ABSTRACT

Title of Document:	DYNAMIC MODELING AND POSITION CONTROL OF A PIEZOELECTRIC FLEXTENSIONAL ACTUATOR
	Benjamin John Nickless, M.S., 2008
Directed By:	Professor James E. Hubbard, Jr. Department of Aerospace Engineering

Many smart material actuators suffer in either insufficient force generation or displacement range, two important performance metrics in actuator design. Piezoelectric flextensional actuators were conceived to bridge the gap between displacement and force, offering acceptable performance in both categories. Their displacement range and load carrying capability make them suitable for many applications requiring micrometer-scale displacements. Typical applications require closed-loop control algorithms to achieve good resolution at these displacement levels. In this thesis, an open-loop model of a commercially available, piezoelectric flextensional actuator and drive system was designed. This model was used to design a feedback control system for reference tracking applications. The control system was built and verified with the physical actuator. Its performance was shown to agree with the model simulations. Both the model and the physical system had negligible overshoot, settling times of less than 30 milliseconds, and zero steady-state error in response to step inputs.

DYNAMIC MODELING AND POSITION CONTROL OF A PIEZOELECTRIC
FLEXTENSIONAL ACTUATOR

By

Benjamin John Nickless

Thesis submitted to the faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Science
2008

Advisory Committee:

    Professor James E. Hubbard, Jr., Chair
    Dr. William Fourney
    Dr. Alison Flatau

Preface

"Circumstances rule men and not men circumstances."

-Herodotus, The Histories

Dedication

      To my father and mother, Randy and Judith Nickless, my sister Jill, and the rest of my family and friends who must go unnamed for fear of missing someone. I hope, despite any flaws that I bear in thought, word, or deed, that you are pleased with who I have become, and that the good in me, and the good that I have done thus far in life, has made you proud to call me family or friend. No matter how near or far I may be, you all have always been, and always will be, in my thoughts and prayers, for I know that I have been in yours. Without your love, support, and the answered prayers you continuously send to God above on my behalf, I could never have made it to where I am, and I would scarcely be able to make it another step.

Acknowledgements

First and foremost, I owe a great debt of gratitude to Dr. Tian-Bing Xu and Dr. Ji Su. Not only did they procure the funding for my graduate education and research, they provided the resources by which I performed it. Furthermore, they provided excellent guidance and insight throughout the entire process. None of this would have been possible without them. Thank you both very much.

I would like to thank Dr. Peter Lillehei for taking time out of his day on several occasions to provide me with much needed, and much appreciated, insight into the expectations and pitfalls of laboratory work and equipment.

Thanks are due to Dr. Alison Flatau and Dr. William Fourney for the patience they have shown in serving on my thesis committee. I would like to extend an extra thanks to Dr. Fourney for providing my first opportunity when I came to the University of Maryland, for it was as his teaching assistant where I met Dr. Hubbard.

Dr. James E. Hubbard, Jr is the director of all of this work. He is the one who gave me this opportunity and introduced me to Dr. Xu and Dr. Su. Furthermore, he was patient and lenient with me when my research and I were going through trying times. He has been a great academic advisor and an excellent mentor, both professionally and personally. I owe him all of the gratitude one can offer to another. Thank you, Dr. Hubbard.

Last but not least, a big thanks to all of my Morpheus people. Nelson Guerreiro, Geoff Slipher, Robyn Harmon, Jared Grauer, Sandra Ugrina, Sarah Haack, David Billingsley, and Kevin Calabro – you all are the best.

Table of Contents

List of Tables

List of Figures

ix

1   Introduction

The research that is presented herein provides a solution to the modeling and precision displacement control of a piezoelectric, flextensional actuator. As this is a multifaceted engineering problem, each aspect of the problem and its solution, as well as any relevant background information, are deserving of their own explanation. Before proceeding to the work itself, it is pertinent to lay groundwork for the ensuing pages to explain why this research was performed and the requisite knowledge needed to do so. The current chapter embodies this groundwork with: 1) a justification for performing this research, 2) a literature review of relevant past and present work, and 3) a summary of the research directive that was followed.

1.1 Rationale for Precision Position Control

The research and technology sector is often application driven. Motivation for the research described herein is rooted in applications that mandate high-precision position control, with accuracy of micro to nanometers. With the rise of fields and devices such as micro and nanoelectromechanical systems (MEMS, NEMS), adaptive optics, optoelectronics, scanning probe microscopy (SPM) and many others, as well as the precision manufacturing required by these fields, applications for high-precision positioning are numerous [1-2]. Among the more prominent applications are probe and sample positions in a SPM, precision cutting and grinding machines, and mirror translation for optical wavefront sensing [1-4]. The research contained

herein warrants an exploration of these applications to understand why they necessitate position control to such accuracy.

1.1.1 SPM Probe/Sample Positioning

In scanning probe microscopy a probe is scanned over a surface in order to generate an image of the surface. Several types of SPM exist. The major traits setting the types of SPM apart are the probe characteristics, e.g. whether it is a contact or non-contact method, or the quantity being detected and used to generate a surface image. Regarding non-contact methods, the probe must be kept within nanometers of the scanned sample without coming into contact with the surface of said sample.

Atomic force microscopy (AFM) and scanning tunneling microscopy (STM) are both types of SPM that operate in a non-contact mode, and both require the high-precision position control described previously. A schematic typical of an AFM is shown in Figure 1.1. As the probe tip comes within close proximity to the sample surface, attractive or repulsive forces at the atomic or molecular level are generated (Van der Waals, magnetic, etc.). These forces can become large enough to result in the collision of probe and sample, thus damaging the probe, the sample, or both. Feedback electronics and piezoelectric actuators must work in harmony to maintain the force at a constant level and prevent such a collision [2].
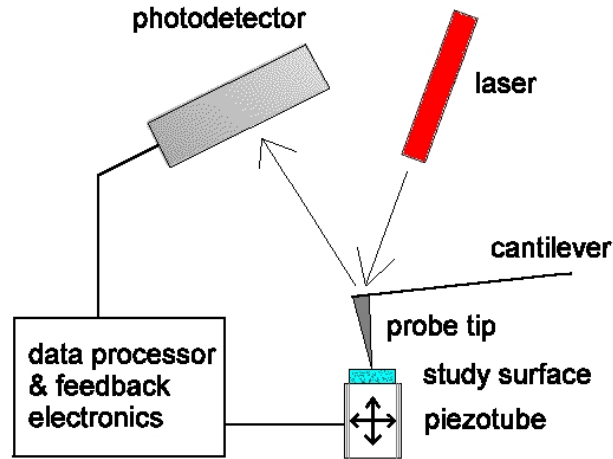
Figure 1.1   SPM schematic typical of AFM [5].

STM shares the general operating principle of AFM with the difference being that the probe resides on the piezoelectric actuator instead of the sample.   The detected quantity for imaging is a weak electric current that is dependent on the distance separating the probe and sample.   This current must remain constant. Therefore, as the surface changes the probe must be moved accordingly to maintain the current at a constant level [5].

Both types of SPM require high-precision positioning. Piezoelectric actuators possess the displacement range needed for both applications, though they typically cannot achieve the requisite accuracy without the use of a feedback control scheme. External disturbances, inherent non-linearities, and stringent specifications often prevent open-loop control schemes from achieving performance goals.

1.1.2 Precision Cutting/Grinding (Micrometer Precision)

Advances in the industries of aerospace, optics, computing, and many others have resulted in component tolerances on the order of micrometers or less.

3

Demanding such accuracy puts great strain on the manufacturing sectors of these industries [3]. Prototyping components to this tolerance can prove difficult without implementing error or vibration compensation systems.

On cutting machines such as lathes or mills, this compensation system can be manifest in the form of mounting the cutting tool onto an actively controlled transducer or tool post. Figure 1.2 illustrates a schematic of such a system including the components most relevant to the actuator.
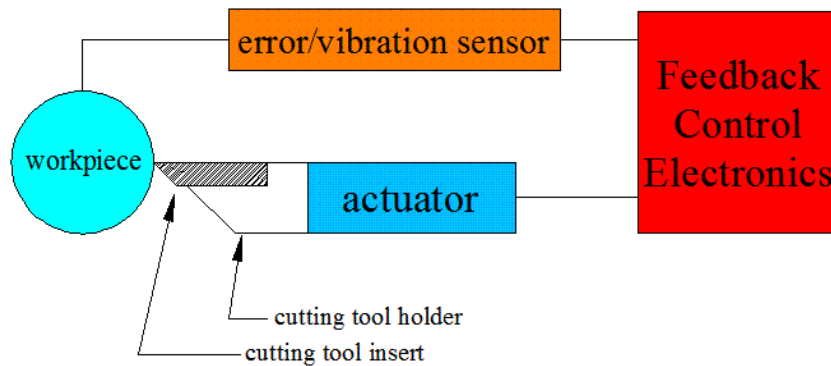
Figure 1.2   Schematic for cutting tool error or vibration compensation system.

Forces on the cutting tool can be significant on these types of machines, especially when working with hard or tough materials such as hardened steels or titanium. This mandates that any error compensating actuator placed between the cutting tip and tool post be of sufficient stiffness to withstand such forces [6].

Piezostack actuators, typically made of ceramic materials, and the closely related metal shell flextensional actuators possess the mechanical properties required for this application. However, just as in the case of AFM and STM, they cannot achieve requisite accuracy (micrometer scale precision in this case) without utilizing a control algorithm. The error or vibration to be compensated must be sensed and fed back into the control electronics in order to properly command the actuator.

1.1.3 Optical Wavefront Sensor (Nanometer Precision)

Images as viewed by Earth-based optical telescopes are the result of gathering light that is either emanating from or reflecting off of a body somewhere in space. Since space is a vacuum, there is virtually nothing to affect the quality or intensity of the light waves. Thus, light waves can traverse the many light-years to Earth nearly unchanged. In the last few hundred miles to Earth, however, the light suffers severe degeneration [4]. The light is bent, distorted, and phase-shifted by Earth's atmosphere, causing aberrations that lower the quality and intensity of the light collected for imaging. The ability to sense the optical wavefront is essential to correcting these optical aberrations. Many strategies have been devised for this purpose, one of which is known as a Sampled Twyman-Green Optical Phase-Shifting Digital Interferometer [4], which is a typical two-beam interferometer as illustrated in Figure 1.3.
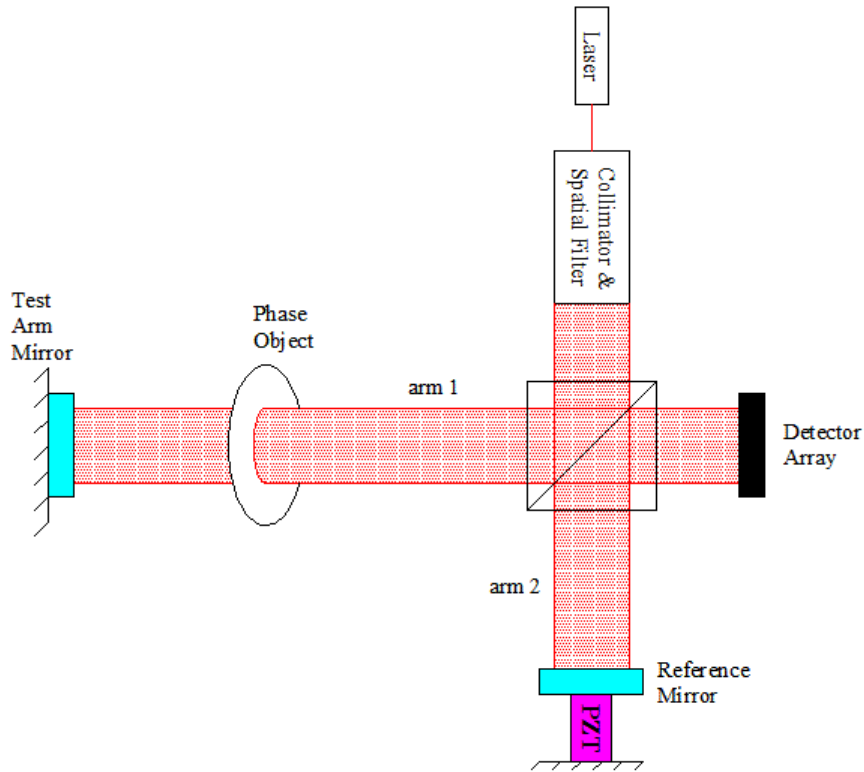
Figure 1.3   Twyman-Green Phase-Shifting interferometer [4].

A single laser beam of known wavelength is split into two beams after being collimated, expanded, and filtered. These are labeled arm 1 and arm 2 in Figure 1.3. Arm one is passed through a phase object with the purpose of determining the aberration it causes, and arm two serves as the reference beam. The beams are reflected back so that their interference may be sensed on the detector array. Translating the reference mirror back and forth via the actuator upon which it rests introduces the controlled phase. As a result of a known phase shift in the reference arm, the wavefront, including aberrations, of the test arm can be determined. Complete wavefront measurement requires scanning the phase for all values between zero and 2π. In order for the phase to assume these values, the mirror, and thus the

actuator, must be scanned from zero to *λ/2*, where λ is the wavelength of the laser light source.

The shape of the scan profile should be mathematically irrelevant so long as it is known. An intricate scan profile could, however, complicate data acquisition and analysis and have adverse effects on the reference mirror. Furthermore, there is little benefit to an intricate scan profile, thus a simple and continuous scan profile like that of Figure 1.4 (a) is typically administered [4]. Like the scan profile, the scan direction does not matter so long as it is known. This property allows repetition of the scan profile without large discontinuities. If the mirror follows the path of Figure 1.4 (b) many phase scans can occur successively where each forward scan, as well as each backward scan, represents a separate set of measurements. The actuator must track the profile of Figure 1.4 (b) as accurately as possible in order to make reliable wavefront measurements.
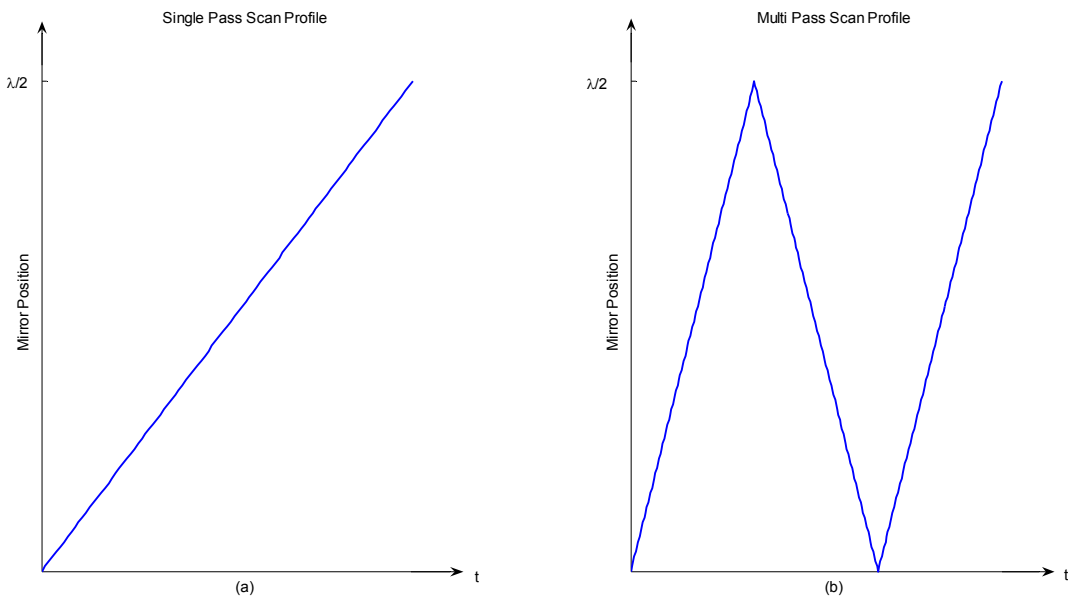
Figure 1.4   Mirror scan profile for actuator tracking.

An actuator is essentially an electromechanical system. The mechanical portion of the system, like any other, will have mass, stiffness and damping properties. If left in an open-loop configuration the tracking performance will likely not meet the specifications of the wavefront sensor. To track a signal such as that of Figure 1.4 (b) typically requires a closed loop control system.

1.1.4 Closing Remarks on Applications

The applications that have just been described have need for high precision positioning actuators. Piezoelectric flextensional actuators have the potential to achieve the precision needed for these applications. Their electromechanical performance must be explored and quantified in order to determine whether or not they are in fact suitable.

1.2 Background on Smart Materials, Modeling, and Control

The work in this thesis on system modeling and control of smart actuators spans three major fields of research. An unabridged review would contain a comprehensive survey of the three separate fields of smart materials and actuators, system modeling, and automatic controls as well as a survey of work done in effort to unite these three. It is the intention herein to present a concise overview of these three fields of knowledge and how they relate to each other, with the result being a clear picture of the state of the art in smart actuator control and its relevance here.

1.2.1 Smart Materials, Smart Structures, and Smart Actuators

The technology community has been exploring the application of smart materials and smart structures for some time. Reasons for this attention to smart materials are that they typically possess dimensional versatility, mechanical simplicity, and high energy density [7]. It is relevant at this time to define and differentiate what makes a material or a structure smart. Brief definitions are that smart materials have the potential to perform both sensing and actuating functions, while smart structures are those that involve smart material actuators and sensors as well as one or more microprocessors to analyze and command the sensors and actuators, respectively, so as to alter the system properties (e.g. stiffness, damping, etc.) or the system response in a controlled manner [8-9]. The research in this document is primarily concerned with smart materials and actuators and not with the more complex smart structures.

Materials that actualize an input/output protocol represented by any of the diagonals in Table 1.1 are considered typical. It is the materials that manifest an input/output connection portrayed by the off-diagonals that are considered smart materials, which are acknowledged to have both sensing and actuating capabilities. Should a material realize one of the off-diagonal relationships in Table 1.1, it could potentially be used as a sensor, an actuator, or both. However, availability does not necessarily imply utility. Some relationships prove to be more useful than others. Those effects commonly used as sensors and actuators are encircled in blue and red, respectively. Of the four effects that can be employed as smart actuators, the

converse piezoelectric effect is the most popular. A piezoelectric actuator is a common thread among the three applications presented in Section 1.1 [2-6]

Table 1.1 Various input/output phenomena for materials [10].

INPUT → MATERIAL/DEVICE → OUTPUT

OUTPUT / INPUT	CHARGE/ CURRENT	MAGNETI-ZATION	STRAIN	TEMPE-RATURE	LIGHT
ELEC. FIELD	Permittivity/ Conductivity	Elect.-mag. Effect	Converse piezo-effect	Elec.-caloric effect	Elec.-optic effect
MAG. FIELD	Mag.-elec. Effect	Permeability	Magneto-striction	Mag.-caloric effect	Mag.-optic effect
STRESS	Piezoelectric effect	Piezo-mag. Effect	Elastic Constant	----	Photoelastic effect
HEAT	Pyroelectric effect	----	Thermal Expansion	Specific Heat	----
LIGHT	Photovoltaic effect	----	Photo-striction	----	Refractive Index

Diagonal = Trivial Material
Off-Diagonal = Smart Material

Sensor | Actuator

The converse piezo effect is when a material responds to an applied electric field with a mechanical deformation. This is the result of electric dipoles within the material realigning in the presence of the applied electric field. A piezoelectric material will not, however, produce a useful deformation if it is not poled. The electric dipoles within the material are generally oriented randomly. Due to this randomness, the responses from individual dipoles will destructively interfere with each other resulting in zero net dimension change. The dipoles must be aligned, or poled, in order to achieve a net displacement.

Piezoelectric materials possess dipoles, and a Curie temperature (or Curie point) above which the dipoles diminish. Applying a large electric field when the material is above its Curie point and then cooling the material below the Curie point before removing the applied field allows the dipoles that formed to be aligned [11]. The material is said to have been poled after undergoing this process and will now exhibit the piezo and converse piezo effects.

1.2.2 System Modeling and Control

Dynamic system control begins with a dynamic model of the uncontrolled or open-loop system. A dynamic model relates the relevant system variables as inputs and outputs in time and typically comes in the form of differential or difference equations in continuous or discrete time respectively. They can be linear or non-linear, lumped or distributed, deterministic or stochastic, etc. The model designer has two generalized methods of model construction from which to choose, those methods being physical modeling or system identification modeling [12].

Physical modeling begins with structuring, which addresses issues such as deciding which variables are of interest for input/output purposes, determining if any variables can be regarded as constant, which quantities are needed to describe the system, and how the variables are affected by changes in other variables. If the system is complex, it is separated into subsystems in this phase as well. A useful way to look at the system is with a free body or block diagram. Applying the relevant laws of physics to this block or free body diagram yields the mathematical formulation of the system that the control engineer seeks. If the system is a high

order or multiple-degree-of-freedom (MDOF) system, this formulation may be organized into a canonical or state space form [12].

Physical modeling is rooted in known laws of physics derived from previous theoretical work and may or may not involve direct experimentation on a real system. System identification relies directly on current data from experimentation with the actual system or device to be modeled. Input and output data are gathered from an experiment and are then analyzed in order to ascertain a model [12-13]. How to acquire this data and which type of analysis to apply are the two general obstacles with which a system identification user is faced. The diagram in Figure 1.5 shows the iterative process of system identification.
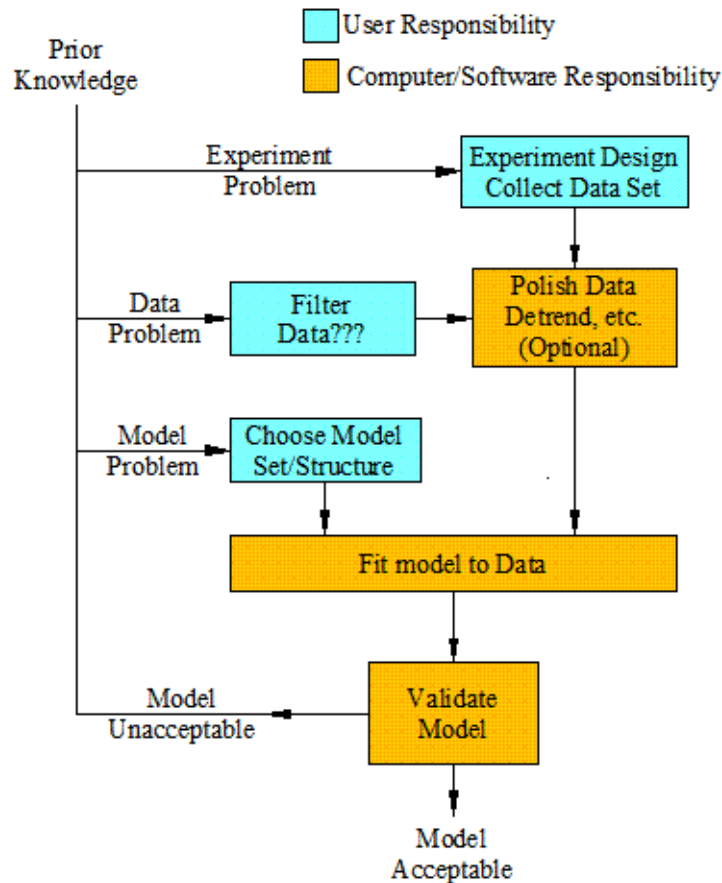
Figure 1.5   System identification loop [12-13].

To begin the system identification process, an experiment must be designed and conducted so that a data set may be obtained. The input often has the most substantial impact on the data set, and the experiment designer must dictate the system input shape. It is in fact, with regard to control design, paramount to be able to command the input so that the output may be commanded to track a desired trajectory.

An ideal input for identification purposes would evoke all of the system properties. Such an input would have to possess infinite frequency and amplitude content. A signal of this nature is not practical to generate. Even if it could be generated, sending it into a physical system or device could have adverse effects, conceivably to the point of destruction. The issue then becomes how to shape the input so the data set is informative enough with regard to a set of potential models and their intended use [13].

It is useful to provide some additional insight into the system and why it is being modeled. For example, consider a plant that is known to be underdamped, implying at least second-order properties, and is potentially either higher order or, perhaps, non-linear. In addition, the model may be used in simulations for predicting the system output. In this case, a data set generated from a step input would not be the best choice. An input signal with only two levels, such as a step, may not reveal any of the suspected higher-order properties or nonlinearities.

To this end, the following guidelines on experiment design, input properties, and some common inputs, all relevant to system identification should be observed when possible [12]:

- Conduct the experiment under conditions as close as possible to those that will be subjected to the model
- Ensure that the input evokes all relevant modes or properties, i.e. that the experiment is informative
- Select the input and output so the estimated model is as sensitive as possible to the estimated parameters

An informative experiment can be conducted by using an input that persistently excites a sufficient order. The input signal $u(t)$ persistently excites to the nth order if the power spectrum $\Phi_u(\omega)$ is non-zero on at least n points in the interval $-\pi < \omega \leq \pi$. The number of parameters to be estimated should not exceed the order of persistence of the input signal, i.e. if there are n parameters in the numerator and m parameters in the denominator, then the input must persistently excite to the order $(m+n)^{th}$ order. Some common input signals used for system identification are: filtered Gaussian white noise, random binary signal, pseudo-random binary signal, sum of sinusoids, swept sinusoids (chirp). Some of these signals involve only two amplitude levels, so it should be noted that if a nonlinear model is to be constructed, the input should be such that it contains more than two input levels [12-13].

According to Figure 1.5, in addition to obtaining a data set the user must select a model structure, that is to say, an appropriate manner in which to parametrize the model. This involves selecting a model type and model order. Selecting a model type addresses the general issue of forming either a tailor-made model or a ready-made model. Tailor-made models are essentially the amalgamation of physical modeling and system identification. After a modeling problem is structured and the relevant physics laws are applied, what is left is an equation(s), some of whose parameters maybe unknown. These unknown parameters may then be estimated using experimental data. Ready-made models, on the other hand, are of a black-box

sort. The parameters do not directly correlate to any of the physical variables in the system. They are simply a means of characterizing the dynamics contained in the given data set [12]. A number of black-box model structures exist, and as experimental data comes in a sampled form, these structures are expressed as difference equations (discrete time) as opposed to differential equations (continuous time) as mentioned previously.

An elementary representation of a linear system, including disturbances, is

$$y(t) = G(q)u(t) + H(q)e(t) \qquad (1.1)$$

where $y(t)$, $u(t)$, and $e(t)$ are the output, input, and white noise, respectively, and together comprise the experimental data set with $y(t)$ and $u(t)$ being selected in the experiment design and $e(t)$ being inherent to data acquisition. The terms $G(q)$ and $H(q)$ are sets of parameters that define the input/output and input/noise relationships, respectively. Even though they are not transfer functions in the continuous time, Laplace transform sense, $G(q)$ and $H(q)$ still map from one signal to another, so it is intuitive to express them as rational functions. This gives rise to a family of linear model structures given by

$$A(q)y(t) = \frac{B(q)}{F(q)}u(t) + \frac{C(q)}{D(q)}e(t) \qquad (1.2)$$

where the terms $A(q)$, $B(q)$, $C(q)$, $D(q)$ and $F(q)$ are the polynomial vectors whose elements are to be estimated for model fit. The combination of these five polynomials defines the particular model structure in the family [13]. There are several possible model structures born from Equation (1.2). Those most common to system identification are listed in Table 1.2. The autoregressive-extra input (ARX)

model structure is the simplest of the model structures and is often the first choice in model structures as it is helpful in determining input delay and model order if more complex models are required. The autoregressive-moving average-extra input (ARMAX) has become a common tool for controller design [12-13].

Table 1.2   Common model structures based on Equation (1.2) [13].

Polynomials Estimated	Model Structure
A, B	*AutoRegressive eXtra input (ARX)
A, B, C	*AutoReg. Moving Average eXtra input (ARMAX)
A, C	AutoReg. Moving Avg. (ARMA)
A, B, D	ARARX
A, B, C, D	ARARMAX
B, F	*Output-Error (OE)
B, F, C, D	*Box-Jenkins (BJ)

*Denotes inclusion in Matlab System Identification Toolbox™ ver. 7.0.

It is often the case that a system identification user is implementing existing algorithms for model structures and their associated parameter estimation routines rather than writing these algorithms autonomously. This implies that one is usually limited to what is included in the chosen software package. The system identification package available for this work is the System Identification Toolbox™ version 7.0 by Matlab®. Those model structures available are indicated in Table 1.2. This package also facilitates tailor-made models for single-input, single-output (SISO) systems in the form of linear, continuous-time transfer function models of up to an order of three (four if an integrator is added).

One could elaborate on the concept of system identification and provide much more detail and information than given above. However, as this work is not solely concerned with modeling via system identification, this level of explanation is

sufficient for the scope of this thesis. Only the facets required to make use of available system identification software were highlighted. For a more, in-depth explanation of the concepts upon which system identification modeling is built and their implementation, the reader is referred to [12] and [13].

After a model of the plant has been deemed sufficient, controller design can commence. The first step toward arriving at a suitable control system is to define the overall system specifications. Specifications can include, but are not limited to, requirements on dynamic performance metrics (rise time, settling time, overshoot, damping ratio, etc.), steady-state tracking error, sensitivity to plant parameter changes, and system stability [14].

There are two rudimentary forms of control, those being open-loop control and closed-loop, or feedback, control. A third form known as feedforward control exists, but is usually implemented in concert with feedback control. When implemented as a stand-alone control scheme it becomes a special case of open-loop control where the controller is the inverse plant model [15]. Block diagrams for both open- and closed-loop schemes are illustrated in Figure 1.6

The first question to be addressed is which of these control formats to implement. Answering this question requires understanding details of the plant, the overall system specifications, and if applicable, the real-world implementation of hardware. If stability is the issue, an open-loop control scheme will not introduce instability to an already stable system, but cannot stabilize an inherently unstable system. In this case a feedback scheme is usually the better choice. If minimal hardware implementation is the key and performance specifications are not stringent,

open-loop control would be a better choice as it does not require as many sensors. Furthermore, a feedback scheme will, in almost all cases, realize better performance in the face of model and/or disturbance uncertainty as well as better capability of meeting stringent system specifications when compared to open-loop control [14-15].
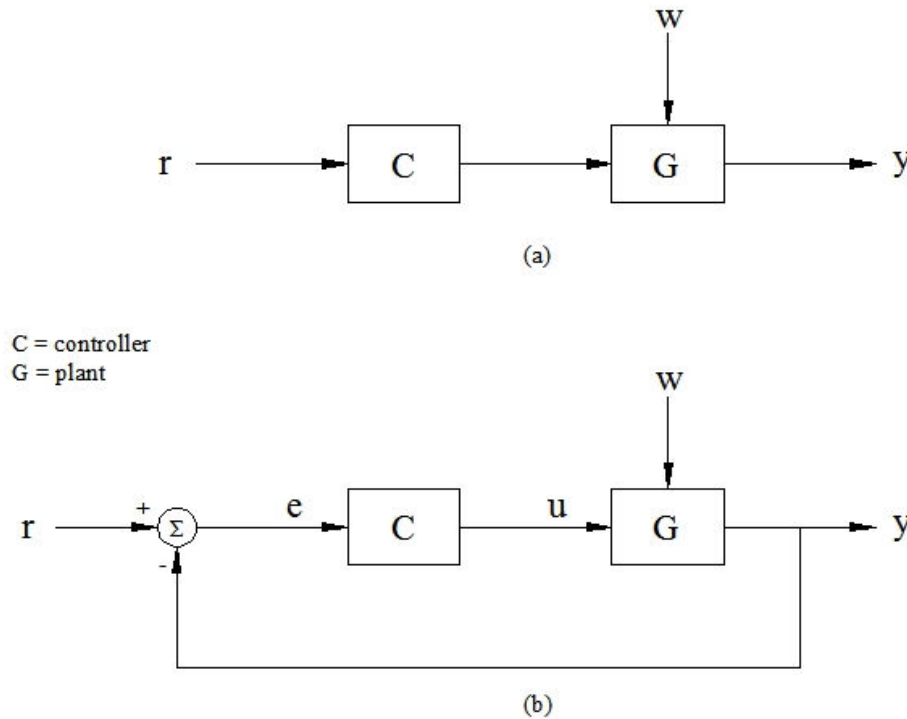
Figure 1.6   (a) Open-loop and (b) closed-loop control block diagrams with r, w, e, u, and y being the reference input, disturbance, error, controller output, and output signals respectively.

The discussion of feedback control schemes begins with the classical three-term proportional-integral-derivative (PID) controller. A PID controller implementation is a sum of the elements shown in Figure 1.7, and has a continuous-time transfer function given by Equation (1.3), in which the values K_P, K_I, and K_D are the proportional, integral, and derivative gains respectively.
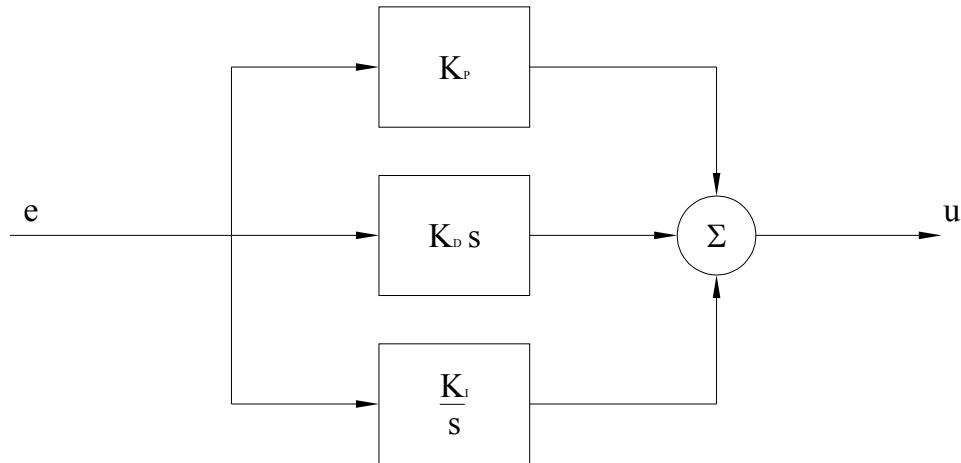
Figure 1.7   Block diagram of PID controller.

$$C(s) = K_p + \frac{K_I}{s} + K_D s \qquad (1.3)$$

Setting any of the gains equal to zero negates that element, e.g. setting K_D equal to zero results in a PI controller. Hypothetically, any combination of the three terms could be implemented via this process of zeroing gain(s) of unwanted component(s), yielding seven possible configurations. Not all seven combinations, however, are practical. The most often used are P, PI, PD, and PID control.

The three terms each have an impact on the system response. Some are favorable, and some are adverse. Increasing the proportional gain will decrease the steady-state error, but may increase the overshoot, and may lead to instability if increased excessively. Adding the integral term can eliminate the steady-state error, but may decrease stability. The derivative term can improve the dynamic response, i.e. alter damping, overshoot, etc., but does so at the cost of steady-state error performance [14, 16].

Another group of controllers whose names stem from their effect on the output phase are lead and lag compensators. Lead and lag compensators share a generalized transfer function given by Equation (1.4), in which z and p are the controller's zero and pole locations, respectively.

$$C(s) = K \frac{s + z}{s + p} \tag{1.4}$$

The output will lead if $z<p$ and lag if $z>p$. Lead compensators are similar to PD controllers in that they quicken the response time and decrease the overshoot without having much impact on the steady-state error. Lag compensators are akin to PI controllers because of their effect on the steady-state response. While they may not be able to drive the error to exactly zero, it will be reduced by a factor equal to z/p. Serially applying a compensator of each type results in a lag-lead compensator, and in the same vein as a PID controller, has the ability to affect both the transient and static responses [14,16]. As compared to Equation (1.4), which introduces a single pole and zero to the system, a lag-lead compensator introduces a pair of poles and zeros. The lag-lead transfer function is shown in Equation (1.5), in which z and p represent zero and pole locations respectively, just as before.

$$C(s) = K \frac{(s + z_{lag})(s + z_{lead})}{(s + p_{lag})(s + p_{lead})} \tag{1.5}$$

It has been assumed thus far that the controller will have been implemented as in Figure 1.6 (b), which is known as cascade compensation. Placing a controller in the feedback loop offers some benefits over cascade compensation, especially if the

20

controller has derivative action, or an approximation thereof, and the reference signal has sudden changes as in the cusp of a triangular wave or a step input. The derivative of the signal at points such as these will approach infinity, thus acting as an impulse to the system. PD controllers are often implemented as in Figure 1.8 when this is the case with the reference signal. This prevents the reference from being differentiated yet still yields the desired effect on the system response.
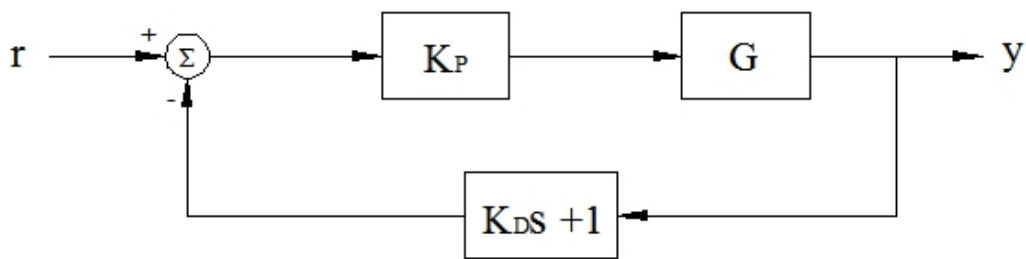
Figure 1.8   Alternative PD controller configuration [14].

The discussion now turns to optimal control methods. Like previously mentioned control techniques, optimal methods are linear and assume the plant to be linear as well. They differ in that the plant model is required to be in the state-space form given by Equation (1.6), in which x, u, and y are the state, control input, and output vectors, and A, B, C, and D are the state matrices. Many optimal control methods exist, but only the common and well-developed will be discussed.

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + B\mathbf{u}(t)$$
$$\mathbf{y}(t) = C\mathbf{x}(t) + D\mathbf{u}(t)$$

(1.6)

The most prominent and basic optimal method is the linear quadratic regulator (LQR). It is considered optimal because it seeks the linear feedback control law u that minimizes a quadratic cost measure J, both of which are given in Equation (1.7).

$$u = -Kx$$

$$J = \int_0^\infty (x^T Q x + u^T R u)\, dt \qquad (1.7)$$

The variables x and u are the state vector and control vector from the state space equations, and K is the controller gain matrix. Performance-measure matrices Q and R serve as design parameters when calculating the controller gain matrix, K, to provide a means of compromise between control performance and effort. Their selection can be tedious and usually requires iteration to achieve a controller that meets all performance goals.

Closely related to the LQR controller is linear quadratic Gaussian control (LQG). It assumes the dynamics of Equation (1.6) with the addition of both process and measurement noise corruption. As the name suggests, both are estimated as Gaussian white noise processes. Solving for the LQG controller amounts to determining an optimal state feedback gain matrix (essentially LQR control) and an optimal state estimator, i.e. a Kalman filter [15, 17]. Performance matrices Q and R must still be chosen. Guidelines for starting points exist, e.g. Bryson's rule of starting with values equal to the inverse of maximum acceptable state and control input values. They are rules of thumb at best, so the selection can still be difficult. Experience plays an important role in both.

There are many more control methods and configurations than those mentioned here. The aforementioned represent tried and true, well-developed methods that have found widespread application.

1.2.3 Modeling and Control of Smart Actuators

Modeling and control as it applies to smart material actuators is relatively young field. Yet, due to the diversity of the three separate fields of modeling, control, and smart materials, a tremendous amount of technical literature is available. The problem of how to model and then control a device is addressed for a variety of smart materials and devices, and is done so in a variety of different manners [7-9].

It is generally accepted that all systems contain some level and type of non-linearity, and smart materials, piezoelectrics in particular, are no exception. This gives rise to the question of whether to use an approximated linear model of the system, or to attempt modeling the non-linearities. Many testify to the necessity of modeling non-linearities, while others have been successful with linear models.

The IEEE published a Standard on Piezoelectricity [18], wherein linear constitutive relations are defined that lay a foundation upon which dynamic models of piezo actuators can be, and in fact were, built. Hagood et al. and Liang et al., among others, did just that [19-20]. The former cast the constitutive relations into a dynamic state space model, while the latter used Laplace and inverse Laplace transforms to develop an expression for the transient response of a piezo actuator. Main et al. extended the relations to a multilayer piezostack actuator [21]. Two continuous-time transfer functions were developed sharing the trait of having mechanical displacement and applied voltage as the input and output respectively. They differ in that one model has the equations configured to have applied voltage directly control the displacement, whereas the other has charge as the intermediary controlling input between voltage and displacement.

23

Piezoelectric actuators are often implemented into other devices whose dynamics may need to be modeled as well. In [22], Gorman et al. show results of modeling and controlling a piezoelectric actuator driving a beam-steering flexure mechanism based on an analytic, closed-form model. The flexure mechanism is modeled simply as a lever arm. Without regard to the constitutive relations in [18], the piezo is modeled as a force actuator with a linear relationship between input voltage and output force defined by a transformer ratio. Two PD controllers were designed based on this model, one including estimation for base excitation. The tracking error for these controllers to a 10 Hz, 5 μm sine wave is shown in Figure 1.9. At first glance the tracking errors appear large, but note that the y-axis scale for both is in nanometers. The claimed tracking errors, after the transients have died out, are ±65 nm and ±10 nm for the two controllers, respectively [22]. This error magnitude may still be considered large depending on the magnitude of the reference signal. Also, note that no experimental validation was performed for these simulations.

(a) PD   (b) PD w/base ex. estimates
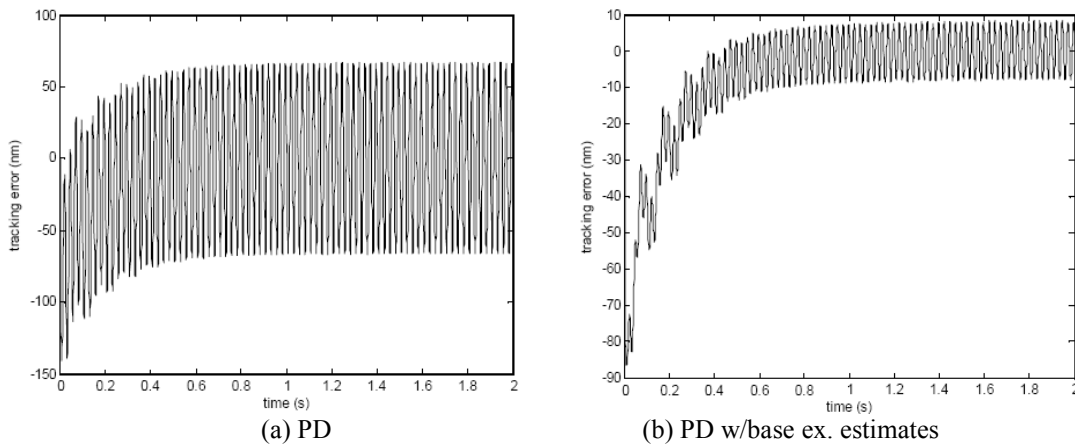
Figure 1.9   Tracking error results from [22].

Controllers have been designed based on linear, black-box system identification models as well. Salapaka et al. fit such a model to a one-dimensional

nanopositioning stage driven by piezoelectric stack actuators and subsequently designed controllers for it [23]. Their model, given by Equation (1.8), was fit to experimentally taken frequency data via a sine sweep input with 50 mV amplitude and a 2 kHz final frequency.

$$G(s) = \frac{9.7 \times 10^4 (s - (7.2 \pm 7.4i) \times 10^3)}{(s + (1.9 \pm 4.5i) \times 10^3)(s + (1.2 \pm 15.2i) \times 10^2)} \quad (1.8)$$

Based on this model, PI and H_∞ (a more advanced type of optimal control algorithm not discussed)controllers were designed. The PI controller exhibited excellent reference tracking capabilities on the scale in which it was experimentally tested, but based on simulation suffered in bandwidth. The H_∞ controller maintains tracking ability and makes significant improvements in bandwidth when compared to the PI controller. Schitter et al. adapted a similar approach to the modeling and control of a piezo actuator for an AFM. The dynamics were modeled by applying a band-limited white noise signal to the piezo actuator and fitting a model to the resulting input-output data [24].

