Flexure-Based Nanopositioning Systems: Integrated Design and Control

by

Vijay Shilpiekandula

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2010

Author ....
Department of Mechanical Engineering
September 10, 2009

Certified by...
Kamal Youcef-Toumi
Professor of Mechanical Engineering
Thesis Supervisor

Accepted by ....
David E. Hardt
Chairman, Department Committee on Graduate Students

Flexure-Based Nanopositioning Systems: Integrated Design and Control

by

Vijay Shilpiekandula

Submitted to the Department of Mechanical Engineering
on September 10, 2009, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Mechanical Engineering

Abstract

This thesis deals with the design and control of flexure-based mechanisms for applications requiring multi-degree-of-freedom positioning and alignment. Example applications include positioning a probe or sample in atomic force microscopy, alignment of tool and sample in stamping processes, and fine-positioning of wafer steppers in semiconductor manufacturing. Such applications necessitate nanopositioning systems that satisfy critical functional requirements, such as load-capacity, bandwidth, resolution, and range. Therefore, a systematic approach for design and control is an important tool for research and development for flexure-based nanopositioning systems.

In this thesis, a novel methodology is presented for generating flexure-based topologies that can meet performance requirements, such as those dictating structural strength or dynamical behavior. We present performance metrics that allow for the generation of topologies that are tuned for a desired level of structural strength or modal separation. The topology generation is aimed as a valuable addition to the design toolkit, facilitating novel designs that could not have been conceived otherwise. The parameters within any particular topology could be adjusted at a subsequent phase through a detailed shape and size optimization.

The thesis also proposes a controller generation approach. Unlike existing controller parameterizations, a novel parameterization formulated in this thesis allows for directly tuning the sensitivity transfer function of the closed-loop system. Tuning sensitivity is critical in mitigating the effects of disturbances affecting the system, as well as those arising from cross-coupling and parasitic error motions.

Further, an integrated methodology for design and control is presented. This methodology uses the design topology generation approach and controller generation approach proposed in the thesis. The key distinction of our design for control approach is that the design is iterated over topologies and not just parameters within a selected topology. A simple one-degree-of-freedom positioning system example is worked out to detail the steps of the proposed integrated design and control methodology. A novel design topology that is ideally suited for achieving a desired design and control performance is derived using this methodology.

Finally, the hardware design and control of a novel flexure-based nanopositioner implementation for scanning probe microscopy are presented to illustrate the effectiveness of the approaches discussed in this thesis.

Thesis Supervisor: Kamal Youcef-Toumi
Title: Professor of Mechanical Engineering

Acknowledgements

ఎందరో మహానుభావులు, అందరికి వందనములు
[Transliteration: *Endaro mahaanubhaavulu Andariki vandanamulu*
Translation: There are many experienced noble souls, Salutations to all]
— Sri Raaga Pancharatna Krithi, St. Thyagaraja, 18th Century, India

Like in long-distance running, there comes a moment of epiphany in writing a Ph.D. thesis. With the finish line in sight, while crossing the t's and dotting the i's for the print-ready version, you know you are there, but not *quite* yet — until you pause and reflect on the journey that preceded, until your mind jogs down the memory lane, thinking of the many experienced noble souls that held your hand in your baby steps, taught you how to walk long before you even dreamt of running.

First on the list is Amma. As a primary school teacher in Vijayawada, India, my mother worked day and night throughout her career, just so that my brother and I could get a decent education. In seemingly-troubling times, if I could put things in perspective it is because of a simple example of never-give-up never-complain attitude that she has always set for us. Pillars of strength for Amma are my aunt (Sriviri Peddamma) and uncle (Peddanaannagaru) who have been with us at every hour, in joy and sorrow, weakness and strength. Peddamma is one *peddabaalasikhsa*, a veritable encyclopedia of values in life, and I always seek her advice and guidance. Vijji Mama, my mother's brother, has always encouraged us to pursue the studies of our choice. Mama supported me throughout my education and for that I am ever thankful to him.

What can I say about my brother Vikas? If there is a rock, solid and strong, that's him. Sangeeta, my sister-in-law, is a person more balanced than I have ever known of anyone. If there's someone that surpasses the attention I give to this couple in my life, it is my 15-month-old niece Hema. Kaaka kuThe is what she asks in Marathi, something she had done a zillion times before when I had gone underground writing this thesis, and now as I engage her in another game of peekaboo. I can only wait to see the wonders of life ahead of us with Hema taking the lead.

As Peddamma says, gratitude is the language of the Gods. In my offerings of thanks to the pantheon of my personal deities, I must first mention my Guru, Professor Kamal Youcef-Toumi. Kamal has been a father-like figure, showing me the right from wrong in life as much as in mechatronics, control theory, and hardware design. Our discussions have spanned topics as diverse as Kharitonov theorem in robust control, or predicting non-minimum phase zeros in mechanical systems, to the healing power of timed breathing exercises, and the correct way to breathe while, say, singing or swimming. Kamal's work ethic, his emphasis on clinical and crisp writing, and an even-keeled approach to scientific research will remain legacies that I have internalized over my graduate student life.

An ensemble of faculty, student colleagues, and staff were part of my thesis journey. Professors Harry Asada and Martin L. Culpepper served on my committee, each giving me valuable advice, be it about avoiding roadblocks or meeting deadlines.

Professor Asada has been a pioneer in integrated design and control for high speed robotic manipulators for over three decades. His advice and feedback were highly useful in shaping the integrated design and control methodology proposed in this thesis. Professor Culpepper is an expert on flexures, who helped us formulate the topology generation problem. I am often amazed by the depth and breadth of his knowledge of current research and researchers in precision engineering. I am thankful to both these faculty members for their interest in the progress of the thesis, as reflected from their focused one-on-one meetings, prompt replies to emails, and an open walk-in accessibility in spite of their busy schedules.

Dan, Georg, and Kwang were my colleagues on the AFM nanopositioner project. They are the AFM Gurus I have had the pleasure of working with, learning from, and discussing various design and control issues, problems, and strategies. Group meetings on this project were action-packed with in-depth discussions and debates, and I will miss them. Other colleagues in the Mechatronics Research Laboratory that added to the light-hearted humor (come on, we were working in a basement glass lab!) include Osamah, Khalid, Pablo, Mauricio, Dr. Xu, Adam, Anirban, Dimitris, Heejin, and Sang-il. I am also thankful to the opportunity of recruiting, mentoring, defining structured research projects for many undergraduate research assistants (UROPs) and M. Engg. graduate students. These include Justin, Samarth, Anupam, and Paul, who all offered a helping hand one time or the other at the hobby shop, Edgerton student shop, and the Graduate student shop in our lab projects.

A few of my best friends who stood by me through thick and thin. Ajay Deshpande is more like an elder brother than a friend, a friend than a colleague, a colleague than a quals-buddy. I cherish the coffee breaks of many afternoons of graduate school, discussing dynamic systems, new research trends, and life in general, with him. Ankur Sinha has this infectious optimistic approach to life that makes you forget the hard constraints and limitations and focus more on the givens, the taken-for-granteds that you otherwise miss.

And then there are some friends you don't meet often. But then you bond more than the space-time continuum can permit. Among them are my first-year roommate of IITM and MIT - the evercool Harsh, an Italian maestro Meher, and the very Techpassion-ate Kodi. And then there are those that bore the brunt of my humor more often than most would have preferred. Here's to the list that never ends: Navakanth, Pavithra, Prahladh, Kripa, Manasa, Rajappa, Sriram, Sreekar, Pradeep, Karthik Dantu, Kavitha, Sangeetha, Vikram, Sivaram, Hari, Madhusudhan, Lavanya, Mythili, Jaykumar, Aditi, Bhaskaran, Shashi, Nikhil, Sumeet, Abishek, Isha, Ajay Somani, Vivek S, Vivek R, Vivek J, Sudhish — To them that make a stop and chat, a simple skype call, an email exchange, or a drop-in at their nook and corner in the world worthwhile, here's a shout out, a big thank you!

Any note of thanks is incomplete without referring to some of the administrators that made students' lives simple and easy. Trust Lois Graziano for going the extra mile in making an after-hour purchase on McMaster, or making all necessary arrangements for a conference registration or a journal article payment. Include my post-defense celebration that she had diligently planned and executed so well! Leslie Regan is our Graduate Administrator, who oversees most of our graduate requirements and guidelines. There isn't a time when I had been to Leslie's office and returned without her having called someone of dispatched an email in real-time so that a problem at hand got instantly resolved. Thank you Leslie for all the help and advice you have given us for navigating through the maze of MIT graduate school!

The hardware prototypes detailed in the thesis were built at the MIT Hobby Shop, Edgerton Student Shop, and the ME Department Graduate Student Shop. Ken Stone and Hayami Arakawa trained me on the OMAX abrasive waterjet machine and the Prototrak CNC, and their endless patience and diligence in getting me on speed with the hardware prototyping is highly appreciated. Mark Belanger of the Edgerton Student Shop is the Guru of design and machining, and I miss all the fun hours working with him.

Many funding sources provided the monetary support backing this thesis. I am particularly thankful to the MIT Dean School of Engineering for the Xerox Foundation fellowship that have me the creative freedom of exploring my choicest research, without the hassle of catering to any sponsor stipulations. The American Academy of Mechanics and the Robert and Mary M. Haythornthwaite Foundation provided a seed grant that triggered most of our research on flexures. The Manufacturing Science and Technology division of the Singapore MIT Alliance Flagship Research Program funded the micro and nanomanufacturing parts of our research, which provided the background for our design applications. The AFM nanopositioner project was sponsored by the Mechatronics Researcg and Development Division of Samsung Electronics Co. Ltd. The timely feedback and reviews from Dr. Yongshik Douglas Kim's team at Samsung, represented at MIT by Dr. Yonmook Park and Dr. Dongmin Kim, kept us aligned to the latest trends and needs in industrial metrology.

A special mention must be made about the helpful feedback that I have received from faculty and industry experts at research conferences. The American Society for Precision Engineering supported me with the R.V. Jones Memorial Scholarship based on our paper that won the best student paper award at their 2008 Annual Conference in Portland, OR. It is here that I had the good fortune of discussing design and controls aspects of my thesis with Prof. Shorya Awtar of University of Michigan, Ann Arbor, Dr. Layton Hale of Lawrence Livermore National Laboratories, Dr. Pradeep Subrahmanyan of RAPT Industries, Inc., and Dr. Douglas Watson, Nikon Research. I would also like to thank Prof. Jayaraman Raja of UNC Charlotte, Prof. Thomas Kurfess of Clemson University, and Dr. Ananda Mysore of Seagate Technologies, Inc, and Dr. Shawn Moylan of NIST for their feedback and suggestions based on my talks at the Society for Manufacturing Engineers Micro and Nanomanufacturing, Boston, MA.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Precision positioning and alignment at nanoscale resolutions is often referred to as "nanopositioning." Many design and control groups in the academia and industry have explored this area of research over the past few decades in various contexts, such as semiconductor manufacturing, biological imaging, optical communications, and others [2]-[10]. From among the many different methods of implementing nanopositioning systems, those involving compliant flexure-based mechanisms have gained popularity over the years [14, 48]. Flexure-based mechanisms are slender beam-like spring elements in mechanical designs; they are close to being ideal motion bearings with minimal friction, backlash, and other uncertainties [16]. These advantages make flexure-based designs ideal candidates to implement for precision motion control.

We are interested in the design and control of flexure-based mechanisms for nanopositioning applications. As will be explained in this thesis, existing designs do not address critical functional requirements, such as load-capacity, bandwidth, and range needed for a niche class of nanopositioning applications. Further, to the best of our knowledge, an integrated design and control approach for flexure-based nanopositioning systems is lacking in the current literature. A systems-based methodology can facilitate developing valuable synthesis tools for achieving the desired closed-loop control performance.

1.1 Motivation

Nanopositioning is critical to an emerging class of small-scale precision manufacturing/metrology and numerous motion control applications. The drive for better performance steers high-resolution designs towards satisfying stringent specifications in terms of functional parameters such as range, load-capacity, and bandwidth. Examples of applications needing nanopositioning include (i) high-bandwidth steering of mirrors in optomechanical system applications requiring precision positioning and tracking [1], (ii) sample or probe positioning in scanning probe microscopy methods such as atomic force microscopy (AFM) for measuring topography at nano-scale resolution (iii) tool-sample alignment in stamping applications such as imprint lithography [2, 3] and micro-contact printing [4], and (iv) alignment of optically flat surfaces brought in close proximity to characterize fields and forces on small-scales, such as the Casimir force [5]-[8], or facilitate macromolecular separations [10].

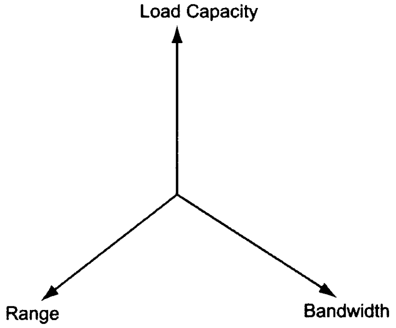


Figure 1-1: Design triad motivating the broad classes of applications. Designs developed primarily for achieving large load-capacity are different from those developed for high bandwidths, or those developed for large-range.

In terms of structural performance, the selection of flexures is dictated by many requirements, examples being range, bandwidth, and load-capacity. A broad classification of high-resolution nanopositioning systems is shown as a design triad in

Fig. 1-1. In this thesis we are interested in emerging nanotechnology applications that all fit within this classification framework.

1.2 Technical Challenges and Issues

While flexure-based nanopositioning systems for such advanced nanotechnology applications have been around for the past few decades [17], designing them for dynamic performance has received little attention. Kinematic arrangement of parallel flexure systems using projection geometry theory has been worked out in [18]. Analysis of the statics [16] and dynamics of flexure-based mechanisms have been extensively studied [21, 49]. However, few publications [22, 23] have appeared in the context of design for dynamic performance. The design of flexures in the context of mechanical advantage is detailed in [22]. A finite-element approach based on Euler-Bernoulli beam bending theory is formulated for analyzing dynamics in [23] and optimizing the design space for precision flexure-based applications in [24].

While dynamic performance of just the flexure-based mechanism or 'plant,' as referred to in the field of control systems, presents one of the performance requirements, a more challenging and critical requirement is achieving an overall desired closed-loop control performance [25] of a system assembled with the mechanism, and suitable actuator and sensor subsystems. Optimal locations for embedding actuators and sensors in a mechanism with distributed compliance to satisfy controllability and observability conditions are discussed in [26]. However, neither the design of the controller nor the influence of a poor design choice on control performance is addressed in this reference. A potentially useful approach in this context should be based on integrating design and control methods right from design conception and validation phase before hardware implementations are tested out. Integrated design and control has been an active area of research spanning applications such as robotic manipulator design and control [27]-[33], motion stages developed using lead-screw drives [35], passive and active vibration isolation platforms [36, 37], and chemical process control [38]. To the best of our knowledge, an integrated approach for the design and control of flexure-based nanopositioning systems is lacking in the existing literature. A common systems-based methodology can facilitate developing valuable synthesis tools for achieving the desired closed-loop control performance.

A related issue is one of lightly damped flexible modes of flexure-based mechanisms. Physical damping is low in flexures made from metals such as aluminium (used in development stages of the design process for ease of machining), or titanium (used in the implementation and testing phase because of its high fatigue strength and other material properties). External damping such as squeeze film damping and foam-damping have been suggested and explored for flexures in the past. Active damping through appropriate selection of control strategies needs to be addressed to tackle the lightly damped resonances in these structures. Since the level of damping in an assembled mechanism is hard to predict before the fabricated product is available for testing, it becomes necessary to iterate the design process with thorough system identification and testing of hardware mechanism implementations.

1.3 Approach

In this thesis, we base our design for control approach on two main contributions. The first is the generation of a design topology library. This step is critical since, before we implement any optimization of a selected design in terms of its shape and size, material choice and other details, we need to be sure we are starting with an appropriate, if not the most optimal, topology to begin with. Once the topology is selected based on an initial screening (such as meeting design specifications, or coupling between motion axes for control, or a desired level of modal separation), the shape and size optimization routines can be deployed to fine-tune the parameters of the selected topology.

The second contribution of the thesis is a method for controller generation. Instead of selecting a brute-force method for generating controllers, first a nominal or baseline is established. The closed-loop sensitivity transfer function obtained with the nominal controller may not have the desired degree of robustness or disturbance rejection capability. In this regard, we propose tuning the sensitivity transfer function of the system using a free parameter transfer function $Q(s)$. Under the novel controller parameterization, synthesis of $Q(s)$ is shown to be one of model-matching, or a linear filter design. This approach is demonstrated via examples and a hardware case study. The suggested approach allowed us to tie in theory with an applied setting in the context of hardware implementations.

1.4 Case Study: AFM Nanopositioner

Many performance requirements dictate the selection of flexures, examples being range, bandwidth, and load-capacity. In this case study, we implement a design for closed-loop control approach for flexure-based mechanisms in nanopositioning applications such as scanning probe microscopy. Here, we regard the design for the control approach as being primarily driven by closed-loop system requirements such as achievable maximum range of motion, resolution and repeatability of positioning, and mitigating the effects of coupling or cross-talk between the axes. In dimensional metrology with application to semiconductor process characterization, high-speed atomic force microscopy can push the state-of-the-art in fast and reliable measurements at a nanoscale resolution.

In particular, we developed a novel scanner for probe positioning in atomic force microscopes. Unlike existing XY scanners, this scanner is designed specifically to improve the dynamic performance in the Z direction. We have built a series of scanner prototypes driven with piezoelectric actuators, and whose resulting motion is sensed by laser interferometers and capacitance probes. We have tested the prototypes for their static range and frequency response functions, both in-plane and out-of-plane. Design trade-offs in achieving competing requirements are minimized with judicious flexure component choices that further improve the potential control performance.

This case study is an application of the design for control approach proposed in the thesis. The final design we converged on fits within a library of design topologies that were generated using the principles derived herein. A hardware photograph of the nanopositioner is presented in Figure 6-15 below.

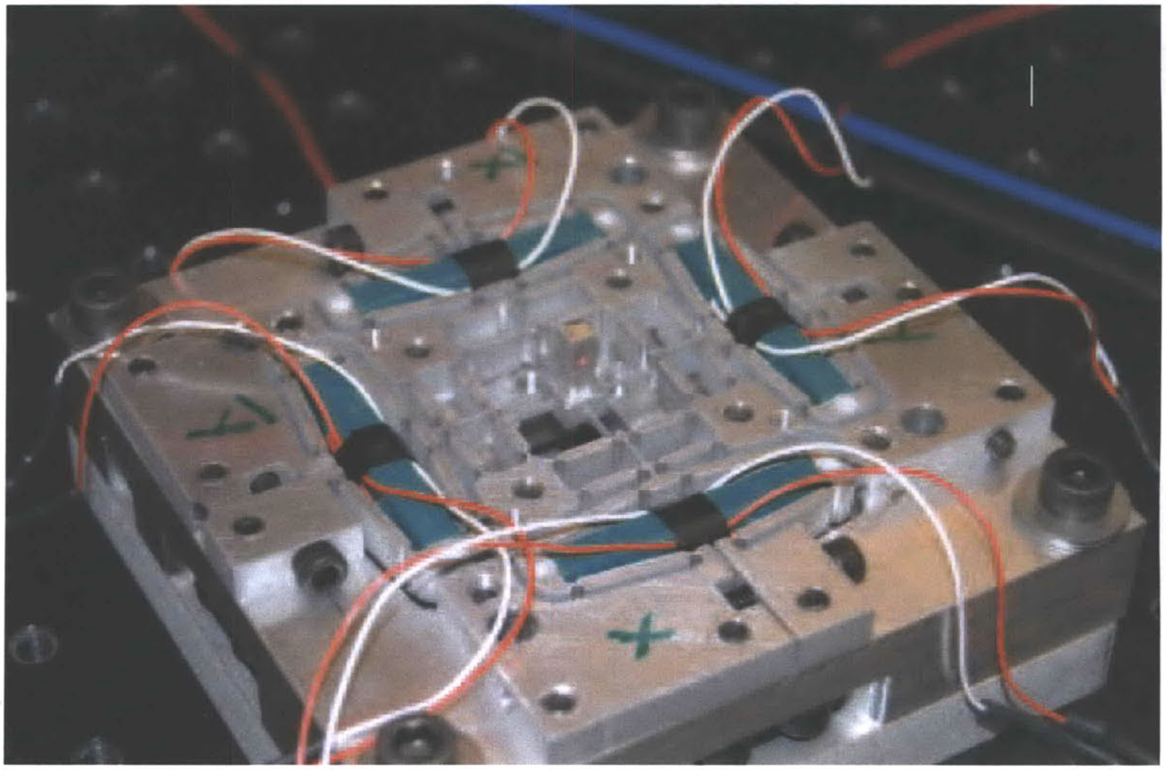


Figure 1-2: Hardware photograph of nanopositioner built as a case study for design for control approach developed in the thesis. The footprint of the device is about 4.5 *in* × 4.5 *in*.

The key results are tabulated in Table 1.1. In comparison with prior art, this nanopositioner allows for triangular wave scans, within a field of view about four

Table 1.1: Experimentally demonstrated results obtained with the novel AFM nanopositioner.

Measured Parameter	Value
Range of deflection	$50\,\mu m \times 50\,\mu m$
Lateral Resonance	$750\ Hz$ (first peak)
Static coupling between lateral axes	$< 44\ dB$ from Y to X $< 30\ dB$ from X to Y
Vertical Resonance	$33.1\ kHz$ (first peak)

times larger than existing art. Unlike resonant mechanisms, our nanopositioner allows for nearly uniform velocity scans over the large field of view. The nanopositioner is statically decoupled, with a measured difference of $44dB$ from Y to X and 30 dB from X to Y in the DC gains of the cross coupling frequency response functions when compared to that of their direct frequency response counterparts. The vertical resonance of as high as 33 kHz allows for significantly large modal separation between the axes, making it suitable for high-speed imaging. In parallel to this nanopositioner design and control, an optical subsystem is being developed in our research group. With the add-on of the optical subsystem the nanopositioner, we expect to achieve high-speed imaging of micron-sized samples, such as silicon features in semiconductor manufacturing. Other applications of this AFM nanopositioner include imaging of biological samples to detect nanoscale topography and understand the associated science.

1.5 Thesis Roadmap

The rest of this thesis is organized as follows. Chapter 2 presents an overview of the dynamic modeling of flexure-based mechanisms. In Chapter 2.2, we formulate performance requirements such as range, resolution, bandwidth, and load-capacity for flexure-based nanopositioning systems. Chapter 3 discusses the generation of design topologies that meet performance requirements such as structural strength and dynamic performance. In chapter 4 we propose a novel controller parameterization, which allows for tuning the sensitivity transfer function of the closed-loop system. In chapter 5, we present a methodology for integrating design and control, and work out in detail an example positioning system that can meet a set of desired design and control performance requirements. The case study of a novel nanopositioner for atomic force microscopy is detailed in Chapter 6. A summary of thesis contributions, possible ideas for future directions of this research are presented in Chapter 7.

Chapter 2

Preliminaries

In this chapter, we review the dynamic modeling of flexures and formulation of performance specifications for positioning systems using the flexure-based mechanisms [1]. Simple lumped parameter models for flexure beams and assemblies of such beams are derived for their mass and stiffness matrices. Flexure materials such as aluminium, steel, and other metals, are typically characterized by low damping. Simple proportional models for damping are mentioned, understanding that predicting damping accurately is infeasible. In Chapter 4, we discuss how to tackle lightly damped resonances through tuning sensitivity transfer functions. For formulating performance specifications, we present a state-space formulation of a sample set of specifications commonly required of positioning systems. The parameters such as range, resolution, bandwidth and load-capacity are derived for flexure-based positioning systems. Any of these specifications can be included in the cost or constraint functions in any optimization formulation.

---

[1] Parts of this chapter have been reprinted from our American Control Conference 2008 article cited as reference [40]. The following copyright/credit notice is presented as stipulated by IEEE. ©2008 IEEE. Reprinted, with permission, from Proceedings of American Control Conference, June 2008 paper "Characterization of Dynamic Behavior of Flexure-based Mechanisms for Precision Angular Alignment," by Shilpiekandula V, Youcef-Toumi K, cited as reference [40].

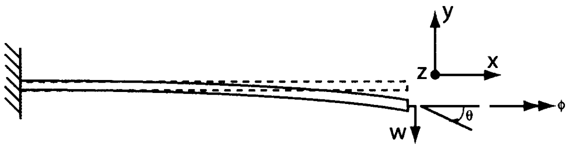


Figure 2-1: Schematic diagram showing a flexural beam element with deflection $w(x,t)$, slope $\theta(x,t)$, and twist about the X-axis by an angle $\phi(x,t)$ [40].

2.1 Dynamic Modeling

Flexures are beam-based motion guidance elements. Unlike friction-based bearings, they allow for achieving smooth elastic motion with minimal nonlinearities such as hysteresis or backlash. A simple flexure building block, or primitive, is shown as a beam in Fig. 2-1. When a vertical force is applied at its free end, the beam is subject to a bending action, that arises from the compliance distributed long its entire length. In implementing positioning systems with flexures, it is critical to capture an accurate dynamic model for developing a controller that satisfies properties like good command following, or rejection of disturbances. In what follows, we detail the dynamic modeling of the flexure beam of Fig. 2-1.[2]

2.1.1 Governing Equations

For the modeling, both Timoshenko beam [52] and Euler-Bernoulli beam models are widely used. A Euler-Bernoulli beam model cannot capture the effects of shear and rotational inertia. These effects become significant for beams with thickness comparable to their length. To model torsion of the flexural element, we use St. Venant's torsion formulation assuming that (i) the effects of restrained warping are negligible and (ii) bending and torsion are decoupled. Further, we assume that the deflections of the flexural beam element are small (an order of magnitude smaller than the beam thickness) and hence, neglect the effects of axial stretching and the

---

[2] Dynamic modeling of other flexure primitives can be derived on similar lines. Refer to [21] for modeling of notch flexures and [36] for modeling of post-buckled springs.

resultant stress stiffening along the length of the beam element.

Under the above-mentioned assumptions, the distributed parameter model for the beam is well-documented in the literature [52] as being depicted by a set of partial differential equations in the deflection $w(x,t)$, slope $\theta(x,t)$, and angle of twist $\phi(x,t)$ as follows [60, 56]:

$$\begin{aligned} EI_{yy}\frac{\partial^2\theta(x,t)}{\partial x^2} + \kappa AG\left\{\frac{\partial w(x,t)}{\partial x} - \theta(x,t)\right\} & \\ -\rho I_{yy}\frac{\partial^2\theta(x,t)}{\partial t^2} &= 0 \qquad (2.1) \\ \rho A\frac{\partial^2 w(x,t)}{\partial x^2} - \kappa AG\left\{\frac{\partial^2 w(x,t)}{\partial x^2} - \frac{\partial\theta(x,t)}{\partial x}\right\} &= 0 \qquad (2.2) \\ GJ_{xx}\frac{\partial^2\phi(x,t)}{\partial x^2} + I_{xx}\frac{\partial^2\phi(x,t)}{\partial t^2} &= 0 \qquad (2.3) \end{aligned}$$

where ρ, E, G, J_{xx} are the density, elastic modulus, and torsional rigidity, respectively; A, I_{yy}, and κ are the cross-sectional area, area moment of inertia about the neutral axis Y, and a geometry-dependent shear-factor, respectively. For a rectangular cross-section κ assumes a value of 0.833 [63].

2.1.2 Lumped parameter modeling

Lumped parameter models approximate the infinite-dimensional behavior governed by the set of partial differential equations. Such models often provide insight into parametric relations that a direct finite element analysis routine may not provide. In the initial design phase, lumped parameter models are often pivotal in making design decisions, either about a layout of flexure constraints, or interconnection with actuators. In this section, we detail the procedure for obtaining lumped parameter models for flexures as well as assemblies of rigid bodies and flexures.

Many approximation methods such as the method of assumed modes [53], a Galerkin approximation or Rayleigh approximation [62] are used for generating the lumped parameter models. For example, using an assumed modes method, the infinite-dimensional behavior of the mechanism is approximated to a finite-series made up of spatially varying mode shape functions (or trial functions) with temporally varying mode amplitudes [55]. Since a one-element model is used for the beam, the distributed properties of the beam are lumped to the node at the guided end of the beam; the fixed node of the beam has no lumped mass or stiffness. Hence, from nodal displacements assumed for the guided end of the beam, a multi-DOF lumped parameter model can be derived.[3] The detailed application of the assumed modes method to the set of partial differential equations governing the motion of the beam can be found in parts in [56] and [60].

To illustrate the lumped parameter modeling, let us consider the dynamic modeling of diaphragm flexures used in precision angle alignment mechanisms [40], such as the one shown in Fig. 2-2. A diaphragm flexure is a mechanism that suspends the payload on a radial or tangential arrangement of flexural beams. The flexure shown in Fig. 2-2 is a parallel kinematic mechanism with a central rigid circular disk centered at the origin and parallel to the horizontal **XY** plane of the cartesian **XYZ** space. In the rest position, the principal axes of the rigid body, i.e. the disk, namely, **X'**, **Y'**, and **Z'** coincide with the cartesian axes **X**, **Y**, and **Z**, respectively. Let the disk be of radius R, thickness T, mass M, and moments of inertia J_{xx} and J_{yy} about the X and Y axes respectively. A number n of slender beam flexures, each of width W, thickness H, and length ℓ, are in the **XY** plane connecting every peripheral point A_i to the ground. (Fig.!2-2 shows $n = 3$ beams connecting the points A_1, A_2, and A_3 on the disk to the ground.) The coordinates of P_i in the **X'Y'** plane are $(Rcos\alpha_i, Rsin\alpha_i)$ with angles $\alpha_i \in [0, 2\pi)$ for $i = 1, 2, 3... n$.

Since the beams provide high axial (and hence in-plane **XY**) stiffness and low out-of-plane stiffness, we expect that the dominant modes correspond to the out-

---

[3] A many-element model can be used for each flexural beam to develop higher-order lumped parameter models, if desired.

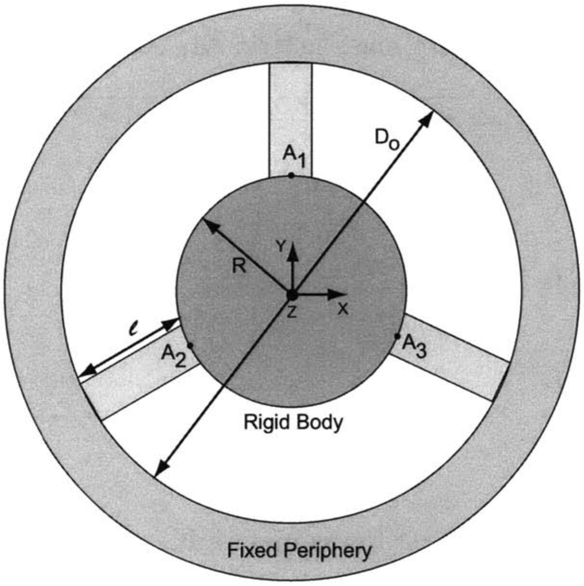


Figure 2-2: A diaphragm flexure as a parallel kinematic mechanism with a central rigid mass connected by $n = 3$ flexural beam units to the ground. The dimension $D_0 = 2\ell + 2R$ is referred in this paper as the footprint of the mechanism. The **Z** axis is shown pointing out of the page [40] .

of-plane motion, namely vertical deflection, pitch, and roll. We hence assume that the out-of-plane motion of the disk is decoupled from the in-plane motion, i.e. the center of the disk always moves only vertically. For small vertical deflection $z(t)$ of the center of mass, and small angular rotations $\theta_x(t)$ and $\theta_y(t)$ about the **X** and **Y** axes respectively, the principal plane **X'Y'** of the disk moves out of the **XY** plane to the one depicted by

$$\mathbf{Z}_p(t) = \theta_y(t)\mathbf{X} + \theta_x(t)\mathbf{Y} + z(t) \tag{2.4}$$

Under the geometric boundary conditions of (i) one end $x = 0$ of the flexural beam being grounded and (ii) the other end $x = \ell$ subject to generalized displacements $\mathbf{V}(t) = [w(\ell,t)\ \theta(\ell,t)\ \phi(\ell,t)]^T$, (where $w(\ell,t)$ is the vertical deflection, $\theta(\ell,t)$ the slope, and $\phi(\ell,t)$ is the angle of twist), the mass $\mathbf{M}_f$ and stiffness $\mathbf{K}_f$ matrices (which represent the kinetic energy, and elastic potential energy storage, respectively,) are

Table 2.1: Flexure mass matrix $\mathbf{M_f}$ component values

$$\boxed{\begin{aligned}
M_1 &= \alpha m_f[\tfrac{13}{35} + \tfrac{7}{5}(\tfrac{c}{\ell})^2 + \tfrac{4}{3}(\tfrac{c}{\ell})^4] + \tfrac{6}{5}\tfrac{\rho\alpha I_{yy}}{\ell} \\
M_2 &= -\alpha m_f \ell[\tfrac{11}{210} + \tfrac{11}{60}(\tfrac{c}{\ell})^2 + \tfrac{1}{6}(\tfrac{c}{\ell})^4] \\
&\quad +\rho\alpha I_{yy}[\tfrac{1}{10} + (\tfrac{c}{\ell})^2] \\
M_3 &= \alpha m_f \ell^2[\tfrac{1}{105} + \tfrac{1}{30}\{(\tfrac{c}{\ell})^2 + (\tfrac{c}{\ell})^4\}] \\
&\quad +\rho\alpha I_{yy}\ell[\tfrac{2}{15} + \tfrac{1}{3}(\tfrac{c}{\ell})^2 + \tfrac{4}{3}(\tfrac{c}{\ell})^4] \\
M_4 &= \tfrac{I_{xx}}{3}
\end{aligned}}$$

Table 2.2: Flexure stiffness matrix $\mathbf{K_f}$ component values

$$\boxed{\begin{aligned}
K_1 &= \alpha\tfrac{12EI}{\ell^3} + 144\alpha\beta \\
K_2 &= -\alpha\tfrac{6EI}{\ell^2} - 72\alpha\beta \\
K_3 &= \alpha\tfrac{4EI}{\ell}[1 + (\tfrac{c}{l})^2 + (\tfrac{c}{l})^4] + 36\alpha\beta \\
K_4 &= \tfrac{GJ_{xx}}{\ell}
\end{aligned}}$$

as given below:

$$\mathbf{M}_f = \begin{bmatrix} M_1 & M_2 & 0 \\ M_2 & M_3 & 0 \\ 0 & 0 & M_4 \end{bmatrix} ; \ \mathbf{K}_f = \begin{bmatrix} K_1 & K_2 & 0 \\ K_2 & K_3 & 0 \\ 0 & 0 & K_4 \end{bmatrix}$$

where the matrix values depend on material properties and geometry, and are listed in Tables 2.1 and 2.2. The parameters used in the tables are $\alpha{=}\left\{\frac{1}{1+2p}\right\}^2$ and $\beta = \frac{EI_{yy}}{6}\frac{c^2}{l^5}$, where, $p = (\frac{c}{\ell})^2$, and $c = \sqrt{\frac{6EI_{yy}}{\kappa AG}}$ representing the length scale at which effects of shear dominate. For a flexural beam with rectangular cross-section of height H, c reduces to $\sqrt{1.1(1+\nu)}H$, where ν is the poisson's ratio of the material.

Zeroes in either matrix result from the decoupling assumed between bending and torsion. The lumped mass and stiffness matrices are used as building blocks for assembling dynamic models of mechanisms involving flexural beam units. Note that we need to restrict these matrix building blocks to parallel kinematic configurations since the geometric boundary conditions corresponding to $x = 0$ have been assumed to be all zero. Formulations for serial kinematic configurations can be developed by altering this set of geometric boundary conditions [56].

For continuity of displacement at each of the nodes P_i, Eq. (2.4) can be used to show that the end-displacements $\mathbf{V}_i(t)$ of every i^{th} flexure are related to the global generalized (rigid body) displacements $\mathbf{V}_R(t)$ as follows:

$$\mathbf{V}_i(t) = \left\{ \begin{array}{c} w_i(\ell,t) \\ \theta_i(\ell,t) \\ \phi_i(\ell,t) \end{array} \right\} = \Re \left\{ \begin{array}{c} z(t) \\ \theta_x(t) \\ \theta_y(t) \end{array} \right\} = \Re \mathbf{V}_R(t)$$

where the transformation matrix $\Re =$

$$\begin{bmatrix} 1 & Rsin\alpha_i & Rcos\alpha_i \\ 0 & -sin\alpha_i & -cos\alpha_i \\ 0 & -cos\alpha_i & -sin\alpha_i \end{bmatrix}$$

Based on the mass and stiffness properties of the individual flexural beam units connected to the central rigid body, we need to derive the mass $\mathbf{M}$ and stiffness $\mathbf{K}$ properties of the assembly shown in Fig. 2-2, which contains the central disk as the rigid body to which the flexures are connected..

By formulating the Lagrangian of the assembly in terms of the rigid body displacements $\mathbf{V}_R(t)$, we develop the lumped mass and stiffness matrices of the overall parallel kinematic mechanism as follows:

$$\mathbf{M} = \sum_{i=1}^{n} \Re^T \mathbf{M}_{fi} \Re + \mathbf{M}_R; \ \mathbf{K} = \sum_{i=1}^{n} \Re^T \mathbf{K}_{fi} \Re \tag{2.5}$$

where $\mathbf{M}_{fi}$ and $\mathbf{K}_{fi}$ are the lumped mass and stiffness matrices, respectively, of the individual flexure building blocks given in Tables 2.1 and 2.2; $\mathbf{M}_R$ is the mass matrix of the rigid body and is given by: $\mathbf{M_R}$=

$$\begin{bmatrix} M & 0 & 0 \\ 0 & J_{Rxx} & 0 \\ 0 & 0 & J_{Ryy} \end{bmatrix}$$

Models such as proportional damping, given by $\mathbf{B} = b_m\mathbf{M} + b_k\mathbf{K}$, are widely used in the literature [61], where b_m and b_k are constants that depend on material properties and are experimentally determined from sine-sweep frequency response measurements.

With this assumption, the equations of motion of the lumped parameter representation, for the free response case, is in the form given below:

$$\mathbf{M}\ddot{\mathbf{V}}_R + \mathbf{B}\dot{\mathbf{V}}_R + \mathbf{K}\mathbf{V}_R = \mathbf{0} \tag{2.6}$$

A solution of the form $\mathbf{V_R(t)} = \psi\mathbf{e^{j\omega t}}$ for the above equation, reduces the above equation to the form of an eigenvalue problem for free vibrations. The eigen modes ϕ appear in the transfer function from the input forcing function to the output displacements of the system. The input-output transfer function of the system can be derived after modal decomposition as [34]:

$$G(s) = \sum_{i=1}^{n} G_i(\omega) \tag{2.7}$$

where each modal transfer function is given by:

$$G_i(\omega) = \frac{\mathbf{C}\psi_i\psi_i^T}{m_i(s^2 + 2s\xi_i\omega_i + \omega_i^2)} \tag{2.8}$$

where ψ_i, ξ_i, ω_i, m_i represent the i^{th} mode shape, modal damping factor, modal frequency, and modal mass, respectively, and C is an output coefficient matrix.

Given the above procedure for assembling lumped parameter models for flexures connected to rigid bodies, we now examine the closed-loop control system implementations that are built around the flexure mechanism, which forms the plant along, with actuators and sensors. A formulation for performance specifications of such systems is presented in the next section.

2.2 Performance Specifications

In this section, we present a formulation for requirements of range, resolution, bandwidth, and load-capacity for flexure-based nanopositioning systems. These formulations become important for developing the cost functions and constraints in an optimization setting. We present here the time-domain state space equations that can help in the analysis. Frequency domain representations are also possible, some of which are formulated in Chapter 4.

Consider a generic closed-loop system, such as the one shown in Figure 2-3. We assume that the plant of the system is composed of an electromechanical subsystem containing a flexure-based mechanism driven by actuators (e.g. piezoelectric stacks). While flexible systems are inherently infinite dimensional, reduced-order lumped parameter models are assumed for the system. Since full state measurements are not readily available in flexible system implementations, an output-based feedback control is chosen. Measurements recorded by sensors (e.g. capacitance probe) are used to design the feedback controller subsystem.

In the presence of stochastic noise sources (a) process noise $\mathbf{w}$ and (b) measurement noise $\mathbf{v}$, we assume the component subsystems have linear time-invariant dynamics

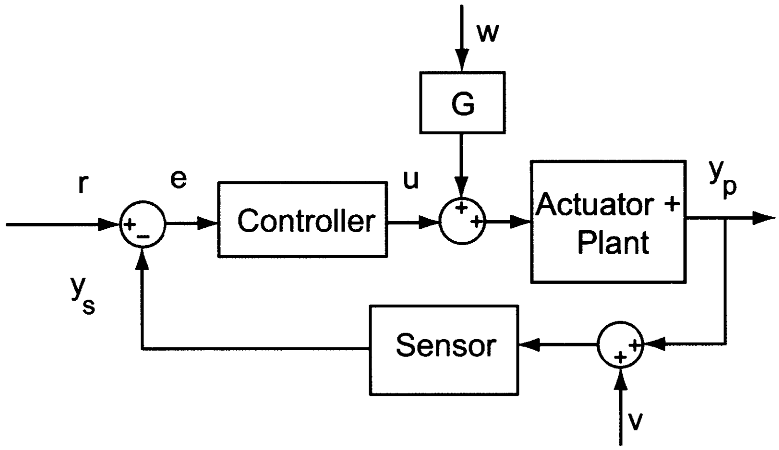


Figure 2-3: A generic closed-loop controlled system block diagram showing component subsystems such as the actuator and plant, sensor and a feedback controller. The reference signal is r, which should be tracked by the output y_s which is fed back to the system. Error signal is e which is amplified by the controller to generate a command u. The output of the actuator/plant susbsystem y_p is measured by the sensor as y_s. A disturbance enters w the system through a block G. A noise v enters the system in the measurement. This system is selected as a common reference system based on which all performance requirements are defined.

defined in the state space according to equations given below. Note that this particular choice of the component subsystems is arbitrary, the underlying approach is meant to be flexible for incorporating any alternative representations.

(i) Electromechanical subsystem (comprising the mechanism and actuator) with a state-space formulation for state vector $\mathbf{x_p}$ of order p, and actuator input u, and disturbance, or process noise, $\mathbf{w}$ in terms of matrices $\mathbf{A_p}, \mathbf{B_p}, \mathbf{G_p}$, and $\mathbf{C_p}$:

$$\dot{\mathbf{x}}_\mathbf{p} = \mathbf{A_p x_p} + \mathbf{B_p u} + \mathbf{G_p w} \tag{2.9}$$

$$\mathbf{y_p} = \mathbf{C_p x_p} \tag{2.10}$$

where $\mathbf{y_p}$ are the plant outputs, without any sensor dynamics and noise, which will include next in the sensor subsystem state equations.

(ii) Sensor subsystem with a state-space formulation for state $\mathbf{x_s}$ of order s, with its input $\mathbf{y_p}$ contaminated by the measurement noise vector $\mathbf{v}$, in terms of matrices $\mathbf{A_s}, \mathbf{B_s}, \mathbf{H_s}$, and $\mathbf{C_s}$:

$$\dot{\mathbf{x}}_\mathbf{s} = \mathbf{A_s x_s} + \mathbf{B_s y_p} + \mathbf{H_s v} \tag{2.11}$$

$$\mathbf{y_s} = \mathbf{C_s x_s} \tag{2.12}$$

(iii) Controller based on output feedback with a state-space formulation for state $\mathbf{x}_c$ of order c, with output $\mathbf{u}$, in terms of matrices $\mathbf{A_c}, \mathbf{B_c}$, and $\mathbf{K_c}$:

$$\dot{\mathbf{x}}_\mathbf{c} = \mathbf{A_c x_c} + \mathbf{B_c}(\mathbf{r} - \mathbf{y_s}) \tag{2.13}$$

$$\mathbf{u} = \mathbf{K_c x_c} \tag{2.14}$$

Defining a new state vector $\mathbf{x} = [\mathbf{x_p}\ \mathbf{x_s}\ \mathbf{x_c}]^\mathrm{T}$ of dimension $n = p + s + c$, the state equations of the overall system reduce to form given below:

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Br} + \mathbf{Gw} + \mathbf{Hv} \tag{2.15}$$

$$\mathbf{y} = \mathbf{Cx} \tag{2.16}$$

$$\tag{2.17}$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{A_p} & \mathbf{0} & \mathbf{B_p K_c} \\ \mathbf{B_s C_p} & \mathbf{A_s} & \mathbf{0} \\ \mathbf{0} & -\mathbf{B_c C_s} & \mathbf{A_c} \end{bmatrix}; \ \mathbf{B} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{B_c} \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{C_p} & \mathbf{C_s} & \mathbf{0} \end{bmatrix}; \ \mathbf{G} = \begin{bmatrix} \mathbf{G_p} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}; \ \mathbf{H} = \begin{bmatrix} \mathbf{0} \\ \mathbf{H_s} \\ \mathbf{0} \end{bmatrix}$$

In what follows, we assume that each of the above component subsystems, and the overall closed-loop system, are stabilizable and reachable. Now, we formulate the performance requirements for the system. Note that we have defined resolution and range based on outputs, the presented framework also applies to unmeasured states, such as a velocity of a flexural probe tip.

2.2.1 Resolution

Resolution for a positioning system can be defined as the smallest incremental change that can be achieved in the desired output of the system. Achieving fine resolution is a critical requirement for most precision applications. This requirement defines the positioning error in regulator problems, where a set point displacement or velocity has to be achieved, or in tracking problems, where a reference input has to be tracked in time.

Within the framework of the system described above, the variance of the output vector provides a measure for the resolution possible with the overall system. For the case when the commanded reference input is deterministic, the input does not affect the resolution. Resolution is only limited by the stochastic noise sources $\mathbf{w}(t)$ and $\mathbf{v}(t)$. We assume that these sources are (i) Gaussian white noise processes with zero mean and variances $\mathbf{R_w}$ and $\mathbf{R_v}$, respectively and (ii) uncorrelated with respect to each other and also to initial conditions of the system.

Under the above assumptions, the resolution $y_{i,min}(t)$ of the i^{th} output $y_i = \mathbf{C_i}\mathbf{x}$ of the output vector $\mathbf{y}$, is given as:

$$y_{i,min}(t) \geq \sqrt{\rho(\mathbf{C_i}\mathbf{P}(t)\mathbf{C_i}^T} \tag{2.18}$$

where $\rho[.]$ denotes the spectral radius (largest eigenvalue) of its matrix argument, and $\mathbf{P}(t) = E(\mathbf{x}\mathbf{x}^T)$ is the variance of the state vector $\mathbf{x}$. The smaller the value of

$y_{i,min}(t)$, the finer the resolution of the system. Note that the state variance matrix $\mathbf{P}(t)$ satisfies the stochastic Lyapunov matrix differential equation:

$$\dot{\mathbf{P}}(t) = \mathbf{A}\mathbf{P}(t) + \mathbf{P}(t)\mathbf{A}^T + \mathbf{G}\mathbf{R}_{\mathbf{w}}\mathbf{G}^T + \mathbf{H}\mathbf{R}_{\mathbf{v}}\mathbf{H}^T \tag{2.19}$$

For an initial condition $\mathbf{P}(0) \geq \mathbf{0}$, the analytical solution of the above equation in terms of the state transition matrix $\phi(t) = e^{\mathbf{A}t}$ is given as follows [87]:

$$\mathbf{P}(t) = \phi(t)\mathbf{P}(0)\phi(t)^T + \int_0^t \phi(t-\tau)\left\{\mathbf{G}\mathbf{R}_{\omega}\mathbf{G}^T + \mathbf{H}\mathbf{R}_{\mathbf{v}}\mathbf{H}^T\right\}\phi(t-\tau)^T d\tau \tag{2.20}$$

The above equation can be numerically solved to determine the bounds on the variance $\mathbf{P}(t)$ of the state vector, and hence the resolution defined in Eq. (2.18) can be computed.

2.2.2 Range

Range for a positioning system can be defined as the largest change that can be achieved in the desired output of the system. Achieving a large range, such as a maximum displacement, velocity, or acceleration can be a critical requirement for most precision applications. This requirement defines the largest opening of a gap that can be used for size-based particle filtration, or the scan area in stages used for positioning measurement samples in scanning probe instruments.

Within the framework of the system described above, the range $y_{i,max}$ of the desired output signal $y_i(t)$ can be defined as a maximum norm in time. Of possible choices, such as 1-norm, 2-norm (energy), we consider the power or RMS semi-norm [65] as a measure of the strength of the signal over time. The range $y_{i,max}$ is defined in terms of the power norm as:

$$y_{i,max} \triangleq \|y_i(t)\|_{pow} = \lim_{T\to\infty} \sqrt{\frac{1}{2T}\int_{-T}^{T} |y_i(t)|^2 dt} \tag{2.21}$$

This choice is justified since the inputs generated by the actuator are bounded in power. For power-bounded reference inputs, the power norm value can be related to the $\mathcal{H}_\infty$ norm, i.e. the singular value measures in the frequency domain, using the matrices **A** and **B** defined from Eq. (2.17) as follows:

$$y_{i,max} \triangleq \|y_i(t)\|_{pow} = max_\omega \; \overline{\sigma}(\mathbf{C_i}(j\omega\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}) \tag{2.22}$$

While the above relation specifies the maximum amplification given an unit upper bound for the reference input, the maximum value of the desired output will scale in magnitude depending on the value of the upper bound on the reference input. In practice, the upper bound that can be specified for the reference input is limited by the physical constraints in the system, such as actuator saturation.

2.2.3 Bandwidth

Bandwidth indicates the frequencies over which command following can be closely achieved with feedback control. This specification dictates the overall response time of the positioning system to a command input. While careful controller choice can improve on the achievable bandwidth of a system [82], it is ultimately limited by the natural frequencies set by the physical plant.

Within the framework of the system described above, the closed-loop system sensitivity transfer function $\mathbf{S}(j\omega)$ is related to product of all gain blocks in the system, i.e. the loop transmission function $\mathbf{L}(j\omega)$, as follows:

$$\mathbf{S}(j\omega) \;=\; (\mathbf{I} + \mathbf{L}(j\omega))^{-1} \tag{2.23}$$

Bandwidth ω_0 can be computed as the frequency where the maximum singular value $\overline{\sigma}(\mathbf{S})$ of the closed-loop system sensitivity transfer function $\mathbf{S}(j\omega)$ crosses $-3\ dB$ from below [65]. This relation is as shown below.

$$\overline{\sigma}(\mathbf{S}(j\omega_0) \quad = \quad \sqrt{\frac{1}{2}} \tag{2.24}$$

2.2.4 Load Capacity

In Chapter 3 we define load-capacity as the maximum force that the flexure-based structure can sustain without undergoing failure. The stress generating within a flexure-based mechanism subject to a given loading needs to satisfy the following condition:

$$S < S_{max} = \frac{S_Y}{SF} \tag{2.25}$$

where S_Y is the yield strength of the material and SF is a safety factor. The load-capacity can be defined as the force that increases the stress levels to S_{max}. For a given flexure topology, the maximum force that can be sustained is a function of the geometry, dimensions, and material of the flexure. In Chapter 3 we determine the load-capacity for common flexure building blocks and propose a generic building block that can achieve large load-capacity while still satisfying a large range requirement.

2.3 Summary

In this chapter, we covered basic modeling of flexure-based mechanisms. Starting from the governing partial differential equation, we can use a lumped parameter approach such as assumed modes to generate the component level mass and stiffness matrices. Proportional damping is assumed, but is likely to be not achieved in practice. Physical damping in flexure materials is low and hence makes controlling them a difficult task. We need to address active means for tackling damping. Chapter 4 discusses this issue and presents a sensitivity tuning method that accounts for the lightly damped harmonic modes.

Further, we also examined resolution, range, bandwidth, and load-capacity as some of the performance requirements specified for most flexure-based nanopositioning systems. We have formulated these requirements using state space analysis. Using Parseval's theorem, the expressions we derived in time domain can be converted to equivalent representations in the frequency domain. The expressions given here can be used in the cost of constraint functions of any optimization routine set up for flexure-based nanopositioning. This will be discussed in Chapters 4 and 5 and an example optimization problem for achieving a desired bandwidth is worked out in Section 5.4.

Chapter 3

Design Topology Generation

In this chapter, we propose a novel framework for the generation of design topologies directed at improving performance in terms of strength, or structural integrity, and dynamic behavior. First, we explain the difference between topology selection and shape and size selection. While the former refers to selecting a design configuration, the latter refers to fine-tuning the parameters of a given configuration. The described method for topology generation is applied to many specific cases for practical applications in motion control and manufacturing. Existing designs that can be explained and derived using our topology generation method are highlighted. The approach described in this chapter will be used in Chapter 6 to develop a library of candidate topologies for the AFM nanopositioner.

3.1 Background and Motivation

Flexure-based mechanisms have been around for many decades. A survey of the history of flexure mechanisms is detailed in [18]. R.V. Jones [17] pioneered some of the early parallelogram flexure designs in high-precision instrumentation during the early 1940's and 1950's. While there have been a number of flexure-based mechanism designs showing up in the field of precision engineering for the past many decades [14], [48], [49] focus of interest has been in the compliant mechanisms community for flexures that generate large ranges of motion. In both the fields, while many designs have been developed over the years on a case-by-case basis using a forward approach involving the sequence of design conception, analysis, and performance optimization, recent efforts have focused on using an inverse approach, i.e. synthesizing designs that meet a given set of end performance requirements.

There have been a number of approaches for design synthesis in the literature for compliant mechanisms [80]. Howell et al proposed a method based on modeling compliant members, such as flexures, as pseudo rigid bodies, i.e. rigid bodies with localized joint stiffnesses. By varying the location and value of the lumped stiffness, different pseudo rigid body models can be generated, and correspondingly different topologies can be mapped out.

Many methods for design synthesis adopted in the precision engineering community are based on arranging the constraints of motion and selecting flexures that implement those constraints. Exact constraint design, as well documented in [77], [78], is based on implementing designs using kinematic principles. Every rigid body has six degrees of freedom. If n constraints are imposed on the rigid body, the available degrees of freedom are $n - 6$. In this approach, the desired number of degrees of freedom are first identified for a given design problem. The curbed degrees of freedom are achieved by implementing as many constraints with suitable flexure building blocks [78].

While exact-constraint methods provide a deterministic approach to design, an alternate method known as elastic averaging provides more of a probabilistic approach [14]. In elastic averaging many redundant constraints are allotted so that the resulting design 'averages' out any imperfections in manufacturing and assembly. A discussion of elastic averaged designs in the context of flexure-based mechanisms is presented in [43]. In what follows, we formulate the design objectives and a synthesis approach for meeting these objectives.

3.2 Our Approach

We present here a synthesis approach for meeting some the design objectives of a precision positioning application. We restrict our attention here to systems implemented with flexure-based mechanisms that are (i) driven by high-resolution actuators, such as piezoelectric stack actuators, and (ii) measured for their fine motions by high-resolution sensors such as laser interferometers or capacitance probes. Within this scope, the design objectives include, for example, meeting performance requirements in terms of (i) strength behavior (avoiding failure by structural limitations such as yield, fatigue, buckling), (ii) modal behavior in frequencies of interest, and (iii) thermal behavior. In this thesis we focus on the former two requirements. Detailed design formulations for these objectives are provided in Sections 3.4 and 3.5.

Before we proceed any further, we need to introduce a few terms that will be repeated throughout the thesis. In synthesis methods, the layout or configuration of a design is often referred to as its topology. In designing mechanisms, a basic distinction needs to be made between selecting a design topology and selecting parameters within a topology. For example, consider a design objective of causing motion along a road. One could select a car as a design configuration. Another choice that also meets the design objective is a bicycle. These two design choices constitute different design topologies. Based on a specified performance requirement one could select one topology over the other. For instance, in case of a requirement of higher speed, a car can be chosen over the bicycle. Fine-tuning of the car design so that a higher speed (among all possible cars) can be achieved constitutes a further optimization step, often referred to as "shape and size optimization."

As another example for design topologies, we consider direct drive as against transmission drives. A direct-drive motor constitutes a design topology that is different from one based on a motor with a transmission, such as a belt drive. Within the topology based on a motor with a transmission, the stiffness and damping in the transmission belt can be tuned for better performance, for example, higher servo bandwidth or disturbance rejection. So, in summary, topology selection refers to the choice between a direct drive motor or a motor with a transmission. Within a selected topology, for e.g. of the motor with the transmission, fine-tuning design parameters such as stiffness or damping in the transmission constitutes shape and size optimization.

In the synthesis of flexure-based mechanisms, in both the precision engineering and complaint mechanisms communities, topology generation is seen to be a pivotal step in the design process [18]. Kim et al [79] proposed a method for synthesizing building blocks based on considering constraints and compliance ellipsoids for achieving the desired kinematics. Hopkins et al [18] have addressed the allotment of design degrees of freedom and constraints to generate design topologies that achieved the desired kinematics and elastomechanics for parallel and serial flexure systems. Using projection geometry, a comprehensive list of design topologies for satisfying multi-degree-of-freedom kinematics were derived. Awtar et al [16] have analyzed fundamental flexural building blocks for their kinematics and elastomechanics. The analysis of dynamical performance of flexures in the context of mechanical advantage is detailed in [50]. A finite-element approach based on Euler-Bernoulli beam bending theory is formulated for analyzing dynamics in [24] and optimizing the design space for precision flexure-based applications in [3].

