

Modeling and Control of a Six Degree of Freedom Magnetic / Fluidic Motion Control Stage

by

Stephen Joseph Ludwick, Jr.

Bachelor of Science in Mechanical Engineering
& Engineering and Public Policy
Carnegie Mellon University
May 1994

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

Master of Science in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 1996

© 1996 Massachusetts Institute of Technology
All rights reserved

Signature of Author.....

Department of Mechanical Engineering
January 31, 1996

Certified by.....

David L. Trumper
Assistant Professor of Mechanical Engineering
rvisor

Accepted by.....

Ain A. Sonin
Chairman, Department Committee on Graduate Students

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

MAR 19 1996

ARCHIVES

LIBRARIES

Modeling and Control of a Six Degree of Freedom Magnetic / Fluidic Motion Control Stage

by

Stephen Joseph Ludwick, Jr.

Submitted to the Department of Mechanical Engineering
on January 31, 1996, in partial fulfillment of the requirements for the degree of
Master of Science in Mechanical Engineering

Abstract

This thesis describes a six degree of freedom positioning stage for fine motion control. The stage achieves a 6σ positioning noise of 0.3 nm at a 1 Hz measuring bandwidth and a controller bandwidth of 5 Hz. Total travel is within a cube of 100 μm . The single moving element is immersed in oil, forming squeeze film dampers between itself and the frame. This design results in a highly overdamped and vibration resistant system. Twelve electromagnets provide the forces necessary to suspend and servo the platen while six capacitance probes sense the position. All of the control algorithms are performed digitally, using a PC-based digital signal processing board. The controller performs two functions that are essential to achieving 0.1 nanometer positioning resolution. First, it uses a detailed model of the stage in a feedback linearization scheme that linearizes and decouples the degrees of freedom. Second, it uses a combination of a digital filter and an estimator to reduce the effect of measurement noise by about two orders of magnitude. The end result is a stage that is suitable for positioning a sample with better than atomic resolution. Envisioned applications of the stage include producing the scanning motions required in scanned probe microscopy or as a motion control stage for integrated circuit metrology.

Thesis Supervisor: David L. Trumper

Title: Assistant Professor of Mechanical Engineering

Contents

1	Introduction	11
1.1	Project Goal	11
1.2	Thesis Summary	13
1.3	Background	13
1.3.1	Scanned Probe Microscopy	13
1.3.2	Examples of Precision Motion Control Devices	15
1.3.3	Prior Uses of Fluid Flotation	17
2	Hardware	21
2.1	Stage Overview	21
2.2	Metrology Loop	24
2.2.1	Capacitance Probes	25
2.2.2	Anti-Aliasing Filter	26
2.2.3	Analog-to-Digital Conversion Board	28
2.3	Force Loop	28
2.3.1	Digital-to-Analog Conversion Board	28
2.3.2	Power Amplifiers	29
2.3.3	Electromagnetic Actuators	31
3	Mechanical Modeling	33
3.1	Introduction	33
3.2	Nomenclature	34
3.3	Fluid Dampers	35
3.3.1	Circular Squeeze Film Dampers (Capacitance Probes)	36
3.3.2	Rectangular Squeeze Film Dampers (Actuators)	36
3.3.3	Shear Damping	38
3.3.4	Combined Damping Model in Six Degrees-of-Freedom	39
3.4	Electromagnetic Actuators	42
3.4.1	Electromagnetics Nomenclature	42
3.4.2	Terminology / Basic Electromagnetics	43
3.4.3	Derivation of Force and Torque for a Non-Uniform Gap	46
3.5	Kinematics	50
3.5.1	Measurement Sensitivity Matrix	51
3.5.2	Modal Force Transformation	51
3.6	Gap Estimation	53