Even those that advocate non-linear modeling usually decouple the non-linearities from the actuator dynamics, i.e. a non-linear estimator or model is created. If the goal is reference tracking, the non-linear model inverse is added to feedback control input in a feedforward layout as shown in Figure 1.10 [25-27], and if the goal is simply to compensate or cancel the effects of the non-linearity, the inverse model is added in an open-loop, cascaded configuration as shown in Figure 1.11 [28-30].
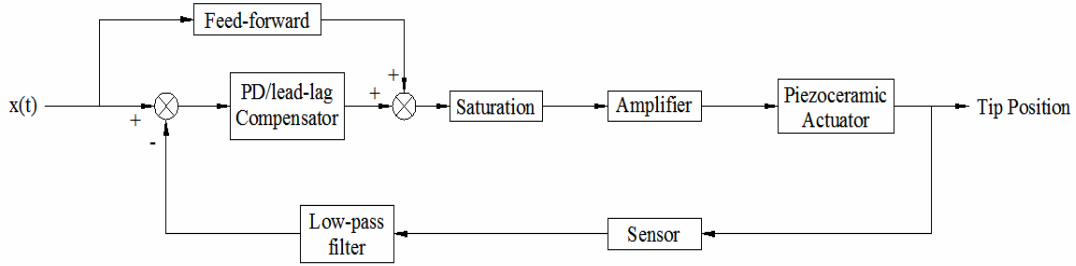
Figure 1.10   Closed-loop controller with feedforward hysteresis compensation [26].
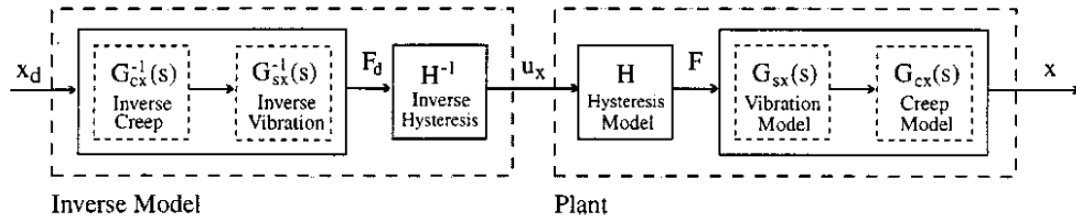
Figure 1.11   Open-loop inverse compensation approach [28].

The non-linearity addressed most often in technical literature is hysteresis. There are several non-linear models/estimators that are used to approximate the hysteresis present in piezoelectric materials. The most common is the one used in [28], the Preisach model. It is based on an elemental hysteron unit, otherwise known as a simple relay operator. Goldfarb et al. propose a model based on generalized Maxwell slip [30], which is the force-displacement relationship of an infinite number of elasto-slide elements placed in parallel. An elasto-slide element is a linear spring connected to a massless block that is subjected to Coulomb friction. For simulation purposes, this basically amounts to using a finite number of elasto-slide elements, the number of which depends on the desired accuracy. The parameters of each element must be determined so that the comprehensive result fits the shape of the hysteresis that is to be predicted. Bashash et al. used a subclass of the Preisach model known as the Prandtl-Ishlinskii operator [31].   The Prandtl-Ishlinskii hysteresis model is

26

superposition of several backlash operators, which are similar to the relay operators used in the Preisach model. These three are just a sample of the non-linear models/shape estimators. They represent the most prominent of those found in the literature.

Even among proponents of non-linear modeling there are some that concede the point that linear models have their merits and that they may be sufficient for prediction, simulation, controller design, etc., depending on operating circumstances. Croft et al. state that non-linear effects can reduce positioning precision when a piezoactuator is actuated over a large displacement range, but that they also have a linear range and can thus be operated so as to avoid these issues [28]. Hughes and Wen agree with the Croft et al that with piezoceramic hysteresis, the larger the input signal, the larger effect [32]. This could imply that at small driving input levels, these effects become much less prominent and could thus be combated by means other than modeling them. In fact, Smith et al. admit that conventional feedback laws successfully mitigate the non-linear effects of piezoceramics in certain operating regimes, which is the reason for the success of AFM technology [33].

1.2.4 Current Challenges and Issues

The purpose of the previous discussion was not to denounce non-linear modeling of smart material actuators or non-linear modeling in general. In fact, the author concedes that non-linear modeling, for systems of any sort, can be a worthwhile pursuit, especially when the main goal is to better grasp the underlying physics of the given system.

The purpose of the previous section was to provide an overview of smart actuator modeling and control and to illustrate the large variance in the approaches taken. This variance highlights the current issues in the smart actuator community. The question of what is the best way to model and control these devices still has no definitive answer, as it depends on the specific parameters and constraints of a particular application.

1.3 Research Directive

Using the modeling techniques previously discussed, a dynamic model of a piezoelectric flextensional actuator was constructed. The model order and structure were guided by experimental data, which was taken such that it revealed low and high-frequency behavior, as well as any modes that needed to be captured by the model. After designing a suitable dynamic model, a controller was synthesized to accurately command the actuator displacement. The real-time controller performance was then assessed by applying the controller the actual system hardware.

The research contained here characterizes only the actuator with which it is concerned. It results in a feedback control algorithm that can be used to control the actuator's displacement. In addition, this work outlines a general process by which other smart material actuators can be characterized, modeled, and controlled.

2 Experimental Setup

Experiments in a laboratory setting, whether they are the sole purpose of the research or just a fraction of the complete work, present a formidable set of problems and difficulties. The potential applications mentioned in the previous chapter require displacements on the order of micrometers to nanometers. Measurements of this scale can encroach upon the capabilities of laboratory equipment. They are also subject to being overcome by vibration noise, which, in most cases, is present to at least some degree. It is then pertinent to make an assessment of the equipment in the laboratory in which it will be used. Doing so yields the limitations on what can be expected of the equipment and the experiment setup with regard to resolution, operating range, data quality, etc.

2.1 The Plant: Piezoceramic-Metal Flextensional Actuator

Actuator research pursues a design that can generate a large force and mechanical stroke over a broad frequency range with minimal power consumption. Piezoelectric actuators are no exception. Typically they suffer, like most actuators, in that they rarely achieve excellent performance in all categories simultaneously. While the bandwidth is usually good, piezoelectric actuators compromise between generative force and displacement, i.e. an actuator with much force usually has little displacement and vice versa.

Two common piezoelectric actuators are the piezostack actuator and the bimorph actuator. These two actuator types reside at opposite ends of this force vs. displacement spectrum, i.e. the piezostack can generate much force but little displacement and the bimorph can span a large displacement but generates little force. A type of flextensional actuator known as ceramic-metal composite was developed that sought to bridge the gap between piezostacks and bimorphs. Well-known devices of this type are the moonie and its offspring the cymbal [34-36]. These consist of a piezoelectric actuator of some sort integrated into a metal shell. As the piezoactuator expands or contracts to flex the metal shell, displacement amplification is had without a significant deterioration in load carrying ability [34-36]. The device used in this work is of this type and is shown in Figure 2.1. The red and black wires are the positive and negative electrical input leads. A voltage applied across these leads yields a mechanical displacement. The device has a footprint of approximately three square centimeters. Overall dimensions are provided in Table 2.1.
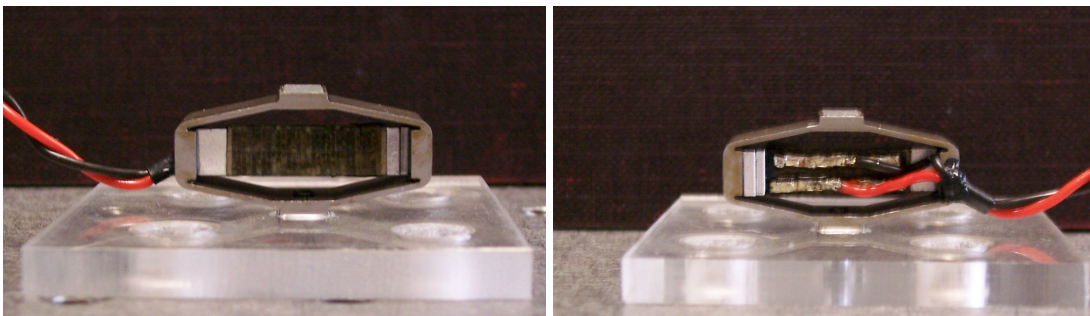
Figure 2.1 **Photographs from either side of the actuator used in all experiments. Active element is a piezostack actuator. The surrounding metal shell, serving as a displacement amplification mechanism, is a steel alloy.**

Table 2.1 Geometric dimensions of the flextensional actuator [36].

	Length (mm)	Width (mm)	Height (mm)	Number of Active Layers	Active Layer Thickness (mm)
Piezostack	15	5	5	30	0.5
Metal Shell	27	10	13.5	N/A	N/A

The usable mechanical displacement direction is perpendicular to the piezoelectric displacement direction. This is shown schematically in Figure 2.2. Applying a positive voltage across the electrical input leads causes an extension of the piezostack. This extension flexes the metal shell, forcing the top and bottom surfaces closer together.

Working
Surface

Piezoelectric
Displacement
Direction

Working
Displacement
Direction
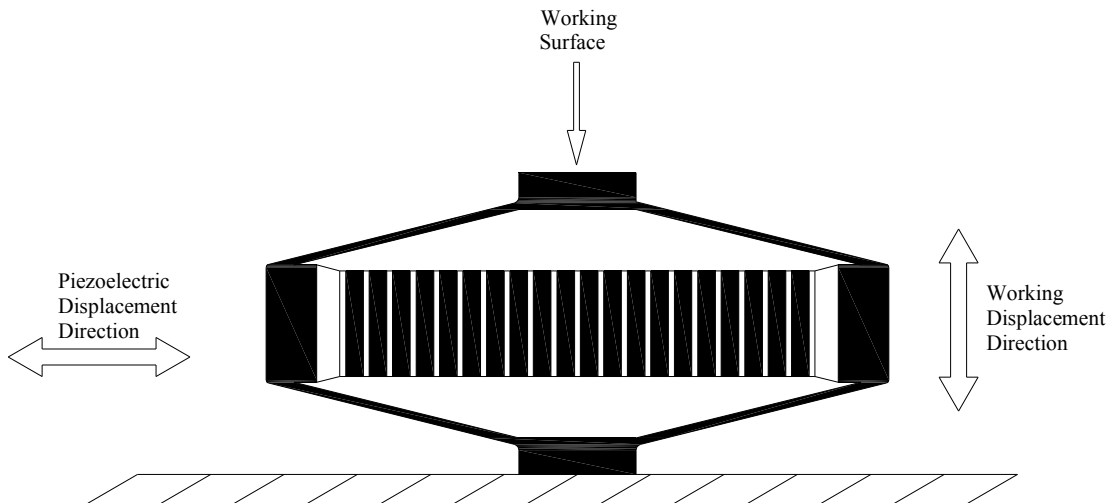
Figure 2.2 Schematic of actuator showing working displacement direction relative to the active piezoelectric element. Working direction is perpendicular to piezo.

Manufactured by TRS Technologies, Inc, the active material in the piezostack is the single-crystal piezoceramic compound lead magnesium niobate-lead titanate (PMN-PT) and the metal shell is a steel alloy. The recommended operating range is 0-500 Volts [36]. This is a safe range, but it can be exceeded per the maximum electric field guidelines in [37] if needed. Given the thickness of each PMN-PT layer

31

in the piezostack from Table 2.1 and the maximum applied electric field range in [37], the absolute maximum applied voltage range is -100 to 1250 V. The range used in this work was -50 to 500 V. Due to the capacitive nature of piezoelectric materials, a negative voltage may be needed to drive the actuator to a lower displacement level in a sufficiently short amount of time.

Single crystal piezoceramics, including the PMN-PT compound used in the actuator here, show superior strain capabilities as compared to typical piezoceramics such as lead zirconate titanate (PZT) [38-41]. Much can be said of single crystal piezoceramics, so the reader will refer to Appendix C for a more thorough discussion of the materials, both in general and as it applies to the specific device being tested.

2.2 Data Acquisition and Driving Waveform Source

Foremost among the equipment used for data acquisition and waveform generation with which to drive the actuator is a commercial system made by dSPACE and its associated hardware and software. The dSPACE system provides the communication link between the analog nature of most experiments and the digital interpretation of a computer. Control Desk Developer Version 3.1.1 is included for real-time interface with Simulink$^®$. Thus, any waveforms that need be generated are done so digitally using Matlab$^®$ and/or Simulink$^®$ and converted to analog signals as necessary.

The digital-to-analog converter (DAC) and the analog-to-digital converter (ADC) are external to the dSPACE system itself. These come in the form of the

DS2102 DAC board and the DS2003 ADC board and are integrated into one cohesive unit. The DAC board offers 16-bit resolution and output ranges of ±10 V, ±5 V, and 0-10 V. Bit number is adjustable with the ADC board and can be set at various values between 4- and 16-bit. Operating ranges are ±10 V and ±5 V. Resolution is variable with both boards and will depend on the range (and bit number as well, in the case of the ADC board) selected. To abate the effects of bit noise, or quantization error, as much as possible, the ADC board resolution was set at the maximum of 16-bit. This results in a voltage resolution of 0.15mV for the ±5 V range and of 0.30mV for the ±10 V range.

2.3 Trek High-Voltage Amplifier

The maximum output level of ±10 V from the DAC board is well below allowable voltage range for the actuator. Using the DAC board alone will limit the maximum obtainable position change, as well as the rate at which that position change can be reached. A high-voltage amplifier was needed to obtain sufficient closed-loop control authority. The Trek, Inc. 609E-6 high-voltage amplifier shown in Figure 2.3 was used for this purpose
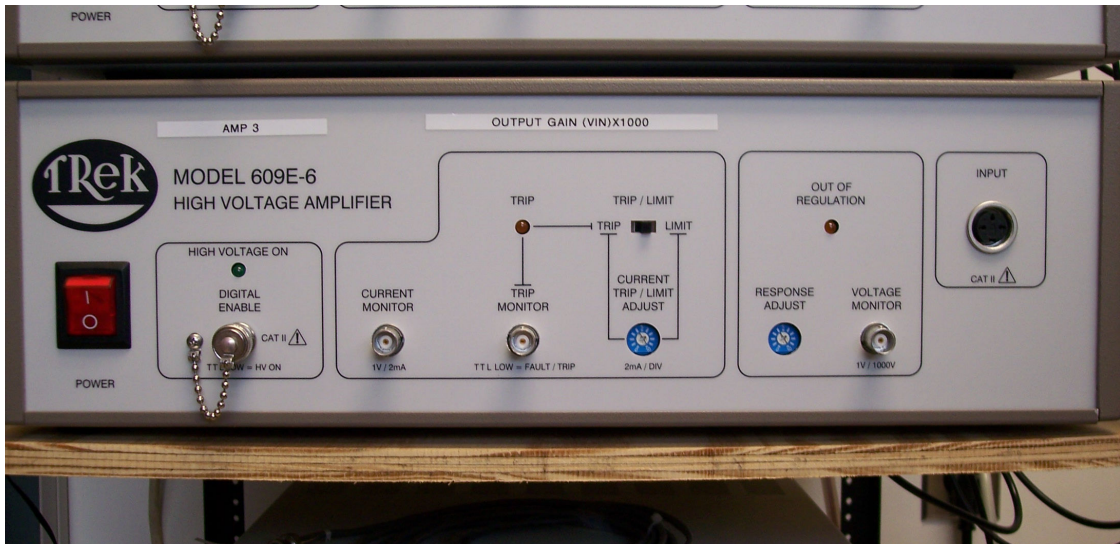
Figure 2.3   Trek 609E-6 high-voltage amplifier.

The 609E-6 accepts a maximum input level of ±4 V. An amplification of 1000V/V yields a maximum output range of ±4000 V. Furthermore, the amplifier is current limited to 20 milliamps. It has a small signal bandwidth of 35 kHz and a slew rate of 150 V/μs [42]. This voltage level is far beyond that which is allowable for the actuator. This amplifier enables usage of the actuator's full potential.

2.4 Laser Vibrometer Measurement System

A laser Doppler vibrometer system made by Polytec is used for making non-contact displacement measurements of the actuator. The complete vibrometer system is comprised of the controller OFV-3001 and the sensor head OFV-512. Interference of the measurement and reference beams occurs in the sensor head, and results in a fringe pattern from which frequency and phase difference signals are gleaned. These signals are sent to the controller where they are decoded. Velocity is proportional to

the frequency difference, and displacement is proportional to the phase difference. This is shown schematically in Figure 2.4, where f_0 and φ_0 are the frequency of reference beam and f_{mod} and φ_{mod} are the frequency and phase of the object beam.
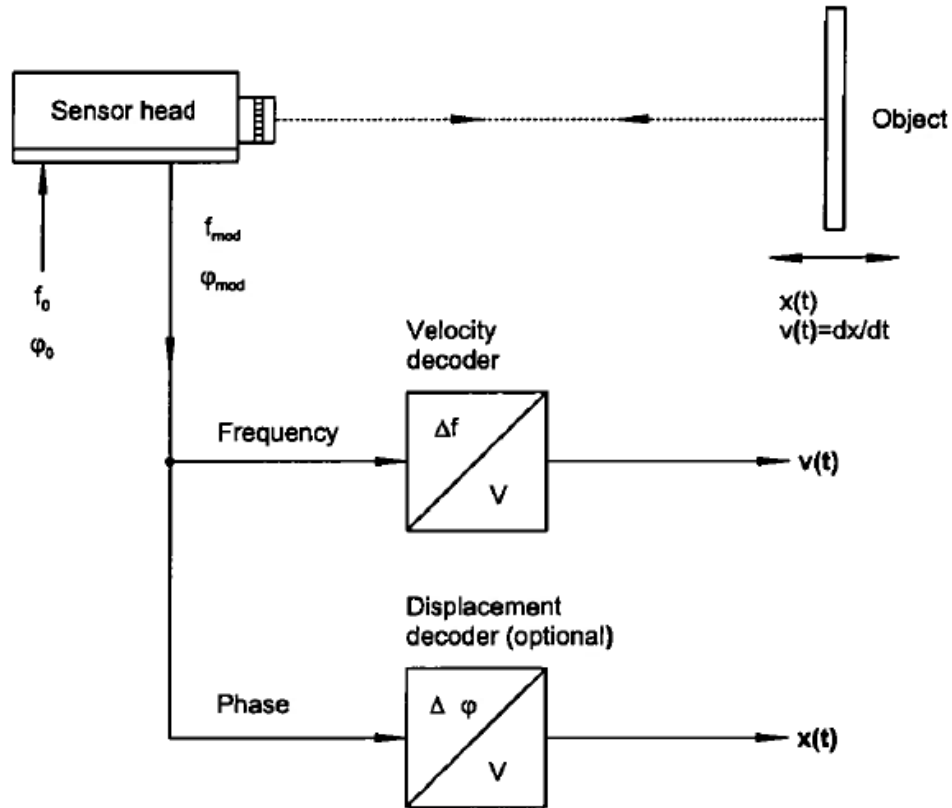
Figure 2.4   Signal progression of laser vibrometer system [43].

Output signals for displacement and velocity are available as analog voltage signals via two BNC jacks at the front of the controller. Proportionality between the actual displacement and velocity and these voltage signals is direct. The user can select scaling factors for velocity and displacement separately at the controller interface. Scaling factors available depend on the decoders that are installed. In this case, the velocity decoder is the OVD-02 and the displacement decoder is OVD-20. See Appendix D for the specifications of these decoders from [43].

2.5 Vibration Isolation

The world is in constant motion and, as such, causes vibrations that threaten the ability to make minute movement detections. Experiments seeking to detect such movements must be isolated from the world to the best of the experiment designer's ability. For this reason experiment hardware resides directly on a Newport RS-4000 optical table, which in turn is supported by four Newport I-2000 pneumatic isolations legs. The representative photo in Figure 2.5 shows the assembly of these two components.

Figure 2.5   RS-4000 optical table and I-2000 vibration isolation legs [44].

While the table and isolator legs function in concert, the two serve individual purposes. The table provides bedrock upon which the experiment can be fixed, while the legs supporting it serve to attenuate external, mechanical vibrations that may emanate from the building, the machinery and equipment in it, or anywhere else for that matter. Ideally, the table should be infinitely stiff so that, despite any disturbance

force to which it is subject, no deviation between any two points on the table would ever occur, and the legs would attenuate vibrations of any amplitude and frequency.

These notions are, as stated, ideal, not realistic. The table is simply two steel sheets separated by a steel honeycomb, and the legs are essentially pneumatic columns with matched valves and chambers such that the attachment points yield an auto-leveling surface. Together, they amount to a well designed beam resting on specifically tuned dampers. Just like other mechanical systems, they will have resonant frequencies and vibration modes. The manufacturer addresses these issues for each component individually. Details come in the form of transmissibility and compliance curves for the isolation legs and optical tables respectively.

Vertical and horizontal transmissibility curves for the I-2000 legs are shown in Figure 2.6, and associated specifications from [45] can be seen in Appendix E. These curves communicate how well the legs abate vibrations in both vertical and horizontal directions. Transmissibility is equal to unity at 0 Hz (not shown because curves are clipped at 0.8 and 0.5 Hz). It peaks at the resonant frequency and then begins to fall. The y-axis values can be pragmatically interpreted as percentages, e.g. the amplitude of vibration noise at 10 Hz will be attenuated by a factor of approximately 0.01. This implies that any vibration noise occurring at frequencies where transmissibility is greater than unity is not only being transmitted to the table, it is being amplified [44, 46-47]. In this frequency range the table behaves, for all practical purposes, as a rigid body, thus vibrations in this range get transmitted to all components on the table top after being amplified by the legs.
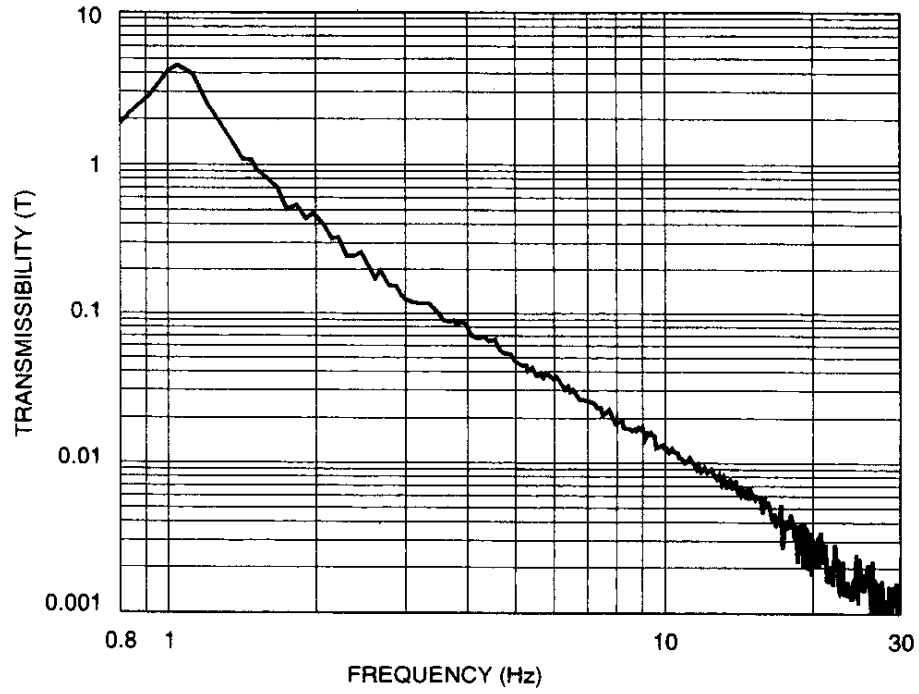
Vertical Transmissibility (at maximum recommended load)
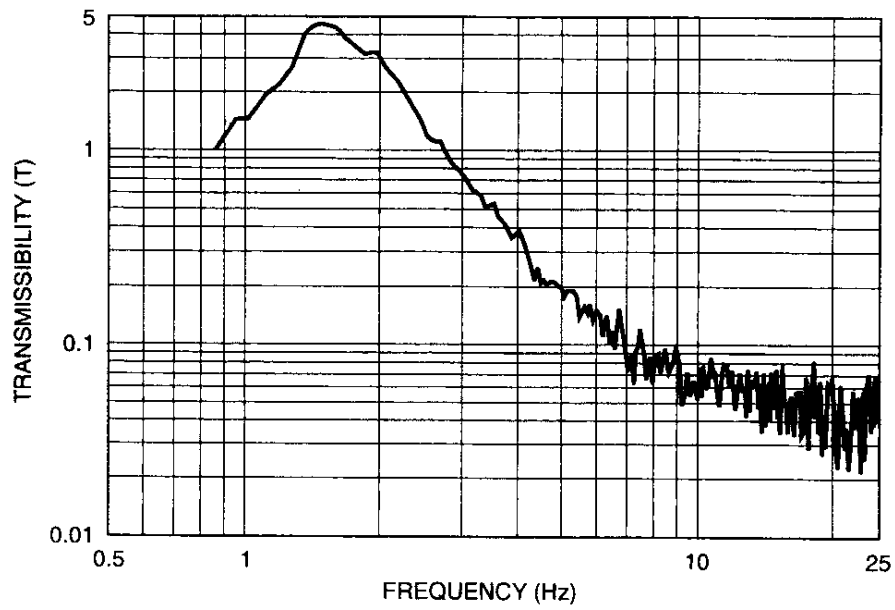
Horizontal Transmissibility

Figure 2.6 Transmissibility curves for I-2000 legs [45].

38

Per the information in [44], the transmissibility curves give no indication of the amplitudes of floor vibration used to generate said curves. Vibrations must overcome friction before the isolation system will function properly. Directly from [44]: In the nanometer region, isolator friction effects become a real problem.

Compliance curves tell a similar story about the table in that they reveal the notable frequencies or frequency ranges. Unlike transmissibility curves, however, compliance curves do give some indication of relative amplitude, but they are specific to the points from which they were measured. So, as the table size increases, so does the number of compliance curves that may be generated. A sample of five compliance curves for the table used in this experiment was supplied courtesy of the manufacturer, each one measured from a different location on the table top [48]. They were used more to give the author a qualitative idea of what to expect than for quantitative calculations.

2.6 Experiment Layout

2.6.1 Geometric Layout of Hardware

Photographs of the experiment details are shown in Figure 2.7 through Figure 2.10, and a schematic of the overall experiment is shown in Figure 2.11. Because it is susceptible to vibration noise, the laser vibrometer is placed on the optics table along with the lens, fiber-optic cables, and the actuator. The actuator is mounted to a three-axis positioning stage with micrometer adjustments for each axis. Figure 2.8

highlights the separation distance between the actuator and laser lens. This dimension is important, as it affects the quality and quantity of laser light being reflected back into the vibrometer. The Y-direction micrometer is used to adjust the distance from the laser lens to the actuator's working surface where the measurements are taken.
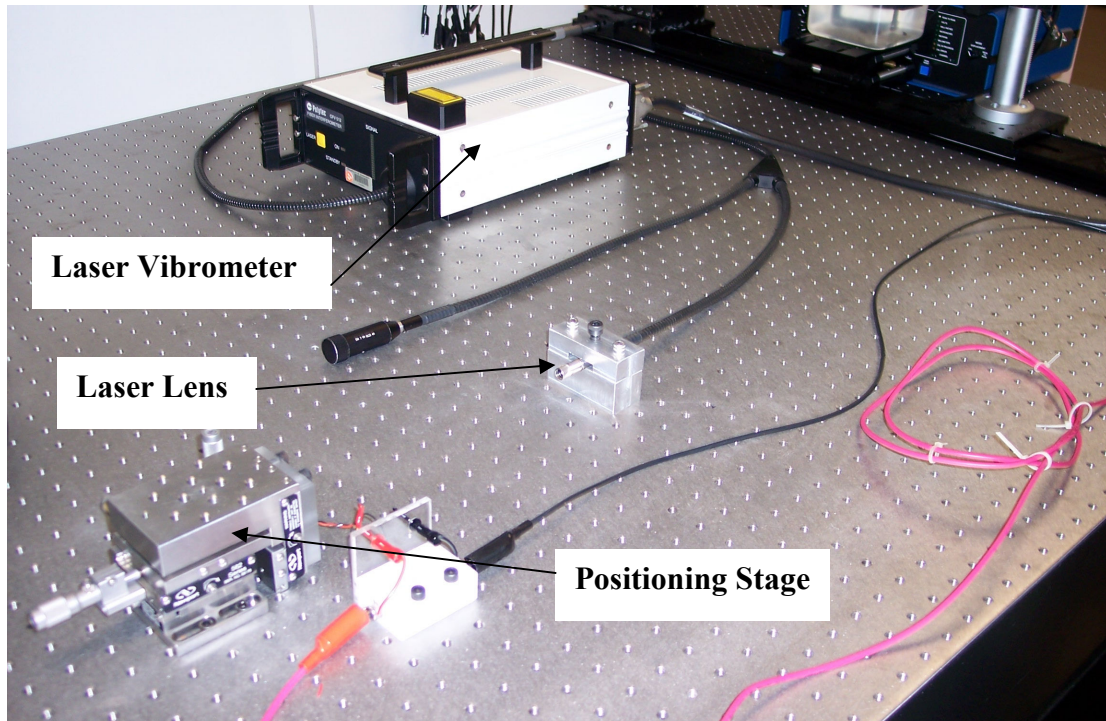
Figure 2.7   Overall experiment setup on optical table.
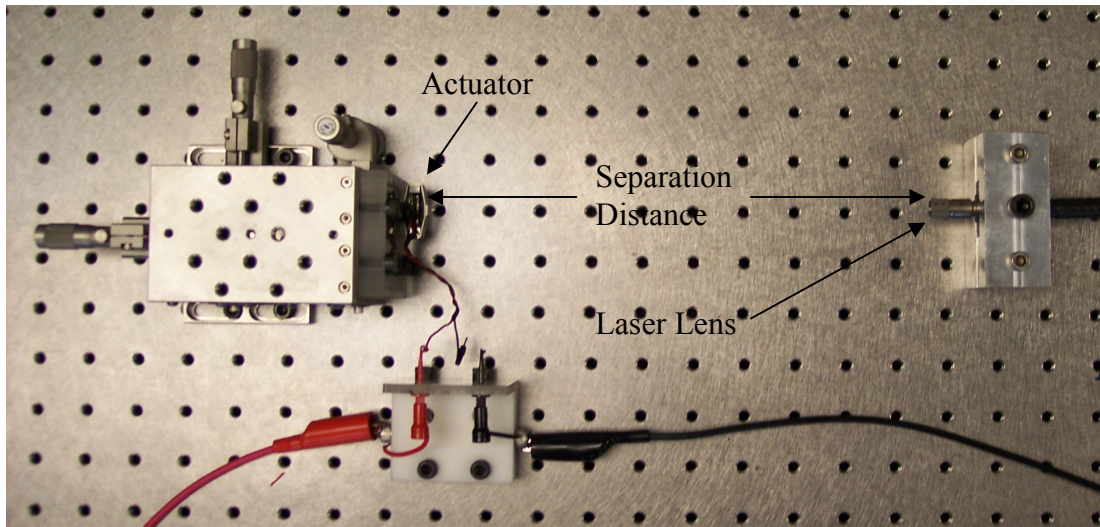
Figure 2.8   Closer view of experiment layout to show relative positions of laser lens and actuator. Distance from lens to object is important. Positioning stage Y-direction micrometer used to alter this dimension to obtain maximum laser signal strength.

Also illustrated by Figure 2.8 is the orientation of the actuator with respect to the laser lens. Recall the actuator schematic in Figure 2.2. Furthermore, consider that movement away from and towards the laser lens is defined as a negative and positive displacement, respectively [43]. The conclusion that can be drawn from this is that a positive input will result in a negative output. This fact will be important in later chapters regarding modeling and control design.

Figure 2.9 shows a closer view of the actuator and positioning stage. Notice the color difference between the metal shell and the working surface. The silver color of the working surface is reflective tape. The laser vibrometer relies on light being reflected back from the measurement point to function. More of the laser light being reflected results in a stronger, cleaner signal from the vibrometer. Reflectivity is the only purpose being served by the reflective tape, since a good signal-to-noise ratio could not be obtained without it.
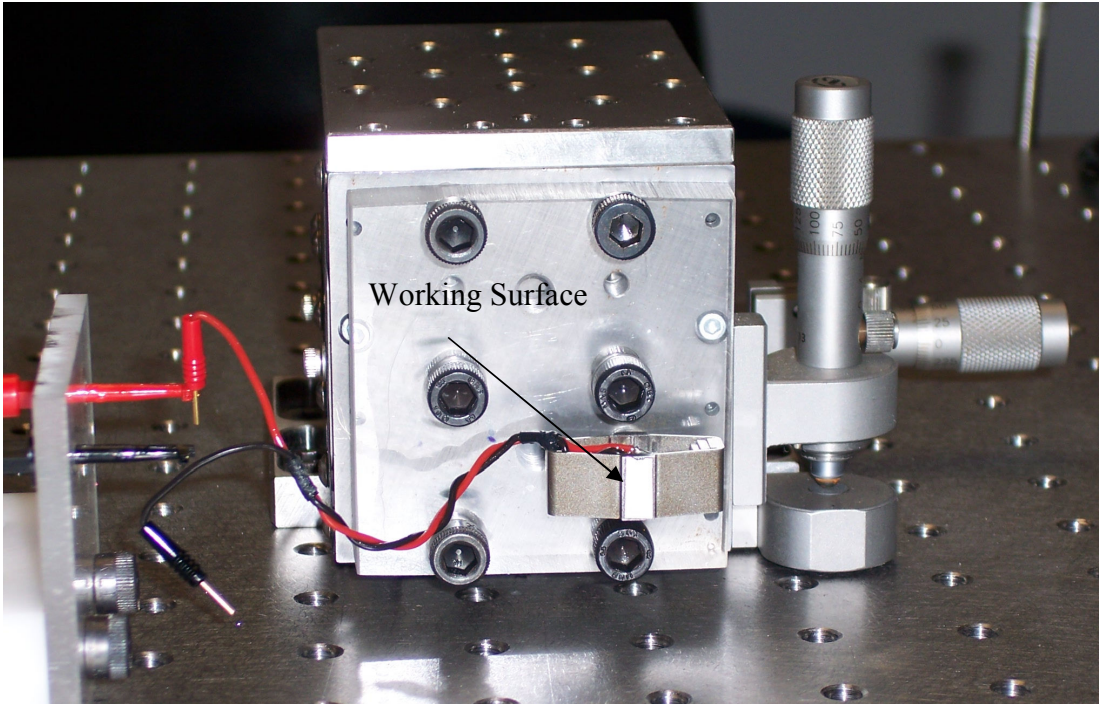
Figure 2.9   Actuator mounted XYZ positioning stage. X- and Z-direction micrometers used to position laser on working surface of actuator.
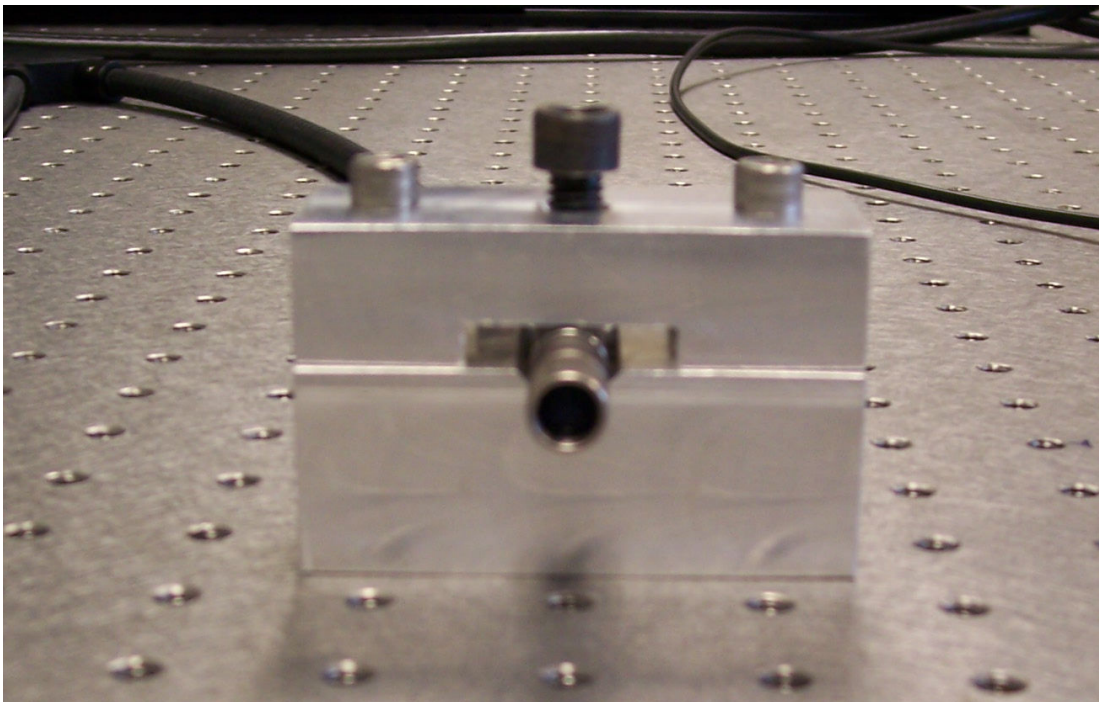
Figure 2.10   Laser lens clamped into a V-block.
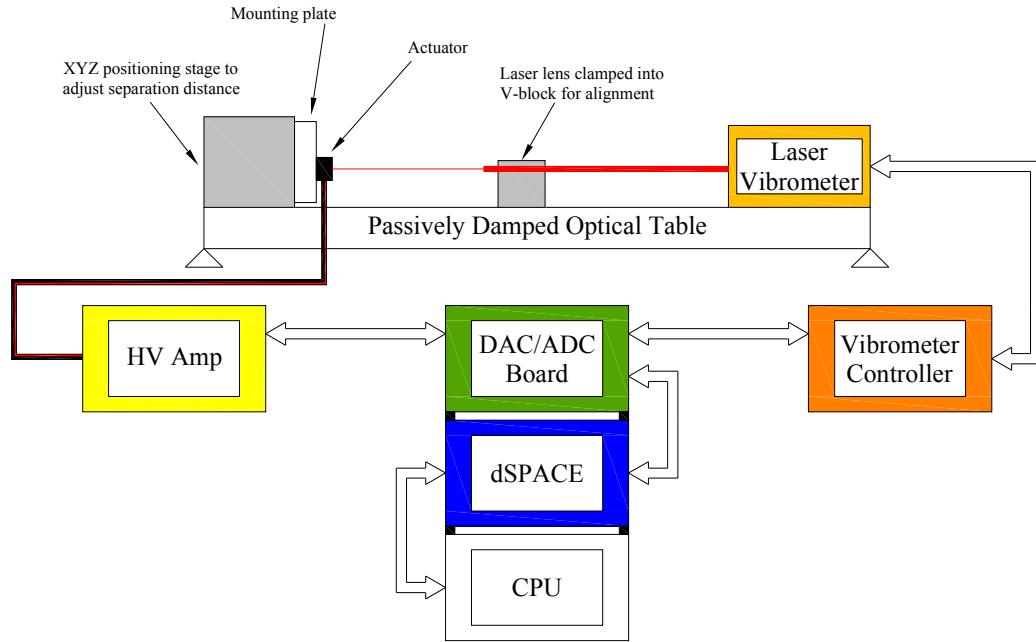
Figure 2.11   Schematic of overall experiment layout.

2.6.2   Software: Simulink Model Layouts

The experiment was controlled using a digital interface created using dSPACE Control Desk, which works with Simulink by making block diagram parameters real-time adjustable.   The block diagrams used for open-loop system identification experiments and closed-loop control experiments are shown in Figure 2.12 and Figure 2.13 respectively.

Gains to cancel gain
of 5 built into DAC board
when set to +/-5 Volts

Chirp Signal → Chirp Gain (3) → Saturation → DACcancel (1/5) → DS2102_B1 [DAC #1, DAC #2, DAC #3, DAC #4, DAC #5, DAC #6]

Ground

Ground 1

MUX_ADC / DS2003_B1 → ADCcancel (5) → DisplacementScale (8) → Displ_um (1)

→ HVAmp_out (2)

Gain of 5 to cancel the 1/5
gain built into the ADC board
when set to +/- 5 Volts
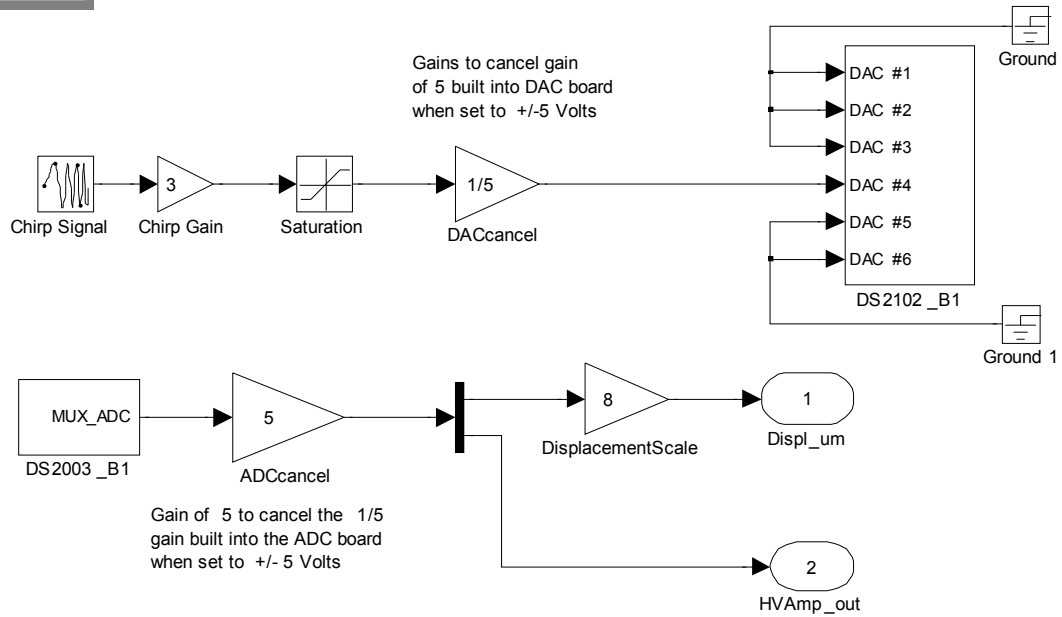
Figure 2.12   Simulink block diagram for open-loop system identification experiments.