To the best of our knowledge, there are no topology generation or synthesis methods that address achieving a desired strength or dynamic performance in the context of closed-loop control. In what follows, we present a novel framework that allows one to generate design topologies based on performance requirements such as strength and dynamic performance. A selected topology can subsequently be fine-tuned with a shape and size optimization procedure.

3.3 Proposed Framework

A novel framework [1] that we propose for topology generation is shown in Figure 3-1. In this framework, we start with a set of primitives. These primitives are subjected to a finite number of operations dictated by performance requirements. These operations could be, for example, a parallel or serial replication, or a geometrical transformation, or adding a redundant constraint that imparts symmetry. The primitives subjected to these operations generate building blocks that meet the desired performance requirement.

Once the building blocks are generated, a library of design topologies can be generated by using the building block as an implementation of the constraints (following a constraint-based synthesis approach) for satisfying the necessary kinematics. In a nutshell, using the performance-tuned building block allows to meet a strength, or modal performance criterion, while the constraint-based arrangement allows for satisfying the required kinematics.

We now examine our framework explaining each of these features in more detail in the following sections.

3.3.1 Primitives

We define primitive as a simple physical implementation of a flexure constraint. A primitive constitutes the start or the initial implementation over which we improve using a set of operations, which we will describe shortly, to create building blocks for

---

[1] Parts of the framework developed in this chapter are built on our work documented in our paper at the American Society for Precision Engineering Annual Conference, Portland OR, October 2008 cited as [39] in this thesis.

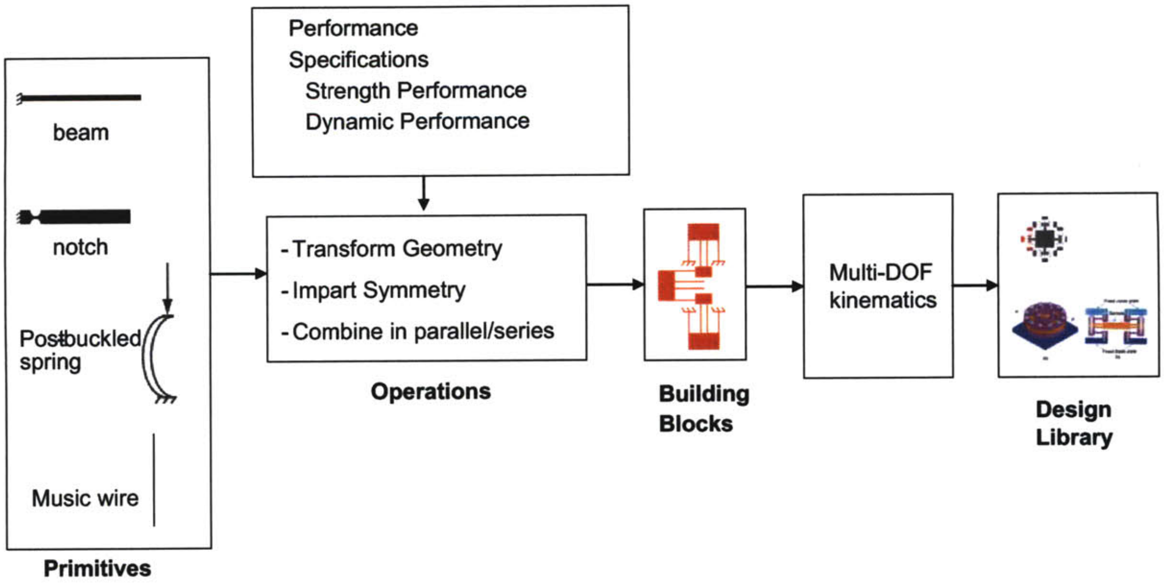


Figure 3-1: A performance-driven design library shown as constructed from building blocks prepared by performance-driven operations on a set of primitives. Novel designs synthesized with this method are schematically shown in the Design Library block.

the design topology.

A set of design primitives are shown to the left in Fig. 3-1. Primitives can be viewed as falling mainly into two categories: those with localized compliance and those with distributed compliance. Example of a primitive with a localized compliance is a notch joint in an otherwise rigid member.

Examples of a primitive with a distributed compliance include:

- A slender beam flexure, also referred to as a blade flexure.[2]
- A wire or rod flexure, i.e a beam with a circular cross-section, as in guitar strings or musicwires.[3]

Some of these primitives can be flexures that have tunable behavior based on how they are physically set up. For example, a flexure member subject to loads exceed-

---

[2] A simple analogy often used to describe a blade flexure is the diving board of a swimming pool.

[3] Spring steel musicwires of about 0.5 mm radius when cut to short lengths (about 0.5 in) are marked by their high longitudinal stiffness. They are ideal flexures for pinning down trampoline-like vertical modes of a motion stage undergoing lateral translation. This idea has been implemented for the AFM Nanopositioner in Chapter 6.

ing its buckling limit can be significantly softer than when the loads are relatively smaller [36].

If a design topology is being developed *ab-initio*, i.e. from scratch, any of the above primitives can be chosen. Or a combination of the primitives can be chosen as well. In case of a design revision of an existing topology, the fundamental flexural building block in the existing topology can be regarded as a primitive on which the performance-driven operations will be run to generate new building blocks.

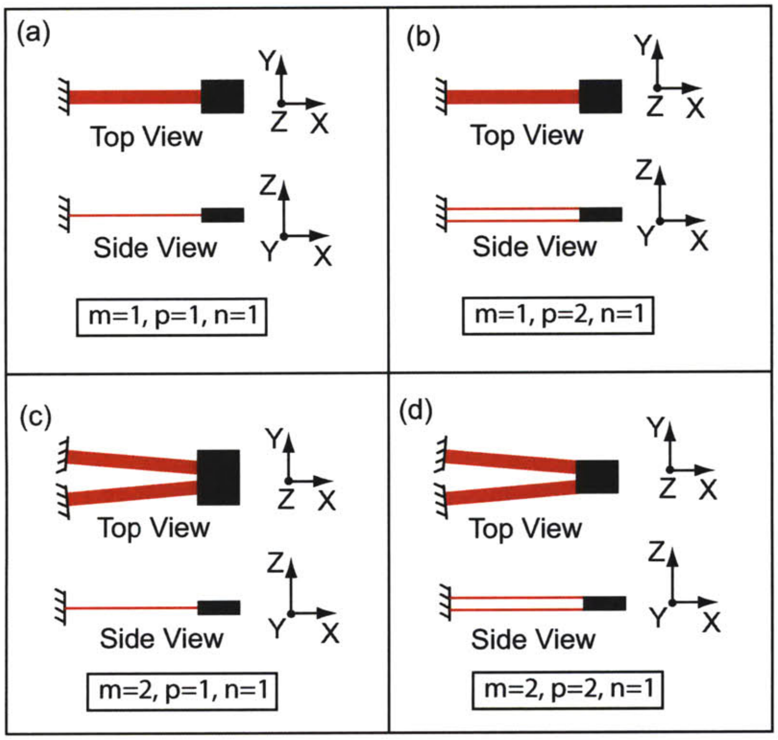


Figure 3-2: Parallel and serial replication of flexure beam to cause redundant constraint that allows for meeting a desired performance requirement. A simple blade flexure primitive is shown in (a). It has stacking indices $m = 1$, $p = 1$, and $n = 1$. The analogy for this flexure is a diving board of a swimming pool. Parallel replication is achieved by stacking in two ways. In (b) the beams are stacked in a plane transverse to the neutral plane of the beam, i.e. they are stacked on top of another. This building block corresponds to $p = 2$ with m and n values the same as its primitive in (a). In (c), the beams are replicated azimuthally, i.e. along the neutral plane of the beam. This stacking corresponds to $m = 2$, while p and n values are the same as the primitive shown in (a). Finally, parallel stacking both azimuthally and transersely, with indices as $m = 2$, $p = 2$, and $n = 1$.

3.3.2 Operations

We define operations on the concept-level as methods of adding a prescribed level of redundancy to a primitive. Redundancy is needed so that a desired performance requirement can be met. When a set of operations are performed on the primitive, the added redundancy will make the building block tuned better to the performance requirement.

The operations we consider here include parallel and serial replications of flexure primitives so that a functional building block can be generated.

Under a parallel replication, the starting primitive and its replica share the applied load. Just as springs arranged in parallel have an increase in the resultant stiffness, parallel replication of flexure primitives results in an increase in the stiffness.

Under a serial replication, the starting primitive and its replica are subject to the same loading in the mechanical load path. Just as springs arranged in series have a decrease in the resultant stiffness, serial replication of flexure primitives results in a decrease in the stiffness.

The parallel and serial replications are illustrated as follows. Here, we introduce indices to parameterize the topology generation. The parallel replication indices are m and p, while the serial replication index is n. Consider the simple beam flexure Fig. 3-2(a). A parallel replication for this flexure constraint can be done in two ways as shown in Fig. 3-2(b) and 3-2(c):

(I) Stacking primitives in a direction normal to the bending neutral plane (plane perpendicular to this page and passing through axis Y in the figure). This method of stacking is as shown in Fig 3-2(b).[4]

Let us denote by p the number of flexures stacked transversely (to the neutral plane) in this manner. In this thesis we will be using this index to refer to

---

[4] Going back to our analogy of swimming pool diving boards for beam flexures, a parallel replication described here for stacking normal to the neutral plane is the same as stacking diving boards one on top of another.

"parallel transverse stacking."

For the building block of Fig. 3-2(b) we have $p = 2$, while m and n remain unchanged with respect to the primitive in Fig. 3-2(a).

(II) Stacking primitives along the bending neutral plane in an azimuthal manner, i.e. flexures spread out in the neutral plane at a set of chosen angles in the total 4π radians possible. This way of stacking is shown in Fig. 3-2(c).[5]

Let us denote by m the number of flexures stacked azimuthally in this manner. In this thesis we will be using this index to refer to "parallel azimuthal stacking."

For the building block of Fig. 3-2(b) we have $m = 2$, while p and n remain unchanged with respect to the primitive in Fig. 3-2(a).

Fig. 3-2(d) shows a building block obtained with both parallel transverse stacking and azimuthal transverse stacking. The indices for this building block are $m = 2$, $p = 2$, and $n = 1$.

We now examine series replication. The building block in Fig. 3-4(e) is obtained by a series replication of the flexure constraint of the primitive in Fig. 3-2(a). By series replication, we mean that the flexure is "folded back," effectively increasing the bending length. The resulting increase in compliance can facilitate an increase in deflection. Note that an intermediate mass is used in this folding back of the flexure.

Now we can combine series replication with the parallel replication that we just described in the context of Fig. 3-2. The building block in Fig. 3-4(f) is obtained by a parallel azimuthal stacking as seen in the top view. The building block of Fig. 3-4(f) uses a parallel transverse stacking of $p = 2$, along with the series replication of $n = 2$. This concept is the well-known double parallelogram flexure [17] used originally by British Scientist Dr. R.V. Jones during the post World War-II era.

---

[5] Again, using the analogy of swimming pool diving boards for beam flexures, a parallel replication described here for replicating in the neutral plane is the same as having two diving boards adjacent to each other.

Finally, the building block shown in Fig. 3-4(h) has parallel transverse stacking, parallel azimuthal stacking, and serial stacking, i.e. $m = 2$, $p = 2$, and $n = 2$.

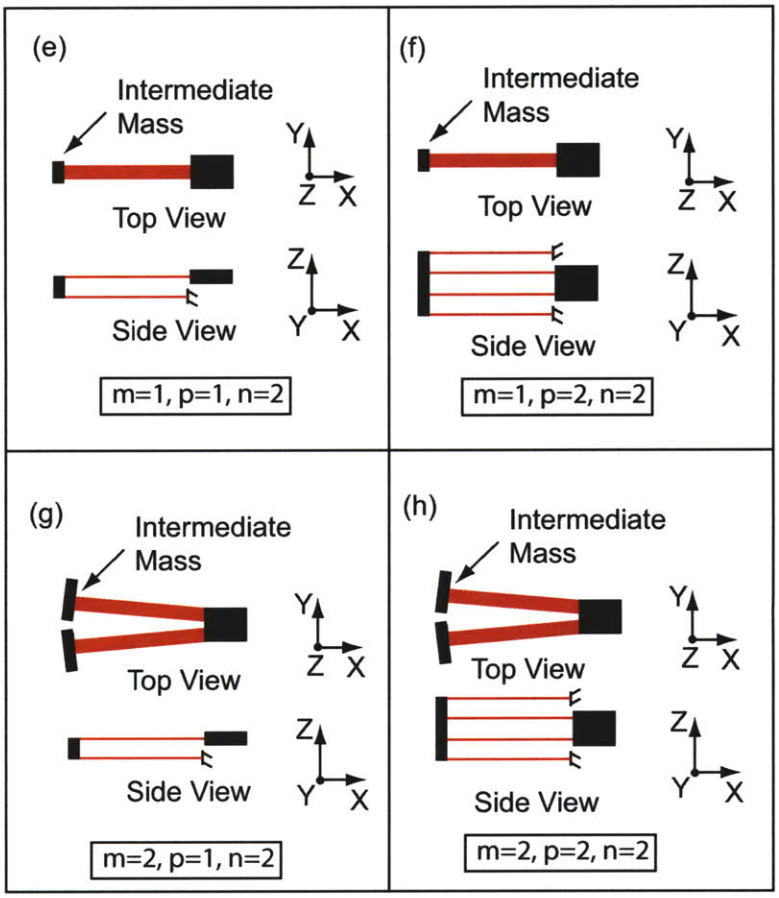


Figure 3-3: Operations on primitive of Fig. 3-2(a) showing series and parallel replication steps. Series replication is achieved in (e) by "folding back" the primitive of Fig. 3-2(a), effectively increasing its bending length. Series replication with $n = 2$ is achieved along with parallel transverse stacking $p = 2$ is the well-known double parallelogram flexure [?] configuration in (f). The concept of (e) is replicated azimuthally, i.e. with $m = 2$ to obtain the concept topology in (g). A building block with parallel azimuthal stacking with $m = 2$, transverse stacking $p = 2$, and series stacking $n = 2$ is shown in (h).

Other operations are illustrated for the case of a XY beam flexure shown in Fig. 3-4(a). The operations are listed as follows:

(i) Adding a redundant constraint to impart symmetry: In Fig. 3-4(b), a redundant beam flexure is added in $-Y$ to make the structure symmetric about X axis.

(ii) Performing a geometric transform: In Fig. 3-4(c), two flexures are placed between

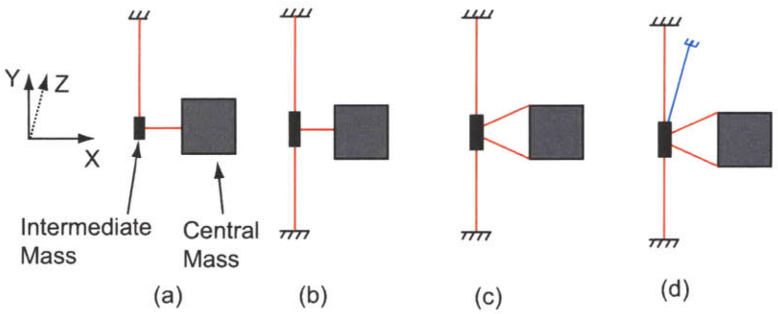


Figure 3-4: Operations on primitives that enhance a strength or dynamic performance requirement.

the masses and they are aligned at a tangent circumscribing the central mass. With this configuration, the θ_Z mode is stiffened to minimized parasitic θ_Z rotations of the central mass. In Fig. 3-4(c), a redundant constraint is added in Z direction to curb the out-of-plane modes of the intermediate mass. These examples serve to illustrate the underlying principles, and in no means constitute the only way to perform the operations.

Imparting symmetry is a general guideline to be followed in designs that need to be thermocentric, and hence avoid warping and parasitic errors. Performing a geometric transform is a useful operation when the topology needs to have large separation in natural frequencies of the modes of interest.

3.3.3 Building Blocks

The operations described above are performed on the primitives to generate building blocks of the design topology. Here, we restrict our attention to planar building blocks. One could also consider stacking versions of planar ones, that can be used in a three-dimensional arrangement to result in a multi-DOF flexure system.

All the operations of parallel and serial replications on the primitive of Fig. 3-2(a) resulted in building blocks of Figs 3-2(b)-(d) and Figs 3-4(e)-(h).

As another example fora building block generated from a primitive, consider the primitives, the operations performed, and the final building block in Fig. 3-5. The primitive in Fig. 3-5(a) is a simple blade flexure. This primitive is reinforced for its bending along X axis (and also imparting symmetry about X axis) to create a double blase flexure building block in Fig. 3-5(b). With another beam flexure that can impart Y compliance, we arrive at the XY-compliant building block of Fig. 3-5(c).

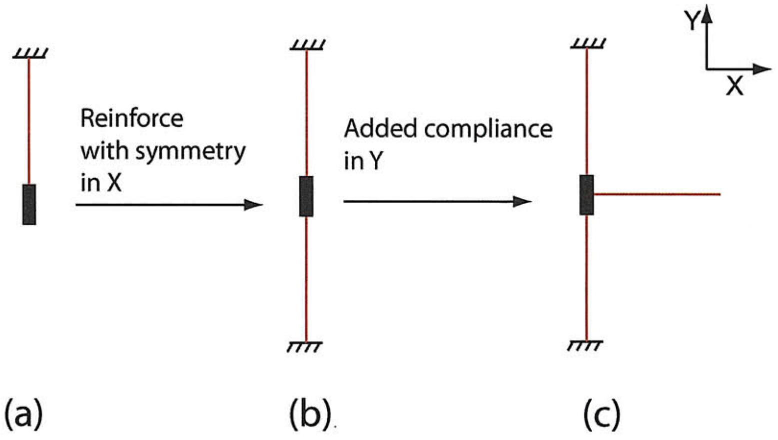


Figure 3-5: Operations on primitives that enhance a strength or dynamic performance requirement.

3.3.4 Multi-DOF Arrangement

Any of the constraint-based methods can be used to arrange the performance-driven building blocks generated in the previous step in multiple directions so that the required kinematics are achieved. This is illustrated for the building block of Fig. 3-5(c)

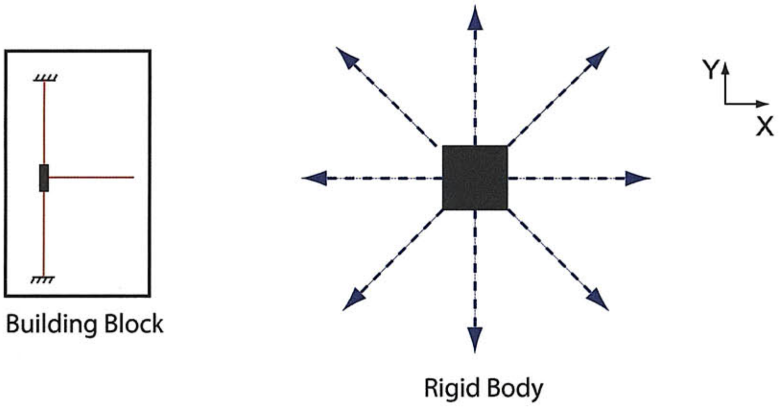


Figure 3-6: Building blocks are placed along multiple directions to obtain the necessary degrees of freedom.

in Fig. 3-6. To impart $X - Y$ degrees of freedom to the rigid body shown in Fig. 3-6, the building block can be placed in one or more of the directions indicated by the dashed arrows.

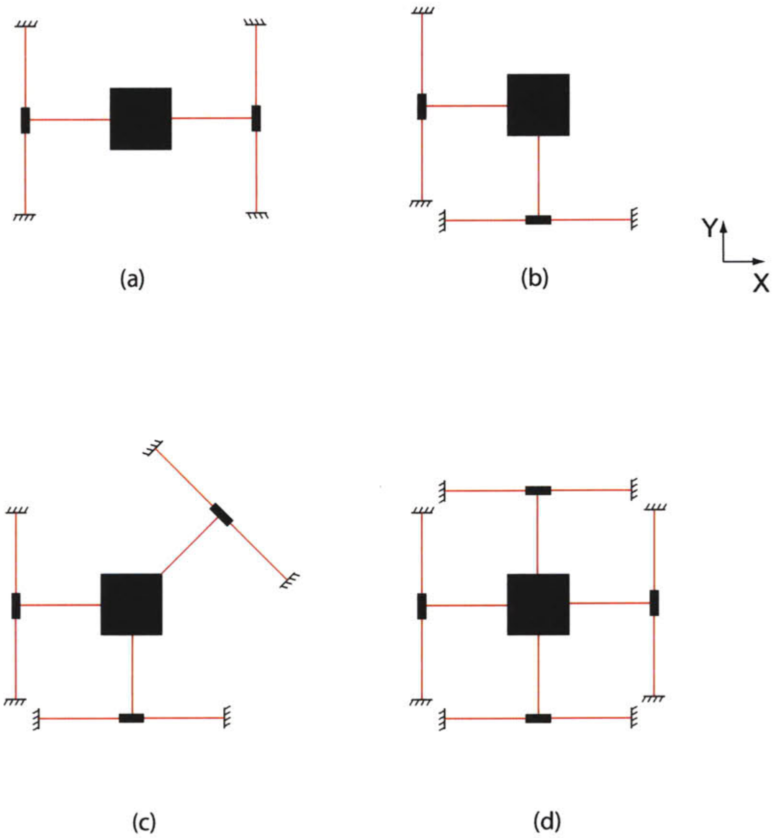


Figure 3-7: Multiple topology concepts generated from placing the building block of Fig. 3-5 around the central rigid body that needs to be guided.

A few such topologies are indicated in Fig. 3-7. Of all the four topologies, the one presented in Fig. 3-7(d) is symmetric with a 2-fold symmetry. The topology has X, Y and also θ_Z degrees of freedom.

To summarize the above section, we have presented a novel framework for design topology generation. We parameterized ways to include redundancies in flexure constraint arrangements such that a desired performance requirement can be met. The stacking indices that we introduced are repeated here, in Table 3.1.

Now that we have the methods of stacking and hence generating different topologies, we need to understand how strength or dynamic performance requirements can

Table 3.1: Stacking indices for Flexure Design Topology Generation.

Index	Description
m	Parallel azimuthal stacking
p	Parallel transverse stacking
n	Serial stacking

make a designer prefer some topologies over the others. In what we covered so far, we also hinted at making motion along one axes much stiffer or much more compliant than any other axes. This relative modal behavior will be captured with another index that will be introduced in the next section.

3.4 Topologies for Strength Performance

In this section, we discuss critical strength performance requirements. We formulate the generation and subsequent selection of design topologies based on these requirements. Strength and stiffness are often confused as being the same [80] goals of design. A high-strength design is not necessarily the design with highest stiffness. We discuss these topics and also provide a method for selecting appropriate design topologies for strength based on the stacking indices that we introduced in the previous section. An illustrative example is also presented.

3.4.1 Motivation

Strength is an important design criterion in flexure-based mechanism synthesis. A number of applications require the structural integrity of the flexures to be intact under the application of large static or dynamic loads. Load-capacity in this context is primarily dictated by the yield limit in the flexures reached by stresses developed under the applied load.

For example, a case of flexure-based mechanism design under large static loads is as follows. In stamping applications such as micro-hot-embossing tool-sample angular alignment is critical for forming channels with uniform depth. Consider the case of 1-micron deep channels being formed over a 75 *mm* × 25 *mm* area (this area corresponds to that of a microscope slide), a 1 *mdeg* (17.4 *μrad*) tool-sample misalignment can result in a 13% variation of depth across the 75 *mm* lateral span. Assuming other variations are held constant, this height variation corresponds to about a 39% variation in the flow rate across the polymer part, using a simplified Poiseuille flow approximation in the channels. Such a flow rate variation is unacceptable for many precise flow control applications, e.g. microfluidic devices designed for controlled drug delivery. A flexure-based mechanism that supports the sample in such applications, sees a large load on the order of a few kN, and it becomes critical to design for strength.

Another example is MEMS implementation of flexure-based mechanisms, where the flexible structures are subjected to cyclical loads, with the number of cycles exceeding millions. In such MEMS devices, structural reliability over long cycles is a critical requirement. In this case as well, design for strength can take into account the fatigue behavior and allow for durable and sustained operation.

In particular, we need to address the following question: does the largest load-capacity design topology have to be the stiffest? Is it possible to design a large load-capacity flexural positioning system that meets the range requirement as well? A design methodology for load-capacity driven design is lacking currently in the literature. A systematic methodology provided here is aimed at facilitating synthesis of new design topologies and comparison of candidate design topologies.

3.4.2 Method

In this section, we present the design decisions and evolution of candidate design topologies that can meet high strength requirements. We fundamentally distinguish strength-based designs from stiffness-based designs. Designing for high strength need not necessarily imply designing for high stiffness.

To explain the above point, first, we consider the relations for load-capacity F_{max}, stiffness K, and range Δ_{max} for the fundamental flexural building blocks of Figure 3-8. The first four rows of Table 3.2 show the relations for these building blocks in terms of the parameters: maximum allowable stress σ_a and E_0 is: (i) E, the Young's Modulus for plane stress and (ii) $\frac{E}{1-\nu^2}$ for plane strain, ν being the poisson's ratio.

The relations given in Table 3.2 can be derived from simple linear beam bending lumped parameter models. For simplicity, we consider only small-deflection behavior, i.e. the beam deflections are an order of magnitude smaller than the beam thickness. While a detailed FEA can refine and be used in advanced design phases, the lumped models presented here enable closed-form solutions and allow for quick topology generation in the initial design phase.

The same requirements for a generic building block that can improve strength performance is as indicated in the last row of Table 3.2. This building block has a set of parallel replications, m in the azimuthal direction, p in the transverse direction, and n serial replications of the flexure constraint. This redundancy allows for meeting the strength performance requirement. Recall that we have parameterized topology generation with parallel and serial replication operations as illustrated in Fig. 3-2 and Fig. 3-4.

Referring to Table 3.2, note here that the allowable stress limit needs to be specified based on the physics of the problem. For a static load application with beams subjected to bending loads, the allowable stress limit can be selected as a value lower than the yield limit by a safety factor. If fatigue performance has to be met, the allowable stress limit can be set to a value lower than the endurance limit of the material by a safety factor. If the maximum stress in the mechanism is less than the endurance limit, infinite life is expected [21], [48]. Both of these limits are highly

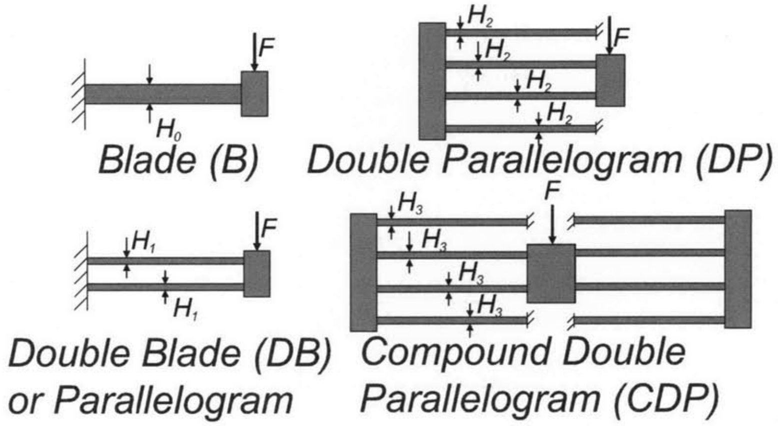


Figure 3-8: Flexural building blocks of length L, width (into the page) B, and beam thicknesses defined as follows: H_0 for the blade flexure, H_1 for the double blade (or the parallelogram) flexure, H_2 for the double parallelogram flexure, and H_3 for the compound double parallelogram flexure.

Table 3.2: Comparison of flexural mechanisms

Flexural Building Block	F_{max}	K	Δ_{max}
Blade	$\frac{1}{6}\frac{BH_0^2}{L}\sigma_a$	$\frac{E_0BH_0^3}{4L^3}$	$\frac{2}{3}\frac{L^2}{H_0}\frac{\sigma_a}{E_0}$
Double Blade	$\frac{2}{3}\frac{BH_1^2}{L}\sigma_a$	$\frac{2E_0BH_1^3}{L^3}$	$\frac{1}{3}\frac{L^2}{H_1}\frac{\sigma_a}{E_0}$
Double Parallelogram	$\frac{2}{3}\frac{BH_2^2}{L}\sigma_a$	$\frac{E_0BH_2^3}{L^3}$	$\frac{2}{3}\frac{L^2}{H_2}\frac{\sigma_a}{E_0}$
Compound Double Parallelogram	$\frac{4}{3}\frac{BH_3^2}{L}\sigma_a$	$\frac{2E_0BH_3^3}{L^3}$	$\frac{2}{3}\frac{L^2}{H_3}\frac{\sigma_a}{E_0}$
Proposed Generic Building Block	$pm\frac{2}{3}\frac{BH^2}{L}\sigma_a$	$\frac{pm}{n}\frac{2E_0BH^3}{L^3}$	$n\frac{1}{3}\frac{L^2}{H}\frac{\sigma_a}{E_0}$

material dependent. For example, Al 6061 alloy has a yield limit of about $70MPa$, on the other hand Al 7075 alloy has a limit of about 550 MPa. For good fatigue behavior, Ti Alloys such as Ti-6Al-4V have an endurance limit of about 500 MPa. A list of commonly used materials for flexure-based designs and their properties is given in [48].

What distinguishes our approach from others is examining the range and stiffness performance space for two different design topologies meeting the same strength requirement. Table 3.3 presents a comparison for range and stiffness values when the flexural topologies of Table 3.2 all have the same load-capacity. For simplicity, we

Table 3.3: Flexural mechanisms of Table 3.2 designed for the same load-capacity.

Flexural Building Block	H	K	Δ
Blade	H_0	K_0	Δ_0
Double Blade	$\frac{1}{2}H_0$	K_0	Δ_0
Double Parallelogram	$\frac{1}{2}H_0$	$\frac{1}{2}K_0$	$2\Delta_0$
Compound Double Parallelogram	$\frac{1}{2\sqrt{2}}H_0$	$\frac{1}{2\sqrt{2}}K_0$	$2\sqrt{2}\Delta_0$
Proposed Generic Building Block	$\frac{1}{2\sqrt{pm}}H_0$	$\frac{1}{n\sqrt{pm}}K_0$	$n\sqrt{pm}\Delta_0$

assume length L and width B of the beams are kept the same. But their thickness H is varied such that the designs have the same load-capacity. In essence, for achieving large load-capacity, reinforcing flexural beam thickness is a poor choice, since an increase in thickness results also in an increase in stiffness and hence reduced range. As is evident from Table 3.3, the new proposed design has a range tunable by the parameters m, n, and p, and can be much larger than that of an equivalent thick beam primitive.

A point to be emphasized here is that it is not necessary that the largest load-capacity design has to be the stiffest. Note that serial replication does not affect the load-capacity since series addition of beams does not result in load-sharing. Sufficient range can be achieved by tuning n, which does not affect the load-capacity since series addition of beams does not result in load-sharing.

In summary, the strength-based design topology generation uses a combination of parallel and serial replication of the flexure constraints. For topologies targeted at achieving maximum range, a simple set of spring mass systems can illustrate the steps to be implemented. Let us assume our primitive is a single thick beam flexure as shown in Fig 3-9(a). This is schematically shown as a stiff spring in Fig. 3-9(A).

A relaxation of the flexure constraint by a parallel combination of softer constraints results in a new building block shown in Fig. 3-9(b). The corresponding spring mass system showing two soft springs is in Fig. 3-9(B). Note when the primitive and the

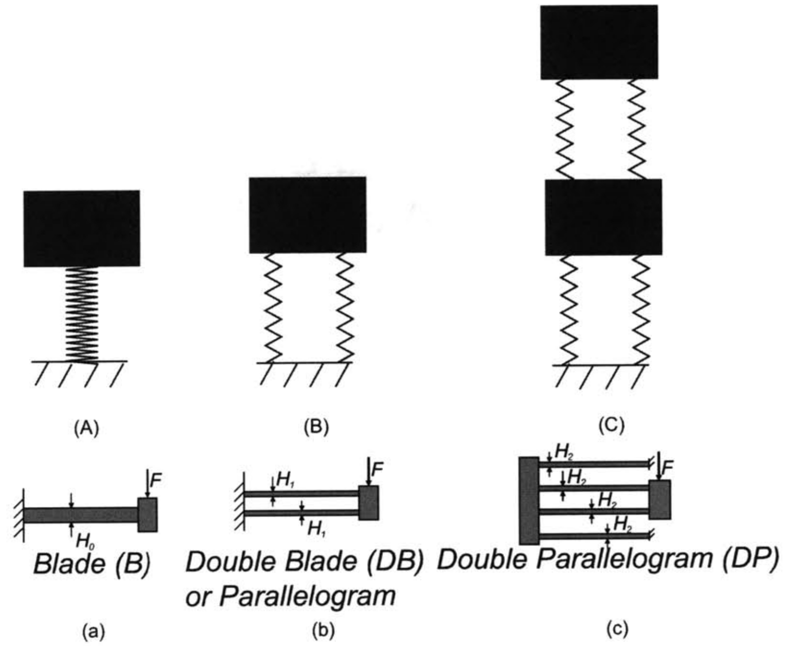


Figure 3-9: Simple spring mass systems shown to explain the underlying physics of parallel and serial replications in strength-based topology design.

building block are designed so that they have the same load-capacity or strength, the building block in Fig. 3-9(b) has a larger range of motion.

Now we can further relax the stiffness by a serial replication of the building block. This replication would work provided each of the repeated units have the strength required to bear the series force load. The serial replication is shown in Fig. 3-9(C) and its corresponding spring mass model in Fig. 3-9(c).

A generic design topology for a $Z\theta_X\theta_Y$ mechanism is depicted in Figure 3-10 for achieving large load-capacity. This concept is generated through parallel and series combination of flexural building blocks, as follows. First, load-sharing is introduced through multiple parallel beams supporting the central mass. As discussed in Section 3.3.2, a parallel combination of the flexural constraints can be obtained in two ways, azimuthally and transversely to the beam bending neutral plane. For m azimuthal replications, as shown in the plan view of the Figure 3-10, the load transferred to each beam is reduced by a factor of m, and hence the maximum stress. The

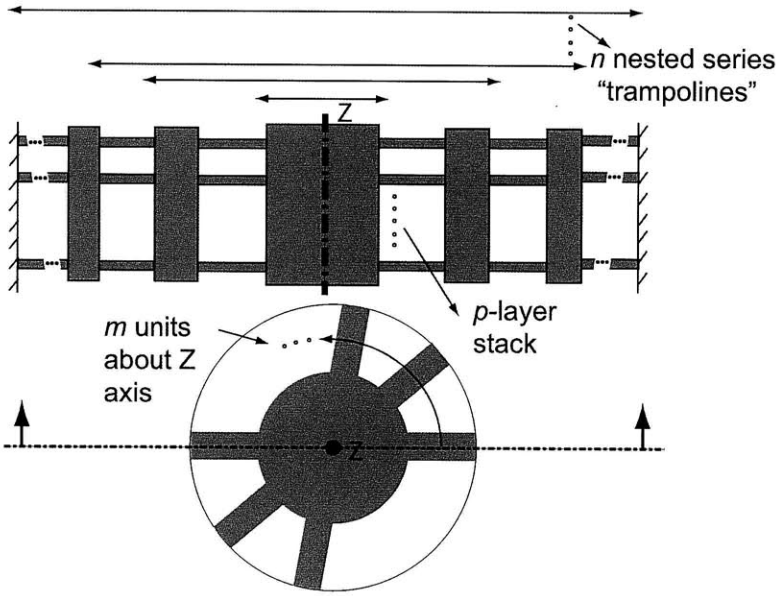


Figure 3-10: Proposed concept for large load-capacity. A plan view of the central ($n = 1$) trampoline is also shown. Serial combination of folded-back flexures is also possible, and is preferred for reducing over-constraint in the design.

resulting configuration is that of a diaphragm flexure.

For p transversely stacked layers, the load is further reduced for each beam by a factor of p.

However, these design changes aimed at an increase in load-capacity result in a over-constrained arrangement. In order to reduce the effect of the over-constraint, a series combination of the flexures is suggested. The resulting reduction in the stiffness allows for an increase in the range. Note that a series combination of diaphragm flexures results in a nested "trampoline" configuration. If n denotes the total number of trampolines, as shown in Figure 3-10, the overall stiffness and the range for the new design are as indicated in the last row of Table 3.2.

Existing design topology concepts can be explained with the two parallel stacking indices: azimuthal stacking index m, transverse stacking index p, and the serial stacking index n. For example, that the planar compound double parallelogram flexure, one of the designs marked in Fig. refone-sided, has the index values of $m = 1$ (since it is planar and that means only one azimuthal replication), $n = 2$ (folded back

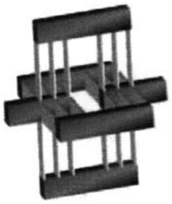


Figure 3-11: A spatial compliant translational joint presented in [47] fits as a special case of our generic design concept. This design corresponds to parallel stacking indices of $m = 2$ and $p = 3$, and serial stacking index of $n = 2$.

structure leads to a serial stacking), and $p = 2$ (each flexure unit within a fold has two beams in it).

The spatial compliant translational joint of [47] shown below in Fig. 3-11 is a special case of the generic design concept with parallel stacking indices of $m = 2$ and $p = 3$, and serial stacking index of $n = 2$. Since the flexure beams are arranged azimuthally with a replication along two directions we have $m = 2$. In each flexure unit, there are three flexure beams, which makes $p = 3$. For reducing stiffness and increasing range along the axis of motion, note that the joint has a serial replication of the flexural beam modules, which results in $n = 2$.

Topology Selection

For topology selection, we present here a non-dimensional approach to capturing design trade-offs all in one single graph. We use this graph to evaluate a design as well as compare different designs. First, we non-dimensionalize the performance requirements of interest: load-capacity f, stiffness k, and range δ of the proposed generic building block. For this we use the following normalizations:

$$f = \frac{F_{max}}{2BL\sigma_a/3} \tag{3.1}$$

$$k = \frac{K}{2E_0B} \tag{3.2}$$

$$\delta = \frac{\Delta_{max}}{\sigma_a L/(3E_0)} \tag{3.3}$$

For a non-dimensional flexure thickness $h = \frac{H}{L}$, the performance relations specified for the block reduce to the following form:

$$f = pmh^2 \tag{3.4}$$

$$k = \frac{pm}{n}h^3 \tag{3.5}$$

$$\delta = \frac{n}{h} \tag{3.6}$$

The f-k-δ design profiles for the non-dimensional thickness $h = \frac{H}{L}$ varied in a typical range $0.05 \leq h \leq 0.1$ are shown in Figure 3-12 for a sample set of values for m, n, and p.

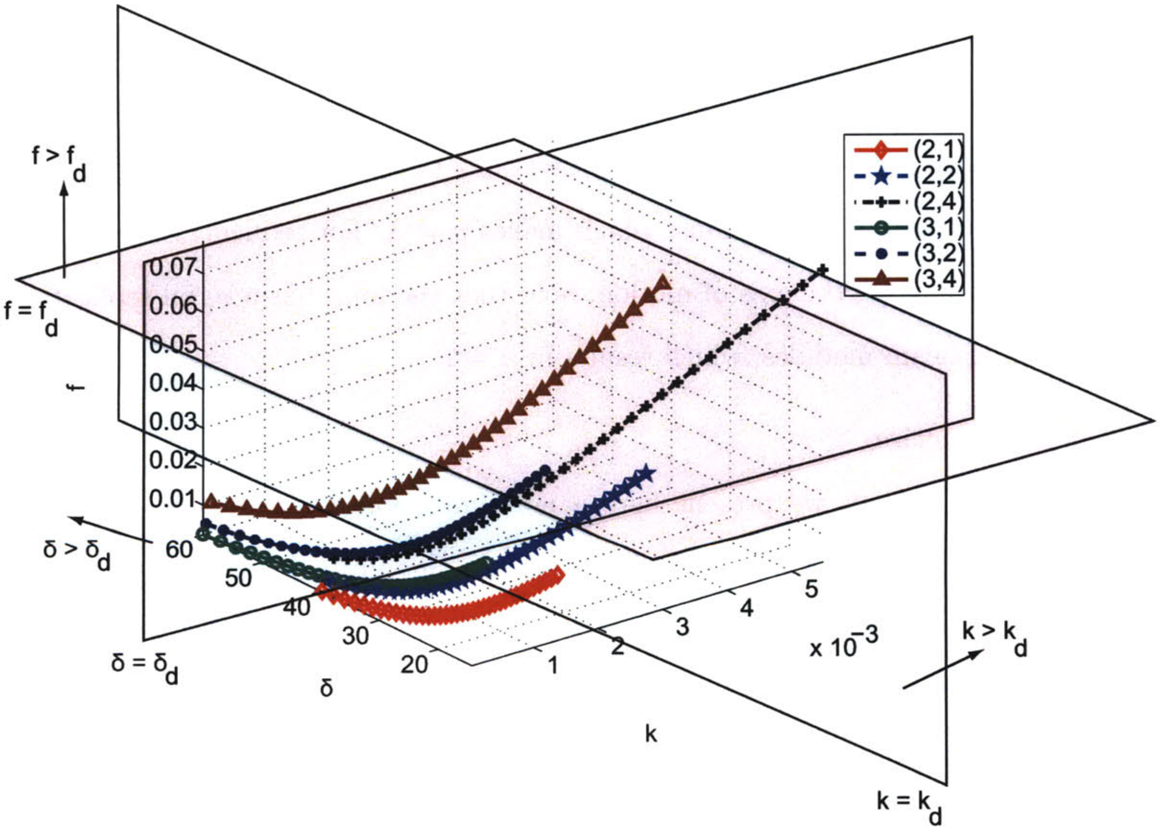


Figure 3-12: Non-dimensional performance space plotted as a function of non-dimensional thickness h for a sample set of $m = 1$, and (n,p) values as indicated.

A design methodology for generating strength-based topologies is proposed with the following steps:

- Specify the desired performance requirements F_{max}, K, Δ_{max} for the application of interest.
- Is the material choice already made? If not, select a material with large σ_a/E, for e.g. Ti-6Al-4V alloy [43].
- Assume a length L and width B for the beams. (Note that this step is assumed for simplicity. A more rigorous approach will iterate on these choices.)
- Find non-dimensional desired requirements: force f_d from F_{max}, δ_d from Δ_{max}, and k_d from K using the normalization relations given above.
- From the graph, view the $f = f_d$ planar section of the non-dimensional performance space.
- Select designs whose m, n, p values are such that the profile intersections in the $f = f_d$ plane have (δ, k) values in the quadrant $\delta > \delta_d, k > k_d$.
- If more than one design outcome is acceptable, use weighted criteria, or other criteria (such as thermal sensitivity, error motions), to obtain an optimal solution in the performance space. Using Eq. (3.6) estimate the corresponding parameter h in the input space. If no acceptable design exists, the steps need to be repeated for either a different L, B, or one of the stiffness or range constraints need to be relaxed. In order to compare two design concepts, we just need to examine the $f = f_d$ plane, and compare the δ and k values of the designs. Any series/parallel combination of flexure building blocks fits in this non-dimensional graph. Building on the treatment presented here, the performance space can be captured including the effects of other variables such as length and width of the beams, and material constants as well.

3.4.3 Example

A flexural motion stage for pure 1DOF vertical guidance for the following specifications is needed to meet the following requirements: vertical height $\leq 4\ in$, load-capacity of $100\ N$, range of at least $50\ \mu m$, and minimal rotational errors. Al 6061 was chosen as a suitable material for this case due to its ease of machining with a conventional abrasive waterjet. For a maximum allowable stress of $15\ MPa$, flexures of length $L = 1.25$", width $B \approx 0.8$", we find $f_d = 0.015$, $\delta_d = 31.5$. Of all the intersections of f_d plane with the performance space profiles of Figure 3-12, the design with $(m,\ n,\ p)$ values of $(1, 2, 4)$ is selected since it gives the largest K value for $\delta > \delta_d$.

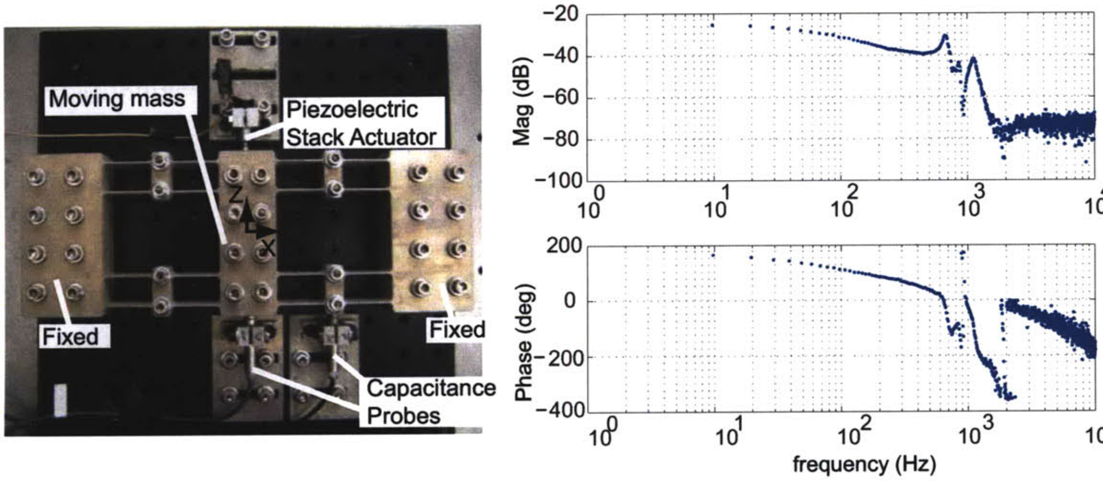


Figure 3-13: Large load-capacity purely 1DOF stage. Frequency response transfer function recorded between a piezoelectric actuator input and capacitance probe output is as plotted.

A design embodiment for the above concept is shown in Figure 3-13, with the double beam units on the top and bottom spaced wide apart so as to reduce the rotational compliance. This design achieves purely $1 - DOF$ motion similar to the design of [45], but has the added benefit of design for adequate range at large loads. The collocated frequency response recorded for the system with a custom-built PRBS system identification program on a dSPACE DS1103 platform shows a fundamental resonance around about 650 Hz. This value is higher than expected, because of the mechanical stiffness of the piezoelectric stack actuator adding in parallel with the mechanism. An improvement on this design embodiment could be one which uses compound double parallelogram flexures to remove the over-constraint imposed in a fixed-fixed beam arrangement.

3.5 Topologies for Modal Performance

In this section, we motivate design topologies which have modal decoupling and enough separation between slow and fast modes. At a topology generation phase, it is infeasible to accurately calculate the modal separation. Many assumptions have to be made in terms of geometry, material choice, and so on. However, after a topology is selected and the geometry, materials have been decided, the modal separation can in fact be predicted or used in the formulation at a shape and size optimization stage. But for generating a number of design topologies, it suffices to know the possible modal separation given the choice of placement of a redundant constraint. As a performance metric in this context, we define a modal separation index. We base our topology changes on this metric in Chapter 6.

3.5.1 Motivation

Flexure-based mechanisms are often affected by parasitic errors and cross-coupling errors. For example, for an XY flexural stage with intended degree of freedom of x and y, the parasitic degrees of freedom are Z, Θ_Z, Θ_X, Θ_Y. Parasitic error is, for example, a yaw Θ_Z error when the stage is driven with a voltage. Cross-coupling error on the other hand is between the desired degrees of freedom, i.e. the error in Y when X is actuated, and vice versa.

In order to minimize parasitic errors and cross-coupling errors, it is critical to design the flexure-based mechanism to be as much statically and dynamically decoupled as possible. Further, sometimes it is important to design the mechanism such that there is wide modal separation between the desired motions. Separating modes is critical so as to minimize energy transfer or leakage because of the invariable coupling from manufacturing errors.

In the case of the AFM nanopositioner, for example, the vertical (Z) dynamics need to be an order of magnitude faster than the lateral (X, Y). Achieving modes that are far apart needs to be input to the topology generation. For this purpose, we introduce an approximate metric below.

3.5.2 Method

Let us consider constraining a rigid body of mass m by attaching a flexure oriented along any direction in 3-dimensional space. A schematic diagram of the rigid body with its cartesian axes and the flexure oriented along any direction, for example, $\hat{e}$, is shown in Fig. 3-14.

We assume that the rigid body has mass symmetry and we fixed a set of cartesian axes at its center. This assumption makes the inertia matrix diagonal, with components J_{xx}, J_{yy}, and J_{zz}. Note that this assumption is valid for most flexure-based designs, since the designer usually builds the topology around a central stag, which can be chosen to be a mass-symmetric rigid body.

We define the modal separation index metric S_{d_1,d_2} for the ratio of natural frequencies of the fundamental modes along the i^{th} and jth directions of total 6 directions possible:

$$S_{d_1,d_2} = \frac{\omega_{d1}}{\omega_{d2}} = \sqrt{\left(\frac{k_{\|}(\hat{e}.\hat{d}_1)^2 + k_{\perp}|\hat{e} \times \hat{d}_1|^2}{k_{\|}(\hat{e}.\hat{d}_2)^2 + k_{\perp}|\hat{e} \times \hat{d}_2|^2}\right)\frac{M_{d2}}{M_{d1}}} \tag{3.7}$$

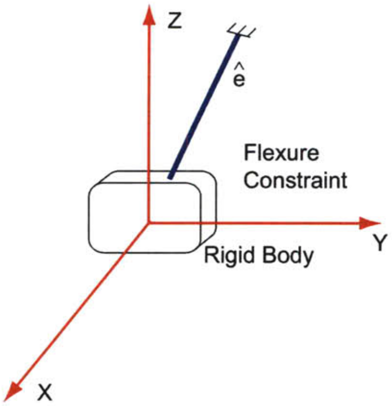


Figure 3-14: Schematic diagram illustrating principle of a modal separation index for a rigid body of mass m and moments of inertia J_{XX}, J_{YY}, J_{ZZ}.

where M_{d1}, M_{d2} denote the lumped mass associated with the kinetic energy stored due to motion along the d_1^{th} or d_2^{th} direction, $k_{\|}$ is the longitudinal stiffness of the flexure, and $k_{\perp}$ is the transverse stiffness of the flexure. The terms $\hat{e}.\hat{d_1}$ and $\hat{e}.\hat{d_2}$ are the direction cosines of the vector $\hat{e}$ along the directions $\hat{d_1}$ and $\hat{d_2}$, respectively.

For example, if d_1 is along the cartesian unit vector $\hat{\imath}$, and d_2 is along the the cartesian unit vector $\hat{\jmath}$, we select $M_{d1} = m$ and $M_{d2} = m$. If d_1 is along the cartesian $\hat{\imath}$, and d_2 is along the unit vector Θ_X representing an angular motion, $M_{d1} = m$, $M_{d2} = J_{XX}/r^2$, where r is the arm radius at which the flexure constraints are placed from X axis.

3.5.3 Example

To illustrate this with an example, consider the case when d_1 is along $\hat{\imath}$ and d_2 is along $\hat{\jmath}$, which leads to the following theorem.

Theorem:

Given directions $d_1 = \hat{\imath}$, $d_2 = \hat{\jmath}$, and stiffnesses $k_{\|}$ and $k_{\perp}$ for the flexure, the direction $\hat{e}$ along which the flexure constraint should be placed to maximize $S_{i,j}$ is $\hat{e} = \hat{\imath}$, and the maximum modal separation obtained is

$$max\ S_{i,j} = \sqrt{\frac{k_{\parallel}}{k_{\perp}}} \tag{3.8}$$

Proof:

We first resolve the flexure constraint direction $\hat{e}$ along the cartesian axes in terms of directional cosines:

$$\hat{e} = cos\theta_x\hat{\imath} + cos\theta_y\hat{\jmath} + cos\theta_z\hat{k} \tag{3.9}$$

Noting that $\hat{e}.\hat{\imath} = cos\theta_x$ and $|\hat{e}\times\hat{\imath}| = cos^2\theta_y + cos^2\theta_z$, and $cos^2\theta_x + cos^2\theta_y + cos^2\theta_z = 1$, and $M_i = M_j = m$, the modal separation index defined in Eq. (3.10) can be shown to be:

$$S_{i,j} = \sqrt{\left(\frac{k_{\parallel}cos^2\theta_x + k_{\perp}sin^2\theta_x}{k_{\parallel}cos^2\theta_y + k_{\perp}sin^2\theta_y}\right)} \tag{3.10}$$

For $k_{\parallel} > k_{\perp}$, this ratio is maximized when θ_x=0, and $\theta_y = \pi/2$, which proves the result that $\hat{e}$ should be aligned along $\hat{\imath}$.

Based on the above analysis, two general guidelines can be stated for deriving flexure topologies that maximize modal separation:

(a) Implement a flexure constraint in axial loading for obtaining a stiffer structure and hence allow for obtaining (i) high natural frequency and (ii) lower parasitic errors) for motions along the constraint direction.

(b) By the rule of reciprocity in design [14], implement a flexure constraint in bending for obtaining a more compliant structure and hence allow for obtaining (i) low natural frequency and (ii) large range for motions along the constraint direction.

As an example illustration of the above guidelines, consider the design topologies in Fig. 3-16(d) and (e). In Fig. 3-16(d) the flexure constraints are placed along the tangents to the sphere circumscribing the central stage. This makes the design much stiffer in Θ_X, Θ_Y. On the other hand, in Fig. 3-16(e), the flexure constraints are placed along the radii of the sphere circumscribing the central stage. This makes the design much more compliant in Θ_X, Θ_Y.

3.6 Examples: Topology Library

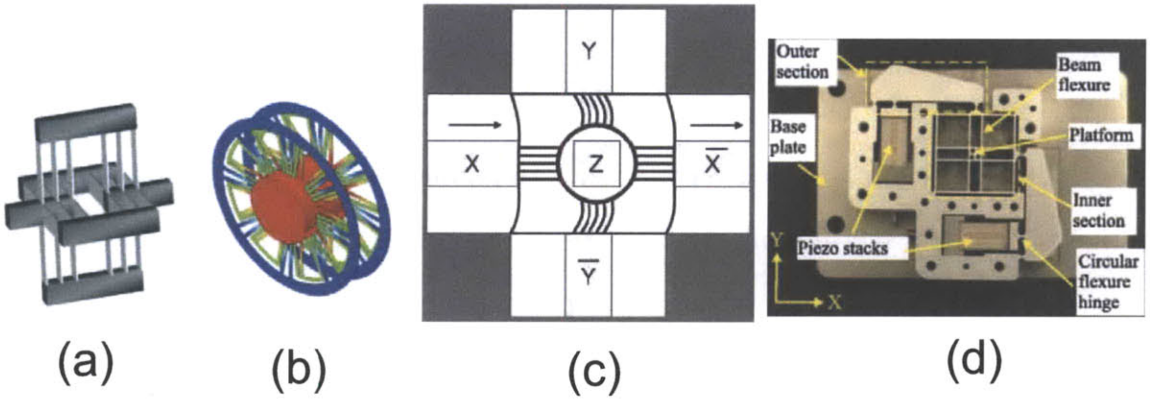


Figure 3-15: Examples of existing designs that fall within the proposed framework. (a) Large displacement compliant joint [47]. (b) A diaphragm flexure with minimal parasitic rotations [45]. (c) Flexure scanner for atomic force microscopy [41]. (d) Flexure scanner with lever amplifier [66].

3.6.1 Existing Design Topologies

A set of designs from the literature that can be derived using our method for specific values of strength metrics (parallel/serial redundancies) and/or dynamic performance metrics (modal separation index) are shown in Fig. 3-15.

A large-displacement compliant joint presented in [47] is shown in Fig. 3-15(a). The underlying design topology corresponds to parallel replications: azimuthal replication $m = 2$, transverse replication $p = 3$, and serial replication $n = 2$.

A diaphragm flexure with minimal parasitic rotations presented by [45] is shown in Fig. 3-15(b). The underlying topology corresponds to parallel replications: azimuthal replication $m = 8$, transverse replication $p = 2$, and serial replication $n = 1$.

A flexure scanner for atomic force microscopy presented by [41] is shown in Fig. 3-15(c). The underlying topology corresponds to parallel replications: azimuthal replication $m = 1$, transverse replication $p = 6$, and increased modal separation $S_{X,Y}$.

A flexure scanner with lever amplifier presented by [66] is shown in Fig. 3-15(d). The underlying topology corresponds to parallel replications: azimuthal replication $m = 1$, transverse replication $p = 4$, and increased modal separation $S_{X,Y}$.

