol{\sigma} \cdot \mathbf{n} = 0, \qquad (\mathbf{x} \in \boldsymbol{\varphi}(\Gamma_{\mathbf{t}})), \tag{4.86}$$

$$\boldsymbol{\varphi} := \bar{\boldsymbol{\varphi}}, \qquad (\mathbf{x} \in \boldsymbol{\varphi}(\Gamma_{\boldsymbol{\varphi}})), \tag{4.87}$$

$$\phi := \begin{cases} \bar{\phi} & (\mathbf{x} \in \boldsymbol{\varphi}(\Gamma_{\phi})), \\ \bar{\phi}_{\text{dc}}, & (\mathbf{x} \in \boldsymbol{\varphi}(\Gamma_{\text{ac}})), \end{cases} \tag{4.88}$$

$$\sigma_q := \mathbf{D} \cdot \mathbf{n} = \bar{\sigma}_q, \quad (\mathbf{x} \in \boldsymbol{\varphi}(\Gamma_q) \setminus \boldsymbol{\varphi}(\Gamma_{\text{dc}})). \tag{4.89}$$

The part of the boundary Γ_{ac}, where an additional electrical fixed boundary condition of $\bar{\phi}_{\text{dc}}$ is applied, is a subset of Γ_q. The problem is restated in the weak form as, find the pair$(\boldsymbol{\varphi}, \phi)$,

$$\boldsymbol{\varphi} \quad \in \quad \mathcal{S}_{\boldsymbol{\varphi}}, \tag{4.90}$$

$$\phi \quad \in \quad \mathcal{S}_{\phi,dc} := \left\{ \phi : H^1(\Omega; \mathbb{C}) : \phi = \bar{\phi} \text{ on } \Gamma_{\phi}, \phi = \bar{\phi}_{\text{dc}} \text{ on } \Gamma_{\text{dc}} \right\} \subset \mathcal{S}_{\phi}, \tag{4.91}$$

such that for any pair $(\delta\boldsymbol{\varphi}, \delta\phi_{\text{dc}})$,

$$\delta\boldsymbol{\varphi} \quad \in \quad \mathcal{V}_{\boldsymbol{\varphi}}, \tag{4.92}$$

$$\delta\phi_{\text{dc}} \quad \in \quad \mathcal{V}_{\phi,dc} := \left\{ \delta\phi : H^1(\Omega; \mathbb{C}) : \delta\phi = \mathbf{0} \text{ on } \Gamma_{\phi} \sqcup \Gamma_{\text{dc}} \right\} \subset \mathcal{V}_{\phi}, \tag{4.93}$$

one has,

$$\Pi_{\boldsymbol{\varphi}}(\boldsymbol{\varphi}_{\text{dc}}, \phi_{\text{dc}})[\delta\boldsymbol{\varphi}] \quad = \quad 0, \tag{4.94}$$

$$\Pi_{\phi}(\boldsymbol{\varphi}_{\text{dc}}, \phi_{\text{dc}})[\delta\phi_{\text{dc}}] \quad = \quad -\int_{\boldsymbol{\varphi}(\Gamma_{\mathbf{q}} \setminus \Gamma_{\text{dc}})} \bar{\sigma}_q \delta\phi_{\text{dc}} \; d\Gamma. \tag{4.95}$$

The difference between this and the original problem is the addition of an electrical fixed potential boundary condition of $\bar{\phi}_{\rm dc}$ on the boundary $\Gamma_{\rm ac}$. $\Gamma_{\rm ac}$ must be a subset of original electrical charge boundary Γ_q. By this inclusion of the boundaries, the solution subspace for this problem $\mathcal{S}_{\phi,\rm dc}$ is a subspace of the original solution subspace $\mathcal{S}_\phi$. The subspace of variations for this problem is also a subspace of the original subspace of variations, $\mathcal{V}_{\phi,\rm dc} \subset \mathcal{V}_\phi$. This problem is essentially solving for the static solution under a given DC bias voltage, similar to the process in Section 4.2. To further clarify the situation we again refer to the schematic of Figure 4.3 of the elastic dielectric sandwiched between two electrodes. For this configuration, the $\Gamma_{\rm ac}$ is chosen as $\Gamma_{q,1}$.

The solution to this boundary value problem $(\boldsymbol{\varphi}_{\rm dc}, \phi_{\rm dc})$ satisfies the relation,

$$\Pi_{\boldsymbol{\varphi}}(\boldsymbol{\varphi}_{\rm dc}, \phi_{\rm dc})[\delta\boldsymbol{\varphi}] = 0, \tag{4.96}$$

$$\Pi_{\phi}(\boldsymbol{\varphi}_{\rm dc}, \phi_{\rm dc})[\delta\phi] = -\int_{\boldsymbol{\varphi}(\Gamma_{\mathbf{q}}\setminus\Gamma_{\rm ac})} \bar{\sigma}_q \delta\phi \; d\Gamma - \int_{\boldsymbol{\varphi}_{\rm dc}(\Gamma_{\rm ac})} \sigma_{\rm q,dc}\delta\phi d\Gamma, \tag{4.97}$$

$$\sigma_{\rm q,dc} := \mathbf{D}(\boldsymbol{\varphi}_{\rm dc}, \phi_{\rm dc}) \cdot \mathbf{n}, \quad (\mathbf{x} \in \boldsymbol{\varphi}_{\rm dc}(\Gamma_{\rm ac})). \tag{4.98}$$

with $\delta\boldsymbol{\varphi} \in \mathcal{V}_{\boldsymbol{\varphi}}$ and $\delta\phi \in \mathcal{V}_\phi$. Compared to Equation (4.95), Equation (4.97) has a non-zero term on the right hand side. This term arises from the difference between the subspace of variations $\mathcal{V}_\phi$ and $\mathcal{V}_{\phi,\rm dc}$. The variations $\delta\phi_{\rm dc} \in \mathcal{V}_{\phi,\rm dc}$ are enforced to be zero on $\Gamma_{\rm ac}$ as opposed to $\delta\phi \in \mathcal{V}_\phi$ for which they are not.

Moving back to Equations (4.82), (4.83), let the boundary charge on $\Gamma_{\rm ac}$ be of the form,

$$\sigma_{\rm q} := \sigma_{\rm q,dc} + \sigma_{\rm q,ac}(t), \tag{4.99}$$

with $||\sigma_{\rm q,ac}|| << ||\sigma_{\rm q,dc}||$ and only $\sigma_{\rm q,ac}$ is time dependent. Under an additional assumption of time-harmonic forcing,

$$\boldsymbol{\varphi}_{\rm ac} = \hat{\boldsymbol{\varphi}}_{\rm ac} \exp(i\omega t), \tag{4.100}$$

$$\phi_{\rm ac} = \hat{\phi}_{\rm ac} \exp(i\omega t), \tag{4.101}$$

$$\sigma_{\rm q,ac} = \hat{\sigma}_{\rm q,ac} \exp(i\omega t), \tag{4.102}$$

Equations (4.82) and (4.83) can be written as,

$$-\omega^2 \int_\Omega \rho \hat{\boldsymbol{\varphi}}_{\rm ac} \cdot \delta\boldsymbol{\varphi} d\Omega +$$
$$\Pi_{\boldsymbol{\varphi}\boldsymbol{\varphi}}(\boldsymbol{\varphi}_{\rm dc}, \phi_{\rm dc})[\delta\boldsymbol{\varphi}, \hat{\boldsymbol{\varphi}}_{ac}] + \Pi_{\boldsymbol{\varphi}\phi}(\boldsymbol{\varphi}_{\rm dc}, \phi_{\rm dc})[\delta\boldsymbol{\varphi}, \hat{\phi}_{ac}] = 0, \tag{4.103}$$

$$\Pi_{\phi\boldsymbol{\varphi}}(\boldsymbol{\varphi}_{\rm dc}, \phi_{\rm dc})[\delta\phi, \hat{\boldsymbol{\varphi}}_{ac}] + \Pi_{\phi\phi}(\boldsymbol{\varphi}_{\rm dc}, \phi_{\rm dc})[\delta\phi, \hat{\phi}_{ac}] = -\int_{\boldsymbol{\varphi}_{\rm dc}(\Gamma_{\rm ac})} \hat{\sigma}_{\rm q,ac}\delta\phi \, d\Gamma \tag{4.104}$$

where the first variations have dropped out. When the forcing frequency ω is close to the eigenfrequency, $\omega_{\rm eig}$, of the system with eigenmode $(\hat{\boldsymbol{\varphi}}_{\rm ac,eig}, \hat{\phi}_{\rm ac,eig})$, the displacement and potential can be approximated as,

$$\hat{\boldsymbol{\varphi}}_{\rm ac} = \hat{u}_{\rm ac}\hat{\boldsymbol{\varphi}}_{\rm ac,eig}, \tag{4.105}$$

$$\hat{\phi}_{\rm ac} = \hat{V}_{\rm ac}\hat{\phi}_{\rm ac,eig}, \tag{4.106}$$

where $u_{\rm ac}$ and $V_{\rm ac}$ are generalized degrees of freedom. Under a Galerkin projection with,

$$\delta\boldsymbol{\varphi} = \hat{\boldsymbol{\varphi}}_{\rm ac,eig}, \tag{4.107}$$

$$\delta\phi = \hat{\phi}_{\rm ac,eig}, \tag{4.108}$$

as the test functions, Equations (4.103) and (4.104) become,

$$-\omega^2 \mathbf{M} \begin{pmatrix} \hat{u}_{\rm ac} \\ \hat{V}_{\rm ac} \end{pmatrix} + \mathbf{K} \begin{pmatrix} \hat{u}_{\rm ac} \\ \hat{V}_{\rm ac} \end{pmatrix} = \begin{pmatrix} 0 \\ -\hat{Q}_{\rm ac} \end{pmatrix} \tag{4.109}$$

with,

$$\mathbf{M} := \begin{bmatrix} m & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \int_\Omega \rho \hat{\boldsymbol{\varphi}}_{\mathrm{ac,eig}} \cdot \hat{\boldsymbol{\varphi}}_{\mathrm{ac,eig}} d\Omega & 0 \\ 0 & 0 \end{bmatrix}, \tag{4.110}$$

$$\mathbf{K} := \begin{bmatrix} k_{\mathrm{dc}} & \eta_{\mathrm{dc}} \\ \eta_{\mathrm{dc}} & -C_{\mathrm{dc}} \end{bmatrix} \tag{4.111}$$

$$= \begin{bmatrix} \Pi_{\boldsymbol{\varphi}\boldsymbol{\varphi}}(\boldsymbol{\varphi}_{\mathrm{dc}}, \phi_{\mathrm{dc}})[\hat{\boldsymbol{\varphi}}_{\mathrm{ac,eig}}, \hat{\boldsymbol{\varphi}}_{\mathrm{ac,eig}}] & \Pi_{\boldsymbol{\varphi}\phi}(\boldsymbol{\varphi}_{\mathrm{dc}}, \phi_{\mathrm{dc}})[\hat{\boldsymbol{\varphi}}_{\mathrm{ac,eig}}, \hat{\phi}_{\mathrm{ac,eig}}] \\ \Pi_{\phi\boldsymbol{\varphi}}(\boldsymbol{\varphi}_{\mathrm{dc}}, \phi_{\mathrm{dc}})[\hat{\phi}_{\mathrm{ac,eig}}, \hat{\boldsymbol{\varphi}}_{\mathrm{ac,eig}}] & \Pi_{\phi\phi}(\boldsymbol{\varphi}_{\mathrm{dc}}, \phi_{\mathrm{dc}})[\hat{\phi}_{\mathrm{ac,eig}}, \hat{\phi}_{\mathrm{ac,eig}}] \end{bmatrix}, \tag{4.112}$$

$$\hat{Q}_{\mathrm{ac}} := \int_{\boldsymbol{\varphi}_{\mathrm{dc}}(\Gamma_{\mathrm{ac}})} \hat{\sigma}_{\mathrm{q,ac}} \hat{\phi}_{\mathrm{ac,eig}} \, d\Gamma \, . \tag{4.113}$$

Let us consider the elastic dielectric presented in Figure 4.3. By normalizing the function $\hat{\phi}_{\mathrm{ac,eig}} = 1$ on $\Gamma_{q,1}$, $\hat{V}_{\mathrm{ac}}$ will denote the voltage potential on $\Gamma_{q,1}$, and $\hat{Q}_{\mathrm{ac}}$ is the corresponding time varying charge on $\Gamma_{q,1}$. $\hat{V}_{\mathrm{ac}}$ is equal to the time varying AC voltage applied from the source. From this, one obtains an approximation for the admittance of the system near the frequency of ω_{eig},

$$\hat{I}_{\mathrm{ac}} := i\omega \hat{Q}_{\mathrm{ac}},$$

$$= Y(\omega) \hat{V}_{\mathrm{ac}}, \tag{4.114}$$

$$Y(\omega) := \left[i\omega C_{\mathrm{dc}} + \frac{1}{\frac{1}{i\omega} \frac{k_{\mathrm{dc}}}{\eta_{\mathrm{dc}}^2} + i\omega \frac{m}{\eta_{\mathrm{dc}}^2}} \right] . \tag{4.115}$$

This expression shows direct correspondence of this system with the parallel plate example presented in Section 4.2. By solving the general problem with the proper energy expression for the electrostatic energy projected onto a 1D subspace, the exact same results to the parallel plate presented in Section 4.2 can be obtained. The advantage of this approach though, is the ability to analyze systems which are not necessarily ideal as this, in a rational manner.

4.4 Electrostatic vs. Piezoelectric forces

The variational approach presented in Section 4.3 to solve the general problem involved nonlinearity in the deformation $\boldsymbol{\varphi}$ and voltage potential field ϕ. These nonlinearities can be removed under proper assumptions. In this section, the forces arising from piezoelectric effects and electrostatic effects are compared. The results reveal that piezoelectric forces dominate electrostatic effects for normal situations. In combination with the justifiable assumption of linear elasticity for MEMS vibration problems this enables one to treat the piezoelectric problem linearly. For the electrostatic problem, it is mentioned that some nonlinearity must be retained for the electrostatic electromechanical coupling effect to exist. An argument is presented to justify the use of linear elasticity in combination with nonlinearity only for the electrostatic portion in solving the electrostatic electromechanically coupled problem.

A general stored energy function $W(\mathbf{F}, \mathbf{E}) = \Psi(\mathbf{C}, \mathbf{E}_r)$ for a nonlinear piezoelectric material can be written as,

$$\Psi(\mathbf{C}, \mathbf{E}_r) = \Psi_{\text{elastic}}(\mathbf{C}) + \Psi_{\text{electrostatic}}(\mathbf{C}, \mathbf{E}_r) + \Psi_{\text{piezo}}(\mathbf{C}, \mathbf{E}_r), \tag{4.116}$$

$$\Psi_{\text{electrostatic}}(\mathbf{C}, \mathbf{E}_r) := -\frac{1}{2}\epsilon_r\epsilon_0 J\mathbf{E}_r \cdot \mathbf{C}^{-1}\mathbf{E}_r, \tag{4.117}$$

$$\Psi_{\text{piezo}}(\mathbf{C}, \mathbf{E}_r) := -\frac{1}{2}\mathbf{E}_r \cdot \mathbf{e}_r : (\mathbf{C} - \mathbf{1}), \tag{4.118}$$

where ϵ_r is the relative permittivity and ϵ_0 is the permittivity of free space. $\mathbf{e}_r$ is the 3rd-rank tensor of piezoelectric stress coefficients in the referential configuration, with the operation,

$$\mathbf{E}_r \cdot \mathbf{e}_r : \mathbf{C} := (\mathbf{E}_r)_A(\mathbf{e}_r)_{ABC}(\mathbf{C})_{BC}. \tag{4.119}$$

From Equations (4.71) and (4.72), the expression for the stress and electric displacement become,

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}_{\text{elastic}} + \boldsymbol{\sigma}_{\text{electrostatic}} + \boldsymbol{\sigma}_{\text{piezo}}, \tag{4.120}$$

$$\boldsymbol{\sigma}_{\text{electrostatic}} := -\frac{1}{2}\epsilon_r\epsilon_0\mathbf{E} \cdot \mathbf{E} + \epsilon_r\epsilon_0\mathbf{E} \otimes \mathbf{E}, \tag{4.121}$$

$$\boldsymbol{\sigma}_{\text{piezo}} := -\mathbf{E} \cdot \mathbf{e}, \tag{4.122}$$

$$\mathbf{D} = \epsilon_r\epsilon_0\mathbf{E} + \frac{1}{2}\mathbf{e} : (\mathbf{1} - \mathbf{b}^{-1}), \tag{4.123}$$

where, $\mathbf{b} = \mathbf{F}\mathbf{F}^T$ is the left Cauchy-Green tensor and $\mathbf{e}$ is the piezoelectric stress coefficient in the spatial configuration,

$$(\mathbf{e})_{ijk} := \frac{1}{J}\mathbf{F}_{iA}\mathbf{F}_{jB}\mathbf{F}_{kC}(\mathbf{e}_r)_{ABC}. \tag{4.124}$$

In Equation (4.120), the electrostatic energy gives rise to two second order terms in $\mathbf{E}$, and the piezoelectric energy gives rise to a single first order term in $\mathbf{E}$. From this, it is clear that when $||\mathbf{E}|| << 1$, the piezoelectric effects dominates the electromechanical coupling. This is also clear by observation of the electromechanical coupling term in Equation (4.103),

$$\begin{aligned}&\Pi_{\boldsymbol{\varphi}\phi}(\boldsymbol{\varphi}_{\mathrm{dc}}, \phi_{\mathrm{dc}})[\delta\boldsymbol{\varphi}, \Delta\phi] \\ &= \int_{\boldsymbol{\varphi}_{\mathrm{dc}}(\Omega)} \nabla_{\mathbf{x}}\delta\boldsymbol{\varphi} : [-\epsilon_r\epsilon_0\mathbf{E}_{\mathrm{dc}} \cdot \Delta\mathbf{E} + \epsilon_r\epsilon_0\mathbf{E}_{\mathrm{dc}} \otimes \Delta\mathbf{E} + \epsilon_r\epsilon_0\Delta\mathbf{E} \otimes \mathbf{E}_{\mathrm{dc}} - \Delta\mathbf{E} \cdot \mathbf{e}]\, d\Omega,\end{aligned} \tag{4.125}$$

where $\Delta\mathbf{E} := -\nabla_{\mathbf{x}}\Delta\phi$ with $\Delta\phi \in \mathcal{V}_\phi$. With typical values for Aluminum Nitride [115],

$$\epsilon_r = 9, \tag{4.126}$$

$$\epsilon_0 = 8.85 \times 10^{-12}[\mathrm{C}^2/\mathrm{Nm}^2], \tag{4.127}$$

$$\mathbf{e}_{333} = 1.55[\mathrm{C}/\mathrm{m}^2], \tag{4.128}$$

and for a typical electrical field at the MEMS scale,

$$\begin{aligned}||\mathbf{E}_{\mathrm{dc}}|| &= \frac{V_{\mathrm{dc}}}{g} \\ &= \frac{10[\mathrm{V}]}{2[\mu\mathrm{m}]} = 5 \times 10^6[\mathrm{V}/\mathrm{m}],\end{aligned} \tag{4.129}$$

one finds that,

$$\frac{||\eta_{\mathrm{electrostatic}}||}{||\eta_{\mathrm{piezo}}||} \approx \frac{||\epsilon_r\epsilon_0\mathbf{E}_{\mathrm{dc}}||}{||\mathbf{e}||} = \frac{3.9 \times 10^{-4}}{1.55} = 2.6 \times 10^{-4}. \tag{4.130}$$

From this, one can conclude that for applications of piezoelectric material, the electromechanical coupling arising from the electrostatic energy is negligible. For piezoelectric materials, an electromechanical coupling from the material properties eliminates the requirement for an electrostatic electromechanical coupling arising from a DC bias voltage, and eliminates the need for solving for

the DC solution $(\boldsymbol{\varphi}_{\mathrm{dc}}, \phi_{\mathrm{dc}})$. This adds further simplifications for MEMS applications where small deformations are assumed, resulting in the assumption $\boldsymbol{\varphi}(\Omega) \approx \Omega$. Thus for the analysis of piezoelectric systems, the electrostatic electromechanical coupling is neglected and small deformations are assumed.

For the analysis of dielectric systems, i.e., non-piezoelectric systems, the electrostatic electromechanical coupling must be included, since excluding the term results in zero linearized coupling. Though analysis can still be simplified by assuming small mechanical deformations and only including the nonlinear contributions arising from the electrostatic electromechanical coupling. Formally, this is equivalent to assuming a Saint-Venant Kirchhoff elastic energy [54],

$$\Psi_{\mathrm{elastic}}(\mathbf{C}) \quad = \quad \frac{1}{8}(\mathbf{C}-\mathbf{1}) : \mathbb{C} : (\mathbf{C}-\mathbf{1}), \tag{4.131}$$

and only including the linear contributions in the displacement,

$$\mathbf{u}(\mathbf{X}) \quad := \quad \boldsymbol{\varphi}(\mathbf{X}) - \mathbf{X}, \tag{4.132}$$

to the force and stiffness contributions. The material stiffness tensor $\mathbb{C}$ is assumed to have major and minor symmetries. The force term in Equation (4.82) is approximated as,

$$\begin{aligned} \Pi_{\boldsymbol{\varphi}}(\boldsymbol{\varphi}, \phi)[\delta\boldsymbol{\varphi}] &= \int_\Omega \frac{1}{4}\delta\mathbf{C} : \mathbb{C} : (\mathbf{C}-\mathbf{1})\ d\Omega + \Pi_{\boldsymbol{\varphi}}^{\mathrm{electrostatic}}(\boldsymbol{\varphi}, \phi)[\delta\boldsymbol{\varphi}] && (4.133) \\ &\approx \int_\Omega \nabla_{\mathbf{X}}\delta\mathbf{u} : \mathbb{C} : \nabla_{\mathbf{X}}\mathbf{u}\ d\Omega + \Pi_{\boldsymbol{\varphi}}^{\mathrm{electrostatic}}(\boldsymbol{\varphi}, \phi)[\delta\boldsymbol{\varphi}], && (4.134) \end{aligned}$$

where

$$\begin{aligned} \mathbf{C} &= \mathbf{F}^T\mathbf{F} = (\mathbf{1} + \nabla_{\mathbf{X}}\mathbf{u})^T(\mathbf{1} + \nabla_{\mathbf{X}}\mathbf{u}) \approx \mathbf{1} + \nabla_{\mathbf{X}}\mathbf{u} + \nabla_{\mathbf{X}}\mathbf{u}^T, && (4.135) \\ \delta\mathbf{C} &\approx \nabla_{\mathbf{X}}\delta\mathbf{u} + \nabla_{\mathbf{X}}\delta\mathbf{u}^T, && (4.136) \end{aligned}$$

is assumed. Under these same assumptions, the stiffness in Equation (4.82) is approximated by,

$$\begin{aligned} \Pi_{\boldsymbol{\varphi}\boldsymbol{\varphi}}(\boldsymbol{\varphi}, \phi)[\delta\boldsymbol{\varphi}, \Delta\boldsymbol{\varphi}] &= \int_\Omega \frac{1}{4}\delta\mathbf{C} : \mathbb{C} : \Delta\mathbf{C}\ d\Omega + [\text{mechanical geometric stiffness}] \\ &\quad + \Pi_{\boldsymbol{\varphi}\boldsymbol{\varphi}}^{\mathrm{electrostatic}}(\boldsymbol{\varphi}, \phi)[\delta\boldsymbol{\varphi}, \Delta\boldsymbol{\varphi}] && (4.137) \\ &\approx \int_\Omega \nabla_{\mathbf{X}}\delta\mathbf{u} : \mathbb{C} : \nabla_{\mathbf{x}}\Delta\mathbf{u}\ d\Omega + \Pi_{\boldsymbol{\varphi}\boldsymbol{\varphi}}^{\mathrm{electrostatic}}(\boldsymbol{\varphi}, \phi)[\delta\phi, \Delta\boldsymbol{\varphi}]. && (4.138) \end{aligned}$$

This assumption of small deformations is justifiable by the following calculation. Assume a block of Aluminum Nitride sandwiched between two parallel plates with seperation g, with the "3-axis" perpendicular to the plane. The stress $\sigma_{\rm dc}$ in the material due to a voltage difference $V_{\rm dc}$ is,

$$\sigma_{\rm dc} \approx \frac{1}{2}\frac{\epsilon_r\epsilon_0}{g^2}V_{\rm dc}^2. \tag{4.139}$$

With the material properties for Aluminum Nitride,

$$\epsilon_r = 9, \tag{4.140}$$

$$\epsilon_0 = 8.85\times 10^{-12}[\mathrm{C}^2/\mathrm{Nm}^2], \tag{4.141}$$

$$\mathbb{C}_{3333} = 389[\mathrm{GPa}], \tag{4.142}$$

and for a typical gap and voltage,

$$V_{\rm dc} = 10[\mathrm{V}], \tag{4.143}$$

$$g = 2[\mu\mathrm{m}], \tag{4.144}$$

the strain in the material is,

$$\varepsilon_{\rm dc} = \frac{\sigma_{\rm dc}}{\mathbb{C}_{3333}} = 2\times 10^{-9}. \tag{4.145}$$

This small value of strain is well in the range of small deformations, validating the assumption for linear elasticity for the mechanical deformations. Thus in solving for the DC solution $(\boldsymbol{\varphi}_{\rm dc}, \phi_{\rm dc})$, one only needs to consider nonlinear contributions from the electrostatic electromechanical coupling. The nonlinear electrostatic electromechanical stiffness expressions for the electrostatic potential $\Psi_{\rm electrostatic}$ are presented in Appendix B.

4.5 Piezoelectric systems

Following the discussion presented in Section 4.4, the analysis of piezoelectric systems can be conducted under the assumption of no electrostatic coupling forces and small deformations. This removes the requirement for solving for the DC voltage solution. Under these assumptions, an appropriate expression for the potential energy is given as,

$$\Pi(\mathbf{u},\phi) \quad := \quad \int_{\Omega} \Psi(\boldsymbol{\varepsilon}(\mathbf{u}),\mathbf{E}(\phi))\ d\Omega + \Pi_{\text{ext}}(\mathbf{u},\phi), \tag{4.146}$$

where,

$$\boldsymbol{\varepsilon} \quad := \quad \text{sym}(\nabla_{\mathbf{X}}\mathbf{u}), \tag{4.147}$$

$$\mathbf{E} \quad := \quad -\nabla_{\mathbf{X}}\phi, \tag{4.148}$$

$$\Psi(\boldsymbol{\varepsilon},\mathbf{E}) \quad := \quad \Psi_{\text{elastic}}(\boldsymbol{\varepsilon}) + \Psi_{\text{electrostatic}}(\mathbf{E}) + \Psi_{\text{piezo}}(\boldsymbol{\varepsilon},\mathbf{E}), \tag{4.149}$$

$$\Psi_{\text{elastic}} \quad := \quad \frac{1}{2}\boldsymbol{\varepsilon} : \mathbb{C} : \boldsymbol{\varepsilon}, \tag{4.150}$$

$$\Psi_{\text{electrostatic}} \quad := \quad -\frac{1}{2}\mathbf{E}\cdot\boldsymbol{\kappa}\mathbf{E}, \tag{4.151}$$

$$\Psi_{\text{piezo}} \quad := \quad -\mathbf{E}\cdot\mathbf{e} : \boldsymbol{\varepsilon}. \tag{4.152}$$

Here $\boldsymbol{\varepsilon}$ is the small deformation strain tensor, $\mathbf{u}$ is the displacement, $\mathbf{E}$ is the electric field defined as the derivative of the potential ϕ with respect to the reference domain coordinates $\mathbf{X}$, $\mathbf{e}$ is the rank-3 piezoelectric stress coefficient tensor, and $\boldsymbol{\kappa}$ is the permittivity tensor.

In this section, the example of a 1D piezoelectric capacitor [117] is used to display the variational approach proposed. The formulas presented in the literature for the electromechanical coupling coefficient in piezoelectric resonators [156] is also presented formally from our variational approach. Our presentation allows one to rigorously derive such relations for the example geometry as wells as for more complex situations.

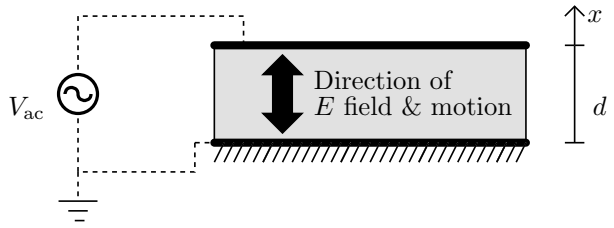


Figure 4.4: Configuration of the 1D piezoelectric problem. Gray denotes the piezoelectric material

4.5.1 1D piezoelectric capacitor

Consider the 1D problem of a piezoelectric resonator defined on the domain $\Omega := [0, d]$. A schematic of the configuration is shown in Figure 4.4. A piezoelectric material is inserted between two electrodes. A time varying AC voltage across the electrodes creates an electrical field in the material. This induces motion in the piezoelectric material in the direction of the electrical field. The energy for this system can be written as,

$$\begin{aligned} \Pi_{\text{total}}(u,\phi) &:= A\int_0^d \frac{1}{2}\varepsilon \cdot E_m \varepsilon \; dx - A\int_0^d \frac{1}{2}\kappa E^2 \; dx - A\int_0^d e \; \varepsilon \; E \; dx \\ &\quad + \Pi_{\text{ext}}(u,\phi), \end{aligned} \tag{4.153}$$

$$\Pi_{\text{ext}}(u,\phi) := Q_{\text{ac}}\phi(d) \; , \tag{4.154}$$

with,

$$\varepsilon = \frac{\partial u}{\partial x}, \tag{4.155}$$

$$E = -\frac{\partial \phi}{\partial x}, \tag{4.156}$$

and boundary conditions,

$$u(0) = 0, \tag{4.157}$$

$$\sigma(d) = 0, \tag{4.158}$$

$$\phi(0) = 0, \tag{4.159}$$

$$D(d) = Q_{\text{ac}}. \tag{4.160}$$

Here, A is the area of the cross section orthogonal to the x direction, E_m is the mechanical Young's modulus, e is the piezo-electric stress coefficient, κ is the permittivity of the material, D is the electric displacement, E is the electric field, σ is the stress, ε is the strain, ϕ is the potential, u is the displacement, and Q_{ac} is the total charge on the top electrode. The constitutive equations are,

$$\begin{pmatrix} \sigma \\ D \end{pmatrix} = \begin{bmatrix} E_m & -e \\ e & \kappa \end{bmatrix} \begin{pmatrix} \varepsilon \\ E \end{pmatrix}. \tag{4.161}$$

Under the assumption of electrostatics for the electrical field, the kinetic energy of the system is,

$$T := \frac{1}{2} A \int_0^d \rho \dot{u}^2 \, dx \, , \tag{4.162}$$

where ρ is the density of the piezoelectric material. From the Lagrangian equations of motion, one obtains,

$$L := T - \Pi_{\text{total}}, \tag{4.163}$$

$$\frac{d}{dt}\left[\frac{\partial L}{\partial(\dot{u},\dot{\phi})}\right] = \frac{\partial L}{\partial(u,\phi)} \, . \tag{4.164}$$

Given a time independent variation $[\delta u, \delta\phi]^T$,

$$\delta u \in \mathcal{V}_u := \left\{\delta u : H^1([0,d];\mathbb{C}) : u(0) = 0\right\}, \tag{4.165}$$

$$\delta\phi \in \mathcal{V}_\phi := \left\{\delta\phi : H^1([0,d];\mathbb{C}) : \phi(0) = 0\right\}. \tag{4.166}$$

one obtains the governing equations of the system,

$$\int_0^d \rho\ddot{u} \cdot \delta u \, dx + \Pi_u(u,\phi)[\delta u] = 0 \tag{4.167}$$

$$+\Pi_\phi(u,\phi)[\delta\phi] = -Q_{\text{ac}}\delta\phi(d), \tag{4.168}$$

where,

$$\Pi_u(u,\phi)[\delta u] = \int_0^d \frac{\partial \delta u}{\partial x}\cdot\ E_m A\frac{\partial u}{\partial x}\ dx + \int_0^d \frac{\partial \delta u}{\partial x}\cdot\ eA\frac{\partial \phi}{\partial x}\ dx \tag{4.169}$$

$$\Pi_\phi(u,\phi)[\delta \phi] = \int_0^d \frac{\partial \delta \phi}{\partial x}\cdot\ eA\frac{\partial u}{\partial x}\ dx - \int_0^d \frac{\partial \delta \phi}{\partial x}\cdot\ \kappa A\frac{\partial \delta \phi}{\partial x}\ dx\ . \tag{4.170}$$

Additionally a time-harmonic motion is assumed:

$$Q_{\mathrm{ac}} = \hat{Q}_{\mathrm{ac}}\exp(i\omega t), \tag{4.171}$$

$$u(x,t) = \hat{u}(x)\exp(i\omega t), \tag{4.172}$$

$$\phi(x,t) = \hat{\phi}(x)\exp(i\omega t). \tag{4.173}$$

When the electromechanical coupling is sufficiently small, a good approximation to the vibrational mode of the mechanical field are the eigenmodes of the pure mechanical modes. For the fixed-free boundary conditions of this problem, these are exactly the sine functions. For the electrical field, a linear field is the simplest assumption one can make.

$$\hat{u}, \delta u \in \left\{\sin\left(\alpha_n \frac{x}{d}\right)\right\}, \tag{4.174}$$

$$\alpha_n := \frac{(2n-1)}{2}\pi, \quad n \in \mathbb{N} \tag{4.175}$$

$$\hat{\phi}, \delta\phi \in \left\{\frac{x}{d}\right\}. \tag{4.176}$$

The solutions are normalized so that the generalized degrees of freedom $\hat{u}_{\mathrm{ac}}$, $\hat{V}_{\mathrm{ac}}$, representing the contribution of the mechanical and electrical mode are the displacement and voltage potential at $x = d$,

$$\hat{u}(x) = \hat{u}_{\mathrm{ac}}\sin\left(\alpha_n \frac{x}{d}\right), \tag{4.177}$$

$$\hat{\phi}(x) = \hat{V}_{\mathrm{ac}}\frac{x}{d}. \tag{4.178}$$

These in combination with Equations (4.167) and (4.168) yield the system of equations,

$$-\omega^2 \begin{bmatrix} \rho A\frac{d}{2} & 0 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} \hat{u}_{\mathrm{ac}} \\ \hat{V}_{\mathrm{ac}} \end{pmatrix} + \begin{bmatrix} E_m A\alpha_n^2 \frac{1}{2d} & \frac{eA}{d} \\ \frac{eA}{d} & -\frac{\kappa A}{d} \end{bmatrix} \begin{pmatrix} \hat{u}_{\mathrm{ac}} \\ \hat{V}_{\mathrm{ac}} \end{pmatrix} = \begin{pmatrix} 0 \\ -\hat{Q}_{\mathrm{ext}} \end{pmatrix}, \tag{4.179}$$

or,

$$-\omega^2 \begin{bmatrix} m_{\rm eq} & 0 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} \hat{u}_{\rm ac} \\ \hat{V}_{\rm ac} \end{pmatrix} + \begin{bmatrix} k_{\rm eq} & \eta_{\rm eq} \\ \eta_{\rm eq} & -C_0 \end{bmatrix} \begin{pmatrix} \hat{u}_{\rm ac} \\ \hat{V}_{\rm ac} \end{pmatrix} = \begin{pmatrix} 0 \\ -\hat{Q}_{\rm ext} \end{pmatrix} . \tag{4.180}$$

From this the admittance,

$$\hat{I}_{\rm ac} := i\omega\hat{Q}_{\rm ext} = Y(\omega)\hat{V}_{\rm ac}, \tag{4.181}$$

$$Y(\omega) := \left[i\omega C_0 + \frac{i\omega\eta_{\rm eq}^2}{k_{\rm eq} - m_{\rm eq}\omega^2} \right], \tag{4.182}$$

$$= i\omega C_0 \left[1 + \frac{2K^2}{\frac{\alpha_n^2}{1+k^2} - \alpha^2} \right], \tag{4.183}$$

$$\alpha := \frac{\omega d}{\sqrt{E_m/\rho}} \frac{1}{\sqrt{1+k^2}}, \tag{4.184}$$

$$k^2 ;= \frac{e^2}{E_m \kappa}, \tag{4.185}$$

$$K^2 := \frac{k^2}{1+k^2}, \tag{4.186}$$

and equivalent circuit parameters,

$$L_{\rm eq} := \frac{m_{\rm eq}}{\eta_{\rm eq}^2} = \frac{1}{2}\frac{\rho d^3}{e^2 A}, \tag{4.187}$$

$$C_{\rm eq} := \frac{\eta_{\rm eq}^2}{k_{\rm eq}}, = 2\frac{e^2 A}{E_m d \alpha_n^2} \tag{4.188}$$

can be computed. The exact solution to this problem yields,

$$Y_{\rm exact}(\omega) = i\omega C_0 \left[1 - K^2 \frac{\tan\alpha}{\alpha} \right]^{-1} . \tag{4.189}$$

The non-dimensionalized admittance $Y/i\omega C_0$ with respect to non-dimensional frequency α is plotted for both the 1 degree of freedom and exact model in Figure 4.5. The range plotted is in the vicinity of the 1st mode of vibration with pole and zero close to $\pi/2$. The zero of the exact admittance is equal to $\pi/2$, and the pole is the solution to the equation $\tan\alpha = 1/K^2\alpha$ closest to $\pi/2$. The case for two values of coupling coefficients $k^2 = \{0.01, 0.1\}$ are presented. One observes that as the coupling increases there is large increase in the difference between the two approximations. Though it is not presented here, it is noted that the situation for higher modes are worse. For better approximations with large coupling and high modes, one must include more degrees of freedom in the approximation.