Specify displacements in micrometers

Error (1)

input voltage (3)

Ground

Reference → (+_) → PI Controller (PID) → (+_) → Saturation → DAC/HVamp cancel (1/5000) → DS2102_B1 [DAC #1, DAC #2, DAC #3, DAC #4, DAC #5, DAC #6]

1/500 gain to cancel gain
of 5 built into DAC board and
the gain of 100 in HV Amp

Ground 1

D Controller (PID)

Displ (2) ← LaserScale (-2) ← ADC cancel (5) ← MUX_ADC / DS2003_B1

Gain of 5 to cancel the 1/5
gain built into the ADC board
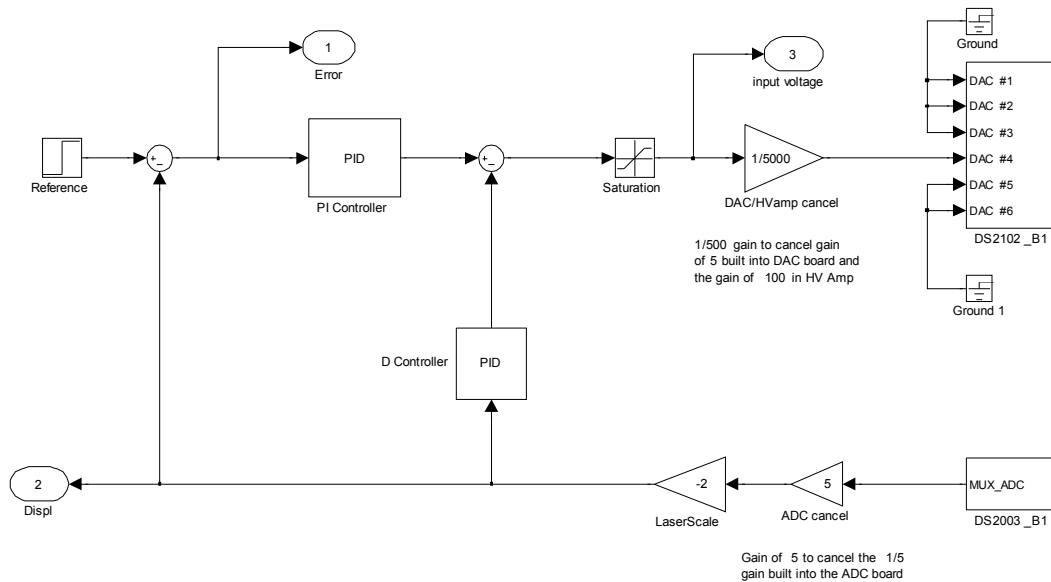
Figure 2.13   Simulink block diagram for closed-loop control experiments.

44

2.7 Noise Spectrum

It is wise to quantify noise or disturbances present in the experiment. Magnitude spectral analysis will reveal the magnitudes of any noise or disturbances associated with particular frequencies. The signal could potentially contain noise or disturbance components of varying amplitudes and at several frequencies, which can also be varying. It is convenient, then, to have a known component of the signal at a specified frequency to provide a frame of reference for judging noise and disturbances. Any frequency content not at the reference frequency is either a noise or a disturbance signal. Furthermore, comparison of the reference amplitude with that of the largest noise component reveals information about the signal to noise ratio.

The actuator was purposely driven with a 210 Hz sinusoid at various amplitudes. This reference frequency was chosen so that it was as far away from multiples of 60 Hz in either direction, and was high enough to avoid power leakage from the expected low frequency noise and disturbances. The peak input magnitude was varied discretely, with arbitrary values chosen as 1, 2.5, 5, 10, and 15 V. Amplitude was varied in order to determine if there was any correlation between noise or disturbance components and input magnitude.

The reference and noise components of each signal were assessed in the frequency domain so as to obtain the magnitude associated with each frequency component. An example of the resulting graph is shown in Figure 2.14. A 5 V peak magnitude at the 210 Hz reference frequency was used. The output shows a strong component at the reference frequency. A 5 V input evokes approximately a 0.3 μm (300 nm) output.
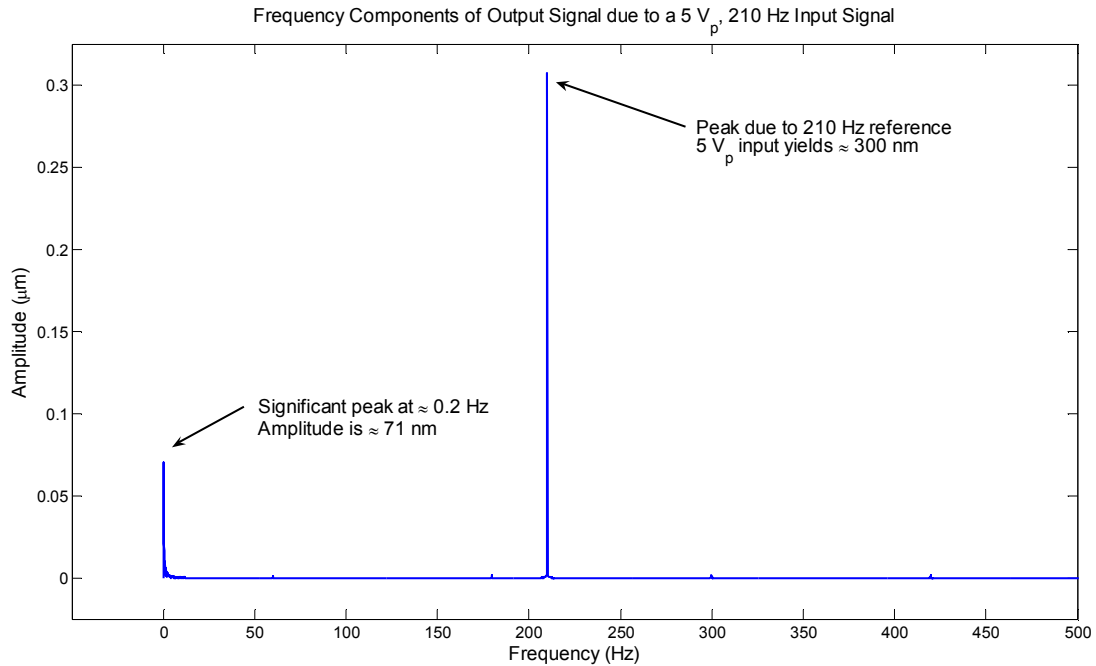
Figure 2.14   Example of frequency domain analysis of a signal. This example is from a 5 V_p input signal. Reference peak is at 210 Hz. Significant peak at 0.2 Hz is either noise or disturbance.

The low frequency noise component apparent in Figure 2.14 is the only one that persistently appears regardless of input magnitude or the number of times the experiment is performed. It is therefore the only noise component that was analyzed. However, it will not always persist at the exact same frequency and magnitude. It will thus disturb the reference signal in a slightly different manner each time the experiment is performed. Ten experiments were performed at each of the peak input magnitudes mentioned above so as to obtain results that carry statistical significance. The results for the reference and noise components of the output signal are given in Table 2.2 and Table 2.3 respectively.

Table 2.2   Statistical results for the reference component of the output signal.

Input Amp (V)	Frequency		Peak Amplitude	
	Mean (Hz)	Std. Dev. (Hz)	Mean (nm)	Std. Dev. (nm)
1	210	0.000	61.91	0.321
2.5	210	0.000	151.12	1.254
5	210	0.000	303.82	2.380
10	210	0.000	625.19	6.888
15	210	0.000	962.51	0.099

Table 2.3   Statistical results for the noise component of the output signal.

Input Amp (V)	Frequency		Peak Amplitude	
	Mean (Hz)	Std. Dev. (Hz)	Mean (nm)	Std. Dev. (nm)
1	0.19	0.120	34.29	17.724
2.5	0.22	0.140	26.70	17.008
5	0.20	0.141	44.48	17.455
10	0.17	0.067	39.78	20.504
15	0.13	0.048	28.76	17.190
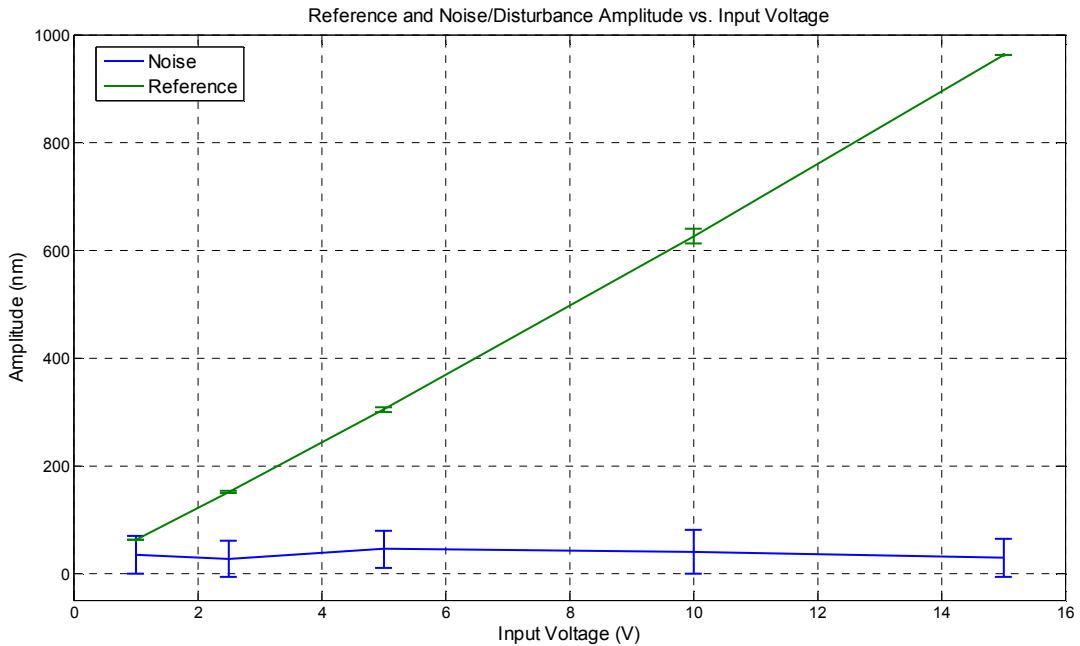
Figure 2.15   Reference and noise component data plotted together.   Reference amplitude increases with increasing input amplitude. Reference component error bars (~2 standard deviations) are small, meaning amplitude is very consistent. Noise amplitude remains nearly constant, implying no correlation with input amplitude. Noise error bars are large, implying some randomness.

Figure 2.15 shows the amplitude data of Table 2.2 and Table 2.3 plotted together to facilitate inferences about the reference signal amplitude, the noise component amplitude, and signal to noise ratio. As input amplitude increases, the reference amplitude increases with near perfect linearity. The reference error bars shown are ±2 standard deviations for a 95% confidence interval assuming a Gaussian distribution. The standard deviations are small regardless of input amplitude, meaning that the amplitude is consistent from one experiment run to the next.

The noise amplitude exhibits neither a steady increase nor a steady decrease with increasing input amplitude. No clearly deterministic correlation appears to exist between noise and input amplitudes. The same ±2 standard deviation error bars are applied to the noise. Large error bars are the result of large standard deviations. The lack of apparent determinism and large standard deviations imply that while the noise amplitude may appear bounded, it is mostly random.

Since there is no apparent deterministic relationship between input amplitude and noise amplitude, it is safe to concatenate all ten experiments from all five input voltage amplitudes. This results in an overall noise amplitude mean and standard deviation of 34.8 nm and 18.5 nm respectively. These comprehensive statistics can be used to assess the ratio of signal amplitude to noise amplitude.
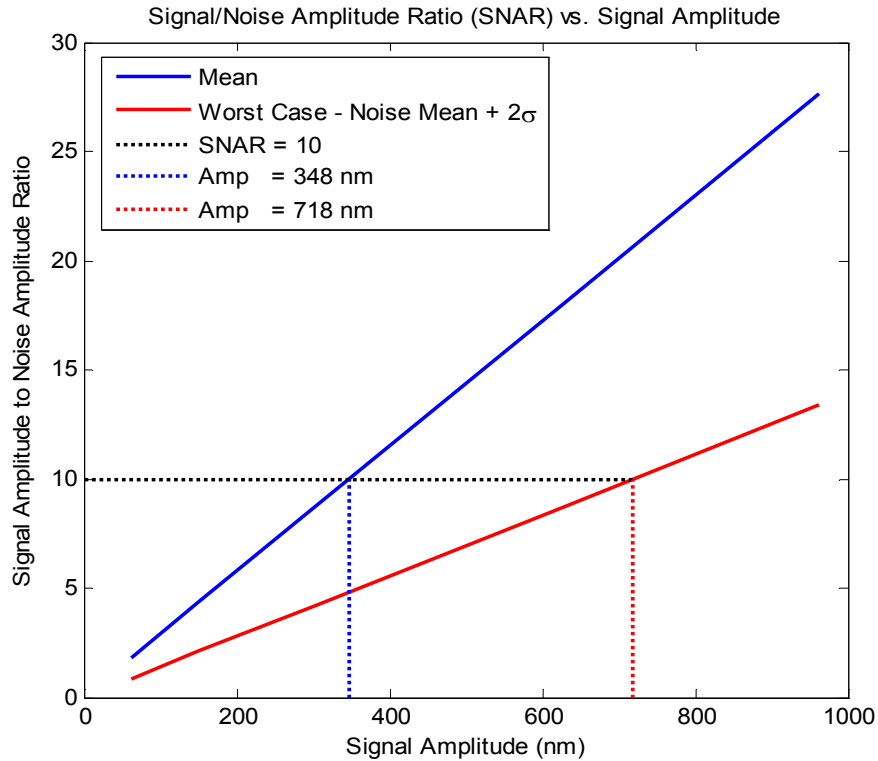
Figure 2.16   Signal amplitude to noise amplitude ratio (SNAR). As expected, SNAR increases with increasing signal amplitude.

Mean and worst-case signal to noise amplitude ratio (SNAR) using the overall noise amplitude statistics are shown in Figure 2.16. The mean case is the signal amplitude mean divided by the noise amplitude mean, and represents the SNAR that can be expected most of the time.   The worst case is the signal amplitude mean divided by the noise amplitude mean plus two standard deviations, and represents the SNAR that results when the largest disturbances within a 95% confidence interval are present.   The SNAR increases with increasing signal amplitude, which is to be expected.   An acceptable value for signal-to-noise ratio depends on the application and desired accuracy. So to provide an example, SNAR = 10 is shown in Figure 2.16 for both the mean and worst cases. In the worst case the SNAR does not exceed ten until the reference signal amplitude is at least 718 nm (0.718 μm).
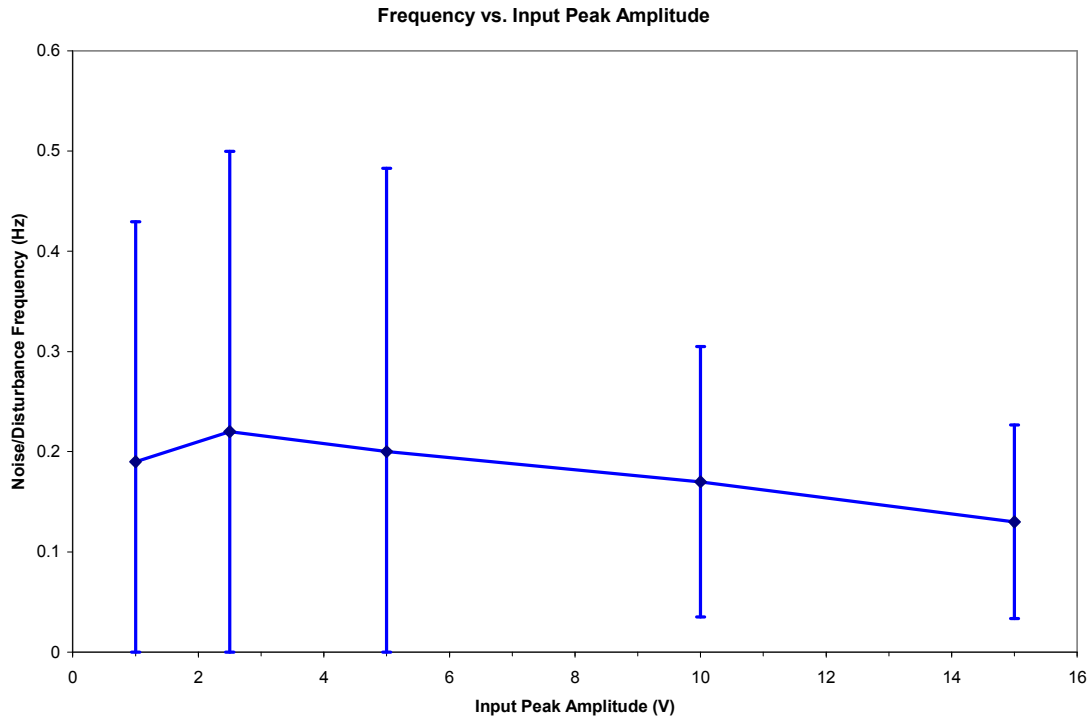
Figure 2.17 **Variance in frequency of noise/disturbance. Error bars are ±2 standard deviations except with the first three data points where this would include negative frequencies. Error bars were then truncated at zero, since negative frequency is an erroneous concept.**

The frequency variance of the noise is also worth addressing in order to give a full picture of what is to be expected. Figure 2.17 shows the noise component frequency data of Table 2.3 as it varies with peak input amplitude. Error bars are ±2 standard deviations, except when this resulted in negative frequencies, in which case the error bars were truncated at zero. While the error bars appear large, considering that the scale on the y-axis is in tenths of a Hertz, the variance is not that large after all. From Figure 2.17, it is expected that the noise component will consistently occur between 0 and 0.5 Hz.

Resolving whether or not the actuator is in fact accurately tracking a desired trajectory becomes increasingly difficult as the magnitude of the trajectory decreases. There is no particular accuracy or signal-to-noise ratio requirement for this work, as

there is not one specific application in mind. Limitations on what to require of the experimental setup must still be chosen. The displacement signal of the laser vibrometer is a voltage with a user-selectable scale. Note in Appendix D that all available scale proportions for the laser sensor are measured in μm/V. The position output of the actuator model presented in the next chapter will thus be measured in μm. Signal-to-noise ratio exceeds 10 at displacement magnitudes of greater than 0.718 μm. A signal-to-noise ratio of greater than 10 is by no means a definitive rule for all measurement cases, but it is assumed to be sufficient here. A 1 μm step yields a signal-to-noise ratio of 14, and conveniently corresponds to the quintessential unit step response in control system evaluation. Therefore, closed-loop step responses of no smaller than 1 μm from a zero set point will be required.

3 Actuator Modeling

A dynamic model is the basis for closed-loop control system design. In some cases, a system model may already be available. In other cases, however, a model must be developed and validated independently. Regardless of the modeling method chosen, the process usually begins with an experimental input/output response. From this response a simple model is constructed with the method of the designer's choice. The model is increased in complexity as needed until it is deemed sufficient for its intended use, which is usually to either predict or simulate the system output resulting from some predetermined input. As for this work, a model sufficient for feedback controller design is sought. The system input and output are applied voltage and displacement (or position) of the actuator respectively.

3.1 Open-Loop Response

The steel shell (see Figure 2.1) is essentially two curved beams, symmetric about a central plane. On that central axis lies the piezostack which is a composite beam of ceramic, electrodes, and epoxy that functions as an electric field-to-force transducer. It will have the inherent mechanical characteristics of mass, damping, and stiffness. As such, the actuator is treated as a mechanical system with the applied force being a function of applied voltage.

3.1.1 Time-Domain Response

A sine sweep was performed manually to determine the approximate frequencies of any resonances that the system may have. Two distinct resonances appear in the frequency decade of 1 to 10 kHz. Beyond this the actuator response appears to roll off until the noise floor of the sensor is reached. Below this range the actuator sustains a sinusoidal response with near constant amplitude. This defines the approximate frequency range over which to perform an automated sine sweep. It also defines the operational range of the system.

Sine sweeps, or chirp signals, with a linear frequency sweep pattern from 100 Hz to 10 kHz and various peak amplitudes between 3 and 9 volts were applied to the actuator to gather data sets. Experiment length was 10.486 seconds, and was sampled at 50 kHz. This time length was chosen coarsely to capture multiple periods at all frequencies in the sweep range, and specifically to create data sets whose lengths are an integer power of two. This is simply because the data analysis techniques rely on discrete Fourier transform methods, which work best if the data record length is an integer power of two. A representative section of the time-domain input signals and its power spectral density (PSD) are shown in Figure 3.1. Only a small time span of the signal is shown. Due to the frequency range and time span, plotting the entire signal results in a plot too dense to visually discern any frequency change. The PSD, however, applies to the entire signal length. Applying the input signal of Figure 3.1 yields the input-output pair shown in Figure 3.2.
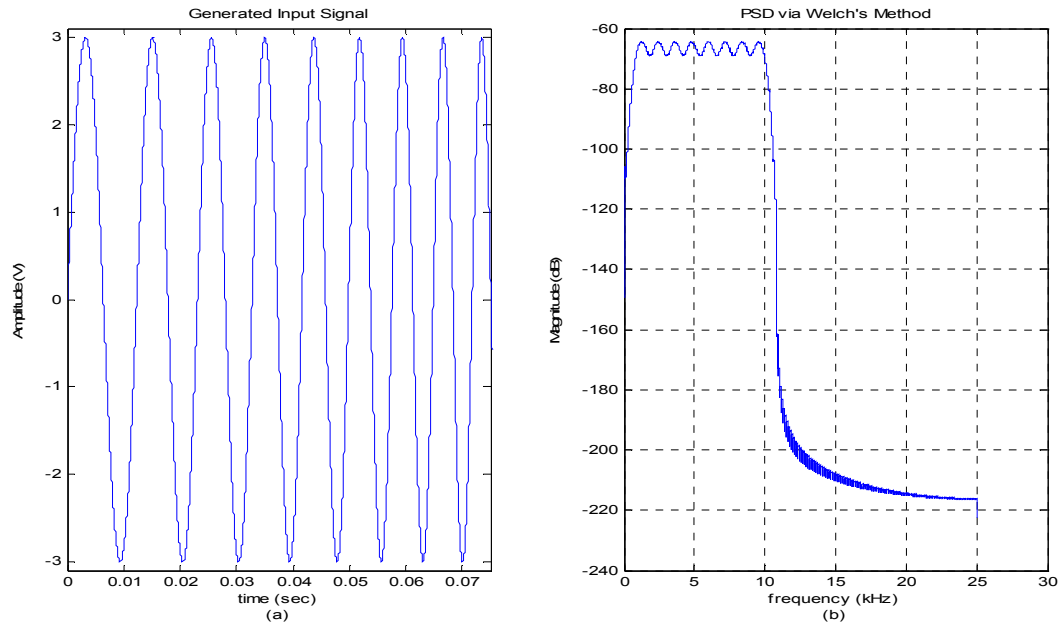
Figure 3.1 (a) Generated sine-sweep input signal and (b) its power spectral density.
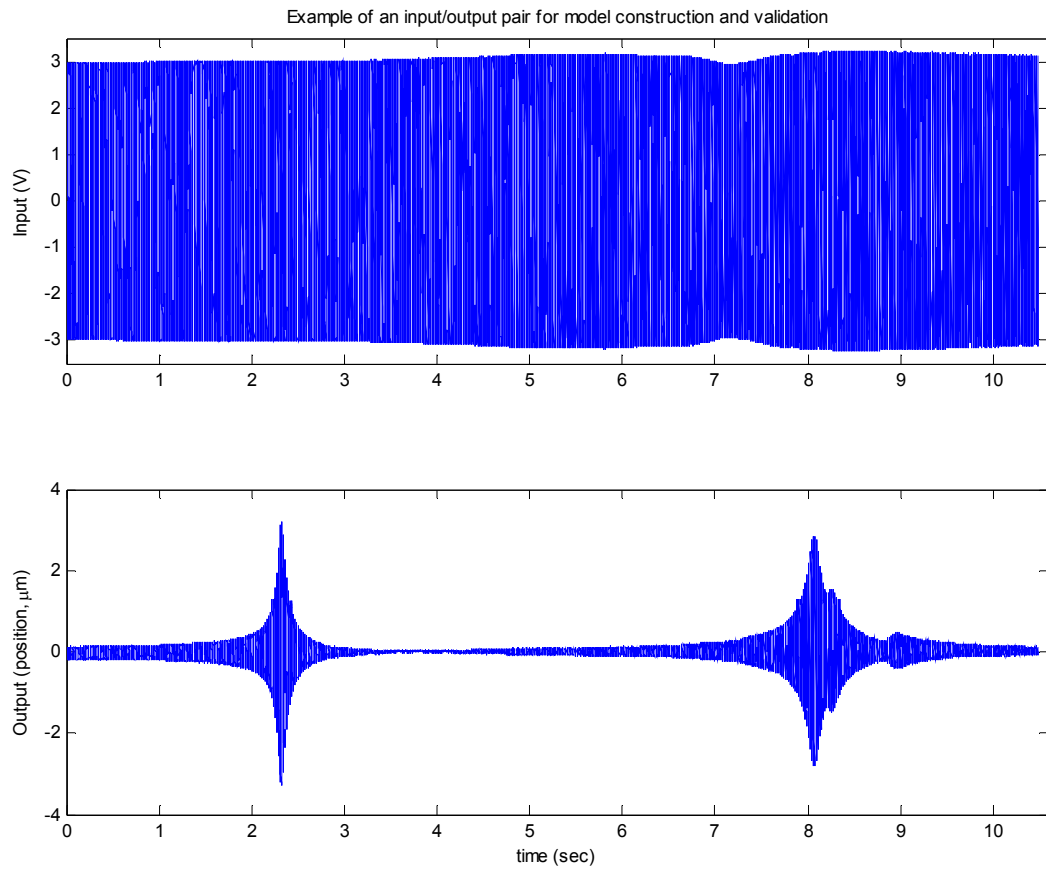
Figure 3.2 An exemplary input/output pair for model development.

3.1.2 Experimental Frequency Response

From the experimental data shown in Figure 3.2, or any other like it, a frequency response, or transfer function estimate, can be obtained. Given an input/output signal pair, a frequency response, or a Bode plot, is given by Equation (3.1) [13, 49-50], wherein Φ_{yu} is the cross power spectral density between the input and output, and Φ_{uu} is the power spectral density of the input.

$$G(\omega) = \frac{\Phi_{yu}(\omega)}{\Phi_{uu}(\omega)} \tag{3.1}$$

Shown in Figure 3.3 is an experimental frequency response estimate from one of these data sets and calculated as in Equation (3.1). The two resonant peaks appear at approximately 2.4 and 8.3 kHz. The frequency of the second peak is not an integer multiple of the first, and is nearly equal in magnitude as well. This is strongly indicative of a two degree-of-freedom (DOF) system, i.e. two coupled, second-order systems [51-52].
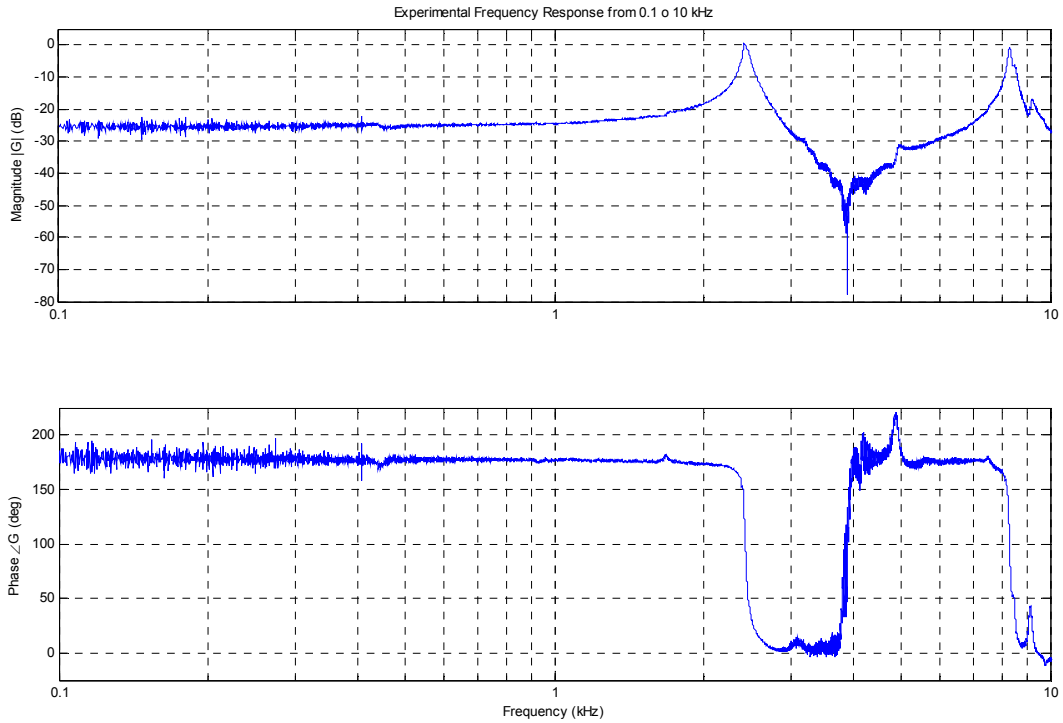
Figure 3.3   An experimental frequency response, 0.1 to 10 kHz.

Figure 3.4 shows the coherency spectrum of the input/output pair used to generate the frequency response given in Figure 3.3. Coherence is a measure of linear correlation between two signals [49-50,53], and is given by Equation (3.2). A coherence value of unity implies perfect linear correlation, whereas a coherence value of zero implies no linear correlation between the two signals whatsoever

$$\gamma_{yu}(\omega) = \frac{\left|\Phi_{yu}(\omega)\right|}{\Phi_{yy}(\omega)\Phi_{uu}(\omega)} \tag{3.2}$$
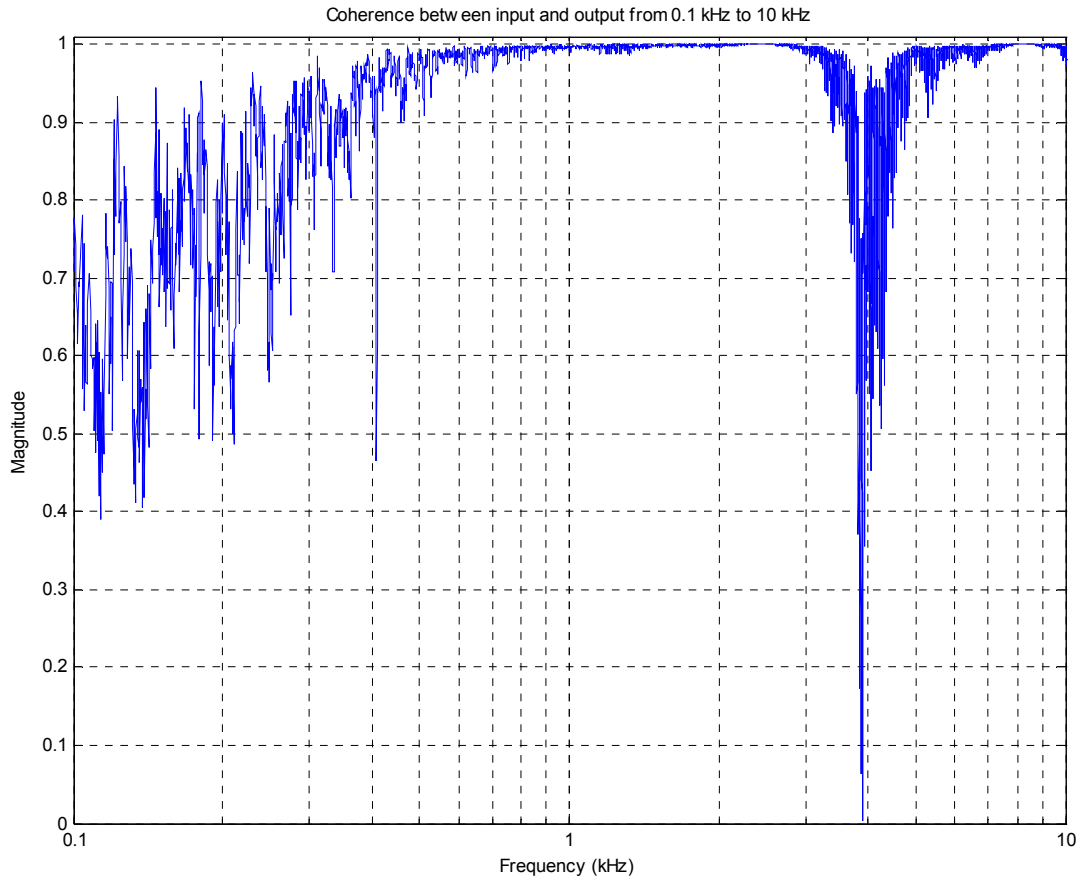
Figure 3.4   Coherence of input/output pair used to create Figure 3.3.

The frequency response shows excellent coherence throughout most of the frequency band in which it was tested. It is near perfect at and around the resonant peaks, and it is good in nearly all of the lower frequency decade of 100 to 1000 Hz. Coherence is low only around the anti-resonant peak at approximately 3.9 kHz. This is to be expected, as the displacements drop below the noise floor of the sensor. In other industries, a general rule of thumb is that if coherence is greater than or equal to 0.6 then the frequency response is sufficiently accurate and it is safe to use linear modeling techniques [49]. This rule is followed here as well. A linear relationship between displacement and applied voltage is assumed.

Also worth mentioning with regard to Figure 3.3 are the relatively small peaks at approximately 4.9 and 9.2 kHz. Others like these, along with the anti-resonant peak, appear and/or diminish dependent upon spatial location of the sensor on the working surface of the actuator. They change in magnitude and location when the laser displacement sensor is shined at different points on the working surface of the actuator. The two dominant peaks, however, accurately maintain their magnitude and location regardless of the sensor location. Several experiments with the laser pointed at different locations were run to verify the previous statements, a sample of which is shown in Figure 3.5.
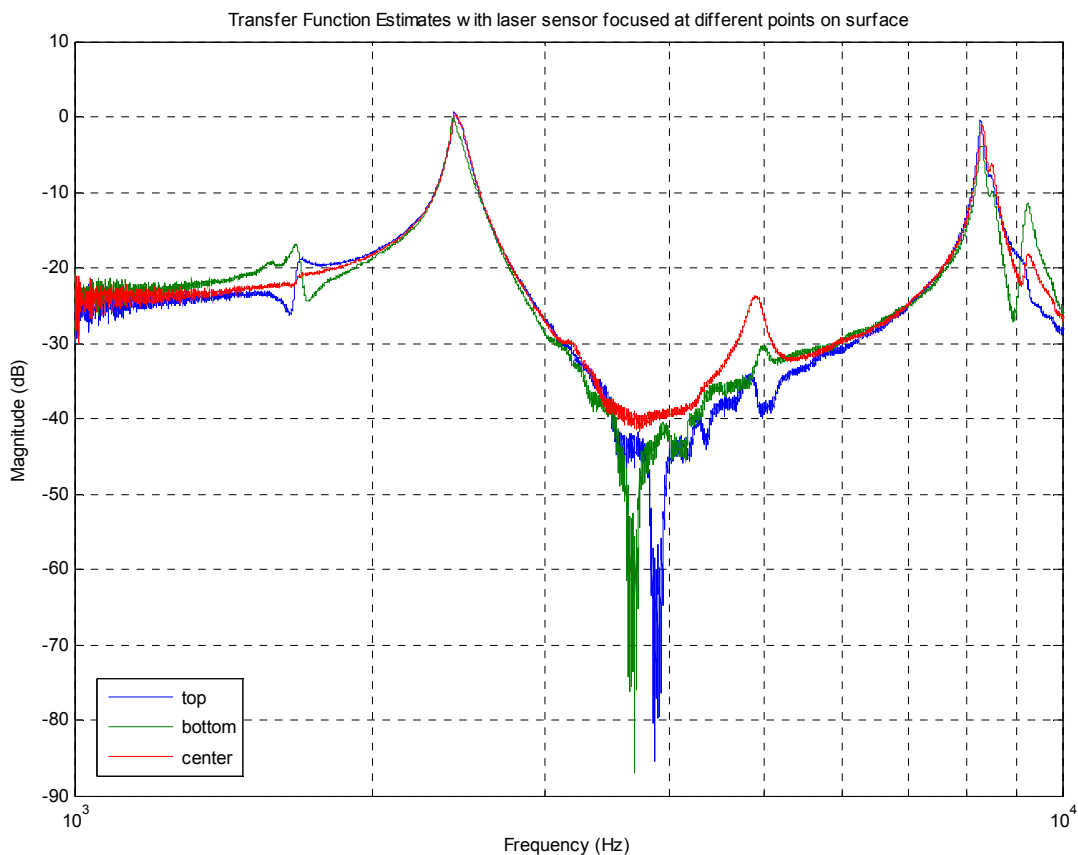
Figure 3.5 **Transfer function estimates from time-domain responses with the laser pointed at top, bottom, and center of actuator's working surface. Small mode at 1.7 kHz inverts when comparing top and bottom locations. Mode at 9.2 kHz diminishes when laser is pointed at the top portion. The two large, dominant modes remain fixed.**

Figure 3.5 shows inversion of a mode at approximately 1.7 kHz when comparing the top (above center) vs. bottom (below center) laser location, as well as that mode diminishing when measurement location approaches the center of the actuator. This confirms that the smaller peaks correspond to a torsion mode of the metal shell component. It also shows the larger modes' magnitude and frequency independence from the sensor spatial location, which confirms that they are associated not with torsion, but with the gross displacement of the actuator.

3.1.3 Assumptions for Model Development

The underlying purpose here is to accurately control the gross displacement of the overall actuator system. Furthermore, including the torsional modes of the metal shell in the model for the overall system is beyond the scope of this work. So, model development proceeds under the following assumptions: the system can be well-approximated linearly, it has two dominant modes and is thus a 2DOF system, and only these two modes are to be captured by the model. The smaller modes not being included are always present to some degree in experimental data. This will degrade model fit estimates, so data sets will be chosen for model estimation and/or validation that have as little of these effects as possible.

3.2 System Identification for Model Construction

There are many facets to system identification and its implementation, and it is wise to be familiar with them before usage. Those which are beyond or within the user's control must be differentiated so that informed decisions can be made when constructing a model. Model characteristics such as order and structure will always be at the user's discretion. How the model parameters are determined, however, may not be. Still, having knowledge of both is often necessary.

3.2.1 Generalized Model Parametrization

The primary purpose of a dynamic model is to describe the relationship between the system input(s) and its output(s). Perhaps the most fundamental expression of such a relationship is given by the difference equation of Equation (3.3), wherein $y(t)$ is the system output, $u(t)$ is the system input, and $e(t)$ is a disturbance term, often modeled as white-noise [13]. The parameters a_i and b_i are coefficients to be adjusted for fitting the model to an input/output data set. By grouping all a_i and b_i together into the vectors $A(q)$ and $B(q)$, Equation (3.3) can be rationalized as Equation (3.4) where q and q^{-1} are the forward and backward shift operators respectively.

$$y(t)+a_1 y(t-1)+\ldots+a_{n_a} y(t-n_a) = b_1 u(t-1)+\ldots b_{n_b} u(t-n_b)+e(t) \qquad (3.3)$$

$$y(t) = \frac{B(q)}{A(q)} u(t) + \frac{1}{A(q)} e(t) \qquad (3.4)$$

Equation (3.4) predicts the current output as a function of past inputs and outputs and the addition of a disturbance term. Note its similarity to the family of transfer functions described by Equation (1.2) and the different family members belonging to it listed in Table 1.2. The model of Equation (3.4) is the familiar ARX model mentioned in Section 1.2.2.

Recall that the coefficients of the polynomials $A(q)$ and $B(q)$, and those of $C(q)$, $D(q)$, and $F(q)$ if a model structure other than ARX is selected, must be estimated. The manner in which they are estimated is often beyond the control of the model designer, as the method may be fixed by an existing identification software package. It is still wise to be aware of methods used.

$$\theta = \left[a_1 \ldots a_{n_a} \quad b_1 \ldots b_{n_b} \quad c_1 \ldots c_{n_c} \quad d_1 \ldots d_{n_d} \quad \ldots \right]^T \qquad (3.5)$$

$$\varphi(t) = \left[-y(t-1) \ldots -y(t-n_a) \quad u(t-1) \ldots u(t-n_b) \quad e(t-1) \ldots e(t-n_c) \quad \ldots \right]^T \quad (3.6)$$

$$\hat{y}(t \mid \theta) = \varphi^T \theta = \theta_1 \varphi_1 + \theta_2 \varphi_2 + \theta_3 \varphi_3 + \ldots \qquad (3.7)$$

$$\varepsilon(t, \theta) = y(t) - \hat{y}(t \mid \theta) \qquad (3.8)$$

$$V_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} l(\varepsilon(t, \theta)) \qquad (3.9)$$

$l(\varepsilon(t, \theta))$ = positive, scalar function

$$\hat{\theta} = \arg\min V_N(\theta) \qquad (3.10)$$

Linear regressions involving constraints on the prediction error are the most commonly employed [12-13]. The unknown parameter and regression vectors, θ and $\varphi(t)$, are given by Equations (3.5) and (3.6). They come together to form the model estimate, $\hat{y}(t\,|\,\theta)$, given by Equation (3.7). Prediction error $\varepsilon(t,\theta)$ is the difference between the experimental value and the value predicted by a model, and is given by Equation (3.8). Prediction error methods, as the name suggests, seek to minimize the prediction error, so a generalized criterion function based on that prediction error is formed and given by Equation (3.9), in which $V_N(\theta)$ is the criterion function and the defining function $l(\varepsilon(t,\theta)$ must be produce positive scalar values. As this is a quantification of the difference between the experimental data and the model, the problem becomes finding the parameter vector estimate $\hat{\theta}$ that minimizes the prediction error criterion function, i.e. Equation (3.10). Note that the previous discussion with the associated equations is a brief excerpt from [12-13], in which a more thorough explanation can be found.

Many methods can be classified as a prediction-error method as described above. A specific method typically has a particular choice of criterion function or minimization method associated with it. For instance, choosing $l(\varepsilon(t,\theta))$ as quadratic in θ results in the familiar least-squares method [13]. Only those used in the Matlab System Identification Toolbox™ v7.0 (MSITB) that were used by the author need be mentioned, as this was the algorithm collection used for model development during this work.

One method employed, the least squares method, has already been mentioned. It is deterministic, and is only used in the case of the ARX model, the simplest model

structure. This is intuitive because the noise/output dynamics share the denominator polynomial of the input/output dynamics. The least squares method is the solution of the criterion function derivative being set equal to zero. There is another, optional method for the ARX model estimation known as instrumental variables. An instrumental variable approach can be used when it is suspect that there is some relationship between the noise present and the regressors used [13]. Instrumental variables appears to be similar to least squares, but was not used here.

The other model structures listed in Table 1.2, variants of the ARX structure, employ a maximum likelihood method. Maximum likelihood, like least squares, is a particular case of the prediction error method and is distinguished by its choice of $l(\varepsilon(t,\theta))$ in the criterion function of Equation (3.9), which is shown in Equation (3.11) wherein $pdf(\varepsilon)$ is the probability density function of the prediction error.

$$l(\varepsilon(t,\theta)) = -\log[pdf(\varepsilon)] \tag{3.11}$$

It can be determined from Equation (3.11) that unlike the least squares approach, which is deterministic, the maximum likelihood approach, as it applies to system identification, is probabilistic. This choice of criterion function can make solving Equation (3.10) difficult. Generally, for cases other than the least squares estimate, the problem cannot be solved analytically [12-13]. In this event, iterative, numerical optimization techniques must be used.

This is in fact the approach taken in the MSITB when using model structures of the form given by Equation (1.2) other than the ARX structure, i.e. ARMAX, OE, BJ models, etc. The initial guess, if not supplied by the user, is estimated with a least

squares method as in the ARX case. From there a search direction is sought that will result in lower criterion function value. This search direction is found using one of three methods, those being the Gauss-Newton, the regularized Gauss-Newton, or the Levenberg-Marquardt method. The user can select a method or allow the algorithm to decide itself.

3.2.2 Model Structure and Order Selection

With a reasonable amount of computational power, many models can be constructed in a short amount of time. This is both a boon and a burden. It is advantageous because of the resulting number of options. There can easily be several models of varying structure, order, and time delay from which to choose. This is, at the same time, disadvantageous because when many of the models have similar performance, the question arises of how compare those models and decide which one is best. This can sometimes be a difficult task, especially since the concept of the best model may vary with the model's intended purpose.

It is often best to begin with the simplest model structure with moderate orders for the polynomials (or number of states if estimating a state-space model) to gain insight, and then build in complexity as necessary. To this end, the following procedure as outlined in [12] can be applied.

1. Begin with an ARX model, the simplest structure. Select moderate orders for all polynomials, fourth-order of often recommended. Vary the time delay to determine which gives the best model performance.