3.6.2 Novel Design Topologies

A sample set of design topologies for multi-DOF flexure systems that can be derived from the proposed topology generation method are shown in Fig 3-16. An explanation of the specific metric (parallel/serial redundancies) and/or dynamic performance metrics (modal separation index) is as follows,

The topology concept shown in Fig. 3-16 is a $XY\theta_Z$ is a high-strength nanopositioner that can withstand large in-plane loads. The fundamental building block, shown in red, is a redundant combination of primitives. Its evolution based on strength criteria is as follows. Consider the design topologies of Fig. 3-17 with their fundamental building blocks all highlighted in red. The topology progression of the

Table 3.4: Stacking indices p and n for flexure building blocks highlighted in red for the topology concepts of Fig. 3-17.

Building Block	p_X (Parallel	n_X (Serial	p_Y (Parallel	n_Y (Serial
	along X)	along X)	along Y)	along Y)
Fig. 3-17(a)	1	1	1	1
Fig. 3-17(b)	2	1	1	1
Fig. 3-17(c)	2	2	2	1
Fig. 3-17(d)	2	2	2	2

fundamental building block from the form in Fig. 3-17(a) to that in Fig. 3-17(d) is as follows.

First, recall that we introduced the building block highlighted in red in Fig. 3-17(a) previously in Figs. 3-5(c). In the context of building blocks populating desired directions and allowing for multi-DOF kinematics, we have seen the topology of Fig. 3-17(a) previously in Fig, 3-7.

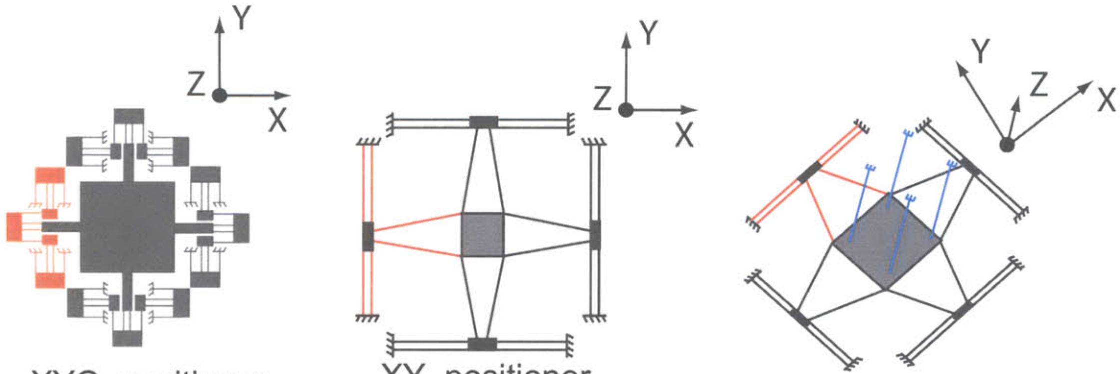
The stacking indices for parallel (m and p) and serial n replications for the building blocks of topologies of Fig. 3-17 are shown in Table 3.4. We can see that as large load-capacities or higher fatigue strength are desired, we need to have parallel replications p for load-sharing. For relieving the increased stiffness, we go for a folded back configuration to implement a series replication. Since $n = 2$, and $p = 2$ along both X and Y, the topology in Fig. 3-17(d) has the highest strength. A topology selection procedure using the non-dimensional force-displacement-stiffness relations of Section 3.4 can be used to select the topolgoy that is likely to work best for a given design problem.

The topology shown in Fig. 3-16(b) is similar to the topology of Fig. 3-17(b) but with two beams instead of just one for connecting the building block to the central rigid body. The two beams of the building block (highlighted in red) that are connected to the rigid body are spaced close to tangents to a circle circumscribing the rigid body. In this way, yaw error motions of the rigid body are minimized. In this topology, the modal separation between modes in X, Y and Θ_Z can be tuned as a function of the angle at which the beam constraints are arranged. A topology using flexural constraints in Z direction is shown in Fig. 3-16(c). This design improves on the modal separation index between X, Y and Z.

As explained in Section 3.5, in the topology of Fig. 3-16(d), the flexure constraints are placed along the tangents to the sphere circumscribing the central stage in Fig. 3-16(d). This makes the design much stiffer in Θ_X, Θ_Y. Given the geometry, the angles of the beams can be selected such that the modal separation between Θ_X and Z or Θ_Y and Z can be maximized. index. On the other hand, in Fig. 3-16(e), the flexure constraints are placed along the radii of the sphere circumscribing the central stage. This makes the design much more compliant in Θ_X, Θ_Y.

3.7 Summary

In this chapter, we examined the difference between topology selection and shape and size selection. Topology generation is a key step in the design process. We presented a framework in which a set of design primitives are converted to building blocks that are better suited to meet a performance requirement, such as fatigue behavior, or separation between vibration modes. We introduced key indices that can distniguish one topology from another. These indices are m for parallel azimuthal stacking, p for parallel transverse stacking, and n for serial stacking, and the modal separation indices S_{d_1,d_2} for any two directions d_1 and d_2, each selected from a total six available for the generalized coordinates of motion of the rigid body. We showed that strength-based topologies need not necessarily have to the stiffest. Examples from the literature can be explained in terms of the proposed method and were presented to illustrate how they can follow from the topology generation methods proposed here for strength and



$XY\Theta_z$ positioner

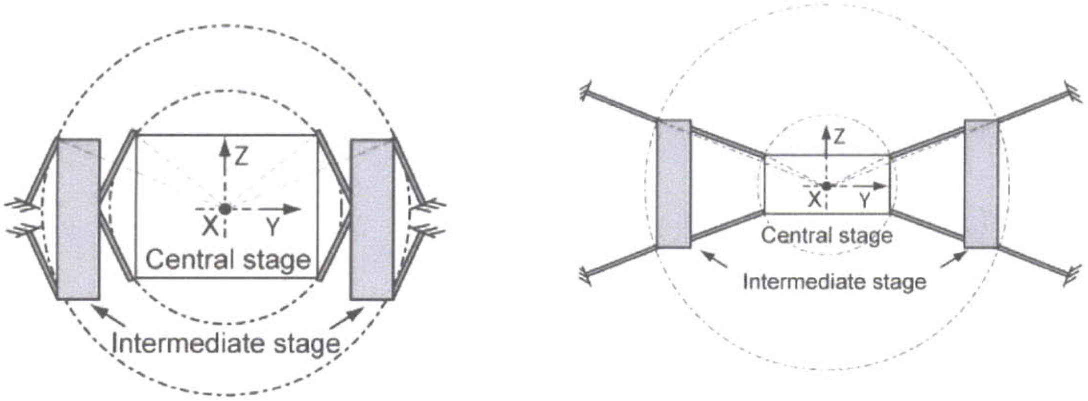
(a)

XY positioner

(b)

XYZ positioner

(c)



Purely-Z positioner

(d)

$\Theta_x\Theta_y$ aligner

(e)

Figure 3-16: Instances of novel design concepts from design library generated using the proposed framework.

dynamic performance. We will be using the topology generation method to examine candidate topologies for the AFM nanopositioner in Chapter 6.

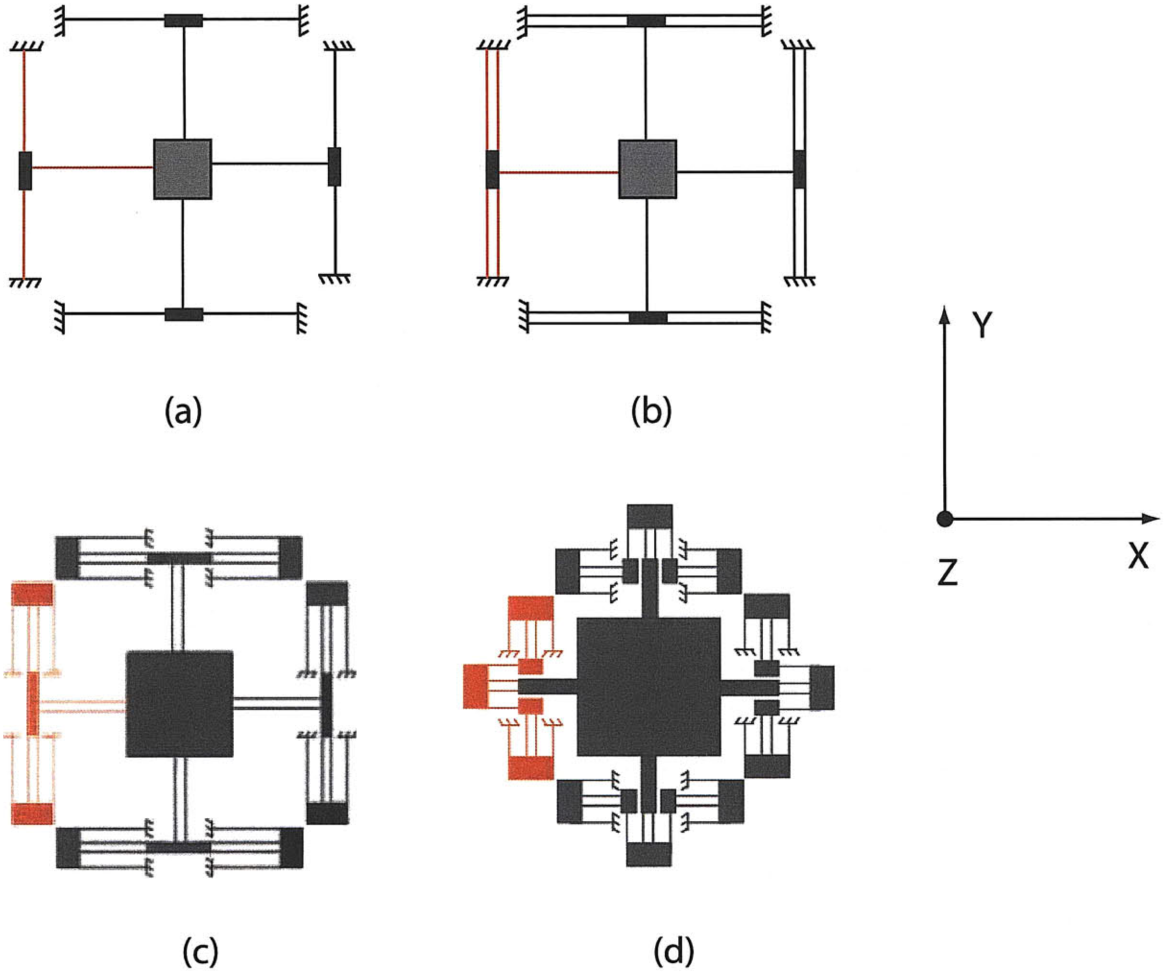


Figure 3-17: Evolution of design topologies explaining how the design (a) of Fig. 3-16 were derived using topology generation method.

Chapter 4

Controller Generation

In this chapter, we propose a control framework within which a control objective can be formulated to meet any of the performance requirements listed in Section 2.2. A key feature of this framework is a novel controller parameterization that allows for tuning the sensitivity transfer function directly with a filter. Tuning the sensitivity transfer function is critical for achieving a desired degree of robustness to parameter variations and modeling uncertainties, or for mitigating the effects of disturbances acting on a lightly damped structure. The proposed parameterization allows for tapping in to the vast knowledge base of filter design or model-matching for achieving a desired sensitivity transfer function. We present an illustrative example highlighting this approach.

4.1 Background and Motivation

Consider the control block diagram shown in the block diagram of Fig. 4-1. As a reference command $r(t)$ drives the system, the goal for the controller $C(s)$ is to make the output $y(t)$ of the plant $P(s)$ track $r(t)$ in the presence of disturbances $d(t)$ and noise $n(t)$.

Denoting the Laplace transform of any signal $h(t)$ by $H(s)$, we obtain the following relation for the output:

$$Y(s) = \frac{P(s)C(s)}{1+P(s)C(s)}R(s) - \frac{P(s)C(s)}{1+P(s)C(s)}N(s) + \frac{1}{1+P(s)C(s)}D(s) \tag{4.1}$$

Using the following definitions for the loop transmission $L(s)$, closed-loop system sensitivity transfer function $S(s)$, and complementary sensitivity transfer function $T(s)$:

$$\begin{aligned} L(s) &\triangleq PC(s) && (4.2) \\ S(s) &\triangleq \frac{1}{1+L(s)} = \frac{1}{1+P(s)C(s)} && (4.3) \\ T(s) &\triangleq \frac{L(s)}{1+L(s)} = \frac{P(s)C(s)}{1+P(s)C(s)} && (4.4) \end{aligned}$$

the output equation can be rewritten as:

$$Y(s) = T(s)R(s) - T(s)N(s) + S(s)D(s) \tag{4.5}$$

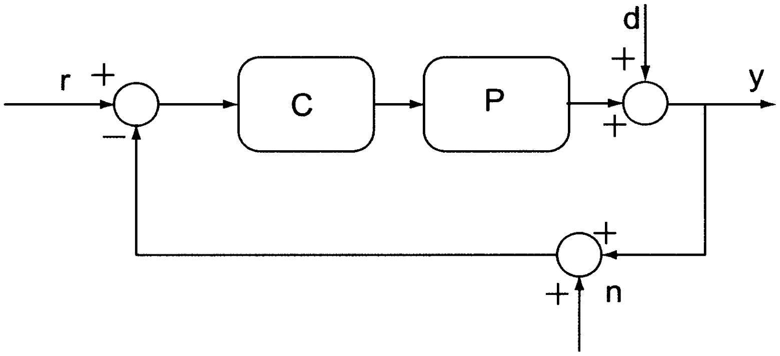


Figure 4-1: control system block diagram showing plant P and controller C.

The objectives of control system design primarily in two categories: robust stability and performance. Since most flexure-based nanopositioners are open-loop stable, our focus is on controllers that allow for maintaining stability even in the presence of parametric and modeling uncertainties. Parametric uncertainties include uncertainties in mass, stiffness, and damping values of modes of interest of the system.

In particular, for flexure-based mechanisms, active control for achieving adequate damping for modes in the frequencies of interest is desired. Physical damping is low in flexures made from metals such as aluminium (used in development stages of the design process for ease of machining), or titanium (used in the implementation and testing phase because of its high fatigue strength and other material properties). External damping such as squeeze film damping and foam-damping have been suggested and explored for flexures in [62]. Active damping through appropriate selection of control strategies can hence tackle the lightly damped resonances in these structures. Since the level of damping in an assembled mechanism is hard to predict before the fabricated product is available for testing, it becomes necessary to iterate the design process with thorough system identification and testing of hardware mechanism implementations. The control design must be robust to uncertainties in parameter estimation, especially damping as predicted from experimentally measured frequency response functions.

4.2 Our Approach

In this section, we propose a control architecture that is specifically motivated by flexible systems, such as flexure-based nanopositioners. Flexure-based systems have unique control issues, such as unpredictable damping in them, sensitivity to modeling errors, and difficulties in tackling disturbances. In obtaining nanoscale resolutions for the motion of these mechanisms, it becomes critical to shape the sensitivity transfer function of the closed-loop system to a desired profile.

Given a plant with uncertainties as described above, a controller with a varying order can be set up to address multiple control objectives in a systematic manner. However, to begin with, we need to ensure that the controller has a basic structure, i.e. a fixed or nominal part that is suitable for the plant. These features could be the inclusion of, say, a free integrator in the controller to ensure a zero steady state error in the step response for a second-order spring mass damper system.

A nominal controller can hence be designed and tested for the nominal plant to set up a baseline in terms of performance. Given such a nominal controller and the resulting closed-loop system transfer functions, it is the task of the part that has the varying order to account for robustness. We refer to this part with varying order as the 'robustifying' controller. A control architecture based on these concepts is presented in the next section.

4.3 Proposed Framework

A control architecture with a baseline block and a robustifying block for the feedback controller is shown in Fig. 4-2. With a suitable parameterization, the robustifying controller generation can iterate over some parameter transfer function, that is free to vary, to generate a class of controllers that can tackle the uncertainty. In the synthesis of the parameter transfer function, we restrict our search to stabilizing controllers. At the end of the robustifying procedure, we are interested in checking what the order of the controller is, and how increasing the order may affect the performance.

Note that the controller finally has the critically needed fixed part (for example, a free integrator) and also the varying order part, the order of which the designer can specify for the required level of robustness. A direct implementation of an optimization routine like μ-synthesis or DK-iteration to generate robust controllers can generate relatively much high order controllers, that may not be easily implemented on hardware [32, 27]. Instead, we pose the control design problem as a sensitivity tuning problem. Sensitivity is a critical aspect of control development for flexible structure systems, including flexure-based nanopositioners. With this motivation for

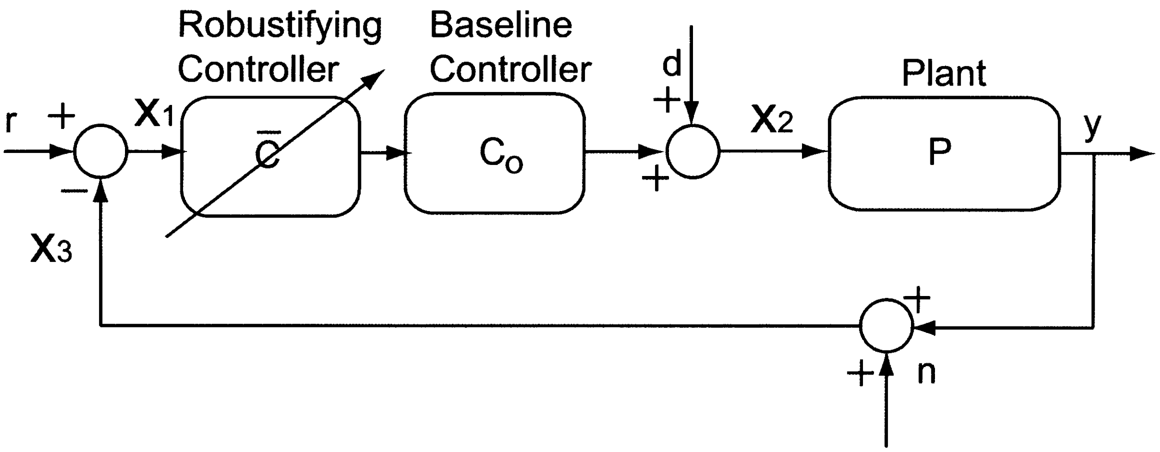


Figure 4-2: control system architecture with a baseline controller $C_o(s)$ and a varying-order controller $\bar{C}(s)$ resulting in a controller $C(s) \triangleq C_o(s)\bar{C}(s)$.

the controller generation, we proceed to a novel control parameterization in the next section, which facilitates a unique way to intuitively tune the sensitivity transfer function.

4.3.1 Novel Controller Parameterization

Existing controller parameterizations, such as the Youla parameterization [81], do not allow to shape the sensitivity transfer function directly. We present here a novel parameterization that preserves convexity and results in a direct design of the sensitivity transfer function. This parameterization is general and encompasses the Youla [81] and alternative to Youla parameterization presented in [85] as well.

We denote by $\mathcal{RH}_\infty$ the set of rational transfer functions that are stable (i.e. they have all poles in $Re(s) < 0$) and proper, i.e. they are bounded-input-bounded-output (BIBO) stable.

Let $P(s) \epsilon \mathcal{RH}_\infty$ represent the transfer function of a stable plant with a coprime factorization[1] $P(s) = \frac{N_p(s)}{M_p(s)}$ where $N_p(s), M_p(s) \epsilon \mathcal{RH}_\infty$.

Let $C_o(s)$ represent a transfer function of a stable controller with coprime fac-

---

[1] A factorization $P(s) = \frac{N_p(s)}{M_p(s)}$ of a plant $P(s)$ is coprime if there are no common right half plane pole-zero cancellations in $\frac{N_p(s)}{M_p(s)}$.

torization $C_o(s) = \frac{N_{C_o}(s)}{M_{C_o}(s)}$, where $N_{C_o}(s), M_{C_o}(s) \in \mathcal{RH}_\infty$. Let us assume that the parameterization $PCo(s) = \frac{N(s)}{M(s)}$ where

$$\begin{aligned} N(s) &\triangleq N_p(s)N_{C_o}(s) && (4.6)\\ M(s) &\triangleq M_p(s)M_{C_o}(s) && (4.7) \end{aligned}$$

is also a coprime factorization. The nominal loop transmission $L_o(s)$, closed-loop system nominal sensitivity $S_o(s)$ and nominal complementary sensitivity functions $T_o(s)$ for the closed-loop system are given by:

$$\begin{aligned} L_o(s) &\triangleq PC_0(s) && (4.8)\\ S_o(s) &\triangleq \frac{1}{1+L_o(s)} = \frac{1}{1+P(s)C_o(s)} && (4.9)\\ T_o(s) &\triangleq \frac{L_o(s)}{1+L_o(s)} = \frac{P(s)C(s)}{1+P(s)C_o(s)} && (4.10) \end{aligned}$$

Consider a parameterization given below for generating a varying-order controller $\bar{C}(s)$:

$$\bar{C}(s) = \frac{R_1(s) + R_2(s)\{\alpha + Q(s)\}}{R_3(s) + R_4(s)\{\alpha + Q(s)\}} \tag{4.11}$$

where $R_1(s), R_2(s), R_3(s), R_4(s), Q(s) \in \mathcal{RH}_\infty$ and α is a real constant. Here, as the order of $Q(s)$ increases, the order of the controller $\bar{C}(s)$ increases as well.[2]

Under this parameterization, the loop transmission $L(s)$, sensitivity $S(s)$ and

---

[2] For a set of selected component transfer functions, we will formulate the relative degree and order of the controller in Section 4.4.2.

complementary sensitivity $T(s)$ transfer functions are given as:

$$L(s) \triangleq P(s)C_o(s)\bar{C}(s) = \frac{N}{M}\frac{R_1(s) + R_2(s)\{\alpha + Q(s)\}}{R_3(s) + R_4(s)\{\alpha + Q(s)\}} \tag{4.12}$$

$$S(s) \triangleq \frac{1}{1+L(s)} = \frac{M(s)\,[R_3(s) + R_4(s)\{\alpha + Q(s)\}]}{\{M(s)R_3(s) + N(s)R_1(s)\} + \{M(s)R_4(s) + N(s)R_2(s)\}\{\alpha + Q(s)\}} \tag{4.13}$$

$$T(s) \triangleq \frac{L(s)}{1+L(s)} = \frac{N(s)\,[R_1(s) + R_2(s)\{\alpha + Q(s)\}]}{\{M(s)R_3(s) + N(s)R_1(s)\} + \{M(s)R_4(s) + N(s)R_2(s)\}\{\alpha + Q(s)\}} \tag{4.14}$$

We have noted that the nominal controller $C_o(s)$ results in the feedback system being internally stable. Now, we need to check if with the inclusion of the robustifying controller $\bar{C}(s)$, the resulting controller $C(s) \triangleq C_o(s)\bar{C}(s)$ will ensure that the feedback system remains internally stable. The following theorem pertains to the internal stability with the robustifying controller placed in series with the nominal controller.

4.3.2 Theorem: Internal Stability

For the plant $P(s)$ and nominal controller $C_o(s)$ that results in internal stability of the closed-loop system, and their respective coprime factorizations in Eqs. (4.6, 4.7), and the robustifying controller $\bar{C}(s)$ defined in Eq. (4.11) for transfer functions $Q(s)$, R_1, R_2, R_3, and $R_4 \in \mathcal{RH}_\infty$, the feedback system is internally stable if the following conditions hold:

(i) $M(s)R_4(s) + N(s)R_2(s) = 0$

(ii) $M(s)R_3(s) + N(s)R_1(s)$ **is a minimum phase transfer function.**

Proof:

From Eqs. (4.13, 4.14), we can show that the closed loop characteristic equation for the system is given by:

$$M(s)R_3(s) + N(s)R_1(s) + \{M(s)R_4(s) + N(s)R_2(s)\}(\alpha + Q(s)) = 0 \qquad (4.15)$$

Using the condition $M(s)R_4(s)+N(s)R_2(s) = 0$ reduces the closed loop characteristic polynomial to $M(s)R_3(s) + N(s)R_1(s)$. To ensure the feedback system is internally stable we consider transfer function matrix between the internal signals x_1, x_2, and x_3 and the exogenous signals r, d, and n in Fig. 4-2. As presented in [81], internal stability can be verified by checking the stability of the nine transfer functions in the transfer matrix **A** given below:

$$\left\{\begin{matrix} x_1 \\ x_2 \\ x_3 \end{matrix}\right\} = \frac{1}{MR_3 + NR_1} \times [A]_{3\times3} \left\{\begin{matrix} r \\ d \\ n \end{matrix}\right\} \qquad (4.16)$$

where

$$A = \begin{bmatrix} M_pM_{C_o}\{R_3 + R_4(\alpha + Q)\} & -N_pM_{C_o}\{R_3 + R_4(\alpha + Q)\} & -M_pM_{C_o}\{R_3 + R_4(\alpha + Q)\} \\ M_pN_{C_o}\{R_1 + R_2(\alpha + Q)\} & M_pM_{C_o}\{R_3 + R_4(\alpha + Q)\} & -M_pN_{C_o}\{R_1 + R_2(\alpha + Q)\} \\ N_pN_{C_o}\{R_1 + R_2(\alpha + Q)\} & N_pM_{C_o}\{R_3 + R_4(\alpha + Q)\} & M_pM_{C_o}\{R_3 + R_4(\alpha + Q)\} \end{bmatrix}$$

As the component transfer functions are all selected to be in $\mathcal{RH}_\infty$, each of the nine entries in A are stable. From Eq. (4.16), we see that the feedback system is internally stable, i.e. the internal signals x_1, x_2, and x_3 are bounded and stable, if the transfer function $\frac{1}{M(s)R_3(s)+N(s)R_1(s)}$ has no poles in the right half plane $Re(s) > 0$. Or in other words if $M(s)R_3(s) + N(s)R_1(s)$ has no zeros in the right half plane. Hence, if $M(s)R_3(s) + N(s)R_1(s)$ is a minimum phase transfer function, each of the nine transfer functions relating internal signals to exogenous signals is stable, and the system is internally stable. This completes the proof of the theorem.

Note that the condition $N(s)R_4(s) + M(s)R_2(s) = 0$ allows for the sensitivity and complementary sensitivity transfer functions in Eqs. (4.13, 4.14) to be convex in the parameter $Q(s)$. Convex optimization is one of the relatively easier ways to pose an optimization problem. The advantage of a convex optimization, for example, a convex minimization problem is that if a local minima exists it is also a global minima [86].

There are further advantages of the above formulation which become apparent by selecting appropriate values for the transfer functions $R_1(s)$, $R_2(s)$, $R_3(s)$, and $R_4(s)$. For instance, the closed-loop sensitivity transfer function can be shaped directly by the $Q(s)$ filter using a set of conditions, as stated below.

We compare here the closed-loop sensitivity obtained using the proposed parameterization and the conventional Youla parameterization in the following two lemmas.

Lemma 1

For the closed-loop system with conditions described above, if $\alpha_0, R_3(s) = 0$ and $R_4(s) = T_o(s)R_1(s)$, the closed-loop sensitivity transfer function $S(s)$, the complementary sensitivity transfer function $T(s)$, and the controller $C(s) \triangleq C_o(s)\bar{C}(s)$ are given by:

$$S(s) = S_o(s)Q(s) \tag{4.17}$$

$$T(s) = 1 - S_o(s)Q(s) \tag{4.18}$$

$$C(s) = \frac{1}{P(s)}\left\{\frac{1 - S_0(s)Q(s)}{S_0(s)Q(s)}\right\} \tag{4.19}$$

Proof:

Using $M(s)R_4(s)+N(s)R_2(s)=0$, $R_3(s)=0$, and $R_4(s)=T_o(s)R_1(s)$ in Eq. (4.13), the result $S(s)=S_o(s)Q(s)$ becomes evident. Using $S(s)+T(s)=1$, we arrive at the expression $T(s)=1-S_o(s)Q(s)$. The expression for the controller $C(s)$ is evident from the relation $T(s)=P(s)C(s)S(s)$ and the expressions for $S(s)$ and $T(s)$ in Eqs. (4.17)-(4.18).

Lemma 2

For the closed-loop system with conditions described above, if $R_4(s)=-N(s)$, $\alpha_0=0$, and $M(s)R_3(s)+N(s)R_1(s)=1$, the resulting controller parameterization is the Youla parameterization. The closed-loop sensitivity transfer function S, the complementary sensitivity transfer function T(s), and the controller $C(s)$ are given by:

$$S(s) = M(s)(R_3(s)-N(s)Q(s)) \tag{4.20}$$

$$T(s) = N(s)(R_1(s)+M(s)Q(s)) \tag{4.21}$$

$$C(s) = \frac{R_1(s)+M(s)Q(s)}{R_3(s)-N(s)Q(s)} \tag{4.22}$$

Proof:

Using $M(s)R_4(s)+N(s)R_2(s)=0$, $R_4(s)=-N(s)$, $\alpha_0=0$, and $M(s)R_3(s)+N(s)R_1(s)=1$ in Eq. (4.13), the result $S(M(s)(R_3(s)-N(s)Q(s))$ becomes evident. Using $S(s)+T(s)=1$, and $M(s)R_3(s)+N(s)R_1(s)=1$, we arrive at the result $T(s)=N(s)(R_1(s)+M(s)Q(s))$. The expression for the controller $C(s)$ is evident from Eq. (4.11) with $\alpha=0$, $M(s)R_4(s)+N(s)R_2(s)=0$,and $R_4(s)=-N(s)$.

4.4 Discussion

In the previous section, it was shown that with the proposed controller parameterization, the Eq. (4.17) has the free parameter $Q(s)$ just multiplying the nominal sensitivity. This makes the synthesis for $Q(s)$ a filter design or a model-matching problem, hence allowing us to tap into the knowledge base of these well-established areas of research.

In contrast, the expression derived for sensitivity in Eq. (4.20) using Youla parameterization has the parameter $Q(s)$ involved as an expression that does not directly allow for designing the sensitivity function. The dynamics of the sensitivity function are dependent on the phase relation between $R_3(s)$ and $N(s)Q(s)$, and the map between a desired sensitivity and the corresponding $Q(s)$ is dependent on the choice of $R_3(s)$, which is not unique.

In what follows, we will discuss some of the aspects involved with controller generation using the Eq. (4.17) for tuning the sensitivity transfer function.

4.4.1 Optimal Synthesis of $Q(s)$ parameter transfer function

The goal for the control design is to shape $S(s)$ according to a disturbance spectrum, or error following dynamics. There is no unique way to define a control cost function that can guide us in the search of $Q(s)$.

One approach is to use a shaping filter $W_s(s)$ that incorporate the disturbance or robustness specifications, and an optimization problem can be formulated as a bound on the ∞-norm of the product $W_s(s)S(s)$ as follows:

$$||W_s(j\omega)S(j\omega)||_\infty \leq 1 \tag{4.23}$$

We can also pose the problem of tuning the sensitivity transfer function as a model-matching problem. $Q(s)$ parameter transfer function can be searched over the class of stable transfer functions so that a chosen norm of the error between the frequency response of the sensitivity transfer function and that of the desired transfer function gets minimized. A formulation using a $\mathcal{H}_2$ norm for the error is as given below:

$$Q(s) = \arg \min_{Q(s)\epsilon\mathcal{RH}_\infty} ||S_d(s) - S_0(s)Q(s)||_2^2 \tag{4.24}$$

where $S_d(s)$ is a desired sensitivity transfer function.

In general, the filter $Q(s)$ can be constructed from a sum, or in some cases a product, of a set of basis filters. One possible (out of many combinations) is a construct showing a finite summation of stable basis filters in $\mathcal{RH}_\infty$ as shown below:

$$Q(s) = a_0 + \sum_{i=1}^{N} \frac{a_i}{s - p_i} \tag{4.25}$$

where N is a finite natural number, a_0 is a real number, coefficients a_i are constants to be selected and p_i are poles of the filters. In general, the poles can be on the negative real axis, i.e. $p_i < 0$, and their corresponding coefficients are real numbers. The poles can also occur as complex conjugate pairs, with $real\{p_i\} < 0$, and the corresponding coefficients are complex conjugate pairs as well so that the overall summation in Eq. (6.26) results in a transfer function with real polynomials in the numerator and denominator. Note that the larger that N is, i.e. more filters are included, the higher the order of the resultant $Q(s)$. Also, if $a_0 \neq 0$, the relative degree of $Q(s)$ as defined in Eq. (6.26) is 0, since after the summation is done the numerator and denominator polynomials are of the same order.

Another construct for $Q(s)$, presented in [20], is shown below:

$$Q(s) = \frac{b_o s^n + b_1 s^{n-1} + b_2 s^{n-2} + ...b_{n-1}s + b_n}{(s+p_1)(s+p_2)(s+p_3)....(s+p_{n-1})(s+p_n)} \tag{4.26}$$

where n is a finite natural number, the coefficients $b_1, b_2, b_3 ... b_n$ are real, and the poles at $s = -p_i$ (for $i = 1..N$) can be negative real, or occur as complex conjugate poles with negative real parts. Here again,the higher the n, higher the order of $Q(s)$. The relative degree of $Q(s)$ is 0, since the numerator and denominator polynomial are of the same order.

Note that the form of $Q(s)$ given in Eq. (6.27) reduces to the form in Eq. (6.26) by using a partial fraction expansion. In the rest of the thesis, we will be using either form, interchangeably.

In the representations of Eqs. (6.26) and (6.27), selecting the basis filters is a critical step. Often, the basis filters must be related to the physics of the system. For example, if the plant has a lightly damped resonance in its frequency response, it is likely to show up as a peak in the sensitivity transfer function. In that case, a notch filter is one possible choice to mitigate the effects of this resonance. From the relation $S(s) = S_0(s)Q(s)$, it is evident that a notch filter placed around the peak in the sensitivity transfer function, essentially reduces the height of the peak.[3]

4.4.2 Controller Generation: Order and Relative Degree

The controller transfer function for the case when we parameterize the controller such that the condition $S(s) = S_0(s)Q(s)$ holds is given in Eq. (4.19). The expression for the controller $C(s)$ can be rewritten in terms of the plant $P(s)$, nominal controller $C_0(s)$ and the parameter transfer function $Q(s)$ as given below:

$$C(s) \quad = \quad \frac{1}{P(s)} \left\{ \frac{1 + P(s)C_o(s)}{Q(s)} - 1 \right\} \tag{4.27}$$

We can work out the order of the controller and its relative degree, defined as the excess of the number of poles over the number of zeros of a transfer function, as follows.

---

[3] We work out an illustrative example in Section 6.7.

Using $F_{num}(s)$ and $F_{den}(s)$ to denote the numerator and denominator polynomials, respectively, of a transfer function $F(s)$, the expression for $C(s)$ becomes:

$$\begin{aligned} C(s) &= \frac{C_{num}(s)}{C_{den}(s)} \\ &= \frac{P_{den}(s)}{P_{num}(s)} \left\{ \frac{P_{den}(s)C_{oden}(s) + P_{num}(s)C_{onum}(s)}{P_{den}(s)C_{oden}(s)} \frac{Q_{den}(s)}{Q_{num}(s)} - 1 \right\} && (4.28) \\ &= \frac{P_{den}(s)}{P_{num}(s)} \left\{ \frac{P_{den}(s)C_{oden}(s)\{Q_{den}(s) - Q_{num}(s)\} + P_{num}(s)C_{onum}(s)Q_{den}(s)}{P_{den}(s)C_{oden}(s)Q_{num}(s)} \right\} && (4.29) \end{aligned}$$

Recall that a proper transfer function has atleast the same number of poles as the number of zeros, while a strictly proper transfer function has more number of poles than the number of zeros.

Let f_{num} and f_{den} denote the order of the numerator polynomial $F_{num}(s)$ and denominator polynomial $F_{den}(s)$, respectively. Under the following assumptions:
(a) the plant $P(s)$ is strictly proper with a relative degree of at least 2, as is the case for most mechanical flexible structure systems, i.e. $p_{num} \leq p_{den} - 2$
(b) the nominal controller $C_0(s)$ is strictly proper, i.e. $c_{num} < c_{den}$, and
(c) the parameter transfer function $Q(s)$ is proper, i.e. $q_{num} \leq q_{den}$,
we see that the relative degree of the transfer function

$$\frac{P_{den}(s)C_{oden}(s) + P_{num}(s)C_{onum}(s)}{P_{den}(s)C_{oden}(s)} \tag{4.30}$$

appearing in Eq. (4.29) is 0.

Further the order of the numerator polynomial $C_{num}(s)$ is given as:

$$c_{num} = 2p_{den} + c_{oden} + order\{Q_{den}(s) - Q_{num}(s)\} \tag{4.31}$$

and the order of the denominator polynomial $C_{den}(s)$ is given as:

$$c_{den} = p_{num} + p_{den} + c_{oden} + q_{num} \tag{4.32}$$

The order c_{den} of the denominator polynomial of $C(s)$ is the order of the controller. However, we need to first check if the controller is proper, so that it can be implemented. Note that the maximum order of the polynomial $Q_{den}(s) - Q_{num}(s)$ is q_{den}. The relative degree of $C(s)$ is defined as $c_{den} - c_{num}$ and is given by:

$$c_{den} - c_{num} = p_{num} - p_{den} + q_{num} - order\{Q_{den}(s) - Q_{num}(s)\} \tag{4.33}$$

If the relative degree of $C(s)$ is negative, we need to multiply the controller with a filter to make it strictly proper. For example, a low pass filter of order at least $c_{num} - c_{den} + 1$ can be included in the controller. Since such filters tend to cause a phase loss at the gain crossover frequency and hence reduce the phase margin, so the filter poles should be placed at frequencies at least 10 times higher than the gain crossover frequency of interest.

If c_{filter} is the order of the filter included to make the controller strictly proper, we see that the resulting controller order increases to:

$$c_{den} = p_{num} + p_{den} + c_{oden} + q_{num} + c_{filter} \tag{4.34}$$

s From Eq. (4.34), it is evident that the order of the controller varies with varying the order of the parameter transfer function $Q(s)$. The extent to which the controller order can increase is regulated by the choice for the order of $Q(s)$, which is in turn dictated by the specification on robustness. The benefit of this approach is that we can intuitively tune the performance, namely, the sensitivity transfer function, which is critical for flexible structure systems having lightly damped harmonics.

In tackling lightly damped resonances, we often desire control design to ensure a prescribed stability and damping, as depicted in Fig. 4-3. For this purpose, we can specify the desired sensitivity transfer function $S_d(s)$ as an inverse of the shaping filter $W_s(s)$, given by:

$$W_s(s) = 1 + (2cos\theta)\frac{\omega}{s} + \frac{\omega^2}{s^2} \tag{4.35}$$

The desired locations for the closed loop poles are shown as the hatched area in the left half s-plane, with damping more than $cos(\theta)$, and damped natural frequency greater than ω. A model-matching problem such as the one shown in Eq. (6.26) can be formulated to solve for $Q(s)$ that pushes the sensitivity close to the desired sensitivity. Note that usual pole placement algorithms are not robust with respect to modeling uncertainties. The proposed approach based on the novel parameterization presented in Section4.3.1 allows us to directly address shaping the sensitivity function, unlike existing approaches using Youla parameterization [84].

In summary, with this approach controllers of varying order can be generated with the parameter transfer function $Q(s)$, which is obtained from an optimal synthesis procedure that tunes the sensitivity transfer function. To summarize, the controller generation is carried out in the following steps:

Step 1: Select a nominal controller.

Step 2: Obtain the nominal sensitivity transfer function. From frequency response plots, note the limitations that need to be addressed. For example, does the sensitivity show a sharp peak, indicating poor robustness?

Step 3: Formulate the desired sensitivity transfer function that addresses the limitations of the nominal sensitivity transfer function.

Step 4: Formulate and solve the optimal synthesis problem for the parameter transfer

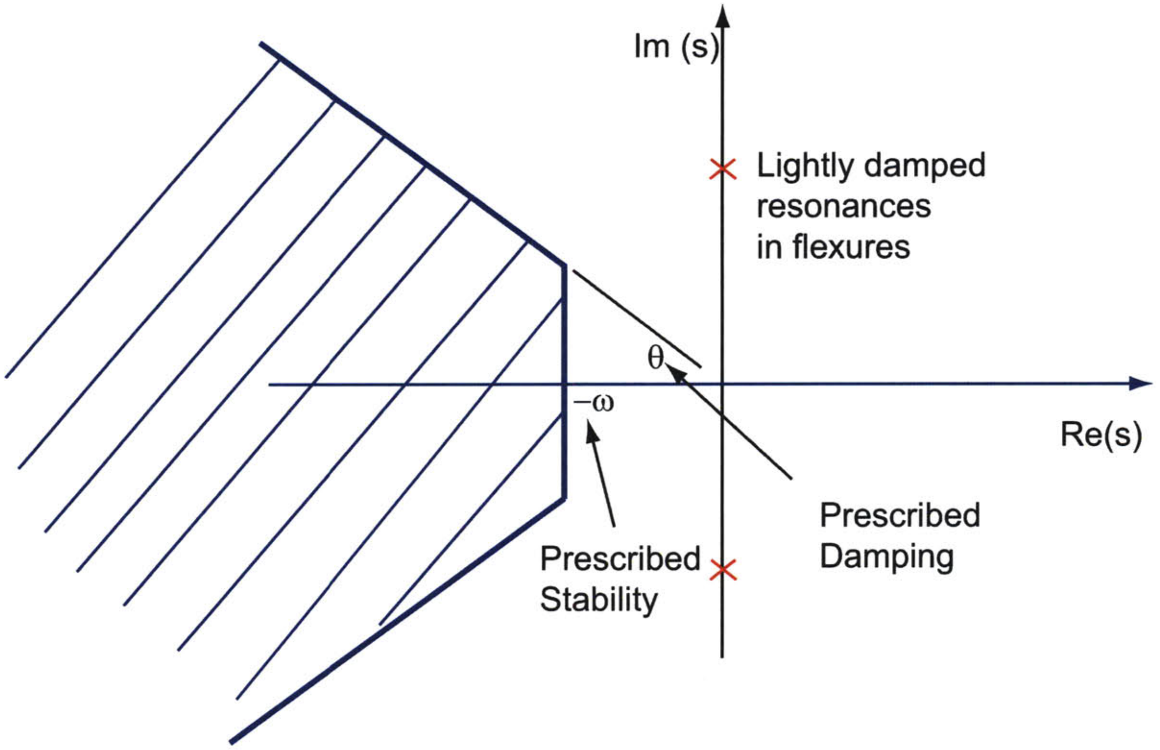


Figure 4-3: Synthesis of Q for prescribed level of damping and stability.

function $Q(s)$. Example approaches include model-matching and filter design.

Step-5: Determine the controller $C(s)$ resulting from $Q(s)$. Check for the controller being proper. Modify suitably with filter, and reduce its order, so that it can be implemented on real-time hardware control prototyping platforms. Compare in simulation the performance of the controller before and after the modifications.

4.5 Summary

In this chapter, we proposed a control architecture using a fixed order baseline controller and variable order robustifying controller. The controller generation is motivated from the application viewpoint of tuning sensitivity in flexible structure systems. A novel parameterization of the controller was proposed to address directly tuning the sensitivity with a parameter transfer function $Q(s)$. The relation $S(s) = S_o(s)Q(s)$

allows for application of filter design and model-matching techniques for the optimal synthesis of $Q(s)$. In this approach, the order of the controller generated varies depending on the choice of $Q(s)$, which can be selected based on the desired degree of robustness. An example model-matching approach for solving for $Q(s)$ is presented in Section 6.7 and a controller is generated and tested, illustrating the steps detailed in this chapter.

Chapter 5

Integrated Design and Control Methodology

In this chapter, we present an integrated design and control methodology tailored for flexure-based nanopositioning systems. First, we examine existing literature on integrated design and control in allied areas of research. We discuss the need for varying design topology and order of a controller in design and control optimization. We then propose an intgerated design and control methodology that utilizes the design topology generation described in Chapter 3 and the controller generation method described in Chapter 4, for meeting a desired set of performance requirements. A simple 1-DOF example is worked out to illustrate the steps involved in using this methodology. The details of optimization problem formulation and solutions for design and control are presented. The outcome of the exercise is a novel design topology, with it shape and size optimized, and a controller synthesized such that a desired control bandwidth and design requirements of strength and modal separation are met.

5.1 Background and Motivation

Integrated structure design and control has been an active area of research spanning diverse fields of aerospace systems, robotics, motion control, and chemical process control over the past few decades. A survey of some of the most relevant methods in the literature is as follows.

The optimal design and control of flexible structures has been studied for (i) improving a mass efficiency metric (defined as mass moved per unit work output) in [9], (ii) a quadratic control performance index in [15], and (iii) a weighted sum of structural mass and the energy of the controlled mechanism in [37]. The integrated design and control problem was formulated as a multi-objective optimization involving design and proportional-integral-derivative (PID) controller parameters in [42] for mechatronic systems. A similar approach optimizing proportional-derivative (PD) controller parameters and design parameters for four-bar linkages was studied in [29]. A non-linear optimization formulation including design costs and a robust performance constraint on the weighted sum of sensitivity and complementary sensitivity functions is considered for a chemical distillation column in [38]. Decentralized control techniques were used to solve for the optimization of passive (design parameters) and active (control parameters) for vibration isolation platforms in [36]. Different approaches for integrated design and control have been studied from an optimization theory standpoint in [27] and [25]. These approaches include (i)sequential optimization with design optimization followed by control optimization, (ii)simultaneous design and control optimization, and (iii)an iterative combination where the design is initially optimized without affecting the controller, then the controller is optimized, and such a cycle is iterated until performance requirements are met.

Motion stages developed using lead-screw drives were characterized for their dynamics and controlled with classical lead-lag compensators in [35]. In this reference, the design and control performance space in terms of performance requirements, such as (i) the positioning error and (ii) control bandwidth of the drive and (iii) the maximum acceleration of the carriage, were captured for the entire range of geometry, material, and other parameters. Since lightly damped harmonics hinder control performance, achieving robust passive damping with foam-based materials is proposed by the same research group in [62].

An integrated design and control methodology for high-speed control of robotic manipulators is presented in [32, 33]. Since unmodeled dynamics in the control bandwidth can adversely affect the performance, it is necessary to account for model-truncation errors in the design and control optimization. In this context, a constraint condition on the Hankel norm of the truncated modes is formulated in the optimization problem [32].

A few references address changing the design structure or configuration, i.e. the topology, so that control performance is enhanced. We examine here two specific cases from the literature that illustrate the importance of selecting an appropriate design topology before deploying any optimization routine.

Consider the example of a robotic system shown in Fig. 5-1 addressing the end-point control of a flexible link. The actuator is a rotary servomotor that generates a torque required for moving the end-point of the link. The feedback signal is the end-point position, which can be recorded by a sensor such as an accelerometer. Since the actuator and the sensor are not at the same location in space, i.e. the system is non-collocated. For the non-collocated system, the flexibility of the link is known to cause non-minimum phase zeros in the transfer function between the voltage applied to the motor and the measured end-point displacement [57].

In order to avoid the occurrence of the non-minimum phase zero, the actuation point shown in Fig. 5-1 (a) can be moved away from the motor closer to the end-point, as shown in Fig. 5-1(b). With the actuation location moved closer to the end-point, the portion of the link from the new actuation point to the sensor location is shorter,

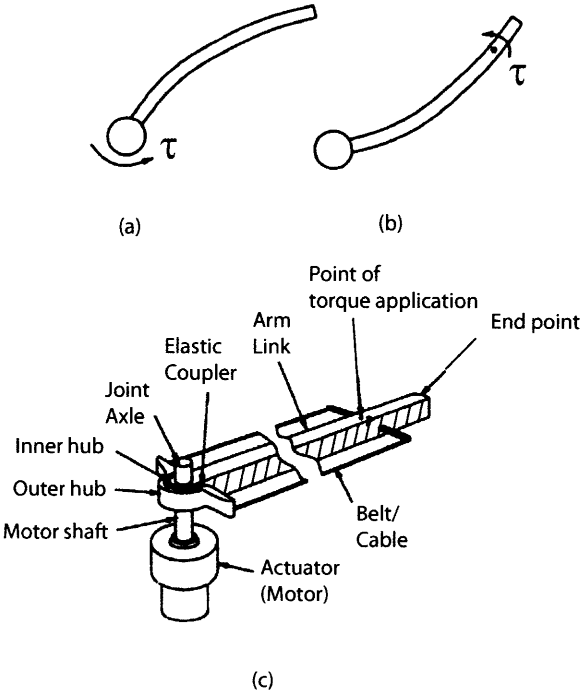


Figure 5-1: Design for control example from [57]. (a) Moving the torque application point away from the hub and closer to the end point of the flexible manipulator results in minimum-phase dynamics, and hence allows for higher control bandwidths. (b) A belt transmission is used on a motor to vary the location of the torque application point.

and hence stiffer. It is shown in [57] that, under certain geometry conditions, this topology change results in moving the zeros from the real-axis on to the imaginary axis, making the system minimum-phase. The design topology change shown in Fig. 5-1(b) is implemented in Fig. 5-1(c) using a cable transmission from the motor. Without this topology change, with the actuator just as the motor and sensor at the end-point, the system would be non-minimum phase and pose critical control challenges.[1]

We next consider the example of a hard disk drive actuator subsystem in Fig. 5-2. As shown in Fig. 5-2 (a), this subsystem positions the read (or write) head at the end of an arm pivoting about a rotary bearing. A lorentz-force F_m generated by voice coil motor at an distance R_e causes the arm to rotate about the pivot. However, the

---

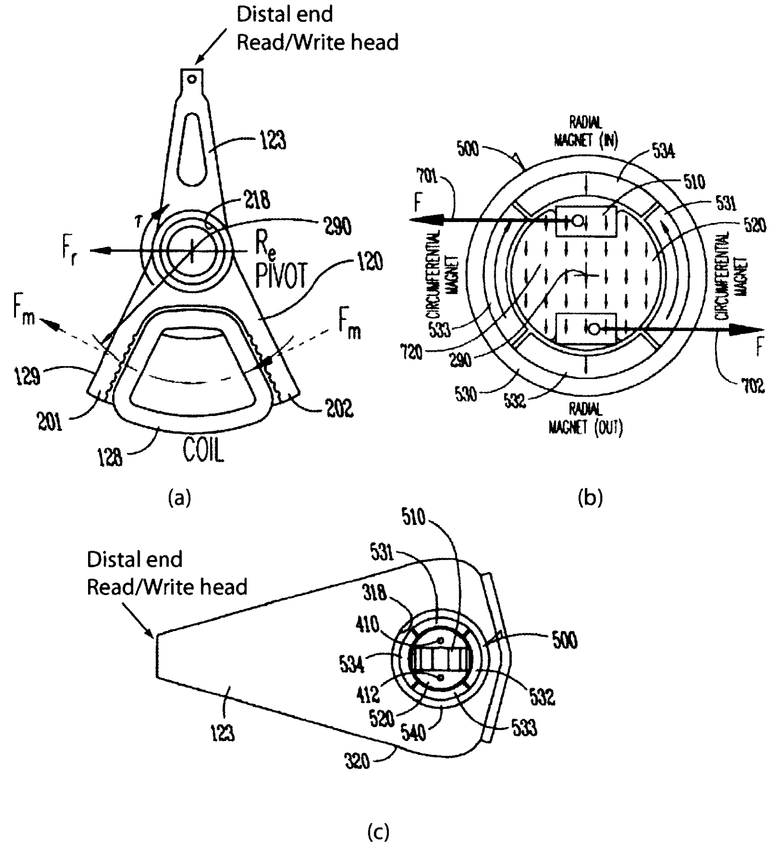
[1] The constraint on control bandwidth imposed by non-minimum phase zeros is worked out for an example positioning system in Section 5.4.6.

applied force F_m also exerts a force F_r at the bearing, exciting its translation mode. The displacement at the read head is composed of the difference of modal responses arising from the rigid body rotation and the bearing translation mode. The presence of the bearing translation mode is undesirable for two reasons: (i) the translation shows up in the displacement at the read head and (ii) the transfer function between the applied force and the measured displacement at read head can be non-minimum phase under certain geometry conditions [59]. A novel actuator (see Fig. 5-2(b)) based on a set of magnetic arrays called Hallbach arrays is designed in [54] to form a voice coil motor that generates only a torque and now net translational force. The new design topology is shown in Fig. 5-2(c) with the purely-torque motor mounted in the pivot itself, without the need for the linear force F_m applied at the arm distance R_e. Without this design topology change, the translation of the bearing and the non-minimum phase zero would limit the performance of the read head.

In summary, the two examples discussed above emphasize the need for developing suitable design topologies before any optimization is attempted. An interesting extension of this problem is one of identifying a set or library of topologies from which we can select an appropriate topology. In what follows, we discuss our integrated design and control approach that builds the case for optimizing over a library of topologies, instead of optimizing parameters within a given topology.

5.2 Our Approach

Based on the examples of integrated design and control described in Section 5.1, we identify the four possible cases for integrated design and control in Table 5.1. As indicated in the table, in any design for control approach, the design (or plant) and the controller need to be judiciously chosen in the problem formulation step. The options listed in the table are based either on a *fixed* or a *varying* topology/order for



(a) (b)

(c)

Figure 5-2: Design for control example from [54] modifying actuator design to eliminate translational loading in voice-coil motors. (a) In disk drives, the actuator for the read/write head is a typical lorentz-force voice coil motor that produces a force at an offset. The force also excites the translational mode of the bearing. (b) A novel design for the voice coil motor, based on a magnetic array called as Hallbach array, which is commonly found in linear motors, is used in a rotary configuration to produce a unidirectional magnetic field in the hub. (c) The resulting actuator is a pure torque motor that minimizes the effect of the translational mode of the bearing.

a design/controller. Before we proceed any further, we present our definitions of these terms as relevant to the integrated design and control methodology we will propose shortly.

A *fixed* controller is one with a pre-specified order and parameters to be selected appropriately. A basic knowledge of the plant dynamics can facilitate a nominal choice for the controller order. The problem of maximizing performance reduces to selecting the best possible controller parameters. However, such a *fixed* structure for the controller limits in most cases the freedom in maximizing performance. On the other hand, we define a *varying order* controller as one in which the order is not pre-selected. Rather, the controller order evolves in the integrated design and control iterations.

Similarly, we define a *fixed* design topology as one in which the overall structure is initially chosen and the iterations are performed to tune the parameters. Tuning the parameters of a *fixed* design cannot alter the design structure at all. On the other hand, by a *varying* design topology option, we iterate with changes in design topology, or configuration, each time creating a different design altogether. As explained in Section 3.2, for example, a design topology can be varied to go from a parallel kinematic design to a serial kinematic design, or from an exact-constraint design to an elastic-averaging design.

Table 5.1: All possible cases for integrated design and control.

Case	Design Topology	Controller Order
I	Fixed	Fixed
II	Fixed	Allowed to Vary
III	Allowed to Vary	Fixed
IV	Allowed to Vary	Allowed to Vary

The cases presented in the table work out as follows. In Case I, the integrated approach optimizes performance over a combination of design and controller param-

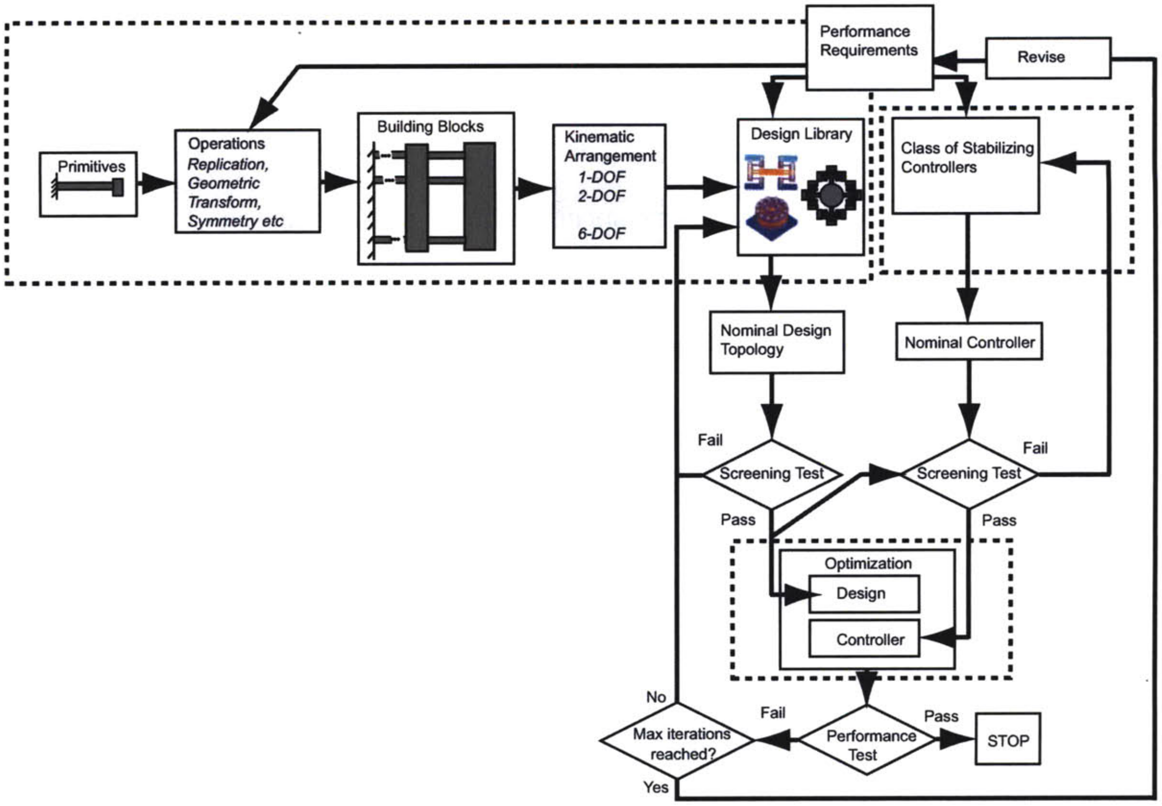


Figure 5-3: Integrated design and control methodology for meeting performance requirements.

eters. The final outcome after the iterations is a design and a controller of the same structure as at the beginning of the iterations, but with the selection of the most promising parameters. Hence, in this case, both design and control structure are fixed and cannot be altered. A poor choice of design topology or controller structure can leave critical performance requirements unfulfilled.