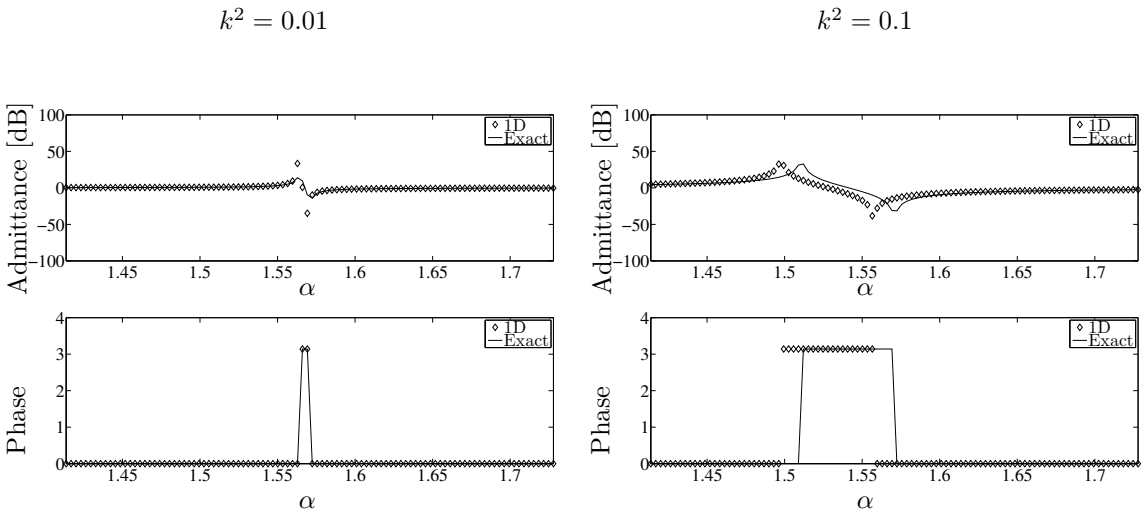


Figure 4.5: Non-dimensionalized admittance for the 1D piezoelectric resonator with coupling coefficients $k^2 \in \{0.01, 0.1\}$

4.5.2 2D plane stress with out of plane forcing

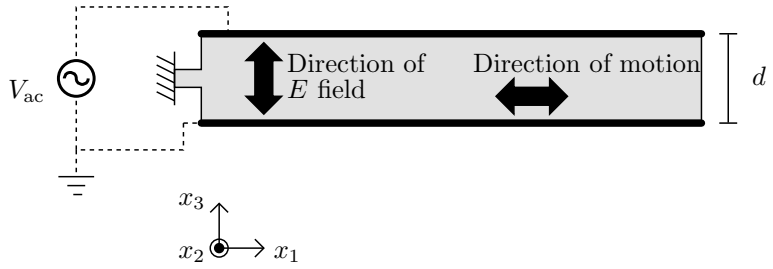


Figure 4.6: Configuration of the 2D piezoelectric problem. Gray denotes the piezoelectric material

In the presented 1D piezoelectric resonator, the thickness of the piezoelectric device d, defines the resonance frequency. For such resonator geometries, one cannot fabricate resonators with different resonant frequencies side by side in current lithographic fabrication processes. To overcome this difficulty, resonators with a resonance frequency defined by the in-plane geometry have been developed [155]. These 2D planar structures are actuated by an out of plane "3 direction" electric field. This causes in-plane forces with the e_{113}, e_{223} components of the piezoelectric stress coefficient

tensor, as opposed to using the e_{333} component used in the thin-film bulk piezoelectric resonators of the type presented in the previous section. A schematic of this configuration is shown in Figure 4.6. This planar design is relatively insensitive to the thickness of the process layer, as well as allows several structures with different resonating frequency to coexist on a single fabrication layer.

For the calculation of these planar systems, plane stress is assumed,

$$\sigma_{33} = 0, \quad \sigma_{13} = 0, \quad \sigma_{23} = 0, \tag{4.190}$$

with the x_3 direction pointing out of the plane. In the following, Voigt notation is used to express the rank-4 elasticity tensor $\mathbb{C}$ and the rank-3 piezoelectric stress coefficient $\mathbf{e}$ and rank-3 piezoelectric strain coefficient $\mathbf{d}$,

$$v(i,i) \to i \quad \text{for } i \in \{1,2,3\}, \quad v(1,2) \to 4, \quad v(2,3) \to 5, \quad v(3,1) \to 6, \tag{4.191}$$

such that,

$$\mathbb{C}_{ijkl} = \mathbb{C}_{v(i,j)v(k,l)}, \tag{4.192}$$

$$\mathbf{e}_{ijk} = \mathbf{e}_{iv(j,k)}, \tag{4.193}$$

$$\mathbf{d}_{ijk} = \mathbf{d}_{iv(j,k)}. \tag{4.194}$$

Under Voigt notation, the tensors are matrices of the corresponding size, $\mathbb{C} \in \mathbb{R}^{6\times6}$, $\mathbf{e} \in \mathbb{R}^{3\times6}$, $\mathbf{d} \in \mathbb{R}^{3\times6}$. The indices $1 \to 6$ are grouped into the set of in-plane components and the set of out-of-plane components,

$$\begin{aligned} \text{in-plane}: \quad p &:= \{1,2,4\}, \\ \text{out-of-plane}: \quad o &:= \{3,5,6\}. \end{aligned} \tag{4.195}$$

$\mathbb{C}$, $\mathbf{e}$, $\mathbf{d}$, $\boldsymbol{\sigma}$, and $\boldsymbol{\varepsilon}$ can be divided into the components as,

$$\mathbb{C} := \begin{bmatrix} \mathbb{C}_{pp} & \mathbb{C}_{po} \\ \mathbb{C}_{op} & \mathbb{C}_{oo} \end{bmatrix}, \tag{4.196}$$

$$\mathbf{e} := \begin{bmatrix} \mathbf{e}_p & \mathbf{e}_o \end{bmatrix}, \tag{4.197}$$

$$\mathbf{d} := \begin{bmatrix} \mathbf{e}_p & \mathbf{e}_o \end{bmatrix}, \tag{4.198}$$

$$\boldsymbol{\sigma} := \begin{bmatrix} \boldsymbol{\sigma}_p \\ \boldsymbol{\sigma}_o \end{bmatrix}, \tag{4.199}$$

$$\boldsymbol{\varepsilon} := \begin{bmatrix} \boldsymbol{\varepsilon}_p \\ \boldsymbol{\varepsilon}_o \end{bmatrix}. \tag{4.200}$$

The constitutive equations are,

$$\begin{pmatrix} \boldsymbol{\sigma}_p \\ \boldsymbol{\sigma}_o \\ \mathbf{D} \end{pmatrix} = \begin{bmatrix} \mathbb{C}_{pp} & \mathbb{C}_{po} & -\mathbf{e}_p^T \\ \mathbb{C}_{op} & \mathbb{C}_{oo} & -\mathbf{e}_o^T \\ \mathbf{e}_p & \mathbf{e}_o & \boldsymbol{\kappa} \end{bmatrix} \begin{pmatrix} \boldsymbol{\varepsilon}_p \\ \boldsymbol{\varepsilon}_o \\ \mathbf{E} \end{pmatrix}, \tag{4.201}$$

and for the plane stress case they reduced to,

$$\begin{pmatrix} \boldsymbol{\sigma}_p \\ \mathbf{D} \end{pmatrix} = \begin{bmatrix} \mathbb{C}^\sigma & -\mathbf{e}^{\sigma,T} \\ \mathbf{e}^\sigma & \boldsymbol{\kappa}^\sigma \end{bmatrix} \begin{pmatrix} \boldsymbol{\varepsilon}_p \\ \mathbf{E} \end{pmatrix}, \tag{4.202}$$

$$\mathbb{C}^\sigma := \mathbb{C}_{pp} - \mathbb{C}_{po}\mathbb{C}_{oo}^{-1}\mathbb{C}_{op}, \tag{4.203}$$

$$\mathbf{e}^\sigma := \mathbf{e}_p - \mathbf{e}_o\mathbb{C}_{oo}^{-1}\mathbb{C}_{op} = \mathbf{d}_p\mathbb{C}^\sigma, \tag{4.204}$$

$$\boldsymbol{\kappa}^\sigma := \boldsymbol{\kappa} + \mathbf{e}_o\mathbb{C}_{oo}^{-1}\mathbf{e}_o^T. \tag{4.205}$$

The quantities with the superscript σ denote the equivalent quantities in the plane stress case.

The expression for the energy becomes identical to Equations (4.146) - (4.152), but with the

replacements,

$$\boldsymbol{\sigma} \rightarrow \boldsymbol{\sigma}_p, \tag{4.206}$$

$$\boldsymbol{\varepsilon} \rightarrow \boldsymbol{\varepsilon}_p, \tag{4.207}$$

$$\mathbf{u} := [u_1, u_2]^t, \tag{4.208}$$

$$\mathbb{C} \rightarrow \mathbb{C}^\sigma, \tag{4.209}$$

$$\boldsymbol{\kappa} \rightarrow \boldsymbol{\kappa}^\sigma, \tag{4.210}$$

$$\mathbf{e} \rightarrow \mathbf{e}^\sigma. \tag{4.211}$$

Given these expressions, one can compute the piezoelectric electromechanical coupling term,

$$-\int_\Omega \mathbf{E} \cdot \mathbf{e}^\sigma : \delta\boldsymbol{\varepsilon}_p \; d\Omega \; . \tag{4.212}$$

For a first approximation, two assumptions which are identical to those made in the analysis conducted in the previous section for the 1D piezoelectric capacitor are made. Namely,

- The potential field in the x_3 direction is assumed linear,

$$\phi(x, y, z) := -V_{\text{ac}} \frac{z}{d} \; , \tag{4.213}$$

$$\mathbf{E} = \frac{V_{\text{ac}}}{d} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{4.214}$$

- The mechanical mode $\mathbf{u}$ as well as its variation $\delta\mathbf{u}$ is approximated by the purely mechanical eigenmodes,

$$\mathbf{u}, \delta\mathbf{u} \in \left\{ \mathbf{u}^{mech,eig} \right\}. \tag{4.215}$$

These assumptions lead to the expression,

$$-\int_\Omega \mathbf{E} \cdot \mathbf{e}^\sigma : \delta\boldsymbol{\varepsilon}_p^{\text{mech,eig}} \; d\Omega = -\int_0^d \int_{xy} \frac{V_{\text{ac}}}{d} [\mathbf{e}^\sigma \delta\boldsymbol{\varepsilon}_p^{\text{mech,eig}}]_3 \; dxdy \; dz \tag{4.216}$$

$$= -V_{\text{ac}} \int_{xy} [\mathbf{e}^\sigma \delta\boldsymbol{\varepsilon}_p^{\text{mech,eig}}]_3 \; dxdy, \tag{4.217}$$

and thus an expression for the electromechanical coupling is obtained.

$$\eta_{\text{eq}} := -\int_{xy} [\mathbf{e}^{\sigma} \delta\boldsymbol{\varepsilon}_p^{\text{mech,eig}}]_3 \; dxdy, \tag{4.218}$$

$$= -\int_{xy} [\mathbf{d}_p \delta\boldsymbol{\sigma}_p^{\text{mech,eig}}]_3 \; dxdy. \tag{4.219}$$

Here, $\delta\boldsymbol{\sigma}_p^{\text{mech,eig}} := \mathbb{C}^{\sigma}\delta\boldsymbol{\varepsilon}_p^{\text{mech,eig}}$, is the stress distribution arising from the pure mechanical mode. These are the two expressions one encounters in the literature [156], with the exception that there the plane stress piezoelectric stress coefficient $\mathbf{e}^{\sigma}$ is replaced with $\mathbf{e}$. For Aluminum Nitride the values are approximately $\mathbf{e}_{31} = -0.58$ and $\mathbf{e}_{31}^{\sigma} = -0.9746$. There is a factor of two difference which is non-negligible.

4.6 Numerical evaluation

In the previous sections, variational expressions to evaluate the equivalent circuit parameters for electromechanically coupled systems have been presented. Unless the geometry is special, a closed form expression for the mechanical eigenmode or potential eigenmode do not exist. Additionally for cases where the electromechanical coupling is strong, the purely mechanical eigenmodes no longer are accurate enough to represent the actual vibrational modes. In such cases, the evaluation of the expressions must be done numerically.

Discretization of Equations (4.82) and (4.83) by the finite element method yield the linear system of equations,

$$\begin{bmatrix} \mathbf{M_{uu}} & 0 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} \ddot{\mathbf{u}}_{\mathrm{ac}} \\ \ddot{\boldsymbol{\phi}}_{\mathrm{ac}} \end{pmatrix} + \begin{bmatrix} \mathbf{K_{uu}} & \mathbf{K}_{\mathbf{u}\boldsymbol{\phi}} \\ \mathbf{K}_{\boldsymbol{\phi}\mathbf{u}} & \mathbf{K}_{\boldsymbol{\phi}\boldsymbol{\phi}} \end{bmatrix} \begin{pmatrix} \mathbf{u}_{\mathrm{ac}} \\ \boldsymbol{\phi}_{\mathrm{ac}} \end{pmatrix} = \begin{pmatrix} 0 \\ -\mathbf{q}_{\mathrm{ac}} \end{pmatrix}, \tag{4.220}$$

where with an abuse of notation $\mathbf{u}_{\mathrm{ac}}$ denotes the nodal displacement vector, $\boldsymbol{\phi}_{\mathrm{ac}}$ is the nodal potential vector, and $\mathbf{q}_{\mathrm{ac}}$ is the nodal charge vector. Assume that the boundary $\Gamma_{\mathrm{ac}} \subset \Gamma_q$, where the time varying boundary charge $\hat{\sigma}_{\mathrm{q,ac}}$ is applied, is divided into n_{ac} disjoint sections,

$$\Gamma_{\mathrm{ac}} := \Gamma_{\mathrm{ac},1} \sqcup \cdots \sqcup \Gamma_{\mathrm{ac},n_{\mathrm{ac}}}. \tag{4.221}$$

Assume that on each $\Gamma_{\mathrm{ac},i}$, the potentials are fixed to the value $V_{ac,i}$. By denoting the vector of nodal potentials on $\Gamma_{\mathrm{ac},i}$ by $\boldsymbol{\phi}_{V,i}$, one has

$$\boldsymbol{\phi}_{V,i} := \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} V_{\mathrm{ac,i}}. \tag{4.222}$$

The physical interpretation of such a condition is that the points on $\Gamma_{\mathrm{ac},i}$ are connected to each other by a conductor. Define the vector $\mathbf{V}$ and $\boldsymbol{\phi}_V$ as,

$$\mathbf{V}_{\mathrm{ac}} := \begin{pmatrix} V_{\mathrm{ac},0} \\ \vdots \\ V_{\mathrm{ac},n_q} \end{pmatrix}, \qquad \boldsymbol{\phi}_{\mathrm{ac},V} := \begin{pmatrix} \boldsymbol{\phi}_{V,0} \\ \vdots \\ \boldsymbol{\phi}_{V,n_q} \end{pmatrix} = \mathbf{G}\mathbf{V}_{\mathrm{ac}}, \tag{4.223}$$

where $\mathbf{G}$ is a matrix of ones and zeros expressing the connectivity. Denote the vector of nodal potentials not on Γ_{ac} as $\boldsymbol{\phi}_{\text{ac},0}$, such that $\boldsymbol{\phi}_{\text{ac}} = [\boldsymbol{\phi}_{\text{ac},0}^T, \boldsymbol{\phi}_{\text{ac},V}^T]^T$. With this notation one has,

$$\begin{pmatrix} \mathbf{u}_{\text{ac}} \\ \boldsymbol{\phi}_{\text{ac},0} \\ \boldsymbol{\phi}_{\text{ac},V} \end{pmatrix} = \begin{bmatrix} \mathbf{I} & 0 & 0 \\ 0 & \mathbf{I} & 0 \\ 0 & 0 & \mathbf{G} \end{bmatrix} \begin{pmatrix} \mathbf{u}_{\text{ac}} \\ \boldsymbol{\phi}_{\text{ac},0} \\ \mathbf{V}_{\text{ac}} \end{pmatrix}. \tag{4.224}$$

The system above can be rewritten,

$$\begin{bmatrix} \mathbf{M_{uu}} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} \ddot{\mathbf{u}}_{\text{ac}} \\ \ddot{\boldsymbol{\phi}}_{\text{ac},0} \\ \ddot{\boldsymbol{\phi}}_{\text{ac},V} \end{pmatrix} + \begin{bmatrix} \mathbf{K_{uu}} & \mathbf{K_{u\phi_0}} & \mathbf{K_{u\phi_V}} \\ \mathbf{K_{\phi_0 u}} & \mathbf{K_{\phi_0\phi_0}} & \mathbf{K_{\phi_0\phi_V}} \\ \mathbf{K_{\phi_V u}} & \mathbf{K_{\phi_V\phi_0}} & \mathbf{K_{\phi_V\phi_V}} \end{bmatrix} \begin{pmatrix} \mathbf{u}_{\text{ac}} \\ \boldsymbol{\phi}_{\text{ac},0} \\ \boldsymbol{\phi}_{\text{ac},V} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -\mathbf{q}_{\text{ac},V} \end{pmatrix} \tag{4.225}$$

A Galerkin projection with the relation in Equation (4.224) yields,

$$\begin{bmatrix} \mathbf{M_{uu}} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} \ddot{\mathbf{u}}_{\text{ac}} \\ \ddot{\boldsymbol{\phi}}_{\text{ac},0} \\ \ddot{\mathbf{V}}_{\text{ac}} \end{pmatrix} + \begin{bmatrix} \mathbf{K_{uu}} & \mathbf{K_{u\phi_0}} & \mathbf{K_{u\phi_V}}\mathbf{G} \\ \mathbf{K_{\phi_0 u}} & \mathbf{K_{\phi_0\phi_0}} & \mathbf{K_{\phi_0\phi_V}}\mathbf{G} \\ \mathbf{G}^T\mathbf{K_{\phi_V u}} & \mathbf{G}^T\mathbf{K_{\phi_V\phi_0}} & \mathbf{G}^T\mathbf{K_{\phi_V\phi_V}}\mathbf{G} \end{bmatrix} \begin{pmatrix} \mathbf{u}_{\text{ac}} \\ \boldsymbol{\phi}_{\text{ac},0} \\ \mathbf{V}_{\text{ac}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -\mathbf{Q}_{\text{ac}} \end{pmatrix}. \tag{4.226}$$

The vector $\mathbf{Q}_{\text{ac}}$ is the same size as $\mathbf{V}_{\text{ac}}$, and the entries are the total charges on boundary $\Gamma_{q,i}$. By defining $\mathbf{U}_{\text{ac}} := [\mathbf{u}_{\text{ac}}^T, \boldsymbol{\phi}_{\text{ac},0}^T]^T$, the system is further rewritten as,

$$\begin{bmatrix} \mathbf{M_{UU}} & 0 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} \ddot{\mathbf{U}}_{\text{ac}} \\ \ddot{\mathbf{V}}_{\text{ac}} \end{pmatrix} + \begin{bmatrix} \mathbf{K_{UU}} & \mathbf{K_{UV}} \\ \mathbf{K_{VU}} & \mathbf{K_{VV}} \end{bmatrix} \begin{pmatrix} \mathbf{U}_{\text{ac}} \\ \mathbf{V}_{\text{ac}} \end{pmatrix} = \begin{pmatrix} 0 \\ -\mathbf{Q}_{\text{ac}} \end{pmatrix}. \tag{4.227}$$

Under time harmonic assumptions,

$$\mathbf{U}_{\mathrm{ac}} = \hat{\mathbf{U}}_{\mathrm{ac}} \exp(i\omega t), \tag{4.228}$$

$$\mathbf{V}_{\mathrm{ac}} = \hat{\mathbf{V}}_{\mathrm{ac}} \exp(i\omega t), \tag{4.229}$$

$$\mathbf{Q}_{\mathrm{ac}} = \hat{\mathbf{Q}}_{\mathrm{ac}} \exp(i\omega t), \tag{4.230}$$

$$\begin{aligned} \mathbf{I}_{\mathrm{ac}} &:= \frac{d\mathbf{Q}_{\mathrm{ac}}}{dt} \\ &= \hat{\mathbf{I}}_{\mathrm{ac}} \exp(i\omega t), \end{aligned} \tag{4.231}$$

$$\hat{\mathbf{I}}_{\mathrm{ac}} := i\omega \hat{\mathbf{Q}}_{\mathrm{ac}}, \tag{4.232}$$

an admittance matrix $\mathbf{Y}(\omega)$ relating the voltage potentials on the boundaries $\hat{\mathbf{V}}_{\mathrm{ac}}$ with the currents $\hat{\mathbf{I}}_{\mathrm{ac}}$ can be obtained,

$$\hat{\mathbf{I}}_{\mathrm{ac}} := i\omega \hat{\mathbf{Q}}_{\mathrm{ac}} = \mathbf{Y}(\omega) \hat{\mathbf{V}}_{\mathrm{ac}}, \tag{4.233}$$

$$\mathbf{Y}(\omega) := -i\omega \mathbf{K}_{\mathbf{VV}} + i\omega \mathbf{K}_{\mathbf{VU}} \left(\mathbf{K}_{\mathbf{UU}} - \omega^2 \mathbf{M}_{\mathbf{UU}} \right)^{-1} \mathbf{K}_{\mathbf{UV}}. \tag{4.234}$$

Unfortunately in this form, one has a large linear system in the representation for the admittance, which is not convenient. In order to derive equivalent circuit parameters, one must project the many degrees of freedom system to a system with a small number of degrees of freedom, ideally a 1 degree of freedom system. When one is interested in the resonance behavior of the system, this can be approximated well by the nearest eigenvector. Thus given a desired operation frequency or mode, one can compute the corresponding eigenvector $\mathbf{v}$ and eigenfrequency ω_0 from the generalized eigenvalue problem,

$$\mathbf{K}_{\mathbf{UU}} \mathbf{v} = \omega_0^2 \mathbf{M}_{\mathbf{UU}} \mathbf{v} \ . \tag{4.235}$$

In the case of a non-symmetric pencil $(\mathbf{K}_{\mathbf{UU}}, \mathbf{M}_{\mathbf{UU}})$, one also requires the left eigenvector $\mathbf{w}$,

$$\mathbf{w}^* \mathbf{K}_{\mathbf{UU}} = \omega_0^2 \mathbf{w}^* \mathbf{M}_{\mathbf{UU}} \ . \tag{4.236}$$

The eigenvectors are computed from the coupled problem, reflecting the electromechanical coupling existing in this system. This approximation is more accurate than projection onto the pure

mechanical modes of the system, and requires no assumption on the electrical field.

There are two ways that one can project the system of Equation (4.226) onto these modes. The projected system has the form,

$$\mathbf{Y}(\omega) \quad = \quad \left[\mathbf{C}_{\mathrm{eq},0} + \frac{i\omega}{k_{\mathrm{eq}} - \omega^2 m_{\mathrm{eq}}}\boldsymbol{\eta}_{\mathrm{eq}}\right]. \tag{4.237}$$

where the parameters $\mathbf{C}_{\mathrm{eq},0}$, k_{eq}, m_{eq}, $\beta_{\mathrm{UV,eq}}$, and β_{eq} depend on the method of projection. Let us introduce the decomposition of $\mathbf{v}$ and $\mathbf{w}$ into the mechanical and potential degrees of freedom.

$$\mathbf{v} \quad := \quad \begin{pmatrix} \mathbf{u}^{\mathbf{v}}_{\mathrm{ac}} \\ \phi^{\mathbf{v}}_{\mathrm{ac},0} \end{pmatrix}, \tag{4.238}$$

$$\mathbf{w} \quad := \quad \begin{pmatrix} \mathbf{u}^{\mathbf{w}}_{\mathrm{ac}} \\ \phi^{\mathbf{w}}_{\mathrm{ac},0} \end{pmatrix}. \tag{4.239}$$

The two methods of projection are the following.

1. Under the right and left projection matrices,

$$\mathbf{P}_r \quad = \quad \begin{bmatrix} \mathbf{u}^{\mathbf{v}}_{\mathrm{ac}} & \mathbf{0} \\ \phi^{\mathbf{v}}_{\mathrm{ac},0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}, \tag{4.240}$$

$$\mathbf{P}_l \quad = \quad \begin{bmatrix} \mathbf{u}^{\mathbf{w}}_{\mathrm{ac}} & \mathbf{0} \\ \phi^{\mathbf{w}}_{\mathrm{ac},0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}. \tag{4.241}$$

This leads to the expressions,

$$\mathbf{C}_{\mathrm{eq},0} := -\mathbf{K_{VV}}, \tag{4.242}$$

$$m_{\mathrm{eq}} := \mathbf{w}^{*}\mathbf{M_{UU}}\mathbf{v}, \tag{4.243}$$

$$k_{\mathrm{eq}} := \mathbf{w}^{*}\mathbf{K_{UU}}\mathbf{v}, \tag{4.244}$$

$$\boldsymbol{\eta}_{\mathrm{UV}} := \mathbf{w}^{*}\mathbf{K_{UV}}, \tag{4.245}$$

$$\boldsymbol{\eta}_{\mathrm{VU}} := \mathbf{K_{VU}}\mathbf{v}, \tag{4.246}$$

$$\boldsymbol{\eta}_{\mathrm{eq}} := \boldsymbol{\eta}_{\mathrm{VU}}\boldsymbol{\eta}_{\mathrm{UV}}. \tag{4.247}$$

2. Under the right and left projection matrices,

$$\mathbf{P}_r = \begin{bmatrix} \mathbf{u}_{\mathrm{ac}}^{\mathbf{v}} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}, \tag{4.248}$$

$$\mathbf{P}_l = \begin{bmatrix} \mathbf{u}_{\mathrm{ac}}^{\mathbf{w}} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}. \tag{4.249}$$

The projections are not applied onto Equation (4.226) directly but onto the Schur complement system with respect to $\boldsymbol{\phi}_{\mathrm{ac},0}$. This leads to the expressions,

$$\mathbf{C}_{\mathrm{eq},0} := -\mathbf{K_{VV}} + \mathbf{G}^{T}\mathbf{K}_{\boldsymbol{\phi}_V\boldsymbol{\phi}_0}\mathbf{K}^{-1}_{\boldsymbol{\phi}_0\boldsymbol{\phi}_0}\mathbf{K}_{\boldsymbol{\phi}_0\boldsymbol{\phi}_V}\mathbf{G}, \tag{4.250}$$

$$m_{\mathrm{eq}} := \mathbf{w}^{*}\mathbf{M_{UU}}\mathbf{v}, \tag{4.251}$$

$$k_{\mathrm{eq}} := \mathbf{w}^{*}\begin{bmatrix} \mathbf{K_{uu}} & \frac{1}{2}\mathbf{K}_{\mathbf{u}\boldsymbol{\phi}_0} \\ \frac{1}{2}\mathbf{K}_{\boldsymbol{\phi}_0\mathbf{u}} & \mathbf{0} \end{bmatrix}\mathbf{v}, \tag{4.252}$$

$$\boldsymbol{\eta}_{\mathrm{UV}} := \mathbf{w}^{*}\mathbf{K_{UV}}, \tag{4.253}$$

$$\boldsymbol{\eta}_{\mathrm{VU}} := \mathbf{K_{VU}}\mathbf{v}, \tag{4.254}$$

$$\boldsymbol{\eta}_{\mathrm{eq}} := \boldsymbol{\eta}_{\mathrm{VU}}\boldsymbol{\eta}_{\mathrm{UV}}. \tag{4.255}$$

The two different projections result in a different matrix $\mathbf{C}_{\mathrm{eq},0}$ which represents the static capacitance of the resonator, and k_{eq} which is proportional to the inverse of the capacitance in

the series LRC unit in the equivalent circuit model. As it is shown in the numerical examples in Section 5.4, method 1 results in a static capacitance value which is not very accurate as opposed to method 2 which is. Thus method 2 should be chosen. This difference in behavior can be attributed to the singularity of $\mathbf{M_{UU}}$ induced by the zero block of potentials. The matrix $\boldsymbol{\eta}$ represents the electromechanical coupling and m_{eq} is proportional to the inductance in the series LRC unit in the equivalent circuit model.

The reduced system of Equation (4.237) simulates the behavior of the system accurately near the frequency ω_0. In the following, some specific examples illustrating the use of Equation (4.237) is presented.

Remark:

The classical method of equivalent parameter extraction, where the mechanical modes of vibration are used in the projection, can be recovered from our framework through the following process.