4	Noise Propagation and Modeling	55
4.1	Introduction	55
4.2	Basic Stochastics	56
4.3	Noise Propagation in One Degree-of-Freedom	58
4.3.1	Propagation of Sensor Noise	59
4.3.2	Propagation of Disturbance Forces	62
4.3.3	Combined Noise Propagation Model	63
4.4	Sensor Noise	67
4.4.1	Capacitance Probe Noise	67
4.4.2	A/D Converter Quantization Noise	68
4.4.3	A/D Converter Electrical Noise	69
4.4.4	Bandlimiting of Sensor Noise	72
4.5	Plant Noise	73
4.5.1	Disturbance Forces Due to Base Vibrations	73
4.5.2	D/A Converter Quantization Noise	81
4.6	Final Predicted Position Variance	83
5	Filtering	85
5.1	Arithmetic Average	85
5.2	Linear Regression	86
5.3	FIR Filter	87
6	Control System Design	93
6.1	Open Loop Plant	93
6.2	Continuous Time PI Control	94
6.2.1	Selection of Gains	95
6.2.2	Control Results	96
6.2.3	Effect of the Changes in Damping	100
6.3	Feedback Linearization	102
6.3.1	Electromagnet Linearization	102
6.3.2	Damper Linearization	103
6.3.3	Combined Linearized Plant	104
6.4	Velocity Control	105
6.4.1	Measuring Velocity	106
6.4.2	Control Algorithm	108
6.4.3	Trajectory Following	109
6.5	Discrete Time Plant Model	110
6.6	Estimators	111
6.6.1	Background	112
6.6.2	Position Estimator	113
6.6.3	Bias Estimator	115
6.7	Kalman Filter	118
6.7.1	Minimizing the Squared Error	120
6.7.2	Calculation of the State Error Covariance Matrices	124
6.7.3	Time and Measurement Updates	126

6.7.4	Steady State Solution	129
6.7.5	Application to the Angstrom Stage	130
6.8	Final Control System Design	131
7	Results	133
7.1	Positioning Capabilities	134
7.2	Decoupling of the Six Degrees of Freedom	136
7.3	Velocity Control	137
7.3.1	Feedback of Measured Velocity	137
7.3.2	Updating Commanded Position (Trajectory Following)	140
8	Conclusions	143
8.1	Design Strong Points	143
8.2	Design Weak Points	143
8.3	Suggestion for Further Work	144
8.4	Acknowledgements	145
9	Bibliography	147
A	Derivation of the Damping Matrix	153
A.1	Lateral Motions	153
A.2	Vertical Motions	155
A.3	Combined Model	158
B	Kinematics	159
B.1	Measurement Sensitivity Matrix	159
B.2	Modal Force Transformation	160
B.3	Gaps	162
B.3.1	Actuator Gaps	162
B.3.2	Capacitance Probe Gaps	163
C	Low Pass FIR Filter Coefficients	165
D	Estimator Gain Setting	169
E	Derivation of Discrete Time Noise Propagation Model	171
F	Closing Thoughts	177