In Case II, for a fixed design topology, the controller is allowed to vary. Hence, in the iterations, the design parameters and controller order evolve to facilitate optimizing the performance.

In Case III, the design topology is varied (i.e. many possible design structures are tested) against a controller with a fixed structure. This case fundamentally limits the performance, since for every new plant we are confined to the same controller type. The potential performance of the system can be lost in thus fixing the control structure.

The most intuitive and useful case is Case IV, which uses a varying the design topology and the order of the controller. However, since the number of possible design configurations in typical nanopositioning system applications are finite, the varying design topology problem can be broken down into a number of fixed design (each tested with a controller of varying order) problems. Hence, we formulate our methodology on Case II with a controller of varying order tested for each fixed design topology of all possible design topologies.

5.3 Proposed Methodology

In this section, we detail the steps involved in implementing the integrated design and control approach based on Case IV of Table 5.1, in which both the design topology and the controller structure are allowed to vary. As discussed above, we simplify this to the one of Case II for a *varying order* controller tested for all possible *fixed* design topologies from a design library. The methodology is illustrated as a flow chart diagram in Figure 5-3.

Step 1: Performance Specification: Formulate performance requirements for the flexure-based nanopositioning system. These requirements can be for the structural, thermal, or control, and other aspects of performance.

Step 2: Design Topology Library Generation: Follow the steps detailed in Chapter 3 to generate topology concepts that improve the specified performance requirements.

Step 3: Design Topology Selection/Screening: Every nominal design topology in the topology library is subject to a screening test to eliminate design topologies that obviously do not meet critical requirements. This screening test is necessary before (blindly) feeding the design topology to a shape and size optimization procedure.

Step 4: Controller Selection/Screening: On the controller side, an initial controller is selected as a *nominal* controller from the entire class of stabilizing controllers for the screened nominal design topology. For the screened nominal design and the nominal controller selected above, a screening test is used to weed out controller choices that do not allow for critical requirements to be met. It is important to perform this screening test before (blindly) feeding the design to an optimization procedure. The structure of the nominal controller is revised until it passes the screening test.

Step 5: Optimization: Given that the nominal design and the nominal controller have passed the screening test, we now feed them to an optimization procedure. This procedure collects the design and controller parameters and optimizes them for an objective function defined by the user. The design optimization may target shape and size optimization of the chosen topology. The controller optimization varies the order of the robustifying part of the controller to meet the robust stability and performance specifications on sensitivity transfer function or complementary sensitivity transfer function.

Many choices exist for implementing the design and control optimization. As we discussed in Section 5.1, different approaches for the optimization are discussed in [25]. These include (i) simultaneous design and control optimization (ii) sequential optimization, with design optimization followed by control optimization, and (iii) iterative design and control optimization. While each of these approaches could be applied in our methodology, we select the option (ii) of optimizing design first and then optimizing the controller for the optimized design since (i) it comes closest to what is done in practice when designing and controlling a hardware positioning system and (ii) it is computationally less intensive and feasible. Fig. 5-4 summarizes the details of the optimization block

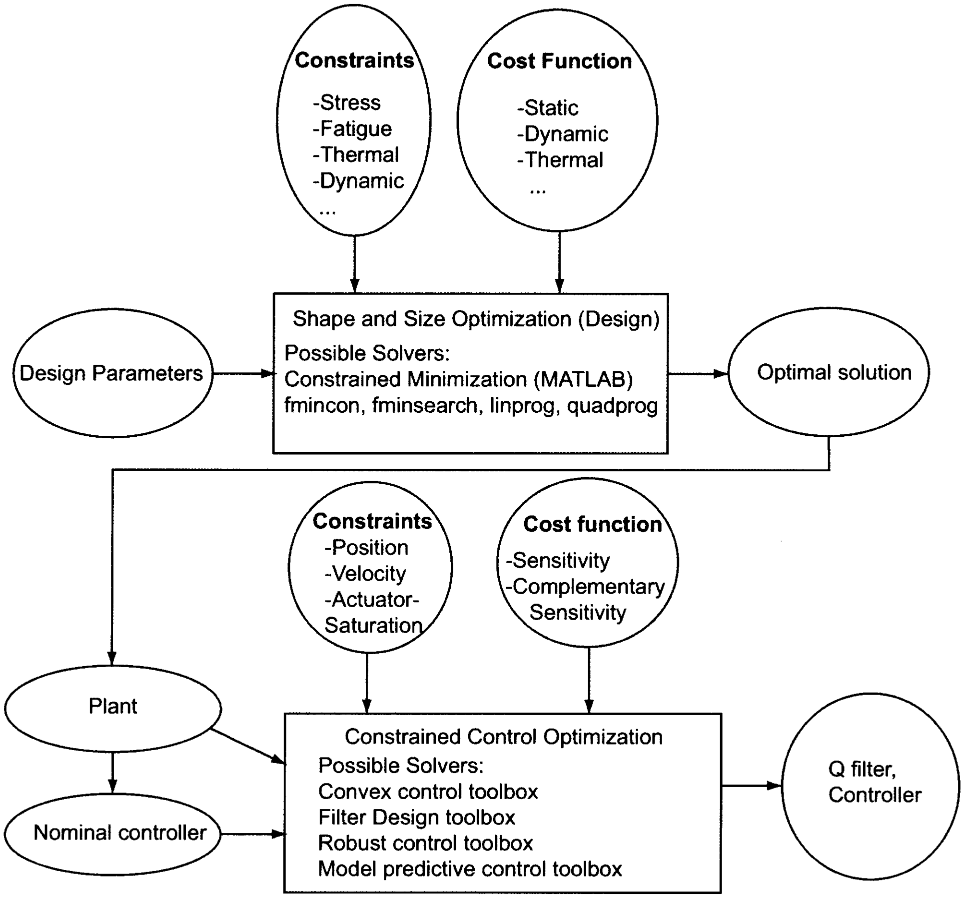


Figure 5-4: Schematic diagram showing constrained optimization in design and control. This is a detailed view of the optimization block shown in Fig. 5-3. In this diagram, we show design optimization performed first, and followed with control optimization.

of Fig. 5-3 highlighting such a sequential approach. The design optimization is formulated to select the design parameters that best allow minimizing or maximizing a desired cost function while meeting design constraints such as stress, fatigue limits. This part of the design optimization is often referred to as shape and size optimization. The optimized plant is fed to the controller optimization block. Once a nominal controller is chosen, it can be enhanced for imparting robustness.

If the performance requirements are met by the outcome of the optimization procedure, the controller is tested on the hardware to see if the performance can be demonstrated. If the performance requirements are not met at the end of the optimization procedure and the maximum number of iterations has not been reached, the nominal design topology is revised. Unless the nominal design topology is revised it is impossible to achieve the desired performance. If the maximum number of iterations has been reached, the only way to proceed any further is by relaxing the performance requirements. The design intuition gained from the optimization or from the hardware application should be used to revise the performance requirements suitably, taking us back to Step 1 listed above.

5.4 Example

In this section, we work out in simulation an example application to illustrate the integrated design and control methodology presented above. A simple, yet practical, positioning system example with a flexure-based mechanism driven by a piezoelectric actuator is considered. A broad overview of the rest of the section is as follows. We present a generic problem description, and then specify a set of critical performance requirements for the problem. The methodology is applied to first generate a set of design concept topologies. Based on design screen tests, a few topologies are ruled out. An optimization problem is formulated in terms of a desired cost function and a set of physical constraints. Design topologies passing the screening test are then input to the optimization problem. A MATLAB-based optimizer is used to fine-tune the shape and size of topology candidates. If the design or control cost function cannot be physically obtained, an optimal solution cannot be obtained and the topology is discarded.

5.4.1 Problem Description

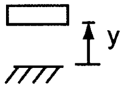


Figure 5-5: Schematic diagram showing a positioning system example. The goal is to vary the gap z over a large range of motion and control bandwidth.

Consider the precision positioning problem shown schematically in Fig. 5-5. The objective is controlling a relative separation z between a moving surface and a fixed surface to form a controllable gap. Such a gap can be used to study physical phenomena at sub-micron scales, such as radiative heat transfer or force interactions such as Casimir forces that occur between metals [8]. Other example applications include size-based filtration for macromolecular separations [7] and characterization of electrochemical properties of gas and liquid molecules [19].

A piezoelectric stack actuator with a lever amplification mechanism is suggested for generating a large displacement range on the order of $100\mu m$ required for the gap z. A schematic diagram showing the concept of a lever mechanism with piezoelectric actuator is presented in Fig. 5-6. The piezoelectric stack, shown as generating an input displacement y_{in}, pushes a lever at a distance L_a away from its pivot. The gap is formed at the distal end, a distance L_s away from the pivot, where the lever output displacement y_{out} is sensed with a laser interferometer. For small-angle motions of the lever about the pivot, the output displacement y_{out} for a input displacement y_{in} is given as

$$y_{out} = y_{in}\frac{L_s}{L_a} \tag{5.1}$$

Our goal here is to illustrate the design and control methodology for the positioning system conceptualized in Fig. 5-6, using flexure-based mechanisms in the design to for the pivot. Unlike friction-based bearings, flexure-based bearings are ideal candidates for the pivot owing to their smooth elastic motion and minimal nonlinearities such as backlash or hysteresis.

Before we proceed any further, we need to make a few assumptions for the relevant parameters. First, we assume the piezoelectric stack actuator has a blocking force of $F_{max} = 850\ N$ and free deflection $y_{piezo,max}$ of $18\,\mu m$. The static force-deflection characteristic of the actuator is shown in Fig. 5-7. In our design, we use two piezoelectric stack actuators held mechanically in series, so that their displacements add up to cause the net displacement input. For a maximum displacement y_{in} of $18\times 2 = 36\mu m$, we need to meet a target of $100\,\mu m$ at the output. Let us assume reasonable values for the distance of the sensor from the pivot $L_s = 2\ in$, and distance from actuator to pivot $L_a = 0.5\ in$. This results in an amplification ratio of 4, and a resultant maximum output displacement y_{out} of $144\,\mu m$, which satisfies our target displacement of $y_{desired}$=$100\,\mu m$. Basing on the stiffness of the structure, the applied force may vary, and the net displacement input can be smaller, so the extra buffer of $44\,\mu m$ is desirable.

Note that the simple model of Fig. 5-6 also depicts the simplified dynamics of a disk drive actuator subsystem example given in [59]. As we will be discussing towards the end of this chapter, in the example of [59], the geometry of a design is altered to improve control performance. Therein, changing the geometry involves changing the parameters within a selected topology. Here, as an alternative approach, we explore the option of varying the design topology to improve on the control performance. The case in which parameters within a design topology are varied is covered in our broad methodology.[2]

---

[2] As explained in Section 5.4.3, where we discuss the dynamics of a few designs represented by the simple model of Fig. 5-6, we motivate the need for altering the design topologies so as to move the non-minimum phase zero outside the range of frequencies of interest. Our approach of integrated design and control is implemented for achieving this feature. In the example of disk drive actuator system given in [59], altering geometry of the given topology eliminates non-minimum phase zeros. In a actual multi-DOF system, given many constraints on geometry, and design requirements, both (i) varying parameters within a topology and (ii) varying the topology (and parameters within each topology) should be explored. As we have seen earlier, the integrated design and control methodology

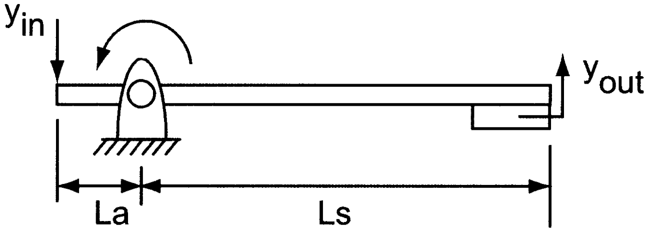


Figure 5-6: Schematic diagram of a piezoelectric actuator with lever amplification mechanism.

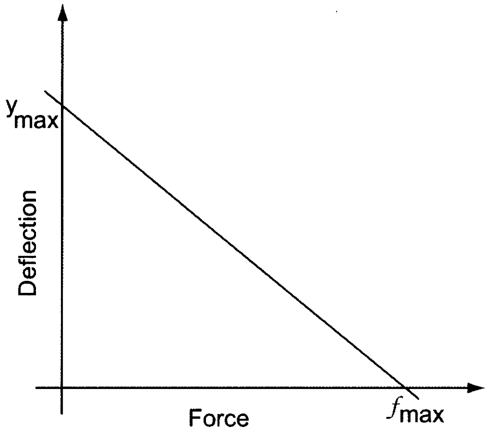


Figure 5-7: Typical static force-deflection characteristic curve of a piezoelectric stack actuator. The piezoelectric stack actuator we select in this application has a maximum force capacity (blocking force) $F_{max} = 850\ N$ and free deflection y_{max} of $18\,\mu m$.

Problem Statement

The problem statement for applying the proposed integrated design and control methodology to the example of the positioning system of Fig. 5-6 is as follows:

Given a lever amplification mechanism of Fig. 5-6 with the following parameters:

(ii) output displacement y_{out} *measured at a distance* $L_s = 2$ *in from the pivot.*

(iii) input displacement y_{in} *provided at a distance* $L_a = 0.25$ *in from the pivot.*

(iv) a piezoelectric stack actuator with a blocking force $F_{max} = 850\ N$ *and free deflection* $y_{piezo,max}$ *of* $18\,\mu m$.

Design a flexure-based pivot that meets the performance requirements given in Table 5.2.

---

proposed in Section 5.3 covers both these cases.

5.4.2 Implementation of Methodology

Given the above parameters for the lever and the piezoelectric stack actuator, we examine the topology, shape-size optimization and control performance of the system when a flexure-based mechanism is used as a pivot for the lever. We now follow the steps of the methodology as presented in Section 5.3 for the integrated design and control of the flexure-based pivot. For simplicity, we restrict our interest to planar implementations, which can be manufactured relatively easily on an abrasive waterjet or a wire-EDM.

Step 1: Performance specifications: The specifications for the positioning are as given in Table 5.2.

Step 2: Design Topology Library Generation: A set of topology concepts derived for flexural pivots are shown in Fig. 5-9. The idea is to use one of these pivots in the amplification mechanism shown in Fig. 5-6. One design topology using the flexure-based mechanism in Fig. 5-9 (c) as a pivot is shown in Fig. 5-8.

The candidate topologies of Fig. 5-9 were generated as follows. The concepts shown in Fig. 5-9(a) and (b) are simple examples of a rotational joint achieved with a lumped rotational compliance. While the notch flexure joint in Fig. 5-9(a) has a localized compliance around its neck, the beam flexure of Fig. 5-9(b)

Table 5.2: Specifications for 1DOF flexure-based positioning system example.

Desired Parameter	Value
Range of motion, max (y_{out})	$> 50\,\mu m$
Control bandwidth	$> 1\ kHz$
$x_{parasitic,max}$ Parasitic horizontal displacement for 5° uncertainty in actuator vertical alignment	$< 1\,\mu m$
Fatigue Performance	Infinite life, i.e. $\approx 10^7 - 10^8$ cycles or more

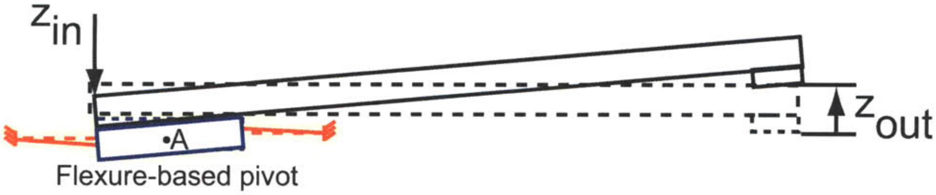


Figure 5-8: Design topology showing a flexure-based mechanism as a pivot for the lever amplification mechanism. Since the piezoelectric stack actuator applies a force offset from the center of mass, the lever rotates about the instantaneous center of rotation A, which forms a virtual hinge or pivoting point in the system. In this figure, flexure-based mechanism of Fig. 5-9(c) is selected as the pivot. Other design topologies under consideration in our example are those that use the flexure-based mechanisms of Fig. 5-9 (d)-(j).

has a compliance distributed over its length.

The rest of the design topologies shown in Figs. 5-9(c)-(j) are obtained as follows. First, we start with a beam flexure as a primitive used to suspend a mass. This primitive is shown in Fig. 5-10(a). To improve on the load-capacity and fatigue performance[3] of the primitive, we add a beam flexure on the other side of the mass. This enhanced primitive is shown in Fig. 5-10(b). A symmetric arrangement of the primitives on either side of the mass results in a redundant constraint, allowing for larger load-capacity, while at the same time ensuring under a vertical load the mass moves vertically, without a lateral motion error, as indicated in Fig. 5-10(b).

Note, however, that there is a second-order effect of over-constraint (the beams fighting with each other), which can be minimized with suitable geometry (for example, longer beam length). An improvement which eliminates the over-constraint is a parallelogram flexure in a folded back configuration [17]. For simplicity, this alternate primitive is not considered in this example.

Further, note also that the primitive flexure choice is not unique to a problem.

---

[3] Both of these are high strength requirements. As explained in Chapter 3, large range of motion can be obtained for the same load-capacity with redundant replication of the flexure constraint, instead of reinforcing a single constraint.

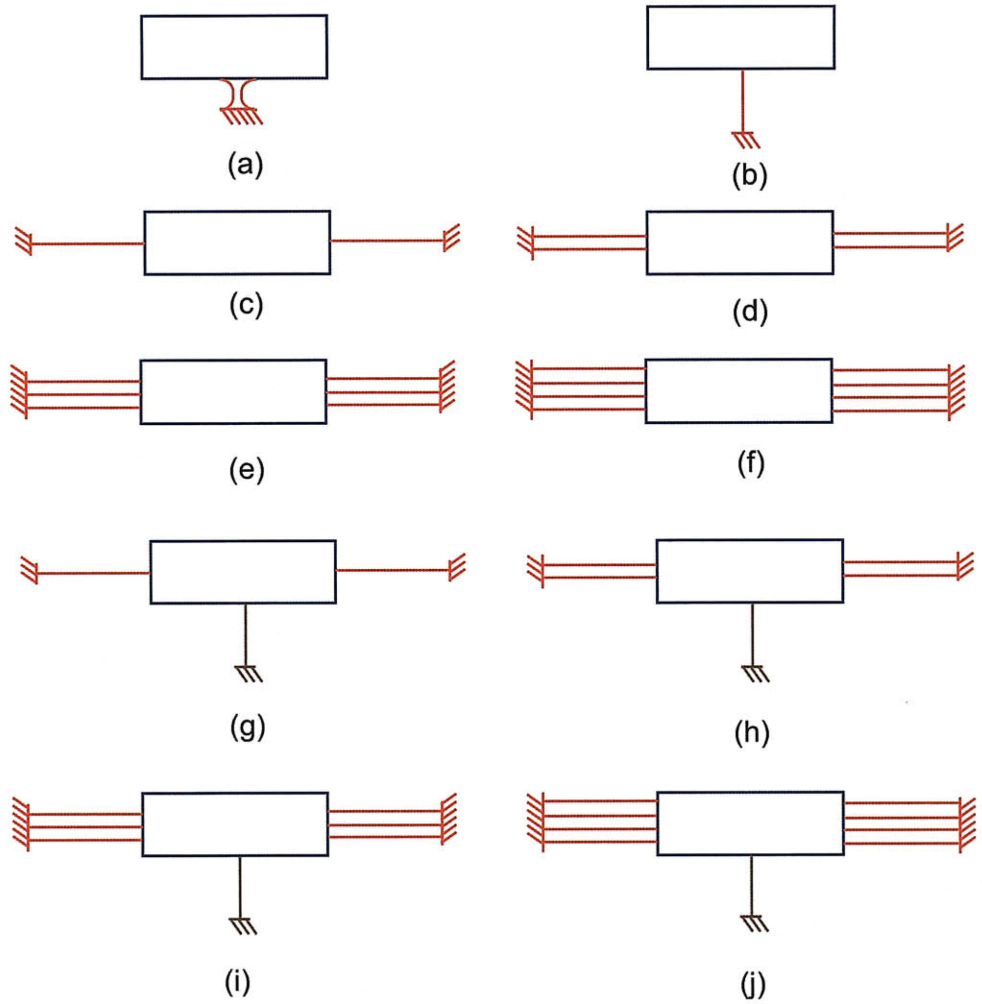


Figure 5-9: A library of candidate design topologies for a flexural pivot.

We converged at the design topologies presented in Fig. 5-9 (c)-(j) starting with a beam flexure of Fig. 5-10(a) as the primitive flexure.

An alternate primitive flexure that can also be considered is a notch flexure. An example double sided notched flexure equivalent of Fig. 5-10 (b) is shown in Fig. 5-11. Unlike the beam flexure which has continuous distribution of compliance, the notch flexure has a localized compliance. Designs featuring localized compliances become over-constrained in the presence of manufacturing errors, while those featuring distributed compliances are known to be more robust in the presence of such errors [44]. In this example, hence, we restrict

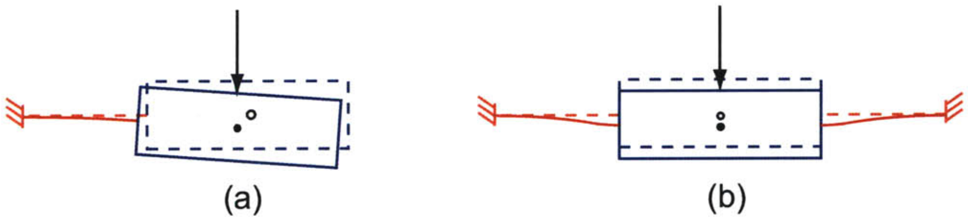


Figure 5-10: Primitive beam flexure shown in (a) is enhanced in its load-capacity by adding a redundant constraint in (b) to produce a double-sided beam flexure primitive.

our attention to the case of distributed compliance, i.e. to design topologies derived from beam flexure primitives as against notch flexure primitives.

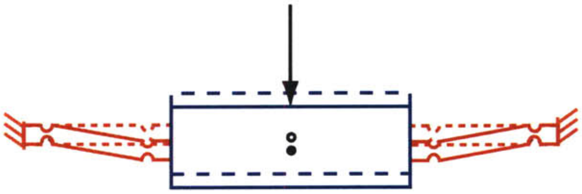


Figure 5-11: Double-sided notched flexure primitive. This primitive is the notched equivalent of the double-sided beam flexure of Fig. 5-10 (b). Unlike the beam flexure, which has a distributed compliance, the notch flexure has a localized compliance at the thin necks of the notch.

The concept topology shown in Fig. 5-9(c) is the double-sided beam flexure primitive of Fig. 5-10(b). With a force applied at an offset from the center of the mass, a rotational motion can be imparted. The pivoting action achieved with this topology is schematically shown in Fig. 5-8. Using the notation introduced in Chapter 3, this topology corresponds to a stacking of the primitive flexure of Fig. 5-10(b) with the following parameter values: a vertical stacking index $p = 1$, azimuthal stacking index $m = 1$, and serial stacking index $n = 1$.

The concept of topology shown in Fig. 5-9(d) has the same mass now suspended on a parallel stacking of two sets of beam flexures on either side. This configuration corresponds to a stacking of the primitive flexure of Fig. 5-10(b) with the following parameter values: a vertical stacking index $p = 2$, azimuthal stacking index $m = 1$, and serial stacking index $n = 1$.

On similar lines, the concept of topology shown in Fig. 5-9(e) and (f) have the same mass now suspended on a parallel stacking of three and four sets, respectively, on either side of the mass. These configurations correspond to a stacking of primitive flexure of Fig. 5-10(b) with the following parameter values: a vertical stacking index $p = 3$, azimuthal stacking index $m = 1$, and serial stacking index $n = 1$ for Fig. 5-9(e), and vertical stacking index $p = 4$, azimuthal stacking index $m = 1$, and serial stacking index $n = 1$ for Fig. 5-9(f).

The concept topology shown in Fig. 5-9(g) is similar to that of Fig. 5-9c, but with a rod flexure pinning down the mass at its center of rotation, and hence curbing the trampoline-like z-mode. Since the rod flexure has a large rotational compliance compared to axial compliance, this topology corresponds to a large modal separation $S_{Z,\theta}$ between the rotational and vertical DOFs.

The concept topologies shown in Fig. 5-9(h)-(j) are the similar center-pinned counterparts of Fig. 5-9(d)-(f), respectively.

Step 3: Design Topology Screening: Screening criterion for topologies can be decided according to the needs of the particular application under consideration. There is no unique way to select a screening criterion. Of many possible screening criteria to select the ideal topologies from the candidate topologies of Fig. 5-9, we select the following criterion that targets the specification of minimal lateral motion errors:

$$S_{X,Z} >> 1 \tag{5.2}$$

where $S_{X,Z}$ specifies the modal separation of the fundamental Z vibration mode from the X (lateral) vibration mode. Recall from Chapter 3 that the modal separation index between any two modes is a ratio of natural frequencies of the modes after a judicious choice of flexure constraints is implemented. The design topologies in Fig. 5-9(a) and Fig. 5-9(b) are more compliant in lateral direction than the vertical direction and hence do not satisfy this criterion.

In other words, for a slight vertical misalignment of the piezoelectric stack actuator, there would be a horizontal force component that will likely cause a large lateral error owing to the small lateral stiffness of the design topologies in Fig. 5-9(a) and Fig. 5-9(b). Hence, these two design topologies are eliminated.

In contrast, the high lateral stiffness of the beam flexures in Fig. 5-9(c)-(j) result in a high modal separation with the lateral DOF occurring at much higher frequencies than the vertical DOF. Hence, these ten candidates are passed to the subsequent shape and size optimization.

Step 4: Controller Selection/Screening: Of all possible stabilizing controllers, we screen for those that ensure both (i) good command following over frequencies up to 1 kHz and (ii) steady state error of zero for a step input. As will be shown later in Section 5.4.3, the plants corresponding to the five nominal design topologies, or plants, have no free integrators. Hence, it is imperative for the controller to have a free integrator in order to satisfy the screening criteria. Many nominal controllers can be constructed to satisfy this screening criterion, such as an integral controller, a proportional-integral controller, a lag controller, and other higher order controllers that have at least one free integrator. In this example, for simplicity, we select a simple integral controller as the nominal controller as given below:

$$C_0(s) = \frac{k}{s} \tag{5.3}$$

where k is a nominal gain selected for the given plant to ensure stability of the nominal closed-loop system.

Step 5: Optimization: We follow a sequential approach with the design optimized first, and the optimized design passed to the controller optimization routine. The formulation of the optimization and discussion of optimization results are presented in Sections 5.4.5-5.4.6. Before we proceed any further, we need to derive lumped parameter models and extract parametric relations needed for the optimization problem formulation.

5.4.3 Lumped Parameter Modeling

A lumped parameter model for the design topologies using flexure-based pivots of Figs. 5-9(c)-(f) is shown in Fig. 5-12. In this model, the flexure-based pivot is shown as a lumped mass suspended on its either side by flexures having a lumped linear stiffness k_y and a rotational stiffness k_θ. The pivoting point, or

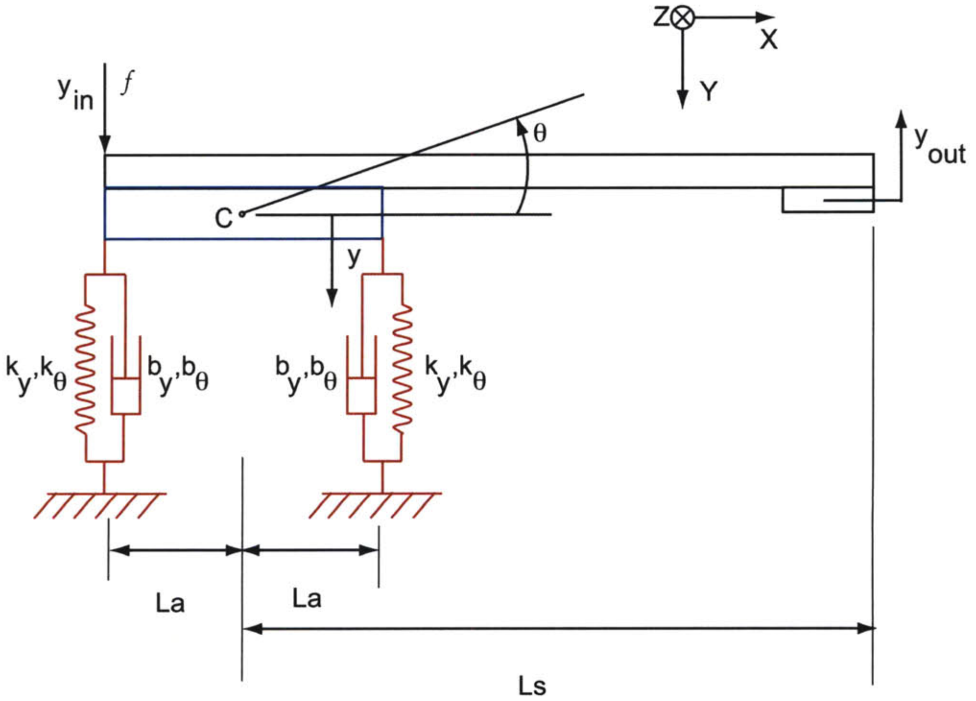


Figure 5-12: Lumped parameter model for depicting dynamic behavior of topology concepts using flexure-based mechanisms of Fig. 5-9(c)-(f) as pivots in the 1-DOF positioning system.

the instantaneous center of rotation of the system is the center of the mass, denoted by C in the figure. For a downward deflection y of the center C and an angle of rotation θ of the mass and the lever (about an axis Z perpendicular to the page and passing through C), the equations of motion for the system are:

$$\begin{aligned} m\ddot{y} &= f - 2b_y\dot{y} - 2k_y y && (5.4)\\ J\ddot{\theta} &= \tau - 2b_\theta\dot{\theta} - 2k_\theta\theta && (5.5)\\ f &= f_{max} - \frac{f_{max}}{y_{max}}y_{in} && (5.6) \end{aligned}$$

where m, J are the total mass and moment of inertia about Z axis passing through C, $\tau = fL_a$ is the moment applied by the force f applied by the piezoelectric stack actuator. Assuming lightly damped harmonics, damping factors b_y and b_θ denoting small damping in the flexures are used.

From the kinematics, as shown in Fig. 5-13, the output displacement y_{out} in terms of the downward deflection y of the instantaneous center of rotation and the angle of rotation θ is given as below:

$$y_{out} = -y + L_s\theta \tag{5.7}$$

After applying Laplace transforms[4] to the equations of motion, the transfer function $\frac{Y_{out}(s)}{F(s)}$ between the applied force input $F(s)$ from the piezoelectric stack actuator to the output displacement $Y_{out}(s)$ is given by:

$$\frac{Y_{out}(s)}{F(s)} = -\left\{\frac{1}{ms^2 + 2b_y s + 2k_y}\right\} + L_s\left\{\frac{L_a}{Js^2 + 2b_\theta s + 2k_\theta}\right\} \tag{5.8}$$

---

[4] We denote the Laplace transform of a time signal $x(t)$ as $X(s)$, i.e. $X(s) = \mathcal{L}\{x(t)\} = \int_0^\infty e^{-st}x(t)dt$

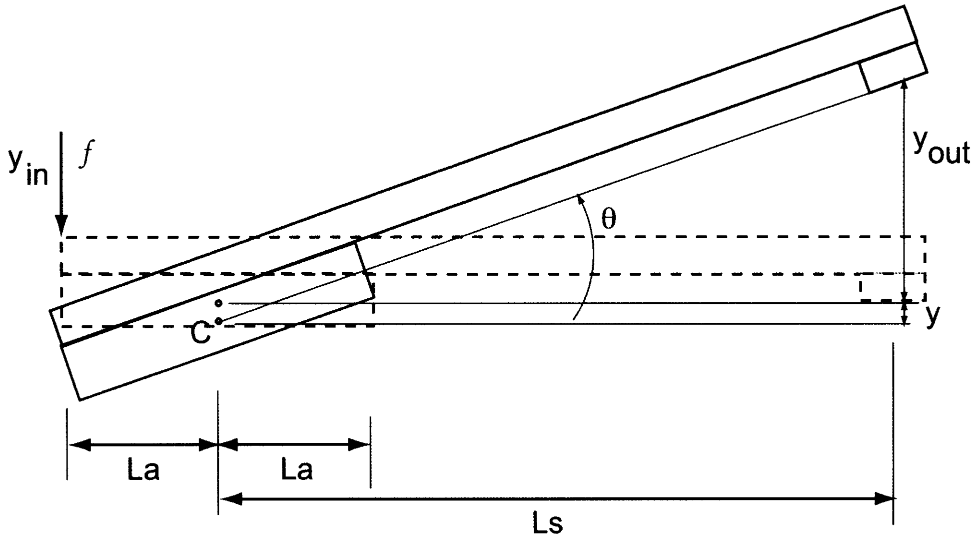


Figure 5-13: Schematic diagram showing kinematic relation between the output displacement y_{out}, the downward displacement y of at the center of rotation C and an angle of rotation θ. For small angle motions, since $y_{out}+y=L_s\theta$, we have $y_{out}=L_s\theta-y$.

The first term in Eq. (5.8) corresponds to the contribution of the fundamental vertical (y) mode of the flexural pivot as seen at the output displacement measurement. Similarly, the second term corresponds to the contribution of the fundamental rotational (θ) mode of the flexural pivot. Note the negative sign premultiplying the vertical mode. This means that, at the output, the difference of these two modes is being measured.

A lumped parameter model for the design topologies using flexure-based pivots of Figs. 5-9(g)-(j) is shown in Fig. 5-14. This model is the same as the model of Fig. 5-12 except for the enhanced lumped stiffness components at the center A. The rod flexure adds a high vertical stiffness k_{0y}, and a mild rotational stiffness

$k_{0\theta}$. The equations of motion for the enhanced system are:

$$
\begin{aligned}
m\ddot{y} &= f - b'_y\dot{y} - k'_y y && (5.9)\\
J\ddot{\theta} &= \tau - b'_\theta\dot{\theta} - k'_\theta\theta && (5.10)\\
k'_y &= k_y + k_{0y} && (5.11)\\
k'_\theta &= k_\theta + k_{0\theta} && (5.12)\\
y_{out} &= -y + L_s\theta && (5.13)\\
f &= f_{max} - \frac{f_{max}}{y_{max}}y_{in} && (5.14)
\end{aligned}
$$

After applying Laplace transforms to the equations of motion, the transfer function $\frac{Y_{out}(s)}{F(s)}$ between the applied force input $F(s)$ from the piezoelectric stack actuator to the output displacement $Y_{out}(s)$ is given by:

$$
\frac{Y_{out}(s)}{F(s)} = -\left\{\frac{1}{ms^2 + b'_z s + k'_y}\right\} + L_s\left\{\frac{L_a}{Js^2 + b'_\theta s + k'_\theta}\right\} \qquad (5.15)
$$

Note again the negative sign premultiplying the vertical mode. This means that at the output, as in the case of the lumped parameter model of Fig. 5-12, the difference of these two modes is being measured.

5.4.4 Parametric Relations

Based on the lumped parameter models presented in Section 5.4.4, we present key parametric relations relevant to the optimization problem. First, the parameter definitions for the system are tabulated in Table 5.3. To simplify the problem, some of the parameters are assumed to be known. For example, aluminium alloy Al 6061 is chosen as the material for the flexure-based pivot and lever. Al 6061 is selected as it is easily machinable on an abrasive waterjet or

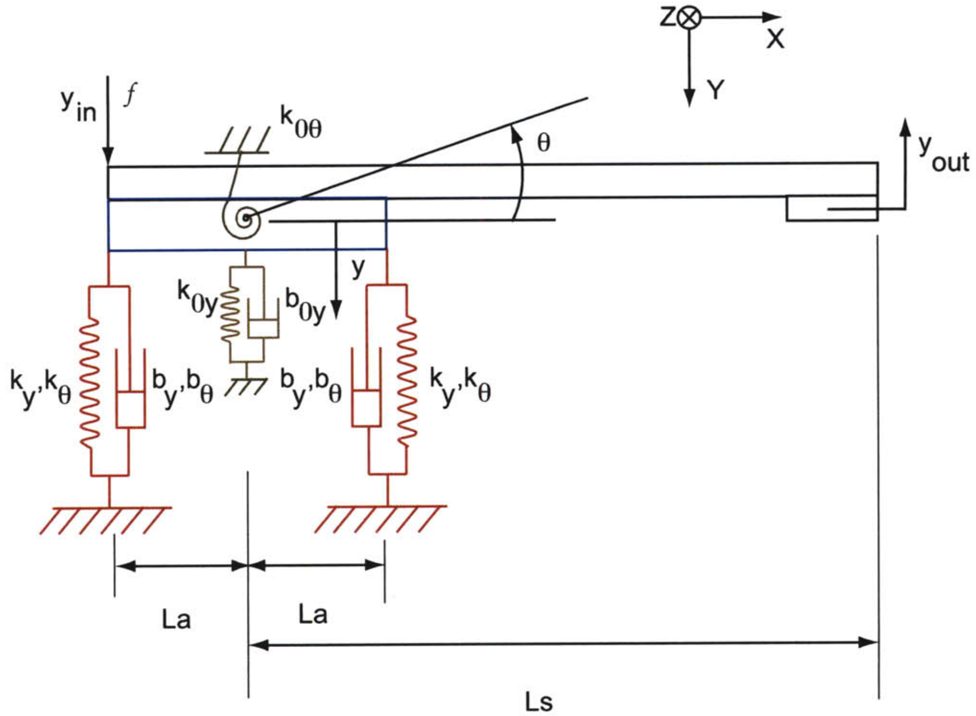


Figure 5-14: Lumped parameter model for depicting dynamic behavior of topology concepts using flexure-based mechanisms of Fig. 5-9(g)-(j) as pivots in the 1-DOF positioning system.

a milling machine. Material properties and other assumed parameters are also listed in Table 5.3.

Using a Lagrangian formulation and assumed modes method for approximation of the motion of the flexure mechanism, as described for a candidate flexure mechanism in Chapter 2, the parametric relations for the linear and rotational stiffnesses are presented in Table 5.4 for the topologies using flexure-based pivots of Fig. 5-9(c)-(f), and in Table 5.5 for the topologies using flexure-based pivots of Fig. 5-9(g)-(j).

5.4.5 Optimization: Problem Formulation

Optimization parameters

For the flexure-based pivots of Figs. 5-9(c)-(f), the optimization parameters

Table 5.3: Parameters of the 1-DOF positioning system configurations with a flexure-based pivot selected from Fig. 5-9(c)-(g).

Parameter	Definition	Value
E	Young's modulus of the material	70 GPa
ρ	Density of the material	2800 kg/m^3
σ_y	Yield stress limit of the material	550 MPa
σ_E	Endurance stress limit of the material	100 MPa
SF	Design safety factor for stress	2
$\sigma_{max} = \frac{1}{SF}\min(\sigma_y, \sigma_E)$	Maximum allowed stress in the material	50 MPa
D_{max}	Maximum footprint of the flexural pivot	4 in
$2L_a$	Length of central disk	1 in
B	Thickness of the flexures and central disk	0.25 in
W	Width of the central disk	0.25 in
$m_d = 2\rho L_a BW$	Mass of central disk	2.9×10^{-3}
$J_d = \frac{1}{12}m_d(4L_a^2 + B^2)$	Moment of inertia of the central disk about pivoting axis centered at A	$1.64 \times 10^{-7} kg.m^2$
L_s	Distance between sensor and pivot	2.0 in
L_a	Distance between actuator and pivot	0.5 in
$L_\ell = L_s + L_a$	Length of lever arm	2.5 in
W_ℓ	Width of lever arm	0.2 in
$m_L = \rho L_\ell B W_\ell$	Mass of lever	5.2×10^{-3} kg
$J_L = \frac{1}{12}m_L(W_\ell^2 + L_\ell^2)$ $+m_L(\frac{1}{2}L_\ell - L_a)^2$ $+m_L\frac{1}{4}(W + W_\ell)^2$	Moment of inertia of lever about pivot A	$4.1 \times 10^{-6} kg.m^2$
$J = J_d + J_L$	Total lumped moment of inertia	$4.26 \times 10^{-6} kg.m^2$
$m = m_d + m_L$	Total lumped mass	8.6×10^{-3} kg
h_0	Width of rod flexure	0.04 in
$I_0 = \frac{\pi}{4}r_0^4$	Bending moment of inertia of rod flexure	5.55×10^{-13} $kg.m^2$
ℓ	Length of beam flexure	To be determined
h	Width of beam flexure	To be determined
ℓ_0	Length of rod flexure	To be determined
$I = \frac{1}{12}Bh^3$	Bending moment of inertia of beam flexure	A function of w

Table 5.4: Linear and rotational stiffnesses of the 1-DOF positioning system configurations with a flexure-based pivot selected from Fig. 5-9(c)-(f).

Flexure-based Pivot used in Design Topology	k_y	k_θ
Fig. 5-9(c)	$12\frac{EI}{l^3}$	$4\frac{EI}{l^3}(l^2+3r^2+3lr)$
Fig. 5-9(d)	$24\frac{EI}{l^3}$	$8\frac{EI}{l^3}(l^2+3r^2+3lr)$
Fig. 5-9(e)	$36\frac{EI}{l^3}$	$12\frac{EI}{l^3}(l^2+3r^2+3lr)$
Fig. 5-9(f)	$48\frac{EI}{l^3}$	$16\frac{EI}{l^3}(l^2+3r^2+3lr)$

Table 5.5: Linear and rotational stiffnesses of the 1-DOF positioning system configurations with a flexure-based pivot selected from Fig. 5-9(g)-(j).

Flexure-based Pivot used in Design Topology	$k'_y = 2k_y + k_{0y}$	$k'_\theta = 2k_\theta + k_{0\theta}$
Fig. 5-9(g)	$24\frac{EI}{l^3}+\frac{EA_0}{l_0}$	$8\frac{EI}{l^3}(l^2+3r^2+3lr)+\frac{EI_0}{l_0}$
Fig. 5-9(h)	$48\frac{EI}{l^3}+\frac{EA_0}{l_0}$	$16\frac{EI}{l^3}(l^2+3r^2+3lr)+\frac{EI_0}{l_0}$
Fig. 5-9(i)	$72\frac{EI}{l^3}+\frac{EA_0}{l_0}$	$24\frac{EI}{l^3}(l^2+3r^2+3lr)+\frac{EI_0}{l_0}$
Fig. 5-9(j)	$96\frac{EI}{l^3}+\frac{EA_0}{l_0}$	$32\frac{EI}{l^3}(l^2+3r^2+3lr)+\frac{EI_0}{l_0}$

are selected as the length ℓ and thickness h of the beam flexures. For the case of Fig. 5-9(g)-(j), another variable, the length ℓ_0 of the rod flexure is also considered.

Constraints

The geometry/dimensional bounds on the parameter for a given footprint of the flexure-based pivot include

$$0.25\ in \leq \ell \leq D_{max} - 2L_a = D_{max} - 1\ in$$
$$0.05\ in \leq h \leq 0.3\ in$$
$$0.25\ in \leq \ell_0 \leq 2\ in$$

Let σ be the maximum stress in the beam flexures, σ_r the maximum stress in the rod flexure, y_{in} the displacement input from the piezoelectric stack actuator, y_{out} is the output displacement, and x_{out} the lateral motion error. The constraints used in the optimization are:

$$Constraint: \quad \sigma < \sigma_{max} \tag{5.16}$$
$$\sigma_r < \sigma_{max} \tag{5.17}$$
$$y_{desired} < y_{out} \tag{5.18}$$
$$y_{in} < y_{piezo,max} \tag{5.19}$$
$$x_{out} < x_{parasitic,max} \tag{5.20}$$

where σ_{max} is the maximum allowed stress in the material, $y_{desired}$ is the 100 μm output displacement requirement, $y_{piezo,max}$ is the maximum piezoelectric stack deflection, and $x_{parasitic,max}$ is the maximum allowed parastic lateral displacement. In order that the mechanism can withstand an infinite number of stress cycles, the maximum strss The cost function is defined for maximizing the output displacement and minimizing the lateral error motion as follows:

$$Cost: \qquad 0.5\frac{y_{out}}{y_{min}} - 0.5\frac{x_{min}}{x_{out}} \tag{5.21}$$

A constrained minimization problem was set up in MATLAB using $fmincon$ with an optimization parameter vector $\mathbf{U} = [\ell;\ h;\ \ell_0]$ with an initial guess $U_0 = \ell^0;\ h^0;\ \ell_0^0]$, and the bounds
$U_{min} = [0.25\ in; 0.05\ in;\ 1\ in]$ and $U_{max} = [D_{max} - 2L_a;\ \frac{D_{max}-2L_a}{10};\ 2\ in]$,
the cost function given in Eq. (5.21), the constraint conditions given in Eq. (5.20) as follows[5]:

```
[U] = fmincon(@cost,U_0,[A],[B],[],[],U_max,U_min,@constraint,[]);
```

where $A = [-0.1, 1, 0]$, $B = [0]$ were chosen to impose a geometry constraint $A\mathbf{U} < B$, i.e. $h < \frac{\ell}{10}$ on the beam flexures.

Since the lumped parameter models used in our parametric relations are based on linear Euler-Bernoulli beam theory that requires slender beams, this condition becomes necessary.[6]

The details of the controller optimization are as follows:

Control Parameter

Following the control parameterization of Chapter 4 the control parameter in our optimization is the parameter transfer function $Q(s)$.

Cost function

For good command following we choose a weight $W_s(s)$, such that a norm, say the ∞-norm, of the weighted sensitivity transfer function $W_s(s)S(s)$ is optimized

---

[5] The full MATLAB script files for this problem are given in Appendix C.1.

[6] A more elaborate optimization formulation could include the results of Timoshenko Beam theory applying the assumed mode approximation approach explained in Chapter 2.

as follows:

$$||W_s(s)S(s)||_\infty \leq 1 \tag{5.22}$$

The weight $W_s(s)$ is chosen as [65]

$$W_s(s) = \frac{(s + M\omega_d)(s + fM\omega_d)}{s(s + fM^2\omega_d)} \tag{5.23}$$

with $M = 1.5$, $f = 10$, and $\omega_d = 1.5\ kHz$ is the desired closed-loop system bandwidth. To make the controller $C(s)$ obtained with this choice of $Q(s)$ strictly proper, a filter with two first-order poles at $200 \times \omega_d$ is used to ensure a roll-off at high frequencies.

5.4.6 Optimization: Results and Discussion

The results of the optimization are presented in Table 5.6. An optimal solution was found for the case of flexure-based pivots of Fig. 5-9(g)-(j) for both the design and control optimization problems. A grid of 1000 points uniformly spread in the three-dimensional optimization parameter space $\mathbf{U}_{min} \leq \mathbf{U} \leq \mathbf{U}_{min}$ were each selected as an initial guess for the optimization. The values of the optimization parameters that gave the best optimal solution, i.e. lowest cost funciton value with no violation of constraints within a numerical tolerance

The optimization problem resulted in an infeasible solution in both design and control problems for the case of flexure-based pivots of Fig. 5-9(c)-(f). Note that the infeasibile solution was an outcome of optimization at each of the 1000 grid points chosen in the three-dimensional optimization parameter space $\mathbf{U}_{min} \leq \mathbf{U} \leq \mathbf{U}_{min}$, for a maximum number 200 sequential quadratic programming iterations at every grid point.

Table 5.6: Results of Optimization of Design and Control for the case of topologies of Fig. 5-9(c)-(j) used as flexure-based pivot. NMP zero stands for "non-minimum phase zero."

Topology	Design Optimal?	Controller Optimal?	Parameter Values	Range (μm)	Control Bandwidth (Hz)	Comments
Fig. 5-9(c)	No	No	—	—	—	Limited by NMP Zero, Fatigue
Fig. 5-9(d)	No	No	—	—	—	Limited by NMP Zero, Fatigue
Fig. 5-9(e)	No	No	—	—	—	Limited by NMP Zero, Fatigue
Fig. 5-9(f)	No	No	—	—	—	Limited by NMP Zero, Fatigue
Fig. 5-9(g)	Yes	Yes	$\ell = 1.5\ in$ $h = 1.25\ in$ $\ell_0 = 1.00\ in$	124.5	1194	
Fig. 5-9(h)	Yes	Yes	$\ell = 1.5\ in$ $h = 0.05\ in$ $\ell_0 = 1.00\ in$	141.1	1194	
Fig. 5-9(i)	Yes	Yes	$\ell = 1.5\ in$ $h = 0.05\ in$ $\ell_0 = 1.00\ in$	139.7	1194	
Fig. 5-9(j)	Yes	Yes	$\ell = 1.5\ in$ $h = 0.05\ in$ $\ell_0 = 1.00\ in$	138.4	1194	

The reason why the design topologies of Fig. 5-9(c)-(f) turned out to be infeasible is discussed as follows. First, we show here that these topologies are hindered in their control performance by the presence of non-minimum phase zeros.

Based on the model for the transfer function $\frac{Y_{out}(s)}{F(s)}$ between the applied force input $F(s)$ from the piezoelectric stack actuator to the output displacement $Y_{out}(s)$ of Eq. (5.8), the topologies made from flexure-based pivots of Fig. 5-9(c)-(f) have a zero in the right-half-plane at frequencies with clear impact on the desired tracking and disturbance rejection over the frequency range of $0 - 1000\ kHz$. This can be shown again from Eq. (5.8) as follows:

$$\frac{Y_{out}(s)}{F(s)} = \frac{\{mL_sL_a - J\}s^2 + 2\{b_yL_sL_a - b_\theta\}s + 2\{k_yL_sL_a - k_\theta\}}{\{ms^2 + 2b_ys + 2k_y\}\{Js^2 + 2b_\theta s + 2k_\theta\}} \tag{5.24}$$

Let us examine the sign (positive or negative) of the three terms in the Eq. (5.24). Given the geometry and other pre-selected parameters of the problem listed in Table 5.3, the term $mL_sL_a - J$ turns out to be positive. For simplicity, let us assume the damping terms $b_zL_sL_a - b\theta$ is negligibly small, as is the case for lightly damped harmonics commonly found in flexure0based mechanisms. Note here that $k_\theta - k_zL_sL_a$ is negative. This results in a non-minimum phase system with zeros in the right-half-plane. Fig. 5-17 shows a typical pole-zero plot for the open-loop plant under the assumptions for parameter values made in Sections 5.4.1 and 5.4.4. Size or shape-optimization accomplished by varying flexure length ℓ or thickness h cannot move the open-loop zeros out of the right half plane.

The presence of the right half-plane zero limits the bandwidth of these design topologies to about half the frequency of the zero, and hence the desired bandwidth of 1 kHz cannot be achieved [65]. The limitation on the bandwidth can be explained from a root-locus viewpoint: higher controller gains resulting from high bandwidth requirements will result in the closed-loop system poles moving toward the right-half-

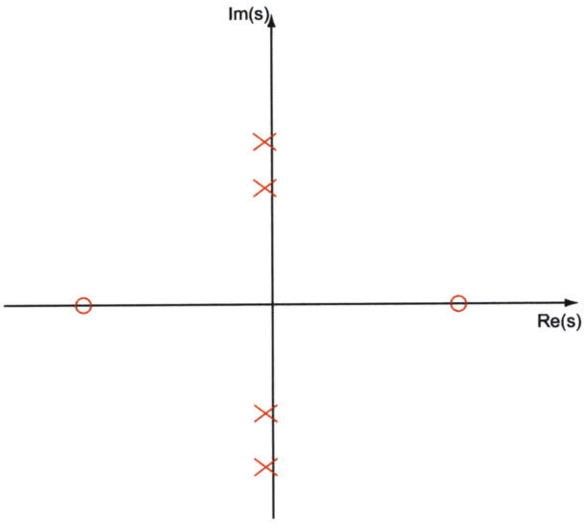


Figure 5-15: Pole-zero plot of open-loop plant corresponding to design topologies using flexure-based mechanisms of Fig. 5-9(c)-(f) as pivots in the 1-DOF positioning system. The zeros of the system are on the real axis and symmetric about the imaginary axis, hence resulting in a non-minimum phase behavior.

plane zero, and hence result in instability. To ensure stability, the gains have to be limited, and hence the bandwidth has to be limited.

The bound on the bandwidth for the topologies achieved with the flexure-based pivots of Figs. 5-9(c)-(f) can be derived as follows [65]. Let ω_0 denote the right half plane zero of the system. Since the system has no right half plane poles, the sensitivity transfer function $S(s)$ should obey the following constraint:

$$||W_s(\omega_0)S(\omega_0)||_\infty \geq |W_s(\omega_0)S(\omega_0)| = |W_s(\omega_0)| \tag{5.25}$$

where, we have used the fact that from its definition given in Eq. (4.3), the sensitivity transfer function assumes a value of 1 at the frequency of the zero. At this frequency, the condition on the upper bound of the weighted sensitivity given in Eq. (5.22) reduces to the following:

$$|W_s(\omega_0)| < 1 \tag{5.26}$$

Using the weighting filter W_s given in Eq. (5.23), the above inequality assumes the form given below:

$$\left|\frac{(\omega_0 + M\bar{\omega}_d)(\omega_0 + fM\bar{\omega}_d)}{\omega_0(\omega_0 + fM^2\bar{\omega}_d)}\right| < 1 \tag{5.27}$$

with $M = 1.5$ $f = 10$, an upper bound on the achievable control bandwidth $\bar{\omega}_d$ is:

$$\bar{\omega}_d < (1 - \frac{1+f}{Mf})\omega_0 = \frac{4}{15}\omega_0 \tag{5.28}$$

For the flexure-based pivots of Fig. 5-9(c)-(f) the maximum value for the frequency of the non-minimum phase zero was found to be $max\{\omega_0\} = 944.49\ Hz$ over the optimization parameter space, which then results in a value for the maximum achievable bandwidth from Eq. (5.28) to be 251.73 Hz. This value is much lower than our target bandwidth of 1000 Hz and hence these design topologies cannot meet the control performance requirements. Further the design constraint of infinite fatigue life were found in the optimization to be too stringent on the beam flexures. Since the central rod flexure is not available in these topologies all the applied load is taken by the beam flexures, which have a limited stress handling capability owing to the material yield limit. Hence, the flexure-based pivots of Fig. 5-9(c)-(f) need to be dicarded in our integrated design and control methodololgy.

One strategy to tackle the non-minimum phase zero is to move it far beyond the frequencies of interest. For this purpose, we can use a root locus argument to see the effect of parameters in the system on the location of the zero. From the numerator of the plant transfer function in Eq. (5.24), the zeros of the system satisfy the equation:

$$\{mL_sL_a - J\}s^2 + \{b_yL_sL_a - b_\theta\}s + \{k_yL_sL_a - k_\theta\} \;=\; 0 \tag{5.29}$$

which can be re-written as:

$$1 + \frac{1}{k_y L_s L_a - k_\theta} \left\{ (mL_sL_a - J)s^2 + (b_yL_sL_a - b_\theta) \right\} s \;=\; 0 \qquad (5.30)$$

and simplified to:

$$1 + \left\{ \frac{1}{k_yL_sL_a - k_\theta} \right\} s(s + \frac{b_yL_sL_a - b_\theta}{mL_sL_a - J}) \;=\; 0 \qquad (5.31)$$

Since, in our example, based on the assumed parameter values of Section 5.4.4-5.4.1, we have $mL_sL_a - J > 0$, and assuming $b_yL_sL_a - b_\theta > 0$, the locus of the zero positions for negative gains is shown in Fig. 5-16. As $|k_yL_sL_b - k_\theta|$ increases,

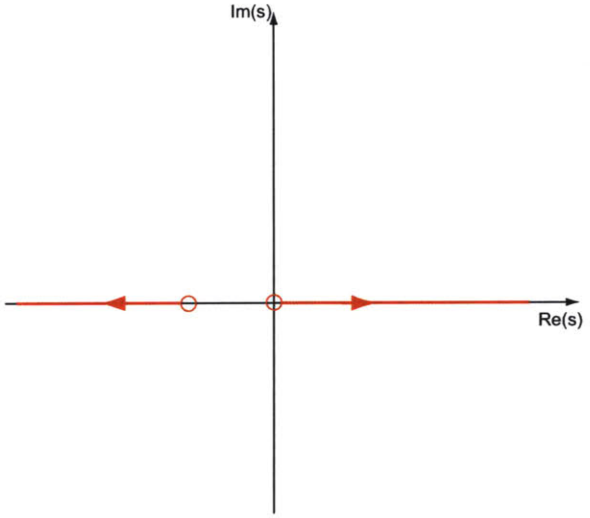


Figure 5-16: Root-locus plot showing the position of zeros of the system for positive increase in the stiffness term.