In the classical method, the electric potential field is assumed a linear field across the electrodes. Under this assumption, the potentials not attached to an electrode, i.e., $\boldsymbol{\phi}_{\text{ac},0}$, can be expressed as a linear combination of the electrode potentials $\mathbf{V}_{\text{ac}}$,

$$\boldsymbol{\phi}_{\text{ac},0} \quad = \quad \mathbf{B}\mathbf{V}_{\text{ac}}, \tag{4.256}$$

where $\mathbf{B}$ is the matrix relating the two vectors. Define $\mathbf{u}^{\mathbf{w}}_{\text{ac,mech}}$ and $\mathbf{u}^{\mathbf{w}}_{\text{ac,mech}}$ as the left and right purely mechanical eigenvectors corresponding to the eigenvalue closest to ω_0. These represent the purely mechanical modes of vibration. Given these expressions, define the right and left projection

matrices,

$$\mathbf{P}_r = \begin{bmatrix} \mathbf{u}^{\mathbf{v}}_{\text{ac,mech}} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}, \tag{4.257}$$

$$\mathbf{P}_l = \begin{bmatrix} \mathbf{u}^{\mathbf{w}}_{\text{ac,mech}} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}. \tag{4.258}$$

By appling these projections to Equation (4.226), one obtains the classical expressions for the equivalent circuit model parameters.

$$\mathbf{C}_{\text{eq},0} := -\mathbf{K}_{\mathbf{VV}} - \mathbf{B}^T\mathbf{K}_{\phi_0\phi_V}\mathbf{G} - \mathbf{G}^T\mathbf{K}_{\phi_V\phi_0}\mathbf{B} - \mathbf{B}^T\mathbf{K}_{\phi_0\phi_0}\mathbf{B}, \tag{4.259}$$

$$m_{\text{eq}} := {\mathbf{u}^{\mathbf{w}}_{\text{ac,mech}}}^*\mathbf{M}_{\mathbf{uu}}\mathbf{u}^{\mathbf{v}}_{\text{ac,mech}}, \tag{4.260}$$

$$k_{\text{eq}} := {\mathbf{u}^{\mathbf{w}}_{\text{ac,mech}}}^*\mathbf{K}_{\mathbf{uu}}\mathbf{u}^{\mathbf{v}}_{\text{ac,mech}}, \tag{4.261}$$

$$\boldsymbol{\eta}_{\text{uV}} := {\mathbf{u}^{\mathbf{w}}_{\text{ac,mech}}}^*\mathbf{K}_{\mathbf{u}\phi_0}\mathbf{B} + {\mathbf{u}^{\mathbf{w}}_{\text{ac,mech}}}^*\mathbf{K}_{\mathbf{u}\phi_V}\mathbf{G}, \tag{4.262}$$

$$\boldsymbol{\eta}_{\text{Vu}} := \mathbf{B}^T\mathbf{K}_{\phi_0\mathbf{u}}\mathbf{u}^{\mathbf{v}}_{\text{ac,mech}} + \mathbf{G}^T\mathbf{K}_{\phi_V\mathbf{u}}\mathbf{u}^{\mathbf{v}}_{\text{ac,mech}}, \tag{4.263}$$

$$\boldsymbol{\eta}_{\text{eq}} := \boldsymbol{\eta}_{\text{Vu}}\boldsymbol{\eta}_{\text{uV}}. \tag{4.264}$$

1-Port case

In the one port configuration, the boundary $\Gamma_{\rm ac}$ consists of one part $\Gamma_{\rm ac,1}$. (See the example in Section 5.4 for details on 1-port configuration). This yields a 1-degree of freedom system,

$$Y_1(\omega) \;=\; \left[\mathbf{C}_{0,\rm eq}\right]_{11} + \frac{i\omega}{k_{\rm eq}-\omega^2 m_{\rm eq}}\left[\boldsymbol{\eta}_{\rm eq}\right]_{11}\;. \tag{4.265}$$

The equivalent circuit parameters are,

$$\eta_{\rm eq} \;:=\; \left[\boldsymbol{\eta}_{\rm eq}\right]_{11}, \tag{4.266}$$

$$L_{\rm eq} \;:=\; \frac{m_{\rm eq}}{\eta_{\rm eq}^2}, \tag{4.267}$$

$$C_{\rm eq} \;:=\; \frac{\eta_{\rm eq}^2}{k_{\rm eq}}, \tag{4.268}$$

$$C_{0,\rm eq} \;:=\; \left[\mathbf{C}_{0,\rm eq}\right]_{11}. \tag{4.269}$$

In the case that $\eta_{\rm eq}$, $m_{\rm eq}$, and $k_{\rm eq}$ are complex, as in the application of PML, an approximation with real parameters can be formulated,

$$L_{\rm eq} \;:=\; {\rm Re}\left(\frac{m_{\rm eq}}{\eta_{\rm eq}^2}\right), \tag{4.270}$$

$$C_{\rm eq} \;:=\; \left[{\rm Re}\left(\frac{k_{\rm eq}}{\eta_{\rm eq}^2}\right)\right]^{-1}, \tag{4.271}$$

$$R_{\rm eq} \;:=\; -i{\rm Re}(\omega_0){\rm Im}\left(\frac{m_{\rm eq}}{\eta_{\rm eq}^2}\right) + \frac{1}{{\rm Re}(\omega_0)}{\rm Im}\left(\frac{k_{\rm eq}}{\eta_{\rm eq}^2}\right), \tag{4.272}$$

$$C_{0,\rm eq} \;:=\; \left[\mathbf{C}_{0,\rm eq}\right]_{11}. \tag{4.273}$$

2-Port case

In the two port configuration, the boundary $\Gamma_{\rm ac}$ consists of two disjoint parts, $\Gamma_{\rm ac,1}$, $\Gamma_{\rm ac,2}$. (See the example in Section 5.4 for details on 2-port configuration). This yields a 2-degree of freedom system,

$$\begin{pmatrix} \hat{I}_{\rm ac,1} \\ \hat{I}_{\rm ac,2} \end{pmatrix} \;=\; \mathbf{Y}(\omega)\begin{pmatrix} \hat{V}_{\rm ac,1} \\ \hat{V}_{\rm ac,2} \end{pmatrix}, \tag{4.274}$$

$$\mathbf{Y}(\omega) \;=\; \mathbf{C}_{0,\rm eq} + \frac{i\omega}{k_{\rm eq}-\omega^2 m_{\rm eq}}\boldsymbol{\eta}_{\rm eq}\;. \tag{4.275}$$

To compute the transmission of this device (See Section 5.4.3 for details on the definition of transmission), one must attach a load resistor R_L to the boundary $\Gamma_{\mathrm{ac},1}$. This enforces the relationship,

$$-R_L \hat{I}_{\mathrm{ac},2} \quad = \quad \hat{V}_{\mathrm{ac},2}, \tag{4.276}$$

between the current and voltage. The transmission then becomes,

$$\text{Transmission} \quad = \quad 20 \log_{10} \left(\frac{-\mathbf{Y}_{21}}{1/R_L + \mathbf{Y}_{22}} \right). \tag{4.277}$$

4.7 Conclusion

In this section a general approach based on a variational framework has been introduced to derive equivalent circuit model parameters for electrostatically or piezoelectrically actuated high-frequency MEMS resonators. These equivalent circuit model parameters are used to model the behavior of the mechanical resonator as an electrical component in an electric circuit in the vicinity of a resonance frequency of the mechanical system. By replacing the mechanical resonator with these equivalent circuit model parameters, the computational expense for evaluating the response of an electrical circuit with a mechanical resonator component can be reduced significantly while retaining sufficient accuracy.

The classical approach taken for equivalent circuit model parameter extraction treats the electromechanical resonator as a combination of a purely mechanical resonator and a separate model incorporating the electromechanical coupling. The equivalent circuit model parameters are extracted with an assumed mode shape for the electric potential field based on engineering intuition and a Rayleigh-Ritz projection of the mechanical resonator system onto the purely mechanical modes of vibrations. This approach completely neglects the interaction that should occur between the electrical and mechanical modes. Additionally, for complex geometry the appropriate model required to represent the electromechanical coupling may not be clear.

The variational approach we present is based on defining the total mechanical and electrical energy of the system defined on an arbitrary domain. This allows one to treat electromechanical resonators with arbitrary geometry. The electrical energy includes both electrostatic and piezoelectric contributions. The formulation incorporates geometric nonlinearity which is crucial in modeling the electromechanical coupling which arises from electrostatically actuated resonators. Since the total electromechanically coupled system is treated, one does not have to worry about constructing the appropriate model to represent the electromechanical coupling. Numerical implementation through the finite element method is also straightforward due to its variational structure. From the dis-

cretized system of equations, one computes the electromechanically coupled mechanical and electric potential modes in the vicinity of a desired frequency. The electromechanically coupled system is then projected onto these coupled modes to obtain the equivalent circuit model parameters. It is shown that if one chooses to project the system, not onto the electromechanically coupled modes, but an assumed mode shape for the electrical potential field and purely mechanical mode shape, one recovers the classical approach. The ability of the equivalent circuit model parameters extracted from our method in accurately modeling the behavior of the dielectric transduced resonator is presented in Section 5.4.

Chapter 5

MEMS examples

5.1 Introduction

In this chapter, the technology that has been developed and presented in the previous chapters for efficient modeling and evaluation of damping in MEMS devices is employed in several numerical examples, to exhibit the effectiveness of the methods. The numerical simulations are conducted using the open-source software *HiQLab* [35]. *HiQLab* has been initiated by Bindel and members of the *SUGAR* [85] group, including myself, have made contributions in developing the software. Some of the contributions that I have made include but are not restricted to implementation of thermoelastic elements, electrical circuit elements, and interfaces to the parallel numerical libraries *Trilinos* [96] and *PETSc* [21]. A brief overview of *HiQLab* is presented in Section 5.2 along with details regarding the interface to the parallel iterative solver library *PETSc* [21], which can be used to solve linear systems on the order of millions and larger. This is followed by the simulation of three MEMS devices. The first in Section 5.3 is the disk resonator with which the damping mechanism of anchor loss is modeled through the application of PML. The method that has been developed in Chapter 2 to compute the quality factor Q from the complex-valued frequencies through the computation of a complex-symmetric generalized eigenvalue problem is presented. The simulations are conducted

on a parallel processor machine to observe the scalability of the proposed method in terms of the compute time for solving large linear systems. Next in Section 5.4 a dielectric transduced resonator is analyzed using the technology developed in Chapter 4 for electrostatic electromechanically coupled systems. The equivalent circuit parameters are extracted and the accuracy of the equivalent circuit model in modeling the behavior of the resonator near the resonance frequency is investigated. The last example in Section 5.5 is a thermoelastic beam resonator and ring resonator, with which the reduced order modeling technology for the thermoelastic problem presented in Chapter 3 is used to compute the transfer function. The accuracy of the reduced order model is investigated.

5.2 *HiQLab*, *PETSc*, and the Fortunato cluster

HiQLab

HiQLab [35] is a finite element tool for simulating resonant MEMS that David Bindel [36] has initiated and continues to develop in collaboration with our research group SUGAR [85] at the University of California, Berkeley. The software's aim is to be able to make available a tool for accurately modeling damping behavior in resonators, since the current widely available CAD tools are able to simulate the resonance frequencies but are not able to model the damping well. Simulation of damping high-frequency resonators is quite different from the modeling of damping in most structural applications where the degree of damping is not so crucial and standard Rayleigh type damping is sufficient. The design goal of high-frequency oscillating resonators, is to minimize the damping to obtain a high quality factor Q. In order to accurately model such small damping, one requires more accurate models of the source of damping. The determination of good damping models is not trivial, since damping phenomenon in general are complex phenomenon, and one must explore the applicability of various type of damping models. Parallel to incorporating accurate damping models, we are also interested in efficient algorithms to simulate damping phenomenon. Standardly available CAD tools do not allow one this type of flexibility. Such constraints have

motivated the development of *HiQLab*.

The current version of *HiQLab* supports the following standard features,

- 1D,2D,3D, and Axisymmetric analysis,
- Steady-state and static analysis,
- Linear elastodynamic and scalar wave problems,
- Thermoelastically and electromechanically coupled problems,

and special features including,

- Anchor loss modeling through PML,
- Thermoelastic damping modeling through the thermoelastically coupled problem,
- Efficient eigenfrequency computation for the thermoelastic problem based on a perturbation approach,
- Arnoldi based Reduced Order Modeling (ROM) for fast transfer function evaluation of elastic and thermoelastic problems.

HiQLab is written mainly in C++, with an interface to the scripting language *LUA* [103] and the popular commercial software *MATLAB* [175]. The interface through *LUA* has been selected for its light weightedness and speed, as well as its open source availability. The interface through *MATLAB*, despite the commercial software license required for its use, enables one full access to its rich numerical libraries and plotting capabilities.

The initial version of *HiQLab* has been developed as a serial code. With this serial version, simulations of axisymmetric simulations of anchor loss in disk resonators and evaluation of thermoelastic damping in beam resonators have been successfully conducted [37, 114]. In the simulation of the disk resonators, the experimental results of a class of resonators had large discrepancies with the simulated results from *HiQLab*. This difference was assumed to arise from anchor misalignment

of the post structure. The verification of this claim cannot be conducted by axisymmetric simulations, and a fully 3D simulation was strongly required. As it has been mentioned in Chapter 2, the modeling of anchor loss through PML can lead to a large linear system of equations that must be solved. The size of these systems for 3D discretizations can be on the order of millions of degrees of freedom, which is no longer feasible to solve on a single processor machine with a serial code. This has initiated the development of the extension of *HiQLab* to a parallel code that can be run on multiple processor machines. (The simulation of the disk resonators mentioned here are explained in more detail along with the results for anchor misalignment in Section 5.3).

PETSc

In order to solve large linear systems on multiple processor machines through parallel computing, an interface to the library *PETSc* [21] has been developed. *PETSc* is a suite of data structures (e.g.,vectors and matrices) and routines (e.g.,iterative linear solution methods and interfaces to external direct linear solution methods) built on top of the MPI standard [141] for facilitation of scalable parallel computing. By using this library, one does not have to manage the low level details of parallel computing and can manipulate linear algebra objects such as vectors and matrices directly. In the parallel version of *HiQLab*, data for the linear system is generated by *HiQLab*, and stored in *PETSc* matrices and vectors for the solution of linear systems.

The details for the solution of linear systems with the proposed geometric multigrid method as a preconditioner to an iterative method are as follows.

1. Serially construct the required data structure required for the system with *HiQLab*.
2. Serially partition the data structure with *METIS* [107], a graph partitioning software, for efficient data distribution between the parallel processes.
3. In parallel read in the data structure on each process and construct the matrices (including coarse grid operators and prolongation operators) and vectors in *PETSc* format.

4. Solve the linear system with a geometric multigrid preconditioned GMRES. For the smoother, we can select between the *PETSc* implemented Gauss-Seidel smoother and our implementation of the Kaczmarz smoother based on component-averaged row projections [82]. The Chebyshev smoother of Chapter 2 cannot be selected since no *PETSc* implementation currently exists. The coarse grid direct solve was conducted by the packages *MUMPS* [11] or *SUPERLU_DIST* [62].

To implement this procedure, one has to take into account efficiency and memory available for each step. The above procedure was selected due to the following design criterion.

- In preliminary attempts, the data structure of the system was not partitioned across processes, i.e., excluding Steps 1. and 2. This produced bottlenecks at the matrix construction in Step 3. As the size of the linear system increased, the total size of the data structure increased proportionally, requiring more and more redundant information to be stored on each process. Thus the preliminary step of serially partitioning the data structure was introduced.

  Still this method has its bottleneck, as is exhibited in Section 5.3 in the disk resonator example. In Step 2, the serial partitioner *METIS* is invoked. In the stable distributed version *METIS* 4.0, there is a restriction on the size of the maximal size of the graph that can be partitioned, due to the index type of 4 byte integers, restricting the size of arrays possible. The version *METIS* 5.0 currently under development, has removed this restriction by implementing `intg64_t` as the array indexing type, but software compatibility with *ParMETIS* 3.1 [108] is still not clear and thus we have selected the use of *METIS* 4.0 for now.

  To increase the size of the system to be partitioned, one can invoke *ParMETIS* on a lexically partitioned data structure of the system, or implement *METIS* 5.0.

To incorporate the geometric multgrid method introduced in Section 2.3 as a preconditioner, the following components had to be developed.

- Efficient construction of prolongation operators between the grids.

The prolongation operator is constructed from the method introduced in Section 2.3.2. The scalability of this scheme is shown in the disk resonator example of Section 5.3.

- Choice of smoother.

  An implementation of the Kaczmarz smoother based on component-averaged row projections [82] has been implemented. It has been shown in Section 2.3.3 that the Kaczmarz smoother performs non-optimally for 3D elastodynamic problems, with slow convergence. Thus even though this smoother has been implemented the convergence rate of the overall multigrid method was slow, as was the implementation in terms of time required for each application. Thus results for this smoother are not presented and only those of the Gauss-Seidel smoother are presented. The Chebyshev smoother of Chapter 2 cannot be selected since it has not been implemented yet in *PETSc*.

The overall multigrid iteration (e.g., multigrid V-cycle) of projecting residuals and interpolating errors, and smoothing as well as coarse grid solves are handled by the preconditioner structure for multigrid `PCMG` supported by *PETSc*.

For the computation of eigenvalues through the geometric multigrid preconditioned Jacobi-Davidson QZ method introduced in Section 2.4, the eigenvalue solver was implemented from scratch utilizing the linear algebra objects from *PETSc*. The correction equation was solved using the GMRES implemented in *PETSc* with our implementation of geometric multigrid as the preconditioner. The steps involved in the whole eigensolve procedure differ from the process of the linear system solve procedure introduced above only in Step 4 which is replaced with an eigensolve.

Fortunato cluster

To solve the linear system or compute the eigenvalues in parallel with the combined *HiQLab* and *PETSc* setup, the program must be run on a multiple processor machine. Our simulations were conducted on the Fortunato cluster located at the Swiss Federal Institute of Technology (ETH) in

Zurich, Switzerland. The specifications of the cluster are summarized below.

Fortunato Cluster specifications

- 16 compute nodes. Each with two dual-core AMD Opteron 2216 processors (theoretical peak performance of 9.6 Gflops), for a total of 64 cores (theoretical total peak performance of 9.6 Gflops per dual core processor $\times$ 2 processors per node $\times$ 16 nodes = 307 Gflops). Each compute node has 16GB of memory for a total of 256GB.
- Nodes are connected via a Quadrics QsNet network which which delivers extreme performance to communication-intensive applications. (14.4 Gbytes/s bisectional bandwidth).

5.3 Disk resonator

In this section, the class of contour-mode vibrating disk resonators are examined. These devices have been fabricated and tested yielding high quality factors Q at frequencies in the MHz to GHz range [188, 187, 189]. Among the many modes possible, the radial contour modes are of interest due to their high frequency. A 3D model of the device geometry along with a schematic diagram of the vertical cross section through the center post is shown in Figure 5.1. The disk resonator consists of two parts, a disk with radius r_{disk} and thickness t_{disk} and a post with radius r_{post} and height h_{post}. By varying the material and geometry of the disk, various resonance frequencies and modes are possible. A good approximation of these frequencies and modes can be obtained through a 2D analysis, but these do not give an idea of how the 2D planar motion couples in motion with the underlying substrate though the post. Proper understanding of the coupling motion between the disk and underlying substrate is crucial in designing devices with high Q values. It has been mentioned in the work of Bindel [33] and Hao and Ayazi [89] that the coupling of the 2D planar motion to the out-of-plane direction through the Poisson ratio, excites motion at the post leading to energy loss. It has also been pointed out and experimentally verified that a coupling between the 2D planar modes and out-of-planes modes can lead to increase and decrease in Q. The axisymmetric simulations conducted by Bindel have been able to predict the quality factor for disks with large radii but have failed for those with smaller radii. The cause of this has been attributed to a potential asymmetry induced by a non-central placement of the post, which may lead to degradation in the mode and Q. The claim that misalignment of the post can lead to Q degradation is further supported in the experiments conducted by Wang et al. [189]. In their experiments, using their post alignment technique, disks with posts intentionally misaligned were tested to yield a lowering of the Q value with increasing misalignment. The explanation given is purely qualitative, such that post misalignment leads to degradation of the mode. Additionally to these experiments, the effect of post thickness has also been tested to yield a result of an increase of Q with decrease in the radius

of the post. In order to investigate the effect of anchor misalignment, one must resort to full 3D calculations.

The effect of anchor misalignment to the quality factor Q is investigated using the geometric multigrid method combined with the Jacobi-Davidson QZ (JDQZ) eigensolver. Before the evaluation of Q, the effectiveness of the geometric multigrid method as a good preconditioner to an iterative method is first verified. This is followed by verification of the performance of the geometric multigrid combined JDQZ eigensolver. Finally the results for the sensitivity of Q with respect to anchor misalignment is presented.

Resonator 3D model

Resonator cross section

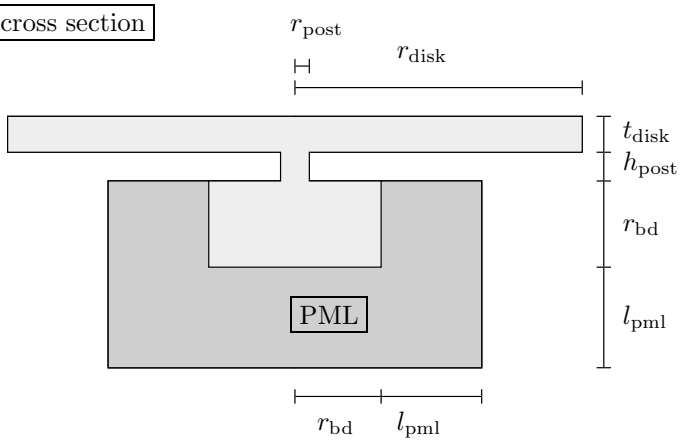


Figure 5.1: 3D model of the disk resonator and cross section

Table 5.1: The 4 levels constructed for the geometric multigrid preconditioned GMRES iteration

Level	n_{dense}	Number of DOFS	$n_{\text{npw,s}}$ at 715.7[MHz]	$n_{\text{npw,s}}$ at 1.140[GHz]
1	1.25	49938	8.69	5.45
2	2.5	197574	17.4	10.9
3	5	977115	34.8	21.8
4	10	6140520	69.5	43.6

5.3.1 Geometric multigrid preconditioned GMRES iteration

The performance of the geometric multigrid method is examined through its application as a preconditioner to GMRES. The disk resonator shown in Figure 5.1 with parameters,

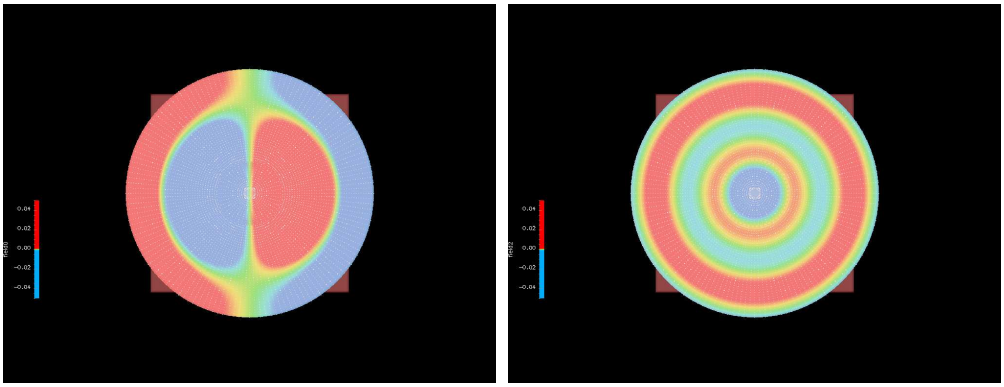
$$\begin{aligned} E: &\quad 150 \quad [\text{GPa}] \\ \rho: &\quad 2330 \quad [\text{kg/m}^3] \\ \nu: &\quad 0.3 \quad [\text{-}] \\ r_{\text{disk}}: &\quad 10 \quad [\mu\text{m}] \\ t_{\text{disk}}: &\quad 2 \quad [\mu\text{m}] \\ r_{\text{post}}: &\quad 1 \quad [\mu\text{m}] \\ h_{\text{post}}: &\quad 0.5 \quad [\mu\text{m}] \\ l_{\text{bd}}: &\quad 2 \quad [\mu\text{m}] \\ l_{\text{pml}}: &\quad 6 \quad [\mu\text{m}] \end{aligned}$$

which is identical to the disk analyzed by Bindel [37] in the axisymmetric case. The device is forced at three frequencies, {0, 715.7, 1140}[MHz]. These frequencies correspond to the static case, 2nd radial contour mode, and 3rd radial contour mode. To give an idea of the radial contour mode, the 2nd radial contour mode of this disk resonator is shown in Figure 5.2. The red denotes positive displacement, and the blue denotes negative displacement. Since the 2nd mode of vibration is presented, one can see one node in the radial displacement across the disk. Note that as the disk resonates there is a coupled z direction bending type of motion. It has been mentioned that the constructive or destructive interaction between these modes can vary the obtained quality factor by orders of magnitudes [37].

The smoother in the geometric multigrid method is restricted to the Gauss-Seidel smoother, since our implementation of the Kaczmarz smoother yielded inadequate convergence rates and time in its application, and the Chebyshev smoother has not yet been implemented in *PETSc*. Several levels of the multigrid have been applied to observe mesh independent convergence behavior of the

Top view: x displacement

Top view: z displacement



Side view: x displacement

Side view: z displacement

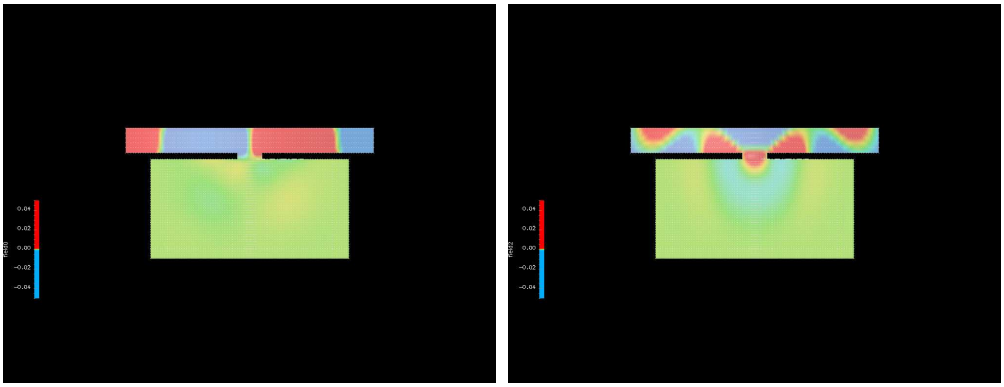


Figure 5.2: The 2nd radial contour mode shape of the 10 μm radius disk resonator

method. 4 grids are constructed with linear element discretization. The 4 levels constructed, and corresponding number of degrees of freedom for each level is summarized in Table 5.1. n_{dense} is defined as the approximate number of nodes within a distance of 1 μm, and is a measure of the discretization. One sees that as the discretization is doubled, the number of degrees of freedom in the mesh increase by approximately 8. $n_{\text{npw,s}}$ is the number of nodes per discretized shear wave, with shear wave length $\lambda_{\text{s}} = 13.0$ μm at a frequency of 715.7 [MHz] for this material. The discretization of the coarse grid was selected so that an adequate representation of the shear wave, which is the

shortest wave in the system, is close to 8 nodes per wave length for the frequency of 715.7[MHz]. As can be seen from Table 5.1, this results in a discretization of the shear wave on the coarsest grid smaller than 8 for the frequency of 1.140[GHz]. This low discretization leads to slower convergence in the multigrid method, as is observed in the simulations. The PML profile of this problem is selected as linear $\gamma(s) = s$.

The geometric multigrid preconditioned GMRES is applied to the linear system of equations presented in Equation 2.101 (on page 68) with a random right hand side replacing the forcing vector $\mathbf{F}$. The number of iterations required for the iterative method to reach a preconditioned residual of 1×10^{-10} is presented in in Tables 5.2- 5.4, for varying forcing frequency, number of levels, and varying PML parameter β. The attained actual residual is presented in parentheses next to the number of iterations required for GMRES convergence. For an ideal multigrid method, the number of iterations required to reach this tolerance should not increase with the number of levels used. The following comments can be made.

- For $\beta = 0$ and $\beta = 1$, the number of iterations for convergence is independent of the number of levels used. The performance of the multigrid method deteriorates when β is larger than 1. This complies with the observations made in Section 2.3.3, regarding the selection of PML parameters for the Gauss-Seidel smoother.

- Due to the coarse level-1 discretization with less than 8 nodes per shear wave length, the number of iterations for the forcing frequency at 1.14[GHz] is large compared to the case of 715.7[MHz].

- The number of iterations for the forcing frequency of 1.14[GHz] is smaller for $\beta = 1$ than $\beta = 0$. This is due to the added shift in the imaginary part of the problematic eigenvalues, improving the convergence rate, as mentioned in Section 2.3.3.

From the obtained results one can state that for ideal multigrid convergence, the PML parameter β should be selected at largest 1 for a linear absorbing function profile $\gamma(s) = s$. The discretization of

the coarsest grid should not be made too coarse either.

The scaling behavior of the geometric multigrid with respect to the number of processes is investigated. The geometric multigrid preconditioned GMRES solver is run for the 4 level case at the frequency 715.7[MHz] on 8, 16, 32, and 64 processes. When selecting the number of processes to run, one can also choose how many processes are run per node, with a maximum of 4 processes per node. This limit exists since the nodes have 2 Dual-core processors per node, resulting in a total of 4 cores. The total time in seconds spent for the solution, including the mesh generation, grid operator construction, prolongator construction, and iterative linear solve, is presented in Figure 5.3 for the possible combinations of number of processes and how many cores per node are used. In general one sees very good scaling since the solution time is approximately halved with a doubling of the number of processes. The time required for solution is the largest when all 4 cores of the node are used. This is understandable from the architecture of the Dual-core AMD Opteron processor. The two cores on each Dual-core processor share the same bus to the memory. Thus when the two cores are used at once, only half the bus can be allocated to each core on average. On the other hand, when only two cores which come from separate Dual-core processors on the same node are used at once, they have separate buses to the memory, resulting in no difference to using one core per node. With this in mind, one should be able to predict identical performance for the 1 core per node and 2 core per node case, for 8 and 16 processes. It is interesting to see that faster results were obtained for the 2 core per node case for 16 processes. This may be due to the MPI optimizing performance by somehow knowing that the two cores on the same node do not have to communicate with each other through the Quadrics interconnect, but internally within the node.

The speedup with respect to the number of processors is presented in Figure 5.4. The fastest time in Figure 5.3 has been chosen for each number of processors as the representative time. The sequential time has been chosen as 8 times that of the 8 processor case. Though one does not have perfect speedup, one observes fairly good increase in performance with respect to increasing number of processors.

Table 5.2: GMRES iterations required to obtain a preconditioned residual of 1×10^{-10} for $\beta = 0$

Levels	1-2	1-3	1-4
0[MHz]	13(2.3e-07)	10(2.0e-07)	9(3.7e-08)
715.7[MHz](2nd mode)	39(1.0e-09)	42(2.5e-09)	38(5.0e-09)
1.140[MHz](3nd mode)	81(8.1e-10)	98(1.2e-09)	94(8.1e-09)

Table 5.3: GMRES iterations required to obtain a preconditioned residual of 1×10^{-10} for $\beta = 1.0$

Levels	1-2	1-3	1-4
0[MHz]	22(3.8e-07)	18(2.6e-07)	13(1.6e-07)
715.7[MHz](2nd mode)	49(1.2e-09)	51(1.6e-09)	41(3.3e-09)
1.140[MHz](3nd mode)	71(9.0e-10)	78(4.3e-10)	72(7.9e-10)

Table 5.4: GMRES iterations required to obtain a preconditioned residual of 1×10^{-10} for $\beta = 1.2$

Levels	1-2	1-3	1-4
0[MHz]	28(3.4e-07)	38(1.1e-06)	101(2.6e-05)
715.7[MHz](2nd mode)	58(1.8e-09)	85(5.7e-09)	160(2.0e-07)
1.140[MHz](3nd mode)	89(1.0e-09)	118(2.8e-09)	195(5.1e-08)

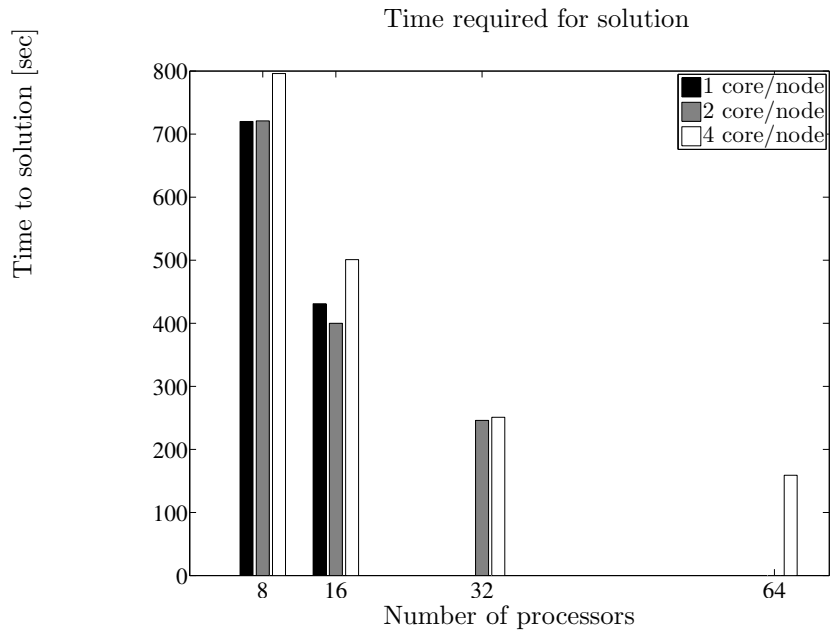


Figure 5.3: Scaling with respect to number of processors for the solution of the disk resonator, 6 million degrees of freedom, linear elements

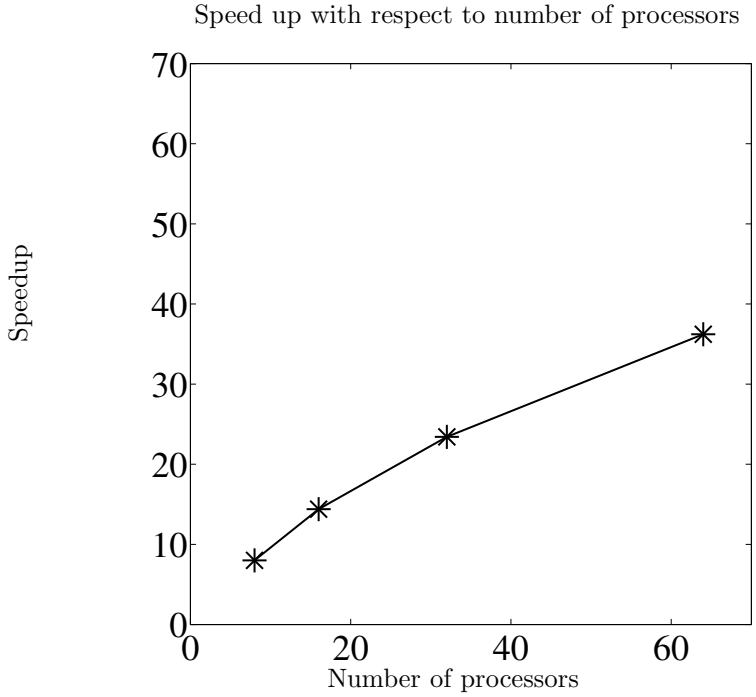


Figure 5.4: Speedup with respect to number of processors for the solution of the disk resonator, 6 million degrees of freedom, linear elements. The fastest time for each number of processors has been chosen from Figure 5.4

5.3.2 Geometric multigrid combined JDQZ

To present the ability of the geometric multigrid combined JDQZ method in computing eigenvalues, the same disk resonator presented in the previous section is examined. The eigenfrequency and corresponding quality factor Q of the 2nd radial contour mode at approximately 715 [MHz] is computed.