List of Figures

2-1	Photograph of the Angstrom Stage on a benchtop vibration isolation table.	22
2-2	Cross sectional view of the Angstrom Stage (from Holmes).	23
2-3	Isometric view of the Angstrom Stage (from Holmes).	23
2-4	Lightweighted platen (from Holmes).	24
2-5	Schematic of a capacitance probe (from Holmes).	25
2-6	Photograph of one of the capacitance probes (penny shown to provide scale).	25
2-7	Schematic diagram of the anti-aliasing filter.	26
2-8	Measured frequency response of the anti-aliasing filter.	27
2-9	Beating effect possibly due to time skew in DAC outputs.	29
2-10	Schematic diagram of a power amplifier (from Holmes).	30
2-11	Drawing of an unpotted electromagnetic actuator (from Holmes).	31
2-12	Photograph of a potted electromagnetic actuator and one lamination with a penny shown to provide scale.	32
3-1	Shape factor for the damping between rectangular plates.	38
3-2	Damping coefficients for a single actuator and capacitance probe	38
3-3	Location of probes and actuators relative to the platen (from Holmes).	40
3-4	Damping coefficients as the stage moves in the translational directions.	41
3-5	Damping coefficients as the stage moves in the rotational directions.	41
3-6	Electromagnetic actuator at a non-uniform gap.	47
3-7	Percent error in the applied force due to actuator rotation.	50
3-8	Torque applied by an actuator due to actuator rotation.	51
3-9	Positions of the force vectors located about the platen (from Holmes).	52
4-1	Free body diagram of a single degree-of-freedom model under proportional control.	59
4-2	Linear block diagram of a single degree-of-freedom model.	59
4-3	Propagation of the sensor noise according to the controller gain setting.	62
4-4	Propagation of the disturbance forces according to the controller gain setting.	64
4-5	Net position variance as a sum of sensor noise and disturbance force transmissions.	64
4-6	Loop transmission magnitude at the noise-optimal gain.	66
4-7	Histogram of the reading from a grounded A/D converter channel.	70
4-8	Autocorrelation of the A/D converter noise data.	71
4-9	Power spectral density plot of the A/D converter noise data.	71
4-10	Free body diagram of the platen-frame system as by base vibrations.	74

4-11	DFT of the floor accelerations from 0–400 Hz.	76
4-12	Time plot of the floor accelerations.	77
4-13	DFT of the floor accelerations from 0–50 Hz.	77
4-14	DFT of the table accelerations from 0–400 Hz.	78
4-15	DFT of the table accelerations from 0–50 Hz.	79
4-16	Time plot of the table accelerations showing high frequency component. .	79
4-17	Time plot of the table accelerations showing low frequency component. .	80
4-18	Two DFT's of the optical table accelerations.	80
4-19	Measured and fitted power spectral density plots of disturbance forces. .	81
4-20	Pie chart showing contribution of each noise source to the position error. .	84
5-1	Frequency response of a moving average filter ($N = 100$).	87
5-2	FIR filtering and downsampling (each of 6 channels).	88
5-3	Impulse response of the input FIR filter ($N = 100$).	89
5-4	Frequency response of the redesigned filter ($N = 100$).	89
5-5	Unfiltered readings from a grounded A/D converter channel.	90
5-6	A/D converter readings filtered with a 100 point moving average.	91
5-7	A/D converter readings filtered with the redesigned FIR filter.	91
5-8	256-pt DFT's of unfiltered, averaged, and filtered A/D noise waveforms. .	92
6-1	Block diagram of a linear PI control scheme.	95
6-2	Continuous time root locus of plant under PI control.	96
6-3	Simulated position step response with the PI controller.	97
6-4	Frequency response of the plant under PI control.	98
6-5	Disturbance force transmission of the PI controlled system.	99
6-6	Gain plots of the pole locations as the damping changes.	101
6-7	Block diagram of the feedback linearization scheme.	105
6-8	Simplified block diagram after feedback linearization.	105
6-9	Predicted and measured step responses with a proportional gain of 5 s^{-1} . .	106
6-10	Frequency response of the velocity filter.	108
6-11	Block diagram of the velocity control structure.	109
6-12	Block diagram of a basic estimator.	112
6-13	Example of estimator convergence.	115
6-14	Block diagram of a bias estimator.	118
6-15	Block diagram of the overall control structure.	132
7-1	Steady state position noise measured through a 1 Hz filter.	135
7-2	Repeated 0.5 nm Step Responses (5 Hz Controller Bandwidth).	135
7-3	An example of a 0.2 nm step response (0.16 Hz controller bandwidth). . .	136
7-4	Error motions in the other axes after a 50 nm X-direction step.	138
7-5	A comparison of scans with both open and closed loop control.	139
7-6	Residual errors from a best fit straight line to the entire scan.	139
7-7	Residual errors from a best fit straight line to a partial scan.	140
7-8	Residual errors from a best fit straight line to a trajectory generated by updating the position setpoint.	141