2. Using this time delay, construct many ARX models of varying order. Select those that give the best model performance.

3. If the model order seems unnecessarily high, look for pole-zero cancellations. The number of poles and zeros that survive serve as an estimate of the order for the input/output relationship. Use this order as a starting point for more complex models such as ARMAX, OE, BJ, etc. with a first- or second-order noise model

It is generally accepted that when developing a system model, it is impossible to perfectly capture its behavior. In light of this fact, no set of guidelines, no matter how effective or well-followed, can ever be expected to force the plant model behavior to converge to the actual system's exact behavior. The above set of guidelines, however, prove useful, as they provide a logical, systematic process to follow, rather than simply a brute-force approach of randomly constructing models until a sufficient model is found.

3.3 Actuator Model

System identification modeling as embodied by the previous sections has been successfully applied to piezoelectric and other smart material systems in several cases [23-24, 28, 54]. It was applied here by following the process for order selection outlined in the previous section. Separate data sets like the one shown in Figure 3.2 were used for model estimation and validation purposes, since this is recommended when using system identification techniques [13]. The resulting model is given by

$$y(t) = \frac{B(q)}{F(q)}u(t) + \frac{C(q)}{D(q)}e(t), \text{ where}$$

$$B(q) = -0.0199q^{-1} + 0.03581q^{-2} - 0.02098q^{-3} + 0.0006357q^{-4}$$

$$C(q) = 1 + 0.5793q^{-1}$$

$$D(q) = 1 - 0.4078q^{-1}$$

$$F(q) = 1 - 2.895q^{-1} + 3.862q^{-2} - 2.845q^{-3} + 0.9694q^{-4}.$$

(3.12)

Using the model given by Equation (3.12), both an input/output simulation and a frequency response, or Bode plot, were generated for comparison to their experimental counterparts. The time-domain output of the model as compared to the experimental output is shown in Figure 3.6 (the experimental input signal is applied to the model, thus input is the same in both cases). The model shows an 82.2% fit where the fit is given by Equation (3.13), in which y is the experimental output, $\hat{y}$ is the simulated output, $\bar{y}$ is the experimental output mean value, and $\|\cdot\|$ denotes the norm value [55].

$$fit = \left(1 - \frac{\|y - \hat{y}\|}{\|y - \bar{y}\|}\right) * 100$$

(3.13)
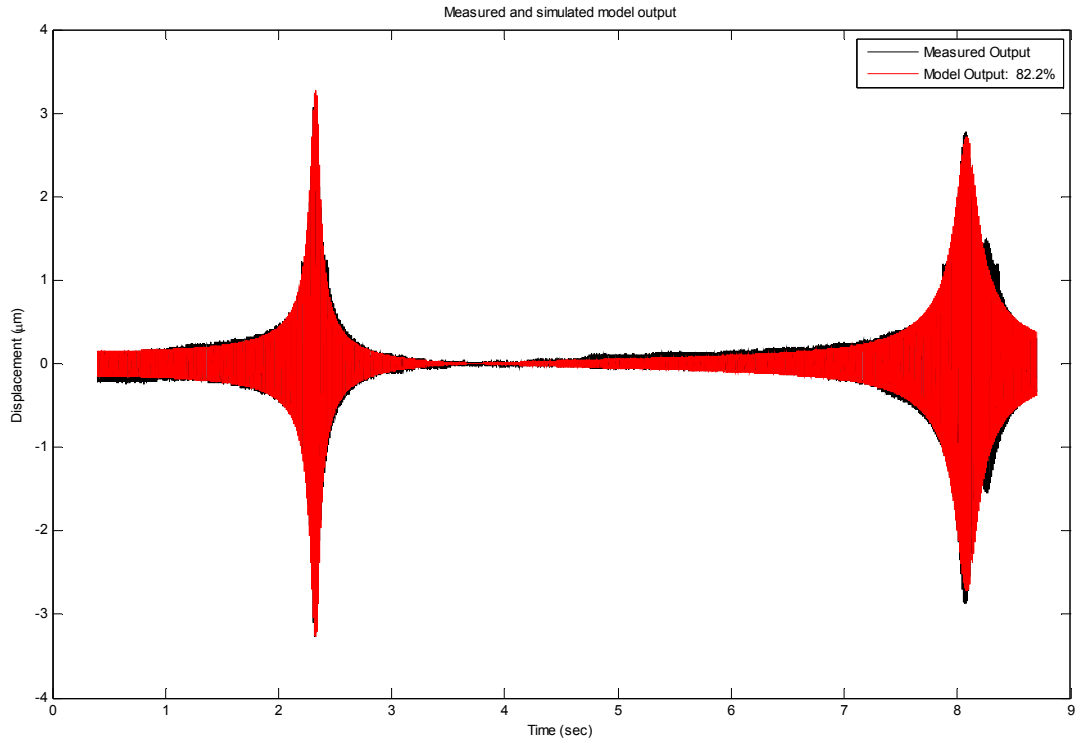
**Figure 3.6   Experimental output response compared to model output response.   The
input signal is the same for both.   The model shows a fit of approximately 82%.**

Shown in Figure 3.7 is the frequency response comparison.   With the
exception of the torsion modes that are not being captured, the correlation is excellent
throughout the entire frequency range of interest.   The model accurately recreates
both resonant peaks and the anti-resonant peak that are being captured.   Together,
Figure 3.6 and Figure 3.7 provide strong evidence that Equation (3.12) is a suitable
model of the actuator.

Experimental Frequency Response and Model Bode plot
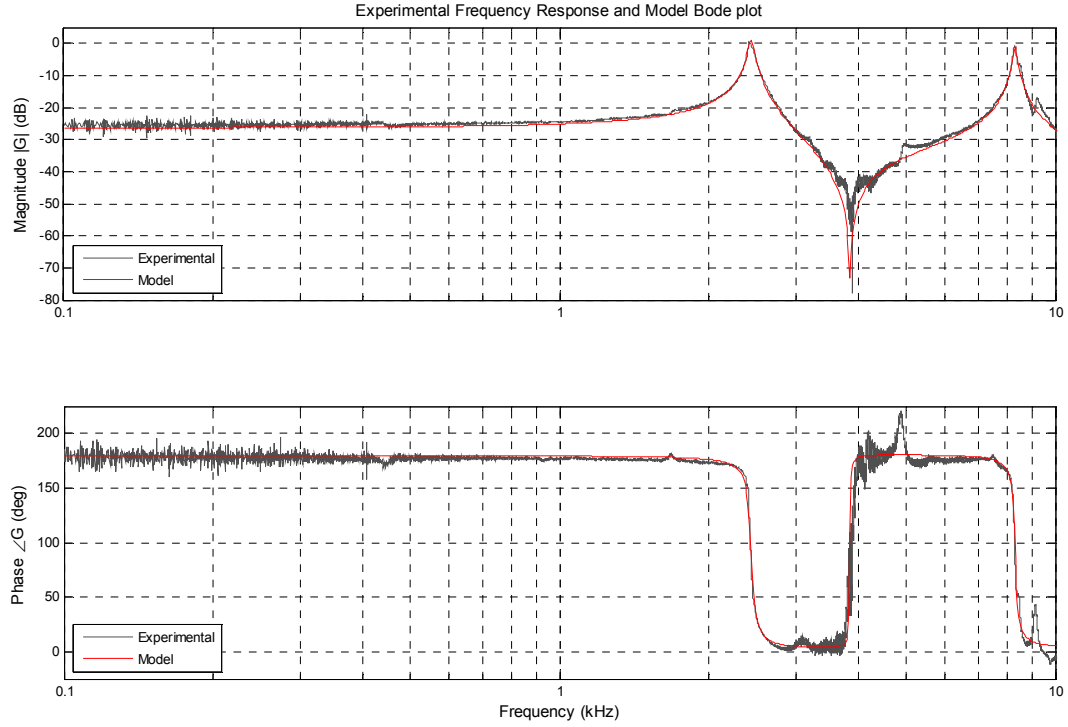
Figure 3.7   Experimental frequency response and model frequency response.

$$z = e^{Ts} \tag{3.14}$$

$$G(s) = \frac{-307.4s^3 - 5.294(10^7)s^2 - 1.867(10^{11})s - 3.105(10^{16})}{s^4 + 1552s^3 + 2.953(10^9)s^2 + 1.487(10^{12})s + 6.37(10^{17})} \tag{3.15}$$

The model of Equation (3.12) was calculated using the MSITB, which results in its form being completely equivalent to the z-transform [56], i.e. the *q* operator is directly equivalent to the *z* operator. The z-transform applies to discrete-time systems in the same manner as the Laplace transform applies to continuous-time systems. The relationship between the two transforms is given by Equation (3.14), in which *T* is the sampling period, s is the Laplace variable, and z is the z-transform variable [57]. This

allows mapping between the discrete-time *z*-plane and the continuous-time *s*-plane. Using Equation (3.14) for a continuous/discrete conversion (or vice versa) is known as the matched pole-zero method [14]. Applying it to Equation (3.12) results in the continuous-time model given by Equation (3.15).

The model of Equation (3.15) will have the same frequency response and input/output reproduction capability as the discrete-time model from which it was derived. The discrete-time model was verified with a sine-sweep data set as shown in Figure 3.6, which is same input type that created the data set used to create the model. It is often wise to test a completely different input shape and its corresponding output to assess model validity. The validity of Equation (3.15) was assessed with 5 and 10 V step inputs, as well as with a 10 V doublet (a doublet consists of two consecutive step inputs of equal magnitude and opposite direction).
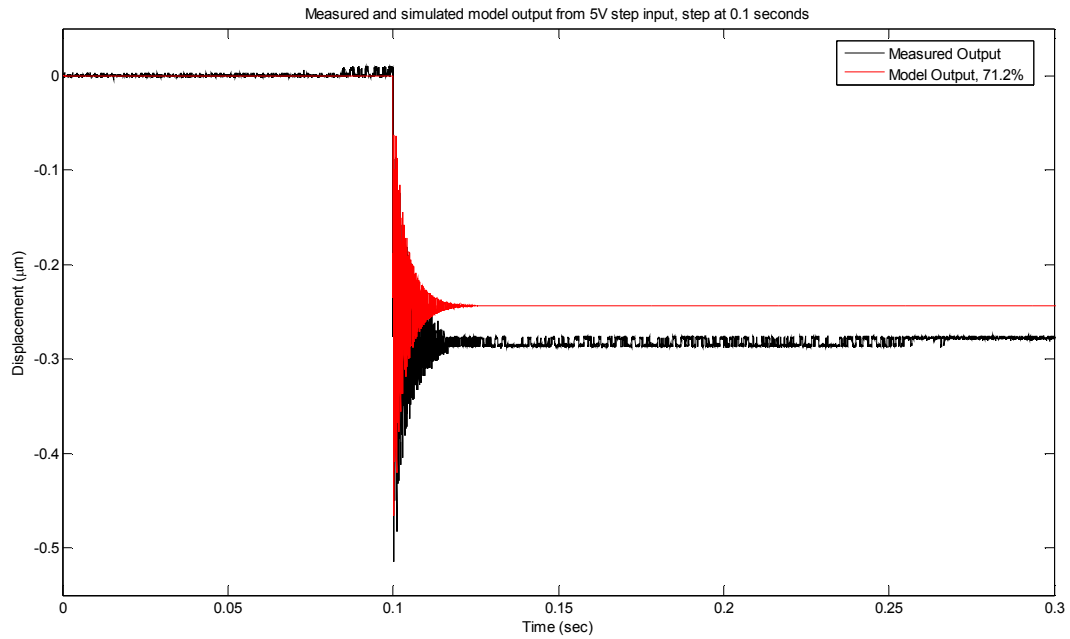
Figure 3.8 **Continuous-time model output compared to measured output, both are output response due to a 5 V step input at 0.1 seconds. Model shows 71.2% match. Most of the deficiency in model fit appears to be the result of a DC gain that has error in estimation.**
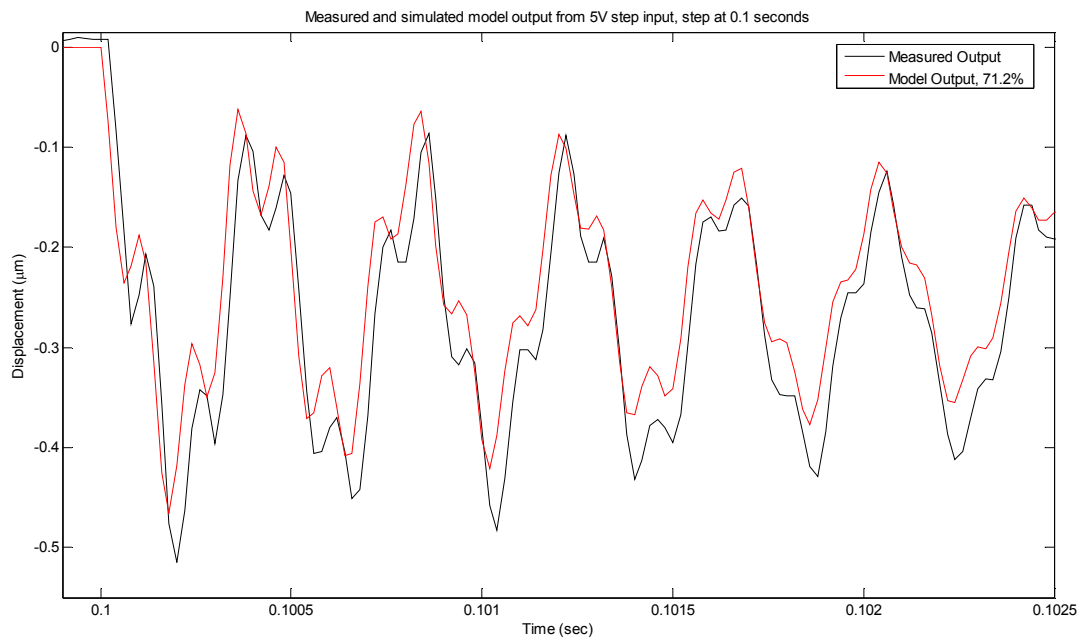
Figure 3.9 **The same model and measured output comparison in Figure 3.8, but zoomed in to better show the first several periods of oscillation. In spite of inaccurate DC gain, transient dynamics seem to be well-captured.**

70

The actual and simulated response to a 5 V step input at 0.1 seconds is shown in Figure 3.8. The model fit according to Equation (3.13) is 71.2%. An error in the DC gain estimate of the model is apparent from the noticeable difference in steady-state values of the measured and simulated outputs. This inaccuracy in the DC gain is the likely cause of the deficiency in model fit. Transient dynamics of the 5 V step response appear to be well-captured, as shown in Figure 3.9. The general shape of the response is the same in both. A small amplitude difference and a minor phase shift are apparent. This provides further confirmation that most of the fit deficiency is due the inaccurate estimation of the DC gain.
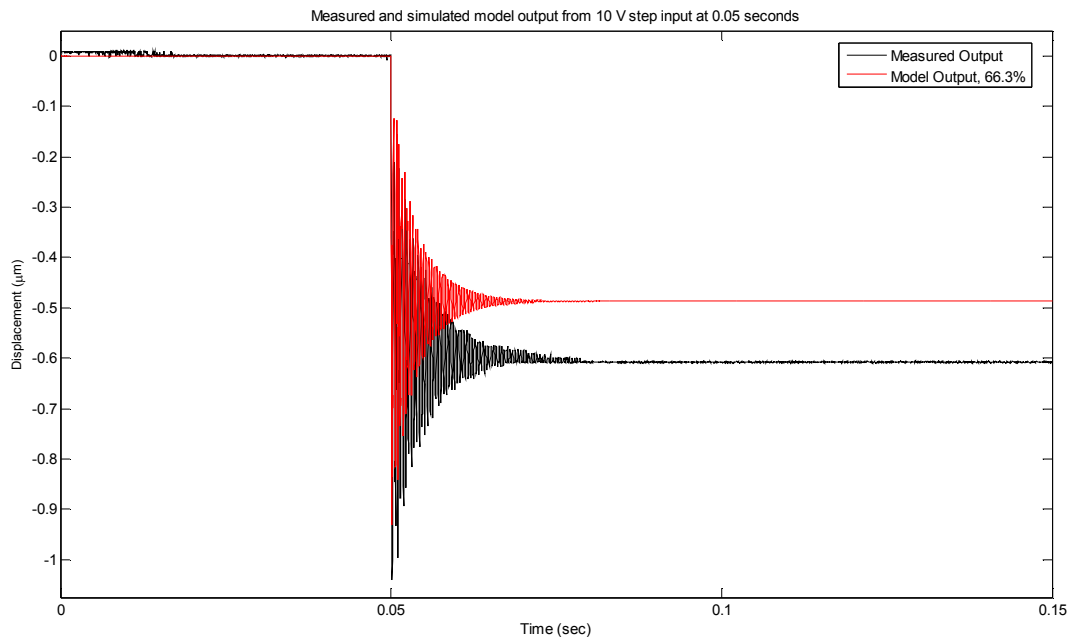
Figure 3.10   Comparison of model and measured output due to a 10 V step input at 0.05 seconds. The model is a 66.3 % match. Just as with the 5 V step case, the deficiency in model fit seems to be due to an inaccurately estimated DC gain.

71
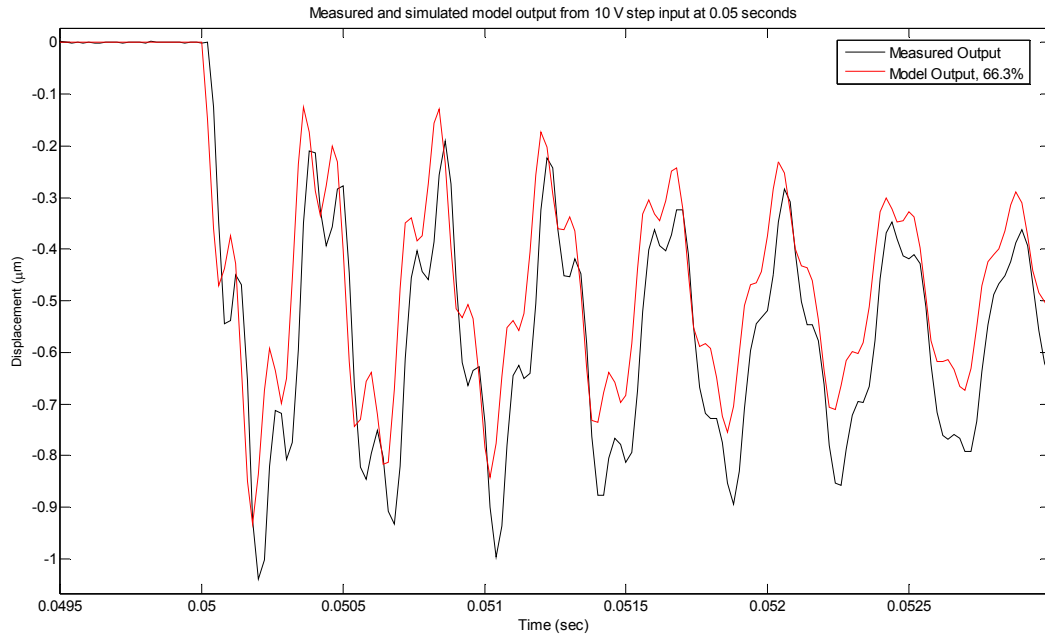
Figure 3.11   Same comparison of 10 V step response as in Figure 3.10, but zoomed in to better show first several oscillatory periods. Transient, vibratory dynamics appear to be well captured in spite of DC gain inaccuracy.
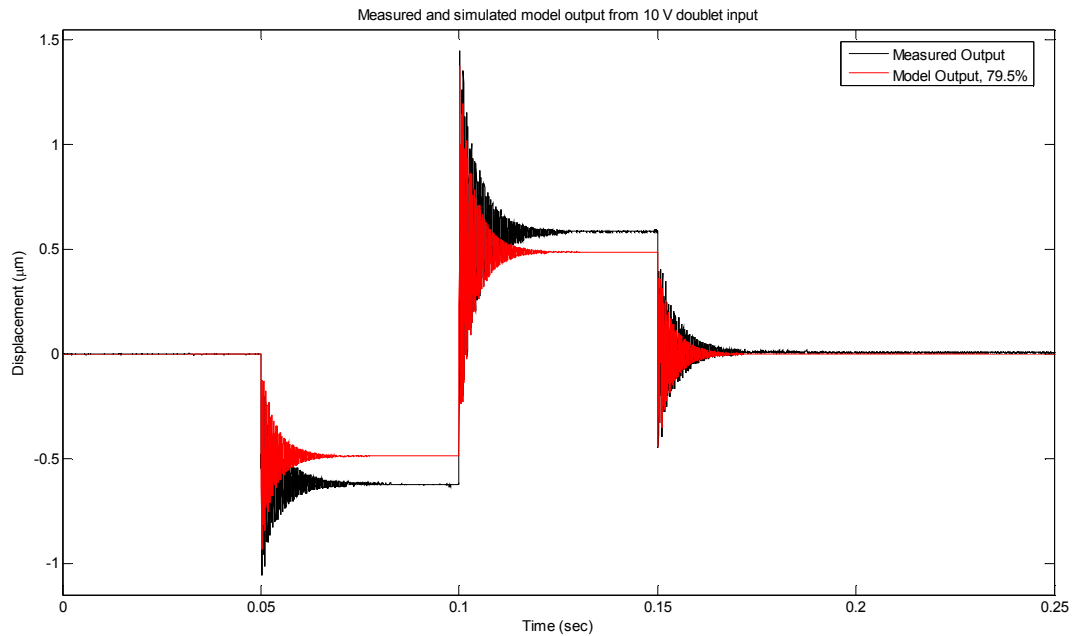
Figure 3.12   Comparison of model and measured output due to a 10 V doublet. Model is a 79.5% match. Just as with 5 and 10 V step cases, most of model fit deficiency appears to be the result of a misestimated DC gain.
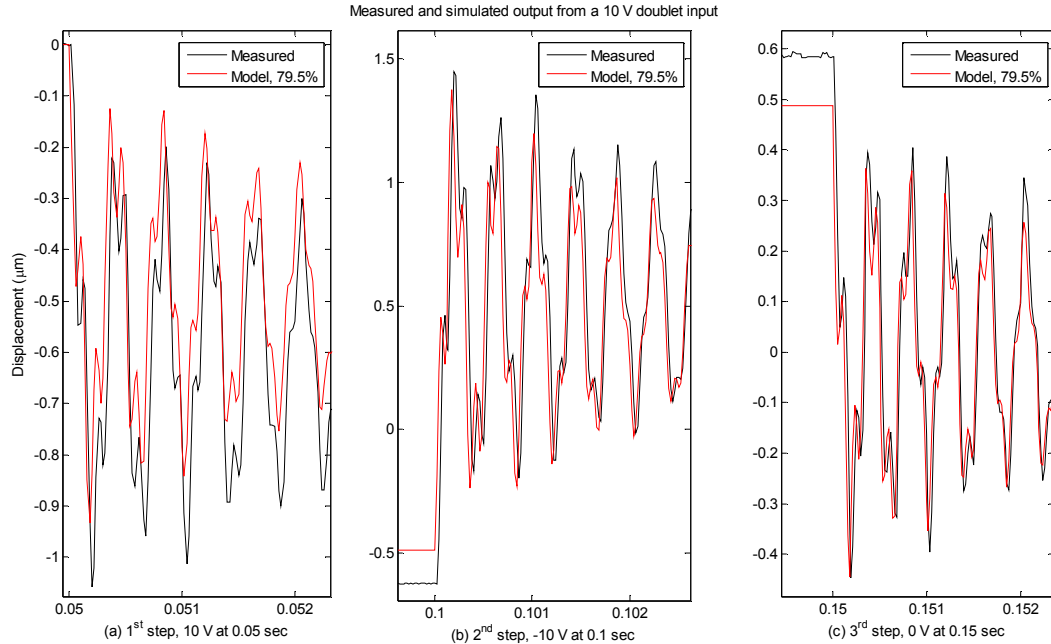
Figure 3.13   Same measurement and model comparison of 10 V doublet response as in Figure 3.12, but zoomed in on (a) first, (b) second, and (c) third steps to better show transient dynamics.   Just as before, transient dynamics appear to be well-represented in spite of DC gain inaccuracy.

The same results occur in the 10 V step and the 10 V doublet responses as illustrated by Figure 3.10 through Figure 3.13. Model fit estimates varied between all three responses. The 10 V step and 10 V doublet showed model fits of 66.3% and 79.5% respectively. Transient dynamics, however, appear to be well-captured in all three cases.

The misestimated DC gain appears to be the source of most of the deficiency in model fit estimates. This problem could be effectively mitigated by simply multiplying the model by a constant. However, closed-loop control adequately compensates for this type of issue. Model error is one the main reasons for using closed-loop control [15]. Therefore, rather than arbitrarily augmenting the model by some constant, the model was deemed sufficient for control design as is.

4 Control Algorithm Synthesis

Transfer function models are well suited to classical control design techniques such as root locus analysis and frequency response analysis. A control scheme that meets the performance specifications can be synthesized utilizing one or both of these design methods.

A controller is typically required to command the system so that it performs in a certain manner. Determining whether or not the system behaves in the desired manner is, like most design problems, judged by performance metrics. Each of the above design methods has its own set of performance metrics [14, 58-59]. The information provided by both sets of performance metrics are correlated, and the designer is usually wise to use both methods [58].

4.1 Time-Domain Analysis and the Root Locus

4.1.1 Time-Domain Specifications

Time-domain specifications are typically embodied by the system step response, and can apply to either the transient or steady-state portion of the step response. An archetypal step response is shown in Figure 4.1 to illustrate these design parameters. Those specifications that apply to the transient portion of the response are explained in Table 4.1.
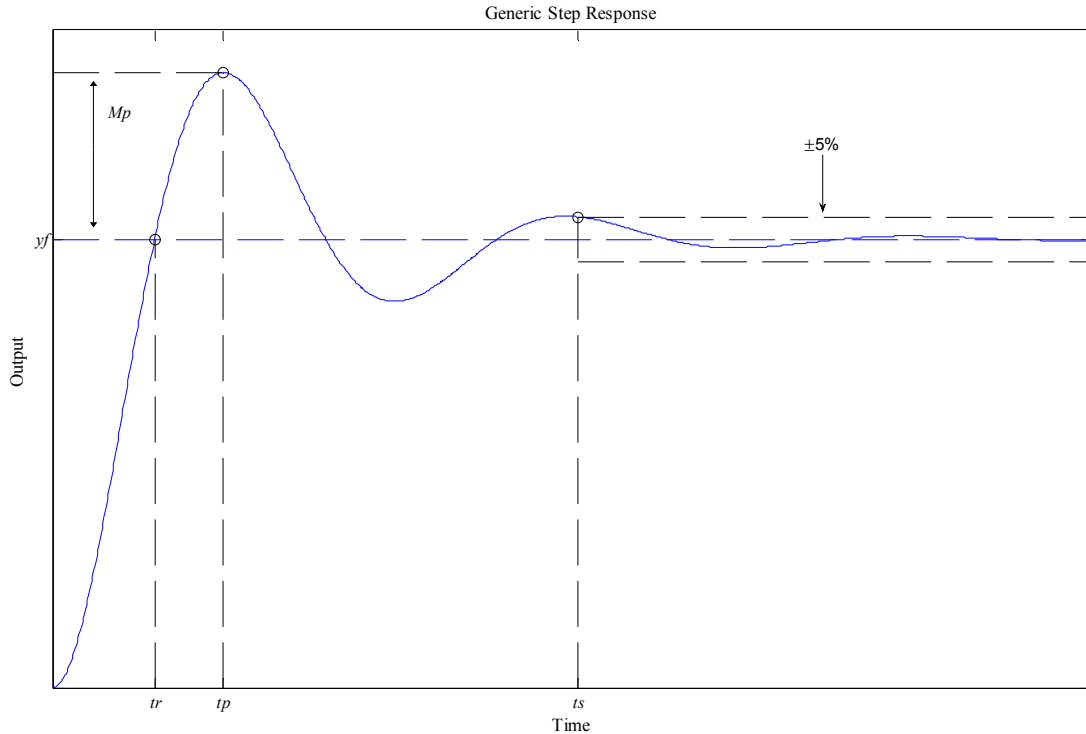
Figure 4.1   Generic step response showing transient time-domain specifications.

Table 4.1   Explanation of time-domain specifications.

Symbol	Description	Explanation
t_r	rise time	Time required to reach vicinity of final value. From 10% to 90% and 0 to 100% are common
M_p	Overshoot	Maximum peak value as measured from the final value. Usually expressed as a percentage.
t_p	peak time	Time at which max. overshoot is reached.
t_s	settling time	Time required for response to reach and stay within a boundary about the final value. Usually expressed as a percentage . 2% and 5% (shown above) are common tolerances.

Controller design begins with defining the specifications that need to be met for the particular application and ends with verifying the in-service performance of the chosen design.   Some typical examples of time-domain design specification statements are:

- Ensure less than 10% overshoot with reasonable settling time.

- Transients must settle in less than 5 seconds with minimal amount of overshoot.
- Set point must be reached in at most 2 seconds with less than 5 percent overshoot

Notice that the first two examples make one quantitative statement and one qualitative statement, while the last example makes only quantitative statements. The specifications can be given in either manner. However, in the latter case, it may happen that the performance objectives may never be met, especially if they are too stringent. This is due to conflicting specifications, e.g. overshoot and rise time are inversely proportional and cannot simultaneously decrease. The result is that the specifications may have to be altered during the design phase [58].

4.1.2 Root Locus Design Method

Consider the typical feedback system shown in Figure 1.6 (b). The plant is given by *G(s)*, and the controller is given by *C(s)*. The closed-loop transfer function from the reference input, *r*, to the output, *y*, is given by Equation (4.1). Setting the denominator of Equation (4.1) equal to zero yields the characteristic equation of the system shown in Equation (4.2), the solutions of which are known as the closed loop poles. Closed-loop pole locations strongly influence the transient response of the system.

$$\frac{y}{r} = \frac{Y(s)}{R(s)} = \frac{C(s)G(s)}{1+C(s)G(s)} \qquad (4.1)$$

$$1+C(s)G(s) = 0 \qquad (4.2)$$

76

Pole locations vary as parameters of the controller transfer function are varied. Root locus analysis seeks to track the path that the poles follow as a controller parameter is varied. To begin, assume that the plant *G(s)* is a ratio of polynomials and unalterable, and that the controller, or compensator, *C(s)* is a pure gain *K* and is variable. As *K* is varied from 0 to infinity, or negative infinity, the closed-loop poles migrate from the open-loop poles to the zeros [58]. Angles of arrival at the zeros and angles of departure from the poles are determined from relative angles between the pole in question and other zeros and poles of the system [58].

Now suppose that *C(s)* still has the same variable gain *K*, but also contains a rational polynomial of poles and zeros, whose locations are also variable. The gain is still varied as before, but the closed-loop poles may follow much different paths depending on the location of the pole and zero additions from the compensator transfer function. The compensator is the component at the control engineer's discretion. By altering the compensator transfer function, the closed-loop pole locations are altered. This changes the transient response as well. The idea behind the root locus design method is to determine compensator pole and/or zero locations that drive the closed-loop poles to some desired location or region in the *s*-plane that will yield a satisfactory time-domain response [58].

The locations or regions in the *s*-plane to which the poles and zeros are to be driven are determined by the control design specifications. Thus, the term 'satisfactory response' is a relative one. Generally speaking, it is usually desired that the rise and settling times be small, and that the overshoot be 20% or less [15]. Alternatively, it is desirable that the system have a damping ratio between 0.4 and

0.8. Systems with damping lower than this range may be sufficient in terms rise and settling times, but exhibit excessive overshoot. Having a damping ratio above this range may yield the opposite problem [58].

4.2 Frequency-Domain Analysis

4.2.1 Frequency-Domain Specifications

Frequency-domain specifications known as gain margin (GM) and phase margin (PM) can be read from the open-loop Bode plot. When the magnitude is expressed in decibels, the gain margin is read from the open-loop bode diagram as the difference between the 0 dB line (a gain of unity) and the value of the magnitude, $|G(j\omega)|$, at the frequency where the phase passes through -180° (known as the phase crossover frequency). The phase margin is the difference between the phase angle $\angle G(j\omega)$ and -180° at the frequency where the magnitude passes through 0 dB from above(known as the gain crossover frequency) [14-16, 58]. A typical bode plot is shown in Figure 4.2 to provide a graphical interpretation of these specifications.
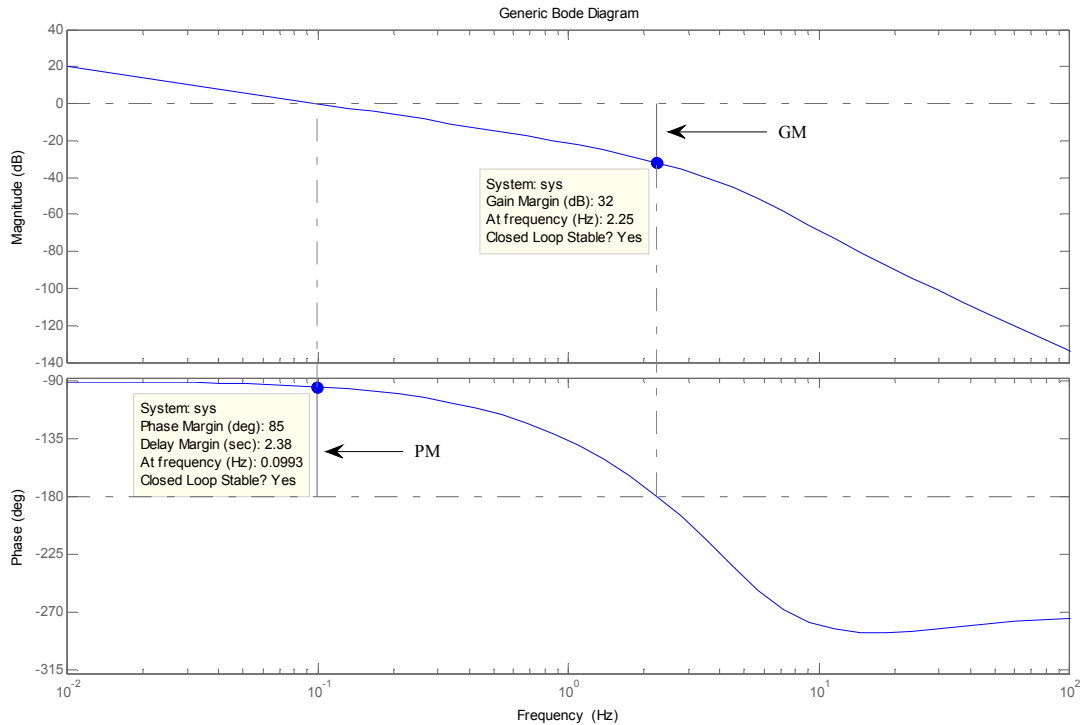
Figure 4.2   Generic Bode plot to illustrate gain margin and phase margin.

More general definitions are that gain margin is a measure of how much the gain can be increased before instability results, and phase margin is a measure of how much negative phase (phase lag) can be added before the onset of instability. Effectively, these margins are measures of system stability [15].

The bode diagram in Figure 4.2 is for an arbitrary system and is not related to the actuator to which this work pertains. It is only shown to graphically present the concepts of gain and phase margin. Note that in Figure 4.2 there is only one gain crossover frequency and one phase crossover frequency. For higher-order systems with multiple modes, it is possible to have multiple gain and phase crossover frequencies, in which case there may be more than one gain or phase margin. In this case, a decision must be made as to which gain or phase margin to use. In some cases, the most conservative estimate is used, i.e. the choice that makes the system

seem the least stable or most unstable. In other cases, the values at the highest crossover frequencies are used. At best the designer must be prudent with regard to this choice [14-15, 58]. Being prudent here involves considering the details of the application for the controller and the desired specifications.

4.2.2 Frequency Response Design Method

The underlying principle of the frequency response design method is finding a compensator that will result in a frequency response shape that meets the phase and gain margin design specifications while also satisfying any steady-state requirements that may exist. Frequency responses are typically presented in either a Nyquist plot, which is a polar plot of the magnitude and phase angle of the open-loop transfer function, or as a Bode diagram. The advantage of the Bode diagram is that for cascaded systems, the individual system Bode diagrams are simply added together to obtain the composite system Bode diagram. This means that, if necessary, a compensator can be designed from experimental frequency response data like that of Figure 3.3, without any knowledge of pole or zero locations of the plant [14].

Frequency-domain compensator design via the Bode diagram approach follows a common process with some minor variations that are dependent on the compensator structure being implemented. It begins by adjusting the open-loop gain as needed to satisfy any steady-state accuracy requirements. At this point, if gain and phase margin specifications are not met by proportional gain adjustment alone, then a compensator is then inserted which will reshape the open-loop system in the requisite manner [58].

Reshaping the loop is reliant on the desired performance and the limitations on reshaping the loop. Disturbance rejection and reference command tracking typically require large gains, while noise mitigation and input signal magnitude limitations usually necessitate small gains. Conflicting objectives often occur in different frequency ranges. A good compromise can usually be made with a compensator that has high gain in the low frequency region below the crossover frequency and low gain in the high frequency region. This is reflected in the closed-loop frequency response as an approximately horizontal asymptote below the crossover frequency and rapid roll-off, or attenuation, above the crossover frequency. Furthermore, the reshaped open-loop frequency response should have sufficient gain and phase margin as well. Typically, a gain margin of at least 6 dB and a phase margin of at least 30° are sufficient to ensure stability and robustness [15].

4.3 Controller Design for the Actuator

The point has been made that designing a controller begins with defining specifications that a controller must satisfy. Performance specifications are typically defined with a particular application in mind. Since there is not a particular application being considered, specifications are used only as metrics to explore the performance that can be obtained from the closed-loop system, rather than goals that must be met.

4.3.1 Open-Loop System

Coordinates, frequencies, and damping ratios (where applicable) of the poles and zeros for the actuator model of Equation (3.15) are given in Table 4.2. A pole/zero map is plotted in Figure 4.3. The pole/zero map shows two lightly damped second-order pole pairs and one second-order zero pair. These correspond to the two resonant peaks and the anti-resonant peak, respectively, in the Bode diagram shown in Figure 4.4. There is also a far-LHP zero contributing to the low-frequency gain and high-frequency roll-off. Steady-state gain, or DC gain, is -26.2 dB according the Bode diagram.

Table 4.2   Details of poles and zeros for continuous-time actuator model.

Description	Location (s-plane)	Frequency (Hz)	Damping Ratio
2nd Order Pole, 1st Resonance	$-226 \pm 15300i$	2435	0.015
2nd Order Pole, 2nd Resonance	$-550 \pm 52100i$	8292	0.011
2nd Order Zero, Anti-Resonance	$-60 \pm 24230i$	3856	N/A
Real Zero	-172070	172070	N/A
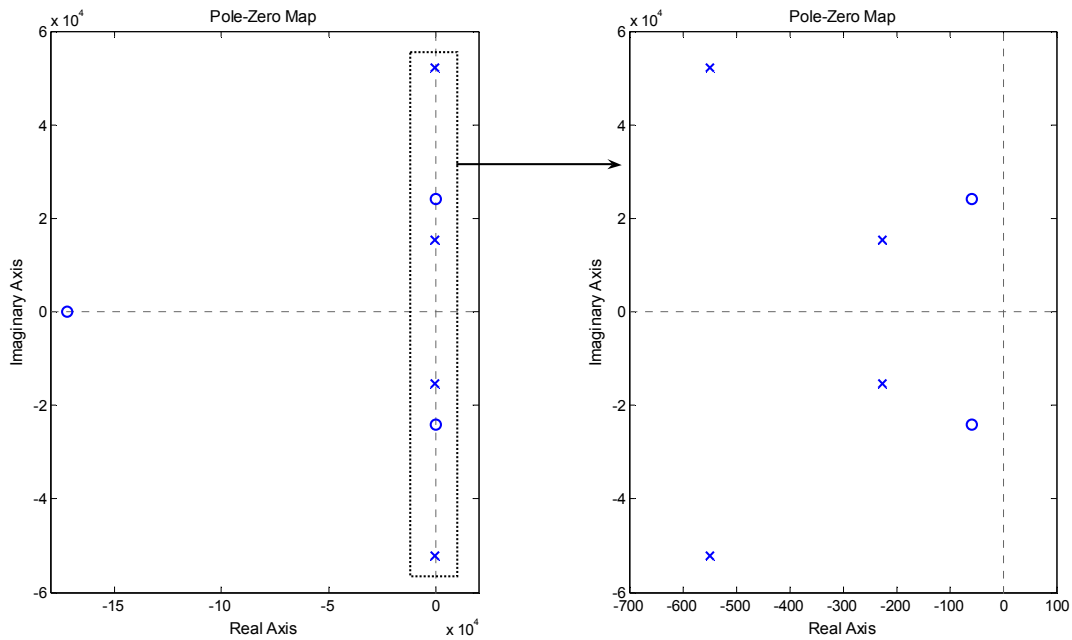
Figure 4.3   Pole/zero map for continuous-time actuator model.
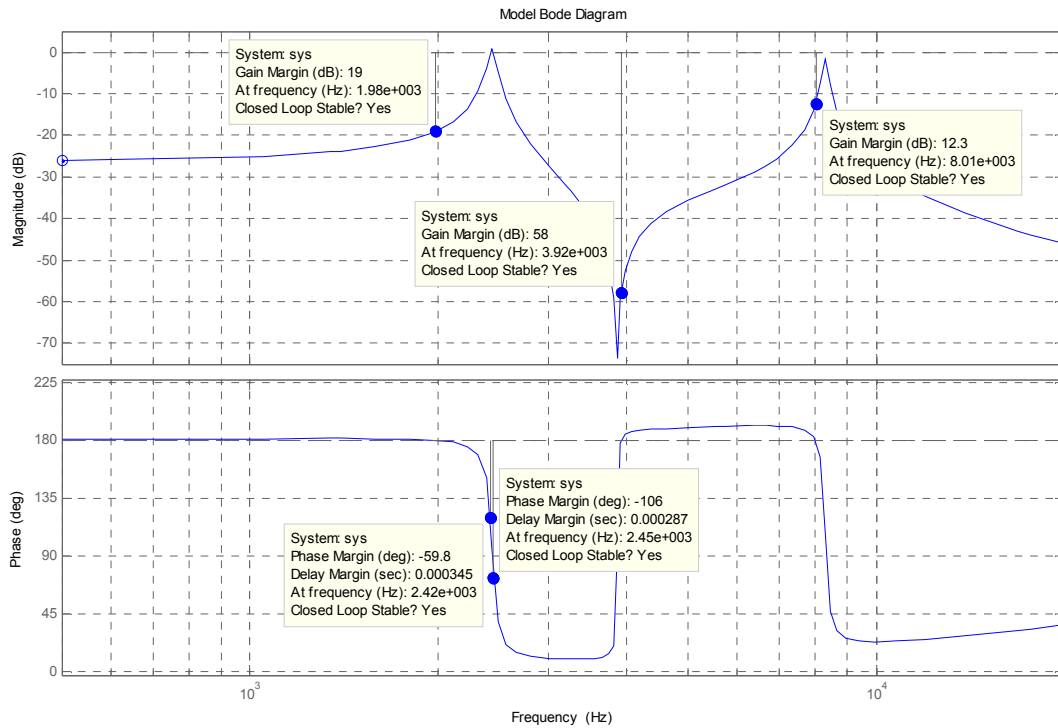
Figure 4.4   Bode diagram for continuous-time model.