As is observed in the previous section, due to the geometric multigrid method, a restriction is placed on the selection of PML parameters. For the linear absorbing function profile $\gamma(s) = s$, β is restricted to a maximum of 1. Additionally due to computational constraints, there is also a restriction on how large one can make the computational domain, i.e., the number of waves in the PML n_{wpml}. For our 3D simulations, the length of the PML is selected as $l_{\text{pml}} = 6\mu$m, which results in $n_{\text{wpml,v}} = 6/13$, since the wave length of the volumetric wave is $\lambda_v = 13$ μm. For the pair $(\beta, n_{\text{pml,v}}) = (1, 6/13)$, a linear absorbing function profile leads to an end termination reflection of $\log_{10}(r_{\text{end}}) = -1.26$.

From the results for the 1D problem presented in Section 2.4.1, regarding the relationship between the reflection (approximately equal to the energy dissipation error) and quality factor Q, we can infer that the quality factor will be accurate to within 1 digit of accuracy, as long as the mesh is fine enough to capture the energy dissipation mechanism.

The 3D simulations are computed using linear, quadratic, and cubic elements with mesh discretization n_{dense} varying from 1 to 10. n_{dense} is the approximate number of nodes in a distance of 1 μm. The limit of 10 arises from the limit of the size of a graph that the *METIS* serial graph partitioner can handle with quadratic element discretization. The size of the graph at this limit is on the order of 8 million. For cubic elements, the number of nonzeros exceeded 1.5 billion for $n_{\text{dense}} \approx 7$ with a graph of size approximately 4 million, which was the limit to which *METIS* could partition. For comparison with the 3D results, the same configuration is also simulated under axisymmetric analysis. The quality factor for the 3D case with respect to the number of degrees of freedom is presented in Figure 5.5. The frequency and quality factor for both the 3D case and axisymmetric

case with respect to n_{dense} is presented in Figure 5.6. The following comments can be made about the results from the 3D simulations.

- For this device and mode of vibration, the mesh discretization for the 3D simulation is not fine enough to fully capture the energy dissipation mechanism. This is observed from the comparison with axisymmetric results. Though one can still see that the obtained value of Q is accurate to within the order of magnitude. With a parallel graph partitioner and more memory, the accuracy could be driven to convergence. For a given n_{dense}, the Q obtained from 3D analysis resembles those obtained from the axisymmetric analysis.

- The frequency converges quite rapidly analogously to the axisymmetric case. This is because the disk part of the resonator determines the frequency and not the energy dissipation mechanism.

- Cubic elements better model the energy dissipation mechanism compared to linear and quadratic elements, analogously to the axisymmetric case.

Since a major part of the time of the JDQZ eigensolver is spent in solving the correction equation, the fast solution of linear systems with the geometric multigrid preconditioned GMRES iteration will lead to a fast eigenvalue evaluation. The speed of convergence of the desired eigenvalue depends on how close the initial approximation of the eigenvalue is to the desired eigenvalue. The same can be said with the initial vector in the JDQZ eigensolver iteration. We have observed that with a moderately good initial eigenvalue approximation, convergence to the desired eigenvalue can be obtained within 10 steps, as long as the correction equations are solved to approximately 8 digits of accuracy. By using the coarser grid eigenvalue and eigenvectors as the initial approximations, as is explained in Section 2.4.2, the JDQZ eigensolver converges in 1-2 steps, which is a huge gain in efficiency.

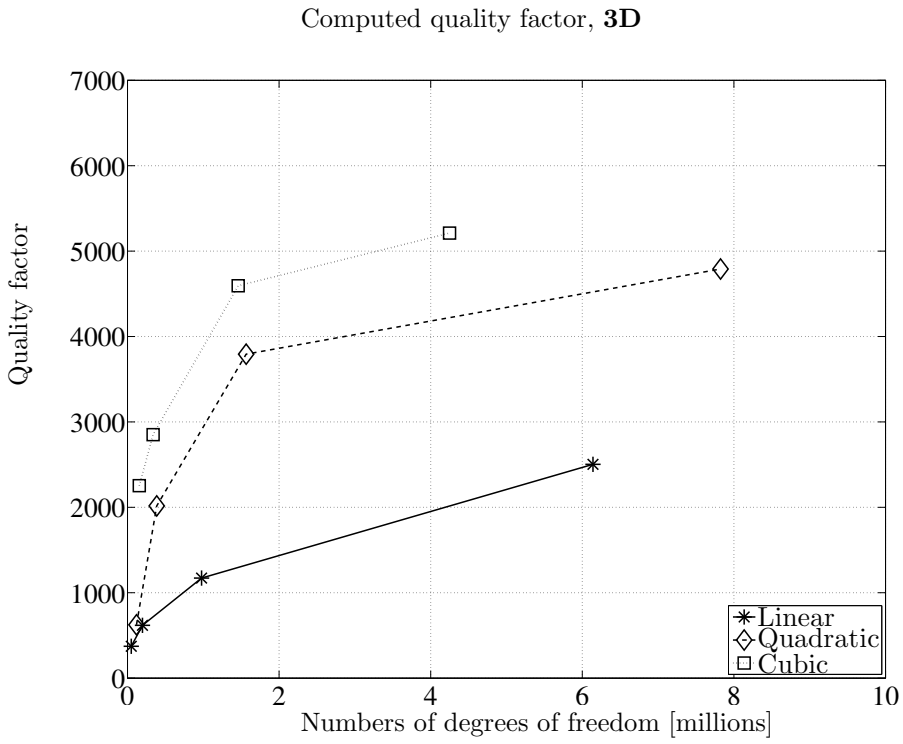


Figure 5.5: Convergence of Q with respect to the number of degrees of freedom

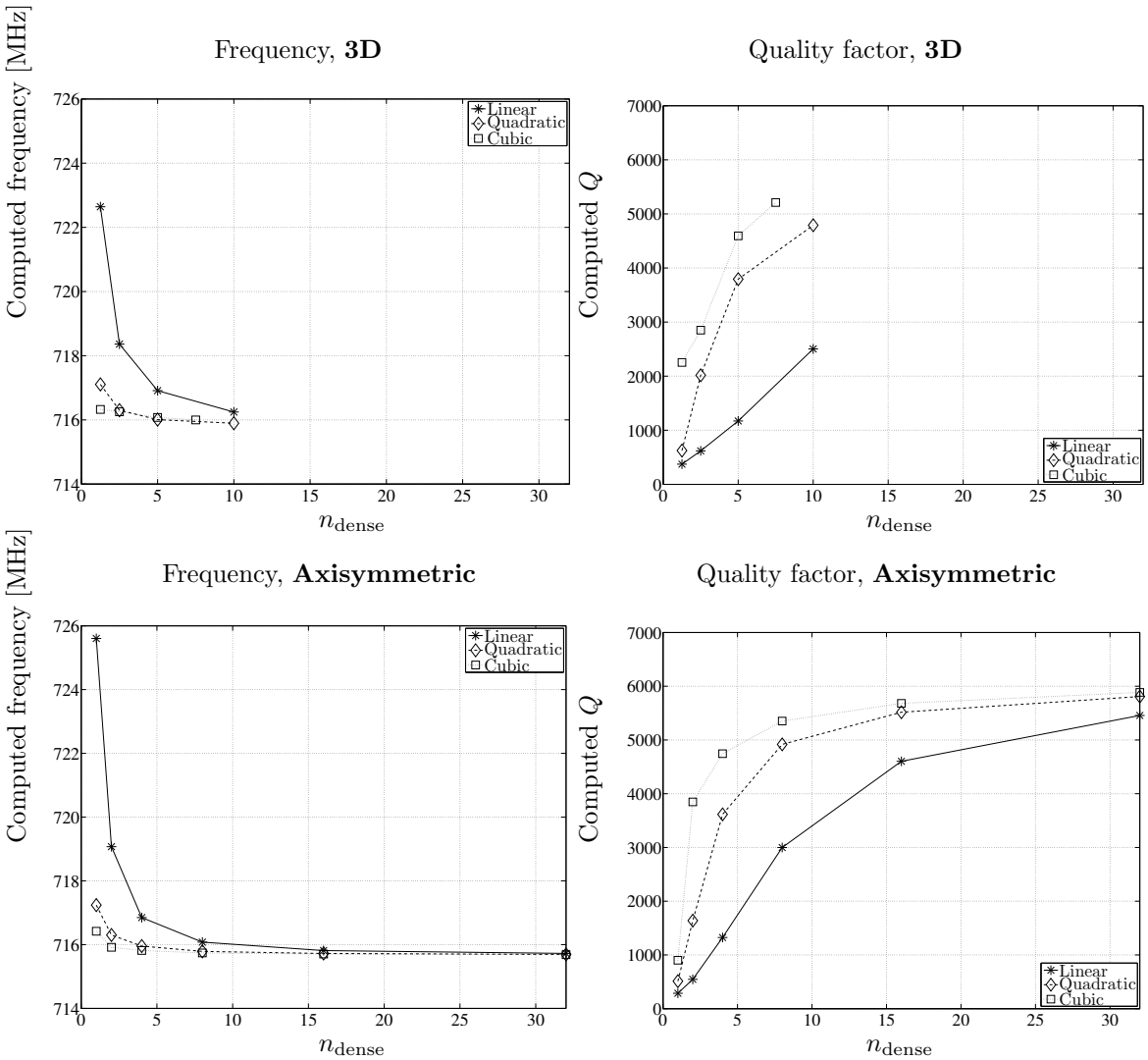


Figure 5.6: Convergence of frequency and quality factor with $(\beta, l_{\mathrm{bd}}, l_{\mathrm{pml}}) = (1, 2\ \mu\mathrm{m}, 6\ \mu\mathrm{m})$, for axisymmetric and 3D analysis

5.3.3 Sensitivity of Q with respect to post geometry

In this section the effect of post geometry on the frequency and quality factor Q are investigated. The disk resonators fabricated by Wang et al. [189] are simulated and compared with the experimentally obtained results. The parameters for the disk resonators are,

E :	150	[GPa]
ρ :	2300	[kg/m^3]
ν :	0.226	[-]
r_{disk}:	{10,18}	[μm]
t_{disk}:	2.1	[μm]
r_{post}:	{0.8,1.0}	[μm]
h_{post}:	0.8	[μm]
l_{bd} :	2	[μm]
l_{pml} :	6	[μm]

The configuration of the disk is shown in Figure 5.3. The simulations are conducted with cubic elements and $n_{\text{dense}} = 7.5$, which should yield the most accurate results for frequency and quality factor Q, subject to the constraints of limited computational resources. A linear absorbing function profile $\gamma(s) = s$ with $\beta = 1$ is selected. For the disk resonators with $r_{\text{disk}} = 10$ μm, the 1st, 2nd, and 3rd modes are computed, and for the disk resonators with $r_{\text{disk}} = 18$ μm, the 2nd and 3rd modes are computed. These restrictions are set from the capability of the PML in properly absorbing the outgoing waves. As the frequency becomes lower, the wave length becomes larger, resulting in a smaller number of waves in the PML layer when the thickness of the PML l_{pml} is fixed. This leads to less absorption. The end termination reflection for the computed frequencies are presented in Table 5.5. For this small value of β, one does not need to worry about the effect of interface reflection, as has been shown in Section 2.2. Since the end termination reflection is fairly large for the low frequency modes, one can only assume that the order of magnitude of Q will be correct. The frequency and Q computed by varying the misalignment of the post measured from the center, from 0 to 0.5 μm is presented in Tables 5.6- 5.9. One can make the following comments on the obtained results.

- Anchor misalignment does not affect the frequency at all. This is contrary to the results obtained experimentally, where the frequency increases with anchor misalignment. One could

attribute the variation in experimentally measured frequency to experimental variation, but we have no method of confirming this.

- Anchor misalignment affects Q only mildly, and not as drastically as with the experimental values where an order of magnitude difference is observed. Strangely, the Q for the 18 μm disk in its 3rd mode increases with increase in misalignment.

The mode shape computed for the 18μm disk with anchor misalignment of 0.5μm resonating in its 2nd mode is presented in Figure 5.7. Red denotes positive displacement and blue negative. One observes from the top view of the disk resonator, that the radial displacement mode shape does not change at all. The computed frequency is mostly defined by the type of motion in the disk portion of the resonator. Since there is no change in the computed frequency with post misalignment, it is expected that the mode shapes do not change either. The zero change in frequency with increasing post misalignment infers this. The z direction displacement mode does change slightly. By looking at the x direction from the side, one sees large horizontal movement in the post. With a small post misalignment, the radial contour mode of the disk does not change at all, and the more compliant stem essentially tags onto the motion of the disk and moves horizontally.

As it has been shown in Section 5.3.2, the convergence of the frequency can be attained quite easily. Since the frequency of the resonator is mostly defined by the disk portion of the resonator, sufficient discretization of the disk leads to accurate results in the frequency. In the simulations results we have presented here, though the quality factors may not be fully converged, we can believe that the frequencies are converged due to their lack of change with respect to increase in post alignment, i.e., if the frequencies were not converged one should see some change in the frequency with respect to change in geometry.

One possible cause for the large difference in behavior of frequency and quality factor Q can be attributed to the nonlinear effects induced by the DC bias voltage required for electromechanical coupling. Another cause can arise from the simplified geometry that we have assumed for the

computational simulations. The experimentally fabricated structures have a post which extends from the substrate up through the disk sticking out from the top surface of the disk. This elongated post is due to the fabrication process used to fabricate the structures. We have neglected this portion under the engineering assumption that the effect is negligible. It has been noted in the experiments that anchor misalignment has lead to smaller pull-in DC bias voltages, which could be due to asymmetry induced by the anchor misalignment. Such behavior could also have an effect on electrical stiffness which can change the resonance frequency slightly. The part of the post that extends above the fabricated disks may contribute to a non-negligible inertial effects when they are placed off-center, leading to a change in resonance mode and frequency.

An additional axisymmetric simulation has been conducted to investigate the change in quality factor Q with respect to the change in post radius. The quality factor of a disk with radius $r_{\text{disk}} = 18\ \mu$m with varying post radius $\{1.0, 0.9, 0.9\}\ \mu$m has been computed. The results are presented in Table 5.10. The obtained quality factors match the trend observed in the experiments. With regards to the frequency, again there is no change at all in the simulations contrary to the experimental results.

Table 5.5: End reflection for PML with $\beta = 1$ and $l_{\text{pml}} = 6\ \mu$m

Disk	Mode	Frequency[MHz]	λ_v	$n_{\text{wpml,v}}$	$\log_{10}(r_{\text{end}})$
$r_{\text{disk}} = 10\mu$m	1st mode	264.0	32.8	0.183	-0.499
$r_{\text{post}} = 0.8\mu$m	2nd mode	706.5	12.3	0.488	-1.33
	3rd mode	1113	7.79	0.770	-2.10
$r_{\text{disk}} = 18\mu$m	2nd mode	393.8	22.0	0.273	-0.745
$r_{\text{post}} = 1.0\mu$m	3rd mode	626.1	13.8	0.435	-1.19

Table 5.6: Disks with radius 10 μm, frequency[MHz]

Misalignment	0 [μm]		0.10[μm]		0.25[μm]	0.50[μm]
	sim.	exp.	sim.	exp.	sim.	sim.
1st mode	264.0	273.6	-	274.0	264.0	264.0
2nd mode	706.5	734.6	-	736.0	706.5	706.5
3rd mode	1113	-	-	-	1113	1113

Table 5.7: Disks with radius 10 μm, Quality factor

Misalignment	0 [μm]		0.10[μm]		0.25[μm]	0.50[μm]
	sim.	exp.	sim.	exp.	sim.	sim.
1st mode	2.132×10^4	9.750×10^3	-	2.020×10^3	2.047×10^4	1.823×10^4
2nd mode	2.914×10^3	7.890×10^3	-	9.80×10^2	2.738×10^3	2.384×10^3
3rd mode	4.415×10^2	-	-	-	4.401×10^2	4.394×10^2

Table 5.8: Disks with radius 18 μm, frequency[MHz]

Misalignment	0 [μm]		0.25[μm]		0.50[μm]	
	sim.	exp.	sim.	exp.	sim.	exp.
2nd mode	393.8	405.2	393.8	405.8	393.9	407.1
3rd mode	626.1	645.0	626.1	646.6	626.1	-

Table 5.9: Disks with radius 18 μm, Quality factor

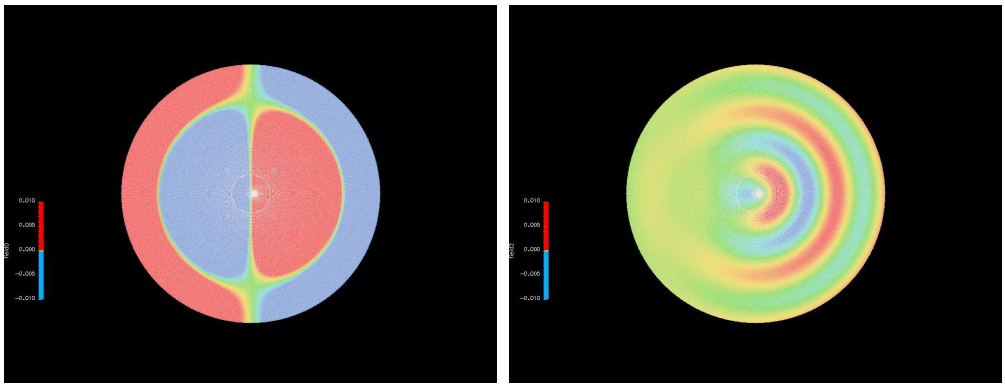
Misalignment	0 [μm]		0.25[μm]		0.50[μm]	
	sim.	exp.	sim.	exp.	sim.	exp.
2nd mode	7.900×10^3	5.551×10^3	7.429×10^3	1.460×10^3	6.325×10^3	6.05×10^2
3rd mode	2.016×10^3	5.730×10^3	2.136×10^3	1.078×10^3	2.474×10^3	-

Table 5.10: Disks with radius 18 μm, 2nd mode

Radius of post	1.0[μm]		0.9[μm]		0.8[μm]	
	sim.	exp.	sim.	exp.	sim.	exp.
Q	8.482×10^3	5.551×10^3	1.233×10^4	1.174×10^4	1.842×10^4	1.466×10^4
Frequency[MHz]	393.7	405.2	393.6	404.6	393.6	404.1

Top view: x displacement

Top view: z displacement



Side view: x displacement

Side view: z displacement

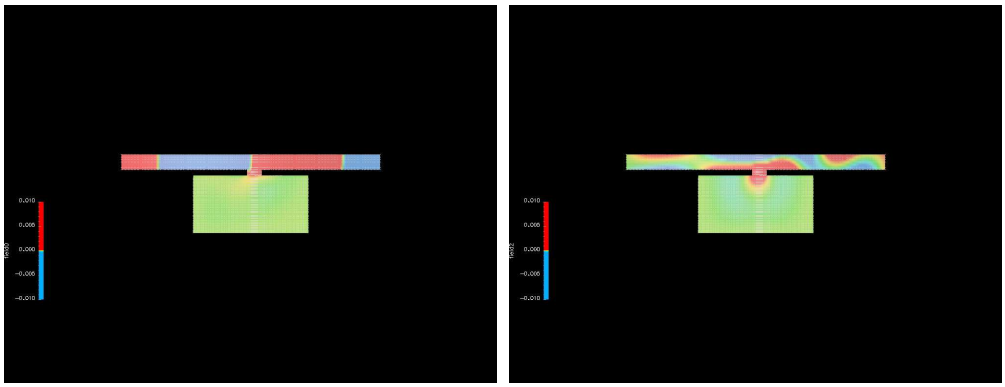


Figure 5.7: The 2nd radial contour mode shape of the 18 μm radius disk resonator with a post misalignment of 0.5 μm

5.4 Dielectric transduced resonator

In this section a resonator forced into motion by internal dielectric transduction is presented to illustrate the applicability of efficient reduced modeling of electromechanically coupled systems through equivalent circuit parameter extraction.

In Section 4.2, the mechanism of electrostatic transduction through a time varying AC voltage superposed onto a DC bias voltage was introduced. The efficiency of electromechanical coupling across an electrostatic gap can be approximated by the parallel plate assumption by,

$$\eta \approx \frac{\epsilon_r \epsilon_0 A V_{\rm dc}}{g_0^2}, \tag{5.1}$$

where ϵ_0 is the permittivity of vacuum, ϵ_r is the relative permittivity of the material in the electrostatic gap, A is the surface area of the parallel plates, g_0 is the gap distance, and $V_{\rm dc}$ is the DC bias voltage. In initial designs of electrostatically actuated resonators, the gap was filled with air, with the idea of reducing the coupling between the resonator and surrounding environment for less energy loss and higher quality factor [189]. With air filled gaps, high electromechanical coupling η was only possible by decreasing the gap g_0 and increasing the DC bias voltage $V_{\rm dc}$. But such a method is limited in the applicable $V_{\rm dc}$ and g_0 due to a pull-in voltage instability or breakdown. The possible gap size g_0 is also limited by the fabrication process, since these air gaps are opened by a chemical release removal process of the sacrificial oxide originally filling the air gaps. With smaller gaps, the sacrificial oxide becomes difficult to remove due to infinite diffusion times required for the etchant to reach the oxide [131].

A new idea introduced to circumvent this problem is to fill the gaps with dielectric material [131, 100, 48, 49]. Dielectric filled gaps can increase the electromechanical coupling by increasing the relative permittivity ϵ_r and by allowing smaller gaps, since a pull-in voltage no longer exists. The applicable DC bias voltage $V_{\rm dc}$ is then limited by dielectric breakdown, which is on the same order as the pull-in voltage for air gaps, so an overall increase in electromechanical coupling η can be obtained. Dielectrically filled gaps also have the advantage of increasing the yield in the fabrication

process by eliminating the sacrificial oxide layer removal step.

The insertion of dielectric material in the gaps however, leads to a coupling between the resonator and actuating electrodes, which did not exist in the case of air gaps. One approach to decreasing the coupling is to place the dielectrically filled gaps at the nodes of the vibrational mode of the resonator [131, 100]. Another approach is to incorporate the placement of the dielectric gap directly into the mode shape of the resonator [49]. In either case, an analysis incorporating the resonator, the dielectric filled gaps, and the electrodes are required for estimating the behavior of these devices. The variational framework presented in Section 4.3 can be applied to these devices to estimate the quality factor Q and frequency as well as the equivalent circuit parameters one desires for efficient simulations.

To illustrate the applicability of the proposed method, the 2D planar resonator depicted in Figure 5.8 is analyzed. The width of the resonator in the x direction is 12[μm], the length in the y direction is 90.4[μm], and thickness in the z direction is 2[μm]. The resonator consists of three parts, electrically separated by two strips of dielectric material, but mechanically coupled. These dielectric strips are 0.025[μm] thick. The resonator connects to the substrate through six serpentine anchors. which also serve as a path for voltage excitation. The resonator is actuated by applying a DC bias voltage on the center piece and applying a time varying voltage across the dielectric gap. The device can be actuated either in a single port configuration or two port configuration depending on the circuitry surrounding the resonator. The location of the anchors were determined by placing them at the maximal strain locations, which are the anti-nodes of a vibrational mode. The vibrational mode that this device is designed for is a y directional longitudinal mode. The mode shape is clarified in the eigenfrequency and mode analysis presented. The resonator is fabricated from Poly-Silicon, and the dielectric material is set as Hafnium Oxide with relative permittivity of 25.

First, the mode shape and corresponding quality factor of the mode of interest is computed and presented. This is followed by equivalent circuit parameter extraction by the two methods proposed in Section 4.6, and their accuracy is investigated by comparing the admittance obtained

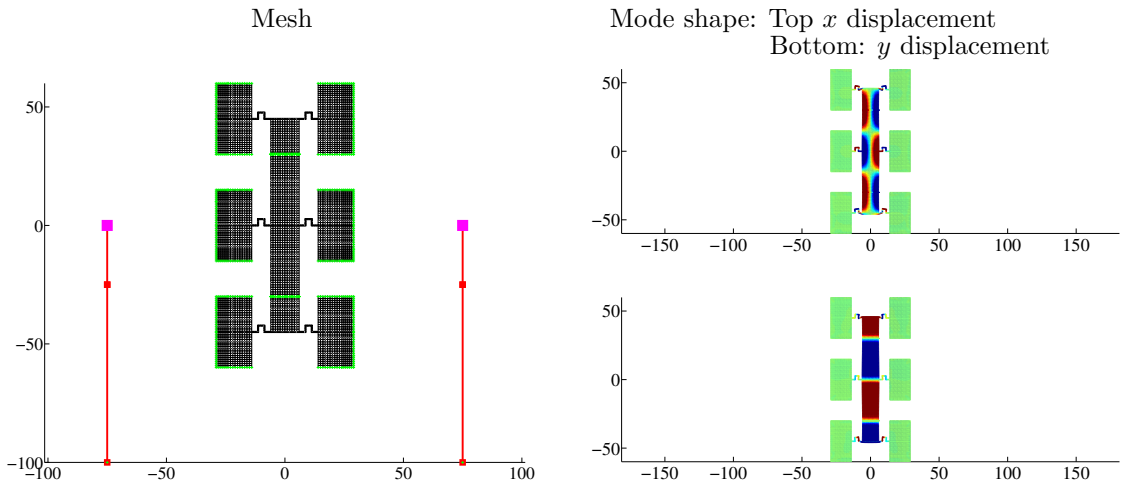


Figure 5.8: Resonator mesh and mode shape at 139.5[MHz]

from these two models with the full finite element model. The section is closed with a computation of the insertion loss of the resonator in 1-port and 2-port configuration, with both the full model and equivalent circuit parameter model for comparison.

5.4.1 Frequency and quality factor evaluation

The mesh of the discretized resonator along with the mode shape at approximately 140[MHz] is shown in Figure 5.8. The model has 71784 degrees of freedom including both mechanical and potential degrees of freedom. The mode of vibration has 3 nodes along the y direction. The dielectric material strips are placed at the anti-nodes for better electromechanical coupling. The obtained frequency and Q for the mode are,

$$f = 1.3954 \times 10^8 + 4.8723 \times 10^1 i \text{ [Hz]}, \tag{5.2}$$

$$Q = 1432000. \tag{5.3}$$

From the mode, one can see that extension and contraction in the y direction leads to motion in the supporting beams and energy dissipation into the substrate.

5.4.2 Equivalent parameter estimation

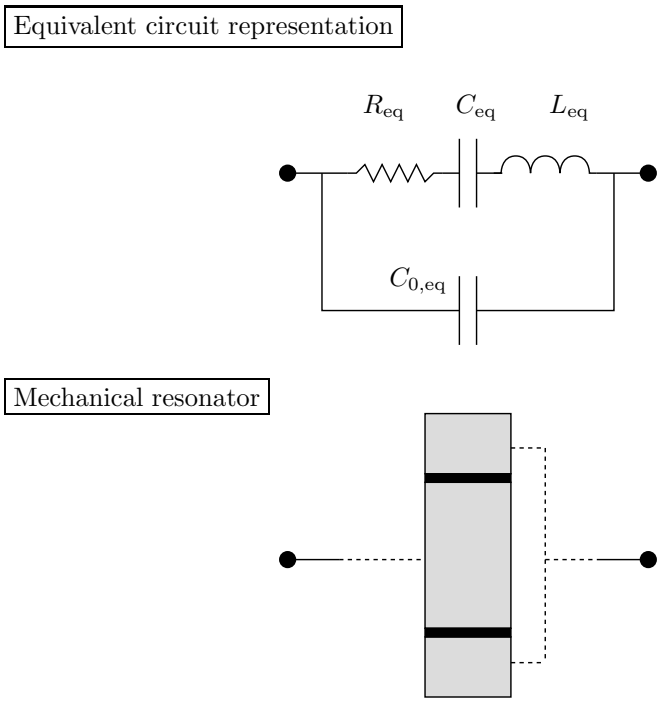


Figure 5.9: Schematic of a resonator in 1-port configuration

The equivalent circuit parameters for the mode of vibration shown in Figure 5.8 can be estimated from the method presented in Section 4.6. The resonator is assumed connected in a 1-port configuration, shown in Figure 5.9, along with its equivalent circuit representation. The top and bottom part are assumed to be of equal potential. In Section 4.6, two methods of equivalent parameter extraction were presented. The values obtained from method 1 are,

$$L_{eq} = 11.54 \text{ [H]}, \tag{5.4}$$

$$C_{eq} = 0.1127 \text{ [aF]}, \tag{5.5}$$

$$R_{eq} = 6.968 \text{ [k}\Omega\text{]}, \tag{5.6}$$

$$C_{0,eq} = 0.4250 \text{ [pF]}. \tag{5.7}$$

The values obtained from method 2 are,

$$L_{\text{eq}} = 11.54 \text{ [H]}, \tag{5.8}$$

$$C_{\text{eq}} = 0.1127 \text{ [aF]}, \tag{5.9}$$

$$R_{\text{eq}} = 6.968 \text{ [k}\Omega\text{]}, \tag{5.10}$$

$$C_{0,\text{eq}} = 0.9917 \text{ [pF]}. \tag{5.11}$$

One sees that the only difference lies in the static capacitance term. The admittance of the resonator defined as,

$$Y(\omega) := \frac{I}{V}, \tag{5.12}$$

is computed from the full model with 71784 degrees of freedom and the equivalent circuit models obtained from method 1 and 2 with one degree of freedom. Recall that method 1 involves projection of both mechanical and potential degrees of freedom onto the eigenmode, as opposed to method 2 which involves a Schur complement of the potential degrees of freedom followed by a projection of the mechanical degrees of freedom on the mechanical part of the eigenmode. The admittance and phase angle are shown in Figures 5.10 and 5.11. One sees that the equivalent circuit model obtained from method 1 is offset and not accurate. On the other hand, the relative accuracy of the equivalent circuit parameter model obtained from method 2 over this region is 10^{-4}. These results imply that method 2 should be chosen to obtain the equivalent circuit parameters.

The time required to compute the 100 data points across this frequency range for the full model is 5 minutes compared to seconds for the equivalent circuit model. For a fair comparison, one must include the time required to obtain to compute the equivalent circuit parameters, which is 2 minutes. Note that the difference between these timings increases linearly with the increase in the number of desired sampling points. Thus one sees that the equivalent circuit model performs well in terms of both time and accuracy.