A-1	Free body diagram of the lateral degrees of freedom of the platen. . . .	154
A-2	Free body diagram of the vertical degrees of freedom of the platen. . . .	156

List of Tables

2.1	Capacitance probe gains.	26
2.2	Amplifier gain factors.	29
3.1	Revised estimates of the null gaps.	54

Chapter 1

Introduction

1.1 Project Goal

The Angstrom Stage project was begun by Dr. David Trumper while he was a faculty member at the University of North Carolina at Charlotte. The goal of this project is to explore the use of electromagnetic actuators in combination with an oil flotation system for motion control applications at the sub-nanometer level. The envisioned application of this stage is in the area of scanned probe microscopy, and the original design specifications set for the stage were [54]:

- Control of six degrees of freedom with no mechanical contact and only a single moving element (the platen).
- Viscous damping of platen motions.
- One angstrom (0.1 nm) resolution with a travel of 100 μm in all three linear degrees of freedom.
- Demonstrate techniques which can in principle be scaled to much larger range of motion.
- Peak acceleration of 5 m/s^2 .
- Dissipate less than 100 mW quiescently and less than 1 W when operating at full acceleration.
- Design sensor system with sub-angstrom resolution and 100 μm linear range which is capable of measuring all six platen degrees of freedom.

- Develop appropriate multivariable control theory.
- Construct prototype to demonstrate performance goals.
- Demonstrate atomic resolution scanning tunneling microscope (STM) images where scanning motion is generated by the magnetic suspension stage.

Most of these design goals have already been met by the individuals originally involved with the design and construction of the stage. Michael Holmes [26] was responsible for the overall design and software work. Under controlled conditions he was able generate atomic resolution images [27]. However, achieving this level of resolution came at the expense of range and bandwidth. Amplifiers with a gain of 1000 along with 1 Hz lowpass filters on the output of the capacitance probes increased the signal-to-noise ratio, but reduced the range to a cube of 100 nm. These modifications to the system proved the validity of the design and pointed to the data acquisition system as the main limitation. In particular, the analog-to-digital conversion board which transduces capacitance probe readings is much noisier than expected. At present, in his doctoral work, Mike is developing another oil-floated stage to again have angstrom resolution but with much larger travel.

There were also several others involved with the design and construction of the Angstrom Stage. David Batchelder [1] designed and calibrated capacitance probes for operation in the oil. No commercially available product had the required million to one dynamic range. The ADE Corporation¹ was particularly helpful in this design step. Roxana Behrouzjou [3] worked on the general design and specifically concentrated on the STM interfacing issues. Alexander Kuzin, a visiting Russian scientist, performed the initial analysis for a noise optimal controller [34]. Mark Williams, a doctoral student, built the electromagnetic actuators [57], and John Murphy, an undergraduate, performed much of the machining work.

This thesis concerns achieving the two remaining unmet design goals, i.e. the demonstration of atomic resolution over the entire 100 μm cube of travel, and the development of an appropriate multivariable control system. The second goal clearly complements the first, since the control system is designed with angstrom resolution as the top priority.

¹Newton, MA

1.2 Thesis Summary

This thesis begins with an overview of scanned probe microscopy and a look at some of the mechanisms that can provide the level of motion control it requires. Chapter 2 details the hardware components of the stage. Some of this information is found in greater detail in Holmes [26], but is included for completeness here. A detailed mathematical model of the stage follows in Chapter 3, and this will later be a key element in the control system design. Chapter 4 contains the models of the noise sources, and Chapter 5 describes the filters used to lessen their effect. The control system for the stage is developed in Chapter 6. This chapter describes the design of a feedback linearization algorithm and an estimator that enables sub-nanometer resolution over the entire 100 μm range of travel. Chapter 7 contains the results of the controller design, and demonstrates the positioning capabilities of the stage. Finally, Chapter 8 concludes the thesis with some thoughts on what has been learned from undertaking this design.