On the other hand, in the topologies of Figs. 5-9(g)-(j) the rod flexure stiffening the trampoline-like y mode of the pivot allows for pushing the non-minimum phase zeros on to the imaginary axis. The resulting system is minimum-phase with no bandwidth limitations imposed by their presence.

The above observation can be explained from Eq. (5.15) which is repeated here with the numerator terms expanded.

$$\frac{y_{out}(s)}{F(s)} = \frac{\{J - mL_sL_a\}s^2 + \{(b_\theta + b_{0\theta}) - (b_y + b_{0y})L_sL_a\}s + \{(k_\theta + k_{0\theta}) - (k_y + k_{0y})L_sL_a\}}{\{ms^2 + (b_y + b_{0y})s + (k_y + k_{0y})\}\{Js^2 + (b_\theta + b_{0\theta})s + (k_\theta + k_{0\theta})\}} \tag{5.32}$$

Note that adding the rod flexures at the center of the flexure-based pivot allows in increasing the resultant vertical stiffness much higher than the relative increase in the rotational stiffness. Hence the term $k_\theta + k_{0\theta} - (k_y + k_{0y})L_sL_a$ turns out to be negative, and since $J - mL_sL_a$ is negative, the zeros are moved away from the right half plane real-axis to to closer to the imaginary axis.

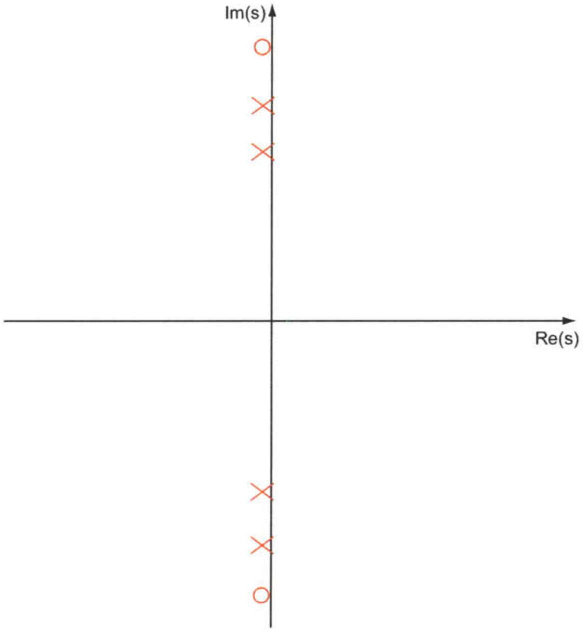


Figure 5-17: Pole-zero plot of open-loop plant corresponding to design topologies using flexure-based mechanisms of Fig. 5-9(g)-(j) as pivots in the 1-DOF positioning system. The zeros of the system are moved closer to the imaginary axis. The system is minimum-phase provided sufficient damping is provided to ensure robustness.

How close the zeros are to the imaginary axis depends on the physical damping in the system. Assuming metallic flexures, the physical damping is typically low. As discussed in Chapter 2 and 4, methods such as foam-damping, squeeze-film damping, and eddy-current damping can provide external modes of energy dissipation, and

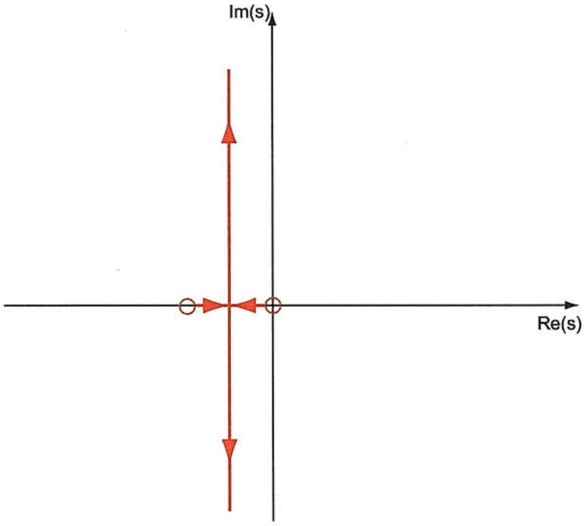


Figure 5-18: Root-locus plot showing the position of zeros of the system for positive increase in the stiffness term.

hence allow for an inherent robustness in the system. Since the closed-loop poles are drawn to the open-loop zeroes, robustness in terms of external damping must be ensured to be able to achieve higher bandwidths without the incidence of instability.

In summary, since non-minimum phase zeros are not found in the output measurement transfer function, we expect the topologies of Figs. 5-9(g)-(j) to be able to meet the specified performance requirements.

Of all the designs, the design topology with four beams in Fig. 5-9(h) has the largest vertical range. The design topology of Fig. 5-9(j) has the lowest lateral displacement error since it has the largest stiffness in X direction. Surpisingly, the design topology of Fig. 5-9(g) does not have the largest vertical range. This result is not obvious, and the optimal parameter vector is different from the rest of the topologies. One possible reason is that since there are fewef beam flexures to carry the load, the stress in the material is a limiting factor. This in fact is reflected in the optimal width of $h = 1.25\ in$ for this topology, as against $h = 0.05\ in$ for the rest of the toplogies.

All the design topologies of Fig. 5-9(g(-(j) meet the control performance requirement of 1000 Hz bandwidth. The control performance of the design topology with the the flexure-based pivot of Fig. 5-9(j) is shown in Fig. 5-19 in terms of the sensitivity transfer function. The nominal sensitivity transfer function resulting from a nominal controller $C_0(s) = \frac{1000}{s}$ has a low bandwidth, while the desired sensitivity has a bandwidth of 1000 Hz, a roll-on of $40db/dec$. Using the model-matching approach that we explained in Chapter 4, the controller parametric transfer function $Q(s)$ has to designed such that the sensitivity transfer function closely achieves the desired sensitivity transfer function of $\frac{1}{W_s(s)}$. Hence,

$$S(s) = \frac{1}{W_s(s)} \quad = \quad S_0(s)Q(s) \tag{5.33}$$

For $S_0(s) = \frac{1}{1+PC_0(s)}$, where $P(s) = \frac{Y_{out}(s)}{F(s)}$ and $C(s) = \frac{1000}{s}$, Q(s) was found to be:

$$Q(s) = \frac{s^4 + 2.121 \times 10^5 s^3 + 1.73 \times 10^8 s^2 + 3.665 \times 10^{13} s + 3.716 \times 10^{10}}{s^4 + 1.55 \times 10^5 s^3 + 2.172 \times 10^9 s^2 + 2.668 \times 10^{13} s + 3.445 \times 10^{17}} \tag{5.34}$$

To ensure a roll-off behavior for the resulting controller, a 2-pole low pass filter with coincident poles at $s = -200\omega_d$ was multiplied with the controller. The resulting sensitivity transfer function matches well with the desired sensitivity transfer function as shown in Fig. 5-19.

For comparison, a robust controller designed with a mixed-sensitivity criterion (allowing for tuning both sensitivity and complementary sensitivity) was also simulated, using *mixsyn* routine in MATLAB Robust Control Toolbox. The sensitivity transfer function obtained with this approach shows in Fig. 5-19 a flat profile at low frequencies and a roll-on of 40 dB/dec starting at about 0.1 rad/s.

The peak response of the sensitivity transfer function obtained with our method has about 2 dB taller peak than obtained from the mixed-sensitivity method, which implies a relatively poorer robustness to uncertainties in the positioning system. The closed-loop system sensitivity developed with our method shows a bandwidth of about 1194 Hz, while that of the mixed-sensitivity approach shows 1430 Hz. The corresponding controller frequency response plots are shown in Fig. 5-20. The mixed-sensitivity approach resulted in a controller of 8^{th} order, and has a lower peak in the sensitivity transfer function. This conforms to the observation that a higer order controller results in a more robust performance for the system.

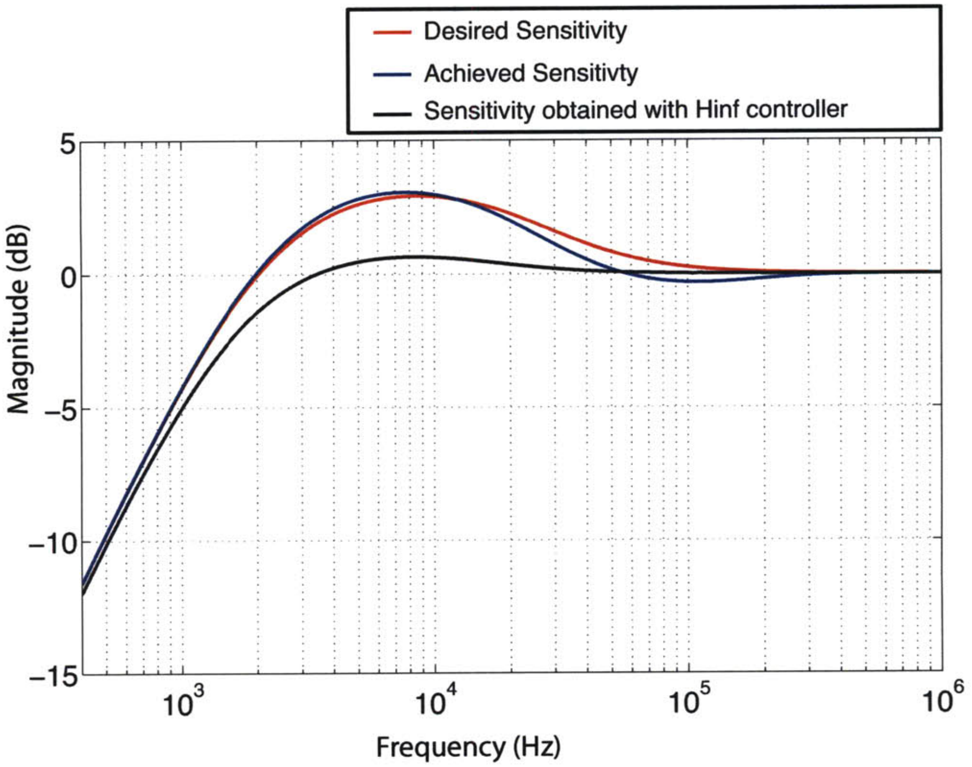


Figure 5-19: Comparison of magnitude response of desired and achieved sensitivity transfer function designed with a model-matching matching procedure. A corresponding response obtained for the case of mixed-sensitivity design is also shown.

In a nutshell, we have converged at a final design topology that meets the specified performance requirements of a bandwidth of greater than 1000 Hz and a range of motion exceeding 50 μm, with infinite stress-cycle life. Further, we have a systematic procedure to develop the topologies, screen them for desired features, and optimize them while dealing with dimensional and material constraints. An outcome of this exercise is the apriori identification of non-minimum phase zeros in flexure-based mechanism designs. Non-minimum phase zeros occur whenever non-collocated actuator and sensor arrangements are implemented. Avoiding the non-minimum phase zero may require reconsidering where to measure relative to where we actuate the system.

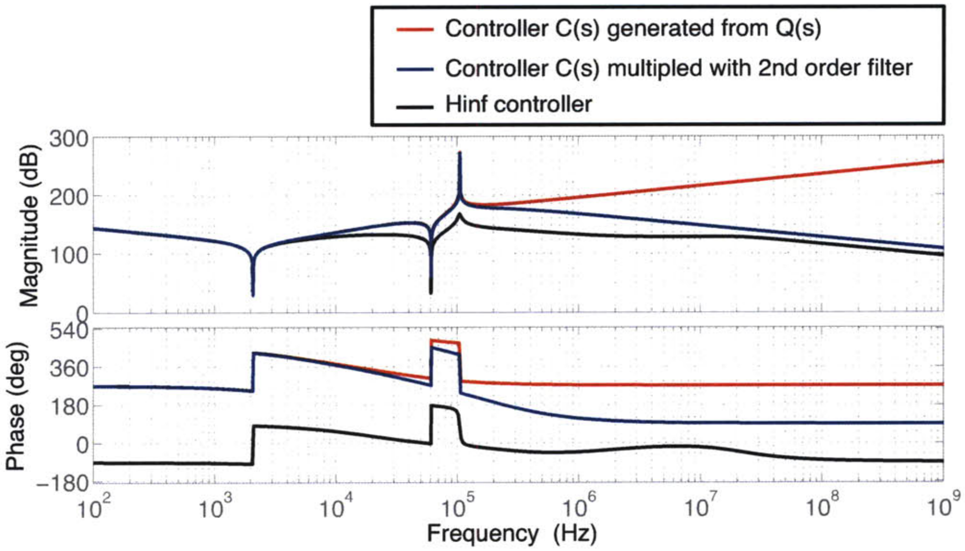


Figure 5-20: Comparison of frequency response of the (i) raw controller transfer function obtained from $Q(s)$ synthesis, (ii) low pass filtered version which is proper and implementable on harwdare and (iii) a $\mathcal{H}_\infty$ controller obtained from a mixed-sensitivity approach.

In our case, we chose not to vary the actuator or sensor location [57], or change the geometry of the design [59], but rather design the mechanism to be stiffened beyond the bandwidth of interest while still meeting the desired motion requirements. This was possible because of judicious design of topology in terms of flexure constraints, while satisfying strength and dynamic performance requirements.

5.5 Summary

In this chapter, we presented a flow chart for iterating on design (plant) and controller to achieve a desired closed-loop system specification. It is emphasized that iterating a design is not just about fine-tuning shape and size of a particular design configuration. Instead, we need to iterate over design topologies. The framework provided here allows to use the topology generation method of Chapter 3 and the controller generation method of Chapter 4.

An example of a flexure-based 1-DOF positioning system was worked out to show the integrated design and control methodology. Parametric relations were derived from lumped parameter models to formulate an optimization problem over the design space and the control performance space. The methodology was worked out step-by-step to cover (i) generation of design topologies (ii) screening of topologies for obvious design choices that cannot work for the given application, (iii) optimization formulation in terms of design parameters, cost functions, and equality and inequality constraints, and (iv) controller generation based on model-matching of a sensitivity transfer function. The infeasibility of a set of topologies was explained by the presence of non-minimum phase zeros that limit the achievable control bandwidth. Based on the intuition gained from this exercise, we also suggested a new screening guideline for checking for non-minmum phase zeros possible for a design topology along with an actuator and sensor placement. This example provides a practical 1-DOF positioning system application and provides a guided approach to converge at a novel design topology that meets all the requirements.

Chapter 6

Case Study: AFM Nanopositioner

This chapter covers the design and control of an atomic force microscope (AFM) nanopositioner. First, we examine a brief overview of atomic force microscopy. In particular, we highlight the need for flexure-based nanopositioners and how they have progressed over the past decade. We then select a design topology for the nanopositioner from a library generated using the principles detailed in Chapter 3. Further, the shape and size optimization of the selected topology is worked out using detailed models of the nanopositioner.

As a practical demonstration of the design, a hardware prototype was fabricated and tested for both static and dynamic performance. Sensitivity tuning of the nanopositioner is demonstrated using the controller parameterization developed in Chapter 4. This case study demonstrates the need for selecting the right topology to begin with before any standard optimization techniques are applied without forethought or physical insight into iterating on the design or controller parameters.

6.1 Background and Motivation

In this section we present an overview of the working principles of an atomic force microscope. We also highlight the advantages of a flexure-based nanopositioner in achieving nanoscale motion resolutions and high bandwidths that are needed for capturing an image.

6.1.1 Principle of Atomic Force Microscope

An atomic force microscope falls under a broader class of metrology tools termed as scanning probe microscopes, which raster scan a sample to form its image at nanoscale resolutions. Conventional optical microscopes are limited in resolution, to an order of half the wavelength of the light source, due to the phenomenon of optical diffraction. In contrast, scanning probe microscopes can achieve nanoscale resolutions, the finest limit being determined by the probe tip radius. The probe tip radius can be as small as 20 nm, or even smaller, and hence scanning probe microscopes have outperformed optical microscopes as imaging tools at the nanoscale.

Example techniques that fall under the class of scanning probe microscopy include mechanical stylus profilometry, scanning electron microscopy, along with atomic force microscopy. While each of these techniques has its own merits and limitations, two decades after its invention in 1986 [71], atomic force microscopy is now widely used as a metrology tool in characterizing nanoscale topographies, examples being biological samples such as DNA strands or bacteria cells, and silicon features in semiconductor manufacturing. Unlike mechanical stylus profilometers, atomic force microscopes are less damaging to sample being measured. Further, unlike scanning electron microscopes, they can image non-conductive samples as well.

The working principle of a conventional atomic force microscope is illustrated in Fig. 6-1. A MEMS cantilever is fixed on one end to a piezoelectric tube actuator and its free end is provided with a probe that has a sharp tip with a radius on the order of 20 nm. The piezoelectric actuator rasters along a scan path at submicron resolutions, allowing the cantilever probe to interact with the sample at each measurement location (pixel) on the scan path. The interaction force is dependent on

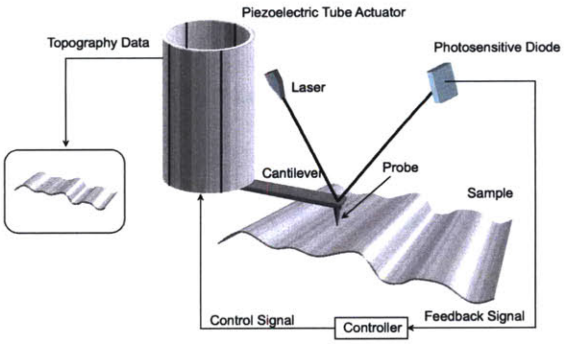


Figure 6-1: Schematic diagram illustrating the principle of an atomic force microscope (AFM). Photo courtesy: Daniel J. Burns, MIT Department of Mechanical Engineering [70].

the material of the sample and the distance from the sample, and depending on this interaction, the cantilever is deflected at each pixel location. There are many modes in which the AFM can be operated. In a constant height mode, the piezoelectric actuator end of the cantilever is controlled so as to maintain a constant height of the cantilever probe tip from the sample surface.

The piezoelectric actuator used in a conventional atomic force microscope is typically a piezoelectric tube, which has both longitudinal (axial tension) and lateral (bending) modes. The raster-scanning of the probe tip and extension in vertical direction are both achieved with the same piezoelectric tube. The fundamental limitation of this design configuration is that the vertical resonance frequencies of the scanner are limited to an order of 1 kHz. Such low frequencies slow down the measurement time to as much as 30 s, which cannot capture the physics of interest in most applications. This limitation led to an active interest in high-speed implementations of atomic force microscopy.

6.1.2 AFM Flexure Nanopositioners

The advantages of smooth elastic motion and absence of friction make flexure nanopositioners ideal for positioning either the probe or the sample in an AFM. Unlike piezoelectric tube actuators, piezoelectric stack actuators have a compact and rigid construction, and can be designed to achieve high natural frequencies along the axis of motion. Given the limitations of piezoelectric tube actuators, flexure-based mechanisms driven by piezoelectric stack actuators can be used for positioning at nanoscale resolutions. Further, with a constraint-based approach for layout of the flexure design, modal separations suitable for AFM can be achieved.

A ground-breaking paper published by Ando et al in 2001 addresses achieving high speeds with a piezoelectric stack actuator for each of the X, Y, Z axes of motion. About 8.5 kHz was achieved for the fundamental Z resonance of the device. A range of 250 $nm \times 250$ nm is reported. While low motion range is a limitation, this design is composed of stacked up axes, which makes it a serial kinematic design. With this design, the axis that drives the moving mass (of the actuator and other axes stacked on it) remains slow. Further, it precludes obtaining scan rotations, i.e. imaging along X and Y axes, or along any set of axes obtained by rotating them about Z.

Since scan rotation is a preferred feature of most atomic force microscope users, a parallel kinematic design of a flexure-based nanopositioner was developed by Fantner et al in [41]. This design uses piezoelectric stack actuators along all the motion axes and flexures as motion bearings. The fundamental resonance frequencies in X and Y axes are as high as 15 kHz. The fundamental resonance in Z direction of this design was measured as 27 kHz, with a dominant 40 kHz resonance reported in the step response of the mechanism. However, the scan range is limited to 15 $\mu m \times 15$ μm, which, in turn, limits the resultant field of view.

On the other hand, a nanopositioner with a large range 5 $mm \times$ 5 mm has been developed by Awtar et al [43]. As we discussed in Chapter 2, typical flexure-based mechanism designs are based on deflections that are smaller in an order of magnitude than the thickness of the flexure beam. Using a compliant design, Awtar et al explored the kinematic and elastomechanic behavior of flexures subjected to deflections on the order of their length. However, the large range is achieved at the expense of limitations in dynamic behavior. The out-of-plane dynamic behavior of this design is not reported, and is likely to be limited by a trampoline-like vibration mode.

A class of designs have been developed around resonant mechanisms that consume less power in their operation. A range of 50 μm has been obtained by resonating a flexure nanopositioner at 1024 Hz by MacKenzie et al [68]. Similar approaches have been followed by Prof. Mervin Miles, University of Bristol, UK and a company he founded [72]. The limitation to this approach is that in resonant scanning, the scan velocities are sinusoidal functions, making them nonlinear and approach zero at the sinusoidal peaks and troughs. This results in a non-uniform sampling along the scan path and further, "reduces the overall scan efficiency." [74]

Yong et al [66] reported a nanopositioner with a range of 25 $\mu m \times 25\ \mu m$ and coupling as lows as $-35\ dB$ between the X, Y axes. A mechanical lever with an amplification ratio of 2 used in the design allows for the large range. However, this mechanism is not designed for adequate modal separation between the in-plane and out-of-plane vibration modes. The out-of-plane dynamics are likely to be slowed down by a trampoline-like vibration mode of the flexures.

While the above-mentioned nanopositioners have advanced the state-of-the-art in AFM flexure designs, they have provided us ample avenues for innovating and approaching the problem from a design for control perspective. A summary of the comparisons between exisiting art and our nanopositioner based on demonstrated results is presented in Table 6.1.

Table 6.1: Comparison of our AFM nanopositioner with existing art. Data unreported in the references is left blank.

Researcher	Lateral Range ($\mu m \times \mu m$)	Lateral Resonance (kHz)	Vertical Resonance (kHz)	Coupling in X-Y (dB)	Comments
Yong (2009) [66]	25×25	2.7	—	-35	Trampoline mode in Z
MacKenzie (2008) [68]	$50 \times$ —	1	5	—	Non-linear scan speed
Awtar (2004) [43]	5000×5000	≈ 0.5	—	—	Trampoline mode in Z
Fantner (2004) [41]	15×15	15	26	—	Short range
Ando (2001) [69]	0.25×0.25	> 1.5	8.5	—	Trampoline mode in Z
Our nano-positioner	50×50	0.72	33.1	-30	

6.2 Selection of Design Topology

A set of design topologies suitable for the flexure nanopositioner for AFMs is presented in Fig. 6-2. In what follows, we examine how these topologies follow from principles listed in Chapter 3. Given our interest in flexure nanopositioners that can facilitate scan rotations, we limit our attention to parallel kinematic topologies.

The topology shown in Fig. 6-2(a) represents a simple flexural constraint arrangement that provides a symmetric design. Further, it also provides isolation between inputs along X from inputs from Y and vice versa. This has been referred to as actuator isolation in [43]. For the case of piezoelectric actuation along X and Y directions, actuator isolation would mean piezoelectric actuator driving the X axis cannot cause a shear load the piezoelectric actuator in Y.

With a parallel replication ($p = 2$) of the vertical beams in the building block, shown in red, of Fig. 6-2(a) we obtain a new building block and a resultant topology shown in Fig. 6-2(b). For wider spacing between the vertical beams in the building block, the modal separation between X, Y and Θ_Y, Θ_X increases.

To improve on the modal separation index between X, Y and Θ_Z, a building block with a pair of tangential beam constraints is shown in Fig. 6-2(c). In this topology, the modal separation between modes in X, Y and Θ_Z can be tuned as a function of the angle at which the beam constraints are arranged. A topology using flexural constraints in Z direction is shown in Fig. 6-2(d). This design improves on the modal separation index between X, Y and Z.

Note here that piezoelectric stack actuators are limited in their maximum deflection by the maximum percentage strain the ceramic material can sustain. As discussed in Chapter 2, typical materials such as lead zirconate can sustain up to about 0.1% of strain. While the topologies above are suitable for direct actuation with piezoelectric stacks, they can not achieve large range of motion. A large range of motion would allow for a large field of view with the AFM. A two-fold increase in scan range results in a four-fold increase in field of view. The increase in range can help to integrate the benefits of high resolution, large range, and high measurement speeds, which pushes the envelope on metrology instrument capabilities known in the literature to this date [70].

To reflect this design requirement of motion amplification, a set of topologies with amplifiers is presented in Fig. 6-3. The simplest amplifier is a lever mechanism shown interfacing between the input and the XY mechanism in Fig. 6-3(a). This amplifier has been used by Yong et al [66]. The limitation with this amplifier is the arcuate motion of the beam and the asymmetry inherent in the design. Instead, a symmetric closed amplifier can allow for a compact design.

A rhombus, or chevron, amplifier consisting of rigid links connected by flexural joints is shown as the amplifier in Fig. 6-3(b). While the chevrons are ideal motion elements along the axes of motion (i.e. diagonals of the rhombus), they form compliant twisting members between the fixed end and the end connected to the rest of the mechanism. To avoid twisting modes along Θ_X, Θ_Y directions, constraints are placed at the end of the chevron amplifiers in Fig. 6-3(c). Better modal decoupling between the Θ_X, Θ_Y of the chevron amplifiers and Z mode of the central stage are achieved with redundant constraints placed directly in Z direction in Fig. 6-3(d). The redundant constraints in this topology are placed at the end of the chevron amplifiers and at the central stage itself, because it is at the central stage that the highest Z stiffness needs to be allotted.

Of all the above topologies, we have opted for Fig. 6-3(d). Further improvements on this topology could be achieved with a momentum-balanced actuation in Z direction, which have been originally presented by Prof. Ando's group in [69].

6.3 Shape and size Optimization

In topology selection, our aim is to use a design configuration (layout of flexural constraints) that is appropriate for achieving the desired control performance. The goal for shape and size optimization is to fine-tune the design topology so we can achieve a feasible solution for a set of chosen constraints and cost functions. In this section, we first develop lumped parameter models to derive parametric relations relevant to the key performance aspects of the nanopositioner, such as lateral range and dynamic performance in the Z direction. We use these relations in formulating a constrained optimization problem, the solution of which is used for building the hardware prototype in Section 6.4.

6.3.1 Lumped Parameter Modeling

In this section, we develop lumped parameter models that will be used to generate parametric relations, that will be used in a shape and size optimization in Section 6.3.

Table 6.2: Chevron amplifier parameter definitions used in the lumped parameter model presented in Fig. 6-5.

Parameter	Definition
$f_{in,j}$	Force input to j^{th} chevron amplifier
$f_{out,j}$	Force output from j^{th} chevron amplifier
k_1,k_2,k_3	Elements of stiffness matrix
α	Angle of the chevron amplifier

Table 6.3: Piezoelectric actuator parameter definitions used in the lumped parameter model presented in Fig. 6-5.

Parameter	Definition
$f_{piezo,j}$	Force generated by series of N piezoelectric stacks in j^{th} chevron amplifier
$\delta_{piezo,j}$	Deflection of piezoelectric stack placed in j^{th} chevron amplifier
V_j	Voltage input to each member of the series of piezoelectric stack actuators placed in j^{th} chevron amplifier
N_j	Number of piezoelectric actuators placed in series in j^{th} chevron amplifier
d_{33}	Piezoelectric coefficient dependent on the ceramic material
F_B	Blocking force of each piezoelectric actuator
k_p	Stiffness of each piezoelectric actuator

Table 6.4: Central stage parameter definitions used in the lumped parameter model presented in Fig. 6-5.

Parameter	Definition
k_b	Bending stiffness of beam flexure connected to the chevron amplifer
k_{wb}	Bending stiffness of flexure between chevron amplifier and central stage
k_{wa}	Axial stiffness of flexure between chevron amplifier and central stage
m_s	Lumped mass of central stage
m_j	Lumped mass at the interconnection between the j^{th} chevron amplifier and stiffness k_{wa}
δ_s	Deflection of lumped mass m_s
δ_1	Deflection of lumped mass m_1
δ_2	Deflection of lumped mass m_2
f_z	Natural frequency of the out-of-plane (Z) mode

Table 6.5: Z rod flexure forest parameter definitions used in the lumped parameter model presented in Fig. 6-5.

Parameter	Definition
n_f	Number of rod flexures
k_{fx}	Lateral stiffness of Z-rod flexures
K_{fz}	Vertical stiffness of Z-rod flexures
d_f	Diameter of Z-rod flexures
l_f	Vertical stiffness of Z-rod flexures
$I_f = \frac{\pi d_f^4}{64}$	Area moment of inertia of Z-rod flexures
E_f	Young's Modulus of material of the Z-rod flexures

A lumped parameter model of the flexure nanopositioner driven by piezoelectric actuators is shown in Fig. 6-5. The parameter definitions are presented in Tables 6.2-6.5. In this model, we assume the chevron amplifier as being a two-port element, as shown on the left schematic diagram in Fig. 6-4. We assume that the amplifier has no energy losses [75], and with constitutive equations given as follows:

$$\left\{ \begin{array}{c} f_{out,j} \\ f_{in,j} \end{array} \right\} = \mathbf{K} \left\{ \begin{array}{c} \delta_{out,j} \\ \delta_{in,j} \end{array} \right\}$$

where the index $j = \{1, 2\}$ for the chevron amplifier placed in the $\{-X, X\}$ directions, and $\mathbf{K} = \begin{bmatrix} k_1 & k_3 \\ k_3 & k_2 \end{bmatrix}$ is a positive definite matrix [75] whose elements are a function of the geometry (such as the angle of the chevron amplifier, shape and size of the notch joints between the links, length of links) and material properties.

The force at the input of the chevron amplifier is from the piezoelectric stack actuators, whose force output is in turn governed by the constitutive force-displacement relationship of the actuator. The force and deflection at the input of the chevron amplifies are given as follows:

$$\begin{aligned} f_{in,j} &= \frac{f_{piezo,j}}{2} = \frac{k_{piezo}}{2N_j}(N_j d_{33} V_j - \delta_{piezo,j}) && (6.1) \\ \delta_{in,j} &= \delta_{piezo,j}/2 && (6.2) \end{aligned}$$

where, the lateral stiffness of the Z-rod flexures is $k_{fx} = 12 n_f \frac{E_f I_f}{l_f^3}$.

From static analysis of the lumped spring mass system model for the flexure part connected to the central stage between the chevron amplifiers, the following relationships can be derived:

$$\begin{aligned} f_1 + f_2 &= [4k_b + (k_{wb} + k_{fx})(1 + 2\frac{k_b}{k_{wa}})]\delta_s && (6.3) \\ \delta_1 + \delta_2 &= [2 + \frac{k_{wb} + k_{fx}}{k_{wa}}]\delta_s && (6.4) \end{aligned}$$

The constitutive relations at the interconnections between the chevron amplifiers and the lumped masses m_1 and m_2 are given by:

$$\begin{aligned} f_1 &= 2f_{01} && (6.5) \\ f_2 &= -2f_{02} && (6.6) \\ \delta_{01} &= -\delta_1 && (6.7) \\ \delta_{02} &= \delta_2 && (6.8) \end{aligned}$$

Using the above constitutive relations for the chevron amplifier and piezoelectric actuator Eqs. (6.1)-(6.2), flexures connected to the central stage Eqs. (6.4)-(6.4), and the interconnection relations Eqs. (6.9)-(6.8), we can show that the range of the central stage m_s is given by:

$$\delta_s = \left[\frac{\frac{k_3}{2(k_2 + kp/N)} F_B}{(k_{wb} + k_{fx}) + [k_b - \frac{k_1 k_2 - k_3^2 + k_1 k_p/N}{k_2 + k_p/N}](4 + \frac{2}{k_{wa}}(k_{wb} + k_{fx}))} \right] (V_2 - V_1) \qquad (6.9)$$

Now let us consider the dynamics along the Z direction. the natural frequency for the fundamental Z mode is given by

$$f_z = \frac{1}{2\pi}\sqrt{\frac{k_{fz}}{m_s}} \tag{6.10}$$

where, the vertical stiffness of the Z-rod flexures is $k_{fz} = n_f \frac{E_f A_f}{l_f}$, is assumed to be the predominant stiffness. The in-plane flexures are relatively compliant when compared to the forest of rod flexures, and hence their contribution to the Z stiffness is neglected.

Further, in terms of imaging functionality of the nanopositioner, the central stage should move laterally with minimal parasitic upward motion. When subject to lateral translation at the free end, the Z rod flexures act like cantilever beams that arc upwards to keep their length constant. This results in a vertical lift of the central stage which is undesirable for imaging. An estimate for the vertical displacement of the ends of the Z-rod flexures at their maximum lateral displacement is given by:

$$l_z = \frac{\delta_s^2}{8 l_f} \tag{6.11}$$

6.3.2 Optimization Problem Formulation and Solution

Given the parametric relations derived in Section 6.3.1, we can now pose a shape and size optimization problem as follows:

Problem Statement

In order to achieve the goals of (i) a X-Y static range of at least $50\,\mu m$*, and (ii) the fundamental natural frequency along Z axes on the order of* $30 - 50$ kHz *or more,* ***determine the following:***

(i) the number n_f ***of Z-rod flexures attached to the central stage,***

(ii) the length of the Z-rod flexures l_f***, and***

(iii) the angle of the chevron amplifier α.

For a parameter vector $\mathbf{U} = [n_f;\ l_f;\ \alpha]$, a possible choice for the cost function and constraints for the shape and size optimization are:

$$\min_{(n_f, l_f, \alpha)} \quad l_z \tag{6.12}$$

under the constraints:

$$\delta_s \geq 50\,\mu m \tag{6.13}$$

$$f_z > 35\ kHz \tag{6.14}$$

A constrained minimization problem was set up in MATLAB using *fmincon* with bounds $U_{min} = [1; 0.25\ in; 5°]$ and $U_{max} = [64; 1\ in; 10°]$, the cost function given in Eq. (6.12), initial guess $\mathbf{U}_0 = (\mathbf{U}_{min} + \mathbf{U}_{max})/2$ and constraint condition given in Eq. (6.14) as follows[1]:

```
[U] = fmincon(@cost,U_0,[],[],[],[],U_max,U_min,@constraint,[]);
```

The optimization converged to a local optimal solution $\mathbf{U}^* = [61;\ 0.3465\ in;\ 8.27°]$ for a mass of $5 \times 10^{-3}\ kg$ for the central stage and Z-rod flexures made from spring steel musicwires of diameter 0.5 mm. The prototype was manufactured using close values of $n_f = 64$, $l_f = 0.35\ in$, and $\alpha = 8°$. With these values for the parameters, a static range δ_s of 50 μm and a vertical natural frequency f_z of about 37 kHz, and a minimal value for lift l_z of about 35 nm was predicted by the model. This completes our selection of the design, which has been subsequently fabricated and tested. The details of the fabrication and assembly are presented next.

---

[1] The full MATLAB script files for this problem are given in Appendix C.2.1.

6.4 Fabrication and Assembly

In this section, we present details of the fabrication and assembly of the flexure nanopositioner design that we converged to in the previous section. An exploded view of a CAD model for the overall assembly is shown in Figure 6-6. The base plate is intended to be bolted to an optical table during the testing phase. The flexure nanopositioner is bolted on top on the base plate. In this model, the Z rod flexures are anchored between the flexure piece and a recessed face on the base plate. In the remainder of this section, we examine how this hardware prototype was fabricated and assembled so that the nanopositioner is suitable for assembling piezoelectric actuators, and for testing for vibration modes with a laser interferometric vibrometer.

6.4.1 Flexure Nanopositioner and Base Plate Subassembly

The flexure piece shown in Fig. 6-7(b) is made in a two-step process, the first involving milling and the second involving abrasive waterjet machining at water pressures of $500MPa$. The notched joints used in the chevron amplifier have radius of $2mm$ and a thickness of 0.2 mm at the thinnest portion. This dimension is too small to be machined directly with a waterjet. Instead, two holes of diameter 2 mm are milled on a CNC at a distance 0.2 mm between their centers. The milled part containing many such closely spaced holes is shown in Fig. 6-7(a). Subsequently, a waterjet toolpath is overlaid on the pattern of holes. To maintain registration between the different instruments, first, an auxiliary plate with holes for three dowell pins was machined under the waterjet cutter. Without changing the XY setting of the waterjet cutter, the milled part was then registered with respect to dowell pins and lowered on to the auxiliary plate. A recess provided in the auxiliary plate allows for machined parts of the waterjet cut to drop down without obstructing the path of the waterjet nozzle.

To achieve a topology in which the out-of-plane twisting modes of the chevron amplifiers (see Fig. 6-7(b)) are curbed, Z rod flexures are inserted in 500 μm holes machined into the flexure nanopositioner at each end of every chevron amplifier, as shown in Fig. 6-7(c). In all, there are 8 such Z rod flexures for curbing the chevron twisting modes. Z rod flexures for curbing the trampoline mode of the flexure nanopositioner are added later, in a subsequent assembly step.

Top and bottom views of the machined base plate are shown in Fig. 6-8 and 6-9, respectively. The flexure piece is bolted on the face shown in the top view. Recesses are provided beneath the footprint of the chevron amplifier and other moving parts of the flexure. A set of 500 μm holes are machined into the base plate to allow for forming a press fit for rod flexures anchored between the base plate and the ends of the chevron amplifier in the flexure piece. These holes were machined with a high density chromium-plate steel drill bit of diameter 0.02 *in* at speeds of 3300 *rpm*, after allowing for sufficient warmup time ($\approx$ 0.5 hour) for the CNC. Care should be taken not to break the drill bit in the base plate. Once a 500 μm drill bit breaks, it is hard to pry out the broken bit from the base plate. An alternative approach would be to turn a cylindrical pin that can be machined for the 500 μm hole and used to anchor the chevrons. The pin would then be press fit into the base plate.

At eight locations of the base plate, Z rod flexures from the flexure are anchored to curb twisting modes of the chevron amplifiers. Two of those eight locations are shown in either figure. The rest are arranged symmetrically around the central axis of the base plate.

The assembly of the base plate and flexure piece is shown in Fig. 6-10. Registration using dowell pins allows for bonding Z rod flexures between the ends of the chevron amplifiers and the base plate such that they remain in pure bending for lateral motions of the nanopositioner. Without the registration, the Z rod flexures may not be perpendicular to the XY plane, hence loading them in both bending and axial directions. Locations marked $1, 2, 3, 4$ on base plate and flexure are marked for dowell pins registration. Only three of the dowell pins are used in a press fit. To avoid over-constraint, the fourth dowell pin fits in to a slip fit.

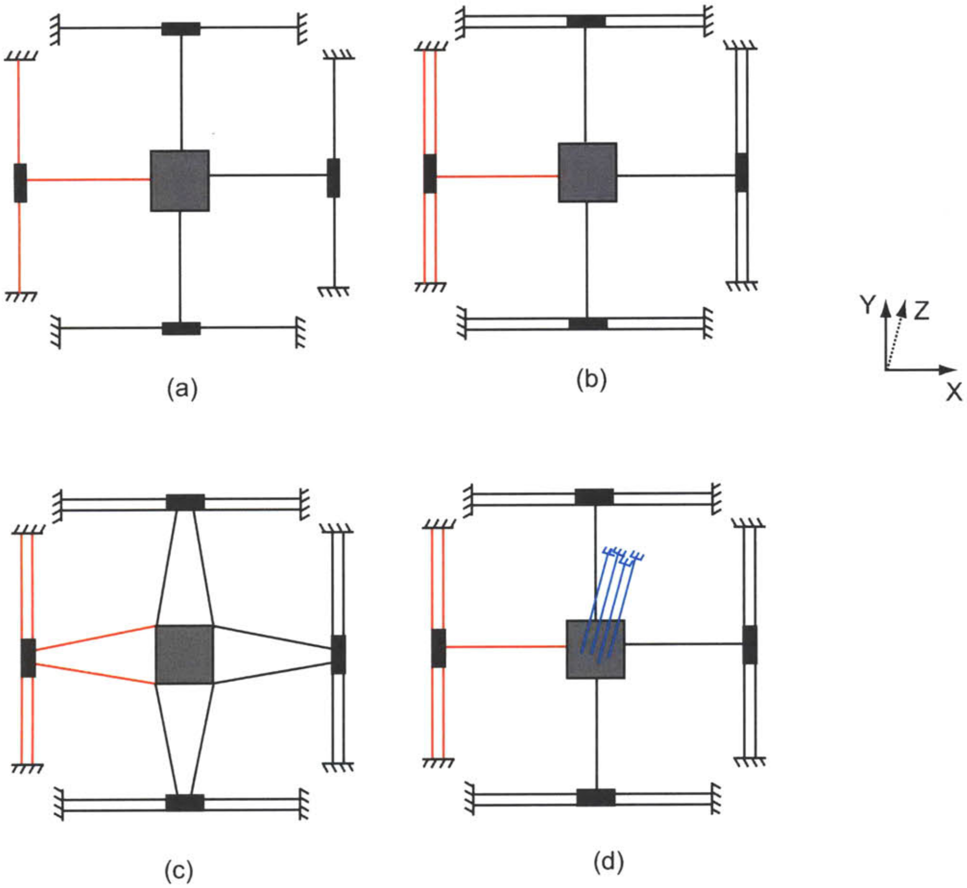


Figure 6-2: An XY positioner is shown in (a). For the purpose of exposition, we highlight one of the flexural building blocks in the color red. The highlighted building block in topology of (a) is replaced by a building block with double beams in (b). The use of widely spaced double beams in the building block of (b) not only improves the strength performance, but also improves the modal separation index between X, Y and Θ_Y, Θ_X. (c) shows a design with a building block chosen to improve the modal separation between X, Y and Θ_Z (d) shows a design with improved modal separation between X, Y and Z.

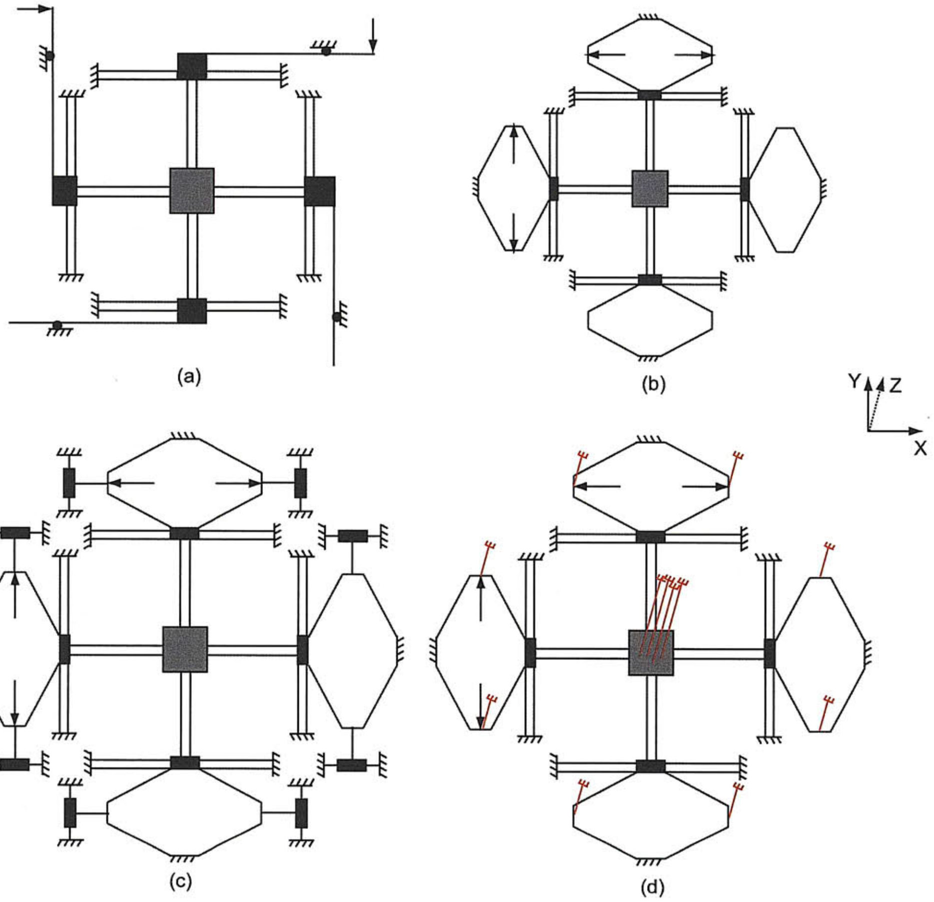


Figure 6-3: (a) shows a simple lever as the mechanical amplifier. A compact and symmetric topology using a chevron amplifier is shown in (b). A topology with constraints placed on the ends of the chevron amplifier is shown in (c) for curbing the Θ_X, Θ_Y modes of the chevron. A topology with improves further on the modal decoupling between the Θ_X, Θ_Y modes of the chevron amplifier and the Z mode of the central stage is shown in (d), which we have chosen as the ideal topology for our nanopositioner.

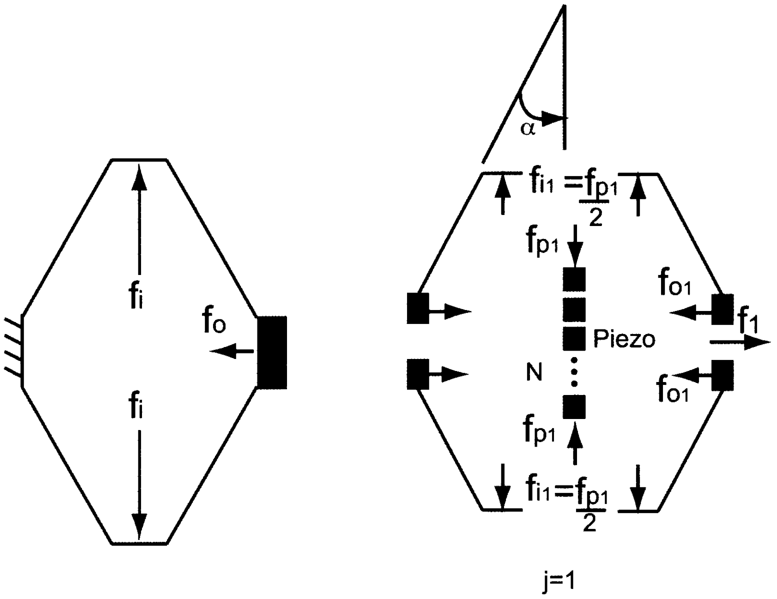


Figure 6-4: Schematic diagram showing chevron amplifier acting on a piezoelectric actuator as a two port element. A free body diagram showing force f_{p1} from the mechanical series of N-number of piezoelectric stack actuators placed back-to-back in the chevron amplifier, which is shown as split in two halves. An equal opposite force f_{p1} acts on the actuator interface of the chevron amplifier, with $f_{i1} = \frac{f_{p1}}{2}$ acting on each half of the chevron.

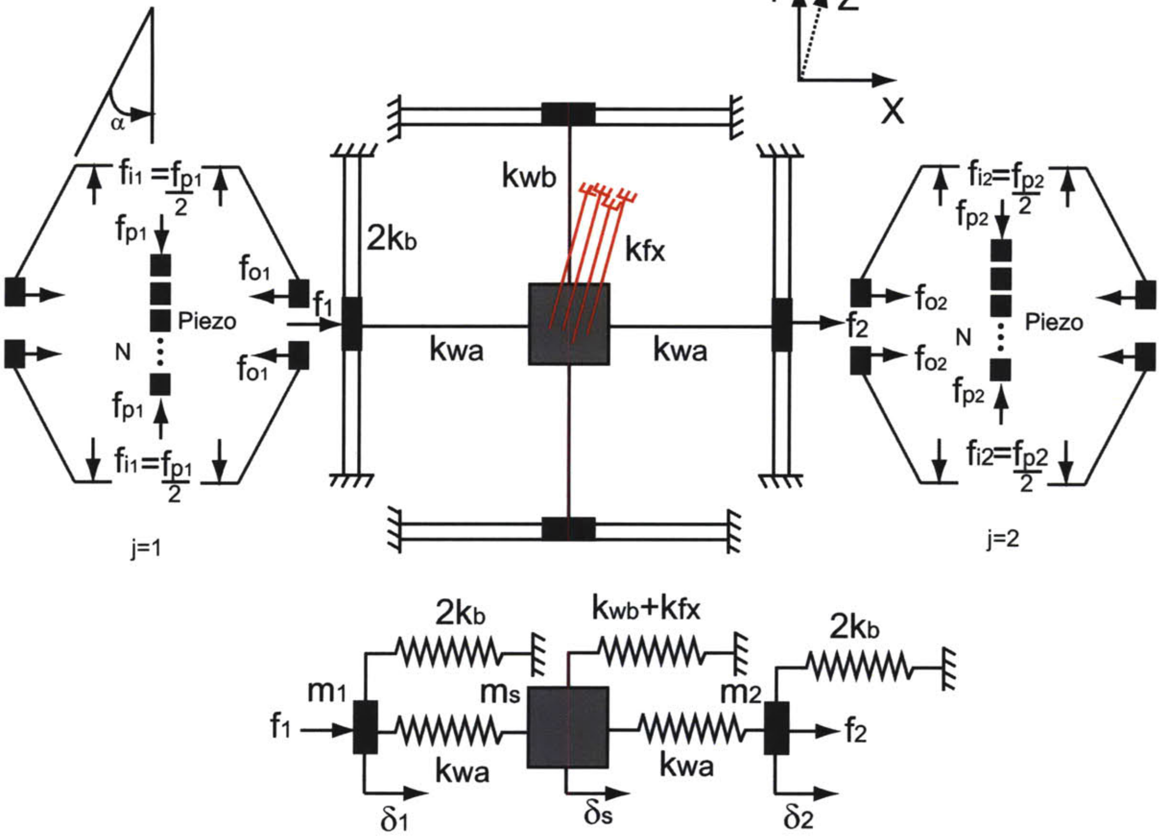


Figure 6-5: Lumped parameter model of the piezo-driven nanopositioner. This model is used to establish parametric relations in the design for shape and size optimization. (a)Model showing the chevron amplifier as a two port lumped element interfacing between N piezoelectric stacks arranged in series and the flexure mechanism. Two such chevron amplifiers are provided on X and $-X$, and the piezoelectric stacks are actuated in a push-pull mode of operation. This model shows the forces and displacements along the X axis for the components.

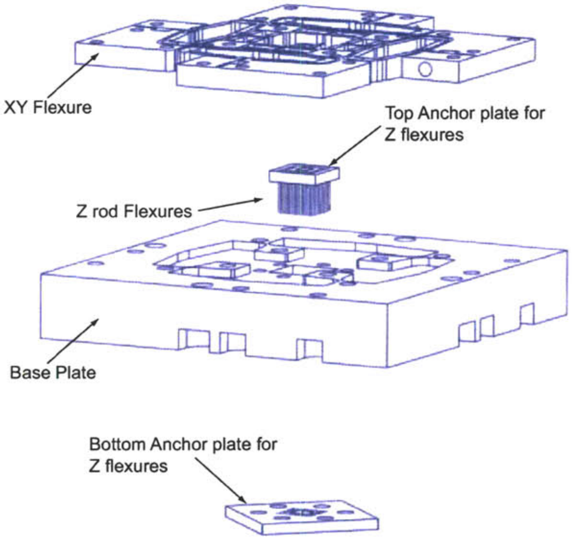


Figure 6-6: Exploded view showing the components of the nanopositioner assembly. The flexure piece is bolted to the base plate. Sixty four Z rod flexures are bonded with high-strength epoxy to a top anchor plate. A bottom plate that is fastened to the base plate anchors the other ends of the Z rod flexures. The forest of Z rod flexures allows for achieving high Z resonance frequency while still meeting the static range requirement. This picture does not show the Z rod flexures bonded between the ends of the chevron amplifiers and the base plate.

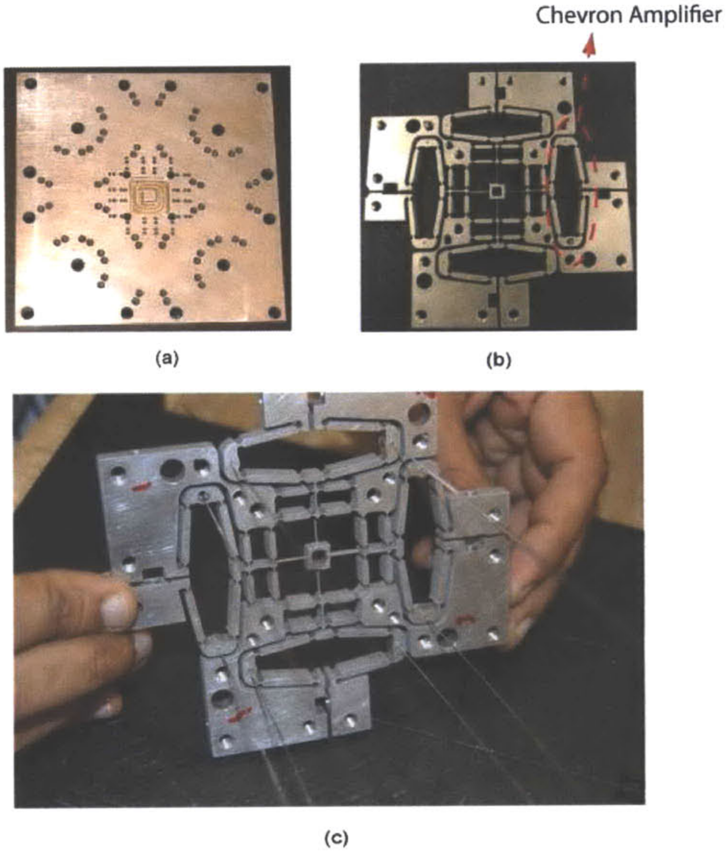


Figure 6-7: (a) Initial template on which the flexure is subsequently waterjetted to obtain the part shown in (b). The flexure piece has four chevron amplifiers, one of which is highlighted in the figure for illustration.(c) Rod flexures inserted at the end of chevron amplifiers in the flexure piece to curb twisting modes of the chevron.

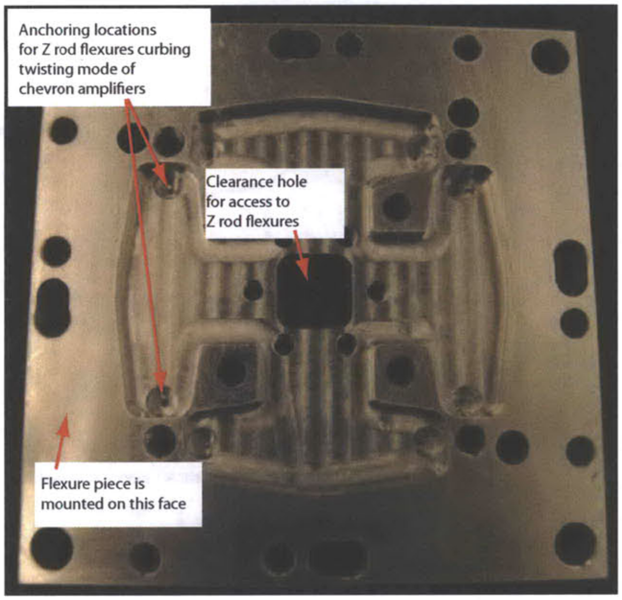


Figure 6-8: Hardware photograph of top view of base plate, which has a footprint of 4.5 *in* × 4.5 *in*. The flexure is mounted on this face of the base plate.

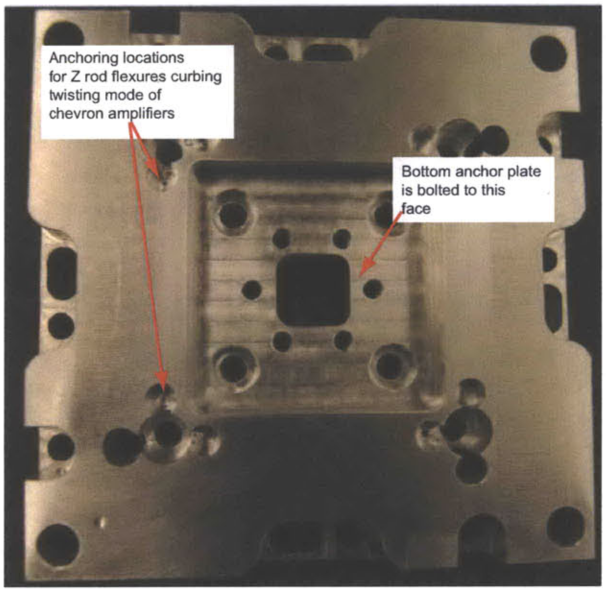


Figure 6-9: Hardware photograph of bottom view view of base plate. The sixty four Z rod flexures curbing the trampoline mode of the flexure are bonded to an anchor plate that is bolted to a central recessed portion on this face.