It has been mentioned in Equation (5.1) that varying of the parameters, the DC bias voltage V_{dc}, the relative permittivity of the dielectric gap material ϵ_r, the surface area A, and the gap distance

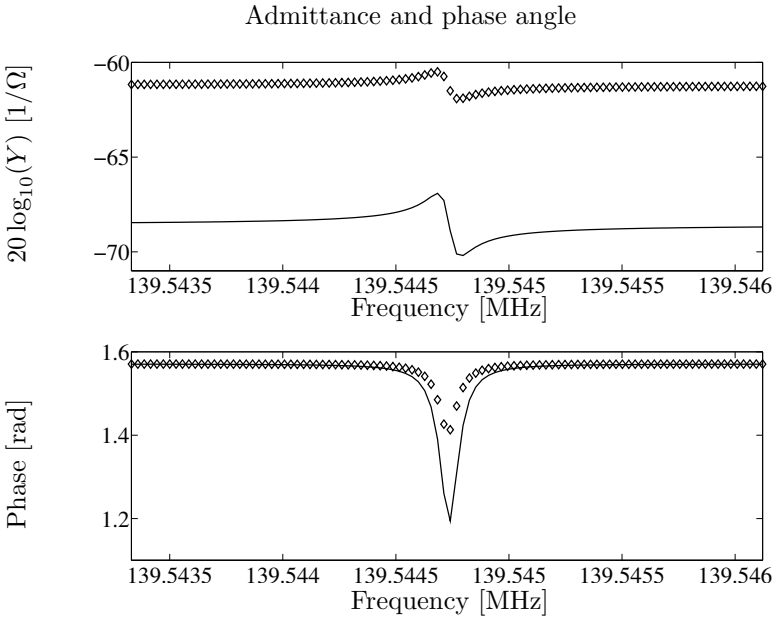


Figure 5.10: Admittance of the resonator obtained from 1st projection method. The solid line is obtained from the full model and the diamond glyphs are obtained from the equivalent circuit parameters.

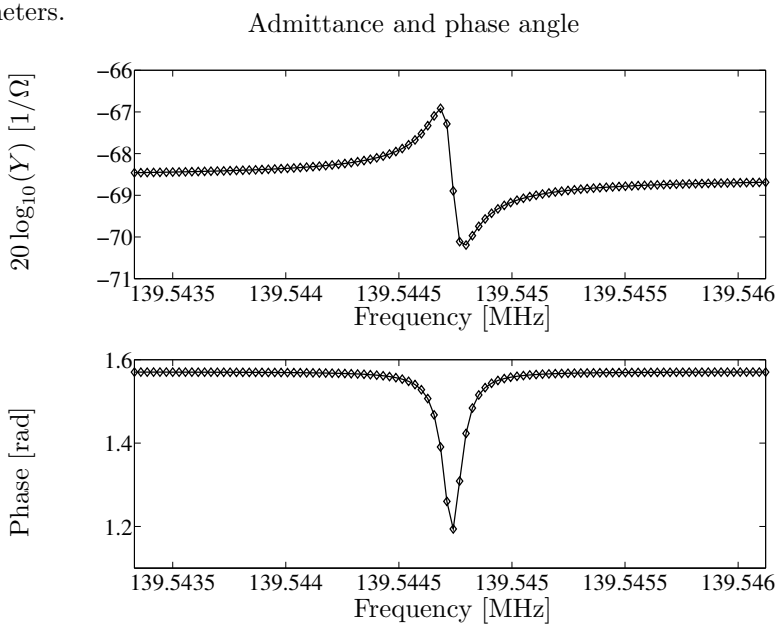


Figure 5.11: Admittance of the resonator obtained from 2nd projection method. The solid line is obtained from the full model and the diamond glyphs are obtained from the equivalent circuit parameters.

g_0 can change the electromechanical coupling η, leading to reduction in the motional resistance R_{eq}. Due to the geometry of the dielectric gaps, the behavior of these gaps can be approximated to some extent as an electromechanical parallel plate capacitor with dielectric material inserted. The motional resistance for the electromechanical parallel plate capacitor is,

$$R_{\text{eq}} := \frac{\sqrt{mk}}{\eta^2 Q} = \frac{\sqrt{mk}}{Q}\left(\frac{g_0^2}{\epsilon_r \epsilon_0 A V_{\text{DC}}}\right)^2. \tag{5.13}$$

The effect of change in several parameters are displayed here. The base parameters for the analysis are,

$$\epsilon_r = 25, \tag{5.14}$$

$$V_{\text{DC}} = 10, \tag{5.15}$$

$$t_{\text{layer}} = 2\ \mu\text{m}, \tag{5.16}$$

$$g_0 = 0.025\ \mu\text{m}. \tag{5.17}$$

In each case only one parameter is varied to see its effect on the motional resistance.

- Table 5.11: $\epsilon_r \in \{25, 50, 100\}$.

  As the relative permittivity is doubled, the motional resistance decreases by a factor of 4. This displays the quadratic dependence of the motional resistance on the relative permittivity.

- Table 5.12: $V_{\text{DC}} \in \{10, 20, 40\}$.

  As the DC bias voltage is doubled, the motional resistance decreases by a factor of 4. This displays the quadratic dependence of the motional resistance on the DC bias voltage.

- Table 5.13: $t_{\text{layer}} \in \{2, 4, 8\}\ \mu$m.

  As the thickness of the device is doubled, the motional resistance decreases by a factor of 2. A doubling of the thickness equals doubling of the area A, and linear dependence contradicts the prediction of quadratic dependence of motional resistance on the area. This is due to the 2D aspect of this problem. As the thickness is increased by 2, the mass and stiffness increase by 2, canceling the quadratic effect of the increase in area.

- Table 5.14: $g_0 \in \{0.025, 0.050, 0.100\}$ μm.

  As the thickness of the gap is increased the motional resistance increases by a factor of 4. This contradicts the prediction of quartic dependence of the motional resistance on the the gap size. The quality factor Q is not a source of this strange behavior, since the Q value only changes in the 3rd or 4th digit due to the change in gap size. The cause of this quadratic dependence lies in the coupling between the gap location and amount of motion. As the gap is doubled, the motion of the parallel plate also doubles, leading to more motional current and less resistance. This leads to a cancellation of the quartic dependence.

From these results, one can see that the parallel plate approximation proves useful to a certain degree, but is not capable of predicting trends such as those seen in the quadratic dependence of gap distance.

Table 5.11: Varying the relative permittivity

ϵ_r	25	50	100
R_{eq}	6.968×10^3	1.741×10^3	4.349×10^2

Table 5.12: Varying the DC Bias voltage

V_{DC}	10	20	40
R_{eq}	6.968×10^3	1.740×10^3	4.325×10^2

Table 5.13: Varying the thickness

t_{layer}	2μm	4μm	8μm
R_{eq}	6.968×10^3	3.484×10^3	1.742×10^3

Table 5.14: Varying the gap size

g_0	0.025 μm	0.050 μm	0.100 μm
R_{eq}	6.968×10^3	2.792×10^4	1.119×10^5

5.4.3 Insertion loss

Here the insertion loss due to the resonator is computed with the full model and the estimated equivalent circuit parameters to display the accuracy and efficiency of the model reduction.

The schematic shown in Figure 5.12 is used to explain the definition of insertion loss. A circuit with a time varying AC voltage $V_{\rm AC}$ and load resistor R_L with a mechanical resonator inserted is depicted. The AC voltage pumps power into the load resistor per cycle. Insertion loss is defined as the decrease in the power supplied to the load resistor by insertion of the mechanical resonator. The original power supplied is,

$$P_0 = \frac{1}{2}\frac{|V_{\rm AC}|^2}{R_L}, \tag{5.18}$$

since the voltage across the load resistor is $V_{\rm AC}$. Let the voltage across the load resistor be V_L after insertion of the mechanical resonator. Now the power supplied is,

$$P = \frac{1}{2}\frac{|V_L|^2}{R_L}; \tag{5.19}$$

the ratio between the two in dB is defined as the insertion loss,

$$\text{I.L.} := 10\log_{10}\left(\frac{P_0}{P}\right) \tag{5.20}$$

$$= 20\log_{10}\frac{|V_{\rm AC}|}{|V_L|}. \tag{5.21}$$

The transmission is defined as the negative of this quantity,

$$\text{Transmission} := -\text{I.L.} \tag{5.22}$$

$$= 20\log_{10}\frac{|V_L|}{|V_{\rm AC}|}. \tag{5.23}$$

1-port configuration

The insertion loss of the resonator in the 1-port configuration shown in Figure 5.9 is computed from the full model and with the equivalent circuit model with the parameters computed in Section 5.4.2 with method 2. Given the equivalent circuit parameters, the transmission for this system

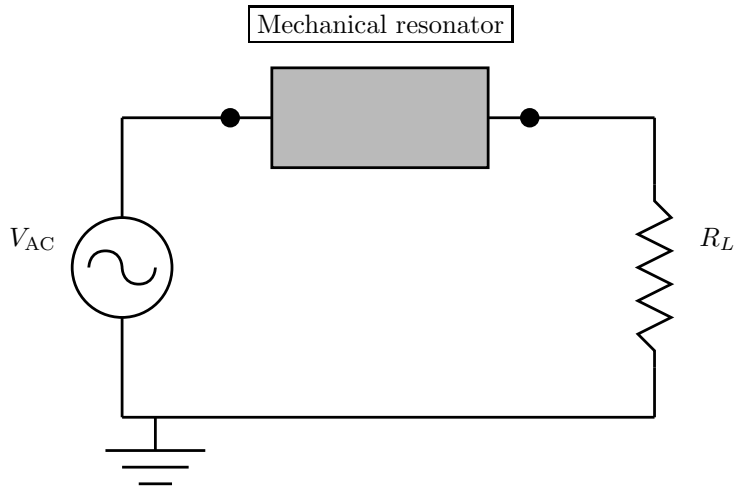


Figure 5.12: Schematic of a resonator in a circuit for measuring insertion loss

is written as,

$$Y(\omega) := i\omega C_{0,\mathrm{eq}} + \frac{1}{\frac{1}{i\omega C_{\mathrm{eq}}} + i\omega L_{\mathrm{eq}} + R_{\mathrm{eq}}}, \tag{5.24}$$

$$\text{Transmission} = 20\log_{10}\left(\frac{R_L}{R_L + 1/Y(\omega)}\right). \tag{5.25}$$

The computed transmission with a load resistor of 50 [Ω] is shown in Figure 5.14. The relative accuracy of the equivalent circuit model is 3 to 4 digits, and the timing results are the same as for the computation of the admittance presented in Section 5.4.2.

2-port configuration

The insertion loss of the resonator in the 2-port configuration shown in Figure 5.13 is computed from the full model and with the equivalent circuit model with the parameters computed in Section 5.4.2 with method 2. The top, middle, and bottom part of the resonator are all set at different potentials. Since the mode of interest is symmetric in motion with respect to the center of the resonator, one can easily compute the equivalent circuit parameters for the two port configuration. These values are shown in Figure 5.13. The factor of 4 arises, from the square of the halving of the

electromechanical coupling coefficient. In the 2-port configuration there is only half of the mechanical coupling to the top and bottom part. Given the equivalent circuit parameters, the transmission for this system is written as,

$$Y_1(\omega) := \frac{1}{4\left(\frac{1}{i\omega C_{\text{eq}}} + i\omega L_{\text{eq}} + R_{\text{eq}}\right)}, \tag{5.26}$$

$$Y_2(\omega) := i\omega\frac{C_{0,\text{eq}}}{2}, \tag{5.27}$$

$$\text{Transmission} = 20\log_{10}\left(-\frac{Y_1(\omega)}{Y_1(\omega) + Y_2(\omega) + \frac{1}{R_L}}\right). \tag{5.28}$$

The computed transmission with a load resistor of 50 [Ω] is shown in Figure 5.15. The relative accuracy of the equivalent circuit model is 4 to 5 digits, and the timing results are the same as for the computation of the admittance presented in Section 5.4.2.

Effect of DC bias voltage and relative permittivity

The effect of changing the DC bias voltage V_{DC} and the relative permittivity ϵ_r are presented. The transmission of the resonator in a 2-port configuration with a load resistor of $R_L = 50$ [Ω] for varying DC bias voltage and relative permittivity is computed.

- Figure 5.16 top: $\epsilon_r \in \{25, 50, 100\}$.

  An increase in the relative permittivity reduces the insertion loss.

- Figure 5.16 bottom: $V_{\text{DC}} \in \{10, 20, 40\}$.

  An increase in the DC bias voltage reduces the insertion loss. The increase in voltage leads to mechanical softening and a decrease in the eigenfrequency.

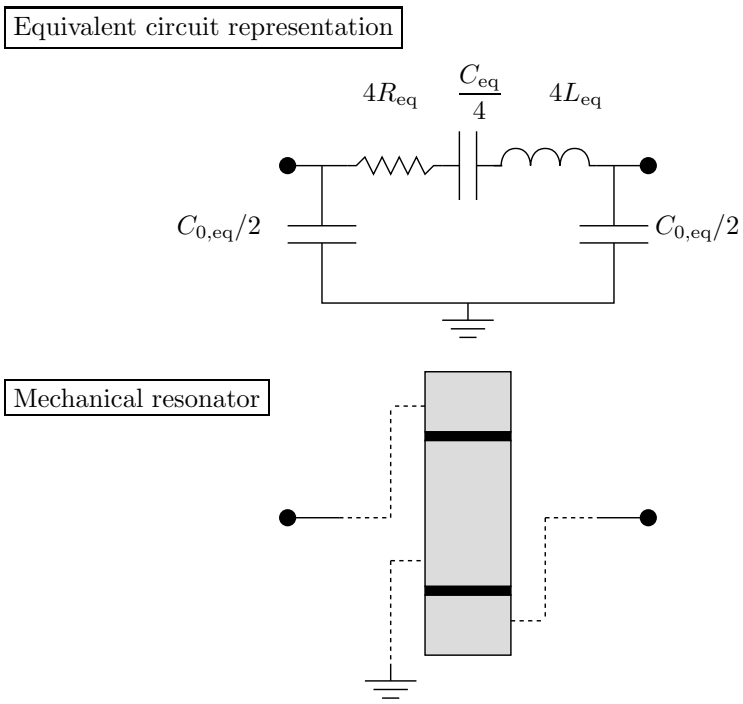


Figure 5.13: Schematic of a resonator in 2-port configuration

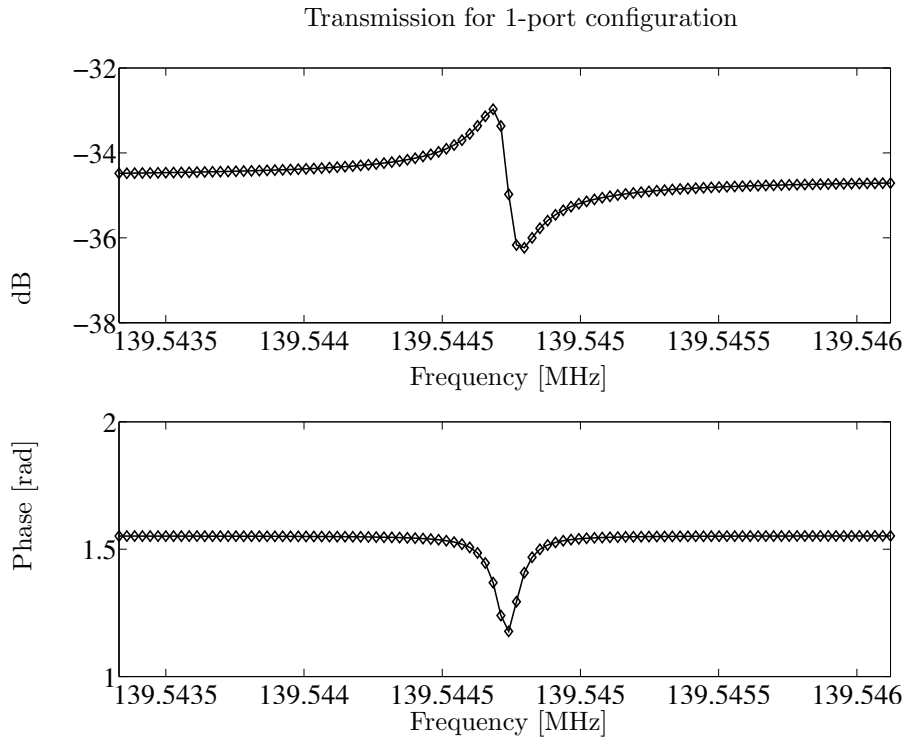


Figure 5.14: Transmission of the resonator in 1-port configuration. The solid line is obtained from the full model and the diamond glyphs are obtained from the equivalent circuit parameters.

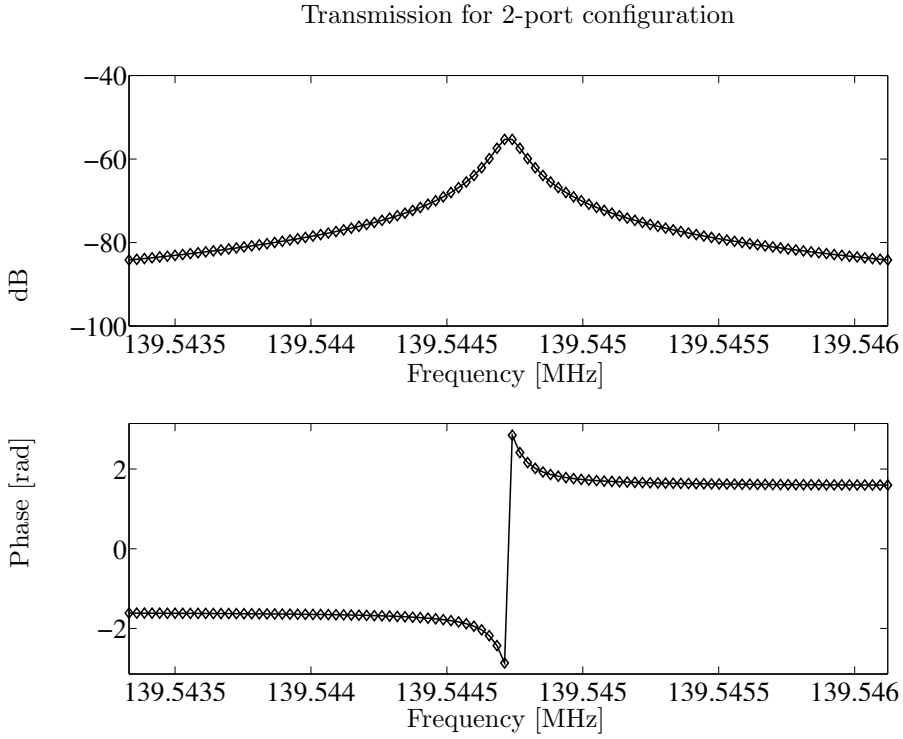


Figure 5.15: Transmission of the resonator in 2-port configuration. The solid line is obtained from the full model and the diamond glyphs are obtained from the equivalent circuit parameters.

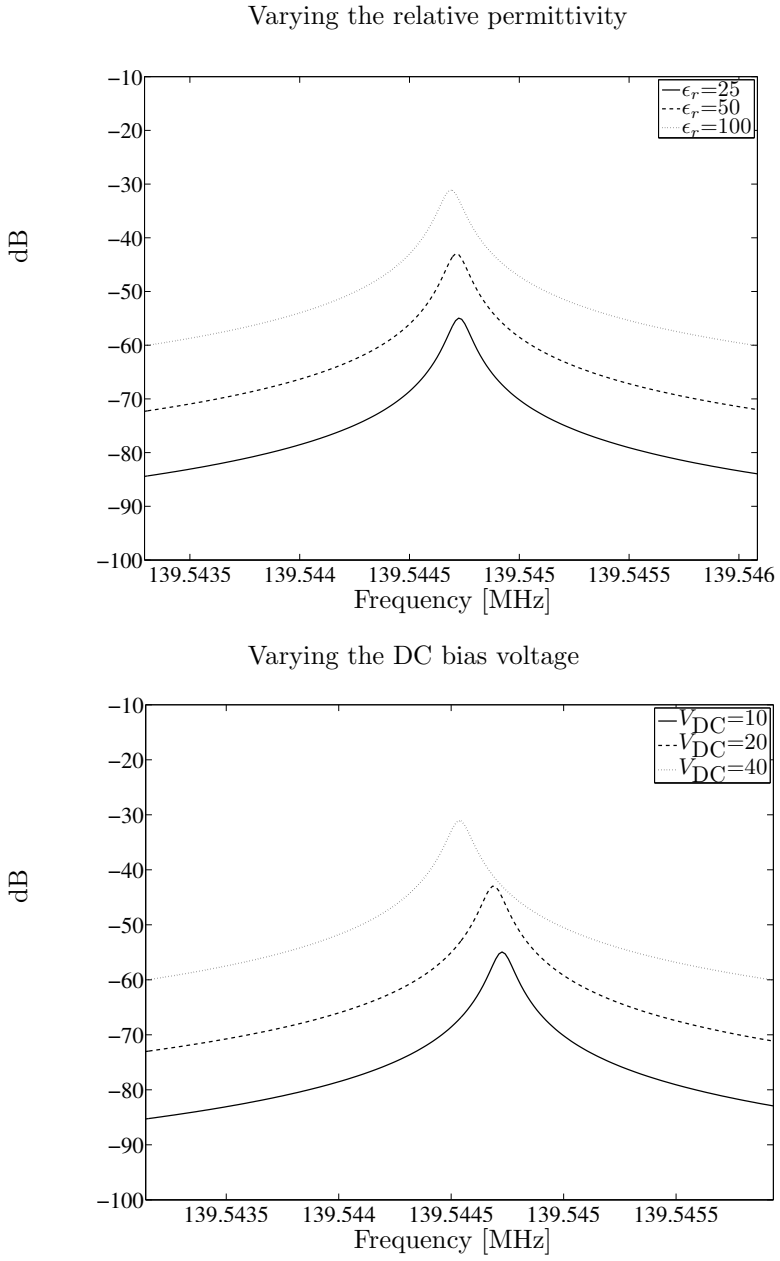


Figure 5.16: Transmission of the resonator. The solid line is obtained from the full model and the diamond glyphs are obtained from the equivalent circuit parameters.

5.5 Michigan free-free beam resonator and ring resonator

Here, we present two examples which will elucidate and verify the theoretical formulation presented in the Chapter 3. The TED contribution to the quality factor Q of the Michigan Free-Free Beam [98], which has been designed to minimize anchor loss, is computed by the normal, SOAR, and structure preserving ROM, SOAR-S, and compared with experimental results. The structure preserving ROM, SOAR-S, yields results which are up to an order of magnitude better than the normal, SOAR, only requiring the same computational effort. The time required to evaluate the transfer function with the two ROMs can be up to 60 times faster than evaluation of the full model. The accuracy of the reduced order model in evaluating the combined effects of TED and PML on a non-beam like geometry is portrayed by a ring resonator example. Due to this irregular geometry and the mode of vibration, a through thickness mode, the classical Zener's formula is inapplicable.

5.5.1 Michigan Free-Free Beam (TED)

A schematic of the Michigan Free-Free Beam (MFFB) is shown in Figure 5.17. The main structure of the 2D planar resonator is the $39.8\mu m \times 2\mu m$ beam in the center, which is actuated by in-plane electrostatic forces at the middle. This beam is designed to vibrate in an in-plane flexural mode similar to its 1st fundamental mode at a frequency of 100 [MHz]. This mode is shown in Figure 5.18, where the colors represent the thermal fluctuations from a reference temperature. Red represents positive and blue represents negative fluctuations. Since the exact material properties of the polysilicon used in the fabricated device are unknown, we use the values summarized earlier in Table 3.1. The thickness of the device is reported to be $2\mu m$, which allows the use of a 2D plane stress assumption.

In evaluating the transfer function, we use a fine finite element discretization of 60318 DOFs. The accuracy of three ROMs produced from SOAR iterations at an expansion point of $s_0 = i9.5926 \times 10^6$ are compared with the full model: (a) a 2-DOF SOAR(2) ROM produced from two iterations

of SOAR, (b) a 4-DOF SOAR-S(4) ROM produced from two iterations of SOAR and structure preservation, and (c) a 4-DOF SOAR(4) ROM produced from four iterations of SOAR. The purely mechanical forcing vector is assumed equal to the sensing vector; i.e., $\mathbf{b}_u = \mathbf{l}_u$ with zero thermal vectors. The 1-norm condition number of the dynamic stiffness matrix $\mathbf{K} + s_0\mathbf{C} + s_0^2\mathbf{M}$ at the expansion frequency s_0 is $O(10^8)$.

Figure 5.19(a) shows the mechanical transfer function for the full model and ROMs centered at the expansion frequency which is indicated by the dotted line. All plots lie on top of each other with $Q_{\text{TED}} = 13861$ for the peak. Comparing this with the experimentally obtained value of $Q_{\text{exp}} = 10743$, we see that the model predicts the actual value relatively close, validating our method. It is clear that we have extremely high accuracy for all ROMs, even with a small number of DOFs, since all results overlap each other. The relative error of the ROMs are shown in Figure 5.19(c), where all ROMs have more that 4 digits of accuracy. Comparison between SOAR(2) and SOAR-S(4) show increased accuracy by the split basis, though the difference is not substantial due to the weak coupling between the mechanical and thermal domains. To confirm this claim, the transfer function is computed for the case when ξ_1, the non-dimensionalized constant representing the degree of coupling, is increased by an order of magnitude. The transfer function and relative error of the ROMS are shown in Figures 5.19(b),5.19(d). It is clear that structure preservation becomes more advantageous when the coupling is increased. Comparison between the SOAR-S(4) and SOAR(4) ROM, which both have 4-DOFs, show that more accuracy can be attained at the expense of an additional two SOAR iterations.

5.5.2 Ring Resonator (TED+PML)

The performance of the ROM in evaluating non-beam like systems involving TED and TED/anchor loss is displayed in the ring resonator device shown in Figure 5.20(a). The 2D planar poly-silicon ring is supported on two ends by beams. The device is actuated by electrostatic forces on the perimeter of the ring in the radial direction. Among the wide range of operating frequencies, we select one in

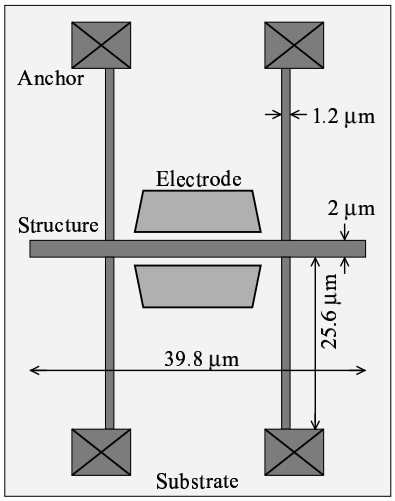


Figure 5.17: Michigan Free-Free Beam Schematic

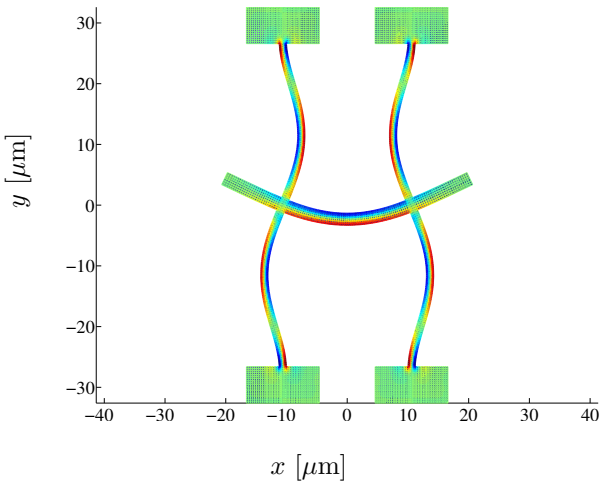


Figure 5.18: Michigan Free-Free Beam Deformed Shape

which the resonator has relatively low insertion loss and vibrates in a mode shown in Figures 5.20(b)-5.20(c). The magnitudes of the x, y displacements and thermal fluctuations are represented by the colors, red for positive and blue for negative values.

The discretization that we select results in a 98274 DOF system. The transfer function obtained at a center frequency of 618.221[MHz] for the TED system is shown in Figure 5.21. The dotted line again denotes the expansion frequency for the ROMs which is equal to the center frequency. The 1-norm condition number of the dynamic stiffness matrix $\mathbf{K}+s_0\mathbf{C}+s_0^2\mathbf{M}$ at the expansion frequency s_0 is $O(10^6)$. The ROMS used have 32 DOFs. Thus, 32 SOAR iterations are needed to produce the SOAR(32) ROM and 16 iterations are needed to produce the SOAR-S(32) ROM. It is seen from the transfer function, that both ROMs perform well. Evaluation of the plot, consisting of 200 data points across the range, takes a little under an hour for the Full Model, compared to approximately a minute for the ROMs (on an Intel(R) Xeon(TM) 3.06GHz CPU). Considering the amount of accuracy achieved and time saved, the efficiency of the ROMs is clear. The relative errors with respect to the full model of the ROMs are displayed in Figure 5.22, where both produce on average 11 digits of accuracy around the expansion point. Locally around the expansion point,

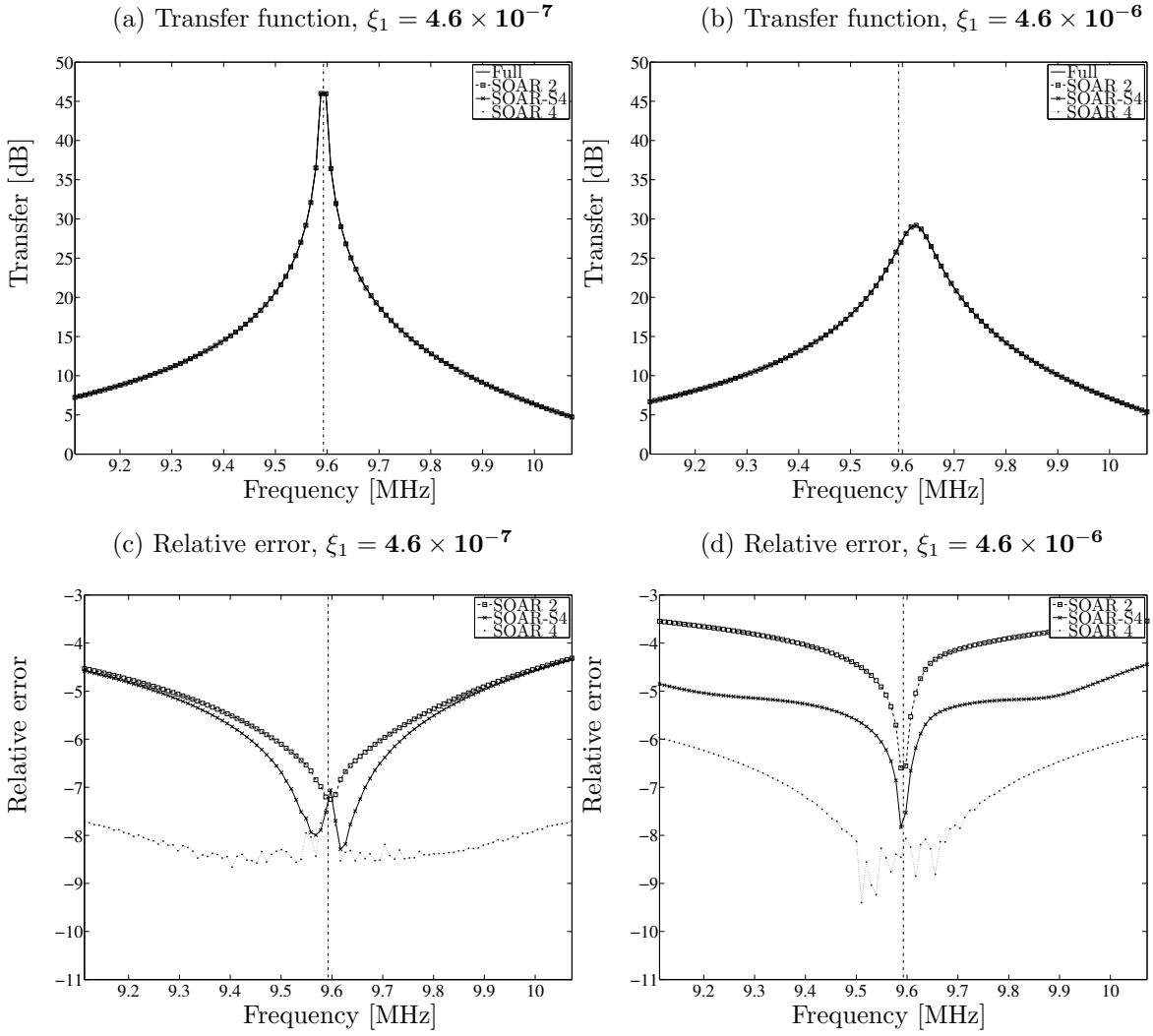


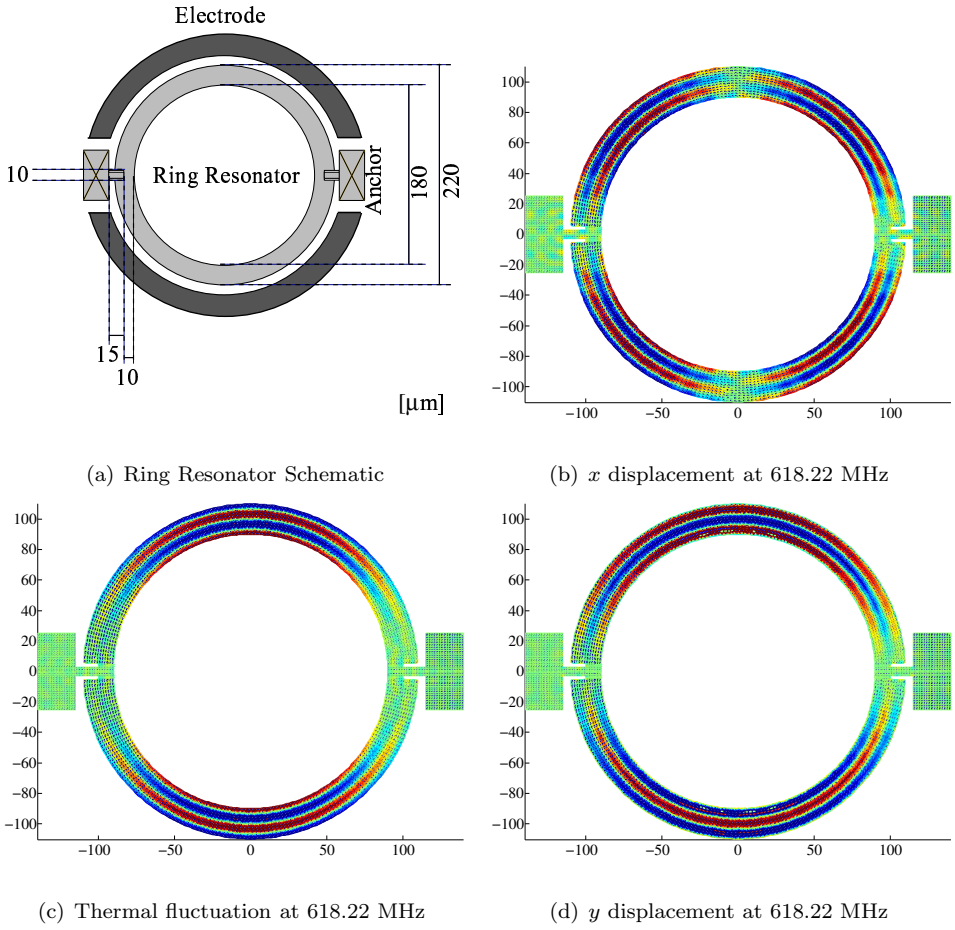
Figure 5.19: Transfer function and relative error of the Michigan Free-Free Beam under weak and strong coupling

we see that structure preservation obtains slightly better accuracy than blindly expanding the size of the ROM. It should be emphasized that SOAR-S(32) requires only half as many iterations as SOAR(32). For this case, one cannot argue that SOAR-S(32) takes half the time of SOAR(32) to compute, since a direct method is used to apply the operator $\mathbf{K}_{s_0}$ in producing the second order Krylov subspaces in Equations 3.56 and 3.57. For direct methods, a majority of the time is spent in the LU factorization required to apply the inverse of the operator $\mathbf{K}_{s_0}$. Still the application of the inverse operator through an upper and lower triangular solve required at each step of the expansion

of the second order Krylov subspace is non-negligible. One still does get a speed up of some degree depending on the ratio between the time required for the LU factorization and the time required for the upper and triangular solve which depends on the sparse direct solve algorithm. When an iterative method is applied instead of a direct method, one can argue that SOAR-S(32) requires half the time of SOAR(32), since the time required to produce the second order Krylov subspace is proportional to its size.