1.3 Background

1.3.1 Scanned Probe Microscopy

In general, most scanned probe microscopes work by rastering a probe across the surface of a sample and measuring some parameter that varies with proximity to the surface. Repeated scans then produce a measure of the surface topology. The first in this family of microscopes was the Topografiner, developed by Young [61] in 1972. It was actually an extension of his Field Emission Ultramicrometer [60] with the addition of scanning capabilities. The basic idea behind these devices is that a fine tip is brought close to the surface of an electrically conductive sample. When a voltage is applied to the tip, a current flows between it and the sample. This current is extremely sensitive to the gap, and feedback is generally used to maintain a constant current through varying the tip position. As the tip is scanned over the surface of a sample, this feedback causes it to follow the surface contours. In this way, repeated scans can be used to generate a topographic image of the sample.

Binnig and Rohrer [5] expanded upon this work to produce the Scanning Tunneling Microscope (STM). They received a Nobel Prize for this work in 1985. The STM was capable of imaging atoms, but only on electrically conductive surfaces. This requirement was eliminated through the development of the Atomic Force Microscope (AFM) [4]. In the AFM, a probe is attached to a cantilevered beam which bends due to interatomic forces when it is brought into close proximity with a surface. The majority of the AFM designs use optical methods to monitor the deflection of the cantilever. This has turned out to be a highly adaptable design, since the basic mechanism can be used to measure characteristics other than interatomic forces. Scanned probe microscopes which measure the electrostatic, magnetic, capacitive, or thermal characteristics of the sample have followed within this basic design. Wickramasinghe [55] has written an overview of these scanned probe methods, and Chetwynd and Smith [9] discuss the evolution of surface metrology from stylus to probe based methods.

Regardless of the surface characteristic being measured, all scanned probe microscopes require some method of producing the associated scanning motions. In order to achieve atomic resolution, such methods must have sub-angstrom resolution and disturbance rejection. A piezoelectric tube is most often used, but these devices necessarily have limited range (on the order of 100 μm maximum). In contrast, the travel of the Angstrom Stage is theoretically limited only by the range of the sensors. Piezoelectrics also have problems with drift, hysteresis, and nonlinearity. These effects can be reduced through charge control [42] or by inserting a capacitor in series with the piezo [30] [29], but a more fundamental problem exists when precise metrological data is required.

Most piezoelectric tube scanners operate open loop. Liu [36] and Griffith [22] demonstrated the accuracy increases that are achievable if capacitive sensors are used to monitor the tube position. Lui used a capacitive sensor to measure the tube extension, and Griffith mounted a capacitive sensor outside the tube to monitor the lateral deflections. These methods have not yet come into widespread use. Even if a closed loop controller is used, the lateral position of the tip cannot be measured directly, and so a reading must be taken at some position further up on the piezoelectric tube. However, the length of the tip is never known exactly, and so neither is this vertical offset. When the tube is

scanned, an estimate of the lateral position of the tip is generated. However, having a vertical offset between the tip and the point where the measurement is taken violates the Abbe principle [52] [48], and leads to a sine error in the traverse length. In addition, producing the scanning motions with a piezoelectric tube requires that the tube move in an arc, and so maintaining a constant tunneling gap requires that the tube lengthen at the extremes of a scan. Therefore, the surface profile is corrupted with this lengthening, and software correction is required in order to separate the two. The Angstrom Stage is designed to scan the sample in the X and Y directions while a piezoelectric tube provides the high bandwidth necessary in the Z-direction. Taking data in this manner keeps the tip at a constant orientation relative to the surface, eliminates many of the “tip effects” that corrupt the data, and also eliminates the need to correct for tube lengthening at the extremes of scans.