4.3.2 Proportional Control: A Simple P Controller

The root locus of the pure-gain, closed-loop system is shown in Figure 4.5. This is the root locus as described in Section 4.1.2 where a pure gain varies from zero to infinity, referred to as a positive root locus. The poles quickly migrate into the right-half plane (RHP) with increasing gain. Poles in the RHP yield an unstable system [14-16].
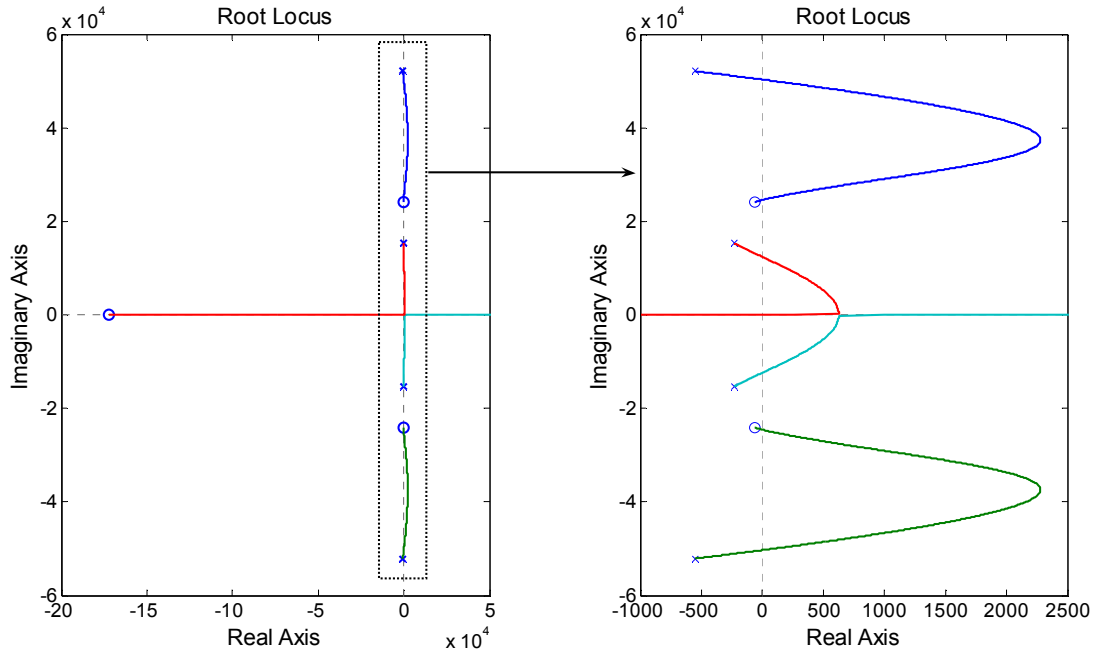
Figure 4.5   Positive root locus, gain varies from zero to infinity. Poles quickly migrate to RHP, i.e. it is easy for system to become unstable here.

The root locus in Figure 4.6 uses negative gain values, varying from zero to negative infinity. In this case the poles remain in the left-half plane (LHP) regardless of gain magnitude, resulting in a stable system. The corresponding Bode diagram is shown in Figure 4.7. Using negative gain shows no change to the magnitude plot, and only shifts the phase plot. The phase plot shape remains, but shifts so that the low-frequency asymptote is at 0°. The phase never passes through -180°, so the GM is infinite, which is intuitive since, in Figure 4.6, the poles never cross the imaginary axis. The magnitude crosses the 0 dB line from above at 2450 Hz, resulting in a 73.5° PM.
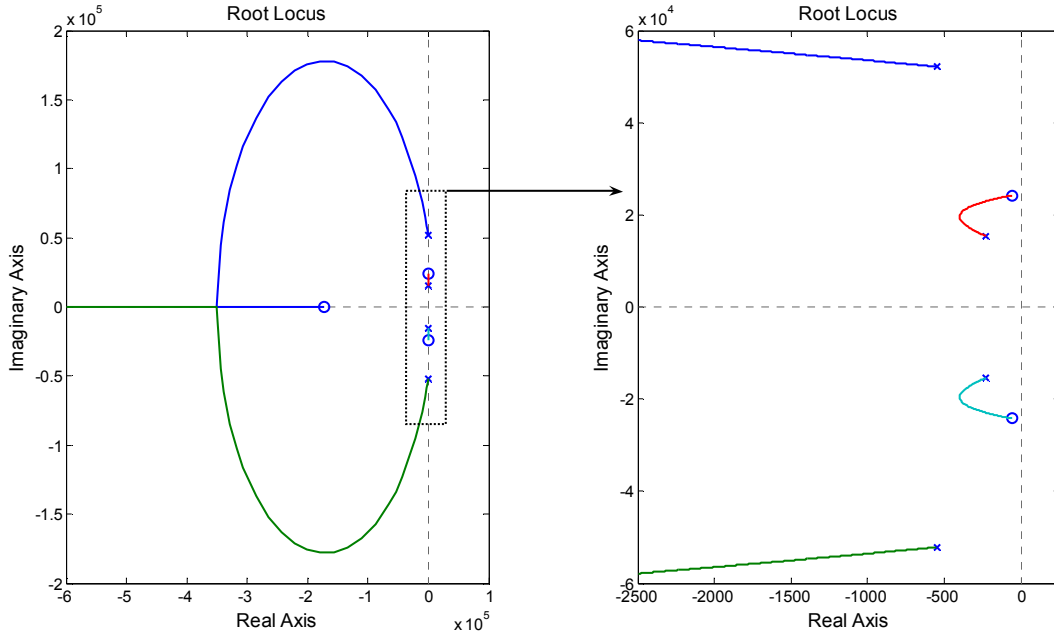
Figure 4.6 Negative root locus, gain varies from zero to negative infinity. Poles stay in LHP, so the system is always stable.
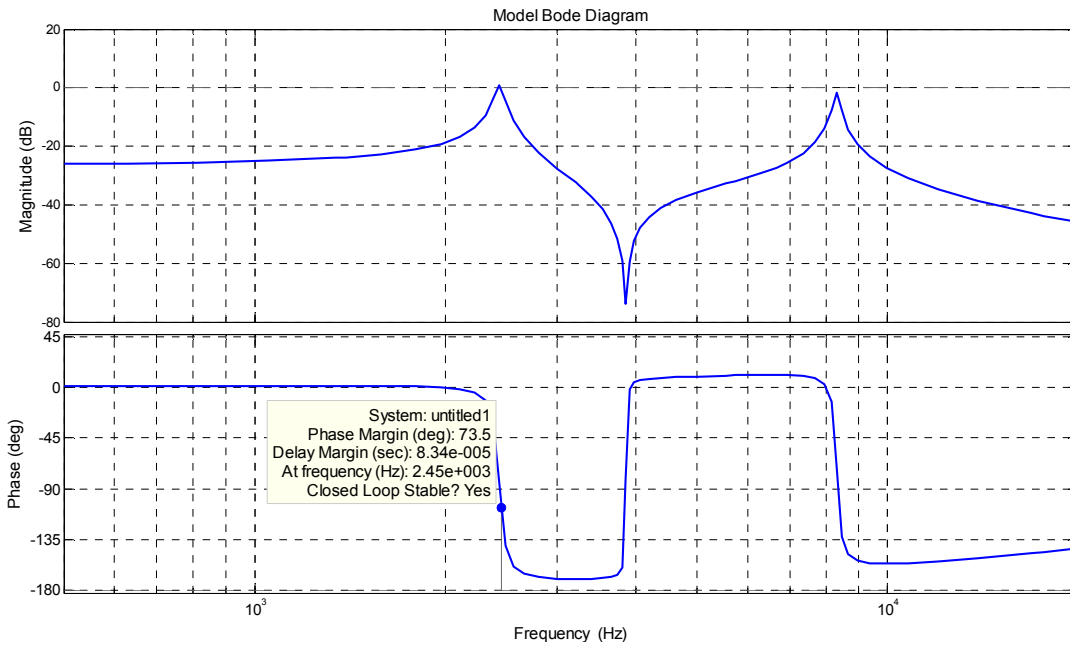
Figure 4.7 Bode diagram for negative-gain system. GM is infinite since the phase plot never crosses -180°. It crosses 0 dB from above once, yielding a 73.5° PM.

Recall the Bode diagram in Figure 4.4, corresponding to positive gains. The low-frequency asymptote is at 180°. This is to be expected after considering the

combination of the actuator and sensor dynamics explained in Section 2.6. Using a negative gain in effect reverses the sensor dynamics, making a positive input yield a positive output in the open-loop system. So it is clear that a negative gain should be implemented, and to this point it has been assumed that the negative gain is implemented in the controller *C(s)*. However, this has notable consequences regarding the closed-loop system. Rather than actually reversing the sensor coordinate system directly, the control output (which is also the plant input) direction is switched. This configuration admits negative input voltages in response the archetypal step input shown in Figure 4.8. Depending on the magnitude of the step input being tracked, the negative voltage limit noted in Section 2.1 could be exceeded, which is undesirable for the piezostack.
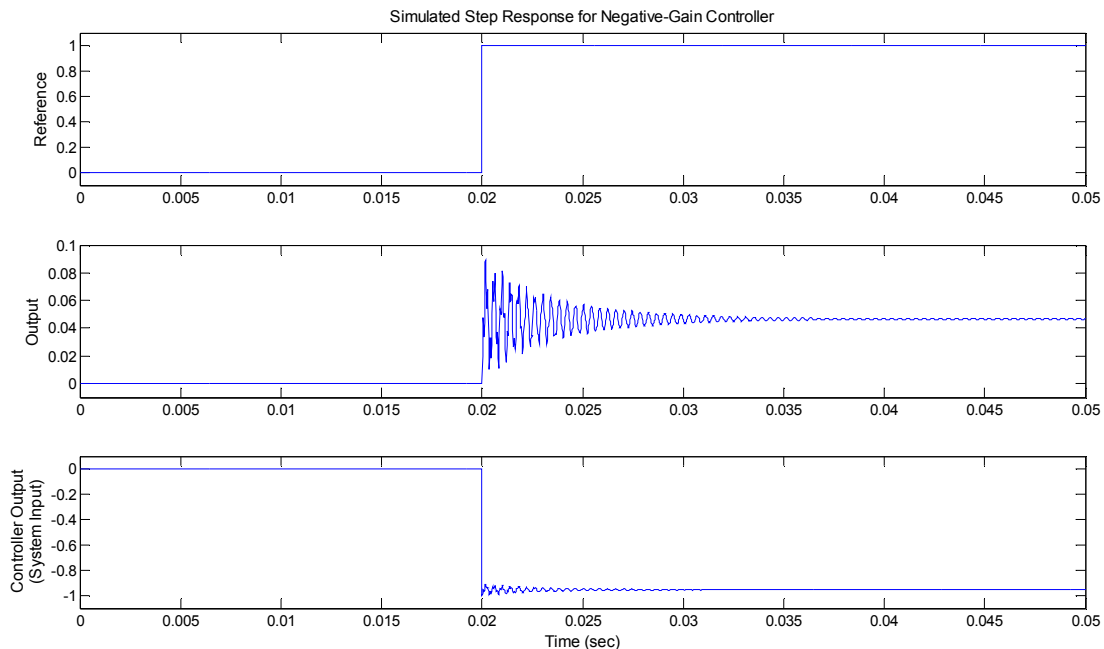
Figure 4.8 **Simulated step response. Reference, output, and input (from controller) signals. When negative gain is placed in the controller, negative inputs are commanded in response to a positive step (position) input.**

The problem that implementing a negative gain seeks to correct, the low-frequency asymptote at 180°, is an artifact of the sensor coordinate system rather than a property of the plant itself. Consider the feedback system in Figure 4.9 (a). This is the system in Figure 1.6 (b) with the addition of the transfer function *H(s)* in the feedback loop, typically referred to as sensor dynamics. Placing the negative gain in *H(s)* reverses the coordinate system without affecting the controller output (or system input). A positive step reference signal results in positive controller and measured output signals as shown in Figure 4.10.

(a)

C = controller
G = plant
H = sensor dynamics

(b)
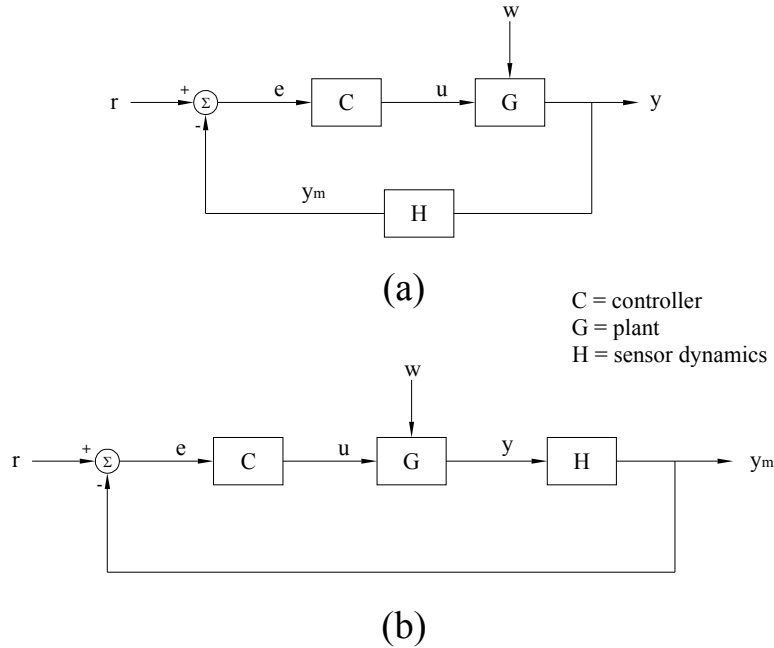
Figure 4.9   Closed-loop feedback system with sensor dynamics *H(s)* in (a) feedback loop or (b) forward loop. Output of *H(s)* is *y_m*, which is the measurement of *y*. Measured output and actual output may have differing coordinate systems.
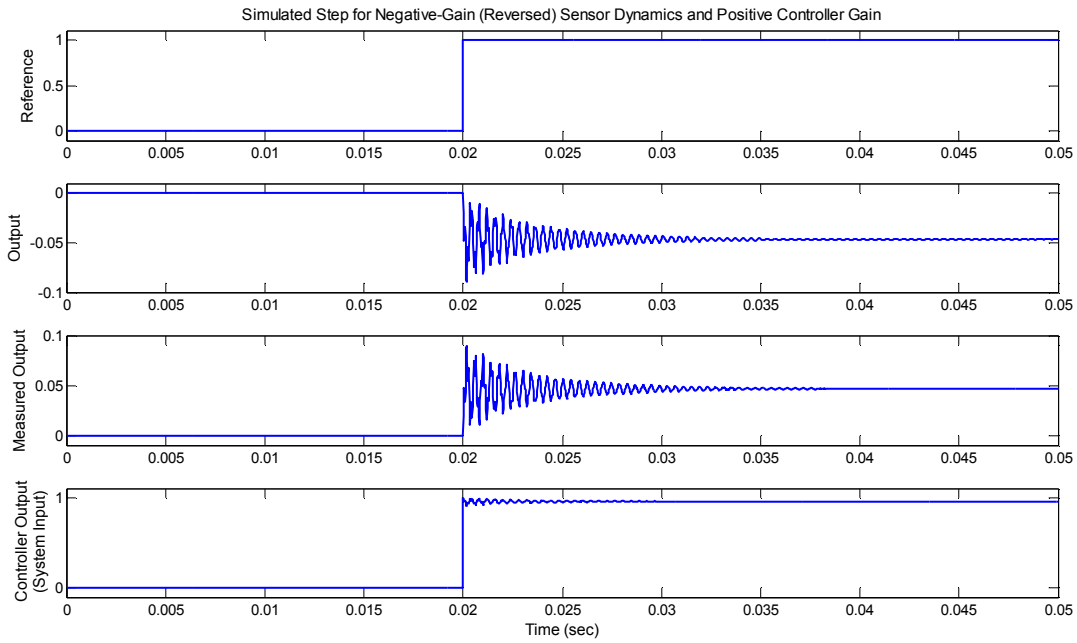
Figure 4.10   Simulated step response with negative gain added to sensor dynamics. Coordinate systems of reference, controller, and measured output are all in the same direction.

Both the actual and measured outputs, y and y_m, are plotted in Figure 4.10. They are now in opposing directions because of the sensor coordinate system reversal. This may seem problematic, but since it is actually the measured output y_m that is being fed back and observed by the user rather than the actual output y, it is of no concern. The local coordinate system of y could be made to match that of y_m if desired by placing the sensor dynamics $H(s)$ in the forward loop, directly after the plant $G(s)$ as in Figure 4.9 (b).

The effect of increasing gain on the closed loop system is illustrated by the closed-loop Bode diagrams with varying gain values shown in Figure 4.11. A proportional gain equal to one yields a closed-loop Bode diagram that is the same as the in the open-loop case. Increasing the gain value increases the low-frequency asymptote and the mode frequencies, and at the same time tends to suppress the modes, i.e. tends to round and flatten the sharp peaks. The effect of mode

suppression is present to a higher degree in the second mode. The close proximity of the anti-resonance and the first mode makes them difficult to suppress.
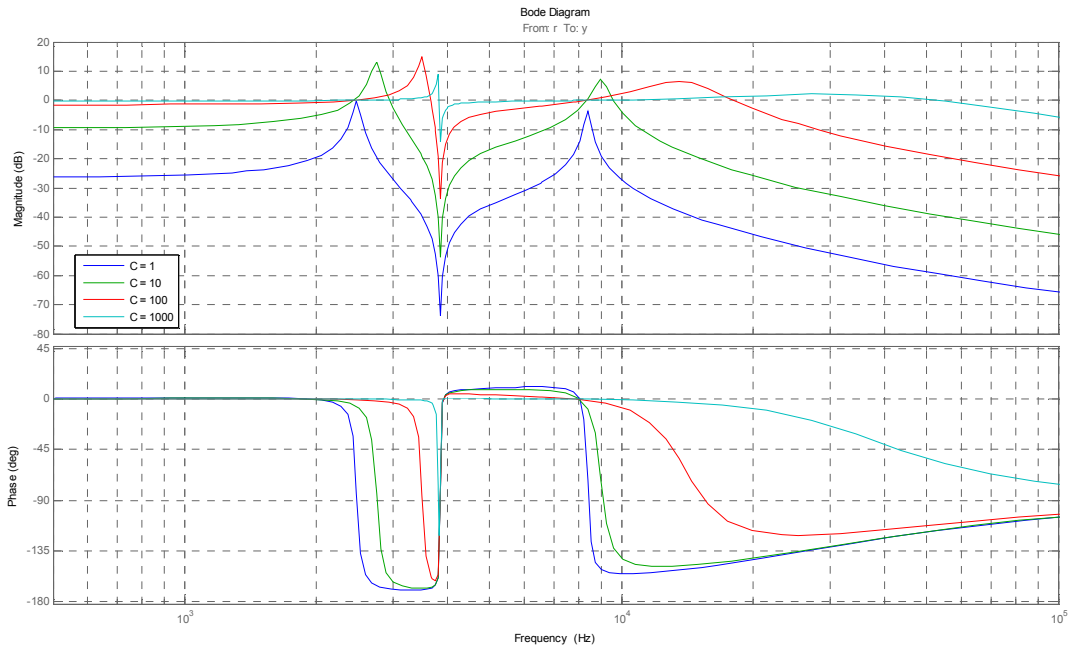
Figure 4.11   Closed-loop Bode diagram with pure proportional controller. Gain value begins at 1 and increases by a factor of 10 until the value reaches 1000.

$$e_{ss} = \lim_{s \to 0} \frac{s}{1 + L(s)} \frac{1}{s} = \lim_{s \to 0} \frac{1}{1 + L(s)} \tag{4.3}$$

Equation (4.3) gives the steady state error of an open-loop system *L(s)* to a unit step input [58]. It is clear from Equation (4.3) that the steady-state error to a unit step input is dependent on the behavior of *L(s)* as *s* approaches zero. This behavior is tied to the concept of system type, which is the number of poles at the origin, i.e. integrators, in *L(s)*. If the system is a rational polynomial with no poles at the origin,

L(0) will be finite, and thus e_{ss} will be finite. Having at least one integrator in the system will drive *L(0)* to infinity, which will drive e_{ss} to zero.

Each integrator in a system will add a -20 dB/dec slope to the Bode magnitude diagram. So the horizontal asymptote at low frequency in Figure 4.7 and the model of Equation (3.15) prove that the system has no inherent integrators. Furthermore, the horizontal asymptote at low frequency present in all plots of Figure 4.11 shows that a pure gain does not add an integrator to the system.

While the steady-state error will approach zero with increasing gain, it will never be equal to zero without adding an integrator to the system. Gain values that yield potentially acceptable steady-state errors result in control input signals far outside the acceptable input range for the device, e.g. a gain of 5000 produces a steady state error of 0.4 %, but creates a maximum control input signal of 5000 V. While this may seem a miniscule amount of error, the title of this work implies that even miniscule error may be unacceptable. A compensator transfer function must therefore be added to the system for better performance.

4.3.3 PI Controller

Compensator transfer functions come in many forms, the most common of which were discussed in Section 1.2.2. The PID controller scheme (or any of its variants: PI, PD, etc.) is attractive because it allows for the addition of a pure integral (I) term, rather than approximation to a pure integral term as in a lag compensator. This will correct the steady-state error issue discussed in Section 4.3.2. The response

was sufficiently fast (settling times <1 millisecond) and well-damped (<20% overshoot with K_p=1000) using pure proportional control. So in keeping with the idea of trying the simpler options first, the integral term was added, yielding a PI controller. Closed-loop performance was then reassessed.

$$C(s) = K_P + \frac{K_I}{s} = \frac{K_P s + K_I}{s} = K \frac{s + z}{s} \tag{4.4}$$

PI controller transfer functions can take any of the forms shown in Equation (4.4). There are two parameters to tune no matter the format chosen. Both parameters will have an effect on the dynamic performance of the system. Tuning multiple parameters in an attempt to meet multiple, and potentially conflicting design constraints can be difficult without an objective method to follow.

$$J = \int_0^\infty |e(\tau)|^2 d\tau$$
$$e(t) = y_{ref}(t) - y(t) \tag{4.5}$$

All of the time-domain performance metrics are expressed relative to the reference input. Acceptable performance metric values are typically aimed at minimizing the error, e.g. less overshoot or smaller settling times decrease the error. Quantifying the metrics into terms error signal is thus intuitive. The integral squared error (ISE) is given in Equation (4.5), which may be recognized as the error signal 2-norm squared. The 2-norm or the integral square error can be minimized with respect to the control parameters. This often results in a reasonable compromise between performance objectives [15]. The general applicability and ease of implementation of PID controllers has resulted in several tuning rules being developed for these

controllers. Tuning parameters to minimize the ISE or other error-integral criterion often gives superior performance as compared to other tuning rules such as Ziegler-Nichols [60-61].

$$C(s) = 3030.85\left(\frac{s + 3.30 \times 10^{-5}}{s}\right) = \frac{3030.85s + 0.10}{s} \tag{4.6}$$

$$L(s) = \frac{931790.5(s + 1.72 \times 10^5)(s + 3.30 \times 10^{-5})(s^2 + 119.1s + 5.87 \times 10^8)}{s(s^2 + 452.2s + 2.34 \times 10^8)(s^2 + 1099s + 2.72 \times 10^9)} \tag{4.7}$$

$$T(s) = \frac{931790.5(s + 1.72 \times 10^5)(s + 3.30 \times 10^{-5})(s^2 + 119.1s + 5.87 \times 10^8)}{(s + 7.0 \times 10^5)(s + 2.32 \times 10^5)(s + 3.28 \times 10^{-5})(s^2 + 147.5s + 5.82 \times 10^8)} \tag{4.8}$$

Tuning a PI controller for minimal ISE gives the controller transfer function in Equation (4.6). Applying this controller to the system model with the sensor coordinate system reversal yields the open and closed-loop transfer functions, from the reference input to the measured output, given by Equations (4.7) and (4.8) respectively. The root-locus shape will be the same, with the controller contributing a pole at the origin and a zero near the origin.

The open-loop Bode diagram of the PI compensated system is shown in Figure 4.12. Since the root locus shape is basically the same, all poles remain the LHP. Gain margin is thus infinite. Phase margin is 47.7 degrees at 3840 Hz. This shows that typical gain and phase margin requirements of greater than 6 dB and 30° to 60°, respectively, are met by the PI controller.
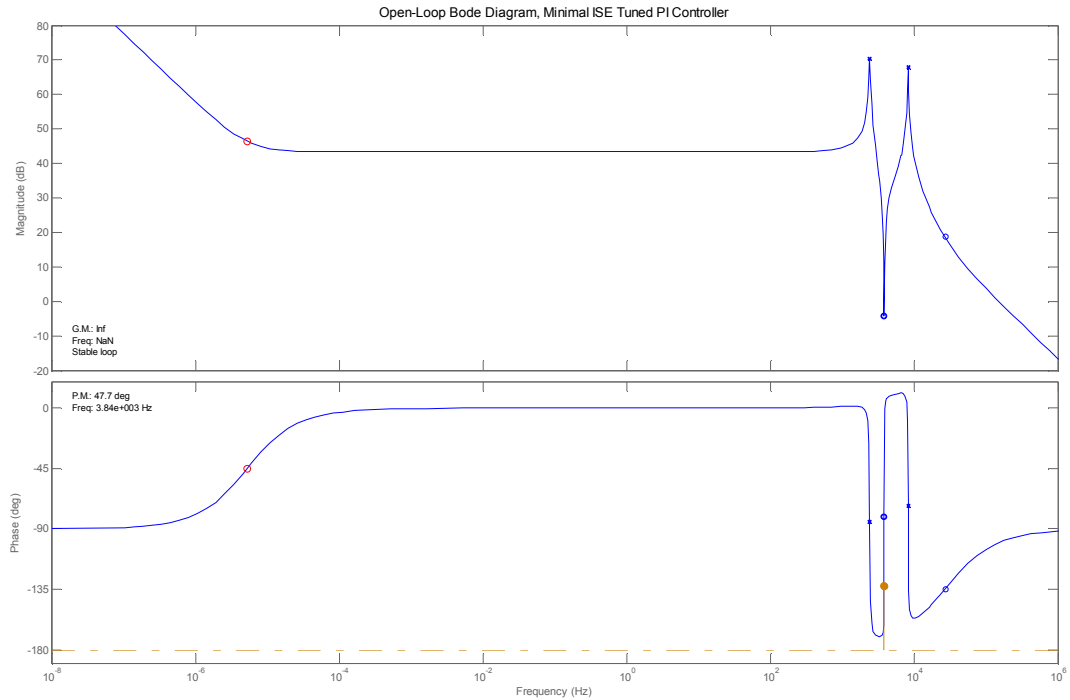
Figure 4.12   Open-loop Bode diagram for PI compensated system. Gain margin is shown to be infinite. Phase margin is 47.7 degrees at 3840 Hz.

The closed-loop Bode diagram is shown in Figure 4.13. With the exception of a dipole at 3900 Hz, the shape fits what is generally desirable in a closed-loop Bode plot The dipole is a result of the first mode and the complex conjugate zero pair being in close proximity. Increasing the controller gain drives them even closer together. The only way to eliminate the dipole is to drive the pole pair all the way to the zero pair. This requires infinite gain and is thus impractical. It can be better mitigated with higher gains, but this is also impractical due to the saturable input limits of the actuator.
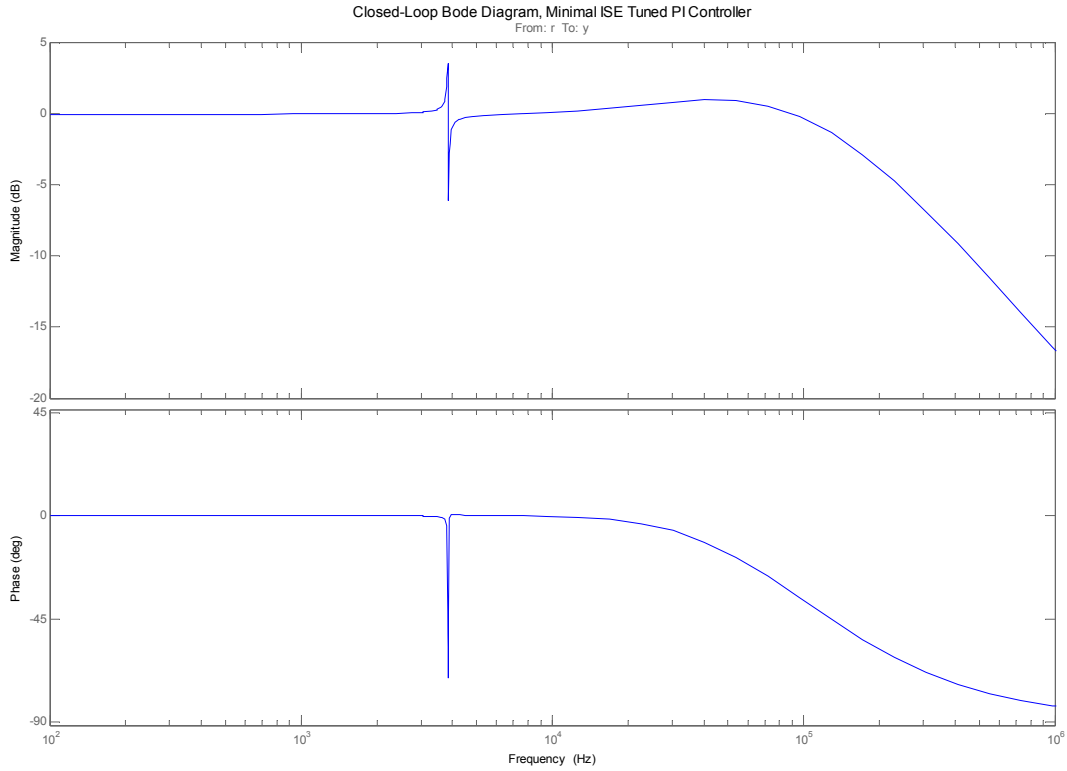
Figure 4.13   Closed-loop Bode diagram for PI compensated system.

A closed-loop, unit step response is shown in Figure 4.14.   Performance as reflected by the position output is excellent.   The peak response is 1.102 µm, making the overshoot 10.2%.   The settling time for a two percent bound is 0.012 milliseconds. A final value of 1 µm shows that adding the integral has eliminated the steady-state error.
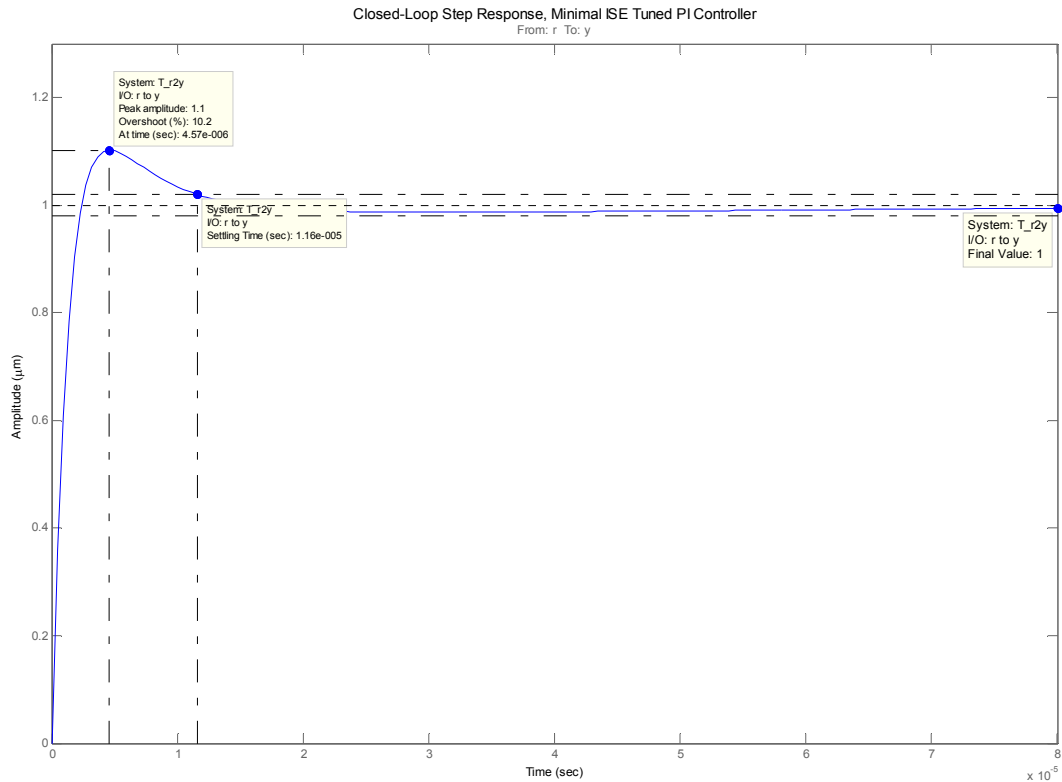
**Figure 4.14   Closed-loop step response of PI compensated system. Overshoot is 10.2%.
Settling time is approximately 0.012 milliseconds. No steady-state error is present.**

The controller output (which is the plant input) that commands the step response of Figure 4.14 is shown in Figure 4.15. At 3000 V, the peak response is well above the 500 V upper limit of the operating range. Furthermore, it is well above the 1250 V maximum input level per the manufacturer specifications. This implies that the performance illustrated in Figure 4.14 is unachievable. Saturation blocks are put into place to enforce the -50 to 500 V operating range. The controller output would thus be clipped at for any amount of time that it is commanding a signal level greater 500 V. Steady-state error will be unaffected, so long as the position set point requires less than 500 V to maintain. The transient specifications will, however, be affected. Rise time and settling time will likely be longer, since a lower input level will evoke a slower response. The effect on overshoot is difficult to predict, but it is

95

likely that it will be lowered as well, because the controller algorithm would have more time to respond due to the slower rise time.
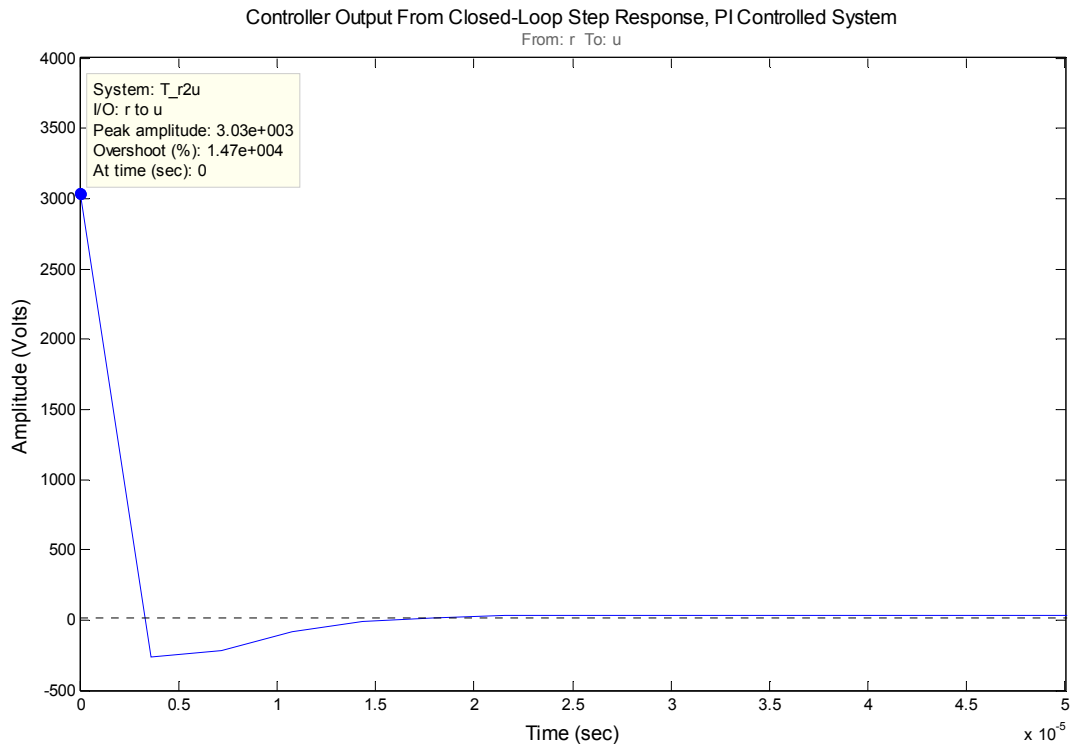
Figure 4.15   Closed-loop controller output (plant input) for step response. Maximum controller out is approximately 3000 V. This is well above actuator input level.

It is clear that while a minimal ISE tune PI controller yields excellent closed-loop performance, that performance may be pragmatically unachievable due to the input levels that it commands. This input level, however, depends upon magnitude of the reference tracking signal. The position output is measured in micrometers, so a unit step response corresponds to a 1 μm reference signal. The input saturation problem may not be an issue for this controller at sub-micrometric reference signal magnitudes. Furthermore, input saturation may also be avoided by using a linear ramp or sinusoidal path when asking for a set point change, since square-edged

reference inputs are impossible for any real system to follow. Therefore, the ISE tuned PI controller was not discounted as a viable option, but addition of the derivative term was considered to further improve closed-loop performance.

4.3.4 PI-D Controller

Derivative (D) control action is proportional to the rate of change of a signal, rather than the signal itself. A large rate of change in a signal is detectable before a large change in the signal magnitude. Large rates of change therefore predict large signal changes. Derivative control action then anticipates the associated overshoot that follows and initiates corrective action to prevent it. In effect, this adds damping and stability to the closed-loop system [58]. Since the damping is added to the closed loop system, it will be reflected everywhere. Adding derivative action thus has the potential to address the input level issue seen with the PI controller.

A disadvantage to adding the D term in the typical PID fashion arises when the reference tracking input contains square edges, i.e. the step input. When differentiating square edged signals, an impulse occurs in the control input due to the rapid change in the error signal. This is called the set-point kick phenomenon. Placing the D term in the feedback loop as in Figure 4.16 adds the desirable qualities of derivative action while avoiding set-point kick.

C1 = controller 1 = PI controller
C2 = controller 2 = D controller
G = plant
H = sensor dynamics
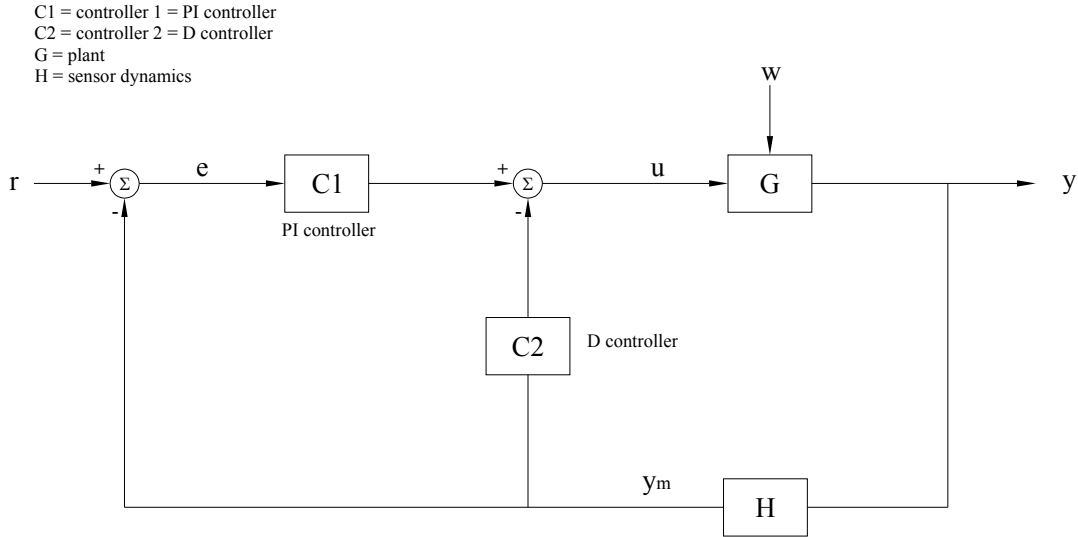
Figure 4.16   PI-D controller block diagram. Derivative action is in feedback loop so that the feedback signal is differentiated, not the error signal. This is desirable because error signal may contain impulses with square-edged reference signals

The overall controller with the addition of the D term was not tuned by an optimal method as with the PI controller. The minimal ISE tuned PI controller was left as it was. The derivative gain was increased until the controller output during a unit step response was brought into the input operating range of the actuator, i.e. less than 500 V.

$$C_1(s) = 3030.85\left(\frac{s + 3.30 \times 10^{-5}}{s}\right) = \frac{3030.85s + 0.10}{s} \tag{4.9}$$

$$C_2(s) = 0.0165s \tag{4.10}$$

$$L(s) = \frac{153430(s + 1.73 \times 10^5)(s + 3.30 \times 10^{-5})(s^2 + 119.1s + 5.87 \times 10^8)}{s(s + 1.41 \times 10^5)(s + 1255)(s^2 + 1576s + 5.92 \times 10^8)} \tag{4.11}$$

$$T(s) = \frac{153430(s + 1.73 \times 10^5)(s + 3.30 \times 10^{-5})(s^2 + 119.1s + 5.87 \times 10^8)}{(s + 3.28 \times 10^{-5})(s^2 + 171.1s + 5.83 \times 10^8)(s^2 + 2.97 \times 10^5 s + 2.68 \times 10^{10})} \tag{4.12}$$

98

The resulting controllers $C_1(s)$ and $C_2(s)$ are given by Equations (4.9) and (4.10) respectively. Note that these correspond directly to C_1 and C_2 in Figure 4.16. The open and closed-loop transfer functions are given in Equations (4.11) and (4.12) respectively.

The open-loop Bode diagram of the PI-D compensated system is shown in Figure 4.17. Gain margin is infinite because the poles never cross the imaginary axis into the RHP. Phase margin is 48.8° at 3840 Hz. The closed-loop Bode diagram for the PI-D compensated system is shown in Figure 4.18. The shape is similar to the PI-compensated case. Low-frequency behavior is unchanged up to the dipole, which is still present to virtually the same degree. High-frequency roll-off begins at a lower frequency than with the PI-compensated system, but occurs at the same roll-off rate.
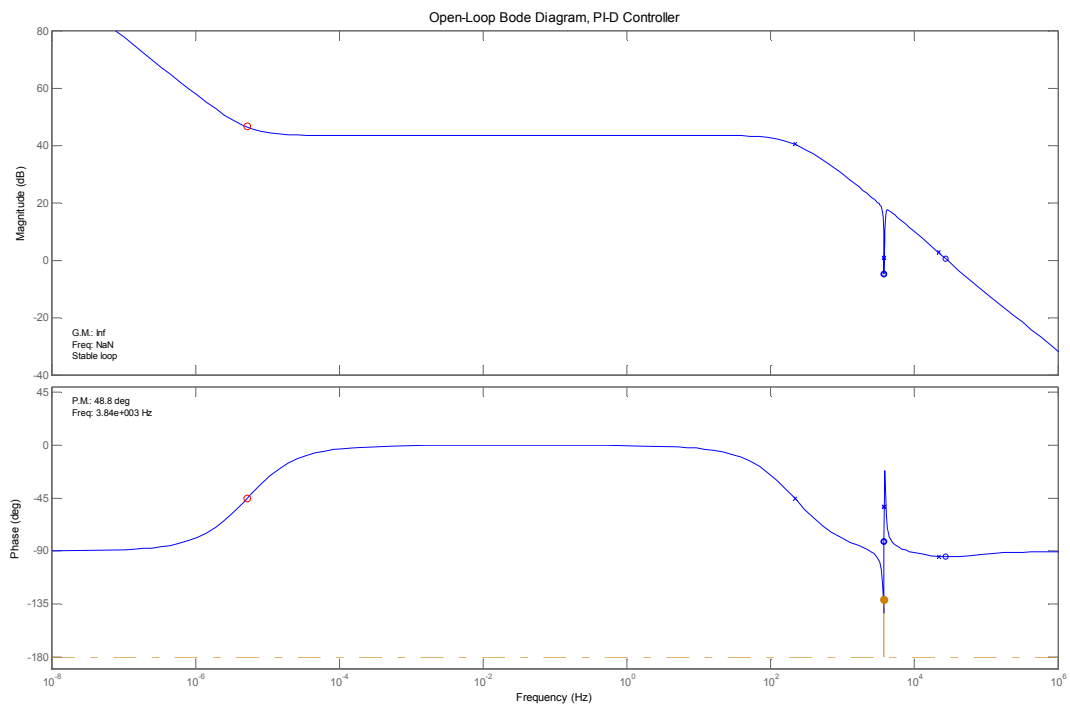

Figure 4.17   Open-loop Bode diagram of PI-D controlled system. Gain margin is shown to be infinite. Phase margin is shown to be 48.8 degrees at 3840 Hz.
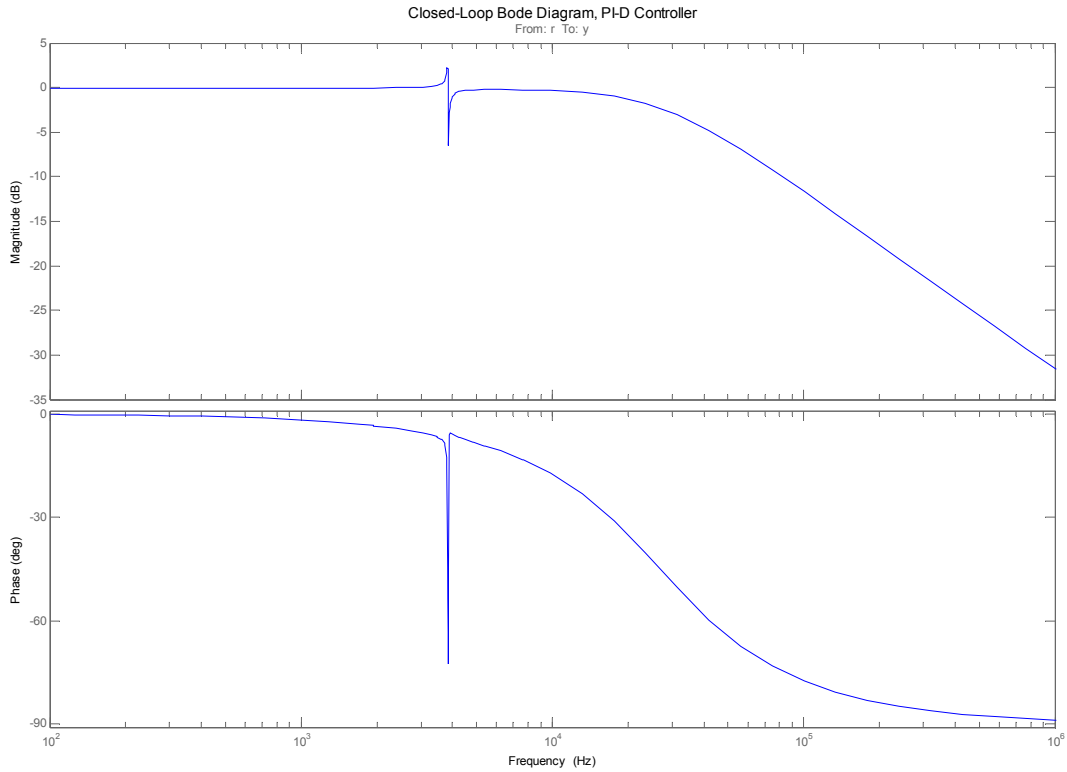
Figure 4.18   Closed-loop Bode diagram for PI-D compensated system.