Next we compare the accuracy of the ROMs for computing the transfer function of a device incorporating both TED and anchor loss. This results in solving a complex symmetric system of equations. The transfer function for the full model and ROMs evaluated at a center frequency of 618.221[MHz] are shown in Figure 5.23. The dotted line again denotes the expansion frequency for the ROMs which is equal to the center frequency. The number of DOFs for the ROMs are matched to 32 requiring 32 SOAR iterations to produce the SOAR(32) ROM, 16 for the SOAR-S(32) ROM and SOAR-R(32) ROM, and 8 for the SOAR-RS(32) ROM. The accuracy of the ROMs are clear and taking into consideration that the time required for evaluation is similar to the TED case, we can extend our claim on the efficiency of the ROM to the TED/anchor loss case.

By examining the ROMs relative errors displayed in Figure 5.24, we see that locally in the neighborhood of the expansion point, the SOAR-RS(32) ROM is the most accurate. Though the SOAR-R(32) and SOAR-S(32) are also quite accurate, they do not do as well as the SOAR-RS(32), since $2k$ moments match only under this form. Thus if one considers the time required for a desired accuracy locally around the point of expansion, we can claim that the SOAR-RS(32) can be up to 4 times as fast as the SOAR(32) when an iterative method is used to apply the inverse operator in generating the second-order Krylov subspace. This is because SOAR-RS(32) only requires 8 SOAR iterations compared to SOAR(32) which requires 32 SOAR iterations.



(a) Ring Resonator Schematic

(b) x displacement at 618.22 MHz

(c) Thermal fluctuation at 618.22 MHz

(d) y displacement at 618.22 MHz

Figure 5.20: A schematic of the ring resonator and its vibrational mode at 618.22 [MHz]

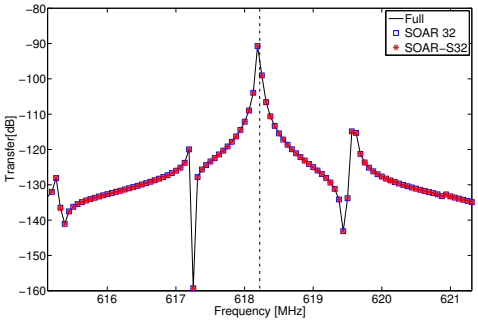


Figure 5.21: Ring Resonator Bode plot (TED)

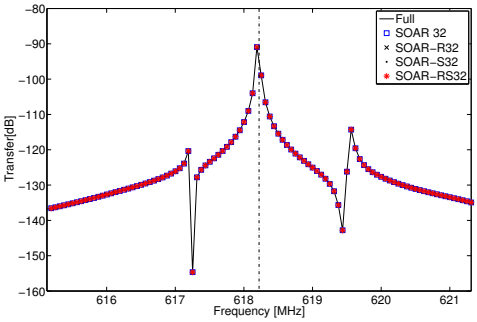


Figure 5.23: Ring Resonator Bode plot (TED/Anchor Loss)

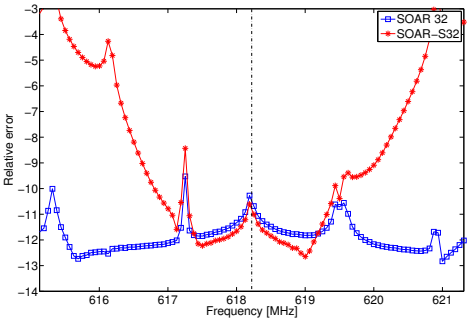


Figure 5.22: Ring Resonator Error plot (TED)

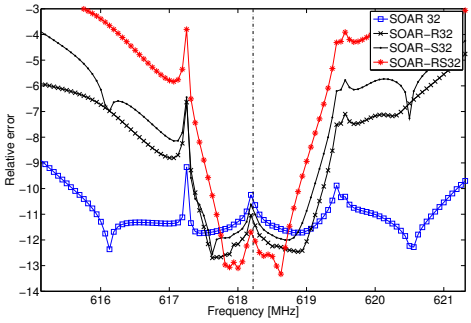


Figure 5.24: Ring Resonator Error plot (TED/Anchor Loss)

5.6 Conclusion

In this chapter, the simulation of a disk resonator, a dielectric transduced resonator, a free-free beam resonator, and a ring resonator have been presented, each exhibiting special features of the technology that has been developed to efficiently simulate damping behavior in MEMS devices.

Through the disk resonator example, the applicability of our proposed geometric multigrid method for solving complex-symmetric linear systems arising from time-harmonic elastodynamics with the application of PML, is confirmed. Though the method cannot solve linear systems with arbitrary PML parameters, i.e., $\beta > 1$ for a linear absorbing function profile, sufficient results can still be obtained under this limitation for resonators operating in the MHz to GHz range. As long as the coarsest grid in the multigrid has sufficient fineness, the method displays fast convergence. Under this condition, the method only requires under 50 preconditioned GMRES iterations for a relative residual of 1×10^{-10} and this number of preconditioned GMRES iterations increases very mildly with respect to an increase in the number of degrees of freedom of the fine grid. The compute time required for a solution also scales relatively well with respect to the number of processors; for a problem with 6 million degrees of freedom, increasing the number of processors from 8 to 64 by a factor of 8 decreases the compute time by a factor of 4. This is a large improvement compared to the current technology where no methods have been proposed for efficient solution of the large scale complex-symmetric linear system arising from the application of PML. The geometric multigrid preconditioned Jacobi-Davidson QZ eigensolver proposed inherits the fast convergence and scalable behavior of the geometric multigrid method, enabling the first 3D computations of the quality factor for the disk resonators. 3D simulations have allowed us to model post misalignment in disk resonators, to confirm the claims made on its effect of degrading the quality factor Q. Simulation results reveal that from a purely mechanical perspective, the anchor misalignment is an infinitesimal perturbation, slightly effecting the quality factor. This suggests that other mechanisms such as the electromechanical coupling introduced by the DC bias voltage may play a role in the quality factor

degradation. Such investigations should be considered.

The dielectric transduced resonator has exhibited the validity of the variational approach to the electromechanically coupled system and the method of equivalent parameter extraction. The extracted equivalent circuit parameters are highly accurate (relative accuracy of 4 digits in the transfer function) and capable of modeling the resonator in the vicinity of its resonance with just one degree of freedom Evaluation of the transfer function with this equivalent circuit model requires only seconds compared to minutes for the full model. The time required to evaluate the equivalent circuit model parameters can be amortized over the number of times the equivalent circuit model parameters are reused. For this device, it has also been shown that simple parallel plate assumptions can give false results, such as the discrepancy between the parallel plate model which if falsely applied predicts quartic dependence of the motional resistance to the gap size, as opposed to the actual quadratic dependence. The proposed simulation framework can analyze and extract equivalent circuit parameters from devices with arbitrary geometry and forcing patterns.

To display the effectiveness of structure preserving reduced order modeling, the free-free beam resonator and ring resonator have been simulated. In terms of the time required to simulate the transfer function across a specified range, the reduced order is superior by an order of magnitude in time compared to evaluation of the full model, and has sufficiently good accuracy. For the thermoelastic coupled problem at the MEMS scale, the structure preservation does not have such a high gain in accuracy. But as it has been displayed, for a system with coupling an order of magnitude larger between the thermal and mechanical domain, the structure preserving reduced order model achieves several orders of magnitude higher accuracy. There is no restriction to the geometry that our structure preserving reduced order model is applicable to, and this is displayed through the reduced order modeling of the ring resonator.

Appendix A

Derivations for Thermoelastic damping

Lemma A.0.1 *Define the matrix valued functions* $f, g : \mathcal{C} \to \mathcal{C}^{N\times N}$,

$$g(s) = \left(\mathbf{I} - s\mathbf{A} - s^2\mathbf{B}\right) \tag{A.1}$$

$$f(s) = g(s)^{-1} \tag{A.2}$$

where $\mathbf{A}, \mathbf{B} \in \mathcal{C}^{N\times N}$. *Then the coefficients* $\mathbf{E}^n \in \mathcal{C}^{N\times N}$ *of the Taylor series expansion of* $f(s)$ *at* $s = 0$,

$$f(s) = \sum_{n=0}^{\infty} \mathbf{E}^n s^n \ , \tag{A.3}$$

are given by the recursion,

$$\mathbf{E}^0 = \mathbf{I} \tag{A.4}$$

$$\mathbf{E}^1 = \mathbf{A} \tag{A.5}$$

$$\mathbf{E}^n = \mathbf{A}\mathbf{E}^{n-1} + \mathbf{B}\mathbf{E}^{n-2} \quad (n \geq 2) \tag{A.6}$$

$$= \mathbf{E}^{n-1}\mathbf{A} + \mathbf{E}^{n-2}\mathbf{B} \quad (n \geq 2) \ . \tag{A.7}$$

Proof Since $g(s)$ is of second order, all derivatives of order 3 and higher are zero. Evaluating these at $s = 0$, we have,

$$g(0) = \mathbf{I}, \quad g'(s) = -\mathbf{A}, \quad g''(s) = -2\mathbf{B} \ .$$

By taking the derivatives of the equation $f(s)g(s) = g(s)f(s) = \mathbf{I}$, and evaluating them at $s = 0$, we obtain the recursion,

$$\begin{aligned} f(0) &= \mathbf{I} \\ f'(0) &= \mathbf{A} \\ f^{(n)}(0) &= nf^{(n-1)}(0)\mathbf{A} + n(n-1)f^{(n-2)}(0)\mathbf{B} \\ &= \mathbf{A}nf^{(n-1)}(0) + \mathbf{B}n(n-1)f^{(n-2)}(0) \ . \end{aligned}$$

Since the coefficients of the Taylor series are defined as,

$$\mathbf{E}^n = \frac{1}{n!} f^{(n)}(0) \ ,$$

the recursion relation holds. □

Lemma A.0.2 *A matrix* $\mathbf{P}$ *that satisfies* $\mathbf{P}^2 = \mathbf{P}$ *is defined as a projector onto* $\mathsf{span}(\mathbf{P})$*. Given matrices* $\mathbf{X}, \mathbf{Y}, \mathbf{K}$*, we define the projectors,*

$$\mathbf{P} = \mathbf{X}\left(\mathbf{Y}^*\mathbf{K}\mathbf{X}\right)^{-1}\mathbf{Y}^*\mathbf{K} \tag{A.8}$$

$$\mathbf{Q} = \mathbf{K}\mathbf{X}\left(\mathbf{Y}^*\mathbf{K}\mathbf{X}\right)^{-1}\mathbf{Y}^* \ , \tag{A.9}$$

where $\left(\mathbf{Y}^*\mathbf{K}\mathbf{X}\right), \mathbf{K}$ *are assumed invertible. Then,*

$$\begin{aligned} \mathbf{P}\mathbf{x} = \mathbf{x} \quad & \textit{for any } \mathbf{x} \in \mathsf{span}(\mathbf{X}) \\ \mathbf{y}^*\mathbf{Q} = \mathbf{y}^* \quad & \textit{for any } \mathbf{y} \in \mathsf{span}(\mathbf{Y}) \end{aligned} \ . \tag{A.10}$$

It is easily verified that these two matrices $\mathbf{P}, \mathbf{Q}$ *satisfy the condition stated above for a projector, and project onto* $\mathsf{span}(\mathbf{X})$*,* $\mathsf{span}(\mathbf{Y})$ *respectively.*

Proof This Lemma is a well known fact and is stated without proof.

Lemma A.0.3 *Let matrices be defined as in Equations (3.27, A.8, A.9) and*

$$\begin{array}{cccc} \mathbf{A}_r = -\mathbf{K}^{-1}\mathbf{D} & \mathbf{B}_r = -\mathbf{K}^{-1}\mathbf{M} & \mathbf{A}_l = -\mathbf{D}\mathbf{K}^{-1} & \mathbf{B}_l = -\mathbf{M}\mathbf{K}^{-1} \\ \mathbf{A}_{r,R} = -\mathbf{K}_R^{-1}\mathbf{D}_R & \mathbf{B}_{r,R} = -\mathbf{K}_R^{-1}\mathbf{M}_R & \mathbf{A}_{l,R} = -\mathbf{D}_R\mathbf{K}_R^{-1} & \mathbf{B}_{l,R} = -\mathbf{M}_R\mathbf{K}_R^{-1} \end{array} . \quad \text{(A.11)}$$

Then,

$$\mathbf{X}\mathbf{K}_R^{-1}\mathbf{b}_R = \mathbf{P}\mathbf{K}^{-1}\mathbf{b} \quad \text{(A.12)}$$

$$\mathbf{X}\mathbf{A}_{r,R} = \mathbf{P}\mathbf{A}_r\mathbf{X} \quad \text{(A.13)}$$

$$\mathbf{X}\mathbf{B}_{r,R} = \mathbf{P}\mathbf{B}_r\mathbf{X} \quad \text{(A.14)}$$

$$\mathbf{l}_R^*\mathbf{K}_R^{-1}\mathbf{Y}^* = \mathbf{l}^*\mathbf{K}^{-1}\mathbf{Q} \quad \text{(A.15)}$$

$$\mathbf{A}_{l,R}\mathbf{Y}^* = \mathbf{Y}^*\mathbf{A}_l\mathbf{Q} \quad \text{(A.16)}$$

$$\mathbf{B}_{l,R}\mathbf{Y}^* = \mathbf{Y}^*\mathbf{B}_l\mathbf{Q} \ . \quad \text{(A.17)}$$

Proof By Equation (A.8), we have

$$\begin{aligned} \mathbf{X}\mathbf{K}_R^{-1}\mathbf{b}_R &= \mathbf{X}\left(\mathbf{Y}^*\mathbf{K}\mathbf{X}\right)^{-1}\mathbf{Y}^*\left(\mathbf{K}\mathbf{K}^{-1}\right)\mathbf{b} \\ &= \mathbf{P}\mathbf{K}^{-1}\mathbf{b} \\ \mathbf{X}\mathbf{A}_{r,R} &= \mathbf{X}\mathbf{K}_R^{-1}\mathbf{D}_R = \mathbf{X}\left(\mathbf{Y}^*\mathbf{K}\mathbf{X}\right)^{-1}\mathbf{Y}^*\left(\mathbf{K}\mathbf{K}^{-1}\right)\mathbf{D}\mathbf{X} = \mathbf{P}\mathbf{K}^{-1}\mathbf{D}\mathbf{X} \\ &= \mathbf{P}\mathbf{A}_r\mathbf{X} \\ \mathbf{X}\mathbf{B}_{r,R} &= \mathbf{X}\mathbf{K}_R^{-1}\mathbf{M}_R = \mathbf{X}\left(\mathbf{Y}^*\mathbf{K}\mathbf{X}\right)^{-1}\mathbf{Y}^*\left(\mathbf{K}\mathbf{K}^{-1}\right)\mathbf{M}\mathbf{X} = \mathbf{P}\mathbf{K}^{-1}\mathbf{M}\mathbf{X} \\ &= \mathbf{P}\mathbf{B}_r\mathbf{X} \ . \end{aligned}$$

Similarly, by Equation (A.9), we have

$$\begin{aligned}
\mathbf{l}_R^* \mathbf{K}_R^{-1} \mathbf{Y}^* &= \mathbf{l}^* \left(\mathbf{K}^{-1}\mathbf{K}\right) \mathbf{X} \left(\mathbf{Y}^* \mathbf{K} \mathbf{X}\right)^{-1} \mathbf{Y}^* \\
&= \mathbf{l}^* \mathbf{K}^{-1} \mathbf{Q} \\
\mathbf{A}_{l,R} \mathbf{l}^* &= \mathbf{D}_R \mathbf{K}_R^{-1} \mathbf{Y}^* = \mathbf{Y}^* \mathbf{D} \left(\mathbf{K}^{-1}\mathbf{K}\right) \mathbf{X} \left(\mathbf{Y}^* \mathbf{K} \mathbf{X}\right)^{-1} \mathbf{Y}^* = \mathbf{Y}^* \mathbf{D} \mathbf{K}^{-1} \mathbf{Q} \\
&= \mathbf{Y}^* \mathbf{A}_l \mathbf{Q} \\
\mathbf{B}_{l,R} \mathbf{Y}^* &= \mathbf{M}_R \mathbf{K}_R^{-1} \mathbf{Y}^* = \mathbf{Y}^* \mathbf{M} \left(\mathbf{K}^{-1}\mathbf{K}\right) \mathbf{X} \left(\mathbf{Y}^* \mathbf{K} \mathbf{X}\right)^{-1} \mathbf{Y}^* = \mathbf{Y}^* \mathbf{M} \mathbf{K}^{-1} \mathbf{Q} \\
&= \mathbf{Y}^* \mathbf{B}_l \mathbf{Q} \ .
\end{aligned}$$

□

Lemma A.0.4 *Let $\{\mathbf{E}^i\}_{i=0}^{\infty}$ be the sequence of matrices defined in Lemma A.0.1. Then,*

$$\mathbf{E}^{i+j} = \mathbf{E}^i \mathbf{E}^j + \mathbf{E}^{i-1} \mathbf{B} \mathbf{E}^{j-1} \tag{A.18}$$

for all $i, j \geq 1$. If we define,

$$\mathbf{E}^{-1} = \mathbf{0} \tag{A.19}$$

then the relation holds for all $i, j \geq 0$.

Proof We prove this by induction. For $i, j = 1$,

$$\mathbf{E}^2 = \mathbf{E}^1 \mathbf{E}^1 + \mathbf{E}^0 \mathbf{B} \mathbf{E}^0 = \mathbf{A}^2 + \mathbf{B} \ .$$

Assume the relation holds for $1 \leq i \leq k$ and $1 \leq j \leq r$. From Equation (A.6),

$$\begin{aligned}
\mathbf{E}^{(k+1)+r} &= \mathbf{A}\mathbf{E}^{k+r} + \mathbf{B}\mathbf{E}^{k+r-1} \\
&= \mathbf{A}\left(\mathbf{E}^k \mathbf{E}^r + \mathbf{E}^{k-1}\mathbf{B}\mathbf{E}^{r-1}\right) + \mathbf{B}\left(\mathbf{E}^{k-1}\mathbf{E}^r + \mathbf{E}^{k-2}\mathbf{B}\mathbf{E}^{r-1}\right) \\
&= \left(\mathbf{A}\mathbf{E}^k + \mathbf{B}\mathbf{E}^{k-1}\right)\mathbf{E}^r + \left(\mathbf{A}\mathbf{E}^{k-1} + \mathbf{B}\mathbf{E}^{k-2}\right)\mathbf{B}\mathbf{E}^{r-1} \\
&= \mathbf{E}^{k+1}\mathbf{E}^r + \mathbf{E}^{(k+1)-1}\mathbf{B}\mathbf{E}^{r-1} \ ,
\end{aligned}$$

and we see that the relation holds for $1 \leq i \leq k+1$ and $1 \leq j \leq r$. Similarly, from Equation (A.7),

$$\begin{aligned}
\mathbf{E}^{k+(r+1)} &= \mathbf{E}^{k+r}\mathbf{A} + \mathbf{E}^{k+r-1}\mathbf{B} \\
&= \left(\mathbf{E}^{k}\mathbf{E}^{r} + \mathbf{E}^{k-1}\mathbf{B}\mathbf{E}^{r-1}\right)\mathbf{A} + \left(\mathbf{E}^{k}\mathbf{E}^{r-1} + \mathbf{E}^{k-1}\mathbf{B}\mathbf{E}^{r-2}\right)\mathbf{B} \\
&= \mathbf{E}^{k}\left(\mathbf{E}^{r}\mathbf{A} + \mathbf{E}^{r-1}\mathbf{B}\right) + \mathbf{E}^{k-1}\left(\mathbf{E}^{r-1}\mathbf{A} + \mathbf{E}^{r-2}\mathbf{B}\right) \\
&= \mathbf{E}^{k}\mathbf{E}^{r+1} + \mathbf{E}^{k-1}\mathbf{B}\mathbf{E}^{(r+1)-1} \ ,
\end{aligned}$$

and we see that the relation holds for $1 \leq i \leq k$ and $1 \leq j \leq r+1$. Combining these two, the relation holds by induction. Additionally, if we define $\mathbf{E}^{-1} = \mathbf{0}$, then

$$\begin{aligned}
\mathbf{E}^{0+0} &= \mathbf{E}^{0}\mathbf{E}^{0} + \mathbf{E}^{-1}\mathbf{B}\mathbf{E}^{-1} = \mathbf{I} \\
\mathbf{E}^{0+1} &= \mathbf{E}^{0}\mathbf{E}^{1} + \mathbf{E}^{-1}\mathbf{B}\mathbf{E}^{0} = \mathbf{A} \\
\mathbf{E}^{1+0} &= \mathbf{E}^{1}\mathbf{E}^{0} + \mathbf{E}^{0}\mathbf{B}\mathbf{E}^{-1} = \mathbf{A} \ .
\end{aligned}$$

Under this definition, we see that the relation holds for all $0 \leq i, j$. □

Theorem A.0.1 *Let matrices be defined as in Equations (3.27, A.11). Let integers $k, r \geq 0$. If,*

$$\mathcal{G}_k\left(\mathbf{A}_r, \mathbf{B}_r; \mathbf{K}^{-1}\mathbf{b}\right) \subset \mathsf{span}\left(\mathbf{X}\right) \tag{A.20}$$

$$\mathcal{G}_k\left(\mathbf{A}_l^*, \mathbf{B}_l^*; \mathbf{K}^{*,-1}\mathbf{l}\right) \subset \mathsf{span}\left(\mathbf{Y}\right) \tag{A.21}$$

then

$$\mathbf{X}\mathbf{E}_{r,R}^{i}\mathbf{K}_R^{-1}\mathbf{b}_R = \mathbf{E}_r^{i}\mathbf{K}^{-1}\mathbf{b} \qquad (0 \leq i \leq k-1) \tag{A.22}$$

$$\mathbf{l}_R^{*}\mathbf{K}_R^{-1}\mathbf{E}_{l,R}^{j}\mathbf{Y}^{*} = \mathbf{l}^{*}\mathbf{K}^{-1}\mathbf{E}_l^{i} \qquad (0 \leq j \leq r-1) \ . \tag{A.23}$$

These two relations imply,

$$\mathbf{l}_R^{*}\mathbf{E}_{r,R}^{i}\mathbf{A}_{r,R}\mathbf{E}_{r,R}^{j}\mathbf{K}_R^{-1}\mathbf{b}_R = \mathbf{l}^{*}\mathbf{E}_r^{i}\mathbf{A}_r\mathbf{E}_r^{j}\mathbf{K}^{-1}\mathbf{b} \tag{A.24}$$

$$\mathbf{l}_R^{*}\mathbf{E}_{r,R}^{i}\mathbf{B}_{r,R}\mathbf{E}_{r,R}^{j}\mathbf{K}_R^{-1}\mathbf{b}_R = \mathbf{l}^{*}\mathbf{E}_r^{i}\mathbf{B}_r\mathbf{E}_r^{j}\mathbf{K}^{-1}\mathbf{b} \tag{A.25}$$

for all $0 \leq i \leq k-1$ and $0 \leq j \leq r-1$. As a result,

$$M_i = M_{Ri} \qquad (0 \leq i \leq k+r-1) \ . \tag{A.26}$$

Proof Let projectors $\mathbf{P}, \mathbf{Q}$ be defined as in Equations (A.8,A.9). We first prove Equation (A.22) by induction. For $i = 0$ and $i = 1$, by multiple application of Lemma (A.0.2), we have,

$$\begin{aligned}
\mathbf{X}\mathbf{E}^0_{r,R}\mathbf{K}_R^{-1}\mathbf{b}_R &= \mathbf{X}\mathbf{I}\left(\mathbf{Y}^*\mathbf{K}\mathbf{X}\right)^{-1}\mathbf{Y}^*\left(\mathbf{K}\mathbf{K}^{-1}\right)\mathbf{b} \\
&= \mathbf{P}\mathbf{K}^{-1}\mathbf{b} \\
&= \mathbf{K}^{-1}\mathbf{b} \\
&= \mathbf{E}^0_r\mathbf{K}^{-1}\mathbf{b} \\
\mathbf{X}\mathbf{E}^1_{r,R}\mathbf{K}_R^{-1}\mathbf{b}_R &= \mathbf{X}\mathbf{A}_{r,R}\mathbf{K}_R^{-1}\mathbf{b}_R \\
&= \mathbf{P}\mathbf{A}_r\mathbf{X}\left(\mathbf{Y}^*\mathbf{K}\mathbf{X}\right)^{-1}\mathbf{Y}^*\left(\mathbf{K}\mathbf{K}^{-1}\right)\mathbf{b} \\
&= \mathbf{P}\mathbf{A}_r\mathbf{P}\mathbf{K}^{-1}\mathbf{b} \\
&= \mathbf{A}_r\mathbf{K}^{-1}\mathbf{b} \\
&= \mathbf{E}^1_r\mathbf{K}^{-1}\mathbf{b}\ .
\end{aligned}$$

Assume the relation holds for all integers smaller than i. Then,

$$\begin{aligned}
\mathbf{X}\mathbf{E}^i_{r,R}\mathbf{K}_R^{-1}\mathbf{b}_R &= \mathbf{X}\left(\mathbf{A}_{r,R}\mathbf{E}^{i-1}_{r,R} + \mathbf{B}_{r,R}\mathbf{E}^{i-2}\right)\mathbf{K}_R^{-1}\mathbf{b}_R \\
&= \left(\mathbf{P}\mathbf{A}_r\mathbf{X}\mathbf{E}^{i-1}_{r,R} + \mathbf{P}\mathbf{B}_r\mathbf{X}\mathbf{E}^{i-2}_{r,R}\right)\mathbf{K}_R^{-1}\mathbf{b}_R \\
&= \mathbf{P}\mathbf{A}_r\left(\mathbf{X}\mathbf{E}^{i-1}_{r,R}\mathbf{K}_R^{-1}\mathbf{b}_R\right) + \mathbf{P}\mathbf{B}_r\left(\mathbf{X}\mathbf{E}^{i-2}_{r,R}\mathbf{K}_R^{-1}\mathbf{b}_R\right) \\
&= \mathbf{P}\mathbf{A}_r\mathbf{E}^{i-1}_r\mathbf{K}^{-1}\mathbf{b} + \mathbf{P}\mathbf{B}_r\mathbf{E}^{i-2}_r\mathbf{K}^{-1}\mathbf{b} \\
&= \mathbf{P}\left(\mathbf{A}_r\mathbf{E}^{i-1}_r + \mathbf{P}\mathbf{B}_r\mathbf{E}^{i-2}_r\right)\mathbf{K}^{-1}\mathbf{b} \\
&= \mathbf{P}\mathbf{E}^i_r\mathbf{K}^{-1}\mathbf{b} \\
&= \mathbf{E}^i_r\mathbf{K}^{-1}\mathbf{b}\ .
\end{aligned}$$

Similarly, we can prove Equation (A.23) by induction. For $i = 0$ and $i = 1$, by multiple application

of Lemma (A.0.2), we have,

$$\begin{aligned}
\mathbf{l}_R^* \mathbf{K}_R^{-1} \mathbf{E}_{l,R}^0 \mathbf{Y}^* &= \mathbf{l}^* \left(\mathbf{K}^{-1}\mathbf{K}\right) \mathbf{X} \left(\mathbf{Y}^* \mathbf{K} \mathbf{X}\right)^{-1} \mathbf{I} \mathbf{Y}^* \\
&= \mathbf{l}^* \mathbf{K}^{-1} \mathbf{Q} \\
&= \mathbf{l}^* \mathbf{K}^{-1} \\
&= \mathbf{l}^* \mathbf{K}^{-1} \mathbf{E}_l^0 \\
\mathbf{l}_R^* \mathbf{K}_R^{-1} \mathbf{E}_{l,R}^1 \mathbf{Y}^* &= \mathbf{l}_R^* \mathbf{K}_R^{-1} \mathbf{A}_{l,R} \mathbf{Y}^* \\
&= \mathbf{l}^* \left(\mathbf{K}^{-1}\mathbf{K}\right) \mathbf{X} \left(\mathbf{Y}^* \mathbf{K} \mathbf{X}\right) \mathbf{Y}^* \mathbf{A}_l \mathbf{Q} \\
&= \mathbf{l}^* \mathbf{K}^{-1} \mathbf{Q} \mathbf{A}_l \mathbf{Q} \\
&= \mathbf{l}^* \mathbf{K}^{-1} \mathbf{A}_l \\
&= \mathbf{l}^* \mathbf{K}^{-1} \mathbf{E}_l^1 \ .
\end{aligned}$$

Assume the relation holds for all integers smaller than i. Then,

$$\begin{aligned}
\mathbf{l}_R^* \mathbf{K}_R^{-1} \mathbf{E}_{l,R}^i \mathbf{Y}^* &= \mathbf{l}_R^* \mathbf{K}_R^{-1} \left(\mathbf{E}_{l,R}^{i-1} \mathbf{A}_{l,R} + \mathbf{E}_{l,R}^{i-2} \mathbf{B}_{l,R}\right) \mathbf{Y}^* \\
&= \mathbf{l}_R^* \mathbf{K}_R^{-1} \left(\mathbf{E}_{l,R}^{i-1} \mathbf{Y}^* \mathbf{A}_l \mathbf{Q} + \mathbf{E}_{l,R}^{i-2} \mathbf{Y}^* \mathbf{B}_l \mathbf{Q}\right) \\
&= \left(\mathbf{l}_R^* \mathbf{K}_R^{-1} \mathbf{E}_{l,R}^{i-1} \mathbf{Y}^*\right) \mathbf{A}_l \mathbf{Q} + \left(\mathbf{l}_R^* \mathbf{K}_R^{-1} \mathbf{E}_{l,R}^{i-2} \mathbf{Y}^*\right) \mathbf{B}_l \mathbf{Q} \\
&= \mathbf{l}^* \mathbf{K}^{-1} \mathbf{E}_l^{i-1} \mathbf{A}_l \mathbf{Q} + \mathbf{l}^* \mathbf{K}^{-1} \mathbf{E}_l^{i-2} \mathbf{B}_l \mathbf{Q} \\
&= \mathbf{l}^* \mathbf{K}^{-1} \left(\mathbf{E}_l^{i-1} \mathbf{A}_l + \mathbf{E}_l^{i-2} \mathbf{B}_l\right) \mathbf{Q} \\
&= \mathbf{l}^* \mathbf{K}^{-1} \mathbf{E}_l^i \mathbf{Q} \\
&= \mathbf{l}^* \mathbf{K}^{-1} \mathbf{E}_l^i \ .
\end{aligned}$$