1.3.2 Examples of Precision Motion Control Devices

At the present, most scanned probe microscopes use a mechanical coarse approach stage coupled with a piezoelectric tube for the fine motions. However, over time, many different mechanisms for motion control at this level have been attempted. In general, these use either piezoelectric, electromechanical, or purely mechanical methods to achieve the required resolution.

The first of these was the “Louse” used by Binnig and Rohrer in their original design [5]. This device consists of three pads attached to the bottom of a piezoelectric frame. The pads can be electrostatically clamped to or released from a planar surface. Additionally, the length of the piezoelectric frame elements can be modified via applied voltages. By coordinating these inputs, a walking motion is created. Binnig and Rohrer reported step sizes as small as 10 nm at a rate of 30 steps/second. In theory, the range of this device is unlimited in two dimensions. The Louse was followed by another device called the “Maggot” [11]. Here the sample rested on a frame which was supported by three spherical feet on a flat surface. Two arms holding permanent magnets extend from the frame into the cores of two electromagnets. Translations in X and Y can be obtained by controlling the current to the two coils. This methods had less of a reliance on the

surface finish than did the Louse. Although capable of 5 nm steps, predictable, reliable steps could only be obtained for displacements of 25 nm or greater.

Others to use electromagnetic approaches are Smith [51] and Davydov [12]. Smith's design is similar to the Maggot in that a sample holder resting on ball bearings is translated across a flat surface through the use of magnetic fields. Again, this is a two dimensional device with resolution claimed at about 5 nm. Davydov uses a one dimensional electromagnetic actuator for the coarse approach in the Z-axis of a cryogenic scanning tunneling microscope. Here the minimum step size was about 28 nm.

There have also been some interesting uses of piezoelectrics other than the usual tube scanner. Matey et al [38] use piezoelectric bimorphs to design a stage with translations in the x, y, and z directions. Low resonant frequencies can be a problem when using the relatively thin bimorphs, but they managed a respectable resonant frequency of 190 Hz and a travel of $\pm 60 \mu\text{m}$. Resolution was better than the $1 \mu\text{m}$ that could be obtained with their instrumentation. Smith et al [50] use piezoelectric pistons and counterweights on stacked slides. When a piezo is rapidly expanded or contracted, the inertia of the counterweight leads to a force on the stage and therefore a step displacement in position. Range is theoretically unlimited, and they claimed single step sizes of between 10 nm and 100 nm.

Others have used purely mechanical means to generate nanometer level position resolution. Carr [6] and Mizumota [40] use twist roller friction drives. In these mechanisms, a block with several cylindrical bearings surrounds a highly polished shaft. The bearings are canted at a slight angle to the shaft such that shaft rotations produce translations of the block. Carr achieved a resolution of 90 nm per degree of shaft rotation with a relatively simple device, and Mizumota demonstrated 0.2 nm steps using a more elaborate setup.

Other interesting systems are the Ball Aerospace designs for actively controlled mirrors.² Although not directed towards scanned probe microscopy, these designs are still relevant here since they use a set of electromagnetic actuators to position a levitated mass. In

²Ball Aerospace Systems Group, Electro-Optics/Cryogenics Division, P.O. Box 1062, Boulder, Colorado 80306

Germann [18], the mirror is supported by flexures and acted upon by electromagnets that position it in two rotational degrees of freedom. Rotational stability is as fine as 10 nrad. Medbery [39] describes a system in which the flexures have been eliminated and electrodynamic actuators position the mirror in a full six degrees of freedom. This system matched the operational capabilities of the flexure mounted design and did so with greater travel. These systems are designed for use in a zero gravity environment. A notable contribution of these references is that they explicitly detail the issues involved in decoupling the motions in all six axes.