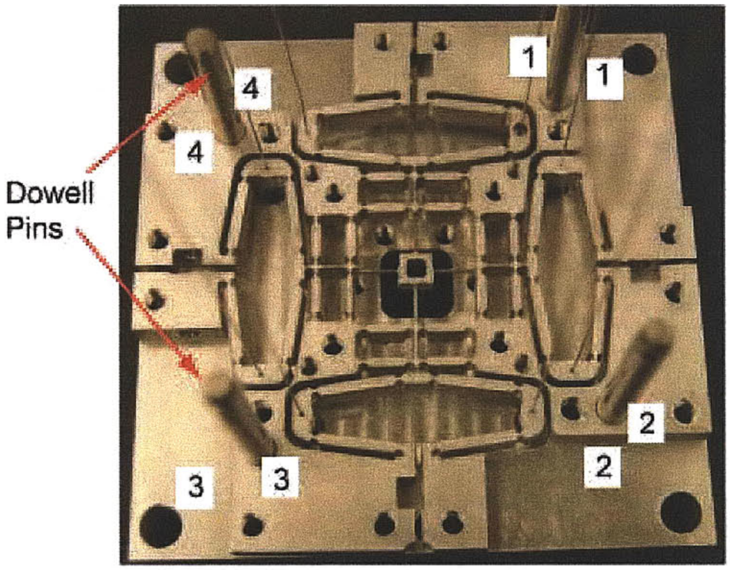


Figure 6-10: Hardware photograph of the flexure nanopositioner mounted on the base plate with registration achieved using dowell pins of .25 *in* diameter.

After the assembly is completed, the rod flexures are bonded in place with a high vacuum epoxy[2]. The whole assembly is baked in an oven at 80°C for three hours to allow for curing and subsequent hardening. The cured and hardened assembly is shown in Fig. 6-11.

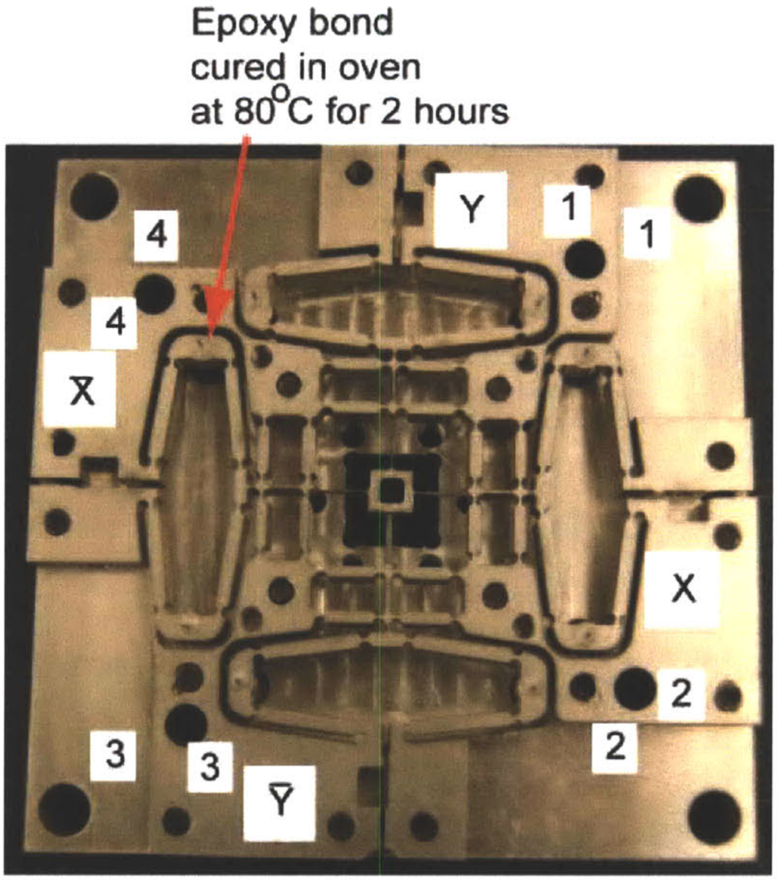
6.4.2 Central Z rod flexures subassembly

As explained in Section 6.3.1, a forest of rod flexures are needed to curb the trampoline mode of the flexure piece and ensure the z natural frequency is as high as possible.

Hardware photograph of the bottom plate anchoring the Z rod flexures to the base plate is shown in Fig. 6-12. Sixty four Z rod flexures were assembled into an area of about 18 mm × 18 mm. Each Z rod flexure is a spring steel cylinder of diameter 500 μm and length 9.2 mm. The rods are anchored in place using two plates, called the 'top anchor plate' and the 'bottom anchor plate' in Fig. 6-12. Note that machining the rods to a precise length is a tedious procedure. Instead our design

---

[2] High vacuum epoxy from Ideal Vacuum Products, LLC., refer to Appendix E.



Epoxy bond cured in oven at 80°C for 2 hours

Figure 6-11: Hardware photograph after Z rod flexures anchoring the chevron ends are bonded by high-strength epoxy. The whole hardware assembly is baked in an oven at $80°C$ for three hours to allow for curing and subsequent hardening.

is based on sandwiching the rods between the anchor plates that are separated by a precise distance.

The assembly steps are follows. First, the Z rod flexures are inserted through both the anchor plates. Sixty four Z rod flexures were assembled into an area of about $18\ mm \times 18\ mm$. The ends closer to the top anchor plate are bonded in place with high-strength epoxy and stripped down to be flush with the rest of the face of the anchor plate. (This anchor plate is subsequently bonded to the bottom of the central stage of the AFM nanopositioner, and hence any rods must not protrude out.)

A few (3-4) rod flexures are left at each corner while the rest are stripped down to be flush with the top face of the anchor plate, as shown in Fig. 6-12(b). The rods at the corners are used to manipulate the anchor plate when aligning it to the central stage under an optical microscope. The precise registration under a microscope is necessary to ensure that the Z rod flexures are positioned symmetrically at the center of the XY central stage.

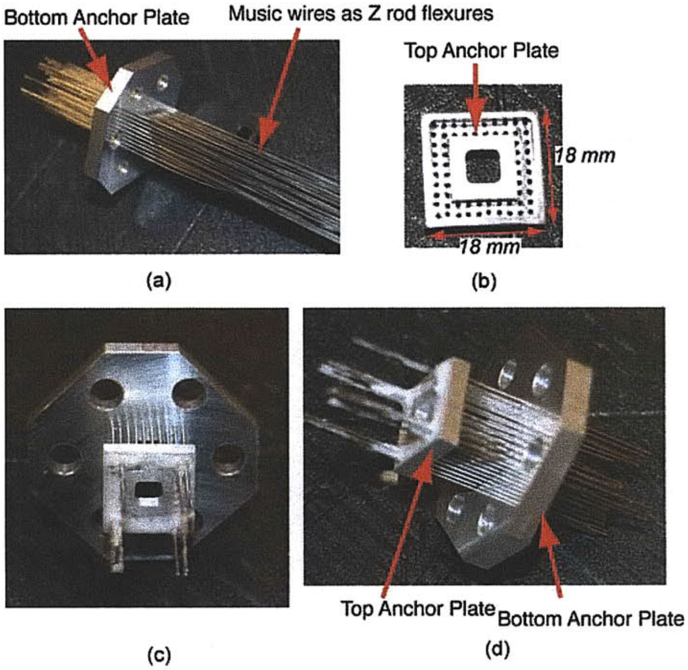


Figure 6-12: Hardware photograph of bottom plate anchoring the Z rod flexures to the base plate.(a) Z rod flexures inserted into bottom anchor plate. (b) Top anchor plate. (c) Z rod flexures inserted into the top anchor plate and bonded with epoxy. (d) A side view of the Z rod flexure assembly.

Figure 6-13: Optical registration against crosswires of a microscope allows for XY alignment of the Z rod flexures before they are bonded to the central stage.

6.4.3 Mounting Piezoelectric Stack Actuators

Piezoelectric stack actuators[3] with a maximum range of 18 μm (at no-load) and blocking force of 850 N were selected for driving the nanopositioner.

The piezoelectric stack actuators were mounted into the nanopositioner as follows. Two piezoelectric stack actuators were bonded back-to-back with a high-strength epoxy [4] are lowered in the angular recess of the chevron amplifier. Once the piezoelectric stack actuators are lowered in place, as shown in Fig. 6-14, a screw on the distal end of chevron amplifier is tightened to cause a press fit and preload the mechanical contact between the expanding piezoelectric stack and the chevron amplifier. To avoid flat on flat contact that can cause shear loads on the piezoelectric actuator, the chevron amplifier is provided with a circular cut. Thus, the actuator and chevron amplifier interface simulates a cylinder (which can be easily made with abrasive waterjetting) on flat configuration. An alternative to this approach is to use piezoelectric

---

[3] Product number AE0505D16 from NEC Tokin, refer to Appendix E

[4] High vacuum epoxy from Ideal Vacuum Products, LLC., refer to Appendix E.

actuators with a hemi-spherical tip[5]. In our case, we found the cylindrical design acceptable, since moment or vertical shear loads at the interface will be likely to produce less effect on the piezoelectric actuators. Most of the vertical shear load would be taken by the music wire rod flexures on the ends of the chevron amplifiers.

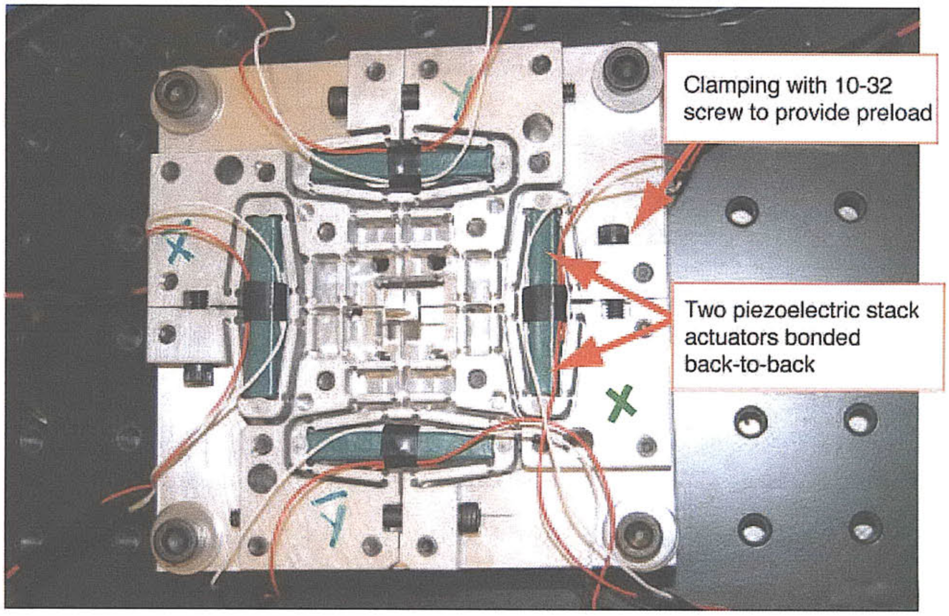


Figure 6-14: Four sets each containing two piezoelectric stack actuators bonded back-to-back are mounted into the nanopositioner and preloaded by tightening a #10-32 screw on an end of the chevron amplifier.

A photograph of the final assembled hardware is shown in Fig. 6-15. The overall dimensions of the nanopositioner assembly are 4.5 *in* × 4.5 *in* × 1 *in*.

6.5 Instrumentation and Testing Setup

The voltage rating of the actuators is $0 - 150\ V$. The capacitance value for the actuators are rated at $1.5\ \mu\ F$. Driving the capacitance load of these actuators to high frequencies necessitated the use of custom-built high-speed power amplifiers from our collaborators [76]. These amplifers are rated to supply $200\ V$ at maximum

---

[5] Physik Instrumente http://pi.com provide piezoelectric stack actuators with a spherical tip.

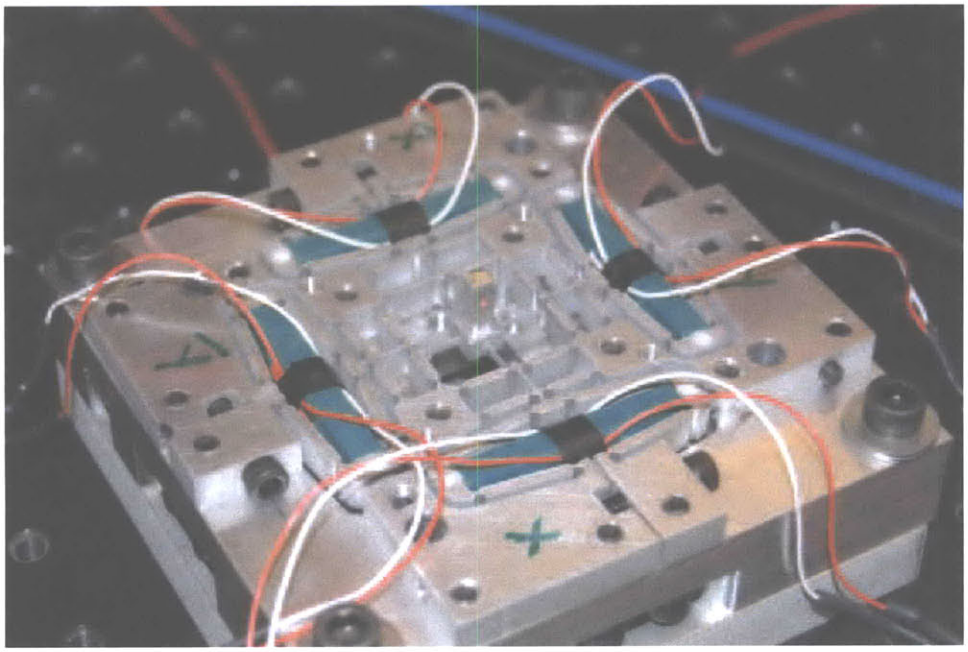


Figure 6-15: Hardware photograph of nanopositioner built as a case study for design for control approach developed in the thesis. The footprint of the device is about 4.5 in $\times$ 4.5 in.

currents of 8 A, up to frequencies of $> 100\ kHz$ for a $1.5\mu F$ load. These specifications meet our needs for driving the piezoelectric actuator.

A point to note here is that piezoelectric actuators cannot contract. The displacement of the central stage of the flexure nanopositioner in, say, the horizontal (i.e. X direction) can hence be achieved as follows. A DC offset voltage V_o of 75 V is chosen at half of the maximum value $V_{max}(= 150V)$, which corresponds to the maximum extension of the piezoelectric actuators housed in the chevron mechanisms. The voltage on the X set of piezoelectric actuators is gradually increased from V_o to V_{max}, while the voltage on the $-X$ set of actuators is simultaneously reduced from V_o down to 0. The displacement of the central stage achieved as described above in the X direction, can also be simultaneously achieved in the Y direction.

A DC offset voltage of 75 V is the default the output of the power amplifier. This voltage is chosen as the baseline V_o discussed above. The amplifier is specifically tuned for the capacitive load of the piezoelectric actuators. The amplifier has a constant gain of 7.5 over the frequencies of interest $(0 - 100\ kHz)$. So a maximum of 10 V can be input to the amplifier. If the input voltage exceeds this limit, the piezoelectric actuator will be subject to a voltage overloading, which could over time potentially result in the loss of its piezoelectric properties, and irreversible damage.

A laser interferometric vibrometer with a factory-calibrated resolution of 1 nm is used as a sensor for our testing. The instrument comes with capabilities for both an interferometer (which can be used as a feedback sensor) with update rates of up to 500 kHz. The interferometer is provided with temperature and humidity sensors that interface with the measurement unit and allow for correcting for the drift in the measurements. The laser interferometer needs a direct line of vision and smooth reflective surface on the target, hence, a $4\ mm \times 4\ mm \times 2mm$ metallic post was fixed to a recess in the central stage. To avoid significant changes in natural frequencies of the scanner measured using the post, the height of the post was kept at the bare minimum needed for grazing incidence of the laser beam.

A typical hardware setup used for measuring the nanopositioner performance is shown in Fig. 6-16.

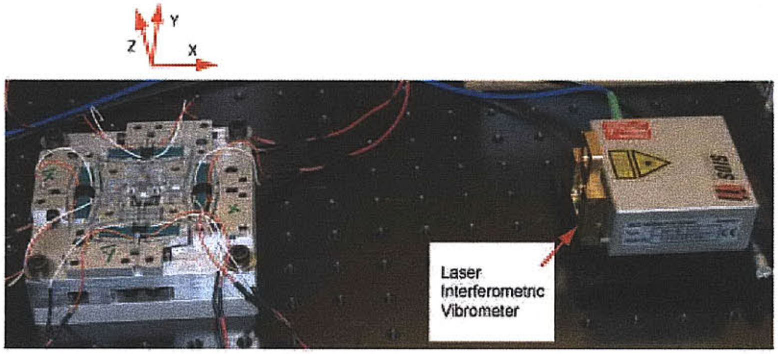


Figure 6-16: Nanopositioner assembled on an optical vibration isolation table and sensed by a laser interferometric vibrometer for both static and dynamic measurements. A temperature and humidity sensor provided with the laser interferometeric vibrometer enables driftless measurement.

For characterizing frequency response between piezoelectric actuator voltage inputs and displacement measurements of the laser vibrometer, we used an Agilent

Dynamic Signal Analyzer 35670A that could provide sine-sweep, chirp, random noise and various other sources of excitation over the frequency range 50 Hz to 50 kHz. Another method for the same included using a LABVIEW-based [6] stochastic pseudo-random binary system identification code. With this method the LABVIEW VI's generated PRBS signals that were fed through a National Instruments hardware multifunction data-acquisition board (m-DAQ) 6048 to the power amplifiers driving the piezoelectric stack actuators. Analog voltage outputs of the laser interferometric vibrometer were fed back to the input channels of the m-DAQ 6048. Since the m-DAQ 6048 has an update rate of 200 kHz, the measurements were limited to upto the Nyquist sampling rate of 100 kHz. The post-processing of the LABVIEW data using FFT-based methods was performed both in Labview and the system identification toolkit of MATLAB.

The control design was performed on a dSPACE DS1103 platform, a rapid controller prototyping platform that interfaces with MATLAB/Simulink to create C-code that runs a real-time DSP provided with the dSPACE controller board. The input channels of DS1103 have an ADC resolution of 16bit and were run at update rates of upto 40 kHz.

6.6 Static and Dynamic Performance: Experimental Characterization

A number of static and dynamic tests were performed on the flexure nanopositioner assembly. The instrumentation described in Section 6.5 were used to measure the lateral range of motion, frequency response functions between voltage applied by the piezoelectric stack actuator and output displacements measured by the laser interfer-

---

[6] The LABVIEW VI's for the system identification were written by Daniel J. Burns, a colleague in the Mechatronics Research Laboratory, Department of Mechanical Engineering, MIT.

ometric vibrometer, and also the cross-coupling.

6.6.1 Static Range Measurements

The results of testing performed for static range of the nanopositioner are shown in Fig. 6-17. The data were recorded for a driving voltage $V_2 - V_1$ in the range $0 - 90V$. The measured maximum displacement δ_s for $V_2 - V_1 = 90V$ is about 30 μm. The maximum deviation of the data from the fit, as indicated by the error bars, is 0.39 μm. As predicted from the least squares fit, the maximum range at full driving voltage $V_2 - V_1 = 150V$ is 49.95 ± 0.39 μm, which is a value close to our target requirement of 50 μm.

The sources of error in the data include inherent nonlinearities such as cross-coupling error motions due to axial shortening of the flexural beams used in the design, piezoelectric actuator hysteresis and creep, and thermal errors in the measurement path arising from mismatched coefficients of thermal expansion. The performance loss in terms of actual imaging will be an intersting aspect to study after an optical subsystem is integrated to the current nanopositioner prototype.

6.6.2 Frequency Response Measurements: Out-of-plane

An important point to be noted here is that mere fine-tuning of this topology, i.e. the XY planar flexure without the Z rod flexures, can not generate a design that can be suitable for controlling Z to high servo bandwidths on the order of $10 - 50$ kHz. To re-emphasize the argument, from an optimization view point, inputting this design to an optimization routine will be a futile exercise unless the optimization routine revises not just the values of beam thicknesses or heights, but also the topology.

The hardware setup showing the nanopositioner fixture assembled with its Z axis aligned to laser beam of the interferometer is presented in Fig 6-18. The frequency response measured for a the topology with the Z rod flexures bonded to the central

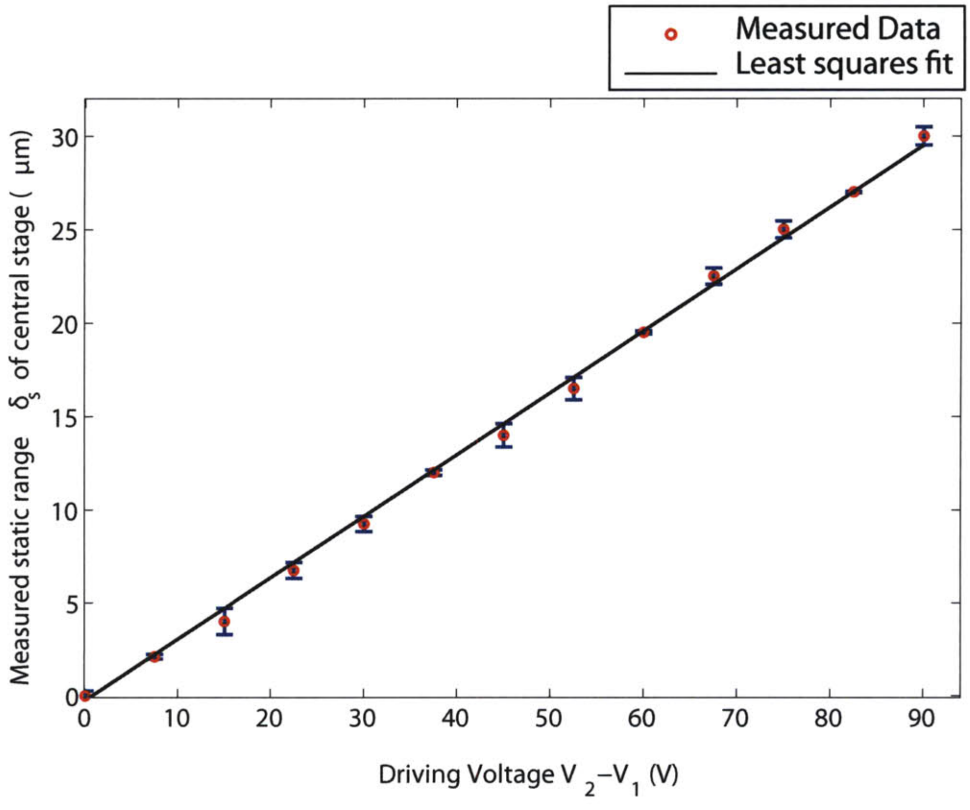


Figure 6-17: Measured data for the lateral range of motion δ_s of the nanopositioner for a driving voltage $V_2 - V_1$ in volts. A least-squares fit of the data points is also shown. The measured maximum displacement δ_s for $V_2 - V_1 = 90V$ is about 30 μm. The maximum deviation of the data from the fit, as indicated by the error bars, is 0.39 μm. As predicted from the least squares fit, the maximum range at full driving voltage $V_2 - V_1 = 150V$ is 49.95 ± 0.39 μm.

stage is shown in Fig. 6-18. To record this measurement a 3 $mm \times 3$ mm mm piezoelectric actuator with a fundamental resonance frequency of > 300 kHz was bonded on top of a 4 $mm \times 4$ $mm \times 2$ mm post fixed on to the central stage. The measured resonance frequency in the Z frequency response function shows a fundamental mode at 33.1 kHz and a second mode at 56 kHz. The fundamental Z mode of the scanner is hence increased 22 fold by bonding the Z rod flexures. This data is promising, and to the best of our knowledge, this makes our nanopositioner the largest in static range with a high fundamental Z mode. The comparison of our scanner in terms of different measured static and dynamic properties is summarized in Table 6.1.

6.6.3 Frequency Response Measurements: In-plane

The in-plane frequency response measured for the flexure nanopositioner with the Z rod flexures bonded to the central stage is shown in Fig. 6-18. To record this measurement a 4 $mm \times 4$ $mm \times 2$ mm aluminium post was fixed on to the central stage to provide a reflecting target for the laser beam. The measured resonance frequency in the Z frequency response function shows a fundamental mode at 750 Hz and a second mode at about 2.33 kHz. An antiresonance mode (or complex zero pair) is seen at about 800 Hz. Physically, this zero pair corresponds to the frequency where the central stage remains almost stationary and all the energy input by the piezoelectric actuators is absorbed in the chevron amplifier and other locations of the flexure.

To test for the uncertainty in the measurement data, the response was captured for repeated trials. As seen in the multiple data sets, the uncertainty in the data is limited to a bounding envelope about 5 dB in the magnitude plot, and about 25° in phase. The uncertainty in the model that concerns us most is centered about the first resonance mode (complex pole pair) and the antiresonance (complex zero pair). This

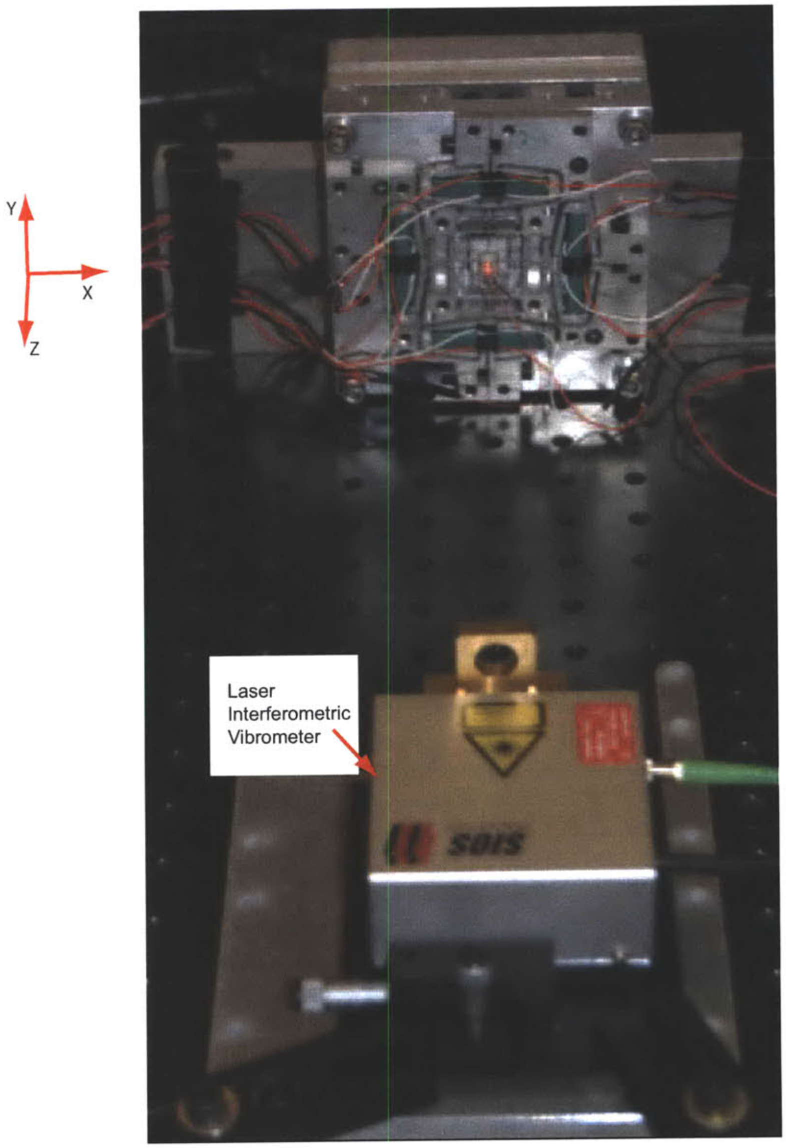


Figure 6-18: Nanopositioner assembled with its normal aligned to laser interferometer for measuring out of plane vibration mode. A 3 mm × 3 mm × 2 mm piezoelectric actuator with a fundamental resonance mode of $> 300\ kHz$ is bonded to a post fixed to the center stage. The frequency response characterization is limited to sampling rates of 200 kHz with a NI m-DAQ card 6048. The resonance from the piezoelectric actuator does not show up in the data collected up to the Nyquist frequency of 100 kHz.

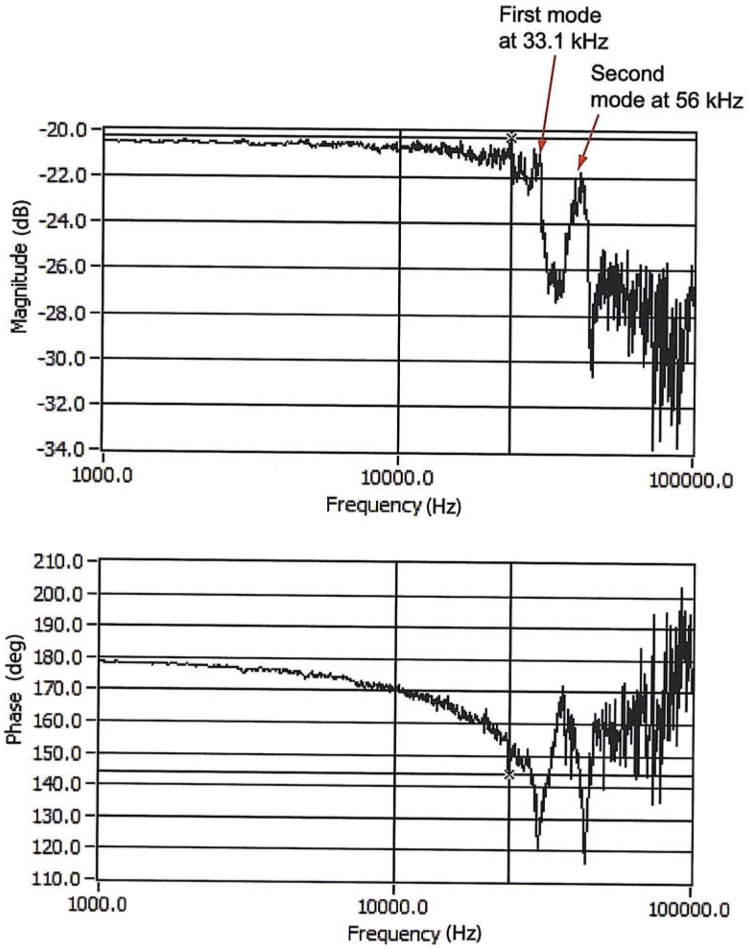


Figure 6-19: Measured frequency response of the transfer function between voltage applied to the z-direction piezoelectric stack actuator and the z-deflection of the flexure nanopositioner. The fundamental Z mode appears at a frequency of 33.1 kHz, while a second dominant mode appears at 56 kHz. The 180^{circ} phase seen in the raw data at DC is an artifact that comes from the laser interferometer output and piezoelectric stack actuator input being completely out-of-phase, i.e. when the Z-direction piezoelectric actuator expands, the gap between the target and the laser interferometer reduces.

uncertainty at the complex zero pair is expected due to the low signal to noise ratio at that frequency.

Note that this frequency response plot is dependent on the preload applied to the flexures using the bolted split clamp in our design. In our experiments, the preload of the flexure was not changed from the set initial value, which is a function of the opening of the split clamp and the number of turns of the clamping screw. Given the repeatability of the resonance frequencies in the measured data, the preload on the chevron amplifiers did not seem to vary appreciably with time over about a week during which the data in Fig. 6-20 was recorded.

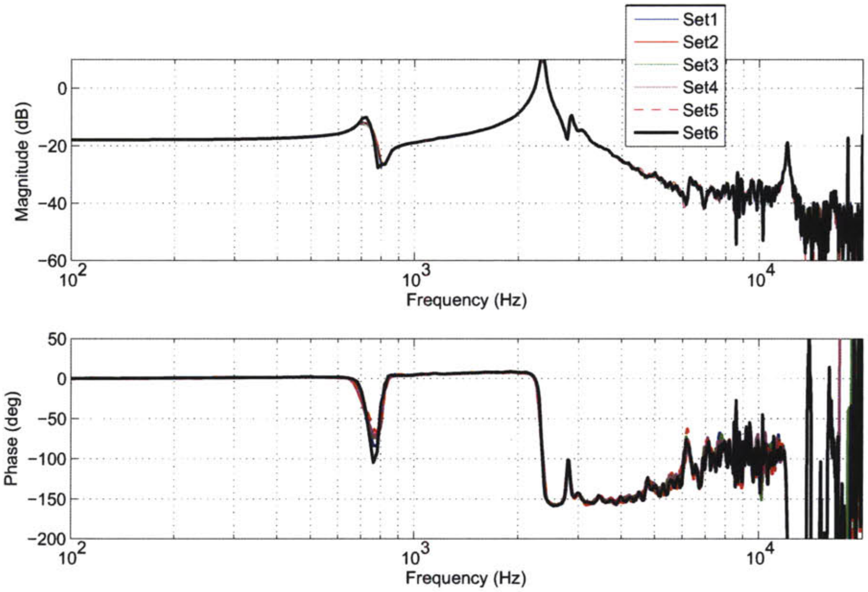


Figure 6-20: Measured in-plane frequency response function of the flexure nanopositioner after Z rod flexure subassembly is bonded to the central stage.

6.6.4 Cross-axis Coupling

Cross-axis coupling motions refer to the error motions that occur along any one axis, when another axis of the system is actuated. The question of coupling between motion axes is critical, since it affects the behavior of most multi-DOF flexure systems.

To characterize the coupling in the AFM nanopositioner, we captured the fre-

Table 6.6: DC Gains of the transfer matrix $\mathbf{P}(j\omega)$.

	$Magnitude$	dB $= 20log_{10}(Magnitude)$
$\lvert Pxx(0)\rvert$	0.1281	-17.85
$\lvert Pxy(0)\rvert$	8.44×10^{-4}	-61.47
$\lvert Pyx(0)\rvert$	0.0018	-55.06
$\lvert Pyy(0)\rvert$	0.053	-17.85

quency responses of the transfer functions between the voltage applied to the X and Y piezos, and the measured displacements along X and Y. There are four possible combinations, as follows: $Pxx(j\omega)$ is the frequency response obtained for actuation and sensing along X. Similarly, $Pyy(j\omega)$ is the frequency response obtained for actuation and sensing along Y. $Pxy(j\omega)$ is the frequency response obtained for actuation along Y and sensing along X. Finally, $Pyx(j\omega)$ is the frequency response obtained for actuation along X and sensing along Y. A 2×2 matrix of these responses is referred to as the transfer matrix, which is given as follows:

$$\mathbf{P}(j\omega) = \begin{bmatrix} Pxx(j\omega) & Pxy(j\omega) \\ Pyx(j\omega) & Pyy(j\omega) \end{bmatrix}$$

The experimentally measured coupling data for the frequency response of the above four transfer functions are as shown in the graphs of Fig. 6-21. The DC gain[7] values measured from the frequency response plots are presented in Table 6.6.

The static decoupling from X to Y defined as $\frac{|Pxy(0)|}{|Pyy(0)|}$ is 0.033, which corresponds to $-29.55\ dB$. Similarly, the static decoupling from X to Y, defined as $\frac{|Pyx(0)|}{|Pxx(0)|}$ is 0.0066, which corresponds to $-46.3\ dB$. These values indicate that the design topology is incorporated with redundancies that, in spite of inevitable manufacturing errors, make the system appear decoupled to a large extent. This was possible because of

---

[7] DC gain of a transfer function is defined as its magnitude at DC, i.e. at the frequency $\omega = 0$.

the inherent modal separation between X and Y achieved with a judicious choice of flexure constraints in the design topology. The static decoupling makes the design suitable at low frequencies for nearly control of the multi-axis system as two decoupled single-input single-output (SISO) systems. However, at higher frequencies, such as around $2-3\ kHz$, the coupling becomes relatively significant, the peak responses in the off-diagonal frequency response functions almost reaching with a couple of dB of the DC gains of their diagonal counterparts.

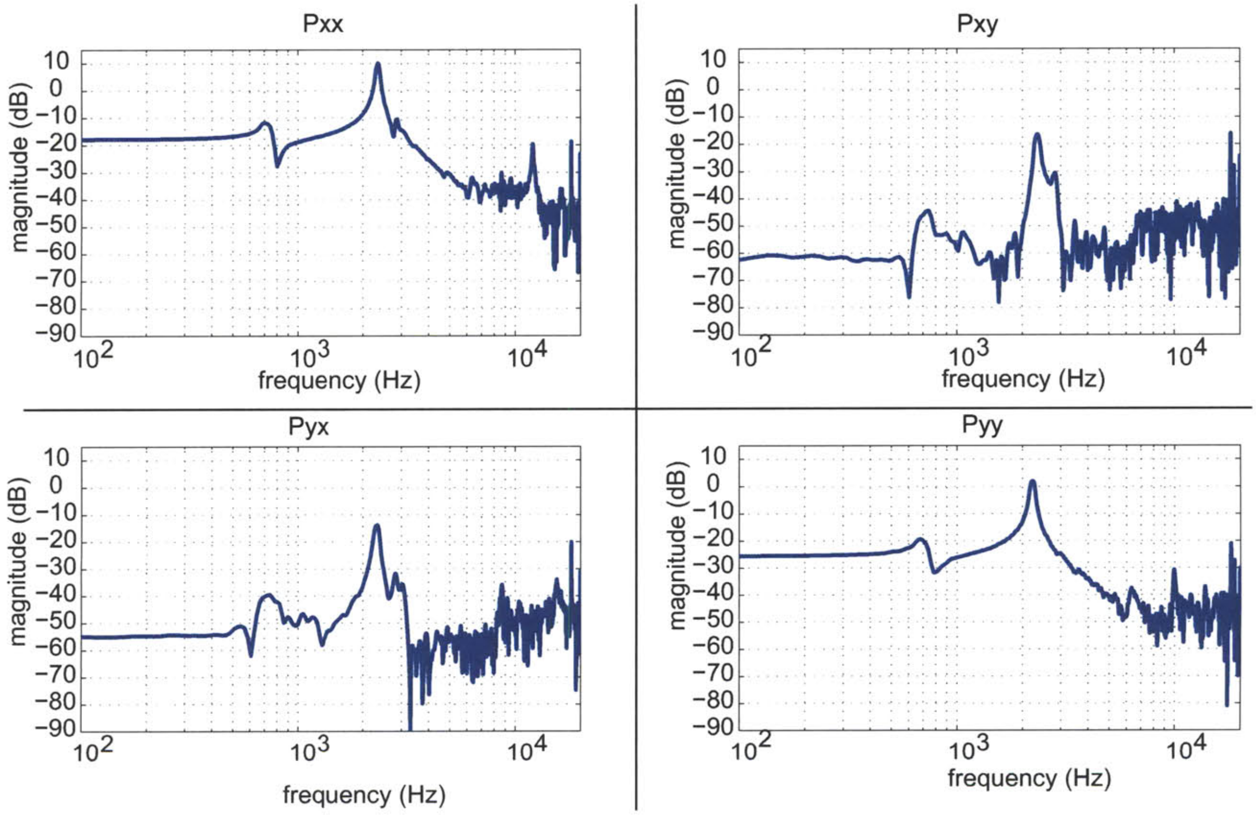


Figure 6-21: Experimentally measured frequency response functions depicting the plant transfer matrix. The off-diagonal frequency response plots indicate the level of cross-coupling in the system. As can be inferred from the DC gains of this figure, by virtue of our design, the plant has static decoupling of 30 dB from X to Y, and decoupling of about 44 dB from Y to X. However, dynamic coupling at resonance frequencies of the structure are evident and need to be addressed in the control formulation.

Given the above frequency responses, a performance metric that quantifies the coupling is desired. One such metric could be the relative gain array, introduced by Bristol [67] in 1966. This metric is a function of frequency ω and is defined as $\mathbf{P}(j\omega) * (\mathbf{P}^{-1})^T(j\omega)$, where $(*)$ denotes the Schur product, i.e. an element-by-element product[8] of the two matrices $\mathbf{P}(j\omega)$ and $(\mathbf{P}^{-1})^T(j\omega)$, where $\mathbf{P}(j\omega)$ is the transfer matrix of the plant.

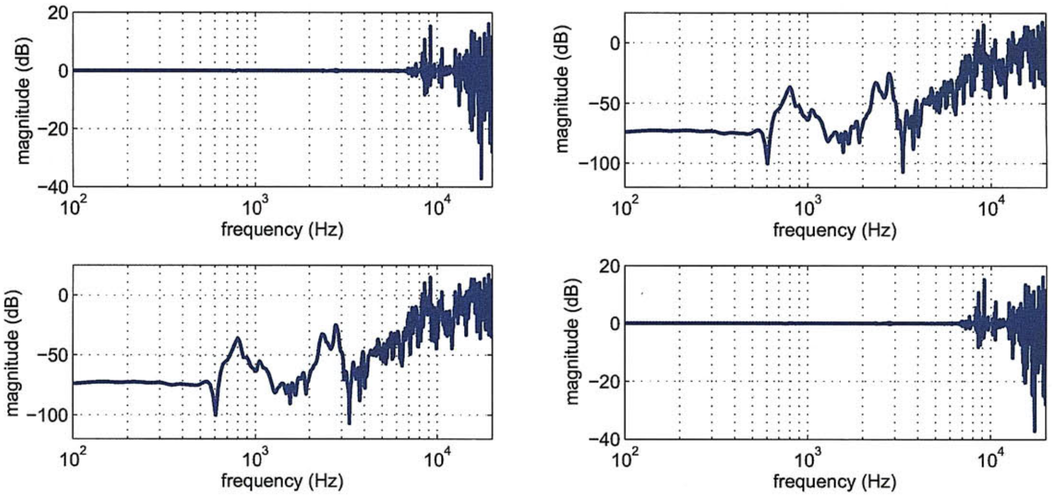


Figure 6-22: Relative Gain Array matrix calculated for the system from experimental frequency response data of the XY transfer matrix $\mathbf{P}(\mathbf{j}\omega)$. At frequencies higher than about 5 kHz the poor signal to noise ratio makes the measurements unreliable, and hence the effect of these data are ignored.

The relative gain array matrix as a function of frequency for the measured plant frequency response data for our nanopositioner is plotted in Fig. 6-22. This matrix should ideally look like a diagonal matrix with unity gain in the frequencies of interest in the diagonal terms. The frequency content of the off-diagonal terms indicates the extent of coupling. In our case, the diagonal terms are close to a magnitude of 1, i.e. 0 dB. The dynamic coupling between the axes shows up at frequencies around 700 Hz and 2.3 kHz. This means energy input into any axis is leaking into the other. Note the symmetry in the data, the off-diagonal frequency response functions $Pxy(j\omega)$ and $Pyx(j\omega)$ look almost identical to each other. The coupling from X to

---

[8] The Schur product of two matrices A= $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and B = $\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$ is given by $A. * B=$ $\begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} \\ a_{21}b_{21} & a_{22}b_{22} \end{bmatrix}$. In MATLAB, we use $.*$ to implement a Schur product.

Y is almost identical as that from Y to X. This inherent symmetry in the data may be resulting from the symmetry achieved in the design.

Currently used methods in the field of precision engineering incorporate the coupling between motion axes in the design error budgets. However, the methods used to characterize the coupling are primarily static. The relative gain array provides a snapshot of the dynamic coupling and could be incorporated into the dynamic error budgets in the design of flexure systems.

As indicated by the low DC gain of the off-diagonal transfer functions $Pxy(j\omega)$ and $Pyx(j\omega)$ in Fig. 6-21, static decoupling has been addressed with using symmetry in the design of the nanopositioner. However, the dynamic coupling that shows up at higher frequencies cannot be addressed by design. There are many ways to minimize the cross-coupling using control approaches. One approach is using decentralized feedback control with high disturbance rejection, as the coupling effects can be regarded as a form of disturbance. This approach is essentially one of sensitivity tuning, which is addressed in Section 6.7.

Prefiltering is another approach to handle the cross-coupling problem. In this approach, the cross-coupling motions are minimized by placing a filter on the inputs driving the undesirable motion axis. Prefiltering is one of the different ways of performing feedforward control, and offers the advantage that it can be performed open-loop without the use of a feedback sensor.

A simple prefilter can be constructed from a low pass filter. When the inputs to the X axis are passed through after prefiltering by a low pass filter, the frequency content of the filtered inputs beyond the cut-off frequency of the low-pass filter are attenuated by an amount dictated by the slope of the frequency roll-off of the filter. The decoupling performance shown in the frequency response of Fig. 6-23 was obtained using a prefiltering approach with a first-order low pass filter with a cut-off frequency of 200 Hz. The data in the figure was collected with actuation along Y

and measurement made along X direction. The data obtained using the prefilter is indicated as $P^*xy(j\omega)$. With the prefiltering, the attenuation in the magnitude gain around $2-3\ kHz$ in the $P^*xy(j\omega)$ relative to that of $Pxy(j\omega)$ is given by:

$$20log_{10}\left\{\sqrt{\frac{1}{\tau_c^2 w_p^2+1}}\right\} \tag{6.15}$$

where $\tau = \frac{1}{2\pi f_c}$ and w_p is the frequency at which the attenuation is desired. For values $w_p = 2\pi(2000)\ rad/s$, and filter cut-off frequency $f_C = 200\ Hz$, attenuation calculated from Eq. (6.15) is $20\ dB$, which matches closely with the difference in the peaks of $Pxy(j\omega)$ and $P^*xy(j\omega)$ in Fig. 6-23.

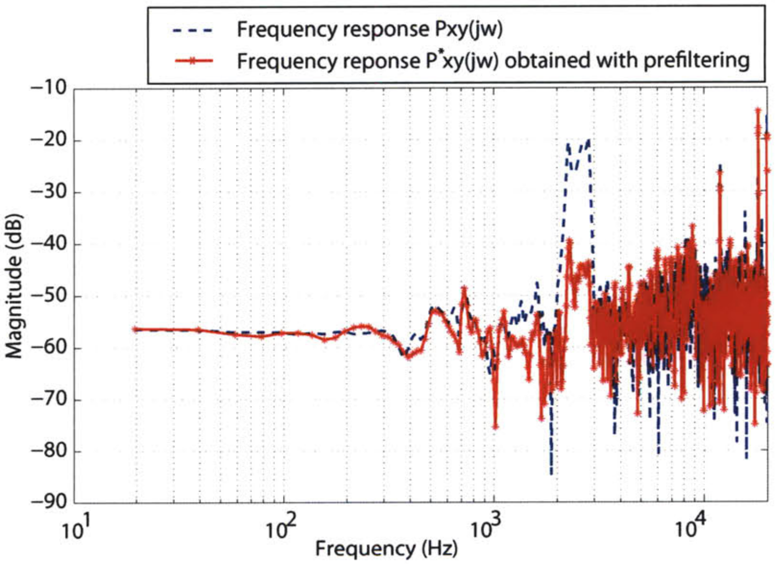


Figure 6-23: Attenuation of the cross-coupling peak in the frequency response of the transfer function $Pxy(j\omega)$ by about $20\ dB$ is achieved with a prefilter acting on inputs driving the Y axis.

The limitation of a prefiltering approach is that the plant model should be reliably known. To impart robustness to modeling uncertainty, the order of the prefilter can be increased. However, increasing the order of the prefilter results in increasing the phase lag between the input and the response of the system [58]. As alternative to prefiltering, feedback control can be applied with tuning of the sensitivity for disturbance rejection, as we will be discussing in the following section.

6.7 Closed-loop Sensitivity Tuning: Simulations and Experiments

Our goal here is to illustrate for the AFM nanopositioner a controller generation method based on sensitivity tuning. As described in Chapter 4, tuning of the sensitivity transfer function is ciritcal in achieving robustness and disturbance rejection. For tuning the sensitivity transfer function, we use the result Eq. (4.17) of Chapter 4, which is reproduced here:

$$S(s) = S_0(s)Q(s) \tag{6.16}$$

Recall that the above relation holds because of the novel controller parameterization that we developed in Section 4.3.1. The simple multiplicative relationship between nominal sensitivity and the $Q(s)$ parameter transfer function allows us to use model-matching and filter-design techniques for tuning $S(s)$.

In other words, we will be searching for the $Q(s)$ parameter transfer function over the class of stable transfer functions $\mathcal{RH}_\infty$[9] so that a chosen norm of the error between the frequency response of the sensitivity transfer function and that of a reference or desired transfer function gets minimized. A formulation using a $\mathcal{H}_2$ norm for the error is as given below:

$$Q(s) = \arg \min_{Q(s)\epsilon\mathcal{RH}_\infty} ||S_d(s) - S_0(s)Q(s)||_2^2 \tag{6.17}$$

where $S_d(s)$ is a desired sensitivity transfer function.

---

[9] We define $\mathcal{RH}_\infty$ as the set of rational transfer functions that are stable (i.e. they have all poles in $Re(s) < 0$) and proper, i.e. they are bounded-input-bounded-output (BIBO) stable.

The rest of this chapter is on formulating, solving, and testing a model-matching problem for tuning a given sensitivity transfer function. By tuning, we would like to make the closed-loop sensitivity transfer function match well with a sensitivity transfer function that, for example, has an increased robustness to modeling uncertainties. First, we develop a plant model that best fits (in a $\mathcal{H}_2$-norm sense) the experimentally captured dynamic behavior. Then, we select a nominal controller which when used to implement a closed-loop control system ensures internal stability, i.e. all internal signals are bound and there are no RHP pole-zero cancellations[10]. The nominal sensitivity may not have the desired features such as a desired level of robustness. Hence, we tune it to match it with a desired sensitivity transfer function that meets the robustness requirement, and we will examine the effects on the resultant control performance. Experimental results performed on the AFM nanopositioner hardware will be discussed and compared with the simulations.

6.7.1 Plant Model

In this section, we derive a plant model from a $\mathcal{H}_2$-optimal fit of the measured dynamic frequency response behavior of the nanopositioner. An example set of frequency response data for the lateral motion of the AFM nanopositioner are shown as discrete dots in Fig. 6-24. This set of data, along with similar frequency response characterizations we discussed earlier in Section 6.6, represent the frequency response of the transfer function between the voltage applied to the X set of piezoelectric stack actuators and the corresponding central stage displacement along the X axis.

The displacement was measured by the laser interferometer whose measurement bandwidth is on the order of 500 kHz. Since this frequency lies well beyond the frequency range of 20 Hz $-$ 20 kHz, we will be ignoring here the dynamics of the laser interferometer. Similarly, the power amplifier used for driving the piezoelectric

---

[10] Refer to an explanation on internal stability given in Section 4.3.1 and Appendix A.

actuator has a bandwidth on the order of 100 kHz and hence its dynamics will also be ignored.

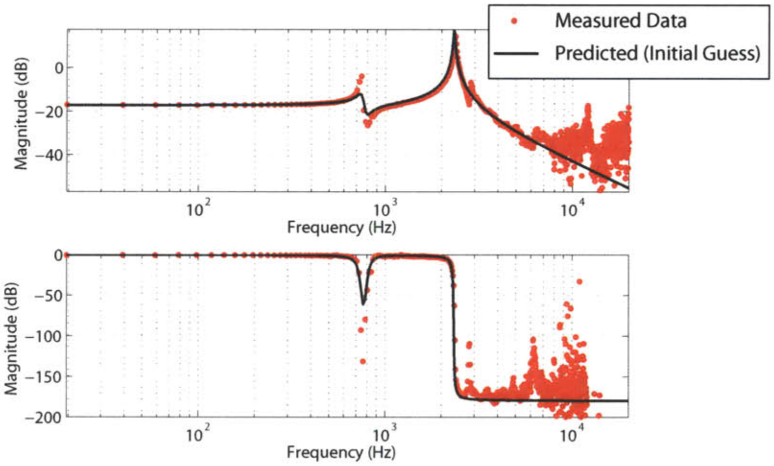


Figure 6-24: Frequency response plot of transfer function between voltage applied to piezoelectric stack actuator and lateral (X) displacement of the central stage measured with a laser interferometer. Measured data are shown with dot markers. An initial guess of a model obtained by trial and error is shown as a dotted line.

Based on the frequency response data, we consider two dominant modes, one at about 750 Hz and another at about 2330 Hz, as capturing the behavior of the nanopositioner. Using the sum of modal dynamics, as depicted in Chapter 2, the dynamic model for the plant can be expressed as the following superposition of the two modes:

$$P(s) \triangleq \frac{X(s)}{V_2(s) - V_1(s)} = \alpha \left\{ \frac{\omega_1^2}{s^2 + 2\xi_1\omega_1 s + \omega_1^2} + \beta \frac{\omega_2^2}{s^2 + 2\xi_2\omega_2 s + \omega_2^2} \right\} \tag{6.18}$$

where ξ_1 and ξ_2 denote the damping factors, ω_1 and ω_2 denote the undamped natural frequencies associated with the two modes, β denotes a relative modal participation factor that denotes the dominance of one mode over the other in the overall plant response, and α is a scalar constant that affects the DC gain.

For an initial guess of values $\xi_1^0 = 0.04$, $\omega_1^0 = 2\pi(750)\ rad/s$, $\xi_2^0 = 0.008$, $\omega_2^0 = 2\pi(2330)\ rad/s$, $\alpha^0 = 0.0138$, and $\beta^0 = 9$, a nominal plant $P_0(s)$ was derived from Eq. ((6.18). The frequency response predicted using $P_0(s)$ is shown as a solid line in Fig. 6-24. With this initial guess, the plant model matches closely with the measured frequency response data. However, it is not an optimal fit for the given data. An optimal fit is desired so as to accurately capture the plant model for subsequent controller design.

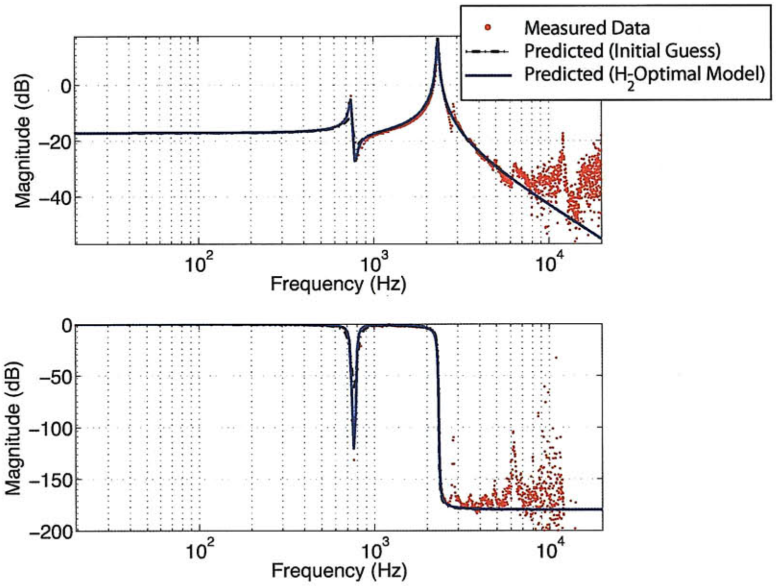


Figure 6-25: Measured data and predicted models for the frequency response plot of transfer function between voltage applied to piezoelectric stack actuator and lateral (X) displacement of the central stage measured with a laser interferometer. Measured data are shown with dot markers. The response predicted by the initial guess of Fig. 6-24 is shown as a dotted line. This initial guess was fed to a nonlinear optimization routine that minimized a norm of the error between the responses of the measured and the predicted model. The frequency response of the optimal plant model is shown as a solid line.

To achieve an optimal fit of the model to the experimental data, we define a cost function that represents a $\mathcal{H}_2$ norm of a weighted error transfer function between the measured and the predicted frequency response as follows:

$$\text{Cost}: \quad \int_0^{\omega_f} ||\left\{P_{exp}(j\omega) - P(j\omega)\right\} W(j\omega)||_2^2 d\omega \tag{6.19}$$

where $P_{exp}(j\omega)$ is the experimentally measured frequency response data, $P(j\omega)$ is frequency response generated from the model (which is a function of the optimization parameters), $W(j\omega)$ is a weighting filter that is designed to reduce the effect of poor data coherence at high frequencies, and ω_f represents the maximum frequency for which the experimental frequency response data were collected.

For a vector of optimization parameters $\mathbf{U} = [\xi_1;\ \omega_1;\ \xi_2;\ \omega_2;\ \alpha;\ \beta]$ who values are bound between the limits $\mathbf{U}_{min}$ and $\mathbf{U}_{max}$, we also impose an equality constraint that fixes the DC gain of the optimal response to be the experimentally measured value. This equality constraint is given below:

$$\text{Constraint}: \qquad |P_{exp}(0)| - |P(0)| = 0 \tag{6.20}$$

Using an initial guess $\mathbf{U}_0 = [\xi_1^0\ \omega_1^0\ \xi_2^0\ \omega_2^0\ \alpha^0\ \beta^0]$, the optimization problem assumes the form:

$$\mathbf{U}^* = arg \min_{\mathbf{U}_{min}<\mathbf{U}<\mathbf{U}_{max}} \int_0^{\omega_f} || \left\{ P_{exp}(j\omega) - P(j\omega) \right\} W(j\omega) ||_2^2 d\omega \tag{6.21}$$

subject to the constraint:

$$|P_{exp}(0)| = |P(0)| \tag{6.22}$$

A constrained minimization problem was set up in MATLAB using *fmincon* with the initial guess $U_0 = [\xi_1^0;\ \omega_1^0;\ \xi_2^0;\ \omega_2^0;\ \alpha^0;\ \beta^0]$, $U_{min} = [.01;\ 2\pi(745);\ 0.005;\ 2\pi(2300);\ 0.013;\ 8]$ and $U_{max} = [.05;\ 2\pi(755);\ 0.01;\ 2\pi(2335);\ 0.015;\ 10]$, the cost function given in Eq. (6.20), and constraint condition given in Eq. (6.22) as follows[11]:

```
[U] = fmincon(@cost,U_0,[],[],[],[],U_max,U_min,@constraint,[]);
```

The optimizer uses a sequential quadratic programming method to solve for the local optima in the vicinity of the initial guess.