The major differences between the PI and PI-D compensated systems are evident in the time-domain. The closed-loop unit step response is shown in Figure 4.19. Virtually no overshoot is present in the step response. Settling time (with 2% bounds) is 0.02 milliseconds.   This is approximately 67% longer than the PI-compensated system. Integral control action is still present, so there is zero steady-state error.
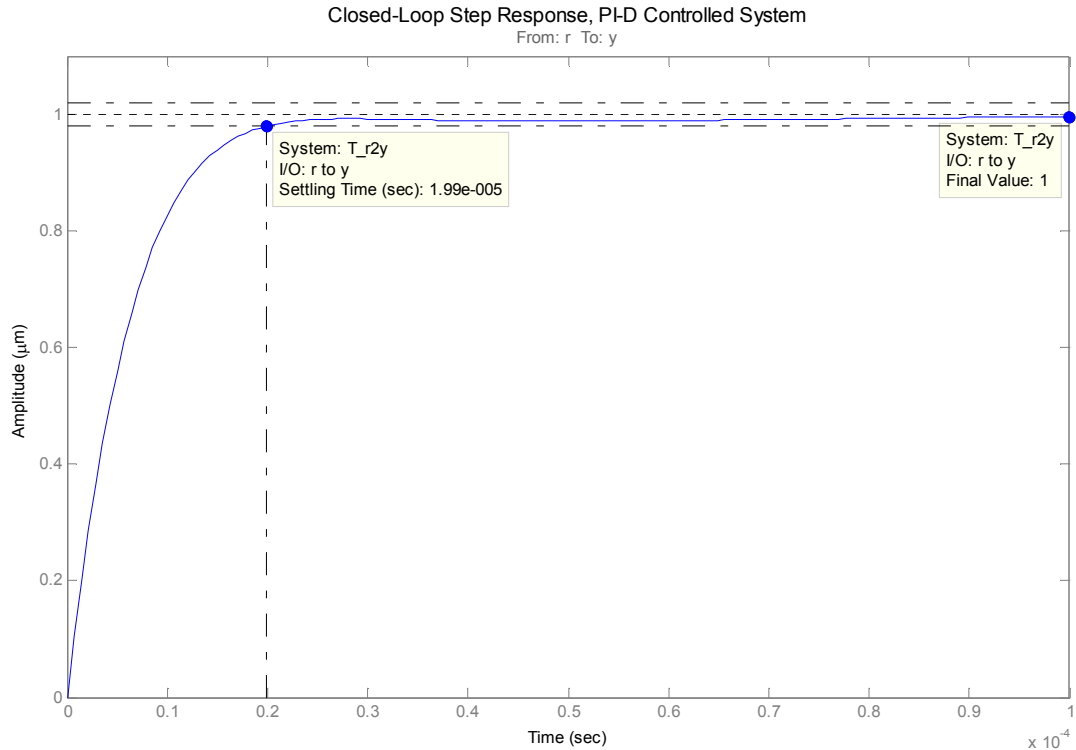
Figure 4.19 Closed-loop step response of PI-D controlled system. Virtually no overshoot is present. Settling time is approximately 0.02 milliseconds. No steady-state error is present.

The controller output that commands the response in Figure 4.19 is shown in

Figure 4.20. Adding derivative control has brought the peak controller output down

into the operating range of the physical actuator. Just as with the PI controller, this is

a unit step response corresponding to a 1 μm step. A larger step would likely evoke a

higher peak in the controller output signal.
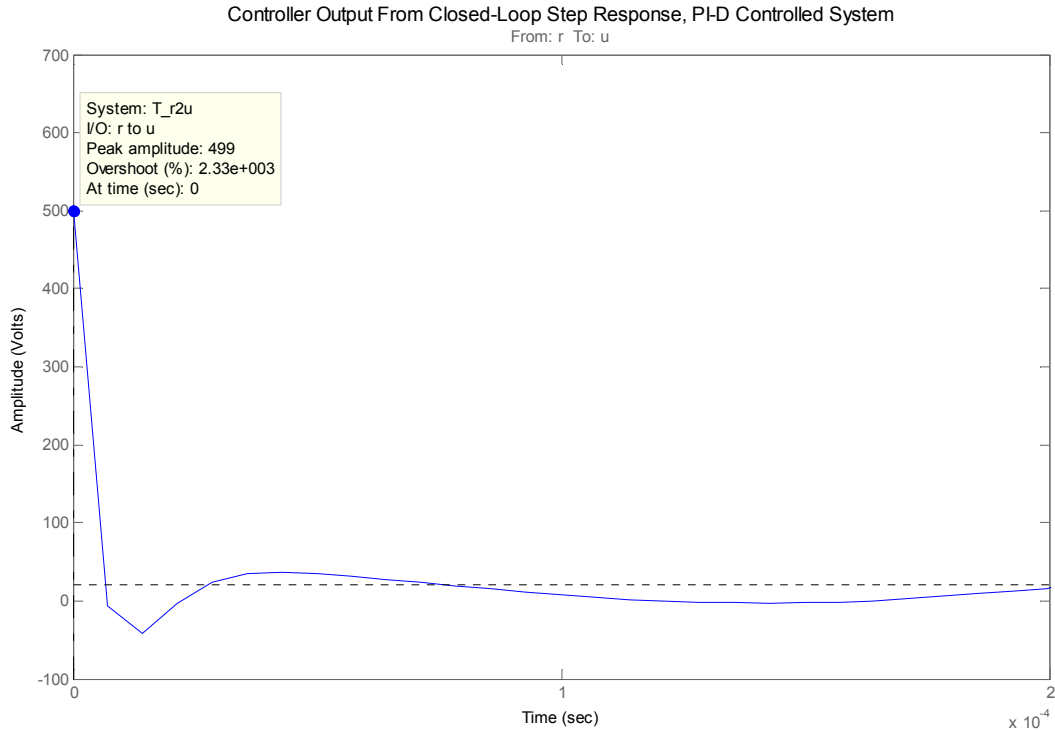
Figure 4.20   Controller output associated with closed-loop step response of PI-D compensated system. Maximum output is within actuator input range.

4.4 Actuator Control Design Conclusions

Many different controllers can be designed to meet the same set of specifications, and exploring all of them would be laborious. Two controllers that yield generally acceptable performance (based on the actuator model) have been presented. What remains is to apply these controllers to the physical system in order to test their ability to command the actuator's working surface to track a particular reference trajectory (step, rounded step, sine wave, etc.). It is then wise to briefly reconsider the controller designs and their predicted performance before taking these controllers online.

The high proportional gain ($K_p \approx 3030$) in the PI controller of Equations (4.6) and (4.9) were noted (recall that Equations (4.6) and (4.9) are the same transfer function, but appear in both of the control schemes discussed). The largest error contribution due to a step response occurs during the span of time it takes to reach the vicinity of the new set point. So it is intuitive that the ISE method commands large proportional gains in an attempt to snap the system to the vicinity of set point as quickly as possible.

This high proportional gain can potentially yield in an oscillatory or unstable system [61-62]. Even though the simulated responses did not show this problem, it could surface when the controller is applied to the real system, where noise and disturbances occur, and where unmodeled dynamics may be present.

5 Experimental Validation

Online, closed-loop experiments with real hardware in the loop present issues that may not have been addressed during modeling and simulated control design process, and can result in lower performance than predicted by simulation with said model. A control system should thus be verified with actual hardware if it is to be of any real use. Doing so yields real-life limitations of the plant and any required hardware, as well as any deficiencies in the plant model or the controller algorithm. Online control testing reveals whether the plant, controller, or the required hardware warrant alteration to obtain a functioning system.

5.1 ISE-Tuned PI and PI-D Controllers

Both the PI and the PI-D controllers designed in the previous chapter were implemented digitally via the model shown in Figure 2.13. In the case of the PI controller, the second controller in the feedback was nullified by setting all of the gains to zero. The unit step response and its associated controller output signal for the PI controller are shown in Figure 5.1. Actual performance is drastically deviant from the predicted performance.

Sustained oscillations about both the zero point as well as the 1 μm set point are the most noticeable deviation from predicted performance. Control authority saturates at the lower bound of -50 V virtually the entire time. Though it does not

saturate at the upper bound, it still reaches relatively high amplitude of approximately 250 V.
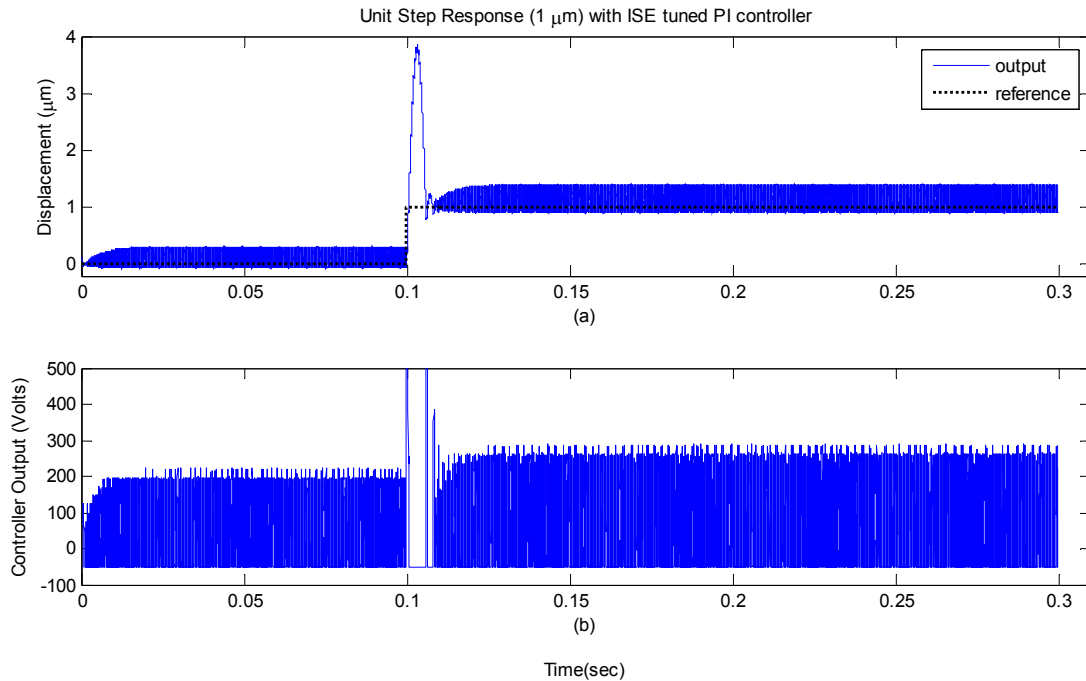
Figure 5.1 **(a) Unit step response with ISE-tuned PI controller and (b) associated controller output signal. There is a sustained oscillation present around set points (even at zero displacement) due to the excessive proportional control gain. This high proportional gain results in high control action, or chatter. Controller output signal saturates at the lower bound only.**

A 10 μm step response and it commanding controller signal are shown in Figure 5.2. The same oscillatory phenomenon in the unit step response of Figure 5.1 is present here as well, in both the displacement and controller signal. In this case the controller completely saturates at both the upper and lower bounds when trying to command the actuator to stay at the 10 μm set point.
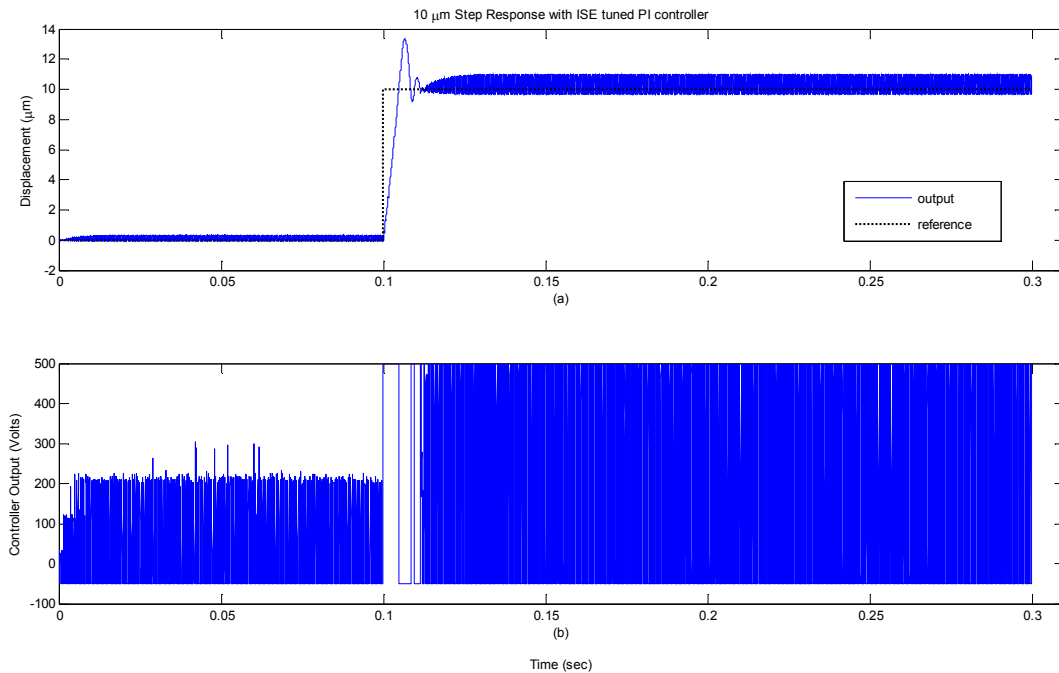
Figure 5.2 (a) Ten-μm step response with ISE-tuned PI controller and (b) associated controller output signal. The oscillatory phenomenon present in the unit step response of Figure 5.1 is seen here as well. Controller completely saturates.

The same responses were examined with PI-D controller. They showed the same oscillatory issue seen with the PI controller. Since the PI-D controller contains the same PI portion as the PI controller alone, and derivative control action was only added to suppress overshoot, this result should not be surprising. Unfortunately, the high proportional gain causes the PI portion of the PI-D controller to dominate the response. Showing the experimental results of the PI-D controller would be fruitless, since the responses are virtually identical to that of the PI controller alone.

This type of behavior is indicative of a marginal stability, with high control action attempting to stabilize the system. Furthermore, that this was not predicted in simulation warrants review of the model from which the controllers were designed.

5.2 Model Design Revisited

The oscillatory response exhibited by the previous controller indicates that there is a complex pole pair approaching the imaginary axis and threatening to cross into the unstable RHP. However, all root loci in Chapter 4 indicate that this is not the case. Further opposition is evident in all Bode diagrams showing infinite gain margin, which is evidence that the gain can be increased indefinitely without inducing instability. Nevertheless, the experimental results indicate otherwise. So there must then be additional dynamics in the true system that have not been included in the actuator model presented in Chapter 3, the model around which the controllers have been designed.

The actuator was identified using chirp signals of amplitudes ranging from 3 to 9 V. This is within the output range of the DAC board, and since every electrical/electronic component in the system will add noise, the system identification experiments were performed without the high-voltage amplifier in the loop. Actuator modes are approximately 2.4 and 8.3 kHz. Recall that the small signal bandwidth of the amplifier is 35 kHz, and that the input range of the actuator is well below the amplifier's output capability. Thus it was assumed that since the bandwidth of the amplifier was much higher than that of the actuator, that the amplifier would not add significant dynamics to the system. This was, however, realized as an incorrect assumption after further consideration of both the type of load associated with piezoelectrics and the current-limited nature of the amplifier.

Applying a mechanical stress causes charge to build up on the surfaces of a piezoelectric material. Conversely, a charge to build up due to applying voltage

across the surfaces of a piezo results in a mechanical deformation [1, 10]. In either case, charge build up is due to capacitance. Current drawn by a capacitive load is proportional to the rate of change in voltage according to

$$i(t) = C_t \frac{dV(t)}{dt} \qquad (5.1)$$

in which C_t is the total capacitance, including the load and internal capacitance [64]. In the case of trying to follow a particular signal shape (as in the case of an amplifier), the rate of voltage change represents the slew rate of the amplifier. Therefore, Equation (5.1) implies that the slew rate, and thus the bandwidth, is inversely proportional to total capacitance if the available current is finite.

The capacitance of the piezoelectric element of the actuator used here is approximately 84 nF. This amount of capacitive load is enormous compared to the amplifier's internal capacitance of 50 pF [42], and drastically affects the bandwidth of the amplifier. Since the actuator was modeled without the amplifier, the amplifier was thus modeled without the actuator. A capacitor of near equivalent capacitance (~90 nF) was put in place of the actuator to simulate the capacitive load.

The resulting input/output frequency response of the Trek 609E-6 amplifier is shown in Figure 5.3 along with a moderately damped, second-order approximation. The experimental response is nearly flat in the low frequency region. Note the low-frequency gain is 0 dB. The gain is normalized with a gain block placed in the Simulink model. This is done simply so that amplifier output is equal to the controller output. The response begins to roll off at approximately 4 kHz. The -3 dB crossing (i.e. the bandwidth) is approximately 5.9 kHz. A second-order transfer

function with DC gain of 0 dB, a natural frequency of 5.9 kHz, and a damping ratio of 0.7 is given by Equation (5.2). The actual frequency response and the second-order approximation are nearly coincident up to approximately 10 kHz, which encompasses the range of interest for the actuator.

$$A(s) = \frac{1.374 \times 10^9}{s^2 + 51900s + 1.374 \times 10^9} \qquad (5.2)$$
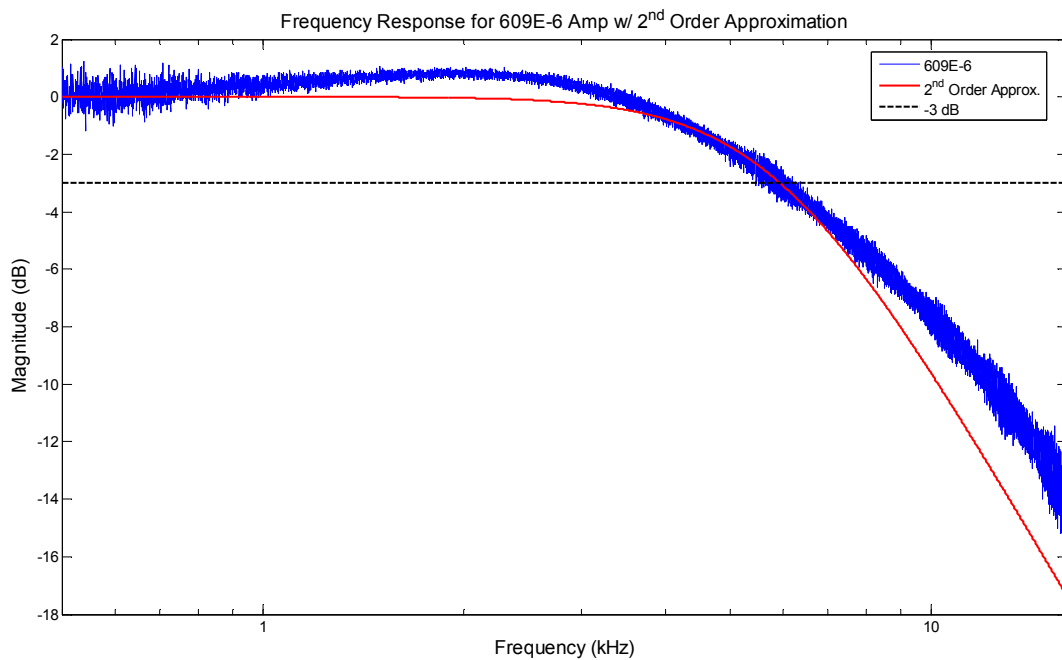
Figure 5.3 **Experimental frequency response of Trek 609E-6 amplifier compared to a second order approximation. A 90 nF capacitor is put in place of the actuator to simulate the capacitive load. This is done in order to maintain consistency with how the actuator was characterized, i.e. without the amplifier. 2nd order approximation parameters are ζ=0.7 and ω_n=5.9 kHz.**

Figure 5.3 indicates that due to the capacitive nature of the actuator, the amplifier adds significant dynamics to the system that were in not captured during identification of the actuator itself. However, it also indicates that the amplifier can

be approximated by the second-order system of Equation (5.2) in the frequency range of interest for the actuator.

5.2.1 Complete System Model

Merging the amplifier model of Equation (5.2), the actuator model of Equation (3.15), and the coordinate system reversal gain discussed in Section 4.3.2 yields the complete, open-loop system model given by

$$A(s)G(s)H(s) = L(s) = \frac{\left(4.2\times10^{11}s^3 + 7.3\times10^{16}s^2 + 2.6\times10^{20}s + 4.3\times10^{25}\right)}{\begin{pmatrix} s^6 + 5.3\times10^4 s^5 + 4.4\times10^9 s^4 + 1.6\times10^{14} s^3 + \ldots \\ \ldots 4.8\times10^{18}s^2 + 3.5\times10^{22}s + 8.8\times10^{11}s^3 \end{pmatrix}} \quad (5.3)$$

The pole and zero locations for this system model are shown graphically in Figure 5.4 and numerically in Table 5.1.   Pole and zero contributions from the actuator remain unchanged and are shown in green, while amplifier contributions are shown in blue.
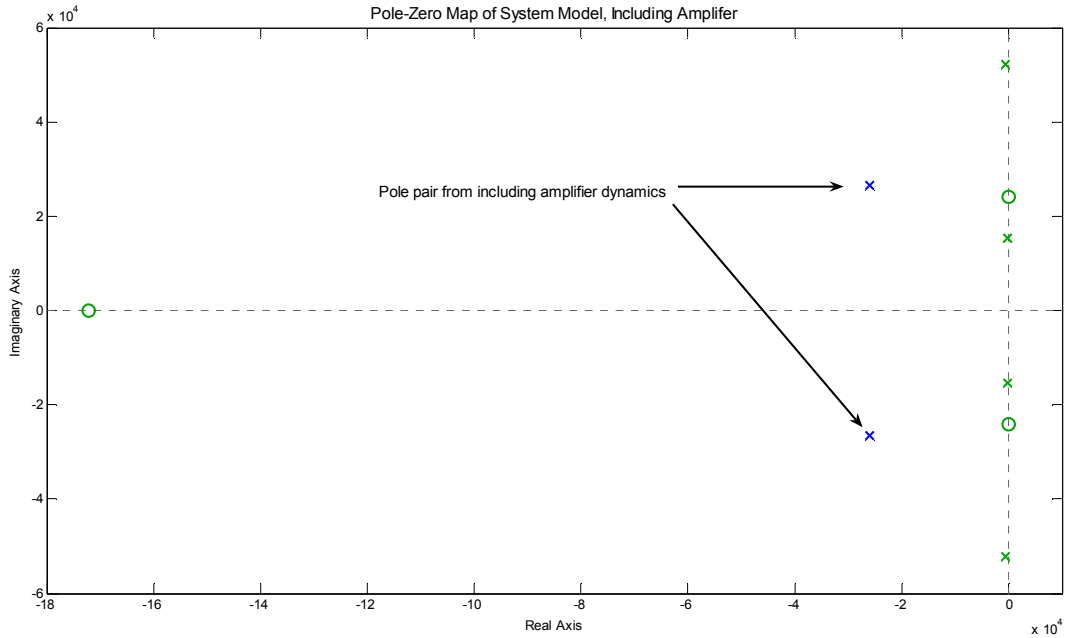
Figure 5.4 **Pole-zero map for open-loop system of Equation (5.3). This includes the actuator model, the sensor coordinate system reversal gain, and the amplifier 2nd order approximation.**

Table 5.1 **Pole and zero details for the complete open-loop system.**

Description	Location (s-plane)			Frequency (Hz)	Damping Ratio
2nd Order Pole	-226	$\pm$	15300i	2435	0.015
2nd Order Pole	-550	$\pm$	52100i	8292	0.011
2nd Order Zero	-60	$\pm$	24230i	3856	N/A
Real Zero	-172070			172070	N/A
2nd Order Pole (Amplifier)	-25900	$\pm$	26500i	5900	0.7

The resulting Bode diagram is shown in Figure 5.5 along with the original Bode diagram of the actuator itself. The difference in the magnitude plots does not become apparent until approximately 4 kHz. This is intuitive because this is where the amplifier begins to roll off dramatically. Phase retains a similar but distorted shape compared to the original. The phase begins to lag sooner than in the original case, but it is by a relatively small amount and is likely due to the damping added by

the amplifier. The phase difference becomes more drastic after approximately 4 kHz due the amplifier roll off.
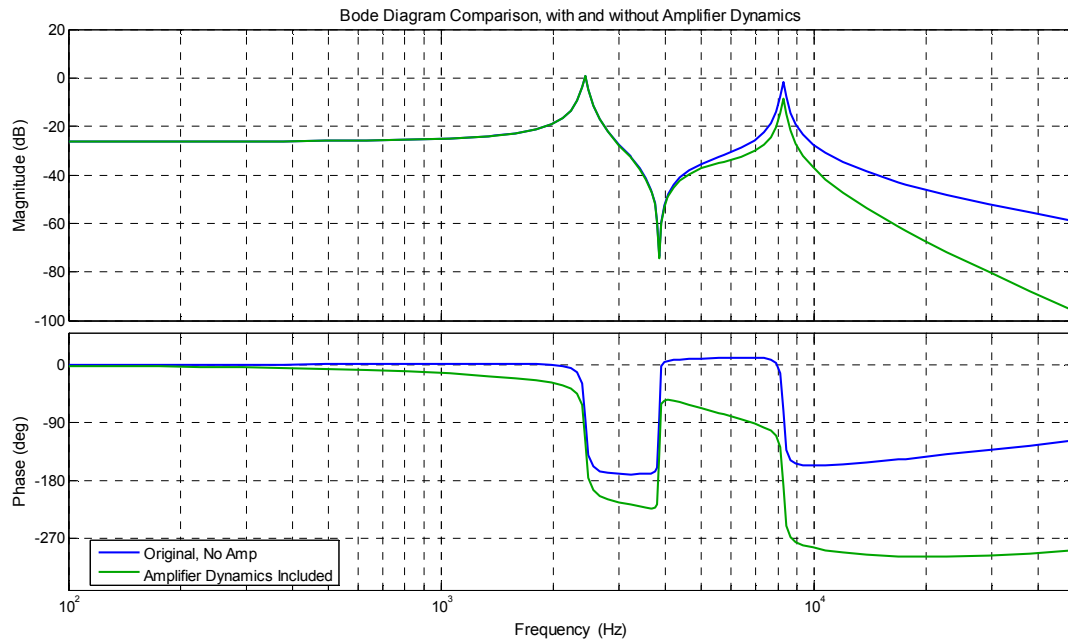
Figure 5.5   Bode diagram comparison between the original model and the complete system model including amplifier dynamics.

The addition of a complex pole pair is a significant change to the open-loop system model and warrants the redesign of the control algorithm that commands the system. All of the techniques (root-locus, Bode, etc.) discussed in Chapter 4 still apply, and the same control architecture can still be used, but the controller parameters, however, must be retuned.

5.3 Controller Redesign and Revalidation

The root-locus of the complete system with pure-gain control is shown in Figure 5.6. It shows that the addition of the pole pair from the amplifier alters the

path of the other two pole pairs from the original model, causing them to cross into the RHP. This was expected after considering the results of Section 5.1.
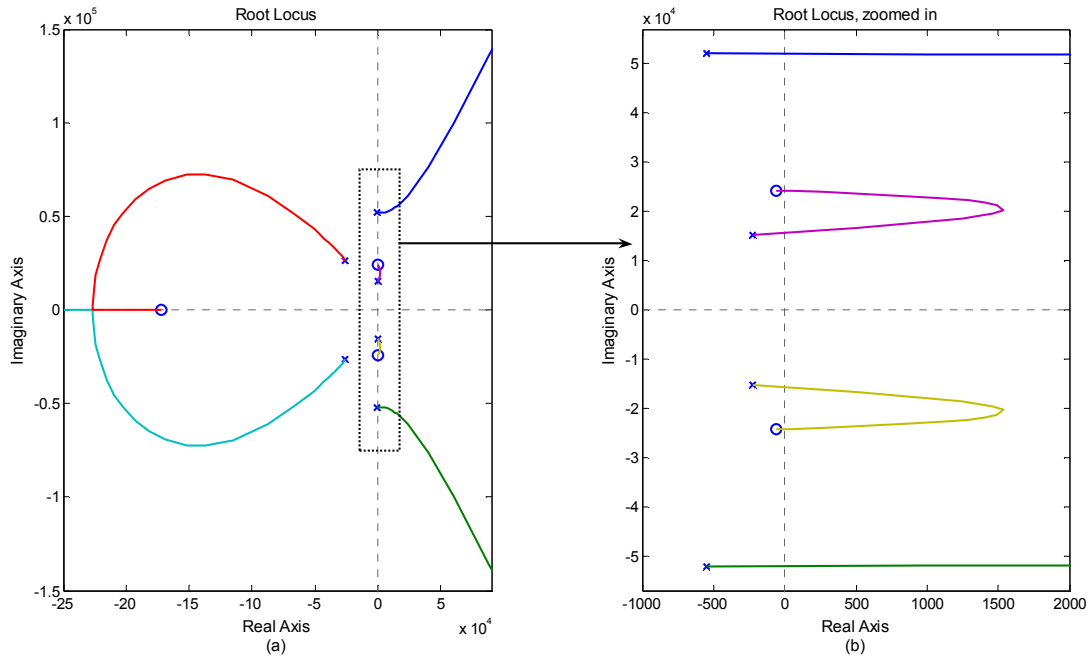
Figure 5.6 (a) Root-locus for complete system and (b) zoomed in to show imaginary axis crossing of original actuator model poles.

The PID control structure was kept. Retuning with the minimal ISE metric resulted in negative controller gains. As explained in Section 4.3.2, this is undesirable. Thus, alternative tuning methods had to be used. Ziegler-Nichols closed-loop tuning method lends itself well in situations where the system can be driven to instability. Originally developed for online tuning or instances where plant dynamics are not well known, the method can be applied to model or simulation-based control design when the user has good knowledge of the system.

Ziegler-Nichols (ZN) closed-loop tuning method is based on the critical gain and period associated with the marginal stability point. The process begins with the

113

system under pure proportional control. Proportional gain is increased until the system is marginally stable, i.e. the system exhibits sustained, steady oscillations. This gain value is known as the critical gain, K_{cr}, and is recorded along with the period of the sustained oscillations, known as the critical period, P_{cr}. A controller transfer function of the form of Equation (5.4) is assumed. Parameter settings as functions of the critical gain and period are shown in Table 5.2 according the terms to be included in the controller.

$$C(s) = K\left(1 + \frac{1}{T_i s} + T_d s\right) \tag{5.4}$$

Table 5.2   Ziegler-Nichols tuning rules based on critical gain and period [58].

Controller Type	K	T_i	T_d
P	$0.5K_{cr}$	∞	0
PI	$0.45K_{cr}$	$0.83P_{cr}$	0
PID	$0.6K_{cr}$	$0.5P_{cr}$	$0.125P_{cr}$

When applying the ZN method to a model, marginal stability occurs at the first instance where a pole (or complex conjugate pole pair) crosses the imaginary axis on the root locus. A closer look at Figure 5.6 (b) is shown in Figure 5.7, which shows the imaginary axis crossing of both unstable pole pairs. Of the two pole pairs that cross the imaginary axis, the first to do so is the lower frequency pole pair that migrates to the system zero. This occurs at a gain of 1.92. The other pole pair does not cross into the RHP until the gain reaches 2.68, and proceeds to migrate to infinity (and negative infinity). So then, the system has a critical gain of 1.92. The frequency at this point is 2.5 kHz, yielding a critical period of 0.4 msec.

$$K_{cr} = 1.92$$

$$P_{cr} = \frac{1}{f_{cr}} = \frac{1}{2500} = 0.4 \quad m\sec \tag{5.5}$$
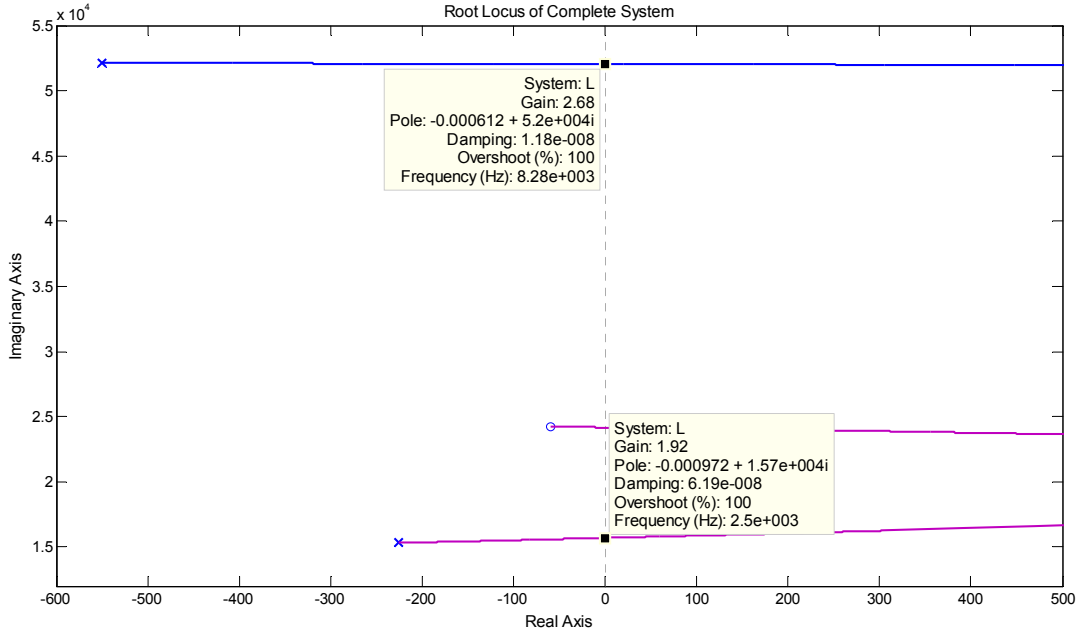
Figure 5.7 A closer look at the root locus shown in Figure 5.6 . Imaginary axis crossings for both unstable poles pairs are shown.

$$C(s) = 0.864 \left(1 + \frac{1}{3.46 \times 10^{-4} s} \right) = 0.864 + \frac{2894}{s} \tag{5.6}$$

$$L(s) = \frac{\left(\begin{array}{c} 3.65 \times 10^{11} s^4 + 6.41 \times 10^{16} s^3 + \ldots \\ \ldots 4.32 \times 10^{20} s^2 + 3.76 \times 10^{25} s + 1.24 \times 10^{29} \end{array} \right)}{\left(\begin{array}{c} s^7 + 5.35 \times 10^4 s^6 + 4.41 \times 10^9 s^5 + 1.57 \times 10^{14} s^4 + \ldots \\ \ldots 4.77 \times 10^{18} s^3 + 3.51 \times 10^{22} s^2 + 8.75 \times 10^{26} s \end{array} \right)} \tag{5.7}$$

$$T(s) = \frac{\left(\begin{array}{c} 3.65 \times 10^{11} s^4 + 6.41 \times 10^{16} s^3 + \ldots \\ \ldots 4.32 \times 10^{20} s^2 + 3.76 \times 10^{25} s + 1.24 \times 10^{29} \end{array} \right)}{\left(\begin{array}{c} s^7 + 5.35 \times 10^4 s^6 + 4.41 \times 10^9 s^5 + 1.57 \times 10^{14} s^4 + \ldots \\ \ldots 4.84 \times 10^{18} s^3 + 3.55 \times 10^{22} s^2 + 9.13 \times 10^{26} s + 1.24 \times 10^{29} \end{array} \right)} \tag{5.8}$$

115

Tuning a PI controller for this system according to the ZN tuning rules in Table 5.2 has the transfer function in Equation (5.6). The controller has low proportional gain and high integral gain. This result is not surprising considering the marginal stability point characteristics in Equation (5.5). The system has a high oscillation frequency and low tolerance for proportional gain. Corresponding open and closed-loop transfer functions are given in Equations (5.7) and (5.8) respectively.

Figure 5.8 shows the open-loop Bode diagram of the retuned system of Equation (5.7). Phase margin is 92° at 22.5 Hz. Gain margin is 4.1 dB at 2480 Hz. The typical phase margin requirement of greater than 30° is sufficiently met, but the typical gain margin requirement of at least 6 dB is not. The 6 dB gain margin requirement is a general one, and not concrete. The 4.1 dB gain margin and is 92° phase margin still yield a system that is closed-loop stable.
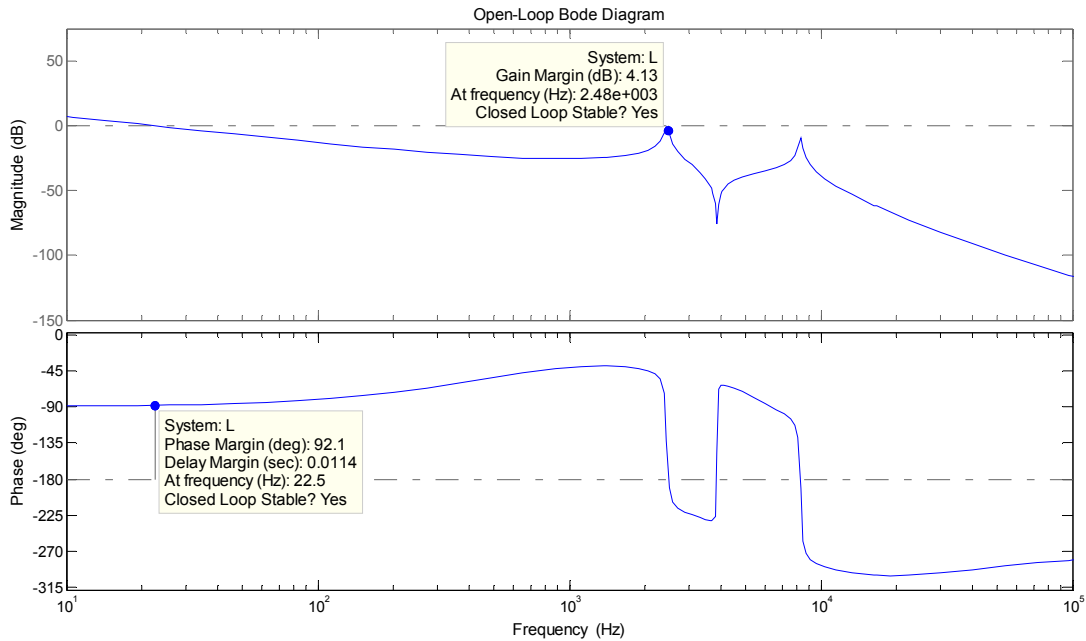
Figure 5.8   Bode Diagram of open-loop system with ZN-tuned PI controller, the system of Equation (5.7). A gain margin of 4.1 dB and a phase margin of 92° result in a system that will be closed-loop stable.

The closed-loop Bode diagram corresponding to Equation (5.8) is shown in Figure 5.9.   Bandwidth is 21.4 Hz.   This is a severe decrease in bandwidth as compared to original closed-loop system that did not include amplifier dynamics. Also, there is a small part of the first actuator mode that is still above 0 dB (a gain of greater 1).   The peak is at 8 dB at a frequency of 2.46 kHz, so the gain is small.   So, Oscillations during the transient portions of the step response could potentially exist, but will likely have small amplitude and die out in steady state.   Disturbances could excite this mode during the steady state portion of operation.   But again, the oscillations will likely have small amplitude and die out quickly.
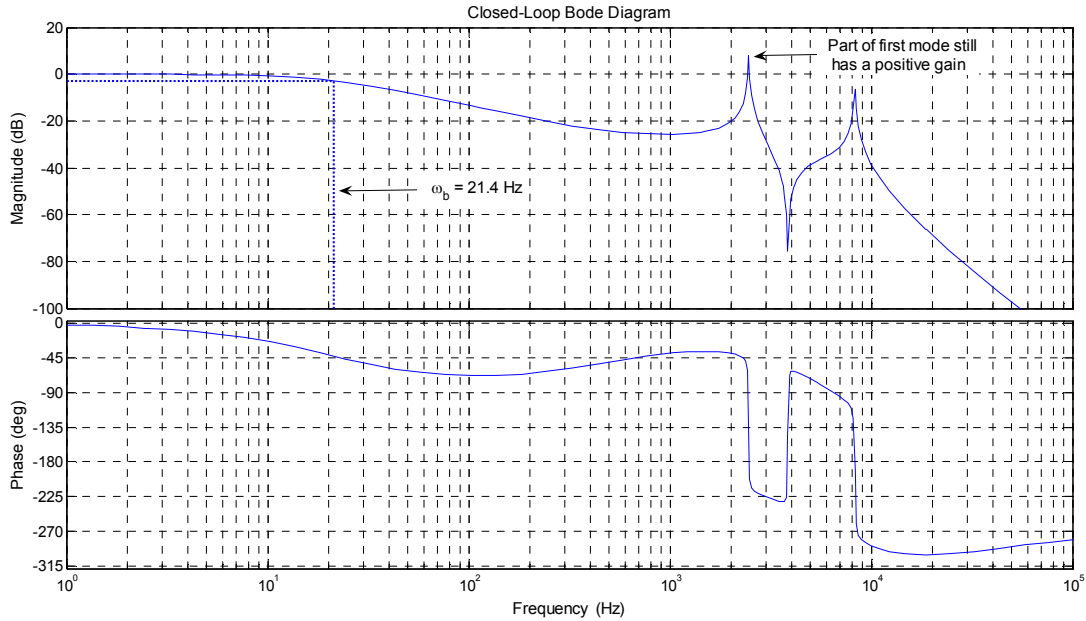
Figure 5.9   Closed-loop Bode diagram with ZN-tuned PI controller, the system of Equation (5.8). Bandwidth (first -3 dB crossing from above) is 21.4 Hz. First mode still has a small part that is above 0 dB (i.e a positive gain). This peak is 8 dB at a frequency of 2.46 Hz.