Since the vectors $\mathbf{b}, \mathbf{l}$ are arbitrary in the expressions for the moments in Equations (3.39,3.40), the equalities imply that,

$$\mathbf{E}_X^i \mathbf{K}^{-1} = \mathbf{K}^{-1} \mathbf{E}_Y^i \tag{A.27}$$

$$\mathbf{E}_{X,R}^i \mathbf{K}_R^{-1} = \mathbf{K}_R^{-1} \mathbf{E}_{Y,R}^i \ . \tag{A.28}$$

Equation (A.24) is obtained from this relation along with the Equations (A.22,A.23).

$$\begin{aligned}
\mathbf{l}_R^* \mathbf{E}_{r,R}^i \mathbf{A}_{r,R} \mathbf{E}_{r,R}^j \mathbf{K}_{r,R}^{-1} \mathbf{b}_R &= \mathbf{l}_R^* \left(\mathbf{K}_R^{-1} \mathbf{K}_R\right) \mathbf{E}_{r,R}^i \left(\mathbf{K}_R^{-1} \mathbf{D}_R\right) \mathbf{E}_{r,R}^j \mathbf{K}_R^{-1} \mathbf{b}_R \\
&= \mathbf{l}_R^* \mathbf{K}_R^{-1} \left(\mathbf{K}_R \mathbf{E}_{r,R}^i \mathbf{K}_R^{-1}\right) \left(\mathbf{Y}^* \mathbf{D} \mathbf{X}\right) \mathbf{E}_{r,R}^j \mathbf{K}_R^{-1} \mathbf{b}_R \\
&= \left(\mathbf{l}_R^* \mathbf{K}_R^{-1} \mathbf{E}_{l,R}^i \mathbf{Y}^*\right) \mathbf{D} \left(\mathbf{X} \mathbf{E}_{r,R}^j \mathbf{K}_R^{-1} \mathbf{b}_R\right) \\
&= \left(\mathbf{l}^* \mathbf{K}^{-1} \mathbf{E}_l^i\right) \mathbf{D} \left(\mathbf{E}_r^j \mathbf{K}^{-1} \mathbf{b}\right) \\
&= \mathbf{l}^* \mathbf{K}^{-1} \left(\mathbf{K} \mathbf{E}_r^i \mathbf{K}^{-1}\right) \mathbf{D} \mathbf{E}_r^j \mathbf{K}^{-1} \mathbf{b} \\
&= \mathbf{l}^* \mathbf{E}_r^i \left(\mathbf{K}^{-1} \mathbf{D}\right) \mathbf{E}_r^j \mathbf{K}^{-1} \mathbf{b} \\
&= \mathbf{l}^* \mathbf{E}_r^i \mathbf{A}_r \mathbf{E}_r^j \mathbf{K}^{-1} \mathbf{b}
\end{aligned}$$

Equation (A.25) is obtained by the same procedure as above with the substitutions,

$$\mathbf{A}_{r,R} \rightarrow \mathbf{B}_{r,R} \qquad \mathbf{A}_r \rightarrow \mathbf{B}_r \quad .$$

Now for $0 \leq i \leq k-1$ and $0 \leq j \leq r-1$, using Lemma A.0.4 we have,

$$\begin{aligned}
M_{Ri+j+1} &= \mathbf{l}_R^* \mathbf{E}_{r,R}^{i+j+1} \mathbf{K}_R^{-1} \mathbf{b}_R \\
&= \mathbf{l}_R^* \left(\mathbf{E}_{r,R}^{i+1} \mathbf{E}_{r,R}^j + \mathbf{E}_{r,R}^i \mathbf{B}_{r,R} \mathbf{E}_{r,R}^{j-1}\right) \mathbf{K}_R^{-1} \mathbf{b}_R \\
&= \mathbf{l}_R^* \left\{ \left(\mathbf{E}_{r,R}^i \mathbf{A}_{r,R} + \mathbf{E}_{r,R}^{i-1} \mathbf{B}_{r,R}\right) \mathbf{E}_{r,R}^j + \mathbf{E}_{r,R}^i \mathbf{B}_{r,R} \mathbf{E}_{r,R}^{j-1} \right\} \mathbf{K}_R^{-1} \mathbf{b}_R \\
&= \mathbf{l}_R^* \mathbf{E}_{r,R}^i \mathbf{A}_{r,R} \mathbf{E}_{r,R}^j \mathbf{K}_R^{-1} \mathbf{b}_R + \mathbf{l}_R^* \mathbf{E}_{r,R}^{i-1} \mathbf{B}_{r,R} \mathbf{E}_{r,R}^j \mathbf{K}_R^{-1} \mathbf{b}_R \\
&\quad + \mathbf{l}_R^* \mathbf{E}_{r,R}^i \mathbf{B}_{r,R} \mathbf{E}_{r,R}^{j-1} \mathbf{K}_R^{-1} \mathbf{b}_R \\
&= \mathbf{l}^* \mathbf{E}_r^i \mathbf{A}_r \mathbf{E}_r^j \mathbf{K}^{-1} \mathbf{b} + \mathbf{l}^* \mathbf{E}_r^{i-1} \mathbf{B}_r \mathbf{E}_r^j \mathbf{K}^{-1} \mathbf{b} + \mathbf{l}^* \mathbf{E}_r^i \mathbf{B}_r \mathbf{E}_r^{j-1} \mathbf{K}^{-1} \mathbf{b} \\
&= \mathbf{l}^* \left\{ \left(\mathbf{E}_r^i \mathbf{A}_r + \mathbf{E}_r^{i-1} \mathbf{B}_r\right) \mathbf{E}_r^j + \mathbf{E}_r^i \mathbf{B}_r \mathbf{E}_r^{j-1} \right\} \mathbf{K}^{-1} \mathbf{b} \\
&= \mathbf{l}^* \left(\mathbf{E}_r^{i+1} \mathbf{E}_r^j + \mathbf{E}_r^i \mathbf{B}_r \mathbf{E}_r^{j-1}\right) \mathbf{K}^{-1} \mathbf{b} \\
&= \mathbf{l}_R^* \mathbf{E}_r^{i+j+1} \mathbf{K}^{-1} \mathbf{b} \\
&= M_{i+j+1} \ .
\end{aligned}$$

□

Appendix B

Electrostatic electromechanical coupling

The force, charge, and consistent tangent stiffness terms are derived from the nonlinear electrostatic electromechanical potential energy introduced in Section 4.4,

$$\Pi^{es}(\boldsymbol{\varphi}, \phi) := \frac{1}{2} \int_{\boldsymbol{\varphi}(\Omega)} \mathbf{E}(\phi) \cdot \mathbf{E}(\phi) \; d\Omega, \tag{B.1}$$

$$= \frac{1}{2} \int_{\Omega} \nabla_{\mathbf{X}} \phi \cdot \mathbf{C}^{-1} \nabla_{\mathbf{X}} \phi \; J \; d\Omega, \tag{B.2}$$

$$\mathbf{E} := -\nabla_{\mathbf{x}} \phi. \tag{B.3}$$

Ω is the reference (undeformed) configuration, $\boldsymbol{\varphi}$ is the deformation mapping, $\mathbf{F}$ is the deformation gradient, $\mathbf{C} := \mathbf{F}^T\mathbf{F}$ is the right Cauchy-Green tensor, $J := \det(\mathbf{F})$ is the Jacobian of the deformation, ϕ is the potential, and $\mathbf{E}$ is the electrical field, Here the permittivity $\epsilon_r \epsilon_0 = 1$ is assumed without loss of generality.

By taking the first variation of this energy, one obtains,

$$\begin{aligned}
\delta\Pi^{es}(\phi,\varphi) &= \Pi^{es}_{\phi}[\delta\phi] + \Pi^{es}_{\boldsymbol{\varphi}}[\delta\boldsymbol{\varphi}], && \text{(B.4)}\\
\mathbf{Q}[\delta\phi] &:= \Pi^{es}_{\phi}[\delta\phi]\\
&= \int_{\boldsymbol{\varphi}(\Omega)} \nabla_{\mathbf{x}}\delta\phi\cdot\nabla_{\mathbf{x}}\phi d\Omega\\
&= \int_{\boldsymbol{\varphi}(\Omega)} \nabla_{\mathbf{x}}\delta\mathbf{E}\cdot\mathbf{E}d\Omega, && \text{(B.5)}\\
\mathbf{F}[\delta\phi] &:= \Pi^{es}_{\boldsymbol{\varphi}}[\delta\boldsymbol{\varphi}]\\
&= \frac{1}{2}\int_{\Omega} \nabla_{\mathbf{X}}\phi\mathbf{C}^{-1}\nabla_{\mathbf{X}}\phi\delta J + \nabla_{\mathbf{X}}\phi\delta\left(\mathbf{C}^{-1}\right)\nabla_{\mathbf{X}}\phi J d\Omega\\
&= \frac{1}{2}\int_{\Omega} \nabla_{\mathbf{X}}\phi\mathbf{C}^{-1}\nabla_{\mathbf{X}}\phi J\mathbf{F}^{-T} : \delta F - \nabla_{\mathbf{X}}\phi\mathbf{C}^{-1}\delta\mathbf{C}\mathbf{C}^{-1}\nabla_{\mathbf{X}}\phi J d\Omega\\
&= \frac{1}{2}\int_{\Omega} \nabla_{\mathbf{x}}\phi\nabla_{\mathbf{x}}\phi J\mathrm{tr}\left(\nabla_{\mathbf{x}}\delta\varphi\right) - \nabla_{\mathbf{x}}\phi\left(\delta\mathbf{F}\mathbf{F}^{-1} + \mathbf{F}^{-T}\delta\mathbf{F}^{T}\right)\nabla_{\mathbf{x}}\phi J d\Omega\\
&= \frac{1}{2}\int_{\Omega} \nabla_{\mathbf{x}}\phi\nabla_{\mathbf{x}}\phi J\mathrm{tr}\left(\nabla_{\mathbf{x}}\delta\varphi\right) - \nabla_{\mathbf{x}}\phi 2\mathrm{sym}\left(\nabla_{\mathbf{x}}\delta\varphi\right)\nabla_{\mathbf{x}}\phi J d\Omega\\
&= \frac{1}{2}\int_{\boldsymbol{\varphi}(\Omega)} \nabla_{\mathbf{x}}\phi\left[\mathrm{tr}\left(\nabla_{\mathbf{x}}\delta\varphi\right)\mathbf{1} - 2\mathrm{sym}\left(\nabla_{\mathbf{x}}\delta\varphi\right)\right]\nabla_{\mathbf{x}}\phi d\Omega\\
&= \int_{\boldsymbol{\varphi}(\Omega)} \nabla_{\mathbf{x}}\delta\boldsymbol{\varphi} : \left[\frac{1}{2}\mathbf{E}\cdot\mathbf{E} - \mathbf{E}\otimes\mathbf{E}\right] d\Omega. && \text{(B.6)}
\end{aligned}$$

$\mathbf{F}$ denotes the linear form representing the force term and $\mathbf{Q}$ denotes the linear form representing

the charge term. The consistent tangent stiffness is obtained by taking the second variation.

$$\Delta\left(\delta\Pi\right) = \Pi_{\boldsymbol{\varphi}\boldsymbol{\varphi}}[\delta\boldsymbol{\varphi},\Delta\boldsymbol{\varphi}] + \Pi_{\boldsymbol{\varphi}\phi}[\delta\boldsymbol{\varphi},\Delta\phi] + \Pi_{\phi\boldsymbol{\varphi}}[\delta\phi,\Delta\boldsymbol{\varphi}] + \Pi_{\phi\phi}[\delta\phi,\Delta\phi], \tag{B.7}$$

$$\begin{aligned}
\mathbf{K}_{\phi\phi}[\delta\phi,\Delta\phi] &:= \Pi_{\phi\phi}[\delta\phi,\Delta\phi], \\
&= \int_{\boldsymbol{\varphi}(\Omega)} \nabla_{\mathbf{x}}\delta\phi\cdot\nabla_{\mathbf{x}}\Delta\phi d\Omega \\
&= \int_{\boldsymbol{\varphi}(\Omega)} \delta\mathbf{E}\cdot\Delta\mathbf{E} d\Omega,
\end{aligned} \tag{B.8}$$

$$\begin{aligned}
\mathbf{K}_{\boldsymbol{\varphi}\phi}[\delta\boldsymbol{\varphi},\Delta\phi] &:= \Pi_{\boldsymbol{\varphi}\phi}[\delta\boldsymbol{\varphi},\Delta\phi], \\
&= \int_{\boldsymbol{\varphi}(\Omega)} \nabla_{\mathbf{x}}\phi \mathrm{sym}\left[\mathrm{tr}\left(\nabla_{\mathbf{x}}\delta\varphi\right)\mathbf{1} - 2\nabla_{\mathbf{x}}\delta\varphi\right]\nabla_{\mathbf{x}}\Delta\phi d\Omega \\
&= \int_{\boldsymbol{\varphi}(\Omega)} \nabla_{\mathbf{x}}\delta\boldsymbol{\varphi} : \left[\Delta\mathbf{E}\cdot\mathbf{E} - \frac{1}{2}\Delta\mathbf{E}\otimes\mathbf{E} - \frac{1}{2}\mathbf{E}\otimes\Delta\mathbf{E}\right] d\Omega,
\end{aligned} \tag{B.9}$$

$$\begin{aligned}
\mathbf{K}_{\phi\boldsymbol{\varphi}}[\delta\phi,\Delta\boldsymbol{\varphi}] &:= \Pi_{\phi\boldsymbol{\varphi}}[\delta\phi,\Delta\boldsymbol{\varphi}] \\
&= \int_{\boldsymbol{\varphi}(\Omega)} \nabla_{\mathbf{x}}\delta\phi \mathrm{sym}\left[\mathrm{tr}\left(\nabla_{\mathbf{x}}\Delta\varphi\right)\mathbf{1} - 2\nabla_{\mathbf{x}}\Delta\varphi\right]\nabla_{\mathbf{x}}\phi d\Omega \\
&= \int_{\boldsymbol{\varphi}(\Omega)} \left[\delta\mathbf{E}\cdot\mathbf{E} - \frac{1}{2}\delta\mathbf{E}\otimes\mathbf{E} - \frac{1}{2}\mathbf{E}\otimes\delta\mathbf{E}\right] : \nabla_{\mathbf{x}}\Delta\boldsymbol{\varphi} d\Omega,
\end{aligned} \tag{B.10}$$

$$\begin{aligned}
\mathbf{K}_{\boldsymbol{\varphi}\boldsymbol{\varphi}}[\delta\boldsymbol{\varphi},\Delta\boldsymbol{\varphi}] &= \Delta\left\{\frac{1}{2}\int_{\boldsymbol{\varphi}(\Omega)} \nabla_{\mathbf{x}}\phi \mathrm{sym}\left[\mathrm{tr}\left(\nabla_{\mathbf{x}}\delta\varphi\right)\mathbf{1} - 2\nabla_{\mathbf{x}}\delta\varphi\right]\nabla_{\mathbf{x}}\phi d\Omega\right\} \\
&= \Delta\left\{\frac{1}{2}\int_{\Omega} \nabla_{\mathbf{X}}\phi\mathbf{F}^{-1}\mathrm{sym}\left[\mathrm{tr}\left(\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\right)\mathbf{1} - 2\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\right]\mathbf{F}^{-T}\nabla_{\mathbf{X}}\phi J d\Omega\right\} \\
&= A_1 + A_2 + A_3 + A_4
\end{aligned} \tag{B.11}$$

$$
\begin{aligned}
A_1 &= \frac{1}{2}\int_{\Omega} \nabla_{\mathbf{X}}\phi\mathbf{F}^{-1}\mathrm{sym}\left[\mathrm{tr}\left(\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\right)\mathbf{1} - 2\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\right]\mathbf{F}^{-T}\nabla_{\mathbf{X}}\phi\Delta J d\Omega \\
&= \frac{1}{2}\int_{\Omega} \nabla_{\mathbf{X}}\phi\mathbf{F}^{-1}\mathrm{sym}\left[\mathrm{tr}\left(\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\right)\mathbf{1} - 2\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\right]\mathbf{F}^{-T}\nabla_{\mathbf{X}}\phi J\mathrm{tr}\left(\Delta\mathbf{F}\mathbf{F}^{-1}\right) d\Omega \\
&= \frac{1}{2}\int_{\Omega} \nabla_{\mathbf{X}}\phi\mathbf{F}^{-1}\mathrm{sym}\left[\mathrm{tr}\left(\nabla_{\mathbf{X}}\Delta\varphi\mathbf{F}^{-1}\right)\mathrm{tr}\left(\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\right)\mathbf{1} - 2\mathrm{tr}\left(\nabla_{\mathbf{X}}\Delta\varphi\mathbf{F}^{-1}\right)\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\right] \\
&\qquad\qquad \mathbf{F}^{-T}\nabla_{\mathbf{X}}\phi J d\Omega \\
&= \frac{1}{2}\int_{\boldsymbol{\varphi}(\Omega)} \nabla_{\mathbf{x}}\phi\mathrm{sym}\left[\mathrm{tr}\left(\nabla_{\mathbf{x}}\Delta\varphi\right)\mathrm{tr}\left(\nabla_{\mathbf{x}}\delta\varphi\right)\mathbf{1} - 2\mathrm{tr}\left(\nabla_{\mathbf{x}}\Delta\varphi\right)\nabla_{\mathbf{x}}\delta\varphi\right]\nabla_{\mathbf{x}}\phi d\Omega \qquad \text{(B.12)} \\
A_2 &= \frac{1}{2}\int_{\Omega} \nabla_{\mathbf{X}}\phi\Delta\left(\mathbf{F}^{-1}\right)\mathrm{sym}\left[\mathrm{tr}\left(\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\right)\mathbf{1} - 2\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\right]\mathbf{F}^{-T}\nabla_{\mathbf{X}}\phi J d\Omega \\
&= -\frac{1}{2}\int_{\Omega} \nabla_{\mathbf{X}}\phi\mathbf{F}^{-1}\Delta\mathbf{F}\mathbf{F}^{-1}\mathrm{sym}\left[\mathrm{tr}\left(\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\right)\mathbf{1} - 2\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\right]\mathbf{F}^{-T}\nabla_{\mathbf{X}}\phi J d\Omega \\
&= -\frac{1}{2}\int_{\Omega} \nabla_{\mathbf{X}}\phi\mathbf{F}^{-1}\left[\mathrm{tr}\left(\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\right)\Delta\mathbf{F}\mathbf{F}^{-1} - 2\Delta\mathbf{F}\mathbf{F}^{-1}\mathrm{sym}\left(\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\right)\right]\mathbf{F}^{-T}\nabla_{\mathbf{X}}\phi J d\Omega \\
&= -\frac{1}{2}\int_{\boldsymbol{\varphi}(\Omega)} \nabla_{\mathbf{x}}\phi\left[\mathrm{tr}\left(\nabla_{\mathbf{x}}\delta\varphi\right)\nabla_{\mathbf{x}}\Delta\varphi - 2\nabla_{\mathbf{x}}\Delta\varphi\mathrm{sym}\left(\nabla_{\mathbf{x}}\delta\varphi\right)\right]\nabla_{\mathbf{x}}\phi d\Omega \qquad \text{(B.13)} \\
A_3 &= \frac{1}{2}\int_{\Omega} \nabla_{\mathbf{X}}\phi\mathbf{F}^{-1}\mathrm{sym}\left[\mathrm{tr}\left(\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\right)\mathbf{1} - 2\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\right]\Delta\left(\mathbf{F}^{-T}\right)\nabla_{\mathbf{X}}\phi J d\Omega \\
&= -\frac{1}{2}\int_{\Omega} \nabla_{\mathbf{X}}\phi\mathbf{F}^{-1}\mathrm{sym}\left[\mathrm{tr}\left(\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\right)\mathbf{1} - 2\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\right]\mathbf{F}^{-T}\Delta\mathbf{F}^{T}\mathbf{F}^{-T}\nabla_{\mathbf{X}}\phi J d\Omega \\
&= -\frac{1}{2}\int_{\Omega} \nabla_{\mathbf{X}}\phi\mathbf{F}^{-1}\left[\mathrm{tr}\left(\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\right)\mathbf{F}^{-T}\Delta\mathbf{F}^{T} - 2\mathrm{sym}\left(\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\right)\mathbf{F}^{-T}\Delta\mathbf{F}^{T}\right]\mathbf{F}^{-T}\nabla_{\mathbf{X}}\phi J d\Omega \\
&= -\frac{1}{2}\int_{\boldsymbol{\varphi}(\Omega)} \nabla_{\mathbf{x}}\phi\left[\mathrm{tr}\left(\nabla_{\mathbf{x}}\delta\varphi\right)\nabla_{\mathbf{x}}\Delta\varphi^{T} - 2\mathrm{sym}\left(\nabla_{\mathbf{x}}\delta\varphi\right)\nabla_{\mathbf{x}}\Delta\varphi^{T}\right]\nabla_{\mathbf{x}}\phi d\Omega \qquad \text{(B.14)} \\
A_4 &= \frac{1}{2}\int_{\Omega} \nabla_{\mathbf{X}}\phi\mathbf{F}^{-1}\mathrm{sym}\left[\mathrm{tr}\left(\nabla_{\mathbf{X}}\delta\varphi\Delta\left(\mathbf{F}^{-1}\right)\right)\mathbf{1} - 2\nabla_{\mathbf{X}}\delta\varphi\Delta\left(\mathbf{F}^{-1}\right)\right]\mathbf{F}^{-T}\nabla_{\mathbf{X}}\phi J d\Omega \\
&= -\frac{1}{2}\int_{\Omega} \nabla_{\mathbf{X}}\phi\mathbf{F}^{-1}\mathrm{sym}\left[\mathrm{tr}\left(\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\Delta\mathbf{F}\mathbf{F}^{-1}\right)\mathbf{1} - 2\nabla_{\mathbf{X}}\delta\varphi\mathbf{F}^{-1}\Delta\mathbf{F}\mathbf{F}^{-1}\right]\mathbf{F}^{-T}\nabla_{\mathbf{X}}\phi J d\Omega \\
&= -\frac{1}{2}\int_{\boldsymbol{\varphi}(\Omega)} \nabla_{\mathbf{x}}\phi\mathrm{sym}\left[\mathrm{tr}\left(\nabla_{\mathbf{x}}\delta\varphi\nabla_{\mathbf{x}}\Delta\varphi\right)\mathbf{1} - 2\nabla_{\mathbf{x}}\delta\varphi\nabla_{\mathbf{x}}\Delta\varphi\right]\nabla_{\mathbf{x}}\phi d\Omega \qquad \text{(B.15)}
\end{aligned}
$$

Summing the terms, $\mathbf{K}_{\boldsymbol{\varphi\varphi}}$ is given as,

$$
\begin{aligned}
\mathbf{K}_{\boldsymbol{\varphi\varphi}}[\delta\boldsymbol{\varphi}, \Delta\boldsymbol{\varphi}] &= \frac{1}{2}\int_{\boldsymbol{\varphi}(\Omega)} \nabla_{\mathbf{x}}\phi \,[\mathrm{tr}\,(\nabla_{\mathbf{x}}\Delta\varphi)\,\mathrm{tr}\,(\nabla_{\mathbf{x}}\delta\varphi)\,\mathbf{1} - 2\mathrm{tr}\,(\nabla_{\mathbf{x}}\Delta\varphi)\,\mathrm{sym}\,(\nabla_{\mathbf{x}}\delta\varphi) \\
&\qquad -\mathrm{tr}\,(\nabla_{\mathbf{x}}\delta\varphi)\,\nabla_{\mathbf{x}}\Delta\varphi + 2\nabla_{\mathbf{x}}\Delta\varphi\mathrm{sym}\,(\nabla_{\mathbf{x}}\delta\varphi) \\
&\qquad -\mathrm{tr}\,(\nabla_{\mathbf{x}}\delta\varphi)\,\nabla_{\mathbf{x}}\Delta\varphi^T + 2\mathrm{sym}\,(\nabla_{\mathbf{x}}\delta\varphi)\,\nabla_{\mathbf{x}}\Delta\varphi^T \\
&\qquad -\mathrm{tr}\,(\nabla_{\mathbf{x}}\delta\varphi\nabla_{\mathbf{x}}\Delta\varphi)\,\mathbf{1} + 2\mathrm{sym}\,(\nabla_{\mathbf{x}}\delta\varphi\nabla_{\mathbf{x}}\Delta\varphi)\,]\,\nabla_{\mathbf{x}}\phi d\Omega \\
&= \frac{1}{2}\int_{\boldsymbol{\varphi}(\Omega)} \nabla_{\mathbf{x}}\phi \,[\mathrm{tr}\,(\nabla_{\mathbf{x}}\Delta\varphi)\,\mathrm{tr}\,(\nabla_{\mathbf{x}}\delta\varphi)\,\mathbf{1} - \mathrm{tr}\,(\nabla_{\mathbf{x}}\delta\varphi\nabla_{\mathbf{x}}\Delta\varphi)\,\mathbf{1} \\
&\qquad -2\mathrm{tr}\,(\nabla_{\mathbf{x}}\Delta\varphi)\,\mathrm{sym}\,(\nabla_{\mathbf{x}}\delta\varphi) - 2\mathrm{tr}\,(\nabla_{\mathbf{x}}\delta\varphi)\,\mathrm{sym}\,(\nabla_{\mathbf{x}}\Delta\varphi) \\
&\qquad +2\mathrm{sym}\,(\nabla_{\mathbf{x}}\delta\varphi\nabla_{\mathbf{x}}\Delta\varphi) \\
&\qquad +2\nabla_{\mathbf{x}}\Delta\varphi\mathrm{sym}\,(\nabla_{\mathbf{x}}\delta\varphi) + 2\mathrm{sym}\,(\nabla_{\mathbf{x}}\delta\varphi)\,\nabla_{\mathbf{x}}\Delta\varphi^T\,]\,\nabla_{\mathbf{x}}\phi d\Omega \\
&= \frac{1}{2}\int_{\boldsymbol{\varphi}(\Omega)} \nabla_{\mathbf{x}}\phi \,[\mathrm{tr}\,(\nabla_{\mathbf{x}}\delta\varphi)\,\mathrm{tr}\,(\nabla_{\mathbf{x}}\Delta\varphi)\,\mathbf{1} - \mathrm{tr}\,(\nabla_{\mathbf{x}}\Delta\varphi\nabla_{\mathbf{x}}\delta\varphi)\,\mathbf{1} \\
&\qquad -2\mathrm{tr}\,(\nabla_{\mathbf{x}}\Delta\varphi)\,\mathrm{sym}\,(\nabla_{\mathbf{x}}\delta\varphi) \\
&\qquad -2\mathrm{tr}\,(\nabla_{\mathbf{x}}\delta\varphi)\,\mathrm{sym}\,(\nabla_{\mathbf{x}}\Delta\varphi) \\
&\qquad +2\mathrm{sym}\,(\nabla_{\mathbf{x}}\delta\varphi\nabla_{\mathbf{x}}\Delta\varphi) \\
&\qquad +2\mathrm{sym}\,(\nabla_{\mathbf{x}}\Delta\varphi\nabla_{\mathbf{x}}\delta\varphi) \\
&\qquad +2\mathrm{sym}\left(\nabla_{\mathbf{x}}\delta\varphi\nabla_{\mathbf{x}}\Delta\varphi^T\right)\,]\,\nabla_{\mathbf{x}}\phi d\Omega
\end{aligned}
\tag{B.16}
$$

Bibliography

[1] M.A. Abdelmoneum, M.U. Demirici, and C.T.-C. Nguyen. Stemless Wine-Glass-Mode Disk Micromechanical Resonators. In *MEMS'03*, pages 698–701, 2003.

[2] R. Abdolvand, G.K. Ho, A. Erbil, and F. Ayazi. Thermoelastic Damping in Trenched-Refilled Polysilicon Resonators. In *Transducers'03*, pages 324–327, 2003.

[3] M. Adams. *Multigrid Equation Solvers for Large Scale Nonlinear Finite Element Simulations*. PhD thesis, University of California, Berkeley, 1999.

[4] M. Adams. Evaluation of three unstructured multigrid methods on 3D finite element problems in solid mechanics. *International Journal for Numerical Methods in Engineering*, 55:519–534, 2002.

[5] M. Adams. Algebraic multigrid methods for direct frequency response analyses in solid mechanics. *Computational Mechanics*, 39:497–507, 2007.

[6] M. Adams, M. Brezina, and J. Hu. Parallel multigrid smoothing: polynomial versus Gauss-Seidel. *Journal of Computational Physics*, 188(2):593–610, 2003.

[7] A. Agrawal and A. Sharma. Perfectly matched layer in numerical wave propagation: factors that affect its performance. *Applied Optics*, 43(21):4225–4231, 2004.

[8] R. Aigner, S. Marksteiner, L. Elbrecht, and W. Nessler. RF-Filters in Mobile Phone Applications. In *Transducers'03*, pages 951–954, 2003.

[9] A. Akhieser. On the Absorption of Sound in Solids. *Journal of Physics*, I(4):277–287, 1939.

[10] F.J. Alexander, A.L. Garcia, and B.J. Alder. Direct simulation Monte Carlo for thin-film bearings. *Physics of Fluids*, 6:3854–3860, 1994.

[11] P. R. Amestoy, I. S. Duff, and J.-Y. L'Excellent. Multifrontal parallel distributed symmetricand unsymmetric solvers. *Comput. Methods Appl. Mech. Eng.*, 184:501–520, 2000.

[12] B. Antkowiak, J.P. Gorman, M. Varghese, D.J.D. Carter, and A.E. Duwel. Design of a High-Q, Low-Impedence, GHz-Range Piezoelectric MEMS Resonator. In *Transducers'03*, pages 841–846, 2003.

[13] P. Arbenz and M.E. Hochstenbach. A Jacobi-Davidson method for solving complex symmetric eigenvalue problems. *SIAM Journal on Scientific Computing*, 25(5):1655–1673, 2004.

[14] N.W. Ashcroft and N.D. Mermin. *Solid State Physics*. Brooks/Cole, 1976.

[15] S. Asvadurov, V. Druskin, M.N. Guddati, and L. Knizhnerman. On Optimal Finite-Difference Approximation of PML. *SIAM Journal of Numerical Analysis*, 41(1):287–305, 2003.

[16] I.M. Babuska and S.A. Sauter. Is the Pollution Effect of the FEM Avoidable for the Helmholtz Equation Considering High Wave Numbers? *SIAM Review*, 42(3):451–484, 2000.

[17] Z. Bai. Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems. *Applied Numerical Mathematics*, 43:9–44, 2002.

[18] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst. *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*. SIAM, Philadelphia, 2000.

[19] Z. Bai and Y. Su. Dimension reduction of second-order dynamical systems via a second-order Arnoldi method. *SIAM Journal of Scientific Computing*, 26(5):1692–1709, 2005.

[20] Z. Bai and Y. Su. SOAR: A second-order Arnoldi method for the solution of the quadratic eigenvalue problem. *SIAM Journal of Matrix Analysis and Applications*, 26(3):640–659, 2005.

[21] Satish Balay, Kris Buschelman, William D. Gropp, Dinesh Kaushik, Matthew G. Knepley, Lois Curfman McInnes, Barry F. Smith, and Hong Zhang. PETSc Web page, 2001. http://www.mcs.anl.gov/petsc.

[22] R.E. Bank. A Comparison of Two Multilevel Iterative Methods for Nonsymmetric and Indefinite Elliptic Finite Element Equations. *SIAM Journal of Numerical Analysis*, 18(4):724–743, 1981.

[23] M. Bao, H. Yang, H. Yin, and Y. Sun. Energy transfer model for squeeze-film air damping in low vacuum. *Journal of Micromechics and Microengineering*, 12:341–346, 2002.

[24] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. Van der Vorst. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods, 2nd Edition*. SIAM, Philadelphia, PA, 1994.

[25] U. Basu and A.K. Chopra. Perfectly matched layers for time-harmonic elastodynamics of unbounded domains: theory and finite-element implementation. *Computer Methods in Applied Mechanics and Engineering*, 192:1337–1375, 2003.

[26] U. Basu and A.K. Chopra. Perfectly matched layers for transient elastodynamics of unbounded domains. *International Journal for Numerical Methods in Engineering*, 59:1039–1074, 2004.

[27] T. Belytschko, W.K. Liu, and B. Moran. *Nonlinear Finite Elements for Continua and Structure*. Wiley, 2000.

[28] M. Benzi. Preconditioning Techniques for Large Linear Systems: A Survey. *Journal of Computational Physics*, 182:418–477, 2002.

[29] J.-P. Bérenger. A Perfectly Matched Layer for the Absorption of Electromagnetic Waves. *Journal of Computational Physics*, 114:185–200, 1994.

[30] J.-P. Bérenger. Perfectly matched layer for the FDTD solution of wave-structure interaction problems. *IEEE Transactions on Antennas and Propagation*, 44(1):110–117, 1996.

[31] A. Bermudez, L. Hervella-Nieto, and R. Rodriguez A. Prieto. An optimal perfectly matched layer with unbounded absorbing function for time-harmonic acoustic scattering problems. *Journal of Computational Physics*, 223:469–488, 2007.

[32] S.A. Bhave and R.T. Howe. Internal Electrostatic Transduction for Bulk-Mode MEMS Resonators. In *Hilton Head 2004*, 2004.