---

[11] The full MATLAB script files for this problem are given in Appendix C.2.2.

For an initial guess used in Fig. 6-24, and the weight $W(jw)$ selected as unity until 5 kHz and zero thereafter, the optimization converged in about 30 iterations for an active set method chosen for the sequential quadratic programming with all constraint conditions met and the final value of the optimum cost function reduced by 51.35% compared to that of the initial guess. The optimization parameter vector was found to be $\mathbf{U}^* = [0.0148;\ 2\pi(745);\ 0.01;\ 2\pi(2335);\ 0.015;\ 8.2565]$ and the corresponding optimal plant model is as given below:

$$P(s) = \frac{2.699 \times 10^7 s^2 + 3.784 \times 10^9 s + 6.549 \times 10^{14}}{s^4 + 431.8s^3 + 2.372 \times 10^8 s^2 + 3.62 \times 10^{10} s + 4.716 \times 10^{15}} \tag{6.23}$$

The frequency response of this $\mathcal{H}_2$-optimal model is shown as the solid line in Fig. 6-25.

6.7.2 Nominal Sensitivity and Desired Sensitivity

In this section, we obtain first a nominal sensitivity and then tune or shape it so that it matches in performance with a desired sensitivity. How well the two responses of the nominal and desired sensitivity transfer functions match is defined in a $\mathcal{H}_2$ sense.

To obtain the nominal sensitivity given the plant described by Eq. (6.23), we need a nominal controller. As in the example of Chapter 5 a simple integral controller was chosen as a nominal controller for the system[12]. This controller is given as:

$$C_0(s) = \frac{k}{s} \tag{6.24}$$

where k is a constant gain. For a value $k = 1000$, the closed-loop system is stable with a phase margin of 90° and a phase crossover frequency of 138 rad/s.

---

[12] There are many possible choices for the nominal controller. For example, a PID, lead, lag, lead-lag, or notch in combination with an integral controller can also be used for this system. The procedure illustrated here is applicable to any such nominal controller that satisfies the condition that the resulting closed-loop system is internally stable.

Let us now examine the magnitude of the frequency response of the nominal sensitivity transfer function of the closed-loop system, which is shown in Fig. 6-26.

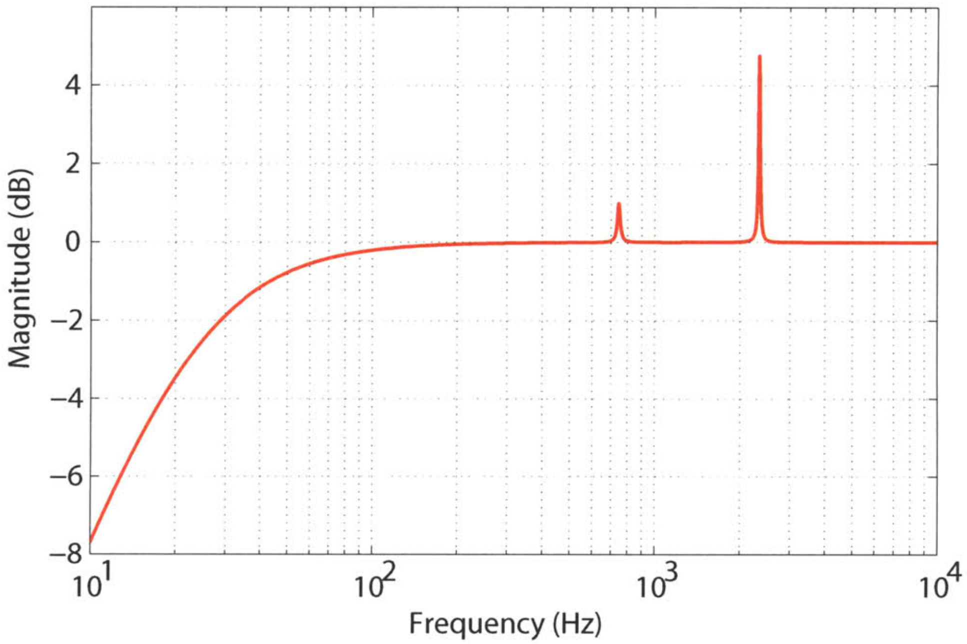


Figure 6-26: Magnitude frequency response plot of nominal sensitivity obtained with an integral controller. The sharp peak around about 2330 kHz results from the sharp peak in the plant frequency response. This peak indicates poor robustness and amplification of disturbance by a gain of 4.7 dB, which is undesirable.

The nominal sensitivity transfer function has a low gain at low frequencies, which results in good command following and disturbance rejection. The low-frequency roll-on or ramp is at a 20 dB/dec, indicating the presence of a free integrator, which comes from the nominal controller $C_0(s)$.

Notice also the sharp peaks in the magnitude response as shown in the figure. The 4.7 dB tall peak around about 2330 kHz results from a corresponding peak in the plant frequency response. This peak is undesirable because of two reasons: (i) a tall peak indicates poor robustness, i.e. any parametric variations can affect stability of the closed-loop system, and (ii) any disturbances at this frequency gets amplified by the magnitude of this peak.

Let us now consider tuning this nominal sensitivity to one that is more robust.

Consider the problem of reducing the peak in the nominal sensitivity by at least a factor of 3 dB, which corresponds to a magnitude reduction by a factor[13] of $\sqrt{2}$. A desired sensitivity transfer function meeting this requirement is shown as constrcuted by a frequency-gridding approach in Fig. 6-27, i.e. from a sample set of magnitude values selected at each frequency and marked by circles. A close-up view of the 2330 Hz peak is shown in Fig. 6-28. As seen in the figure, this desired sensitivity transfer function has its tallest peak reduced in height to a value about 1.5 times, i.e. by 3.5 dB.

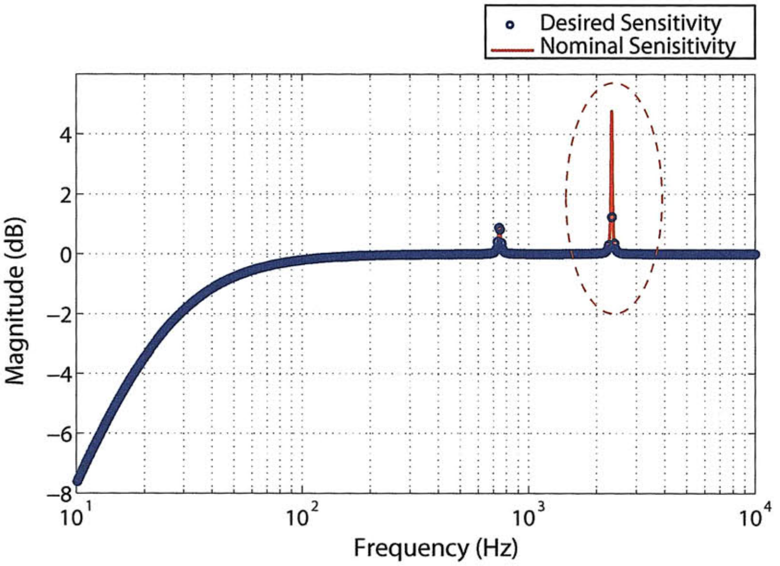


Figure 6-27: Magnitude plot of frequency response of nominal sensitivity transfer function is shown as a solid line. A discrete set of sensitivity values are selected for the desired sensitivity transfer function. Note that the peak of the desired sensitivity transfer function is much lower than the nominal sensitivity transfer function. A zoomed view around the peak is shown in Fig. 6-28. The goal here is to solve a model-matching problem, i.e. solve for the parameter transfer function $Q(s)$ that will minimize the error between the desired and nominal sensitivity transfer function.

Given the nominal and the desired sensitivity transfer functions, the problem of tuning is one of model-matching or filter design of the parameter transfer function $Q(s)$. This is posed as an optimal synthesis problem in the next section.

---

[13] A $\sqrt{2}$ reduction in amplitude for a sinusoidal disturbance signal at this frequency corresponds to 50% reduction in power of the signal

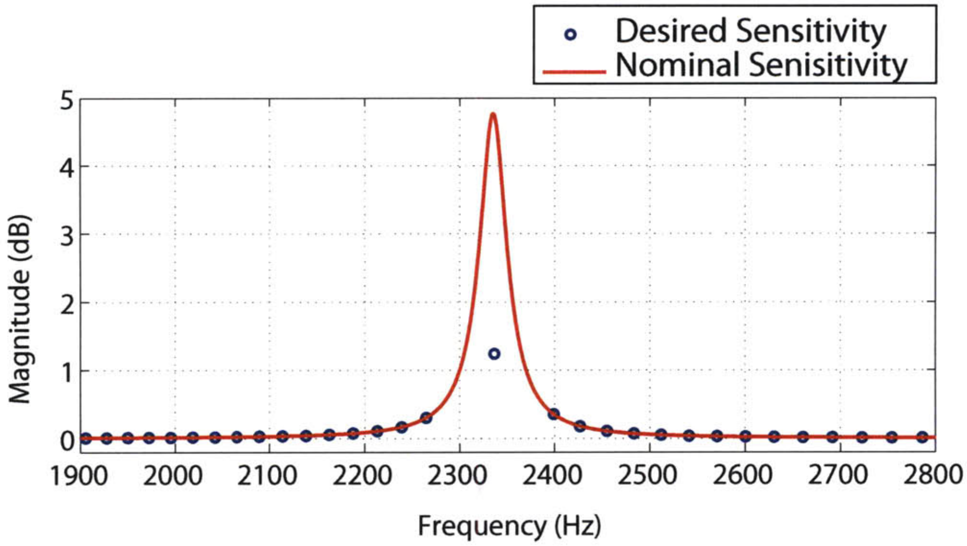


Figure 6-28: Zoomed view of peak of Fig. 6-27.

6.7.3 Optimal Synthesis of Q(s) Parameter Transfer Function

Since the plant is stable and minimum-phase, and using the nominal controller resulted in a stable control system, the relation of Eq. (6.16) is applicable to this problem. The optimal synthesis of the parameter transfer function $Q(s)$ is one of model-matching, i.e. we are searching for $Q(s)$ in the entire class of stable transfer functions $\mathcal{RH}_\infty$ so that an error norm, such as the following gets minimized:

$$Q(s) = \arg \min_{Q(s) \epsilon \mathcal{RH}_\infty} ||S_d(s) - S_0(s)Q(s)||_2^2 \tag{6.25}$$

To formulate the optimization problem, we have two options for assembling the parameter transfer function $Q(s)$. They are reproduced from Chapter 4 as follows:

$$Q(s) = a_0 + \sum_{i=1}^{N} \frac{a_i}{s - p_i} \tag{6.26}$$

$$Q(s) = \frac{b_o s^n + b_1 s^{n-1} + b_2 s^{n-2} + ...b_{n-1}s + b_n}{(s+p_1)(s+p_2)(s+p_3)....(s+p_{n-1})(s+p_n)} \tag{6.27}$$

The first option shown in Eq. (6.26) uses a summation of basis filters, and the second option in Eq. (6.27) uses a similar approach, with a finite set of poles. In both cases, how to select the basis filters needs to be addressed.

The multiplicative nature of Eq. (6.16) allows us to use insights from filter design for lightly damped systems. A simple choice for $Q(s)$ is a notch filter for the 4.7 dB peak seen in the nominal sensitivity transfer function data of Fig. 6-26.

The form of the filter $Q(s)$ following this choice is given as:

$$Q(s) = \frac{s^2 + 2\xi_1\omega_1 s + \omega_1^2}{s^2 + 2\xi_2\omega_2 s + \omega_2^2} \tag{6.28}$$

The above expression for $Q(s)$ can equivalently expressed as:

$$Q(s) = 1 + \frac{\alpha_1 + j\alpha_2}{s - (-\xi_2\omega_2 + j\omega_2\sqrt{1-\xi_2^2})} + \frac{\alpha_1 - j\alpha_2}{s - (-\xi_2\omega_2 - j\omega_2\sqrt{1-\xi_2^2})} \tag{6.29}$$

which is a special case of Eq. (6.26) with three terms chosen in the summation, with $a_0 = 1$, $a_1 = \alpha_1 + j\alpha_2$, $a_2 = \alpha_1 - j\alpha_2$ and $p_1 = -\xi_2\omega_2 + j\omega_2\sqrt{1-\xi_2^2}$, and $p_2 = -\xi_2\omega_2 - j\omega_2\sqrt{1-\xi_2^2}$.

Similarly, we can also rewrite the transfer function in Eq. (6.28) as:

$$Q(s) = \frac{s^2 + 2\xi_1\omega_1 s + \omega_1^2}{\{s - (-\xi_2\omega_2 + j\omega_2\sqrt{1-\xi_2^2})\}\{s - (-\xi_2\omega_2 - j\omega_2\sqrt{1-\xi_2^2})\}} \tag{6.30}$$

which is a special case of Eq. 6.27 with three terms chosen for the numerator polynomial in the summation, $b_0 = 1$, $n = 2$, and $p_1 = -\xi_2\omega_2 + j\omega_2\sqrt{1-\xi_2^2}$, and $p_2 = -\xi_2\omega_2 - j\omega_2\sqrt{1-\xi_2^2}$.

An optimization problem was formulated with a parameter vector $\mathbf{U} = [\xi_1;\ \omega_1;\ \xi_2;\ \omega_2]$ whose values are bounded between the limits $\mathbf{U}_{min}$ and $\mathbf{U}_{max}$, such that the error norm given below is minimized:

$$Cost: \quad \sum_{i=1}^{N} |S_d(j\omega_i) - S_0(j\omega_i)Q(j\omega_i)|^2 \tag{6.31}$$

where ω_i $(i = 1...N)$ are discrete frequencies at each of which a data point is available for the magnitude response of the desired sensitivity transfer function. To guide the optimization, a constraint that the peak reduces by at least 3 dB was imposed. This constraint assumes the form shown below:

$$Constraint: \quad 20log_{10}\left\{\frac{\max|S_{nominal}(j\omega)|}{\max|S_{opt}(j\omega)|}\right\} > 3\text{ dB} \tag{6.32}$$

A constrained minimization problem was set up in MATLAB using *fmincon* with the initial guess $U_0 = [0.006;\ 2\pi(2320);\ 0.008;\ 2\pi(2340)]$, $U_{min} = [0.005;\ 2\pi(2300);\ 0.005;\ 2\pi(2330)]$, $U_{max} = [0.01;\ 2\pi(2350);\ 0.01;\ 2\pi(2350)]$, the cost function given in Eq. (6.31), and constraint condition given in Eq. (6.32) as follows[14]:

```
[U] = fmincon(@cost,U_0,[],[],[],[],U_max,U_min,@constraint,[]);
```

The optimizer uses a sequential quadratic programming method to solve for the local optima in the vicinity of the initial guess.

For an initial guess of $\mathbf{U}_0$, the optimization converged in about 24 iterations for an active set method chosen for the sequential quadratic programming with all constraint conditions met and the final value of the optimum cost function reduced 88-fold compared to that of the initial guess. The optimization parameter vector was found to be $\mathbf{U}^* = [0.005;\ 2\pi(2339.6);\ 0.007;\ 2\pi(2341.1)]$ and the corresponding optimal $Q(s)$ synthesized by this approach, is given below:

$$Q(s) = \frac{s^2 + 147s + 2.161 \times 10^8}{s^2 + 206.8s + 2.164 \times 10^8} \tag{6.33}$$

---

[14] The full MATLAB script files for this problem are given in Appendix C.2.3.

The magnitude frequency response of the $\mathcal{H}_2$-optimal sensitivity $S_{opt}(j\omega)$ is shown as the solid line in Fig. 6-31. The dashed line shows the magnitude of the nominal sensitivity. The square markers are the outcomes of the optimization. At each frequency where the desired sensitivity function has a data point, the square marker represents the optimal solution that minimizes Notice that the peak of $|S_{opt}(j\omega)|$ has not fully reduced by the desired factor of 1.5, but nonetheless, has a reduced peak and hence better than the nominal sensitivity.

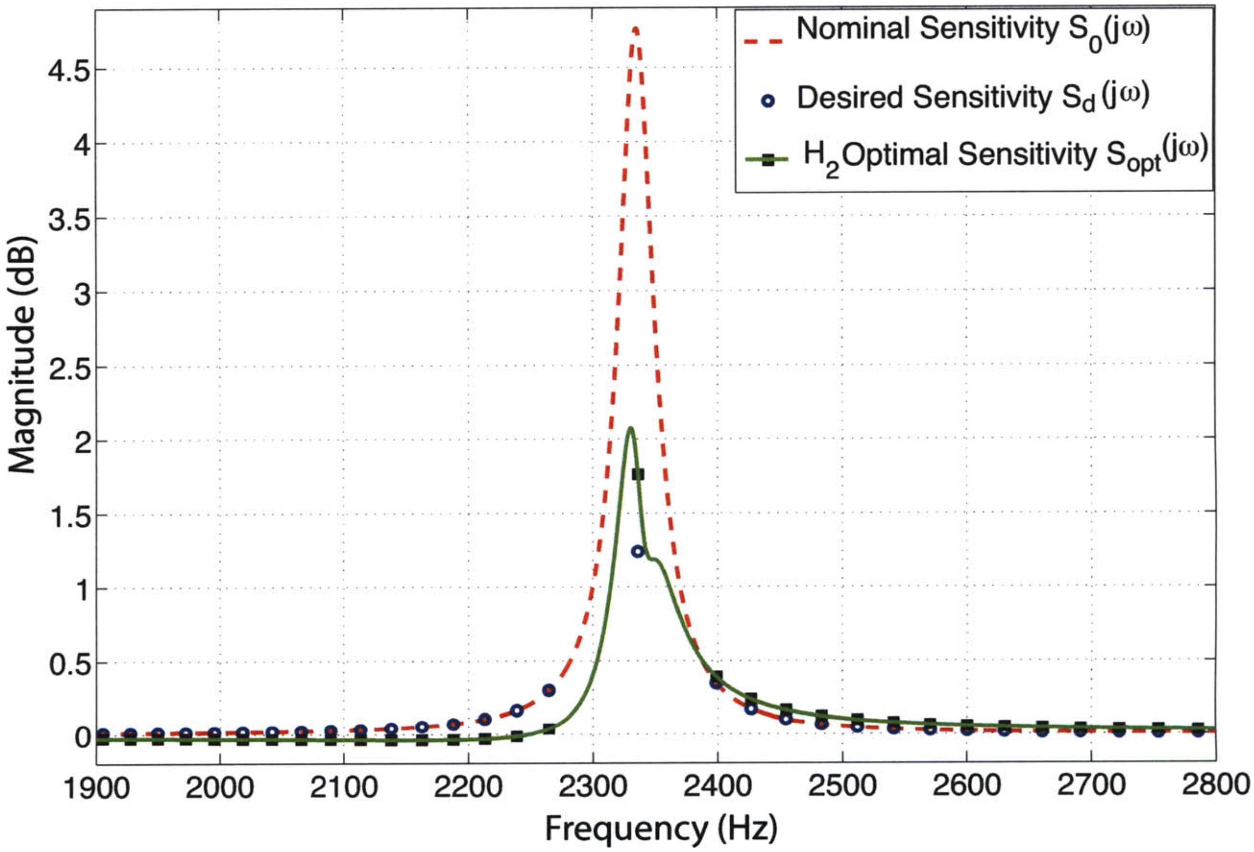


Figure 6-29: Zoomed view of the peak in the optimal sensitivity transfer function obtained from model-matching. The dashed line shows the magnitude frequency response of the nominal sensitivity transfer function. The circle markers denote the discrete points representing the desired or reference sensitivity transfer function for the optimization problem. The outcomes of the otpimization procedure are shown by square markers at the discrete frequencies of the desired sensitivity. The solid line shows the magnitude frequency response of the optimal sensitivity transfer function that corresponds to the square markers.

6.7.4 Controller Generation

Now that we have a value for $Q(s)$ we can use the expression given for the controller in Eq.(4.19), to find the resulting controller $C(s)$ as given below:

$$C(s) = \frac{C_{num}(s)}{C_{den}(s)} \tag{6.34}$$

where the numerator polynomial $C_{num}(s)$ is given as:

$$\begin{aligned} C_{num}(s) = {} & 58s^{10} + 3.437 \times 10^{5}s^{9} + 5.477 \times 10^{10}s^{8} + 1.756 \times 10^{14}s^{7} \\ & +1.679 \times 10^{19}s^{6} + 2.726 \times 10^{22}s^{5} + 1.946 \times 10^{27}s^{4} + 1.235 \times 10^{30}s^{3} \\ & +6.566 \times 10^{34}s^{2} + 1.606 \times 10^{37}s + 6.681e41 \end{aligned} \tag{6.35}$$

and the denominator polynomial $C_{den}(s)$ is given as:

$$\begin{aligned} C_{den}(s) = {} & 2.699 \times 10^{7}s^{9} + 1.861 \times 10^{10}s^{8} + 1.289 \times 10^{16}s^{7} \\ & +6.322e \times 10^{18}s^{6} + 1.808 \times 10^{24}s^{5} + 5.409 \times 10^{26}s^{4} + 6.418 \times 10^{31}s^{3} \\ & +9.34 \times 10^{33}s^{2} + 6.672 \times 10^{38}s \end{aligned} \tag{6.36}$$

The order of this controller can be explained from the consideration of the involved polynomial orders, using the steps detailed in Section 4.4.2. Using the notation from Section 4.4.2 we have the controller order as $c_{den} = p_{num} + p_{den} + c_{oden} + q_{num}$, where p_{num}, p_{den} are orders of the numerator and denominator polynomials, respectively of the plant transfer function, c_{oden} is the order of the controller and q_{num} is the order of the numerator of $Q(s)$. In our case, we have $p_{num} = 2, p_{den} = 4, c_{oden} = 1$, and $q_{num} = 2$, which results in an order of $c_{den} = 9$ for the controller. This is indeed what we see from Eq. (6.36).

Note the free integrator in the controller $C(s)$ (the denominator polynomial does not have a s^0 term). The free integrator comes from the nominal controller. In essence, we have a controller that has a fixed desired part and a variable order part that we obtained to account for improving the robustness of the sensitivity transfer function.

Note also that this controller is not proper. To make it strictly proper, the controller $C(s)$ was multiplied with a low pass filter with double poles on the real axis at $s = -2\pi(5000)\ rad/s$. With this filter included in the controller, the resulting controller $Cp(s)$ has an order of 11. In order to ease the implementation of the controller on a rapid prototyping controller, a lower order controller is desired. For this purpose, the controller was reduced using a balanced truncation in MATLAB as follows: to result in a

```
Cpr=reduce(Cp,'algorithm','balance','MaxError',[0.2]);
```

The reduced-order controller $C_{pr}(s)$ was multipled with a filter that has a first-order pole at $s = -2\pi(5000)\ rad/s$ to further reduce the slope of the roll-off at higher frequencies. The final controller choice $C_f(s)$ is a 3^{rd} order strictly proper controller with a relative degree of 2

$$C_f(s) = \frac{1}{s}\left\{\frac{1.983\times10^{7}s + 1.716\times10^{11}}{s^2 + 3.687\times10^{4}s + 1.714\times10^{8}}\right\} \tag{6.37}$$

We next address the question: do the modifications of low pass filtering and model-order reduction of the raw controller $C(s)$ to genrate the final controller $C(s)$ come with a loss of performance? The comparison of performance between the final controller $C_f(s)$ and the full-order $C(s)$ obtained from the optimal synthesis of $Q(s)$ is shown in Fig. 6-31. The sensitivity resulting from $C_f(s)$ has about 0.25 dB taller peak, which accounts for a slight loss of robustness. The value of 0.25 dB corresponds to about 1.029 in magnitude, i.e. a 3% increase in the height of the peak.[15]

---

[15] The loss of performance coming from multiplication of low pass filters can be predicted ahead.

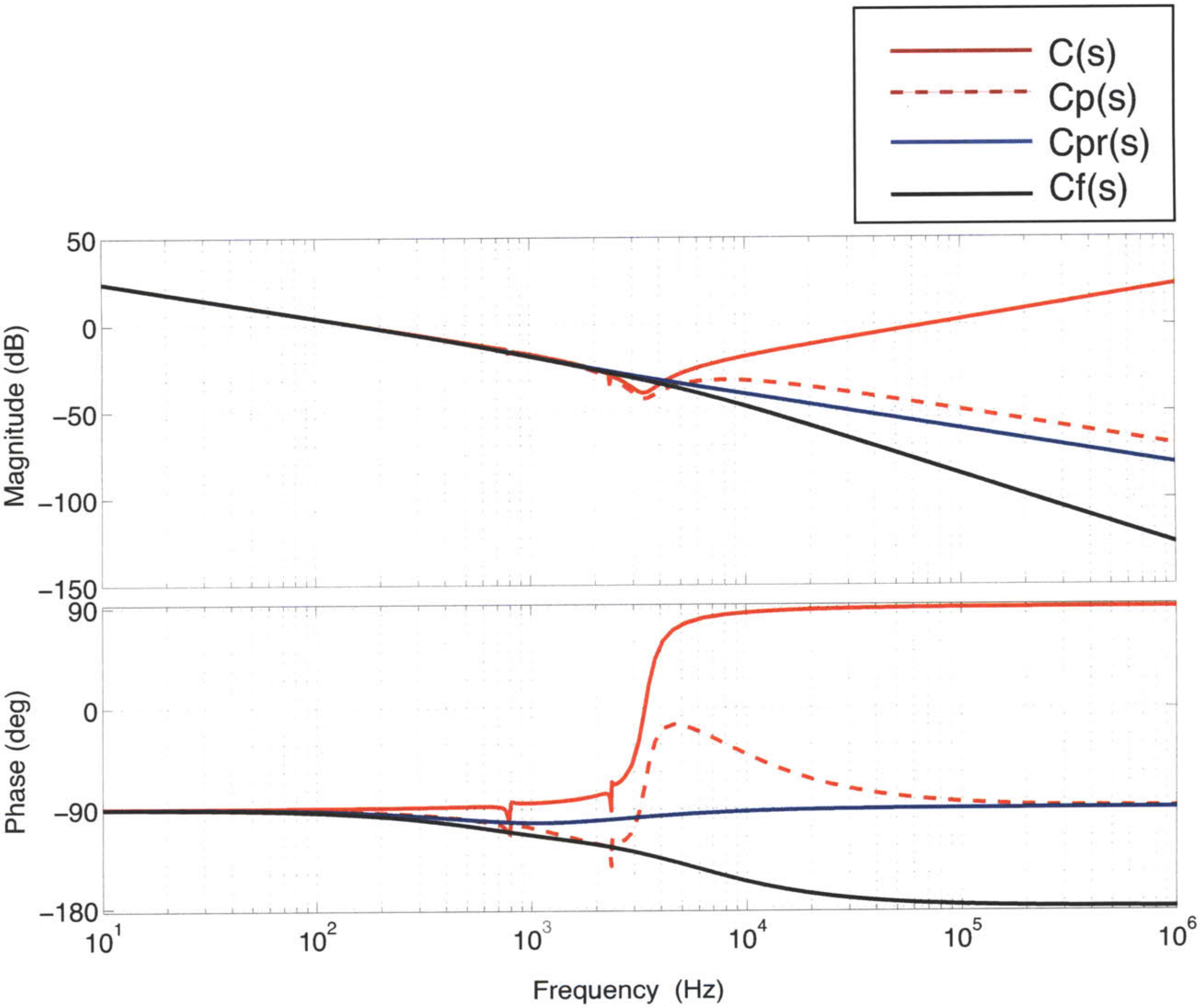


Figure 6-30: Controllers generated with the optimal $Q(s)$ synthesis procedure. $C(s)$ is a 9^{th} order controller which is improper, $Cp(s)$ is a 11^{th} order controller obtained by multiplying $C(s)$ with a double pole filter at $s = -2\pi(5000)$ Hz, $Cpr(s)$ is a reduced order version of $Cp(s)$ obtained from a balanced reduction process, and the final controller choice is $C_f(s)$ which has an additional pole at $s = -2\pi(5000)\ rad/s$ to further reduce the slope of the roll-off at higher frequencies. The final controller choice $C_f(s)$ is a 3^{rd} order strictly proper controller with a relative degree of 2.

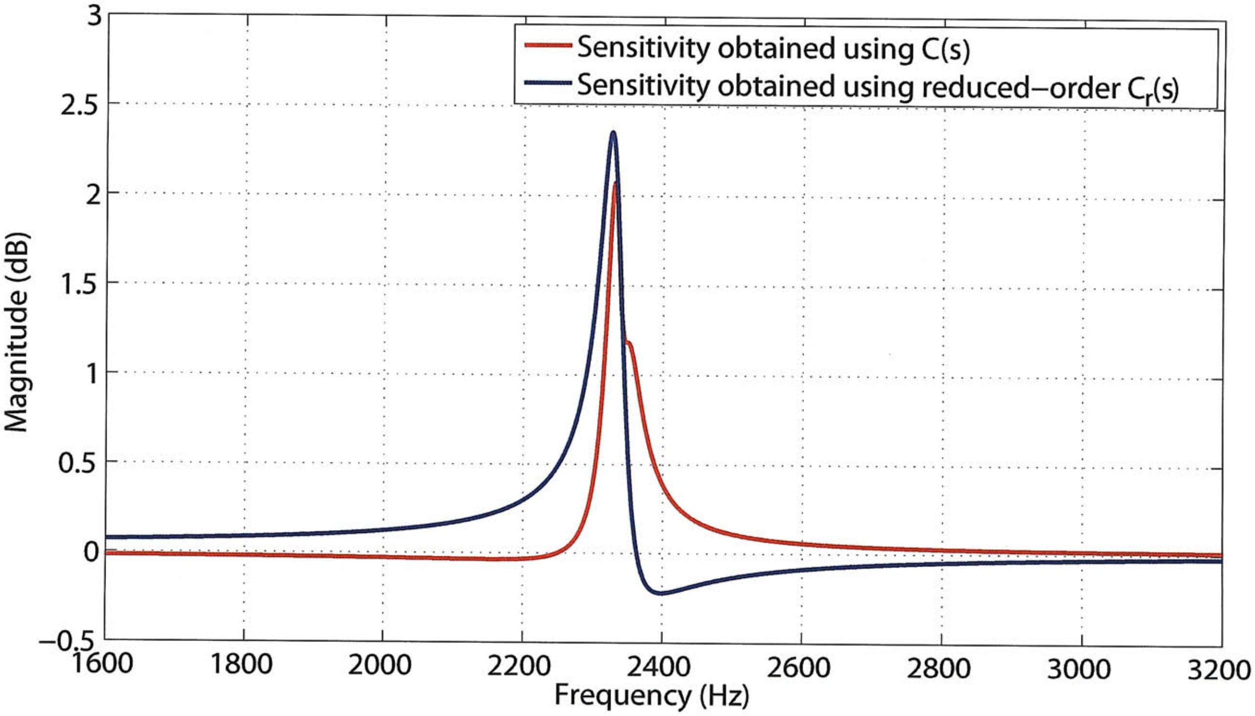


Figure 6-31: Magnitude frequency response of the optimal sensitivity transfer function $S^*((j\omega))$ and the sub-optimal sensitivity transfer function $S^*_{sub}(j\omega)$ obtained after a low pass filter is added to the controller to make it strictly proper, and reduced with a balanced truncation method.

6.7.5 Experimental Results

In this section we present results of control experiments performed on the prototype hardware of our AFM nanopositioner to test for the sensitivity tuning performance.

First, the controller $C_f(s)$ was implemented for the AFM nanopositioner for controlling the motion along the X axis. The instrumentation used include piezoelectric stack actuators and their power amplifiers, a laser interferometric sensor, and a dSPACE DS1103 rapid prototyping controller platform. The details of the instrumentation were discussed in Section 6.5.

The frequency response of the experimental closed-loop system transfer function $T(j\omega)$ obtained between voltage $V_2(s) - V_1(s)$ applied to the piezoelectric stack actuator and displacement $X(s)$ measured by the laser interferometer is shown for the case of the controller choice of $C_f(s)$ in Fig. 6-32. The predicted model closely matches with the locations of peaks in the data in both the magnitude and phase plots. The slope of the data at frequencies beyond the $2330Hz$ resonance matches with the 80 dB/dec roll-off.

To gauge the robustness of the system, the estimated values of sensitivity (derived from the condition $S(j\omega) + T(j\omega) = 1$) from the experimental data of Fig. 6-32 were calculated. These values are indicated as circle markers in Fig. 6-33. A dotted line representing the magnitude of the predicted sensitivity transfer function is overlaid on this data in the figure. The frequency location of peaks observed in the estimated sensitivity match well with those of the predicited sensitivity. The magnitude of the peak in the estimated sensitivity exceeds that of the predicted sensitivity by about 1 dB, but is still smaller than that of the nominal sensitivity. To improve on the performance, we studied the effect on robustness when an extra low pass filter

---

For a given a choice of plant, nominal controller and the parameter transfer function $Q(s)$, the relative degree of $C(s)$ can be predicted using the relations we derived in Chapter 4 in Section 4.4.2. After the relative degree is determined, the order of the low pass filter needed to make the controller proper is known.

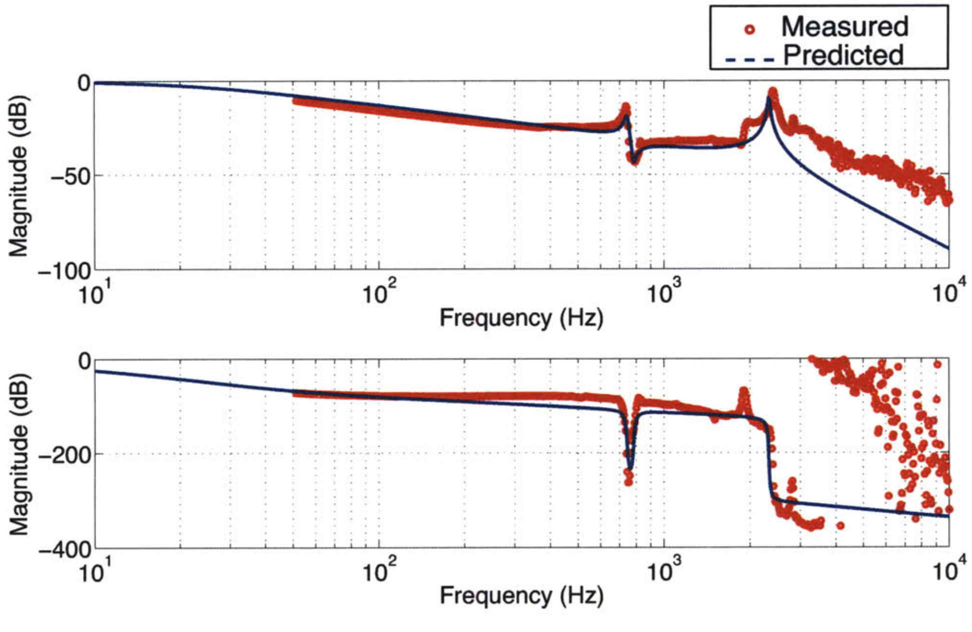


Figure 6-32: Experimental closed-loop control system transfer function between voltage applied to the piezoelectric stack actuator and displacement measured by the laser interferometer. The controller was generated by a sensitivity tuning method described in Chapter 4.

pole at $s = -2\pi(5000)\ rad/s$ was included in the controller $C_f(s)$ to form a new controller $C_{f2}(s)$. The transfer function representing this controller is as follows:

$$C_{f2}(s) = \frac{6.231 \times 10^{11}s + 5.392 \times 10^{15}}{s^4 + 6.829 \times 10^4 s^3 + 1.33 \times 10^9 s^2 + 5.385 \times 10^{12} s} \tag{6.38}$$

This controller can also be achieved by using additional basis filters in $Q(s)$ in our optimal synthesis formulation. The parameter transfer function $Q_2(s)$ that generated[16] C_{f2} is shown as the red solid line in Fig. 6-36. Its frequency response matches closely with a filter $Q_2'(s)$ that has two additional basis filters than $Q(s)$. The frequency response of $Q_2'(s)$ is plotted in black, and is given by:

$$Q_2'(s) = \frac{s^4 + 549s^3 + 4.309 \times 10^8 s^2 + 1.185e11s + 4.641e16}{s^4 + 631.6s^3 + 4.3 \times 10^8 s^2 + 1.361 \times 10^{11}s + 4.62 \times 10^{16}} \tag{6.39}$$

---

[16] Given the controller, nominal controller and plant, Eq. (4.19) can be solved for the parameter transfer function $Q(s)$. In this case, given $C_{f2}(s)$, $P(s)$, and $C_0(s)$, we found $Q_2(s)$ as a 13^{th} order stable transfer function.

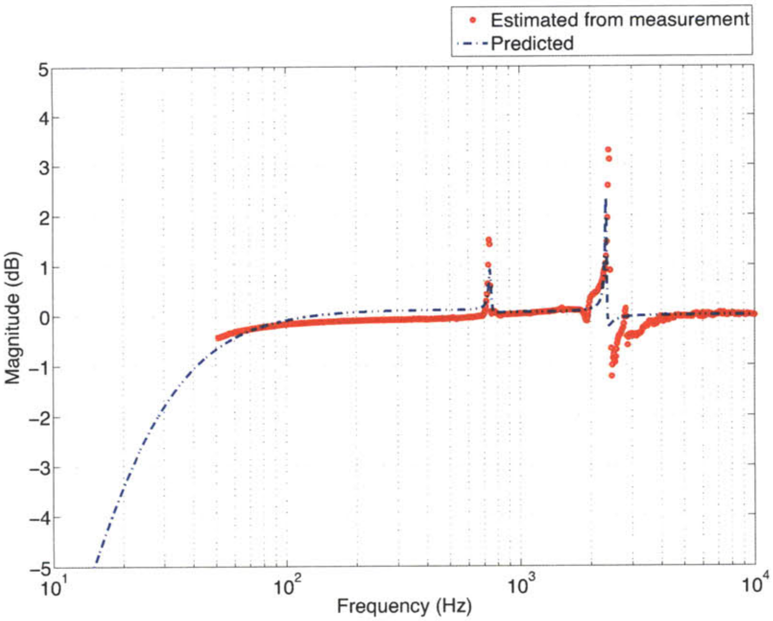


Figure 6-33: Estimates of the frequency response of the closed-loop sensitivity transfer function obtained from the closed-loop system transfer function data of Fig. 6-32 are shown by circle markers. The predicted frequency response of the sub-optimal sensitivity transfer function is shown as a dashed line.

which is a special case of Eq. (6.27) with five terms in the numerator and denominator polynomials.

An equivalent representation for $Q'_2(s)$ obtained from a partial fraction expansion is

$$Q`_2(s) = 1 + \frac{-(11.38 + 43.97j)}{s + (212.4 - 1.461e \times 10^4 j)} \tag{6.40}$$

$$\frac{-(11.38 - 43.97j)}{s + (212.4 + 1.461 \times 10^4 j)} \tag{6.41}$$

$$\frac{-(29.9 - 9.987j)}{s + (103.4 - 1.471 \times 10^4 j)} \tag{6.42}$$

$$\frac{-(29.9 + 9.987j)}{s + (103.4 + 1.471 \times 10^4 j)} \tag{6.43}$$

$$\tag{6.44}$$

which is a special case of Eq. (6.26) with five terms in the summation and basis filters as shown.

In essence, we have obtained a higher order $Q(s)$ parameter transfer function, resulting in an increase in controller order as expected. The closed-loop system frequency response resulting from using $C_{f2}(s)$ as the controller is shown in Fig. 6-34. The magnitude response in this figure has a lower peak at the 2330 Hz resonance. The estimated and predicted sensitivity responses shown in Fig. 6-35 depict an improved robustness for the system. This was possible because the additional pole reduced the sharp peak in the loop transmission, hence adding more gain margin to the system. The experimental data support the fact that an increased controller order will facilitate accounting for robustness.

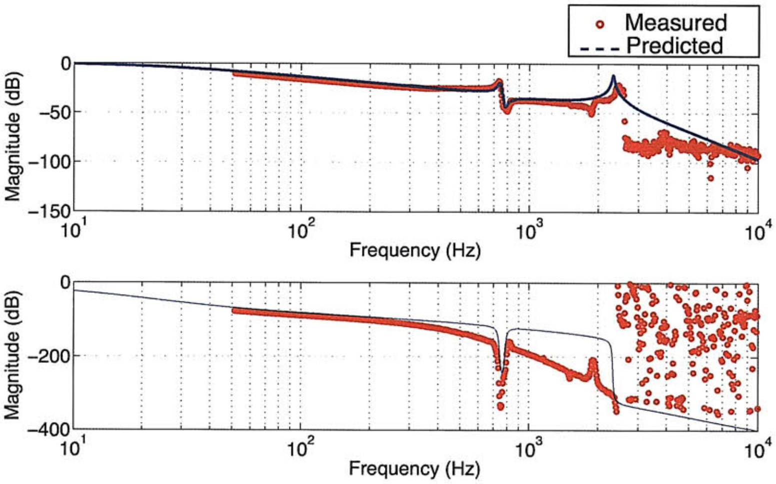


Figure 6-34: An additional pole in the controller results in a reduced magnitude for the peak in the roll-off part of the response located at 2.3 kHz. The higher order controller adds more robustness of the system, resulting in an increase in gain margin at the phase crossover frequency.

6.8 Summary

In this chapter, we worked out the case study of a flexure-based nanopositioner involving integrated design and control. The case study was motivated for the practical application of nanopositioner for probe (or sample) in scanning probe microscopy. Design topologies that allow for the desired control performance generated using the

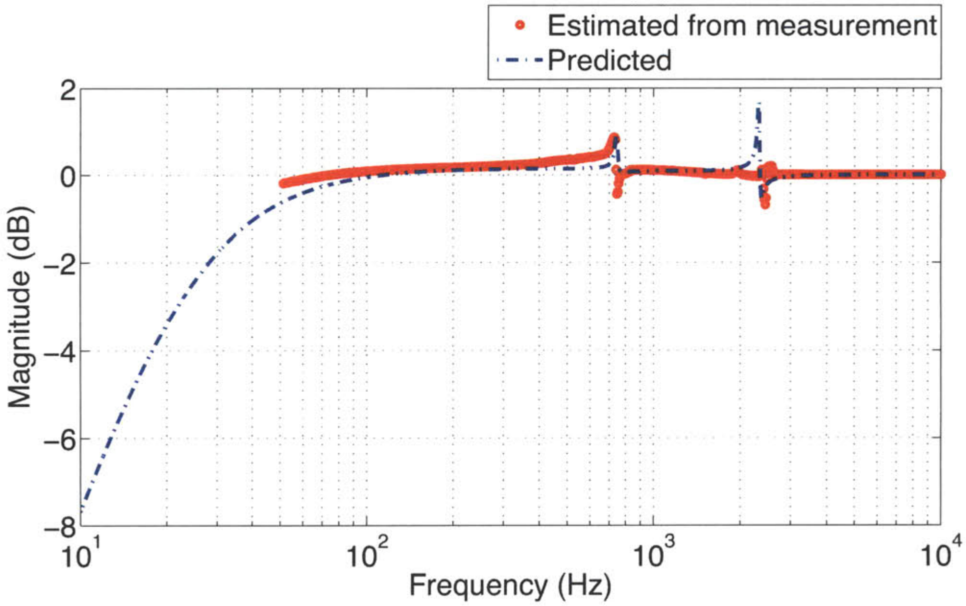


Figure 6-35: Estimates of the frequency response of the closed-loop sensitivity transfer function obtained from the closed-loop system transfer function data of Fig. 6-34 are shown by circle markers. The predicted frequency response of the sub-optimal sensitivity transfer function is shown as a dashed line. This set of data was recorded for the case of the controller $C_{r,new}(s)$, which is obtained from $C_r(s)$ after another low pass filter pole is placed at 5 kHz. In comparison with the response of Fig. 6-33, the peak magnitude of the sensitivity transfer function is further reduced. This is expected as the extra pole of the controller adds to the gain margin at the resonance peak located around 2.3 kHz. Increasing controller order here has increased the robustness of the system.

principles of Chapter 3 were examined. A topology with the XY flexures without the Z-rod flexures pinning the center down to the base plate is likely to have a low Z-resonance. This design topology cannot achieve the desired control performance and still meet the static range requirement. Any iterations in an optimization routine that takes this design as an input and fine-tunes the beam shape or size cannot converge to a feasible solution. On the other hand, a topology with the sixty four rod flexures bonded to the central stage showed a Z fundamental resonance as high as 33.1 kHz.

We also examined the shape and size optimization that followed after the topology selection. Static and dynamic frequency response characterization of the nanoposi-

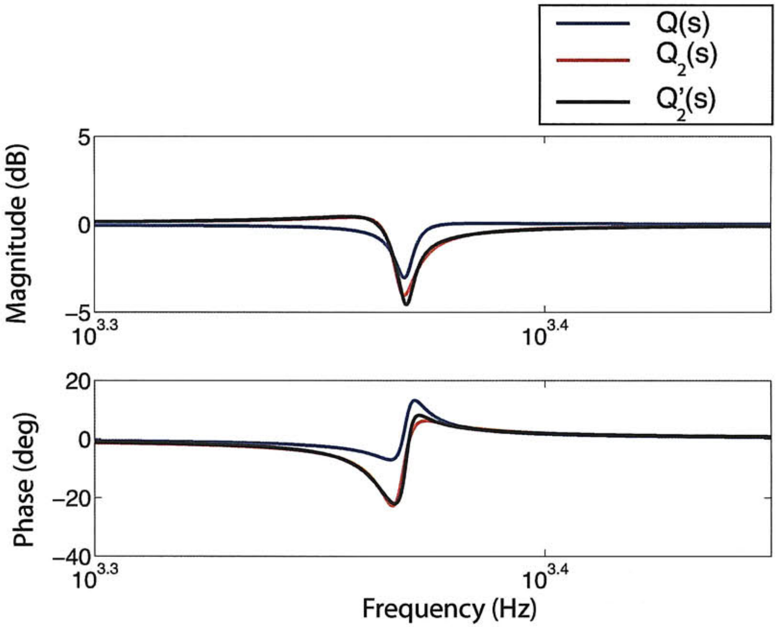


Figure 6-36: Overlaid frequency response plots of $Q(s)$ obtained for controller of Eq. (6.37), $Q_2(s)$ corresponding to controller of Eq. (6.38). The frequency response of the latter can be closely approximated with $Q_2'(s)$ obtained by having a constant and four basis filters in the formulation we used for $Q(s)$. Note that the notch filter obtained for a higher order transfer functions ($Q_2(s)$ and $Q_2'(s)$) is broader than that of $Q(s)$.

tioner were presented. Further, sensitivity tuning was achieved with the controller parameterization that allows for direct model matching to obtain the desired robustness. While the current results obtained with the nanopositioner are promising, we would like to extend the current design to include optical or other sensing subsystem that will allow for measuring the deflection of a cantilever probe on the central stage for measuring nanoscale topography.

Chapter 7

Conclusions and Recommendations

7.1 Summary of Contributions

The key contribution of this thesis is in the area of integrated design and control, with design topologies and controller order varied in order to achieve optimal performance. A novel methodology was tailored and tested for flexure-based nanopositioning systems. As case studies, a simple 1-DOF positioning system was worked out showing the steps involved in applying the methodology in Chapter 5. An experimental case study of a nanopositioner for Atomic Force Microscopy was discussed in Chapter 6. Specifically, the contributions made by the thesis are as follows:

- **Formulation of Performance Specifications**

  Developed a state-space formulation of nanopositioning specs for flexure-based systems: resolution, range, load-capacity, bandwidth. This formulation can be an useful component in setting up an optimization problem, as part of design (structural, thermal, dynamic etc) and closed-loop control (time or frequency domain specifications) objectives.

- **Generation of Design Library**

  Developed and implemented a systematic procedure that starts with design primitives, which when subject to performance requirements, get modified by a set of operations to generate the optimal building blocks that can be used in multi-DOF arrangements for achieving the desired motions and constraints. While existing methods target kinematic and elasto-mechanics for selection of flexures, we focus on a broad set of performance requirements such as (i) modal separation (including stiffness allotment for minimizing parasitic and cross-axis error motions) and structural strength (including high-load capacity and better fatigue performance) [39]. We introduced parameters such as parallel and serial stacking indices for flexure redundancies, and also a modal separation metric, which can facilitate quick design decisions in deriving possible design topologies. We invoked the principle of reciprocity in deriving design concepts that exhibit relatively larger (or reciprocally, smaller) modal separation based on the directions in which flexure constraints are implemented.

- **Controller Parameterization for Sensitivity Tuning**

  Proposed a controller parameterization that has a generic formulation accommodating existing parameterizations, such as the Youla, or Q parameterization. One of the cases in our parameterization allows for a novel way to directly tune the system sensitivity transfer function. In control design, this parameterization can be used to build a robustifying controller that multiples in series with an optional nominal (baseline) controller and allows to account for uncertainties. Since flexure-based mechanisms are usually made from monolithic metallic structures that have low inherent damping, their dynamic behavior is limited by lightly damped resonances. Tuning the sensitivity transfer function allows us to narrow down our controller search in the stabilizing class of controllers, for example, to those that can achieve a prescribed level of damping and prescribed relative stability.

- **Integrated Design and Control Methodology**

  Developed a novel integrated design and control methodology in which both design structure and controller order are varied. Candidate design topologies and controllers are screened for primary criteria, such as a specified requirement for a desired level of modal separation, or a zero steady-state tracking error. The screened design topology and controller are pass through an optimization procedure that fine tunes the design parameters (geometry, material choice etc.), and uses a

- **Novel Atomic Force Microscope nanopositioner**

  Proposed, developed, and tested a novel flexure-based probe scanner that outperforms existing flexure-based probe-scanners in terms of range (field of view in the XY plane) and dynamic performance in vertical (cantilever deflection) direction. This scanner is designed for use as part of a functional atomic force microscope.

- **Novel high-strength designs for heavy-duty flexure stages**

  Presented a systematic procedure for selection of flexure-based mechanism designs that are marked by enhanced strength performance (better load-capacity, fatigue performance). Novel designs critical for tool-sample alignment in high-load processes such as micro-hot-embossing were presented.

7.2 Future Work

Over the course of the thesis, we have benefited from insights from practising engineers and academics alike. While the current work focused on molding the idea of design for control in the context of flexure-based nanopositioners, many interesting topics of research arise. Here are listed a few possible directions for future study:

- **Plant-Controller tradeoffs**

  While this thesis allows for a framework for iterating between a suitable design topology and controller structure for achieving an end closed-loop control performance, it paves the framework for exploring interesting questions. For example, if a certain design (plant) with the control system implemented as suggested in the thesis, or by other suitable means, does not meet the desired closed-loop performance, is the fundamental performance limitation from the plant or the controller?

- **Formulation of Manufacturability in Topology Generation**

  The inherent assumption of the topology generation developed in this thesis was that the building blocks were planar structures that could be formed by traditional machining (e.g. waterjet or wire-EDM) or as MEMS structures in small-scale implementations. The units could be assembled in three-dimensional space to satisfy the kinematics. While this approach simplifies our generation, it would be of critical technical value to explore topology generation with a manufacturability metric that checks if a generated topology can be fabricated with conventional or other advanced manufacturing techniques. Related work in this area has been initiated by Trease [47] and could be a starting point for researchers interested in this topic.

- **Material Choice for Flexures**

  Our current nanopositioner prototypes are machined in Al 7071 alloy, which has high yield strength and hence allows for more deflection before failure. Use of other materials, such as composite fibers could be an alternative option that alleviates design requirements such as damping or stiffness. For instance, composite fibers allow for more internal energy dissipation and hence make them good damping elements.

- **Imaging for Scanning Probe Microscopy**

  The use of the current nanopositioner with a coarse Z-positioning stage allows for lowering an AFM cantilever probe tip on to a sample positioned on a coarse XY-positioning stage. Imaging at relatively higher speeds than conventional piezotube scanners is possible due to the high Z-resonance frequency achieved with our nanopositioner.

  An adjustable optical sensing system based on a laser photodiode and position sensing detector (PSD) is being developed as another thesis project in our lab as an add-on accessory to the flexure-based scanner. Once the development and testing of the optical sensing subsystem is completed, the optical system can be integrated into the current design. An alternative to external sensing with optical sensing is embedded sensing in the AFM cantilever itself. Active cantilevers made with strain sensing and piezoelectric films for actuation are an exciting area of research in this context. The use of dual actuation and embedded sensing for closed-loop control of the Z-axis constitutes a topic of further research.

7.3 Closing Remarks

Designing machines for achieving a desired closed-loop control performance has generated active interest in the precision engineering community. This thesis leveraged our interest in basic flexure design concepts and linear control theory to formulate desired performance requirements, such as range, bandwidth, load-capacity for positioning and alignment applications using flexure-based mechanisms.

The generation of design topologies that are well-suited for closed-loop control is of great interest not only in the nanopositioning community but also those of compliant mechanisms and motion control on a large scale. Physical insights into the dynamic behavior of the system need to be put to use in designing the control system. The use of controller parameterization within this context has been explored with the intent of tuning the sensitivity transfer functions. Many interesting topics of research branch out from the thesis and have been highlighted. We hope this thesis serves as a spring board for new design topologies and controller designs that can facilitate more advances in precision engineering and instrumentation.

Bibliography

[1] Loney GC, “High Bandwidth Steering Mirror,” US Patent 5,110,195, 1992.

[2] Choi BJ, Sreenivasan SV, Johnson S, et al, “Design of orientation stages for step and flash imprint lithography,” Precision Engineering, Journal of the International Societies for Precision Engineering and Nanotechnology, vol. 25 (3), pp 192-199, July 2001.

[3] Choi KB, Lee JJ “Passive compliant wafer stage for single-step nano-imprint lithography,” Review of Scientific Instruments, 76 (7): Art. No. 075106 Jul 2005.

[4] Kendale AM, “Automation of soft lithographic microcontact printing,” SM thesis, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge MA, 2002.

[5] Chan HB, Aksyuk VA, Kleiman RN, et al, “Quantum mechanical actuation of microelectromechanical systems by the Casimir force,” Science, vol. 291 (5510): 1941-1944, Mar 2001.

[6] Slocum AH, “Method of and apparatus for substance processing with small opening gates actuated and controlled by large displacement members having fine surface finishing,” US Patent # 596442, 1999.

[7] White JR, "The nanogate: nanoscale flow control," Ph.D. dissertation, Cambridge, MA: Massachusetts Institute of Technology, Department of Mechanical Engineering, June 2003.

[8] Shilpiekandula V, "Progress through Mechanics: Small-scale Gaps," *Mechanics* (Publication of the American Academy of Mechanics), vol. 35, no. 9-10, pp. 3-6, Sep-Oct 2006.

[9] Prechtl EF and Hall SR, "Design of a High Efficiency, Large Stroke, Electro-Mechanical Actuator," Journal of Smart Materials and Structures, vol. 8, pp. 13-30, 1999.

[10] Shilpiekandula V, Youcef-Toumi K, "Modeling and Control of a Programmable Filter for Separation of Biologically Active Molecules," In Proceedings of *American Control Conference*, pp. 394-399, June 2005.

[11] Yang C, Jiang W, Chen DH, et al, "Estimating contrast transfer function and associated parameters by constrained non-linear optimization," Journal of Microscopy, vol. 233 (3), pp. 391-403, Feb 2009.

[12] Akcay H, Ninness B, "Orthonormal basis functions for modelling continuous-time systems," Signal Processing, vol. 77 (3), pp. 261-274 , Sep 1999.

[13] Randall CA, Kelnberger A, Yang GY, Eitel RE, "High Strain Piezoelectric Multilayer Actuators: A Material Science and Engineering Challenge, "Journal of Electroceramics, Springer Netherlands, ISSN 1385-3449, vol. 14 (3), July 2005.

[14] Slocum AH, "Precision Machine Design," Englewood Cliffs, NJ, Prentice Hall, 1992.

[15] Canfield RA and Meirovitch L, "Integrated structural design and vibration suppression using independent modal-space control,"

AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 34th and AIAA/ASME Adaptive Structures Forum, La Jolla, CA, Technical Papers. Pt. 6 (A93-33876 13-39), pp. 3404-3414, April 1993.

[16] Awtar S, Slocum AH, Sevincer E. "Characteristics of Beam-based Flexure Modules," Journal of Mechanical Design, vol. 129 (6): 625-639 Jun 2007.

[17] Jones RV, "Instruments and Experiences: Papers on Measurement and Instrument Design," John Wiley, 1988.

[18] Hopkins JB, "Design of Parallel Flexure Systems via Freedom and Constraint Topologies," Master of Science Thesis, Massachusetts Institute of Technology, 2005.

[19] White J, Ma H, Lang J, Slocum AH, "An Instrument to Control Parallel Plate Separation for Nanoscale Flow Control," Review of Scientific Instruments, vol. 74, no. 11, November 2003.