Most of the methods described above are only capable of generating motion in one to three degrees of freedom (the Ball Aerospace design is an exception). Positioning in the other degrees of freedom would require a stacked mechanism. Mechanically, the simplest method to achieve these degrees of freedom is to support a single moving element electromagnetically. Trumper [53] used analog control to position a 13 kg mass in five degrees of freedom for his doctoral thesis. Likewise, Williams [58] is using electromagnetic actuators along with a linear motor to design a six degree of freedom wafer stepper stage for photolithography. The linear motor provides one long axis of travel, and electromagnetic actuators allow for fine positioning in the other five degrees of freedom. The concept of the Angstrom Stage grew out of these sorts of designs, but through the use of oil flotation was targeted towards extremely fine motion control at the expense of bandwidth. Because the oil flotation is a dominant part of the stage design, we will now examine some earlier uses of fluid flotation in precision machine design.

1.3.3 Prior Uses of Fluid Flotation

One earlier use of fluid for supporting a moving carriage was by W. R. Horsfield in 1965 with his unique ruling engine for diffraction gratings [28]. In prior systems, motion control in a ruling engine was typically done through the use of a servo controlled leadscrew. The leadscrew provided both the motive power and the positioning information, and so it had to be carefully lapped in order to create the most uniform thread spacing possible. Note that this violates the precision machine design guideline of keeping the force and metrology loops separate. An appropriate closed-loop sensor was not available until laser

interferometry came into practical use in the 1960's, and so periodic errors in the thread spacing still existed after the lapping. These had to be mapped and compensated for with the use of a specially designed cam. Even then, the strain and wear of components over time required that the errors be remapped and the cam redesigned periodically. Horsfield's design differed in that he used a hydraulic drive to provide the motive force, and controlled the position through the use of feedback via a laser interferometer. In doing so, he followed good precision machine design practice by separating the force and metrology loops, and thereby essentially eliminated periodic errors in the line spacing. The similarity between this system and the Angstrom Stage is that the moving carriage was suspended in a bath of mercury which supported between 90% and 95% of the weight of the carriage containing the grating blank. Unfortunately, Horsfield only makes a brief reference to this feature in his design, and says nothing about the damping properties of the mercury.

An early use of an oil damper in combination with an electromagnetic suspension was a balance designed by John Clark [10] in 1947. In his design, an electromagnet supports a tray assembly, and a damping vane at the bottom of this assembly is immersed in a jar of oil. Two photodetectors measure the vertical position of the assembly, and the current in a bias winding is adjusted so that the light falling on each of them is the same. A sample is placed in the tray, and current is passed through a secondary winding in order to return the tray to its bias position. The amount of current required to do so is related to the weight of the sample. The oil damper is used mainly to control oscillations, and appeared to be a concession on the part of the inventor. He argued that a properly designed amplifier circuit for the electromagnet could supply the damping much more neatly. A possibility that the author never mentions is to replace the damping vane by a float. In that way, the oil will apply a buoyant force that will lessen the amount of current required in the bias winding. Nevertheless, the author claims that the prototype device was sensitive enough to be competitive with beam balances, and that it was more suitable for use by semiskilled workers.

Many of the gyroscopes used for inertial navigation are also floated in oil [16]. These instruments generally use a cylindrical float that is mounted inside of a reference cylinder.

A Newtonian fluid fills the gap between them and provides the desired damping through viscous shear and the squeeze film effect. The average density of the float equals the density of the fluid, and closed loop control of the fluid's temperature can be used for the final high-accuracy adjustment of the fluid's density.

Finally, the fluids can be used for more than simply damping and density matching; they can also be used to apply forces directly to submerged objects. *Ferrofluids* are fluids whose pressure distribution alters in relation to an applied magnetic field. Willemsen and Este [56] patented a device that uses a ferrofluid to move a float in order to position an optical element.