The unit (1 μm) step response of both the model and the true system are shown in Figure 5.10. Comparatively, the actual system is slightly faster than the model. Settling times (±2% of set point) are 29 and 22 msec for the model and the true system, respectively. Performance is otherwise similar. Neither has any overshoot worth mentioning. Both show a steady-state tracking error of zero. As expected, oscillations were present at the first mode frequency of 2.48 kHz. For the model, they are most prevalent in the transient portion of the response and non-existent in the steady state portion. This is intuitive since there were no disturbances included in the model that could have excited any mode. In the true system, oscillations are less obvious in the transient portion, and more prevalent in the steady-state portion (e.g. apparent oscillations around 0.05 and 0.06 seconds). Also as expected, the oscillations in both cases are small compared to the signal. As far as

118

those that appear in the true system, they remained within the ±2% settling-time

bounds and died out quickly after they appear. Thus, they are of little concern.
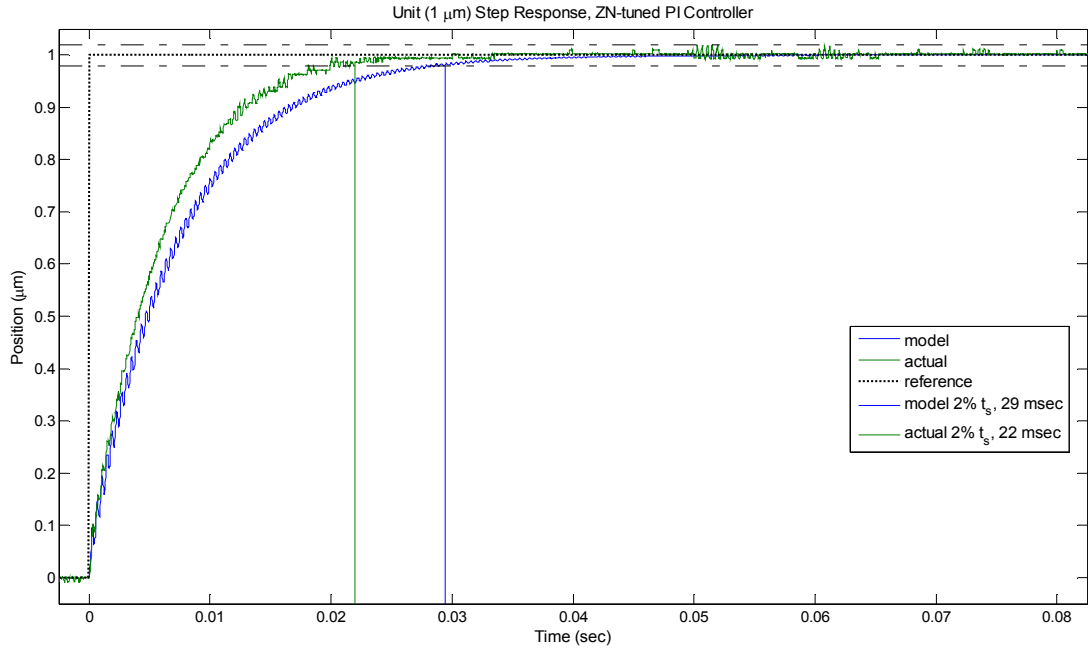
Figure 5.10   Unit (1 μm) step response with ZN-tuned PI controller. Settling times are similar but not equal. Model has a 29 msec settling time, and 22 msec is what actually occurs. Neither has any significant overshoot. Both appear to have zero steady-state error.

Step responses of ten and fifty μm are shown in Figure 5.11 and Figure 5.12

respectively.   The settling time of the model prediction remains at 29 msec as

expected. Settling time of the true system, however, decreases. The system appears

to get faster with increasing step magnitude. Settling times are 12 and 8 msec for the

ten and fifty μm step responses respectively.   The oscillations at the first-mode

frequency are still present in transient portion of the model response, and still have

small amplitude as before.   However, oscillations do not appear to be present in the

true response as before.   The larger set point magnitude makes the oscillations much

less apparent than in the unit step response case.
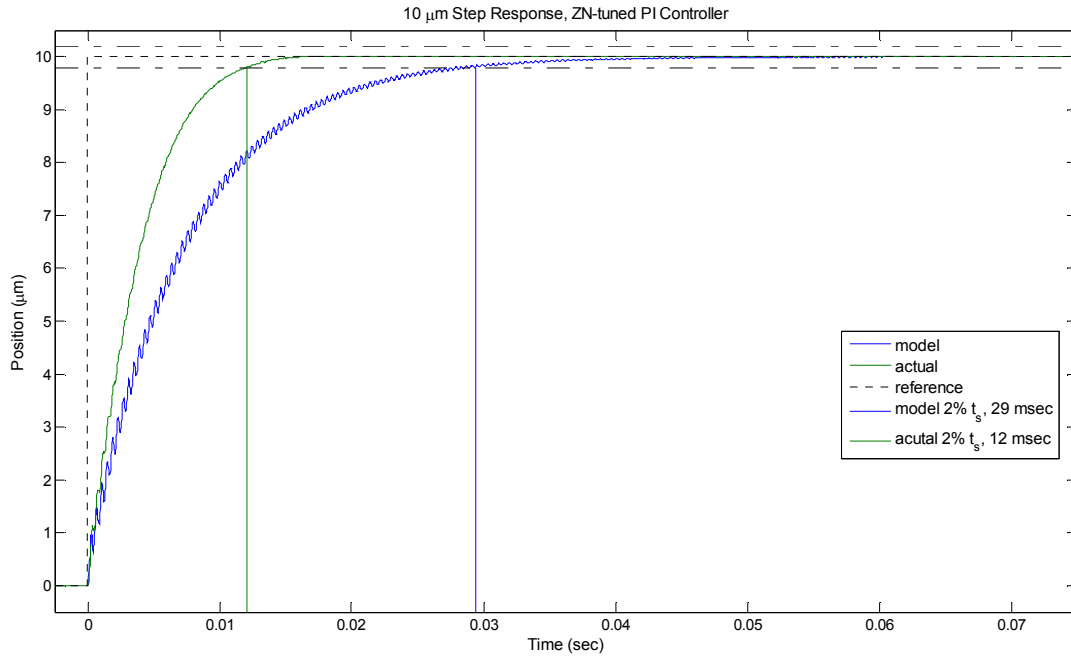
Figure 5.11 Ten μm step response with ZN-tuned PI controller. Settling time is lower (or faster) in the true system as compared to the unit step response case.
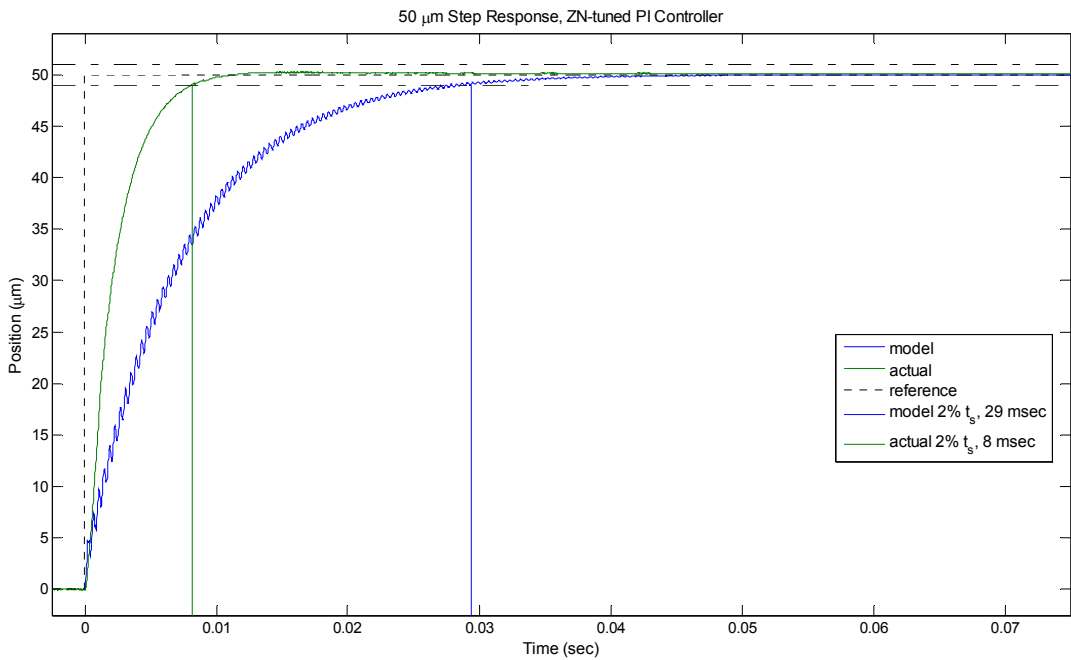
Figure 5.12 Fifty μm step response with ZN-tuned PI controller. Settling time is lower (faster) in the true system as compared to the unit and ten μm step response case

The step response is one of many reference signals used to characterize closed-loop systems (step, impulse, doublet, sinusoid, triangle wave, etc.). To show each and every one at varying amplitudes would be exhausting. A more efficient way to verify the closed-loop system is to compare the closed-loop frequency responses of both the model and the true system. This supersedes the need to show several other closed-loop tracking responses. Closed-loop frequency responses for the model and the true system are shown in Figure 5.13.
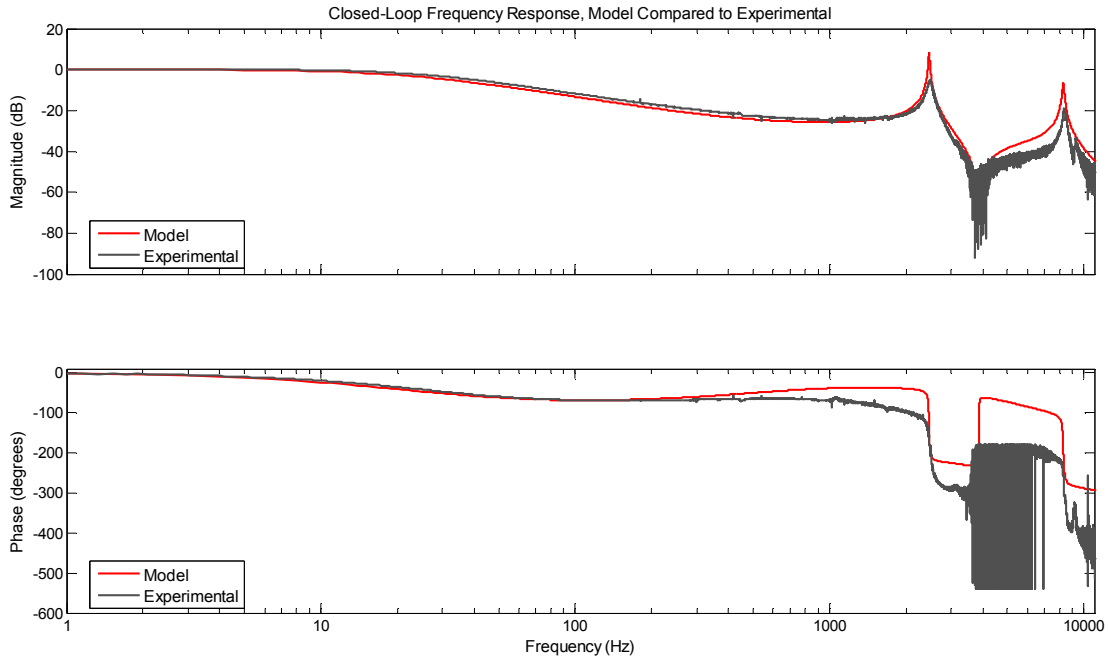
Figure 5.13   Closed-loop frequency response comparison between model and real system. Correspondence of the magnitude is excellent throughout entire frequency range of interest. Phase corresponds well up to approximately 300 Hz, after which it begins to deviate.

Correspondence of the magnitude between model and experiment are excellent throughout the frequency range of interest of up to 10 kHz. The resonant peaks of the physical system are slightly lower and less sharp than their model counterparts, indicating that there is slightly more damping in the system than what

was captured by the model. This explained why oscillations present in the transient portion of the model were not present in the actual response.

The phase of the true system is coincident with the model up to approximately 300 Hz, after which it begins to deviate. Above 300 Hz, the actual phase is lower than predicted by the model, but still maintains the same general shape. The apparent phase oscillation centered on 5 kHz is actually the phase rapidly switching back and forth between -180° and -540°. These two angles are 360° apart, so they essentially amount to the same phase of -180°.

The agreement between the model and the actual frequency responses validates both the open and the closed-loop system models. It verifies that the controller was designed for the proper system and will yield a real closed-loop system that will behave in a predictable manner, i.e. one that is predicted by the model.

As one final test of the closed-loop system's ability, tracking of a signal similar to one that could be seen in the applications discussed in Chapter 1.1 was tested. Closed loop system performance was compared to that of the open-loop system. Shown in Figure 5.14 is the response of both systems to a multi-step reference tracking signal. The 1 μm step allowable for good signal-to-noise ratio was reached in four equal steps of 0.250 μm. This type of signal is representative of what could be encountered in the AFM/SPM application. The closed loop system shows better performance in all but settling time. The open-loop system reaches its setpoint sooner than the closed-loop system. However, the open-loop system shows significant overshoot at each step. More importantly, its output does not step consistently with an input signal that does step consistently, resulting in severe

122

degradation of signal tracking performance. Thus, Figure 5.14 solidifies the argument that applying closed-loop control will result in superior performance as compared to an open-loop system. Table 5.3 summarizes the performance of both the modeled and the physical closed-loop system.
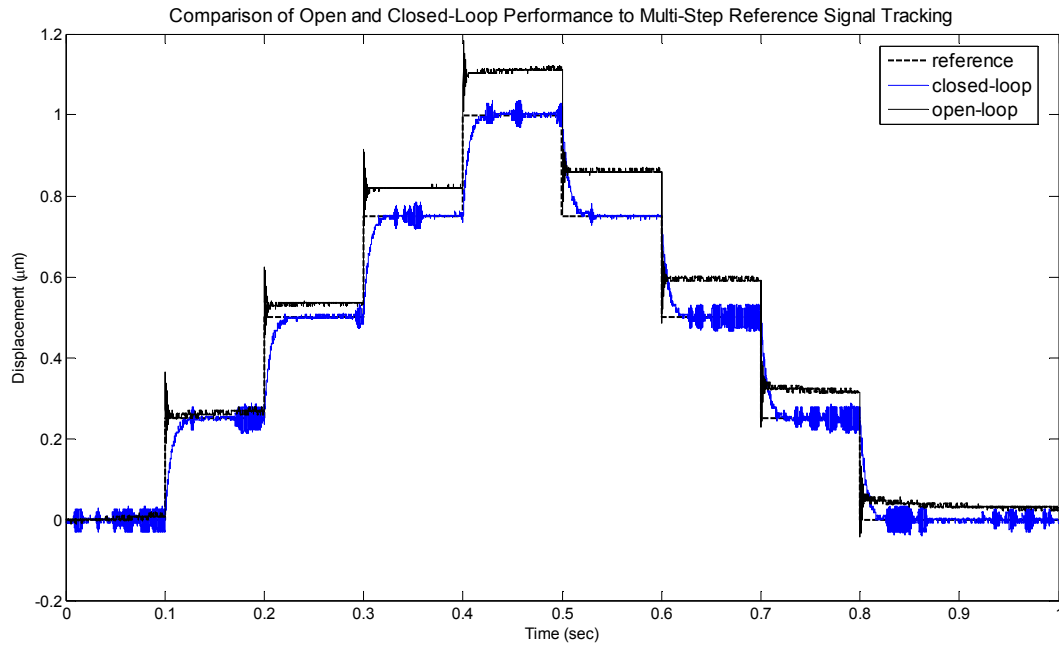
Figure 5.14   Comparison of open and closed-loop schemes' ability to track a multi-step reference signal. The 1 μm resolution limit was reached by four 0.250 μm steps, and then repeated to return to the zero position. This type of signal is more like what could be seen in an AFM trace than a single step response.

Table 5.3   Performance metrics (unit step response and frequency response) using the controller in Equation (5.6). The metrics are shown for both the model and the physcial system. Percent Difference is expressed with the model as the basis.

Performance Metric	Model	Physical System	% Difference
Overshoot	≈0	≈0	0%
Settling Time, 2% bound	29 msec	22 msec	24%
Rise Time, 0 to 90%	16 msec	13 msec	19%
Steady State Error	0	≈0	0%
Gain Margin	4.1 dB	NA*	-
Phase Margin	92°	NA*	-

* NA = not available. Open-loop data w/controller was not taken.

123

6 Conclusions and Recommendations

6.1 Conclusions

The underlying purpose of this work was to model a piezoelectric flextensional actuator, design a closed-loop control algorithm to accurately control actuator displacement, and to verify the closed-loop performance in the real-world system. All tasks were successfully executed. Experiment setup, open-loop model design and verification, and closed-loop control design and verification were thoroughly explained. The result is the outline of a process by which other actuators, of this type and others, can be characterized and controlled. Conclusions can be drawn about each task and how they impact each other.

The entire project began and ended with real-time experiments. Open-loop experimental data was used to create the model, and closed-loop experimental data was used to verify control system performance. The experiment setup and its capability to produce clean, repeatable data directly affected both the quality of system models and the degree of precision that could be demanded of the closed-loop system. Noise and disturbances in the experiment were assessed so that input and output signal magnitudes could be chosen accordingly

Open-loop modeling of the actuator was performed by applying system identification algorithms to input-output data obtained using a chirp signal as the input. The open-loop model captured the dynamics sufficiently. DC gain, however, was not as accurate as desired. Inaccuracy of the DC gain was a problem with the

magnitude of the input signal used to obtain the data sets rather than with the method of system identification. Magnitude of an output due to a chirp input is dependent on the chirp amplitude, which was chosen to be small so that when the system scanned through resonance, instability was avoided. The result was an output magnitude that had good signal-to-noise-ratio at and around the mode frequencies, but a lower signal-to-noise ratio in the low-frequency region that defines DC gain. This was the reason transient dynamics (defined by the modes) were well-captured but the steady-state behavior (defined by low-frequency dynamics) was not captured as well. Feedback control is able to accommodate minor model inaccuracies, so despite the small inaccuracy of the DC gain, the model was deemed sufficient for control design.

The controller for the closed-loop system was designed based on minimizing the integral-squared error. Typical closed-loop performance requirements such as low settling time and minimal overshoot are aimed at reducing the overall response error. Tuning to minimize ISE will usually offer a good compromise between all performance metric requirements. Closed-loop performance with the minimal ISE controllers was excellent. Simulations showed generally acceptable overshoot and sub-millisecond rise and settling times.

Applying the minimal ISE controllers to the true system, however, resulted in an unstable system. This forced the reevaluation of the model, which revealed other model deficiencies. The closed-loop system required a high-voltage amplifier to provide the necessary control authority to command the closed-loop system. It was assumed that since the bandwidth stated in the amplifier manual was far beyond that the actuator, that neither inclusion nor exclusion of its dynamics would have

significant effect. This turned out to be an incorrect assumption because of the dielectric nature of the piezoelectric element. Driving a capacitive load severely deteriorates the bandwidth of the amplifier. It decreases from greater than 35 kHz as stated, to approximately 6 kHz. This frequency is located between the actuator resonances, and thus dramatically alters the system. Addition of a complex pole pair to account for these dynamics was necessary. The controller was redesigned for the completed system, and then retested. Performance comparisons between model and experiment showed good agreement, indicating that the redesigned open and closed-loop models are accurate and that the feedback control system adequately commands the real system to track reference displacements.

6.2 Recommendations for Future Work

Future work should be directed toward improving various aspects of the work contained here. Changes to the experiment could improve its capabilities, which would result in better data sets for model creation and better resolution and precision from the closed-loop system. Alternative modeling techniques and control design methods could improve accuracy and performance. The actuator itself could be modified for better resonant and low-frequency behavior.

All experiments were performed on a composite optical table with passively damped vibration isolation legs. Moving the experiment to a solid optical table with active vibration isolation could offer better noise and disturbance properties, allowing better signal-to-noise ratios for smaller input-output signal magnitudes, thus

increasing the precision and resolution of measured position. Alternative sensors for measuring displacement could be explored, e.g. LVDT, piezoresistive, etc. This may or may not prove useful, because laser vibrometry already provides excellent feedback.

With improved displacement resolution, models could be developed around smaller operating points, perhaps at the nanometer scale. Developing multiple models around progressively smaller and/or larger operating points (i.e. models centered on 0.01 μm, 0.1 μm, 1 μm, 5 μm, 25 μm, etc.) could offer insight into the degree of nonlinearity inherent to the system. If it is decided that nonlinearity is too great for one controller to function over the entire displacement range of the actuator, then several controllers can be designed for different operating ranges in order to implement gain scheduling.

Regarding the actuator itself, several design changes are viable and that could improve performance and render the device more controllable. The actuator has two vibration modes and one zero, all in close proximity. Changing the type of metal from which the shell is made could offer improvements by causing one or both of the modes and the zero to change locations or diminish altogether. Using a less stiff metal such as aluminum could also offer increased displacement without much loss of load carrying capability. A trade study as in [68] could provide insight into benefits of altering either the material or geometry of both components of the actuator.

Adding a viscoelastic polymer layer to the inside of the metal could also add passive damping to the system. This would make the resonant peaks less sharp and easier to suppress and control. Furthermore, if said viscoelastic polymer layer were

also electrically active, as in PVDF, there would then be the option of adding active damping to the system. Adding an active polymer for damping would result in a multi-input system, for which optimal controls methods would be better suited. This would offer a more rigorous approach to control design, and increased ability to move the system poles to desired locations.

No matter what options are chosen for future work with piezoelectric flextensional actuators, all efforts should be directed toward increasing the displacement range, obtaining better displacement resolution, and yielding a more easily controllable system. As always, there must be compromise between these objectives. However, increasing the device performance in any or all of these areas will give the device wider applicability and better marketability.

Appendices

Appendix A.   A Computational Method for Materials Selection in a Hybrid Actuation System[*]

Benjamin J. Nickless[a], Ji Su[b], Tian-Bing Xu[c], James E. Hubbard Jr.[a]

[a]University of Maryland, 3181 Glenn L. Martin Hall, College Park, MD, 20742

[b]NASA-Langley Research Center, Hampton, VA, 23681

[c]National Institute of Aerospace, 100 Exploration Way, Hampton, VA, 23666

ABSTRACT

Researchers at NASA-LaRC have developed a hybrid actuation system (HYBAS) that cooperatively employs an electroactive polymer and an electrostrictive single crystal. Experimental measurements and theoretical model predictions have been in good agreement thus far. To date, current research has only explored the usage of one electroactive polymer and one electrostrictive single crystal. A computational model was created based on this theoretical model. It implements the equations necessary to predict the actuator displacement profile and maximum displacement. Among the model variables are the actuator material properties. Changing the actuator materials has notable effects on actuator performance. As many viable materials as could be found were compiled into a database which can serve as a building block upon which a larger database can be built. Using these materials, a trade study was performed to determine which combination of materials demonstrates the best performance. As more electroactive materials are compiled, more extensive trade studies can be performed. Thus, the work in this paper will serve as a guideline for future HYBAS designs.

---

[*] Original document reprinted here with permission from the publisher [68-69].

1. INTRODUCTION

A growing trend in micro- and nano-technology and multifunctional materials has caused designers to look towards smart materials for use in actuators because of their dimensional versatility, mechanical simplicity, and high energy density. These are all desirable qualities, but the pursuit of actuation design is to obtain large actuation force and high displacement without requiring a relatively high operating voltage or current. However, actuators made from smart materials are usually found wanting in at least one of these three categories. Despite the shortcomings, a great effort has been made to amplify their displacement, increase their actuation force, and enhance their overall efficiency. Among the first and most popular resulting developments are the multilayer stacks and bimorphs. Newnham et al later introduced the ceramic metal composite actuator. Many devices aimed at amplifying displacement have followed[1-6].

Su et al at NASA-Langley Research Center have conceptualized and fabricated one such device[6]. This hybrid actuation system, known as HYBAS, harmoniously employs two electroactive components in order to achieve enhanced electromechanical performance and efficiency. The HYBAS is currently in the prototype design stage. Different dimensional and geometric combinations have been tested in search of optimal performance. The two types of materials utilized are an electroactive polymer (EAP) and an electrostrictive single crystal (ESC). Only one of

each type of these materials has been exploited to date. The material employed for the EAP component is uni-axially stretched and high energy electron irradiated 68/32 mol.% poly(vinylidene-fluoride-trifluoroethylene) copolymer (PVDF-TrFE), and for the ESC component is lead zinc niobate-lead titanate (PZN-PT single crystal). A proposed theoretical model has also been put forth and compared to experimental data[6]. Thus far, theoretical and experimental results have been in good agreement.

There are many materials that exhibit an electric-field induced strain, yet no significant database of such materials exists. Using a computational version of the theoretical model of HYBAS by Xu et al[7], shape profile and maximum displacement results can be obtained. The computational model allows changes to be made to design parameters, including material properties. After refining the computational model, its predictions were compared to those of the theoretical model to assess the correctness and accuracy of both models. After arriving at a sufficiently accurate computational model, a trade study regarding materials was conducted using a database of materials gathered by the authors. All other design parameters were held constant. The HYBAS models, both theoretical and computational, and their similarities and differences will be addressed in this paper. Comparative results of the trade study will follow.

2. THE CURRENT HYBAS ACTUATOR

The HYBAS has two active components and three inactive components. Active components consist of the ESC and the active EAP layer. Inactive components are comprised of electrodes, a plastic frame, and an inactive EAP layer.

A diagram of the HYBAS illustrating its relative geometric layout and different components is shown in Fig. 1.
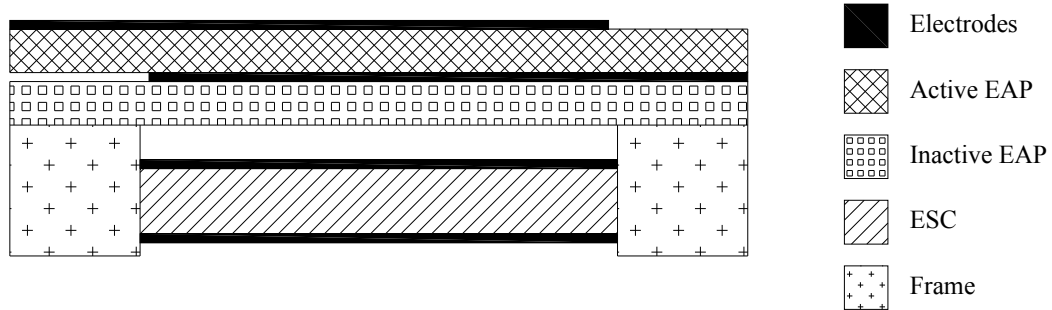
Fabrication of a HYBAS begins with the ESC component. The upper and lower surfaces are coated with gold electrodes. The frame consists of two sections of plastic bar or rod, which are bonded to each end of the ESC component respectively. A frame is necessary in order to couple the ESC element and the EAP element in order to take advantage of both of their desired qualities. After coating the active EAP layer with gold electrodes, an inactive EAP layer of the same polymer is bonded to it forming the complete EAP component. The complete EAP component is then bonded to the top edges of each side of the plastic frame as shown in Fig. 1.

The PZN-PT single crystal used in the original and other possible materials for application in the ESC element are typically ceramic or exhibit ceramic-like properties. They are moderately stiff and do not tolerate large bending displacement. Electroactive polymers, including the PVDF used in the original, are usually less stiff compared to the single crystal materials. They also are able to tolerate large bending displacements without adverse effects. The ESC materials experience a dimensional

decrease with applied electric field while the EAP materials experience a dimensional increase in their respective length directions perpendicular to the applied field.

The HYBAS was designed with careful consideration in order to take advantage of all the aforementioned properties of both material types. Electrically activating either or both components creates a moment in both components. The ESC element is made sufficiently stiffer than the EAP element (higher Elastic modulus and larger cross-sectional area than the EAP) so that the ESC remains planar. As a result, the EAP will buckle when either or both elements are activated. This is known as flextensional actuation and offers increased displacement compared to traditional piezostack actuators[2-4, 8]. This creates an actuator displacement profile in the HYBAS shown in Fig. 2.

Initially the EAP is planar and parallel to the ESC. The moment created by either the ESC contracting or the EAP elongating could cause the EAP to buckle in either the positive or negative z-direction. However, only displacement in the positive z-direction is desired. The inactive EAP layer serves this purpose. It is employed to bias the EAP element so that it will always buckle and displace in the positive Z-direction[6-7, 9]. The inactive EAP layer is the same PVDF copolymer as the active EAP layer.
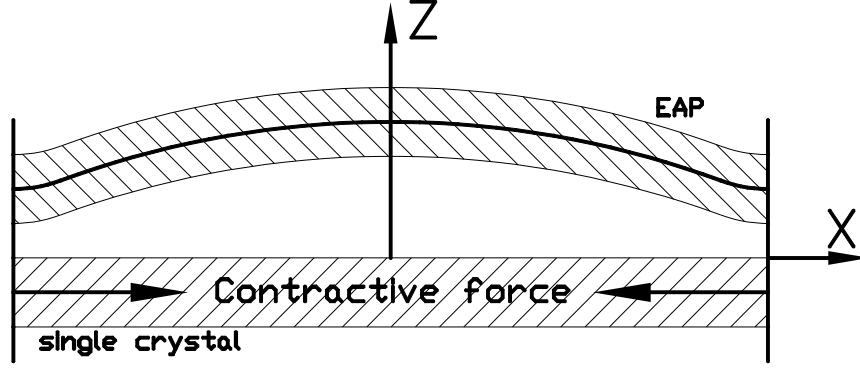
Fig. 2. Actuation response, displacement is in the z-direction[9].

3. THEORETICAL MODEL

The HYBAS was modeled as a rectangular beam fixed at both ends subjected to a uniformly distributed load in the Z-direction[7]. The ESC component realizes displacement in the x-direction only and is much stiffer than the EAP component as mentioned before. As such it defines the dynamic length L_d of the actuator which is given by

$$L_d = L_0\left(1 + s_{ESC}^e\right) \tag{1}$$

where L_0 is the initial length of the ESC when the applied electric field is zero and s_{ESC}^e is the effective electrostrictive strain in the ESC which is a function of the applied electric field[7]. The change in length exerts a contractive force on the EAP element. The contractive force causes the EAP element to buckle. This coupled with the effects of the inactive EAP layer induce a uniformly distributed load on the EAP element. The displacement in the z-direction of such a beam is given by

$$w(x) = \frac{p}{24EI}\left[\left(\frac{L_d}{2}\right)^2 - x^2\right]^2 = c\left[\left(\frac{L_d}{2}\right)^2 - x^2\right]^2 \tag{2}$$

135

where E and I are the Young's modulus and moment of inertia of the EAP element, respectively, and p is the uniformly distributed load per unit length, which is dependent on the strain in both the EAP and ESC elements[10]. Equation (2) differs from that which appears in the reference in that it has been shifted so that $x=0$ is at the center of the beam. By using standard extrema finding techniques the maximum displacement is found to occur at the center of the beam and is given by

$$w_{max} = \frac{cL_d^4}{16} .$$
(3)

It is obvious from Eqs. (2) and (3) that the parameter $c = p/24EI$ plays an important role in the magnitude of the actuation response. The uniformly distributed load p, and thus c, is dependent on the strains in both the EAP and the ESC components[7]. To obtain a value for c necessitates considering the calculation of the total length of the EAP element in its displaced form. The total length can be expressed in two manners, and both must be equal, yielding

$$\int_{-L_d/2}^{L_d/2} \sqrt{\left(\frac{dw}{dx}\right)^2 + \left(\frac{dx}{dx}\right)^2} \, dx = L_0\left(1 + s_{EAP}^e\right) .$$
(4)

Inserting the derivative of Eq. (2) into Eq. (4) yields

$$\int_{-L_d/2}^{L_d/2} \sqrt{c^2\left(4x^3 - L_d^2 x\right)^2 + 1} \, dx = L_0\left(1 + s_{EAP}^e\right)$$
(5)

in which s_{EAP}^e is the effective strain in the EAP component. A value for c can be obtained using Eq. (5). The left-hand side of Eq. (5) cannot be evaluated using analytical techniques. It requires numerical integration to calculate which implies guessing at a value for c until Eq. (5) is satisfied. Once c is obtained, it can be

substituted into Eqs. (2) and (3) to obtain predictions for the displacement profile and maximum value.

In turn, the right-hand side of Eq. (5), and thus s_{EAP}^e, must also be evaluated in order to get a value for c. The ESC component contains only active layers, whereas the EAP component contains an inactive layer that constrains its motion in both the X- and Z-directions. As such s_{EAP}^e is affected by these factors and is less than but related to the free strain of the EAP component. Xu et al considered these effects when deriving the theoretical model of the HYBAS and found the relationship between the free strain s_0 and effective strain in this case to be[7]

$$s_{EAP}^e = \frac{s_0}{1+k} \tag{6}$$

where k is the ratio of inactive layer stiffness to active layer stiffness and is referred to as the clamping ratio. It can be expressed as[9]

$$k = \frac{\sum_{i=1}^{q}(E_n t_n b_n)_i}{E_a t_a b_a} \tag{7}$$

where the subscript a denotes active layers, the subscript n denotes inactive layers, and q denotes the total number of inactive layers that are to be include in the calculation of k. Included in the calculation of k in the theoretical model are the inactive EAP layer, the epoxy layer bonding the inactive and active layers, both of the gold electrodes, and the margins of the active layer not covered by electrodes[7]. The geometric and pertinent material properties used in the theoretical model are shown in Table 1.

Table 1. Dimensions and necessary material properties used in the theoretical model.

Component	Effective Length (mm)	Thickness, t (μm)	Width, b (mm)		Young's Modulus (GPa)
			Total	Effective	
ESC	5.50	470.0	3.00	3.00	20.0
Active EAP layer	5.50	16.0	4.50	3.00	1.0
Inactive EAP layer	5.50	15.0	4.50	4.50	1.0
epoxy layer	5.50	1.0	4.50	4.50	5.0
gold electrodes on EAP	5.50	0.1	3.00	3.00	74.0
unelectroded margins	5.50	16.0	0.75	0.75	1.0

4. COMPUTATIONAL MODEL

The computational model is largely based on the theoretical model described above. All of the previous equations are implemented, although some assumptions had to be made in order to make their implementation possible. Before going any further it is necessary to redefine the nomenclature of components of HYBAS. Thus far the PZN-PT single crystal has been referred to simply as the ESC (electrostrictive single crystal) component. Electrostrictive suggests a particular strain vs. electric field relationship and single crystal refers to a material whose formation is tightly controlled so that all of the crystals form in a particular orientation. It is not likely that all of the materials used in the trade study will exhibit this particular relationship or have this particular crystal structure. For the HYBAS to function properly, the material performing the function of the ESC component must contract when an electric field is applied, and the material functioning as the EAP component must elongate when an electric field is applied. So from this point forward the ESC component will be referred to as the negative strain component and the EAP component will be referred to as the positive strain component.

The response to an applied electric field is assumed to be piezoelectric for all materials included in the study. This implies that the relationship between free strain

and the applied electric field is linear and the slope is the piezoelectric constant of the material. As noted in the previous section, the effective strain in the positive strain component is different from its free strain. The effective strain of the negative strain component, however, is comparable to its free strain[7]. As such, the effective strain in the negative strain component is estimated as being equal to its free strain, thus no clamping ration is calculated for it. A piezoelectric material typically has many piezoelectric constants depending on the applied electric field direction and the output direction being examined. In this case the applied field is in the *Z*-direction and the working displacement direction is the *X*-direction. So, only the d_{31} piezoelectric constant for each active material is needed here. The free strain will be given by

$$s^f = d_{31} \frac{V}{t}$$

(8)

where *V* is the applied voltage and *t* is the material thickness. The quantity *V/t* is the applied electric field. Under the prior assumptions, the negative strain component effective strain will be given by Eq. (8) with appropriate parameters embedded. The positive strain component effective strain will still be given by Eq. (6), but with the free strain s_0 given by Eq. (8) with the appropriate parameters being used.

The main goal of the computational model is to be able to compute Eqs (2) and (3) using different material properties in order to compare different HYBAS configurations. The inherent problem previously mentioned in calculating these two equations is obtaining a value for the parameter *c*. Obtaining this value requires finding that which satisfies Eq. (5). Once the properties of the desired material are supplied, an initial guess for the value of *c* must be made by the program user. The left-hand side of Eq. (5) is then numerically integrated with this value of *c* using a

left-hand Riemann sum algorithm with 1000 subintervals. After this computation is finished, the value is compared to the value computed for the right-hand side of Eq. (5). If the desired accuracy of this comparison is met, this value of c is kept and the computation is complete. If the accuracy condition is not satisfied, the program will change the value of c as necessary and repeat the process. Iteration upon c continues until the desired accuracy is achieved. Here, accuracy refers to the equality of Eq. (5), not the accuracy of the numerical integration itself. Exact equality of Eq. (5) cannot be expected as numerical integration is involved, so the value of c is sought that makes Eq. (5) as true as desired by asking for a certain level of accuracy (i.e. specifying a higher accuracy forces equality to more decimal places).

The values obtained for c are where the computational model shows the largest amount of deviation from the values obtained by Xu et al for the original HYBAS configuration. Table 2 gives a comparison between the two models in this respect. Percent error refers to the accuracy described earlier and represents the amount of difference between the left- and right-hand sides of Eq. (5), that is to say, the error in the true value of c. The values reported for c under the theoretical model appear in Ref. [7], but their associated percent errors were calculated later, during the writing of this paper.

It can be seen from Table 2 that the percent error in the theoretical model never drops below 0.0066%, and that by virtue of the algorithm the percent error with the computational model is consistently less than 0.001%. Thus, under the given assumptions, the values for c found by the computational model are more accurate in all cases. But in spite of its accuracy deficiencies the theoretical model shows

remarkable correlation with experimentally measured data[7]. While it is safe to say

that the theoretical model is superior in displacement prediction for the original

HYBAS configuration, the same cannot be said for the general case. While the

computational model does not match the experimental data for the original HYBAS

as well as the theoretical model, it implements a fundamentally sound and reliable

algorithm for HYBAS performance prediction and can be applied with various

HYBAS configurations. The inconsistency with the original experimental is

inconsequential because the same algorithm will be applied to each configuration. As

all the results will be subject to the same influences they will be sufficient for

comparison.

Table 2. Comparison of theoretical and computational model regarding values computed for *c*.

Voltage (V_{RMS})	Active Elements	Theoretical Model		Computational Model	
		c (10^6 m^{-3})	percent error (%)	c (10^6 m^{-3})	percent error (%)
200	EAP	0.1786	0.0066	0.5591	
	ESC	0.3774	0.0384	1.2579	accuracy set so iteration stops when percent error is less than 0.001%
	HYBAS	0.5714	0.0409	1.3692	
400	EAP	1.3158	0.0308	0.7667	
	ESC	1.0000	0.0580	1.7924	
	HYBAS	2.0964	0.0160	1.9510	
800	EAP	3.3340	0.2627	1.0511	
	ESC	1.2500	0.1278	2.5540	default starting point for *c* in all cases is 1.000
	HYBAS	3.8910	0.1951	2.7523	
1600	EAP	5.5866	0.7582	1.4978	
	ESC	3.2250	0.0697	3.6392	
	HYBAS	6.5876	0.7117	3.9138	

The materials that were used in the trade study along with their constants are

shown in Table 3 which comprises the list of materials that was gathered and whose

material constants could be considered reliable. While the list is not long, it provides

a sufficient database for the current research and a building block upon which to

augment for future purposes.

Table 3. Materials and their properties used in HYBAS trade study[11-13].

	Material	d_{31} (pC/N)	dielectric strength or positive depoling field (kV/cm)	Elastic Modulus (GPa)
negative strain components	Hard PZT (TRS100HD)	-150	20-50	79
	Soft PZT (TRSHK1HD)	-360	20-50	67
	PZN-4.5%PT single crystal	-970	20-50	12
positive strain components	Uni-axial PVDF	20	1600	2
	Bi-axial PVDF	8	1600	2

The positive depoling fields for the negative strain component materials are applicable only to positive applied electric fields. The negative depoling field, or the coercive field, applies to negative applied fields and are typically lower than the positive depoling fields. For this reason, the program assumes a DC bias is applied so that the minimum of the applied AC field is zero and so that higher voltages can be applied.

5. RESULTS

All viable combinations of materials in Table 3 were simulated at 100 V_{RMS} and 650 V_{RMS}. These two voltages were chosen based on the electric field capability of the materials. When working with electroactive materials, their dielectric strength or depoling field must be considered. If a material does in fact have such properties, the material will fail if this electric field is exceeded. Thus any results from simulations above this electric field would be erroneous. As the results of this study are intended to be comparative, the applied voltages (and thus applied electric fields due to fixed geometry) need to be the same for each material combination. Given the thickness of both negative and positive strain components and the dielectric strength

and field properties in Table 3, the peak voltage that can be applied during this study is 940 V, which mandates 665 V_{RMS}. This is governed by the negative strain components. In spite of their larger thickness, their depoling field is much lower than the positive strain component materials. Since the limit is 665 V_{RMS}, 650 V_{RMS} is chosen for simulation to avoid virtual material destruction. As for the lower limit, 100 V_{RMS} is chosen arbitrarily. The true lower limit would obviously be 0 V_{RMS}, but since that is the trivial case, it is not considered.