[33] D. Bindel. *Structured and Parameter-Dependent Eigensolvers for Simulation-Based Design of Resonant MEMS*. PhD thesis, University of California, Berkeley, 2006.

[34] D. Bindel, Z. Bai, and J. Demmel. Model reduction for RF MEMS simulation. In *Proceedings of PARA 04*, 2004.

[35] David Bindel. HiQLab. http://cims.nyu.edu/ dbindel/hiqlab/.

[36] David Bindel. Homepage of David Bindel. http://cims.nyu.edu/ dbindel/.

[37] D.S. Bindel and S. Govindjee. Elastic PMLs for resonator anchor loss simulation. *International Journal for Numerical Methods in Engineering*, 64:789–818, 2005.

[38] D.S. Bindel, E. Quévy, T. Koyama, S. Govindjee, J.W. Demmel, and R.T. Howe. Anchor Loss Simulation in Resonators. In *MEMS'05*, pages 133–136, 2005.

[39] B. Bircumshaw, G. Liu, H. Takeuchi, T.-J. King, R. Howe, O. Reilly, and A. Pisano. The Radial Bulk Annular Resonator: Towards a 50Ω RF MEMS Filter. In *Transducers '03*, pages 875–878, 2003.

[40] Dietrich Braess. *Finite Elements*. Cambridge, 2002.

[41] J.H. Bramble, D.Y. Kwak, and J.E. Pasciak. Uniform Convergence of Multigrid V-Cycle Iterations for Indefinite and Nonsymmetric Problems. *SIAM Journal of Numerical Analysis*, 31(6):1746–1763, 1994.

[42] A. Brandt and S. Ta'asan. Multigrid method for nearly singlular and slightly indefinite problems. *Lecture Nots in Mathematics: Multigrid Methods II, Springer-Verlag*, 1985.

[43] J. Bustillo, R.T. Howe, and R.S. Muller. Surface Micromachining of Microelectromechanical Systems. *Proceedings of the IEEE*, 86(8):1552–6 and 1559–63, 1998.

[44] D. Calvetti, G.H. Golub, and L. Reichel. An adaptive Chebyshev iterative method for nonsymmetric linear systems based on modified moments. *Numerische Mathematik*, 67:21–40, 1994.

[45] R.N. Candler, A. Duwel, M. Varghese, S.A. Chandorkar, M.A. Hopcroft, W.-T. Park, B. Kim, G. Yama, A. Partridge, M. Lutz, and T.W. Kenny. Impact of Geometry on Thermoelastic Dissipation in Micromechanical Resonant Beams. *Journal of Microelectromechanical Systems*, 15(4):927–934, 2006.

[46] R.N. Candler, H. Li, M. Lutz, W.-T. Park, A. Partridge, G. Yama, and T.W. Kenny. Investigation of Energy Loss Mechanisms in Micromechanical Resonators. In *Transducers'03*, pages 332–335, 2003.

[47] Y. Censor, D. Gordon, and R. Gordon. Component averaging: An efficient iterative parallel algorithm for large and sparse unstructured problems. *Parallel Computing*, 27(6):777–808, 2001.

[48] H. Chandrahalim, S. Bhave, E.P. Quévy, and R.T. Howe. Aqueous Transduction of Poly-SiGe Disk Resonators. In *Transducers and Eurosensors '07*, 2007.

[49] H. Chandrahalim, D. Weinstein, L.F. Cheow, and S.A. Bhave. High-κ dielectrically transduced MEMS thickness shear mode resonators and tunable channel-select RF filters. *Sensors and Actuators A*, 136:527–539, 2007.

[50] W.C. Chew and Q.H. Liu. Perfectly Matched Layers for Elastodynamics: A New Absorbing Boundary Condition. *Journal of Computational Acoustics*, 4(4):341–359, 1996.

[51] Y.-H. Cho, B.M. Kwak, A.P. Pisano, and R.T. Howe. Viscous damping model or laterally oscillating microstructures. *Journal of Microelectromechanical Systems*, 3:81–87, 1994.

[52] E. Chow. A priori sparsity patterns for parallel sparse approximate inverse preconditioners. *SIAM Journal of Scientific Computing*, 21(5):1804–1822, 2000.

[53] R.G. Christian. The theory of oscillating-vane vacuum gauges. *Vacuum*, 16(4):175–178, 1966.

[54] P.G. Ciarlet. *Mathematical Elasticity*. Elsevier, 1988.

[55] F. Collino and P.G. Monk. Optimizing the perfectly matched layer. *Computer Methods in Applied Mechanics and Engineering*, 164:157–171, 1998.

[56] F. Collino and C. Tsogka. Application of the perfectly matched absorbing layer model to the linear elastodynamic problem in anisotropic heterogeneous media. *Geophysics*, 66(1):294–307, 2001.

[57] T.A. Davis. *Direct Methods for Sparse Linear Systems*. SIAM, 2006.

[58] A. DAX. The Convergence of Linear Stationary Iterative Processes for Solving Singular Unstructured Systems of Linear Equations. *SIAM Review*, 32(4):611–635, 1990.

[59] D. Day. An Effecient Implementation of the Unsymmetric Lanczos Algorithm. *SIAM Journal of Matrix Analysis and Applications*, 18(3):566–589, 1997.

[60] S.K. De and N.R. Aluru. Theory of thermoelastic damping in electrostatically actuated microstructures. *Physical Review B*, 74(14):144305–1–13, 2006.

[61] M.U. Demirci and C.T.-C. Nguyen. Mechanically Corner-Coupled Square Microresonator Array for Reduced Series Motional Resistance. *Journal of Microelectromechanical Systems*, 15(6):1419–1436, 2006.

[62] James Demmel and Xiaoye Li. SuperLU_Dist, 1997. http://www.cs.berkeley.edu/ demmel/SuperLU.html.

[63] James W. Demmel. *Applied Numerical Linear Algebra.* SIAM, 1997.

[64] M. Dubois and P. Muralt. Stress and piezelectric properties of Aluminum Nitride thin films deposited onto metal electrodes by Pulsed Direct Current Reactive Sputtering. *Journal of Applied Physics*, 89(11):6389–6395, 2001.

[65] A. Duwel, R.N. Candler, T.W. Kenny, and M. Varghese. Engineering MEMS Resonators With Low Thermoelastic Damping. *Journal of Microelectromechanical Systems*, 15(6):1437–1445, 2006.

[66] A. Duwel, J. Gorman, M. Weinstein, J. Borenstein, and P. Ward. Experimental study of thermoelastic damping in MEMS gyros. *Sensors and Actuators A*, 103:70–75, 2003.

[67] H.C. Elman, O.G. Ernst, and D.P. O'Leary. A multigrid method enhanced by Krylov subspace iteration for discrete Helmholtz equations. *SIAM Journal on Scientific Computing*, 23(4):1290–1314, 2001.

[68] Y. Erlangga, C. Oosterlee, and C. Vuik. A novel multigrid based preconditioner for heterogeneous Helmholtz problems. *SIAM Journal on Scientific Computing*, 27(4):1471–1492, 2006.

[69] R. Falgout. An Introduction to Algebraic Multigrid. *Computing in Science and Engineering*, 8(6):24–33, 2006.

[70] C. Farhat, J. Li, and P. Avery. A FETI-DP method for the parallel iterative solution of indefinite and complex-valued solid and shell vibration problems. *International Journal for Numerical Methods in Engineering*, 63:398–427, 2005.

[71] C. Farhat, A. Macedo, M. Lesoinne, F.-X. Roux, F. Magoules, and A. de La Bourdonnie. Two-level domain decomposition methods with Lagrange multipliers for the fast iterative solution of acoustic scattering problems. *Compuer methods in applied mechanics and engineering*, 184:213–239, 2000.

[72] Y.T. Feng, D. Peric, and D.R.J. Owen. A non-nested Galerkin multi-grid method for solving linear and nonlinear solid mechanics problems. *Computer methods in applied mechanics and engineering*, 144:307–325, 1997.

[73] B. Fischer and R. Freund. Chebyshev Polynomials Are Not Always Optimal. *Journal of Approximation Theory*, 65:261–272, 1991.

[74] D.R. Fokkema, G.L.G. Sleijpen, and H.A. Van der Vorst. Jacobi-Davidson Style QR and QZ Algorithms for the Reduction of Matrix Pencils. *SIAM Journal of Scientific Computing*, 20(1):94–125, 1998.

[75] R.W. Freund and N.M. Nachtigal. QMRPACK: a package of QMR algorithms. *ACM Transactions of Mathematical Software*, 22(1):46–77, 1996.

[76] H. Fujita. Microactuators and Micromachines. *Proceedings of the IEEE*, 86(8):1721–32, 1998.

[77] S. Fukui and R. Kaneko. Analysis of ultra-thin gas lubrication based on linearized Boltzmann equation. *ASME Journal of Tribology*, 110:253–261, 1988.

[78] S.D. Gedney. An Anisotropic Perfectly Matched Layer-Absorbing Medium for the Truncation of FDTD Lattices. *IEEE Transactions on Antennas and Propagation*, 44(12):1630–1639, 1996.

[79] H. Goldstein, C. Poole, and J. Safko. *Classical Mechanics*. Addison Wesley, 2002.

[80] G.H. Golub and C.F. van Loan. *Matrix Computations*. John Hopkins, 1983.

[81] D. Gordon. Parallel ART for image reconstruction in CT using processor arrays. *The International Journal of Parallel, Emergent and Distributed Systems*, 21(5):365–380, 2006.

[82] D. Gordon and R. Gordon. Component-Averaged Row Projections: A Robust, Block-Parallel Scheme for Sparse Linear Systems. *SIAM Journal of Scientific Computing*, 27(3):1092–1117, 2005.

[83] J.P. Gorman. Finite-Element Model of Thermoelastic Damping in MEMS. Master's thesis, Massachusetts Institute of Technology, 2002.

[84] K.F. Graff. *Wave motions in elastic solids*. Dover, 1975.

[85] SUGAR group. SUGAR: A MEMS simulation tool. http://mems.sourceforge.net.

[86] I. Guy, S. Muensit, and E. Goldys. Extensional piezoelectric coefficients of Gallium Nitride and Aluminum Nitride. *Appliled Physics Letters*, 75(26):4133–4135, 1999.

[87] W. Hackbusch. *Multi-grid methods and applications*. Springer, 1985.

[88] L.N. Hand and J.D. Finch. *Analytical Mechanics*. Cambridge, 1998.

[89] Z. Hao and F. Ayazi. Support loss in the radial bulk-mode vibrations of center-supported micromechanical disk resonators. *Sensors and Actuator A:Physical*, 134:582–593, 2007.

[90] Z. Hao, A. Erbil, and F. Ayazi. An analytical model for support loss in micromachined beam resonators with in-plane flexural vibrations. *Sensors and Actuator A:Physical*, 109:156–164, 2003.

[91] I. Harari. A survey of finite element methods for time-harmonic acoustics. *Computer methods in applied mechanics and engineering*, 195:1594–1607, 2006.

[92] I. Harari and U. Albocher. Studies of FE/PML for exterior problems of time-harmonic elastic waves. *Computer Methods in Applied Mechanics and Engineering*, 195:3854–3879, 2006.

[93] I. Harari, M. Slavutin, and E. Turkel. Analytical and Numerical Studies of a Finite Element PML for the Helmholtz Equation. *Journal of Computational Acoustics*, 8(1):121–137, 2000.

[94] F.D. Hastings, J.B. Schneider, and S.L. Broschat. Application of the perfectly matched layer (PML) absorbing boundary condition to elastic wave propagation. *Journal of the Acoustical Society of America*, 100(5):3061–3069, 1996.

[95] E. Heikkola, T. Rossi, and J. Toivanen. Fast direct solution of the Helmholtz equation with a perfectly matched layer or an absorbing boundary condition. *International Journal for Numerical Methods in Engineering*, 57:2007–2025, 2003.

[96] Michael Heroux, Roscoe Bartlett, Vicki Howle Robert Hoekstra, Jonathan Hu, Tamara Kolda, Richard Lehoucq, Kevin Long, Roger Pawlowski, Eric Phipps, Andrew Salinger, Heidi Thornquist, Ray Tuminaro, James Willenbring, and Alan Williams. An Overview of Trilinos. Technical Report SAND2003-2927, Sandia National Laboratories, 2003.

[97] H. Hosaka, K. Itao, and S. Kuroda. Damping characteristics of beam-shaped micro-oscillators. *Sensors and Actuators A*, 49:87–95, 1995.

[98] W.-T. Hsu, J.R. Clark, and C.T.-C. Nguyen. Q-Optimized Lateral Free-Free Beam Micromechanical Resonators. In *Transducers'01*, pages 1110–1113, 2001.

[99] W. Huang, D.B. Bogy, and A. Garcia. Three-dimensional direct simulation Monte Carlo method for slider air bearings. *Physics of Fluids*, 9:1764–1769, 1997.

[100] L.-W. Hung, C.T.-C. Nguyen, Y. Xie, Y.-W. Lin, S.-S. Li, and Z. Ren. UHF Micromechanical Compound-(2,4) Mode Ring Resonators with Solid-Gap Transducers. In *Frequency Control Symposium, 2007 Joint with the 21st European Frequency and Time Forum. IEEE International*, pages 1370–1375, 2007.

[101] S. Hutcherson and W. Ye. On the squeeze-film damping of micro-resonators in the free-molecule regime. *Journal Micromechanics and Microengineering*, 14:1726–1733, 2004.

[102] C.-C. Hwang, R.-F. Fung, R.-.F. Yang, C.-I. Weng, and W.-L. Li. A New Modified Reynolds Equation for Ultrathin Film Gas Lubrication. *Transactions in Magnetics*, 32(2):344–347, 1996.

[103] R. Ierusalimschy, L. H. de Figueiredo, and W. Celes. Lua 5.0 Reference Manual. November 2003.

[104] Jr. J. Douglas, J.L. Hensley, and J.E. Roberts. An alternating-direction iteration method for Helmholtz problems. *Applied Mathematics*, 38:289–300, 1993.

[105] R. Johnson. *Mechanical Filters in Electronics*. Wiley, 1983.

[106] J.S. Juntunen, N.V. Kantartzis, and T.D. Tsiboukis. Zero Reflection Coefficient in Discretized PML. *IEEE Microwave and Wireless Components Letters*, 11(4):155–157, 2001.

[107] George Karypis and Vipin Kumar. METIS, A Software Package for Partitioning Unstructured Graphs, Partitioning Meshes, and Computing Fill-Reducing Orderings of Sparse Matrices, 1998. http://www.cs.umn.edu/ karypis.

[108] George Karypis, Kirk Schloegel, and Vipin Kumar. PARMETIS Parallel Graph Partitioning and Sparse Matrix Ordering Library, 2003. http://www.cs.umn.edu/ karypis.

[109] R. Kechroud, A. Soulaimani, and Y. Saad. Preconditioning techniques for the solution of the Helmholtz equation by the finite element method. *Mathematics and Computers in Simulation*, 65(4–5):303–321, 2004.

[110] S. Kim and S. Kim. Multigrid Simulation for High-Frequency Solutions of the Helmholtz Problem in Heterogeneous Media. *SIAM Journal of Scientific Computing*, 24(2):684–701, 2002.

[111] L.E. Kinsler, A.R. Frey, A.B. Coppens, and J.V. Sanders. *Fundamental of acoustics*. Wiley, 2000.

[112] C. Kittel. *Introduction to Solid State Physics*. Wiley, 1996.

[113] Attay Kovetz. *Electromagnetics*. Oxford Science Publications, 2000.

[114] T. Koyama, D.S. Bindel, W. He, E. Quevy, J.W. Demmel, S. Govindjee, and R.T. Howe. Simulation tools for damping in high frequency resonators. In *In 12th International Conference on Solid-State Sensors, Actuators, and Microsystems (Transducers 03)*, 2005.

[115] L. Xinjiao and X. Zechuan and H. Ziyou and C. Huazhe and S. Wuda and C. Zhongcai and Z. Feng and W. Enguang. On the properties of aln thin films grown by low temperature reactive r.f. sputtering. *Thin Solid Films*, 139:261–274, 1986.

[116] D. Lahaye, H. De Gersem, S. Vandewalle, and K. Hameyer. Algebraic Multigrid for Complex Symmetric Systems. *IEEE Transactions on Magnetics*, 36(4):1535–1538, 2000.

[117] K.M. Lakin. Modeling of Thin Film Resonators and Filters. In *IEEE MTT-S Digest*, pages 149–152, 1992.

[118] L.D. Landau and E.M. Lifshitz. *Theory of Elasticity.* Pergamon, 1986.

[119] L. Landua and G. Rumer. ——-. *Physik. Z. Sowjetunion*, 11:18, 1937.

[120] T.H. Lee. *The Design of CMOS Radio-Frequency Integrated Circuits.* Cambridge University Press, 2nd edition, 2004.

[121] S. Lepage and J.-C. Golinval. Finite Element Modeling of Thermoelastic Damping in Filleted Micro-Beams. In *Proceedings of EUROSIME 2007: THERMAL, MECHANICAL AND MULTI-PHYSICS SIMULATION AND EXPERIMENTS IN MICRO-ELECTRONICS AND MICRO-SYSTEMS*, pages 264–270, 2007.

[122] R. Lerch. Simulation of Piezoelectric Devices by Two- and Three- Dimensional Finite Elements. *IEEE Transactions of Ultrasonics, Ferroelectrics, and Frequency Control*, 37(2):233–247, 1990.

[123] M. Levinshtein, S. Rumyantsev, and M. Shur. *Properties of Advanced Semiconductor Materials.* John Wiley & Sons, 2001.

[124] R.-C. Li and Z. Bai. Structure-Preserving Model Reduction Using a Krylov Subspace Projection Formulation. *Communications in Mathematical Sciences*, 3(2):179–199, 2005.

[125] S.-S. Li, Y.-W. Lin, Z. Ren, and C.T.-C. Nguyen. Disk-Array Design for Suppression of

Unwanted Modes in Micromechancial Composite-Array Filters. In *Tech. Digest, MEMS06*, pages 866–869, 2006.

[126] S.-S. Li, Y.-W. Lin, Z. Ren, and C.T.-C. Nguyen. An MSI Micromechanical Differential Disk-Array Filter. In *Proceedings of the 14th International Conference on Solid-State Sensors and Actuators (Transducers'07)*, pages 307–311, 2007.

[127] W.-L. Li. Analytical modeling of ultra-thin gas squeeze film. *Nanotechnology*, 10:440–446, 1999.

[128] W.-L. Li and C.-I. Weng. Modified average Reynolds equation for ultra-thin film gas lubrication considering roughness orientations at arbitrary Knudsen numbers. *Wear*, 209:292–300, 1997.

[129] R. Lifshitz. Phonon-mediated dissipation in micro- and nano-mechanical systems. *Physica B*, 316-317:397–399, 2002.

[130] R. Lifshitz and M.L. Roukes. Thermoelastic damping in micro- and nanomechanical systems. *Physical Review B*, 61(8):5600–5609, 2000.

[131] Y.-W. Lin, S.-S. Li, Y. Xie, Z. Ren, and C.T.-C. Nguyen. Vibrating Micromechanical Resonators with Solid Dielectric Capacitive Transducer Gaps. In *Proceedings, Joint IEEE International Frequency Control/Precisioon time & time Interval Symposium*, pages 128–134, 2005.

[132] I. Livshits and A. Brandt. Accuracy Properties of the Wave-Ray Multigrid Algorithm for Helmholtz Equations. *SIAM Journal of Scientific Computing*, 28(4):1228–1251, 2006.

[133] T. Makkonen, A. Holappa, J. Ellä, and M.M. Salomaa. Finite Element Simulations of Thin-Film Composite BAW Resonators. *IEEE Transactions of Ultrasonics, Ferroelectrics, and Frequency Control*, 48(5):1241–1258, 2001.

[134] J. Mandel. Domain decomposition preconditioning for p-version finite elements with high aspect ratios. *Applied Numerical Mathematics*, 8:411–425, 1991.

[135] J. Mandel, M. Brezina, and P. Vanek. Energy Optimization of Algebraic Multigrid Bases. *Computing*, 62:205–228, 1999.

[136] T.A. Manteufel. The Tchebychev Iteration for Nonsymmetric Linear Systems. *Numerische Mathematik*, 28:307–327, 1977.

[137] T.A. Manteufel. Adaptive Procedure for Estimating Parameters for the Nonsymmetric Tchebychev Iteration. *Numerische Mathematik*, 31:183–208, 1978.

[138] J.E. Marsden and T.J.R. Hughes. *Mathematical Foundations of Elasticity.* Dover, 1983.

[139] L. McNeil, M. Grimsditch, and R. French. Vibrational Spectroscopy of Aluminum Nitride. *Journal of the Americaln Cermaic Society*, 76(5):1132–36, 1993.

[140] R.B. Morgan and M. Zeng. Harmonic projection methods for large non-symmetric eigenvalue problems. *Numerical Linear Algebra with Applications*, 5(1):33–55, 1998.

[141] MPI Standard. http://www-unix.mcs.anl.gov/mpi/.

[142] R. Mullen and T. Belytschko. Dispersion Analysis of Finite Element Semidiscretizations of the Two-Dimensional Wave Equation. *International Journal for Numerical Methods in Engineering*, 18:11–29, 1982.

[143] C. T.-C. Nguyen. Integrated micromechanical circuits fueled by vibrating RF MEMS technology. In *Proceedings, IEEE Int. Ultrasonics Symposium*, pages 953–962, 2006.

[144] C.T.-C. Nguyen. Transceiver front-end architectures using vibrating micromechanical signal processors. *Digest of Papers,Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems*, pages 23–32, 2001.

[145] C.T.-C. Nguyen. Vibrating RF MEMS for Next Generation Wireless Applications. In *Proceedings of the 2004 IEEE Custom Integrated Circuits Conference*, pages 257–264, 2004.

[146] C.T.-C. Nguyen. Integrating Micromechanical Circuits Fueled By Vibrating RF MEMS Technology. In *Proceedings of the 2004 IEEE International Ultrasonics Symposium*, pages 953–962, 2006.

[147] C.T.-C. Nguyen. MEMS technologies and devices for single-chip RF front-ends. In *2005 IMAPS/ACerS Int. Conf. on Ceramic Interconnect and Ceramic Microsystems Technologies (CICMT), Tech. Dig.*, 2006.

[148] A.S. Nowick and B.S. Berry. *Anelastic Relaxation in Crystalline Solids*. Academic Press, INC, 1972.

[149] C.C. Paige, B.N. Parlett, and H.A. van der Vorst. Approximate Solutions and Eigenvalue Bounds from Krylov Subspaces. *Numerical Linear Algebra with Applications*, 2(2):115–133, 1995.

[150] A.K. Pandley and R. Pratap. Coupled nonlinear effects of surface roughness and rarefaction on squeeze film damping in MEMS structures. *Journal of Micromechanics and Microengineering*, 14:1430–1437, 2004.

[151] Y.-H. Park and K.C. Park. High-Fidelity Modeling of MEMS Resonators Part I: Anchor Loss Mechanisms Through Substrate. *Journal of Microelectromechanical Systems*, 13(2):238–247, 2004.

[152] Y.-H. Park and K.C. Park. High-Fidelity Modeling of MEMS Resonators Part II: Coupled Beam-Substrate Dynamics and Validation. *Journal of Microelectromechanical Systems*, 13(2):248–257, 2004.

[153] B.N. Parlett and H.C. Chen. Use of Indefinite Pencils for Computing Damped Natural Modes. *Linear Algebra and its Applications*, 140:53–88, 1990.

[154] P.Chadwick and I.N. Sneddon. Plane Waves in an Elastic Solid Conducting Heat. *Journal of the Mechanics and Physics of Solids*, 6:223–230, 1958.

[155] G. Piazza, P.J. Stephanou, and A.P. Pisano. AlN Contour-Mode Vibrating RF MEMS for Next Generation Wireless Communications. In *Proceeding of the 36th European Solid-State Device Research Conference, 2006. ESSDERC 2006*, pages 61–64, 2006.

[156] G. Piazza, P.J. Stephanou, and A.P. Pisano. Piezoelectric Aluminum Nitride Vibrating Contour-Mode MEMS Resonators. *Journal of Microelectromechanical Systems*, 15(6):1406–1418, 2006.

[157] Sairam Prabhakar and Srikar Vengallatore. Thermoelastic damping in bilayered micromechanical beam resonators. *Journal of Micromechanics and Microengineering*, 17:532–538, 2007.

[158] S. Reitzinger, U. Schreiber, and U. van Rienen. Algebraic multigrid for complex symmetric matrices and applications. *Journal of Computational and Applied Mathematics*, 155:405–421, 2003.

[159] T.V. Roszhart. The Effect of Thermoelastic Internal Friction on the Q of Micromachined Silicon Resonators. In *IEEE Solid State Sensor and Actuator Workshop,HiltonHead*, pages 489–494, 1990.

[160] Y. Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, 2003.

[161] S. Senturia. *Microsystem Design*. Kluwer, 2001.

[162] S.D. Senturia and B.D. Wedlock. *Electronic Circuits and Applications*. Krieger, 1975.

[163] Y. Shapira. Multigrid methods for 3-D definite and indefinite problems. *Applied Numerical Mathematics*, 26:377–398, 1998.

[164] H.D. Simon. Analysis of the Symmetric Lanczos Algorithm with Reorthogonalization Methods. *Linear Algebra and its Applications*, 61:101–131, 1984.

[165] G.L.G. Sleijpen, A.G.L. Booten, D.R. Fokkema, and H.A. Van der Vorst. Jacobi-Davidson

Type Methods for Generalized Eigenproblems and Polynomial Eigenproblems. *BIT*, 36(3):595–633, 1996.

[166] G.L.G. Sleijpen and J. van den Eshof. On the use of harmonic Ritz pairs in approximating internal eigenpairs. *Linear Algebra and its Applications*, 358:115–137, 2003.

[167] G.L.G. Sleijpen and H. van der Vorst. A Jacobi-Davidson Iteration Method for Linear Eigenvalue problems. *SIAM Review*, 42(2):267–293, 2000.

[168] V.T. Srikar and S.D. Senturia. Thermoelastic Damping in Fine-Grained Polysilicon Flexural Beam Resonators. *Journal of Microelectromechanical Systems*, 11(5):499–504, 2002.

[169] P.J. Stephanou, G. Piazza, C.D. White, M.B.J. Wijesundara, and A.P. Pisano. Piezoelectric aluminum nitride MEMS annular dual contour mode filter. *Sensors and Actuators A*, 134:152–160, 2007.

[170] A. Bunse-Gerstnerand R. Stover. On a conjugate gradient-type method for solving complex symmetric linear systems. *Linear Algebra and Applications*, 287:105–123, 1999.

[171] T. Strouboulis, I. Babuska, and R. Hidajat. The generalized finite element method for Helmholtz equation: Theory, computation, and open problems. *Computer methods in applied mechanics and engineering*, 195:4711–4731, 2006.

[172] T. Strouboulis, R. Hidajat, and I. Babuska. The generalized finite element method for Helmholtz equation. Part II: Effect of choice handbook functions, error due to absorbing boundary conditions and its assesment. *Computer methods in applied mechanics and engineering*, 197:364–380, 2008.

[173] H. Takeuchi, E. Quévy, S.A. Bhave, T.-J. King, and R.T. Howe. Ge-Blade Damascene Process for Post-CMOS Integration of Nano-Mechanical Resonators. *IEEE Electron Device Letters*, 35(8):529–531, 2004.

[174] K. Tanabe. Projection Method for Solving a Singular System of Linear Equations and its Applications. *Numerishe Mathematik*, 17:203–214, 1971.

[175] Inc. The MathWorks. MATLAB. www.mathworks.com.

[176] L.L. Thompson. A review of finite-element methods for time-harmonic acoustics. *Journal of the Acoustical Society of America*, 119(3):1315–1330, 2006.

[177] H.A.C. Tilmans. Equivalent circuit representation of electromechanical transducers: I. Lumped-parameter systems. *Journal of Micromechanical Engineering*, 6:157–176, 1996.

[178] H.A.C. Tilmans. Equivalent circuit representation of electromechanical transducers: II. Distributed-parameter systems. *Journal of Micromechanical Engineering*, 7:285–309, 1997.

[179] S.P. Timoshenko and J.N. Goodier. *Theory of Elasticity*. McGraw-Hill, 1934.

[180] F. Tisseur and K. Meerbergen. The Quadratic Eigenvalue Problem. *SIAM REVIEW*, 43(2):235–286, 2001.

[181] W.S.N. Trimmer. Microrobots and Micromechanical Systems. *Sensors and Actuators*, 19:257–87, 1989.

[182] U. Trottenberg, C. Oosterlee, and A. Schueller. *Multigrid*. Academic Press, 2001.

[183] K. Tsubouchi, K. Sugai, and N. Mikoshiba. AlN Material Constants Evaluation and SAW Properties on AlN/Al_2O_3 and AlN/Si. In *Ultrasonics Symposium Proceedings*, 1981.

[184] H.A. van der Vorst and J.B.M. Melissen. A Petrov-Galerkin type method for solving Ax=b, where A is symmetric complex. *IEEE Transactions in Magnetics*, 26:706–708, 1990.

[185] P. Vanek, M. Brezina, and J. Mandel. Convergence of algebraic multigrid based on smoothed aggregation. *Numerische Mathematik*, 88(3):559–579, 2001.

[186] P. Vanek, J. Mandel, and M. Brezina. Two-level Algebraic Multigrid for the Helmholtz Problem. *Contemporary Mathematics*, 218:349–356, 1998.

[187] J. Wang, J.E. Butler, T. Feygelson, and C.T.-C. Nguyen. 1.51-GHz Nanocrystalline Diamond Micromechanical Disk Resonator with Material-Mismatched Isolating Support. In *Proc. IEEE MEMS'04*, pages 641–644, 2004.

[188] J. Wang, Z. Ren, and C.T.-C. Nguyen. Self-Aligned 1.14-GHz Vibrating Radial-Mode Disk Resonators. In *Transducers'03*, pages 947–950, 2003.

[189] J. Wang, Z. Ren, and C.T.-C. Nguyen. 1.156-GHz Self-Aligned Vibrating Micromechanical Disk Resonator. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 51(12):1607–1628, 2004.

[190] K. Wang, A.-C. Wong, and C.T.-C. Nguyen. VHF Free-Free Beam High-Q Micromechanical Resonators. *Journal of Microelectromechanical Systems*, 9(3):347–360, 2000.

[191] S.J. Wong, C.H.J. Fox, and S. McWilliam. Thermoelastic damping of the in-plane vibration of thin silicon rings. *Journal of Sound and Vibration*, 293:266–285, 2006.

[192] T.O. Woodruff and H. Ehrenreich. Absorption of Sound in Insulators. *Physical Review*, 123(5):1553–1559, 1961.

[193] A. Wright. Elastic preoperties of zinc-blende and wurtzite AlN, GaN, and InN. *Journal of Applied Physics*, 82(6):2833–2839, 1997.

[194] Y. Xie, S.-S. Li, Y.-W. Lin, Z. Ren, and C.T.-C. Nguyen. UHF Micromechanical Extensional Wine-Glass Mode Ring Resonators. In *Tech. Digest, 2003 IEEE International Electron Devices Meeting*, pages 953–956, 2003.

[195] J. Yang, T. Ono, and M. Esashi. Energy Dissipation in Submicrometer Thick Single-Crystal Silicon Cantilevers. *Journal of Microelectromechanical Systems*, 11(6):775–783, 2002.

[196] L.-Y. Yap, L.-K. Yap, and W. Ye. Air Damping in an Ultra-High-Frequency Disk Resonator. *Technical Proceedings of the 2003 Nanotechnology Conference and Tradeshow*, 1:316–319, 2003.

[197] K.Y. Yasumura, T.D. Stowe, E.M. Chow, T. Pfafman, T.W. Kenny, B.C. Stipe, and D. Rugat. Quality Factors in Micron- and Submicron-Thick Cantilevers. *Journal of Microelectromechanical Systems*, 9(1):117–125, 2000.

[198] W. Ye, X. Wang, W. Hemmert, D. Freeman, and J. White. Air damping in laterally oscillating microresonators: a numerical and experimental study. *Journal of Microelectromechanical Systems*, 12:557–566, 2003.

[199] Y.B. Yi. Geometric effects on thermoelastic damping in MEMS resonators. *Journal of Sound and Vibration*, 309:588–599, 2008.

[200] C. Zener. Internal Friction in Solids: I.Theory of Internal Friction in Reeds. *Physical Review*, 52:230–235, 1937.

[201] C. Zener. Internal Friction in Solids: II.General Theory of Thermoelastic Internal Friction. *Physical Review*, 53:90–99, 1938.

[202] C. Zener. Internal Friction in Solids: III.Experimental Demonstration of Thermoelastic Internal Friction. *Physical Review*, 53:100–101, 1938.

[203] C. Zener. Internal Friction in Solids: IV.Relation Between Cold Work and Internal Friction. *Physical Review*, 53:582–586, 1938.

[204] C. Zener. Intercrystalline Thermal Currents as a Source of Internal Friction. *Physical Review*, 56:343–349, 1939.

[205] C. Zhang, G. Xu, and Q. Jiang. Characterization of the squeeze film damping effect on the quality factor of a microbeam resonator. *Journal Micromechanics and Microengineering*, 14:1302–1306, 2004.

[206] O.C. Zienkiewicz and R.L. Taylor. *The Finite Element Method:Volume 1 The Basis*. Butterworth-Heinemann, 2000.