[20] Azzam M, Mohammed YS, "Robust controller design for automatic generation control based on Q-parameterization," Energy Conersion and Management, vol. 43 (13), pp. 1663-1673, Sep. 2002.

[21] Lobontiu N, "Compliant Mechanisms : Design of Flexure Hinges," Boca Raton, CRC Press, 2003

[22] Li Z, and S. Kota, "Dynamic Analysis of Compliant Mechanisms, "Proceedings of DETC, 27th Biannual Mechanisms and Robotics Conference," Sep 29 - Oct 2, Montreal, Canada, 2002.

[23] Choi KB, "Dynamics of a Compliant Mechanism based on Flexure Hinges," Proceedings of the Institution of Mechanical Engineers, Part-C, Journal of Mechanical Engineering Science 219 (2): 225-235 Feb 2005.

[24] Choi KB, Han CS, "Optimal Design of a Compliant Mechanism with Circular Notch Flexure Hinges," Proceedings of the institution of Mechanical Engineers, Part-C, Journal of Mechanical Engineering Science 221 (3): 385-392 Mar 2007.

[25] Fathy HK, Reyer JA, Papalambros PY, Ulsoy AG, "On the Coupling between the Plant and Controller Optimization Problems," In Proceedings of American Control Conference, Arlington VA, June 25-27, 2001.

[26] Trease BP, Kota S, "Synthesis of Adaptive and Controllable Compliant Systems with Embedded Actuators and Sensors," In Proceedings of DETC 2006, International Design Engineering Technical Conference, Philadelphia, PA, Sept 10-13, 2006.

[27] Fu K, Sun D, Mills JK, "Simultaenous Mechanical Structure and Control System Design: Optimization and Convex Approaches," In Proceedings of IEEE International Symposium on Intelligetn Control, Vancouver, Canada, October 27-30, 2002.

[28] Rai S and Asada HH, "Integrated Structure/Control Design of High Speed Flexible Robots Based on Time Optimal Contro", ASME Journal of Dynamic Systems, Measurement and Control, Vol. 117, No. 4, pp. 503-512, 1995.

[29] Wu FX, Zhang WJ, Li Q, Quyang PR, "Integrated Design and PD Control of High-Speed Closed-loop Mechanisms," Journal of Dynamic Systems Measurement and Control, vol. 124, (4), pp. 522-528, Dec 2002.

[30] Pil A and Asada HH,"Rapid Recursive Structure Re-Design for Improved Dynamic Performance", ASME Journal of Dynamic Systems, Measurement and Control, July 1994, Vol. 117, December 1995.

[31] Rai S, Asada HH, "Computer-Aided Structure Modification of Electromechanical Systems Using Singular Value Decomposition", ASME J. Mechanical Design, Vol. 116, December 1994.

[32] Savant SV, Asada HH, "Integrated Structure/Control Design Based on Model Validity and Robustness Margin," In Proceedings of American Control Conference, vol. 4, pp. 2871-2875, June 2-4, 1999.

[33] Savant SV, Asada HH, "Shaping Structure Dynamics with Truncation-Error Bounded Reduced-Order Models for Integrated Mechanism/Control Design," In Proceedings of American Control Conference, vol. 6, pp. 3540-3544, June 24-26, 1998.

[34] Savant SV, "Structure Design Fpor Control Robustness Against Truncation Uncertainty," Ph.D. Thesis, Department of Mechanical Engineering, MIT, 1999.

[35] Varanasi KK, "On the Design of a Precision Machine for Closed-loop Performance," M.S. Thesis, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge MA 02139, USA.

[36] Zuo L, "Element and System Design for Active and Passive Vibration Isolation," Ph.D. Thesis, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge MA 02139, USA.

[37] Xianmin Z, Jianwei L, Yunwen S, "Simultaneous Optimal Structure and Control Design of Flexible Linkage Mechanism for Noise Attentuation," Journal of Sound and Vibration, vol. 299, pp. 1124-1133, 2007.

[38] Chawankul N, Budman H, Douglas PL, "The Integration Of Design and Control: Internal Model Control And Robustness," Computers and Chemical Engineering vol. 29 (2), pp. 261-271, Jan 15, 2005.

[39] Shilpiekandula V, Youcef-Toumi K, "A Flexure-based Mechanism for Precision Angular Alignment at Large Loads," In Proceedings of American Society for Precision Engineering Annual Conference, Portland OR, October 2008.

[40] Shilpiekandula V, Youcef-Toumi K, "Characterization of Dynamic Behavior of Flexure-based Mechanisms for Precision Angular Alignment," In Proceedings of American Control Conference, June 2008.

[41] Fantner GE, Schitter G, Kindt JH, et al, "Components for high speed atomic force microscopy," Ultramicroscopy, vol. 106, no. 8-9, pp. 881-887, Jun-July 2006.

[42] da Silva MM, Bruls O, Desmet W, Van Brussel H, "Integrated structure and control design for mechatronic systems with configuration-dependent dynamics," Mechatronics, vol. 19, (6), pp. 1016-1025, Sep 2009.

[43] Awtar S, "Synthesis and Analysis of Parallel Kinematic XY Flexure Mechanisms," Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge MA 02139.

[44] Awtar S and Sevincer E, 2006, "Elastic Averaging in Flexure Mechanisms: A Multi-Parallelogram Flexure Case-study," In Proceedigns of ASME IDETC/CIE 2006, Philadelphia, PA, Paper No. 99752.

[45] Chen S, Golda D et al, "Design of an Ultra Precision Diaphragm Flexure Stage for Out-of-plane Motion Guidance," Proceedings of the ASME IDETC, Salt Lake City, UT, Sept,2004, pp. 1015-21.

[46] Park JH and Asada HH, "Concurrent Design Optimization of Mechanical Structure and Control for High Speed Robots", ASME J. Dynamic Systems, Measurement, and Control, Vol. 116, September, 1994.

[47] Trease BP, Moon Y, Kota S, “Design of Large Displacement Compliant Joints,” Journal of Mechanical Design, vol. 127, pp 788-798, July 2005.

[48] Smith ST, “Flexures: elements of elastic mechanisms,” Amsterdam, Gordon and Breach, 2000.

[49] Howell LL, “Compliant mechanisms,” New York , Wiley 2001.

[50] Li Z, and S. Kota, “Dynamic Analysis of Compliant Mechanisms, “Proceedings of DETC, 27th Biannual Mechanisms and Robotics Conference,” Sep 29 - Oct 2, Montreal, Canada, 2002.

[51] Pil A and Asada HH, “Integrated Structure/Control Design of Mechatronic Systems”, IEEE/ASME Transactions on Mechatronics, September 1996.

[52] Weaver W, Timoshenko SP, Young DH, “Vibration problems in engineering,” New York, Wiley, 1990.

[53] Meirovitch L, “Computational Methods in Structural Dynamics,” Sijthoff Noordhoff, Rockville MD, 1980.

[54] Subrahmanyan PK, “Moving Magnet Voice Coil Motor Using Halbach Arrays,” US Patent 703116B2.

[55] Dwivedy SK, Eberhard P. “Dynamic analysis of flexible manipulators, a literature review,” Mechanism and Machine Theory, vol. 41 (7): 749-777, July 2006.

[56] Prezmieniecki JS, “Theory of matrix structural analysis,” New York, McGraw-Hill, 1968.

[57] Park JH, Asada H, “Design and control of minimum-phase flexible arms with torque transmission mechanisms,” IEEE International Conference on Robotics and Automation, pp 1790-1795, May 1990.

[58] Shilpiekandula V, Song YS,"A Music-based Mechatronic System for Teaching Modeling and Control," Paper IMECE-66817, In Proceedings of ASME International Mechanical Engineering Congress and Exposition, IMECE, Boston MA, October 2008.

[59] Subrahmanyan PK, "Control and design interactions in precision machines," Tutorial Notes, American Society for Precision Engineering Annual Conference, Portland OR, Oct 2009.

[60] Ganesan N and Engels RC, "A dynamic finite element model for the Timoshenko beam," AIAA-1990-3547 Space Programs and Technologies Conference, Huntsville AL, Sep 1990.

[61] Inman DJ, "Engineering vibration," Englewood Cliffs, NJ, Prentice Hall,1994.

[62] Varanasi KK and Nayfeh SA, "Damping of Flexural Vibration Using Low-Density, Low-Wave-Speed Media," Journal of Sound and Vibration, vol. 292, pp. 402-414, 2006.

[63] Hsu JC, Lee HL, Chang WJ, "Flexural vibration frequency of atomic force microscope cantilevers using the Timoshenko beam model," Nanotechnology, vol.18 (28): Art. No. 285503, July 2007.

[64] Varanasi KK, Nayfeh SA, "The Dynamics of Lead-Screw Drives: Low-Order Modeling and Experiments," Journal of Dynamic Systems, Measurement, and Control, Transactions of the ASME, vol. 126, pp. 388-396.

[65] Skogestad S and Postlethwaite I, "Multivariable feedback control : analysis and design," Hoboken, NJ, John Wiley, 2005.

[66] Yong YK, Aphale SS, and Moheimani SOR, "Design, Identification and Control of a Flexure-based XY stage for Fast Nanoscale Positioning," IEEE Transactions on Nanotechnology, vol. 8., no. 1, Jan 2009.

[67] Bristol EH, "On a new measure of interactions for multivariable process control," IEEE Transactions on Automatic Control AC-11, pp. 133-134, 1966.

[68] MacKenzie RI, Gawlik AJ, Trumper DL, "High-scan Rate Positioner for Scanned Probe Microscopy," In proceedings of American Society for Precision Engineering, Portland OR October 2008.

[69] Ando T, Kodera N, Takai E, et al, "A High-Speed Atomic Force Microscope for Studying Biological Macromolecules," Publication of the National Academy of Sciences (PNAS), vol. 98, no. 22, pp. 12468-12472, October 2001.

[70] Shilpiekandula V, Burns DJ, Li S, Xu Z, et al, "Fusion of Metrology Data for High-Volume Large-Scale Manufacturing of Polymer-Based Microfluidic Devices," International Journal of Nanomanufacturing, vol. 3, no. 4, 2009.

[71] Binning G, Quate CF, Gerber Ch, "Atomic Force Microscope," Physics Review Letters, vol. 56, 930, 1986.

[72] Infinitesima Limited, Innovations House, Mill Street, Oxford, OX2 0JX, UK. Phone +44-1865-811-171.

[73] Humphris ADL, Miles MJ, Hobbs JK. "A Mechanical Microscope: High-Speed Atomic Force Microscopy," Applied Physics Letters, vol. 86, no. 034106, 2005.

[74] Burrer GJ, "Dual-mode Resonant Scanning System," US Patent 4859846, Aug 1989.

[75] Ueda J, Secord TW, Asada HH, "Static Lumped Parameter Model for Nested PZT Cellular Actuators with Exponential Strain Amplification Mechanisms," In

Proceedings of IEEE Conference on Robotics and Automation, pp. 3582-3587, May 2008.

[76] Techproject In, Elisenstrasse 23, A-1230 Wien, Osteerreich, (0043) 676 427 7064, http://www.techproject.at

[77] Blanding DK, "Exact Constraint: Machine Design Using Kinematic Principles," ASME Press, 1999.

[78] Hale LC, "Principles and Techniques for Designing Precision Machines," Ph.D. Thesis, MIT Department of Mechanical Engineering, 1999.

[79] Kim C, Moon YM, Kota S, "A Building Block Approach to the Conceptual Synthesis of Compliant Mechanisms utilizing Compliance and Stiffness Ellipsoids," ASME Journal of Mechanical Design, vol. 130, 022308, 2008.

[80] Awtar S, Kim C, "Building Block Design of Compliant Mechanisms" ASME, International Design Engineering Technical Conference (IDETC), Brooklyn NY, 2008.

[81] Doyle JC, Francis BA, Tannenbaum AI, "Feedback Control Theory," Macmillan Publishing Co., 1999.

[82] Sebastian A, Salapaka SM, "Design methodologies for robust nano-positioning," IEEE Transactions on Control Systems Technology, vol. 13, no. 6, pp. 868-876, 2005.

[83] Lee C, Salapaka SM, "Robust broadband nanopositioning: fundamental trade-offs, analysis, and design in a two-degree-of-freedom control framework," Nanotechnology, vol. 20, no. 3, pp. 035501, 2209.

[84] Prochazka H, Landau LD, "Pole placement with sensitivity function shaping using 2nd order digital notch filters," Automatica, vol. 39, no. 6, pp. 1103-1107, 2003.

[85] Khatibi H and Karimi A, "Robust control of polytopic systems by convex optimization," Automatica, vol. 43, no. 8, pp.1395-1402, 2007.

[86] Boyd SP, Barratt CH, "Linear Controller Design: Limits of Performance," Prentice Hall, NJ, 1991.

[87] Gajic Z and Qureshi MJT, "Lyapunov Matrix Equation in System Stability and Contol," Mathematics in Science and Engineering, Academic Press, 1995.

[88] Jabben L, "Mechatronic Design of a Magnetically Suspended Rotating Platform," Ph. D. Thesis, Delft University, Netherlands, 2007.

Appendix A

Internal Stability

Internal stability is a critical requirement of any control design. By internal stability, we mean that the transfer function from any exogenous input to any internal signal in the feedback system remains stable all the time. Right half plane pole-zero cancellations result in violating the internal stability conditions and must be avoided as much as possible.

The internal stability of the feedback system show in Fig. B-1 can be inferred from the following matrix relation given in [81] between the internal signals x_1, x_2, and x_3 to the exogenous inputs r, d, n below. The feedback system is said to be internally stable iff each of the nine transfer functions in this matrix are stable.

$$\left\{\begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array}\right\} = \frac{1}{1+PC}\begin{bmatrix} 1 & -P & -1 \\ C & 1 & -C \\ PC & P & 1 \end{bmatrix}\left\{\begin{array}{c} r \\ d \\ n \end{array}\right\}$$

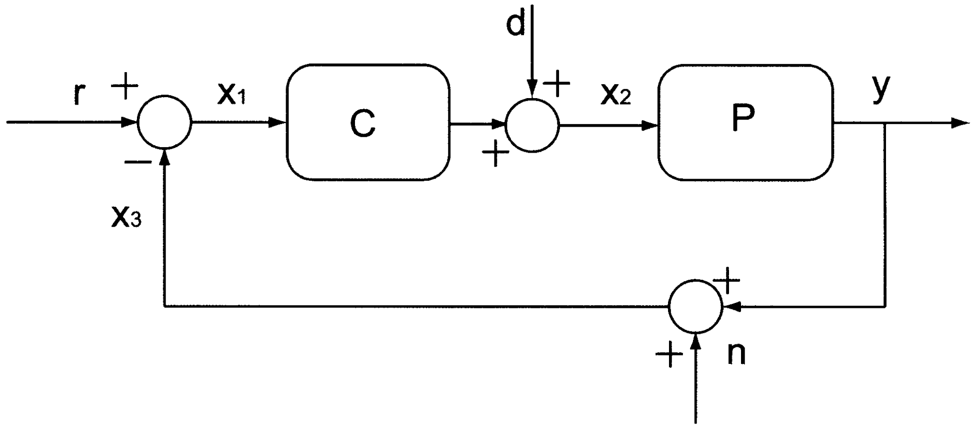
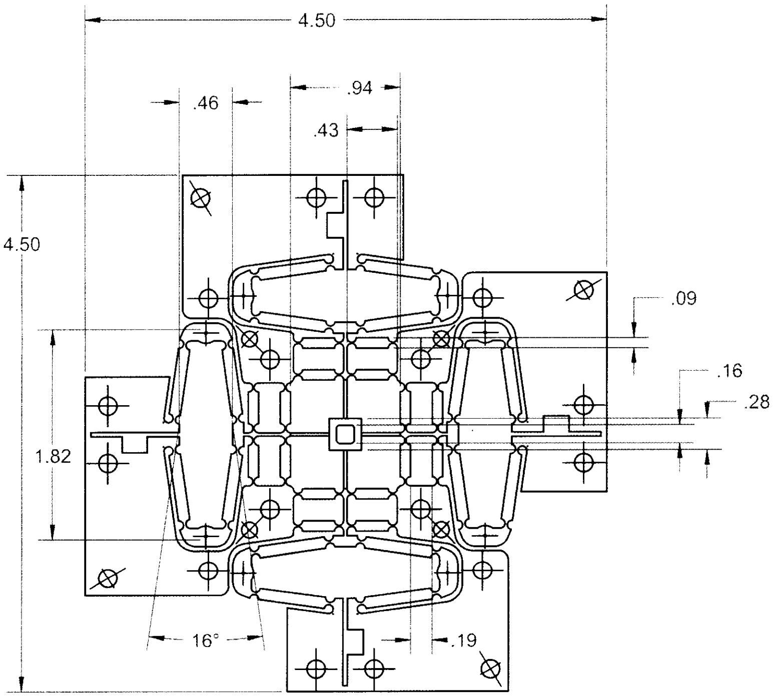


Figure A-1: Internal stability of a feedback system illustrated with a one-DOF control architecture. Figure reproduced from [81].

Appendix B

Design Drawing



MIT AFM Nanopositioner
All dimensions are in inches

Figure B-1: CAD Drawing of AFM Flexure Nanopositioner.

Appendix C

MATLAB Code

In the following pages, we present the MATLAB mfiles and function files used for implementing the design and control optimizations of the thesis.

C.1 One-DOF Positioning System Example

The following .m-files implementing the integrated design and control methodology for simple 1-DOF positioning system detailed in Chapter 5:

```
%%%%%%%%%%%%%%%%%%%%%%%
%main file calling MATLAB non-linear optimization routine fmincon
%%%%%%%%%%%%%%%%%%%%%%%
clear all;
close all;
global P Nz N;
global rho E Es sigma_max sigma_rod_max;
global La Ls B W Wl F r0 D0 m J
global wbs xmax fmax ymin piezoymax;
```

```
P=4; % P=1,2,3,4 parallel stacking index
Nz=1; %Nz=1 when rod flexure is added, otherwise 0



%material constants of Al 70705
rho=2800; %density of Al7075
E=70e9;%Young's Modulus of Al 7075
sigma_yield=550e6; %Yield stress limit of Al 7075
sigma_endurance=100e6;%Endurance stress limit of Al 7075
sf=4; %Factor of Safety
sigma_max=min(sigma_yield,sigma_endurance)/sf;
ymin=50e-6; %Minimum output displacement
piezoymax=18e-6; %Maximum piezo deflection

%material constants of steel for rod flexure (music wired)
Es=200e9;%Young's Modulus of steel
sigma_yield_steel=500e6; %Yield stress limit of steel
sigma_endurance_steel=200e6;%Endurance stress limit of steel
sigma_rod_max=min(sigma_yield_steel,sigma_endurance_steel)/sf;



%assumption
La=0.5*25.4e-3; %actuator placement %half length of central disk
B=0.25*25.4e-3; % thickness of central mass
W=0.25*25.4e-3; % width of central mass
Ls=2*25.4e-3; % sensor placement
```

```
md=rho*2*La*B*W; %mass of central disk
Wl=.2*25.4e-3; %width of lever arm
ml=rho*(Ls+La)*B*Wl; %mass of lever arm
m=ml+md; %total lumped mass

Jd=md/12*(4*La^2+W^2);
Jl=1/12*ml*(Wl^2+Ls^2)+ml*(W/2+Wl/2)^2+ml*(Ls-La)^2/4;
J=Jd+Jl;

r0=0.02*25.4e-3; %radius of rod flexure
D0=4*25.4e-3; %maximum footprint

wbs=1.5e3*2*pi; %desired bandwidth
fmax=850; %piezo blocking force
xmax=1e-6; %maximum lateral displacement error

%bounds
lmin=0.25*25.4e-3;
hmin=0.05*25.4e-3;
l0min=.25*25.4e-3;



lmax=(D0-2*La)/2;
hmax=lmax/12;
l0max=1*25.4e-3;
```

```
lb=[lmin;

    hmin;

    l0min;

    ];


ub=[lmax;

    hmax;

    l0max;];


%calling fmincon in a for loop for a grid of initial guesses.


Npoints=10;

max=0;

number_optimal=0;

exitflagnegative=0;

options = optimset('Display','iter','MaxFunEvals', 1e8,...

'MaxIter',200, 'TolCon',1e-7,'Tolfun',1e-8,'TolX',1e-8);


%linear constraints: inequality

Aineq=[-0.1,1,0;

       ];

Bineq=[0];


figure(1);

for i=1:Npoints+1

    for j=1:Npoints+1

     for k=1:Npoints+1
```

```
         l=lmin+(i-1)*(lmax-lmin)/Npoints;
         h=hmin+(j-1)*(hmax-hmin)/Npoints;
         l0=l0min+(k-1)*(l0max-l0min)/Npoints;
         initial_guess=[l;h;l0;];
        [optvar,fval,exitflag,output,lambda] = fmincon...
        (@costfnew2,initial_guess,Aineq,Bineq,[],[],lb,ub,...
        @nonlinearcnew2,options);
        l_optimal=optvar(1);
        h_optimal=optvar(2);
        l0_optimal=optvar(3);

%        scatter3(l/25.4e-3,h/25.4e-3,l0/25.4e-3,50,'bo');
 %       hold on;

if exitflag>0 optimal_sol(:,i,j,k)=[l_optimal;h_optimal;l0_optimal];
    initial_guess_for_optimal(:,i,j,k)=[l;h;l0];
    output_detail(:,i,j,k)=output;
    yout_optimal(i,j,k)=-fval;
    number_optimal=number_optimal+1;
    exitflag_vector(i,j,k)=exitflag;
        scatter3(l_optimal/25.4e-3,h_optimal/25.4e-3,l0_optimal...
        /25.4e-3,50,'bo');
        hold on;
    if max<-fval
        max=-fval;
        max_optimal_sol=[l_optimal;h_optimal;l0_optimal];
        max_optimal_sol_indices=[i;j;k];
```

```
    end;


    if max==-fval scatter3(l_optimal/25.4e-3,h_optimal/25.4e-3,...
    l0_optimal/25.4e-3,50,'r*'); hold on;
    end;
end;
if exitflag<0 exitflagnegative=exitflagnegative+1;
end;


     end;
    end;
end;


%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%
%Outputs for optimal solution case
l=max_optimal_sol(1);
h=max_optimal_sol(2);
l0=max_optimal_sol(3);


%geometry for side flexure
I=1/12*B*h^3;
%geometry of y flexure
A0=pi*r0^2;
I0=pi*r0^4/4;


%%Stiffness
```

```
%side flexures

ky=P*24*E*I/l^3;

kt=P*(8*E*I/l^3)*(l^2+3*La^2+3*l*La);


%y flexure

k0y=Es*A0/l0;

k0t=Es*I0/l0;

%total stiffness

ky1=ky+Nz*k0y;

kt1=kt+Nz*k0t;


%damping

by=1e-2;

bt=1e-6;

by1=Nz*2e-2;

bt1=Nz*2e-6;


f=fmax/(1+(fmax/piezoymax)*(1/ky1+La^2/kt1));

y0=(Ls*La/kt1-1/ky1)*f;

yi=(1/ky1+La^2/kt1)*f;

%yield stress in side flexures

sigma=f*La*h/4/I/P;

%yield in y flexures

theta=f*La/(kt1);

sigma_y=2*(k0t*theta)*r0/I0;

%buckling critical load in y flexure

Fb=pi^2*Es*I0/(0.7*l0)^2;
```

```
Fvertical_max=Fb/4; %safety factor =4;

%nonlinear constraints: inequality
C1=sigma-sigma_max;
C2=sigma_y-sigma_rod_max;
C3=f-Fvertical_max;
C4=ymin-y0;
C5=yi-piezoymax;
kx=P*2*E*B*h/l+Nz*12*E*I0/l^3;
C6=-xmax+f*asin(5*pi/180)/kx+Ls*theta^2/2;

c=[C1;C2;C3;C4;C5;C6];
ceq=[];

[c,ceq]=nonlinearcnew2(max_optimal_sol);


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Controller design: sensitivity tuning

A1=[0 1 0 0;
   -ky1/m -by1/m 0 0;
   0 0 0 1;
   0 0 -kt1/J -bt1/J];
```

```
B=[0;
   1/m;
   0;
   La/J];

C=[-1 0 Ls 0;];
D=[0];

s=tf('s');
I44=eye(4,4);
G=minreal(C*inv(s*I44-A1)*B);
[pG,yG]=pzmap(G);

f=10; M=1.5; A=1e-1;
Ws=(s+M*wbs)*(s+f*M*wbs)/(s)/(s+f*M^2*wbs);
C0=1e3/s;
S0=1/(1+G*C0);
Sd=1/Ws;
Q=minreal(Sd/S0);
tau=1e3*2*pi*100;
Sf=S0*Q;
C=minreal((1/Sf-1)/G)*1/(s/tau+1)^2;

Marg=allmargin(G*C)
 [Gm,Pm,Wcg,Wcp1]=margin(G*C);
Ws=(s+M*wbs)*(s+f*M*wbs)/(s+A)/(s+f*M^2*wbs);
Wt=1000^2*(s+wbs)*(s+f*wbs)/(s+1000*f*wbs)/(s+1000*wbs);
```

```
[khinf,cl,gam]=mixsyn(G,Ws,[],Wt);

khinf=minreal(khinf);

Marg=allmargin(G*khinf)

[Gm,Pm,Wcg,Wcp2]=margin(G*khinf);


figure(2);

bode(S0,'b');

hold on;

Sf2=minreal(1/(1+G*C));

bode(1/Ws,'m');

hold on;

bode(Sf,'g');

hold on;

bode(Sf2,'r');

hold on;

L=G*khinf;

S=1/(1+L);

bode(S,'k');


figure(3);

bode(khinf);

hold on;

bode(C,'r');


figure(4);

pzmap(khinf);

hold on;
```

```
pzmap(C,'r');

%%%%%%%%%%%%%%%%%%

%function nonlinearcnew2

function [c,ceq]=nonlinearcnew2(optvar);;

global P Nz N;

global rho E Es sigma_max sigma_rod_max;

global La Ls B W Wl F r0 D0 m J

global wbs xmax fmax ymin piezoymax;


l=optvar(1);

h=optvar(2);

l0=optvar(3);



%geometry for side flexure

I=1/12*B*h^3;

%geometry of y flexure

A0=pi*r0^2;

I0=pi*r0^4/4;


%%Stiffness

%side flexures

ky=P*24*E*I/l^3;

kt=P*(8*E*I/l^3)*(l^2+3*La^2+3*l*La);


%y flexure

k0y=Es*A0/l0;
```

```
k0t=Es*I0/l0;

%total stiffness

ky1=ky+Nz*k0y;

kt1=kt+Nz*k0t;


%damping

by=1e-2;

bt=1e-6;

by1=Nz*2e-2;

bt1=Nz*2e-6;


f=fmax/(1+(fmax/piezoymax)*(1/ky1+La^2/kt1));

y0=(Ls*La/kt1-1/ky1)*f;

yi=(1/ky1+La^2/kt1)*f;

%yield stress in side flexures

sigma=f*La*h/4/I/P;

%yield in y flexures

theta=f*La/(kt1);

sigma_y=2*(k0t*theta)*r0/I0;

%buckling critical load in y flexure

Fb=pi^2*Es*I0/(0.7*l0)^2;

Fvertical_max=Fb/4; %safety factor =4;


%nonlinear constraints: inequality

C1=sigma-sigma_max;

C2=sigma_y-sigma_rod_max;

C3=f-Fvertical_max;
```

```
C4=ymin-y0;

C5=yi-piezoymax;

kx=P*2*E*B*h/l+Nz*12*E*I0/l^3;

C6=-xmax+f*asin(5*pi/180)/kx+Ls*theta^2/2;


c=[C1;C2;C3;C4;C5;C6];

ceq=[];

%%%%%%%%%%%%%%%%%%%%

%function costfnew


function [y0]=costfnew2(optvar);

global P Nz N;

global rho E Es sigma_max sigma_rod_max;

global La Ls B W Wl F r0 D0 m J

global wbs xmax fmax ymin piezoymax;


l=optvar(1);

h=optvar(2);

l0=optvar(3);



%geometry for side flexure

I=1/12*B*h^3;

%geometry of y flexure

A0=pi*r0^2;

I0=pi*r0^4/4;
```

```
%%Stiffness
%side flexures
ky=P*24*E*I/l^3;
kt=P*(8*E*I/l^3)*(l^2+3*La^2+3*l*La);


%y flexure
k0y=Es*A0/l0;
k0t=Es*I0/l0;
%total stiffness
ky1=ky+Nz*k0y;
kt1=kt+Nz*k0t;


%damping
by=1e-2;
bt=1e-6;
by1=Nz*2e-2;
bt1=Nz*2e-6;


%cost
f=fmax/(1+(fmax/piezoymax)*(1/ky1+La^2/kt1));
y0=-(Ls*La/kt1-1/ky1)*f;
```

C.2 AFM Nanopositioner Example

C.2.1 Shape and Size Optimization of AFM Nanopositioner

The following are the *.m* files used for the shape and size optimization of the AFM nanopositioner described in Chapter 6.

```
%main file calling fmincon function

close all;

clear all;


global Ef N A B C kb df kwa kwb kp d0 fb;

global fzdesired xrangedesired;

global m;


%objective

fzdesired=35e3;

xrangedesired=50e-6;


%material properties

Ef=2.1e11;

Ea=72e9;

Sa=550e6;


fb=850;

N=2;

kp=50e6;

m=5e-3;


%assumed geometry parameters

df=0.5e-3;

bw=1.35e-3;

hw=0.6e-3;
```

```
lw1=sqrt(6*Ea*xrangedesired*hw/Sa)+2e-3

lw=8.5e-3;


%wire flexurs between chevron and central stage

kwb=2*Ea*bw*hw^3/lw^3;

kwa=Ea*bw*hw/lw;


%box flexures in x

%notch stiffness

b_notch=.25*25.4e-3;

r_notch=5/128*25.4e-3;

t_notch=.4e-3;

l_b=8.98e-3;


s=r_notch./t_notch;

%geometry factor that appears in notch stiffness

f21=2*s.^3.*(6*s.^2+4*s+1)./((2.*s+1).*(4*s+1).^2);

f22=(12*s.^4.*(2*s+1)./(4*s+1).^2.5).*atan(sqrt(4*s+1));

f2=f21+f22;


knotchb=(Ea*b_notch.*r_notch.^2/12).*(1./f2);

kb=(4*knotchb./(l_b.^2));



%Calculation of stiffness matrix of chevron

%Deflections based on FEA of chevron with...

%varying angle alpha
```

```
%k2=fin/din, with dout=0;

%A=[alpha,din]

A=[0.1 2.22E-08

2 2.19E-08

4 2.17E-08

6 2.16E-08

8 2.15E-08

10 2.15E-08

12 2.16E-08

14 2.17E-08

16 2.19E-08

18 2.22E-08

20 2.26E-08

22 2.30E-08

24 2.36E-08

26 2.43E-08

28 2.51E-08

30 2.61E-08

32 2.73E-08

34 2.87E-08

36 3.04E-08

38 3.25E-08

40 3.48E-08

];


%k1=fout/dout with din=0;
```

```
%B=[alpha,dout]

B=[0.1 6.99E-06

2 4.43E-06

4 2.15E-06

6 1.17E-06

8 7.17E-07

10 4.82E-07

12 3.46E-07

14 2.62E-07

16 2.04E-07

18 1.65E-07

20 1.38E-07

22 1.16E-07

24 9.99E-08

26 8.73E-08

28 7.79E-08

30 6.99E-08

32 6.45E-08

34 6.11E-08

36 5.67E-08

38 5.18E-08

40 4.97E-08

];


%Kout calculated from Fin=0;Fout=1N

%C=[alpha,dout]

C=[0.1 1.42E-05
```

```
2 1.44E-05

4 1.41E-05

6 1.40E-05

8 1.41E-05

10 1.43E-05

12 1.41E-05

14 1.45E-05

16 1.45E-05

18 1.47E-05

20 1.50E-05

22 1.54E-05

24 1.61E-05

26 1.65E-05

28 1.70E-05

30 1.80E-05

32 1.88E-05

34 1.95E-05

36 2.07E-05

38 2.28E-05

40 2.41E-05

];


%Calculation of k1, k2, k3 as a function of

%chevron angle

phi=0:0.01:40*pi/180;

k2poly=polyfit(A(:,1),1./A(:,2),4);

k2=polyval(k2poly,phi*180/pi);
```

```
k1poly=polyfit(B(:,1),1./B(:,2),4);
k1=polyval(k1poly,phi*180/pi);

koutpoly=polyfit(C(:,1),1./C(:,2),2);
kout=polyval(koutpoly,phi*180/pi);

k3=-sqrt(k1.*k2-k2.*kout);

%initial guess
LB=[1;.25;5];
UB=[64;1;10];
X0=(LB+UB)/2;

options = optimset('Display','iter','TolFun',1e-9);

[x,fval,exitflag,output,lambda] = FMINCON(@costfunction,...
    X0,[],[],[],[],LB,UB,@nonlinearconstraints_scanner,options)

n=ceil(x(1));
lf=x(2)*25.4e-3;
phi=x(3)*pi/180;

Af=pi/4*df^2;
kfz=n*Ef*Af/lf;
fz=(1/(2*pi))*sqrt(kfz/m);
```

```
k1=polyval(k1poly,phi*180/pi);

k2=polyval(k2poly,phi*180/pi);

kout=polyval(koutpoly,phi*180/pi);

k3=-sqrt(k1.*k2-k2.*kout);


%x stiffness

If=pi/64*df^4;

kfx=n*12*Ef*If/lf^3;

kL=(kwb)*(1+2*kb/kwa)+4*kb+kfx*(1+2*kb/kwa);

s=(kwb+kfx)/kwa+2;

ks=(k1*k2-k3^2+k1*kp/N)/(k2+kp/N);

xrange=((-2*k3*fb./(k2+kp/N))/(kL+2*s.*ks))*1e6


%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%

%Function file specifying cost function

function f=costfunction(x);


global Ef N A B C kb df kwa kwb kp d0 fb;

global fzdesired xrangedesired;

global m;


%assign

n=x(1);

lf=x(2)*25.4e-3;

phi=x(3)*pi/180;

%
```

```
k2poly=polyfit(A(:,1),1./A(:,2),4);
k2=polyval(k2poly,phi*180/pi);

k1poly=polyfit(B(:,1),1./B(:,2),4);
k1=polyval(k1poly,phi*180/pi);
%
koutpoly=polyfit(C(:,1),1./C(:,2),4);
kout=polyval(koutpoly,phi*180/pi);

k3=-sqrt(k1.*k2-k2.*kout);

%x stiffness
If=pi/64*df^4;
kfx=n*12*Ef*If/lf^3;

kL=(kwb)*(1+2*kb/kwa)+4*kb+kfx*(1+2*kb/kwa);

s=(kwb+kfx)/kwa+2;
ks=(k1*k2-k3^2+k1*kp/N)/(k2+kp/N);

m=5e-3;

If=pi/64*df^4;
kfx=n*12*Ef*If/lf^3;

kL=(kwb)*(1+2*kb/kwa)+4*kb+kfx*(1+2*kb/kwa);
s=(kwb+kfx)/kwa+2;
```

```
ks=(k1*k2-k3^2+k1*kp/N)/(k2+kp/N);
xrange=((-2*k3*fb)/(k2+kp/N))./(kL+2*s*ks);

theta_arc=asin(xrange/2/lf);
lift=lf*(1-cos(theta_arc));

fx=(1/(2*pi))*sqrt((kL+2*s*ks)/m);
Af=pi/4*df^2;
kfz=n*Ef*Af/lf;

f=lift;

%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%
%Function file specifying nonlinear constraints
function [c,ceq]=nonlinearconstraints_scanner(x);

global Ef N A B C kb df kwa kwb kp d0 fb;
global fzdesired xrangedesired;
global m;

n=x(1);
lf=x(2)*25.4e-3;
phi=x(3)*pi/180;

k2poly=polyfit(A(:,1),1./A(:,2),4);
k2=polyval(k2poly,phi*180/pi);
```

```
k1poly=polyfit(B(:,1),1./B(:,2),4);
k1=polyval(k1poly,phi*180/pi);

koutpoly=polyfit(C(:,1),1./C(:,2),4);
kout=polyval(koutpoly,phi*180/pi);

k3=-sqrt(k1.*k2-k2.*kout);

%x range
If=pi/64*df^4;
kfx=n*12*Ef*If/lf^3;

kL=(kwb)*(1+2*kb/kwa)+4*kb+kfx*(1+2*kb/kwa);
s=(kwb+kfx)/kwa+2;
ks=(k1*k2-k3^2+k1*kp/N)/(k2+kp/N);

xrange=((-2*k3*fb)/(k2+kp/N))./(kL+2*s*ks);

c(1)=xrangedesired-xrange;

% fz
m=5e-3;
Af=pi/4*df^2;
kfz=n*Ef*Af/lf;
fz=(1/(2*pi))*sqrt(kfz/m);
c(2)=fzdesired-fz;
```

```
ceq=[];
```

C.2.2 AFM $\mathcal{H}_2$-optimal Model-Matching

```
close all;
clear all;
global magGxx G fs sizefs;


A=[
];
%Data collected for AFM lateral frequency response
%Not included here because of space constraints


sizeA=size(A);
s=tf('s');
fs=A(:,2);
magGxx=20*log10(A(:,3));
phaseGxx=A(:,4);


figure(1);
subplot(2,1,1)
semilogx(fs,magGxx,'r.');
grid on;
hold on;


subplot(2,1,2)
semilogx(fs,phaseGxx,'r.');
grid on;
```

```
hold on;

sizefs=size(fs);

for j=1:sizefs(1)
 realG(j)=A(j,3)*cos(A(j,4)*pi/180);
 imagG(j)=A(j,3)*sin(A(j,4)*pi/180);
 G(j)=complex(realG(j),imagG(j));
end;

lb=[.01;2*pi*745;0.005;2*pi*2300;.013;8];
ub=[.05;2*pi*755;0.01;2*pi*2335;.015;10];

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
e10=0.04;
w10=2*pi*750;
e20=0.008;
w20=2*pi*2330;

a0=10^(-37.23/20);
b0=9;

initial_guess=[e10;w10;e20;w20;a0;b0];

P10=tf([w10^2],[1 2*e10*w10 w10^2]);
P20=tf([w20^2],[1 2*e20*w20 w20^2]);
P0=a0*(1*P10+b0*P20);
```

```
[m0,p0]=bode(P0,2*pi*fs);
m0=squeeze(m0);
p0=squeeze(p0);
subplot(2,1,1);
semilogx(fs,20*log10(m0),'k-.','linewidth',2);
grid on;
hold on;

axis tight;
subplot(2,1,2);
semilogx(fs,p0,'k-.','linewidth',2);
grid on;
hold on;
axis tight;

cost0=costmatch3(initial_guess)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%5
options = optimset('Display','iter','MaxFunEvals', 1e8,...
'MaxIter',200, 'TolCon',1e-7,'Tolfun',1e-13,'TolX',1e-8);
[optvar,fval,exitflag,output,lambda] = fmincon(@costmatch3,...
initial_guess,[],[],[],[],lb,ub,@nonlinearmatch3,options);

e1=optvar(1);
w1=optvar(2);
e2=optvar(3);
w2=optvar(4);
```

```
a=optvar(5);

b=optvar(6);


P1=tf([w1^2],[1 2*e1*w1 w1^2]);

P2=tf([w2^2],[1 2*e2*w2 w2^2]);

P=a*(P1+b*P2);


[m,p]=bode(P,2*pi*fs);

m=squeeze(m);

p=squeeze(p);


subplot(2,1,1);

semilogx(fs,20*log10(m),'b','linewidth',2);

grid on;

hold on;

legend('Measured Data','Initial Guess','H2-optimal Model');

axis tight;

subplot(2,1,2);

semilogx(fs,p,'b','linewidth',2);

grid on;

hold on;

axis tight;

axis([19.53 20e3 -360 0]);


for j=1:sizefs(1)

 realP(j)=m(j)*cos(p(j)*pi/180);

 imagP(j)=m(j)*sin(p(j)*pi/180);
```

```
 Pc(j)=complex(realP(j),imagP(j));
end;



cost=0;
delta_f=sizefs/20000;
for j=1:sizefs(1)
if fs(j)>5e3 weight(j)=0;
else weight(j)=1;
end;
cost=cost+(G(j)-Pc(j)).*conj(G(j)-Pc(j))*weight(j)*delta_f;
end;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%5
%Function file specifying cost function
function [cost]=costmatch3(optvar);
global magGxx G fs sizefs;

e1=optvar(1);
w1=optvar(2);
e2=optvar(3);
w2=optvar(4);
a=optvar(5);
b=optvar(6);

s=tf('s');
```

```
P1=tf([w1^2],[1 2*e1*w1 w1^2]);

P2=tf([w2^2],[1 2*e2*w2 w2^2]);

P=a*(1*P1+b*P2);


[m,p]=bode(P,2*pi*fs);

m=squeeze(m);

p=squeeze(p);


for j=1:sizefs(1)

 realPc(j)=m(j)*cos(p(j)*pi/180);

 imagPc(j)=m(j)*sin(p(j)*pi/180);

 Pc(j)=complex(realPc(j),imagPc(j));

end;


%minimization of H2norm defined as integral of (G-Pc).*conj(G-Pc) over frequency
fs

cost=0;

delta_f=sizefs(1)/20000;

for j=1:sizefs(1)

if fs(j)>5e3 weight(j)=0;

else weight(j)=1;

end;

cost=cost+(G(j)-Pc(j)).*conj(G(j)-Pc(j))*weight(j)*delta_f;

end;


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Function file specifying constraint function
```

```
function [c,ceq]=nonlinearmatch(optvar);
%global A magGxx fs a b;
global magGxx G fs sizefs;

e1=optvar(1);
w1=optvar(2);
e2=optvar(3);
w2=optvar(4);
a=optvar(5);
b=optvar(6);

s=tf('s');
P1=tf([w1^2],[1 2*e1*w1 w1^2]);
P2=tf([w2^2],[1 2*e2*w2 w2^2]);
P=a*(1*P1+b*P2);

[m,p]=bode(P,2*pi*fs);
m=squeeze(m);
p=squeeze(p);
c=[];
ceq=20*log10(m(1))-magGxx(1);
```

C.2.3 AFM Sensitivity Tuning

```
close all;

clear all;
```

```
global S0 fs ms0 fsd fsdsize Sdesired;

s=tf('s');

%plant from H2-optimal model based on experimental
%lateral frequency response data

P=(2.699e07*s^2+3.784e09*s+6.549e14)...
/(s^4+431.8*s^3+2.372e08*s^2+3.62e10*s+4.716e15);



%Nominal Sensitivity
C0=1000/s;
S0=(1/(1+P*C0));
fs=10.^(1:0.0001:4);

[ms0,ps0]=bode(S0,2*pi*fs);

ms0=squeeze(ms0);

ps0=squeeze(ps0);

PC0_performance=allmargin(P*C0);

%Desired Sensitivity obtained by gridding
fsd1=10.^(1:0.005:3.357);
```

```
fsd2=10.^3.3685;

fsd3=10.^(3.38:0.005:4);

[msd1,psd1]=bode(S0,2*pi*fsd1);

msd1=squeeze(msd1);

psd1=squeeze(psd1);

[msd2,psd2]=bode(S0,2*pi*fsd2);

msd2=max(ms0)/1.5;

msd2=squeeze(msd2);

psd2=squeeze(psd2);

[msd3,psd3]=bode(S0,2*pi*fsd3);

msd3=squeeze(msd3);

psd3=squeeze(psd3);

msd=[msd1; msd2; msd3];

phsd=[psd1; psd2; psd3];


fsd=[fsd1'; fsd2; fsd3'];


fsd1size=size(fsd1);


peakfrequency=fsd1size(1)+1;


fsdsize=size(fsd);

for j=1:fsdsize(1);

realSd(j)=msd(j).*cos(phsd(j)*pi/180);

imagSd(j)=msd(j).*sin(phsd(j)*pi/180);

Sdesired(j)=complex(realSd(j),imagSd(j));

end;
```

```
figure(1);
semilogx(fsd,20*log10(msd),'bo');

grid on;
hold on;
semilogx(fs,20*log10(ms0),'r');

hold on;

figure(2);
plot(fsd,20*log10(msd),'bo','linewidth',2);

hold on;
plot(fs,20*log10(ms0),'r','linewidth',2);

hold on;
axis([2100 2700 -0.2 4.7]);

%calling optimization routine
e10=0.006;
w10=2.32e3*2*pi;
e20=0.008;
w20=2.34e3*2*pi;

initial_guess=[e10;w10;e20;w20];
lb=[0.005; 2.30e3*2*pi; 0.005; 2.33e3*2*pi];
```

```
ub=[0.01; 2.35e3*2*pi; 0.01; 2.35e3*2*pi];

options = optimset('Display','iter','MaxFunEvals', 1e8,...

'MaxIter',200, 'TolCon',1e-7,'Tolfun',1e-13,'TolX',1e-8);

[optvar,fval,exitflag,output,lambda] = fmincon...
(@cost_sensitivity,initial_guess,[],[],[],[],...
lb,ub,@nonlinear_constraint_sensitivity,options);

%optvar=initial_guess;
e1=optvar(1);
w1=optvar(2);
e2=optvar(3);
w2=optvar(4);

Q=(s^2+2*e1*w1*s+w1^2)/(s^2+2*e2*w2*s+w2^2);

S=S0*Q;
[ms,ps]=bode(S,2*pi*fsd);
ms=squeeze(ms);
ps=squeeze(ps);
[mss,ps]=bode(S,2*pi*fs);
mss=squeeze(mss);

figure(3);
semilogx(fsd,20*log10(msd),'bo','linewidth',2);
```

```
grid on;
hold on;
semilogx(fs,20*log10(ms0),'r','linewidth',2);

hold on;
semilogx(fs,20*log10(mss),'k','linewidth',2);

hold on;

figure(4);
plot(fs,20*log10(ms0),'r','linewidth',2);

hold on;
plot(fsd,20*log10(msd),'bo','linewidth',2);

hold on;
plot(fs,20*log10(mss),'k','linewidth',2);

hold on;
legend('Nominal Sensitivity S_0(j\omega)','Desired...
Sensitivity S_d (j\omega)','H_2 Optimal...
Sensitivity S_{opt}(j\omega)');
axis([2100 2700 -0.2 4.7]);

optvar=initial_guess;
e1=optvar(1);
```

```
w1=optvar(2);

e2=optvar(3);

w2=optvar(4);

Q0=(s^2+2*e1*w1*s+w1^2)/(s^2+2*e2*w2*s+w2^2);

SQ0=S0*Q0;

[msq0,psq0]=bode(SQ0,2*pi*fsd);

msq0=squeeze(msq0);

psq0=squeeze(psq0);


maxwithQ0=20*log10(max(msq0))


figure(5);

plot(fsd,20*log10(msd)-20*log10(ms),'linewidth',2);


hold on;


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Cost function for minimzing H2-norm of error between sensitivity

$and desired sensitivity:


function [cost]=cost_sensitivity(optvar);

global S0 fs ms0 fsd fsdsize Sdesired;


e1=optvar(1);

w1=optvar(2);

e2=optvar(3);

w2=optvar(4);
```

```
s=tf('s');


Q=(s^2+2*e1*w1*s+w1^2)/(s^2+2*e2*w2*s+w2^2);


S=minreal(S0*Q);
[ms,ps]=bode(S,2*pi*fsd);
ms=squeeze(ms);
ps=squeeze(ps);

for j=1:fsdsize(1);
realSf(j)=ms(j).*cos(ps(j)*pi/180);
imagSf(j)=ms(j).*sin(ps(j)*pi/180);
Sf(j)=complex(realSf(j),imagSf(j));
end;

cost=0;
delta_f=10^.005;
for j=1:fsdsize(1)
cost=cost+(Sdesired(j)-Sf(j)).*conj(Sdesired(j)-Sf(j))*delta_f;
end;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Non-linear constraint function for AFM nanopositioner sensitivity tuning

function [c,ceq]=nonlinear_constraint_sensitivity(optvar);
global S0 fs ms0 fsd fsdsize Sdesired;
```

```
s=tf('s');


e1=optvar(1);

w1=optvar(2);

e2=optvar(3);

w2=optvar(4);


Q=(s^2+2*e1*w1*s+w1^2)/(s^2+2*e2*w2*s+w2^2);


S=S0*Q;


[ms,ps]=bode(S,2*pi*fsd);

ms=squeeze(ms);

ps=squeeze(ps);


c=3-(20*log10(max(ms0))-20*log10(max(ms)));

ceq=[];
```

Appendix D

Piezoelectric Actuators: Guidelines

In this section, we present a practical overview of piezoelectric actuators. We highlight some of the practical guidelines (do's and don'ts) when handling them.

D.1 Working Principle

Piezoelectric actuators are known for their high-resolution (on the nano-scale) motion capabilities. They are based on the principle of piezoelectricity, a property of materials such as Lead Zirconate Titanate (PZT). Piezoelectricity is the ability of generating electric field from a stored mechanical energy, such as strain energy stored in the material under an applied mechanical load. The reverse effect of generating a mechanical strain under an applied electrical field, or voltage, is also possible, and is the underlying principle of piezoelectric actuators.

From an energy viewpoint, neglecting losses, the piezoelectric material is a transformer that converts a mechanical energy input into an electrical energy output. In the mechanical domain the energy gets stored in the linear stiffness of the stack elements

Example configurations of piezoelectric actuators are stack actuators and tube actuators. In piezoelectric stack actuators, multiple layers of the material are added mechanically in series so that their displacements add up, and electrically in parallel so that they are driven under the same applied voltage. They typically are used for linear actuation.

In piezoelectric tube actuators, a thin walled tube is provided with discrete electrodes on the inner and outer surfaces of the tube. Longitudinal actuation is achieved by applying a voltage to the inner electrodes. Lateral actuation is achieved by a differential voltage applied to opposing outer electrodes.

D.2 Force Displacement Characteristic Curves

Similar to torque-speed characteristic curves of motors, piezoelectric actuators have a force-displacement characteristic curve. An example force-displacement characteristic graph for piezoelectric stack actuators is schematically shown in Fig. D-1. A few terms need to be explained here. When no force is applied to the free end of the piezoelectric stack actuator, it expands to a maximum of y_{max}, which is called the free deflection and noted as the y-intercept of the line in Fig. D-1. On the other hand, if the piezoelectric actuator is blocked completely from expanding by imposing rigid boundary conditions, it generates a maximum force, known as the blocking force. This is indicated as F_{max} in the figure. This is equivalent to the stall torque of a motor.

For any operating point chosen (δ, f) on this characteristic, the stiffness of the structure is set at $k_{load} = \frac{f}{\delta}$. Note that both f_{max} and δ_{max} scale linearly with the applied voltage. The area under the force displacement characteristic is a maximum when the voltage applied to the piezoelectric actuator is its rated maximum value.

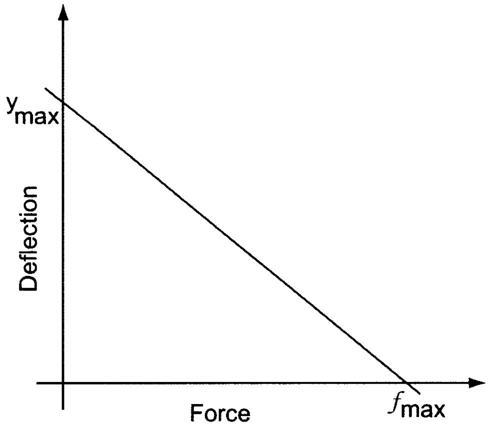


Figure D-1: Typical static force-deflection characteristic graph of a piezoelectric stack actuator. The piezoelectric stack actuator represented in this characteristic graph has a maximum force capacity (blocking force) F_{max} and free deflection y_{max}.

D.3 Sizing of Piezoelectric Actuators

Selecting a piezoelectric stack actuator suitable for a particular application involves multiple decisions. They are listed below:

- Deflection: Typical piezoelectric crystals like PZT are marked by strains of about 0.1%. To get large extensions of the piezoelectric actuator, higher strain materials, such as presented in [13], are available. However, such materials are still subjects of research with limited commercial suppliers. To get large deflections, a series combination of stacks be used to can sum up their displacements, at the same blocking force. Further. amplification mechanisms can be designed for improving on the available deflection. But this comes at the cost of reduced output force.

- Preload: Most commercial piezoelectric actuators come with a casing in which the stack is preloaded against a metal casing with flexure or spring elements. If a raw stack is used, sufficient preload must be provided to prevent loss of interface contact between the actuator and the load. Preload is also needed for protecting the piezoelectric actuator from tension loads. A typical value suggested for preload is $20 - 50\%$ of the blocking force.

- Natural frequency: The shorter the stack the higher the natural frequency. Ideally the stack should be chosen such that its natural frequency is higher than the bandwidth of interest.
- Configuration selection: One design choice for using piezoelectric actuators in flexure-based mechanisms is selecting a stack versus tube. Piezoelectric tubes, found in most commercial AFMs, have a bending mode in addition to the longitudinal mode. Piezoelectric stacks are preferred since they tend to be more compact and are often used when motion along a single axis is desired. Since flexure design topologies can be chosen to parallel-kinematic with decoupled motion axes, single-axis piezoelectric stack actuators along each motion axis is one preferred approach for driving the mechanism.
- Capacitance: The piezoelectric actuator has an electrical capacitance that stores the electrical energy of system. Power amplifiers have a bandwidth that is limited by the capacitance of the piezoelectric stack actuator. Typically, a roll-off such as 20 dB/dec is observed after cut-off frequencies, which depend on the capacitance of the actuator. The larger the capacitance to be driven, the earlier the cut-off frequency occurs. Hence, it is increasingly hard to drive piezoelectric actuators that have high capacitances.

D.4 Practical Aspects

The following is a list of practical do's and don'ts that are helpful when using piezoelectric actuators in a hardware implementation.

- Piezoelectric actuators cannot withstand shear and moment loads. Hence, a rounded tip or a flexure piece should be used to guide the piezoelectric actuator without being affected by the parasitic loads.

- The capacitance between the leads of the piezoelectric actuator must be checked with a voltmeter, before connecting them across the channel of a power amplifier. This is necessary to ensure that short-circuiting is avoided when driving with a power amplifier. If short-circuiting protection is not available on the amplifier, it is likely the amplifier fuse will blow up.

- Soldering leads to piezoelectric stacks should be done carefully, since any heat applied to the stacks can irreversibly affect the piezoelectricity property.

- Commercial ceramic piezoelectric stacks are sometimes available with a resin coating that protects the delicate surfaces of the stack. The stacks that come without the resin coating should be handled carefully since the material tends to brittle and powdery. A stack with a broken surface, or chipped surface, can have short-circuiting of the inner layers and is hence unsuitable for connecting to power amplifiers.

- Exceeding the prescribed voltage range of piezoelectric actuators is likely to cause excessive heating, and irreversibly damage the material.

Appendix E

Vendors

dSPACE Inc.
28700 Cabot Drive - Suite 1100
Novi, MI 48377
Telephone — (248) 567-1300
Web address — http://www.dspaceinc.com/index.htm
Product used: DS1103 Controller Board

The MathWorks, Inc.
3 Apple Hill Dr.
Natick, MA 01760-2098
Telephone — (508)647-7000
Web address — http://www.mathworks.com
Products used: MATLAB, Simulink

NEC Tokin
1600 Golf Road, Corporate Center-Suite 1228, Rolling Meadows, IL 60008 Telephone — (847)981-5047
Web address — http://www.nec-tokin.com
Product used: Piezoelectric Actuator AE0505D16
Distributor in USA: Thorlabs, Inc.

Physik Instrumente
16 Albert Street, Auburn, MA 01501
Telephone — (508)832-3456
Web address — http://www.pi-usa.com
Product used: Piezoelectric Actuator PL022.21, PL033.31, PL885.51

National Instruments Corporation
11500 N Mopac Expressway, Austin, TX 78759-3504
Telephone — (512) 683-0100
Web address —http://www.ni.com
Products used:
Hardware: PXI Chassis 1042, PXI Embedded Controller 8183, Multi-function DAQ 6048
Software: LabVIEW 2009

ADE Tech
80 Wilson Way, Westwood, MA 02090
Telephone — (781)467-3515
Web address — http://www.adetech.com
Product used: Capacitance Probe AD2805

SIOS
SISO Mechtecnik GmbH, Am Vogelherd 46, D98693, Limenau Germany
Telephone — 49(0)3677-64470
Web address — http://www.siso.com
Product used: Single-Axis Laser Interferometeric Vibrometer

TechProject Inc.
Elisenstrasse 23, A-1230 Wien, Osteerreich
Telephone — (0043) 676 427 7064
Web address — http://www.techproject.at
Product used: Six-channel power amplifier tuned for NEC Tokin piezos

Agilent Technologies.
5301 Stevens Creek Blvd, Santa Clara CA 95051
Telephone — (877)424-4536
Web address — http://www.agilent.com
Product used: Dynamic Signal Analyzer 35670A

Ideal Vacuum Products, LLC.
2401B Phoenix Ave NE Albuquerque, NM 87107
Telephone — (505) 872-0037
Web address — http://www.pchemlabs.com/contact.asp
Ideal Vacuum Products, LLC, formerly, PChemLabs Product used: High Vacuum Expoy