Either one or both of the active components can be electrically activated at any given time. Figures 3 shows the displacement profiles of the fully active HYBAS (both components are active) at 100 V_{RMS}. The HYBAS displacement profiles at 650 V_{RMS} are presented in Fig. 4. It can be gathered from Figs. 3-4 that with the present set of materials, there are more gains to be had by varying the negative strain component material. There is not a significant difference in piezoelectric constant of the two positive strain component materials as compared to the negative strain component materials. Additionally, the piezoelectric constant of the negative strain component materials are an order of magnitude higher than those of the positive strain component materials. The implications are that the displacement gains from varying the positive strain component material are miniscule as compared to varying negative strain component. Figures 3-4 illustrate this point well when comparing plots in the first column with plots in the second column.
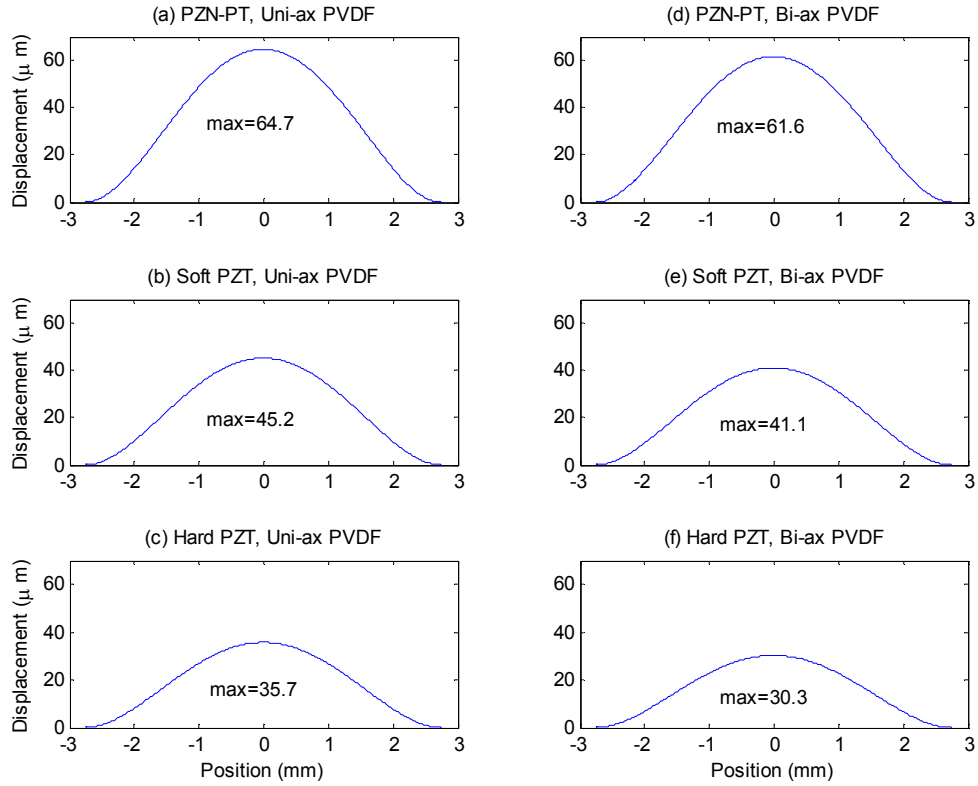
Fig. 3. Displacement profiles with both components active at V_{RMS} = 100 V. (a) PZN-PT single crystal and uni-axial PVDF, (b) Soft PZT and uni-axial PVDF, (c) Hard PZT and uni-axial PVDF, (d) PZN-PT single crystal and bi-axial PVDF, (e) Soft PZT and bi-axial PVDF, and (f) Hard PZT and bi-axial PVDF.
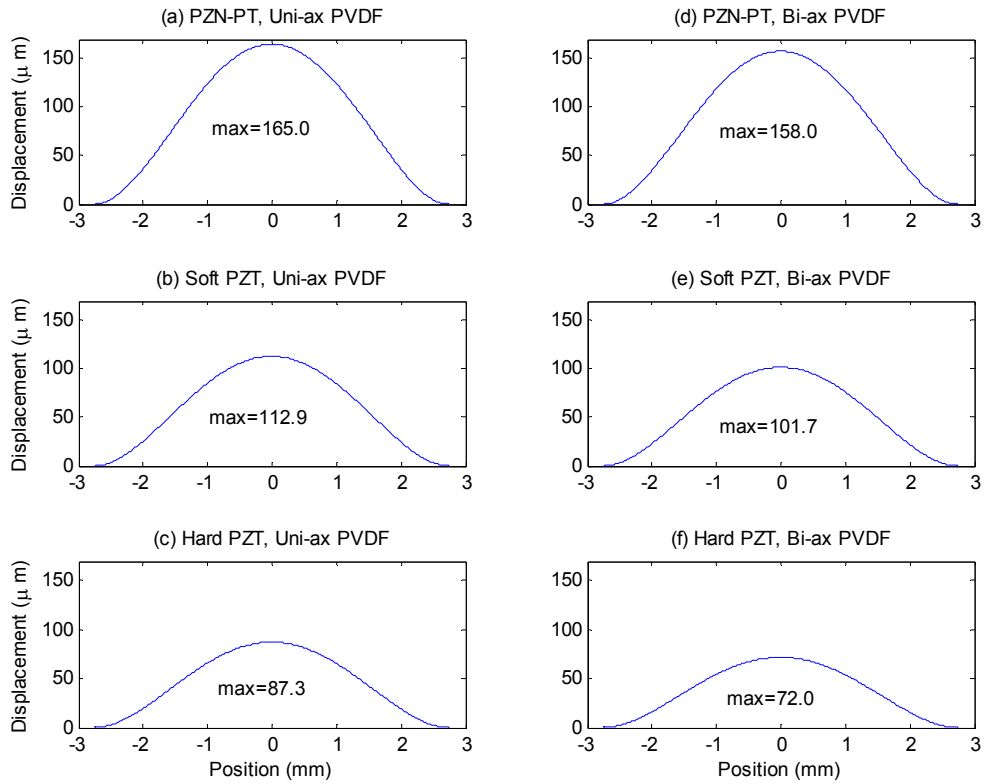
Fig. 4. Displacement profiles with both components active at $V_{RMS} = 650$ V. (a) PZN-PT single crystal and uni-axial PVDF, (b) Soft PZT and uni-axial PVDF, (c) Hard PZT and uni-axial PVDF, (d) PZN-PT single crystal and bi-axial PVDF, (e) Soft PZT and bi-axial PVDF, and (f) Hard PZT and bi-axial PVDF.
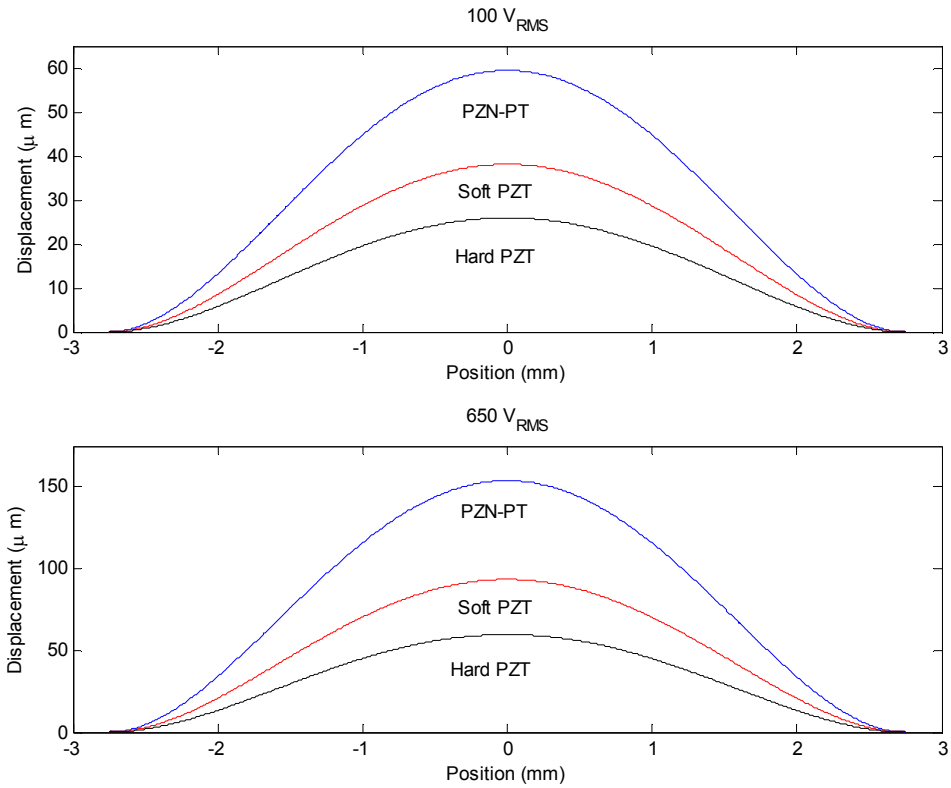
Fig. 5. Displacement profiles with only the ESC component active at $V_{RMS} = 100$ and 650 V.

Figure 5 shows the displacement profiles at 100 and 650 V_{RMS} with just the ESC component active, and Fig. 6 shows the profiles with just the EAP component active at the same voltages. Figures 5-6 reassert the conclusions drawn from Figs. 3-4 that varying the negative strain component has more impact on displacement than varying the positive strain component.

146
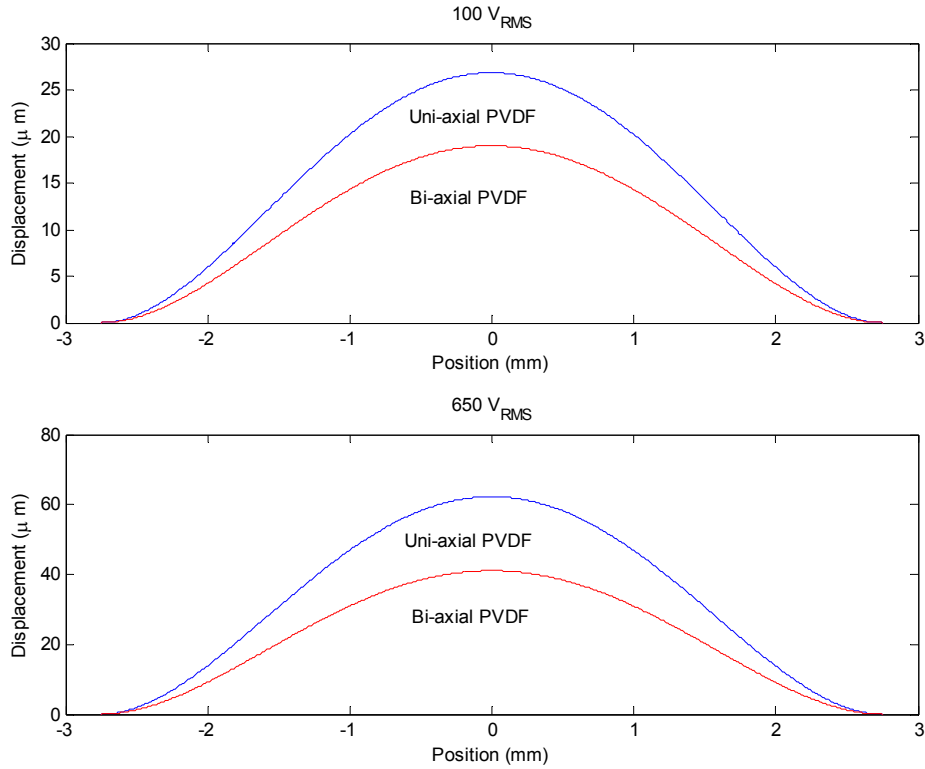
Fig. 6. Displacement profiles with only the EAP component active at V_{RMS} = 100 and 650 V.

Displacement as a function of applied RMS voltage for all viable material combinations with both components active is shown in Fig. 7. It is obvious from Fig. 7 that the best combination of materials for achieving maximum displacement is PZN-PT single crystal and uni-axial PVDF. These materials have the highest piezoelectric constants in the negative and positive strain component groups respectively. If achieving maximum displacement is the only goal, then the materials with the highest absolute value of piezoelectric (or electrostrictive, etc.) constants will always be best suited for the application.
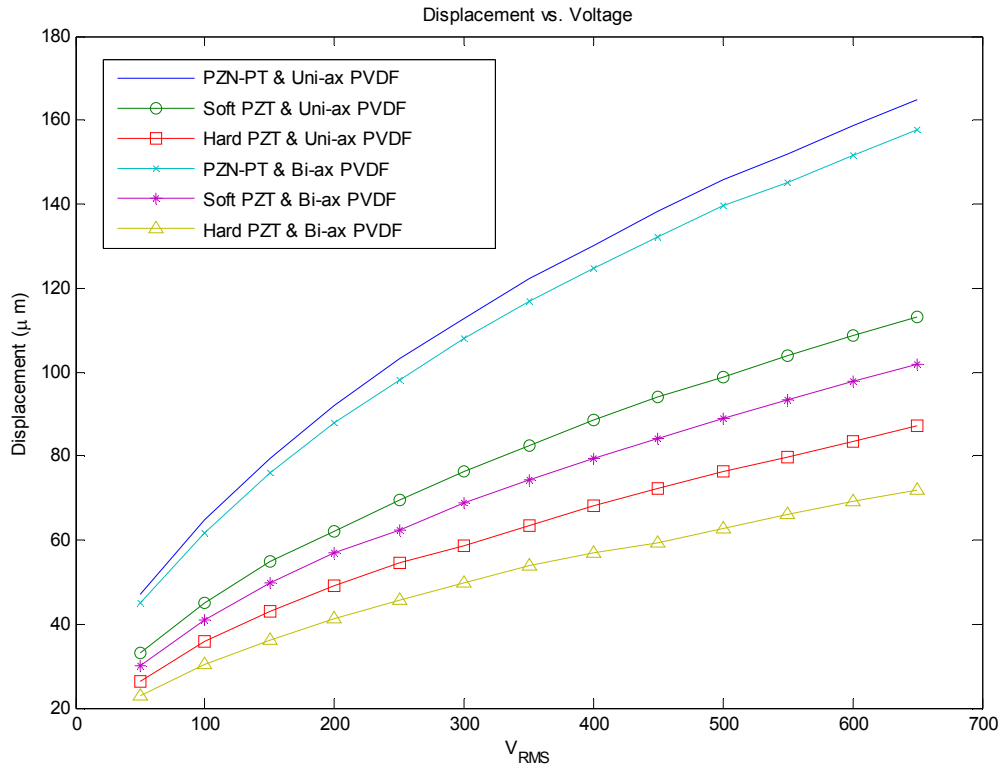
Fig. 7. Displacement vs. RMS Voltage for all material combinations, both components active.

6. CONCLUSIONS AND FUTURE WORK

A computational model was created based on an actuator concept developed by researchers at NASA-Langley Research Center. The model was developed such that design parameters could be altered thus facilitating trade studies to be performed. The underlying purpose of this paper was to use this model to explore the usage of different materials via such a trade study. A theoretical model of the device was the foundation on which the computational model was built. Subtle differences, however, between the theoretical and computational models existed in the form of some necessary assumptions and the manner in which some of the parameters are calculated. After comparing predictions of the two models and careful consideration of how the assumptions might affect said predictions, the computational model was

148

deemed reliable and the trade study was carried out. The ensuing results showed the most suitable material combination was PZN-PT single crystal and uni-axially stretched PVDF.

As mentioned before, this material combination is superior because its constituents have the highest piezoelectric constants in their respective categories of positive and negative strain components and displacement is the only performance parameter being considered here. And while this conclusion may have been obvious from the beginning, maximum displacement is not always the only goal. Even when it is one of many goals, it may not always be the most important. This work serves as a building block for further HYBAS design refinement. Future work should implement more parameters used to select the best material for the application. Parameters such as material cost and availability, fatigue limit and cycles until failure, and electrical power consumption could be included. Further considerations may also involve not only maximum displacement but displacement resolution as well. All of these parameters, including maximum displacement, are application dependent. If the HYBAS is to be tailored to an application as intended, then many of these factors need to be included.

for making this research possible. Also, the authors would like to thank the Morpheus Laboratory for their ongoing support and motivation.

REFERENCES

1. K. Uchino, *Ferroelectric Devices*, Marcel Dekker, Inc., New York, 2000.
2. K. Uchino, *Piezoelectric Actuators and Ultrasonic Motors*, Kluwer Academic Publishers, Boston, 1997.
3. R.E. Newnham, A. Dogan, Q. C. Xu, K. Onitsuka, J. Tressler, S. Yoshikawa, "Flextensional Moonie Actuators," *1993 Ultrasonics Symposium*, IEEE, 509-513.
4. C. Niezrecki, D. Brei, S. Balakrishnan, A. Moskalik, "Piezoelectric Actuation: State of the Art," *The Shock and Vibration Digest*, **33**, No. 4, July 2001, 269-280.
5. I. Chopra, "Review of State of Art of Smart Structures and Integrated Systems," *AIAA Journal*, **40**, No. 11, November 2002, 2145-2187.
6. J. Su, T. B. Xu, S. Zhang, T. R. Shrout, Q. Zhang, "An electroactive polymer-ceramic hybrid actuation system for enhanced electromechanical performance," Applied Physics Letters, **86**, No. 6, August 2004.
7. T. B. Xu, J. Su, "Theoretical modeling of electroactive polymer-ceramic hybrid actuation systems," *Journal of Applied Physics*, **97**, 2005.
8. R. J. Meyer, Jr., A. Dogan, C. Yoon, S. M. Pilgrim, R. E. Newnham, "Displacement amplification of electroactive materials using the cymbal flextensional transducer," *Sensors and Actuators A*, **87**, No. 3, January 2001, pgs. 157-162.
9. J. Su, T. B. Xu, S. Zhang, T. R. Shrout, Q. Zhang, "A Hybrid Actuation System (HYBAS) and Aerospace Applications," *Materials Research Society Symposium V 2005 Fall Proceedings*, **888**, 2006.
10. R. A. Walsh, *Electromechanical Design Handbook*, Mcgraw-Hill, New York, 2000, p. 5.34.
11. TRS Technologies, Inc., http://www.trstechnologies.com/, accessed on February 16, 2007.
12. Goodfellow Corporation, http://www.goodfellow.com/csp/active/gfHome.csp, accessed on February 16, 2007.
13. W. Hackenberger, TRS Technologies, Inc., private communication, February 22, 2007.

Appendix B. Actuator Material Info from Manufacturer

The content of [37], manufacturer's guidelines to the customer for actuator use.

TRS Technologies, Inc.
2820 East College Ave., State College, PA 16801
814-238-7485, fax: 814-238-7539
info@trstechnologies.com

Use Guidelines for TRS Single Crystal Based Stacked Actuators

Although TRS piezoelectric single crystals possess exceptional properties compared to conventional ceramic materials, the crystals require special consideration to ensure that the high performance properties can be realized. Due to the unique nature of the single crystal materials, TRS would like to provide some guidelines for operating and driving the single crystal based stack actuators. The guidelines are summarized in Table I with a more in depth discussion below.

Table I. Guideline Summary Table

Maximum Unipolar Electric Field	< 25 kV/cm
Maximum Negative Electric Field	< 2 kV/cm
Maximum Temperature	<70°C
Maximum Tensile Stress	0 MPa
Maximum Compressive Stress	20 MPa

Physical/Handling

The single crystals are mechanically weaker than polycrystalline ceramics due to the fact that there are no grain boundaries, which can help to arrest a propagating crack. Therefore the crystals require additional handling care. If dropped they may break very easily.

Driving Conditions

The stacks should not be driven at a unipolar electric field greater than 25 kV/cm due to the internal stresses generated within the crystal as well as the morphotropic phase boundary (MPB). At fields approaching 30 kV/cm the crystals will pass through an electric field induced phase transition from rhombohedral to tetragonal and the crystals can partially depole causing a drop in properties. The crystals also have a relatively low coercive field ($E_c \sim 2.5$ kV/cm), which limits the maximum negative field, which can be applied in bipolar drive conditions. Due to the mechanical weakness of the crystals they should not be placed in tensile stress conditions. Under compressive loads the crystals behave similarly to ceramics, but it is recommended that the maximum applied compressive stress be limited to <20 MPa. The crystals also have a very large dielectric constant and therefore it is important to realize that under dynamic driving conditions (i.e. high frequency and high electric fields) a higher current may be observed and your application configuration may require the proper drive electronics to handle these conditions.

Temperature

Again due the presence of the MPB (see above) the crystals are also temperature limited. If the crystals are heated above approximately 80-90°C, they will pass through the phase transition from rhombohedral to tetragonal and partially depole causing a drop in properties. Thermal shock should also be avoided if possible due to the possibility of causing micro-cracking in the crystals, which can further degrade the performance of the crystals.

Contacts for Technical Information:
Xiaoning Jiang, xiaoning@trstechnologies.com
Paul W. Rehrig, paul@trstechnologies.com

Appendix C.   Piezoelectric Single Crystal Materials

The need for devices such as micropositioning actuators (linear and rotary), vibration sensors and suppressors, and others have driven the research and development of piezoelectric materials, as they have the capability to fulfill the demands of these applications. Relaxor-based ferroelectric ceramics materials have been found suitable for actuator applications. More specifically, the single crystal versions of lead magnesium niobate-lead titanate (PMN-PT) and lead zirconate niobate-lead titanate (PZN-PT) with morphotropic phase boundary (MPB) composition are attractive due to high piezoelectric constants [1, 10, 38-41].

A conventionally used piezoceramic material is lead zirconate titanate (PZT) [10]. The strain behavior of PZT as compared to PMN-30%PT is shown in Figure C.1. In this case, both show approximately linear behavior, but the PMN-30%PT shows nearly five times the displacement at the maximum driving field level [41]. The d_{33} piezoelectric constant for a common PZT material is 330 pm/V, while that for a common PMN-32%PT single crystal is 2000 pm/V.

Single crystal materials, like typical piezoceramic materials, exhibit some level of hysteretic behavior. In the case of PMN-x%PT, the level of hysteresis varies with the percentage x of PT. Figure C.2 illustrates this variation for the approximate MPB range of PT content. The level of hysteresis increases up to 35% PT, after which both hysteresis and strain values decrease rapidly [41].
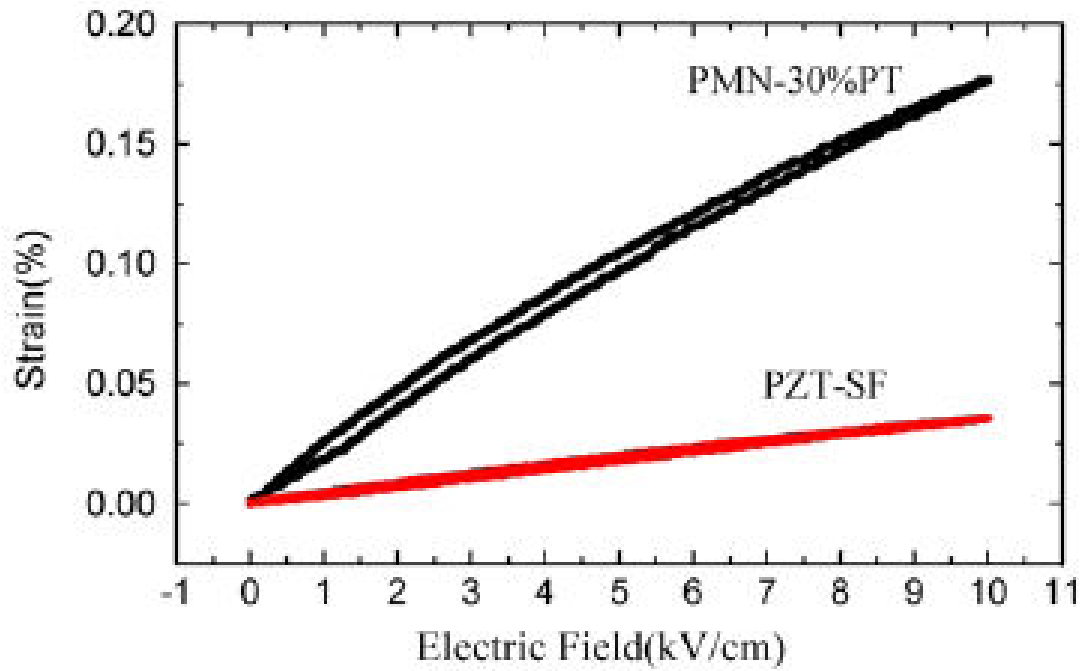
Figure C.1   Strain vs. electric field for PMN-PT and PZT, same driving field. From [41].
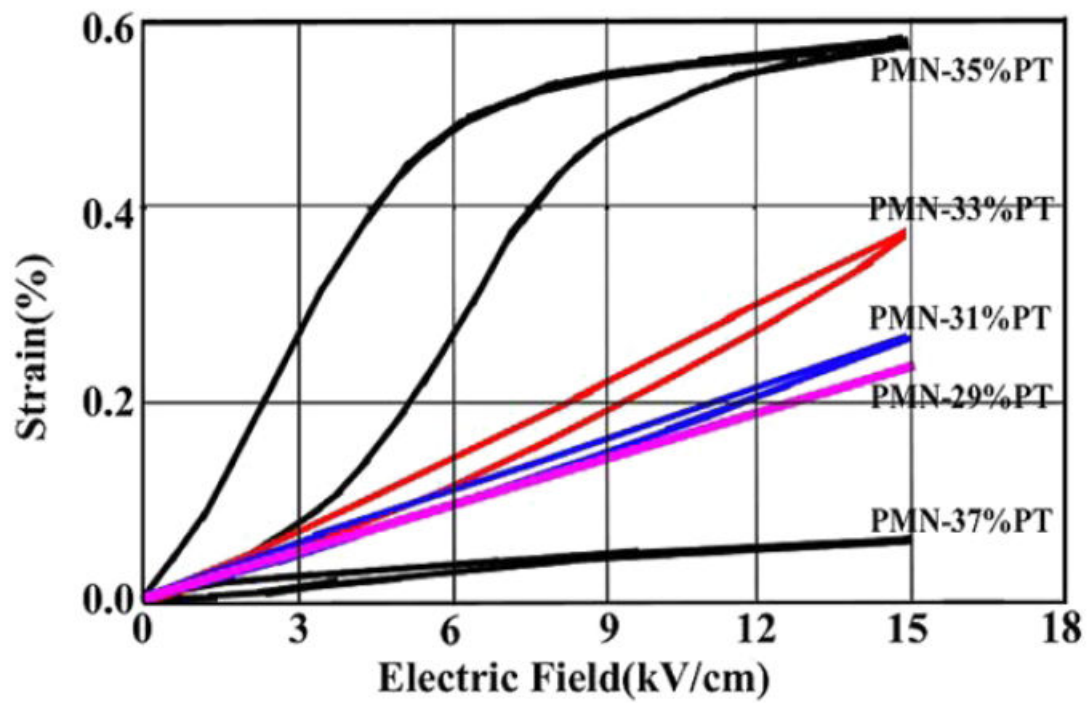
Figure C.2   Dependence of hysteresis on PT content, from [41].

153

In addition to superior strain capabilities, single crystal materials offer performance increases in bandwidth, temperature range, and power consumption. Compared to PZT, single crystals have shown a bandwidth increase of approximately 20%. The displacement capability of PZT drops considerably at cryogenic temperatures. PMN-32%PT will perform at cryogenic temperatures with performance similar to that of PZT at room temperature. Furthermore, with increased displacement capability, a single crystal actuator will consume less power to achieve the same displacement level as conventional piezoceramics [38].

Single crystal materials such as PZN-x%PT and PMN-x%PT offer several performance benefits over conventional piezoceramics like PZT. Their characteristics and manufacture are areas that continue to be investigated.

Appendix D.   Laser Vibrometer Decoder Specifications

From page 7-2 of [43].

7.1.4 Signal Voltage Output VELOCITY OUTPUT

General Data

Output swing:	$20V_{p-p}$
Output impedance:	50Ω
Minimum load resistance:	$10k\Omega$ (-0.5% additional error)
Overrange indicator threshold:	typ. 95% of full scale
Maximum DC offset:	$\pm20mV$

Measurement Ranges

Velocity decoder	Measurement range (scaling factor) $\frac{mm}{s}/V$	Full scale output (peak-peak) $\frac{mm}{s}$	Resolution[1] $\frac{\mu m}{s}$	Maximum frequency[2] kHz	Maximum acceleration g
OVD-01	1	20	0.3	20	150
	5	100	0.6	50	1,600
	25	500	0.8	50	8,000
	125	2,500	1.0	50	25,000
	1,000	20,000	2.0	50	200,000
OVD-02	5	100	0.5	250	8,000
	25	500	1.5	1,500	240,000
	125	2,500	2.0	1,500	1,200,000
	1,000	20,000	5.0	1,500	9,600,000

[1] Resolution is defined as the signal amplitude (rms) at which the signal-to-noise ratio is 0dB in a 10 Hz spectral bandwidth (RBW), measured at 3M Scotchlite Tape®.

[2] -1dB maximum error

Calibration Accuracy

Velocity decoder	Measurement range $\frac{mm}{s}/V$	Amplitude error	
		@ T = (25 ±5)°C (T = (77 ±9)°F) % of rms reading	in the temperature range 5°C...40°C (41°F...104°F) % of rms reading
OVD-01	1...1,000	±1.0	±1.2
OVD-02	5	±1.0	±1.5
	25	±1.0	±2.0
	125 and 1,000	±1.0	±2.5

Conditions: sinusoidal vibration, f =1 kHz, amplitude 70% of full scale range, load resistance ≥ 1 MΩ

From page 7-5 of [43].

7.1.5 Signal Voltage Output DISPLACEMENT OUTPUT (optional)

General Data

Voltage swing: $16V_{p\text{-}p}$
Output impedance: 50Ω
Minimum load resistance: $10k\Omega$ (-0.5% additional error)

Measurement Ranges

Displacement decoder	Measurement range (scaling factor)	Full scale output (peak-peak)	Resolution[1]	Maximum velocity	Bandwidth	Max. frequency for specified accuracy
	µm/V	µm	µm	m/s	kHz	kHz
OVD-10	20	320	0.08	2.5	0...250	100
	80	1,280	0.32	10	0...250	100
	320	5,120	1.3	10	0...250	100
	1,280	20,480	5.0	10	0...250	100
	5,120	81,920	20	10	0...250	100
OVD-20	0.5	8	0.002	0.06	0...25	10
and	2	32	0.008	0.25	0...75	15
OVD-40	8	128	0.032	1	0...75	25
	20	320	0.08	2.5	0...250	100
	80	1,280	0.32	10	0...250	100
	320	5,120	1.3	10	0...250	100
	1,280	20,480	5.0	10	0...250	100
	5,120	81,920	20	10	0...250	100

[1] The resolution is defined as 1 increment of the fringe counter output which is equivalent to a 4mV output voltage step.

Calibration Accuracy

Amplitude error: $\pm1\%$ of rms reading ±1 increment
Conditions: Sinusoidal vibration, f = 100Hz, amplitude 50% of full scale range, load resistance $\geq 1M\Omega$

Amplitude Linearity

Maximum linearity error: Ranges 20µm/V...5,120µm/V: ±1 increment
    Ranges 0.5µm/V...8µm/V: ±2 increments

Trigger (CLEAR Input)

Threshold: +10mV (typ.), rising edge
Hysteresis: 5mV (typ.)
Maximum input voltage: $\pm14V$
Pulse rate: 40Hz...16kHz
Input impedance: > $0.5k\Omega$ (depending on the pulse frequency)

156

Appendix E.   Vibration Isolation System Specifications

From page *v* of [45].

Specifications

Table Tops:

Flatness:	±0.005in. (0.13mm)*
Compliance:	Consult your Newport Catalog or Newport directly for the specific compliance and other pertinent table top specifications of your particular table top model.

Isolators: Stabilizer™ I-2000

Vertical Resonant Frequency:	<1.1 Hz at 80 psi
Horizontal Resonant Frequency:	<1.5 Hz
Recommended Load Range: (per 4 isolators)	660 to 8,000 lb (300 to 3,600 kg)
Automatic Leveling Accuracy:	±0.010 inch (0.25 mm) standard, higher accuracy available on special order
Vertical Adjustment Range:	1.3 inches (33 mm)
Settling Time: (after 5-lb. weight removal)	<1.5 sec.
Typical Air Pressure Range:	10 to 85 psi (0.7 to 6.0 kg/cm²)

Bibliography

1. Uchino, K., *Piezoelectric Actuators and Ultrasonic Motors*, Kluwer Academic Publishers, Boston, 1997.

2. Devasia, S., Eleftheriou, E., Moheimani, S. O. R., "A Survey of Control Issues in Nanopositioning," *IEEE Transactions on Control Systems Technology*, **15**, No. 5, September 2007.

3. Xu, W. L., Han, L., "Piezoelectric actuator based active error compensation of precision machining," *Measurement Science and Technology*, **10**, No. 2, February 1999, pgs. 106-111.

4. Petivan, J. L., *The Design, Construction, and Evaluation of a Real Time Optical Wavefront Sensor*, Master of Science Thesis, Massachusetts Institute of Technology, May 1988.

5. Cross, J. W., http://www.mobot.org/jwcross/spm/, revised on March 10, 2007, accessed on September 27, 2007.

6. Woronko, A. Huang, J., Altintas, Y., "Piezoelectric tool actuator for precision machining on conventional CNC turning center," *Precision Engineering*, **27**, No. 4, October 2003, pgs. 335-345.

7. Niezrecki, C., Brei, D., Balakrishnan, S., Moskalik, A., "Piezoelectric Actuation: State of the Art," *The Shock and Vibration Digest*, **33**, No. 4, July 2001, pgs. 269-280.

8. Newnham, R. E., "Piezoelectric Sensors and Actuators: Smart Materials," *Proceedings of the 1992 IEEE Frequency Control Symposium*, May 1992, pgs. 513-524.

9. Chopra, I., "Review of State of Art of Smart Structures and Integrated Systems," *AIAA Journal*, **40**, No. 11, November 2002, pgs 2145-2187.

10. Uchino, K., *Ferroelectric Devices*, Marcel Dekker, Inc., New York, 2000.

11. Srinivasan, A. V., McFarland, D. M., *Smart Structures Analysis and Design*, Cambridge University Press, Cambridge, 2001.

12. Ljung, L., Glad, T., *Modeling of Dynamic Systems*, Prentice Hall, Upper Saddle River, New Jersey, 1994.

13. Ljung, L., *System Identification Theory for the User, Second Edition*, Prentice Hall PTR, Upper Saddle River, New Jersey, 1999.

14. Franklin, G. F., Powell, J. D., Emami-Naeini, A., *Feedback Control of Dynamic Systems, 4th Ed.*, Prentice Hall, New Jersey, 2002.

15. Skogestad, S., Postlethwaite, I., *Multivariable Feedback Control Analysis and Design*, John Wiley & Sons, Chichester, 1996.

16. Nise, N. S., *Control Systems Engineering, Fourth Edition*, John Wiley & Sons, Inc., Hoboken, New Jersey, 2004.

17. Dorato, P., Abdallah, C. T., Cerone, V., *Linear Quadratic Control An Introduction*, Krieger Publishing Company, Malabar, Florida, 2000.

18. Standards Committee of the IEEE Ultrasonics, Ferroelectrics, and Frequency Control Symposium, *An American National Standard: IEEE Standard on Piezoelectricity*, The Institute of Electrical and Electronics Engineers, ANSI/IEEE Std. 176-1987, New York, 1988.

19. Hagood, N. W., Chung, W. H., von Flotow, A., "Modelling of Piezoelectric Actuator Dynamics for Active Structural Control," *Journal of Intelligent Material Systems and Structures*, **1**, July, 1990, pgs. 327-354.

20. Liang, C., Rogers, C. A., "Transient Electro-Mechanics of Piezoelectric Actuator-Driven Active Mechanical Systems," *Journal of Intelligent Material Systems and Structures*, **8**, April, 1997, pgs. 374-379.

21. Main, J. A., Garcia, E., "Piezoelectric Stack Actuators and Control System Design: Strategies and Pitfalls," *Journal of Guidance, Control, and Dynamics*, **20**, No. 3, May-June, 1997, pgs. 479-485.

22. Gorman, J. J., Dagalakis, N. G., "Modeling and Disturbance Rejection Control of a Nanopositioner with Application to Beam Steering," *Proceedings of the ASME International Mechanical Engineering Congress and Exposition*, Washington D. C., 2003.

23. Salapaka, S, Sebastian, A., "Control of the Nanopositioning Devices," *Proceedings of the 42nd IEEE Conference on Decision and Control*, **3**, December, 2003.

24. Schitter, G., Allgower, F., Stemmer, A., "A new control strategy for high-speed atomic force microscopy," *Nanotechnology*, **15**, January, 2004, pgs. 108-114.

25. Ge, P., Jouaneh, M., "Tracking Control of a Piezoceramic Actuator," *IEEE Transactions on Control Systems Technology*, **4**, No. 3, May, 1996, pgs. 209-216.

26. Song, G., Zhao, J., Zhou, X., De Abreu-Garcia, J. A., "Tracking Control of a Piezoceramic Actuator With Hysteresis Compensation Using Inverse Preisach Model," *IEEE/ASME Transactions on Mechatronics*, **10**, No. 2, April, 2005, pgs. 198-209.

27. Ru, C., Sun, L., "Improving positioning accuracy of piezoelectric actuators by feedforward hysteresis compensation based on a new mathematical model," *Review of Scientific Instruments*, **76**, September, 2005.

28. Croft, D., Shed, G., Devasia, S., "Creep, Hysteresis, and Vibration Compensation for Piezoactuators: Atomic Force Microscopy Applications," *Journal of Dynamic Systems, Measurement, and Control*, **123**, No. 1, March, 2001, pgs. 35-43.

29. Lee, S. H., Royston, T. J., Friedman, G., "Modeling and Compensation of Hysteresis in Piezoceramic Transducers for Vibration Control," *Journal of Intelligent Material Systems and Structures*, **11**, October, 2000.

30. Goldfarb, M., Celanovic, N., "A Lumped Parameter Electromechanical Model for Describing the Nonlinear Behavior of Piezoelectric Actuators," *Journal of Dynamic Systems, Measurement, and Control*, **119**, No. 3, September, 1997, pgs. 478-485.

31. Bashash, S., Jalili, N., "Robust Multiple Frequency Trajectory Tracking Control of Piezoelectrically Driven Micro/Nanopositioning Systems," *IEEE Transactions on Control Systems Technology*, **15**, No. 5, September, 2007.

32. Hughes, D., Wen, J. T., "Preisach modeling of piezoceramic and shape memory alloy hysteresis," *Smart Materials and Structures, Instutute of Physics Publishing*, **6**, No. 3, 1997, pgs 287-300.

33. Smith, R. C., Bouton, C., Zrostlik, R., "Partial and Full Inverse Compensation for Hysteresis in Smart Material Systems," *Proceedings of the American Control Conference*, **4**, Chicago, June 2000, pgs. 2750-2754.

34. Newnham, R. E., Dogan, A., Xu, Q. C., Onitsuka, K., Tressler, J., Yoshikawa, S., "Flextensional Moonie Actuators," *Ultrasonics Symposium, 1993, Proceedings of the IEEE 1993*, **1**, October-November, 1993, pgs. 509-513.

35. Meyer, R. J. Jr., Dogan, A., Yoon, C., Pilgrim, S. M., Newnham, R. E., "Displacement amplification of electroactive materials using the cymbal

flextensional transducer," *Sensors and Actuators A: Physical*, **87**, No. 3, January 2001, pgs. 157-162.

36. Xu, T. B., Su, J., Jiang, X., Rehrig, P. W., Hackenberger, W. S., "The load capability of piezoelectric single crystal actuators," *Materials Research Society Symposium Proceedings*, **888**, 2006, pgs. 145-150.

37. Jiang, X., Rehrig, P. W., "Use Guidelines for TRS Single Crystal Based Stack Actuators," TRS Technologies, Inc., Customer Reference Document, http://www.trstechnologies.com/pdf/Crystal_Guidelines.pdf, accessed on May 15, 2008.

38. Woody, S. C., Smith, S. T., Jiang, X., Rehrig, P. W., "Performance of single-crystal $Pb(Mg_{1/3}Nb_{2/3})$-32%$PbTiO_3$ stacked actuators with application to adaptive structures" *Review of Scientific Instruments*, **76**, July 2005.

39. Liu, S. F., Park, S. E., Cross, L. E., Shrout, T. R. "Temperature dependence of electrostriction in rhombohedral $Pb(Zn_{1/3}Nb_{2/3})O_3$-$PbTiO_3$ single crystals," *Journal of Applied Physics*, **92**, No. 1, July 2002, pgs. 461-467.

40. Liu, S. F., Park, S. E., Shrout, T. R., Cross, L. E., "Electric field dependence of piezoelectric properties for rhombohedral $0.955Pb(Zn_{1/3}Nb_{2/3})O_3$-$0.045PbTiO_3$ single crystals," *Journal of Applied Physics*, **85**, No. 5, March 1999, pgs. 2810-2814.

41. Feng, Z., He, T., Haiqing, X., Luo, H., Yin, Z., "High electric-field-induced strain of $Pb(Mg_{1/3}Nb_{2/3})$-$PbTiO_3$ crystals in multilayer actuators," *Solid State Communications*, **130**, No. 8, 2004, pgs. 557-562.

42. Trek, Inc., *Operator's Manual Model 609E-6 High-Voltage Amplifier*, Revision B1, July 2001.

43. Polytec, *Laser Doppler Vibrometer User Manual*, Controller OFV-3001 and Sensor Heads OFV-303/-353 and OFV-511/-512.

44. Newport Corporation, *Seminar Notes: Compliance and Transmissibility Curves*, http://www.newport.com/Seminar-Notes-Compliance-and-Transmissibility-Cur/154650/1033/catalog.aspx, accessed on November 18, 2007.

45. Newport Corporation, *Vibration Control System Instruction Manual*, 1991.

46. Lillehei, P., NASA-Langley Research Center Advanced Materials and Processing Branch, *Personal Communication*, November 5, 2007.

47. Wong, G., Newport Corporation, Technical Support, November 13, 2007.

48. Stearns, R. C., Newport Corporation, Sr. Staff Applications Engineer, Compliance curves for RS4000-46-8 optical table, sent via e-mail, November 14, 2007.

49. Tischler, M. B., Remple, R. K., *Aircraft and Rotorcraft System Identification Engineering Methods with Flight Test Examples*, AIAA, Reston, VA, 2006.

50. Bendat, J. S., Piersol, A. G., *Random Data Analysis and Measurement Procedures, Third Edition*, John Wiley & Sons, Inc., New York, 2000.

51. Juang, J. N., Phan, M. Q., *Identification and Control of Mechanical Systems*, Cambridge University Press, New York, 2001.

52. Meirovitch, L., *Fundamentals of Vibrations*, McGraw-Hill, New York, 2001.

53. Stoica, P., Moses, R., *Introduction to Spectral Analysis*, Prentice Hall, Upper Saddle River, New Jersey, 1997.

54. Yun, K., Kim, W. J., "Microscale position control of an electroactive polymer using an anit-windup scheme," *Smart Materials and Structures*, **15**, No. 4, August 2006, pgs. 924-930

55. The Mathworks, Inc., Matlab R2007a System Identification Toolbox version 7.0 help files, *Model Output Plots: Plotting and Validating Models (System Identification Toolbox*, 2007.

56. The Mathworks, Inc., Matlab R2007a System Identification Toolbox version 7.0 help files, *Introduction to System Identification Toolbox: Estimating Dynamic Models*, 2007.

57. Ogata, K., *Discrete-Time Control Systems Second Edition*, Prentice Hall, Upper Saddle River, New Jersey, 1995.

58. Ogata, K., *Modern Control Engineering Fourth Edition*, Prentice Hall, Upper Saddle River, New Jersey, 2002.

59. Inman, D. J., *Vibration With Control, Measurement, and Stability*, Prentice Hall, New Jersey, 1989.

60. Sahraian, M., Kodiyalam, S., "Tuning PID controllers using error-integral criteria and numerical optimization," *Proceedings of 6th AIAA, NASA, and ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Bellvue, WA, Sept. 4-6, 1996, American Institute of Aeronautics and Astronautics, 1996, pgs. 237-236.

61. Zhuang, M., Atherton, D. P., "Tuning PID controllers with integral performance criteria," International Conference on Control 1991, March 25-28, 1991, pgs.

62. Lee, J. Edgar, T. F., "ISE tuning rule revisited," *Automatica*, **40**, No. 8, August 2004, pgs. 1455-1458.

63. Cominos, P., Munro, N., "PID controllers: recent uning methods and design to specification," *IEE Proceedings-Control Theory and Applications*, **149**, No. 1, June 2002, pgs. 46-53.

64. Horowitz, P., Hill, W., *The Art of Electronics, Second Edition*, Cambridge University Press, Cambridge, 1989.

65. Su., J., Xu, T. B., Zhang, S., Shrout, T. R., Zhang, Q., "An electroactive polymer-ceramic hybrid actuation system for enhanced electromechanical performance," *Applied Physics Letters*, **86**, No. 6, August 2004.

66. Xu, T. B., Su., J., "Theoretical modeling of electroactive polymer-ceramic hybrid actuation systems," *Journal of Applied Physics*, **97**, 2005.

67. Su., J., Xu, T. B., Zhang, S., Shrout, T. R., Zhang, Q., "A Hybrid Actuation System (HYBAS) and Aerospace Applications," *Materials Research Society Symposium V, 2005 Fall Proceedings*, **888**, 2006.

68. Nickless, B. J., Su, J., Xu, T. B., Hubbard, J. E., Jr., "A Computational Method for Materials Selection in a Hybrid Actuation System," *Proceedings of SPIE*, **6529,** April 6, 2007.

69. Lindelhof, J., SPIE, *Personal Communication*, June 4, 2008, publisher's permission to reprint granted via e-mail.