Chapter 2

Hardware

In general, active magnetic bearings achieve mechanical simplicity at the expense of more complex control system designs. This is especially true when the electromagnetic device is compared to an entirely mechanical device with the same resolution and number of degrees-of-freedom. For example, the Angstrom Stage contains only one moving part. However, the hierarchy of control which positions the stage is quite complex. This is a reasonable tradeoff since the complexity occurs in the electronic and software domains where it is more easily managed. The next section provides a brief description of the mechanical design of the stage, followed by descriptions of the force and metrology loops. The reader is referred to Holmes [26] for a more complete documentation of the mechanical design of the stage, and the author is grateful for the reuse of many of the figures contained therein.

2.1 Stage Overview

The Angstrom Stage consists of a platen floating in an oil filled chamber formed by the frame. Figure 2-1 shows a photograph of the Angstrom Stage, Figure 2-2 shows a cross sectional view, and Figure 2-3 shows an isometric view. In the photograph of Figure 2-1, the stage is shown sitting on a benchtop vibration isolation table. The frame measures approximately $30 \times 30 \times 20$ cm, and is constructed from 6061-T6 aluminum. The platen is also made from aluminum, measures $17 \times 17 \times 8$ cm, and is designed with a hollow

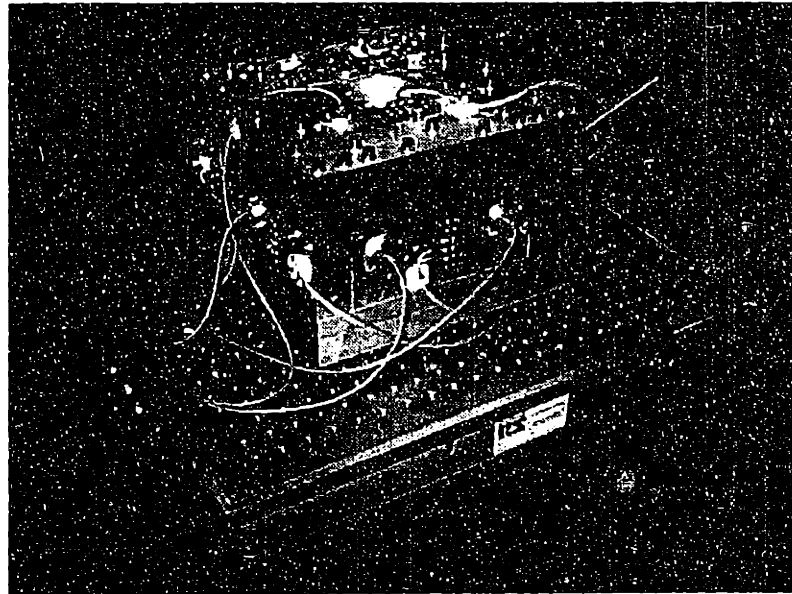


Figure 2-1: Photograph of the Angstrom Stage on a benchtop vibration isolation table.

internal structure. The average density of the platen is chosen to approximately equal the density of the fluid which it displaces, thus making it neutrally-buoyant in the oil. Figure 2-4 shows the details of the ribbed platen. In our prototype, the oil actually offloads about 98% of the 3.09 kg mass of the platen.

The fluorosilicone oil¹ used to float the platen yields some significant advantages in the stage dynamics. First, the buoyant force from the oil offsets the gravitational force on the platen. This reduces power consumption in the electromagnets, and also reduces the associated thermal distortions of the frame. Second, oil flotation attenuates relative motion between the platen and frame resulting from translational base vibrations. If the platen was not neutrally bouyant, it would move relative to the frame when the frame is under acceleration. Third, the oil has a dielectric constant of about 7 relative to air, and this increases the signal-to-noise ratio of the capacitance probes. Finally, the oil provides damping for the stage by forming squeeze film dampers between the platen and the frame-mounted actuators and capacitance probes. With these dampers, the system is highly overdamped, and so exhibits no problems with structural resonances.

The stage is digitally controlled using a TMS320C30-based digital signal processing

¹Dow Corning FS-1265 Fluid, Dow Corning Corporation, Midland MI. (800) 248-2481

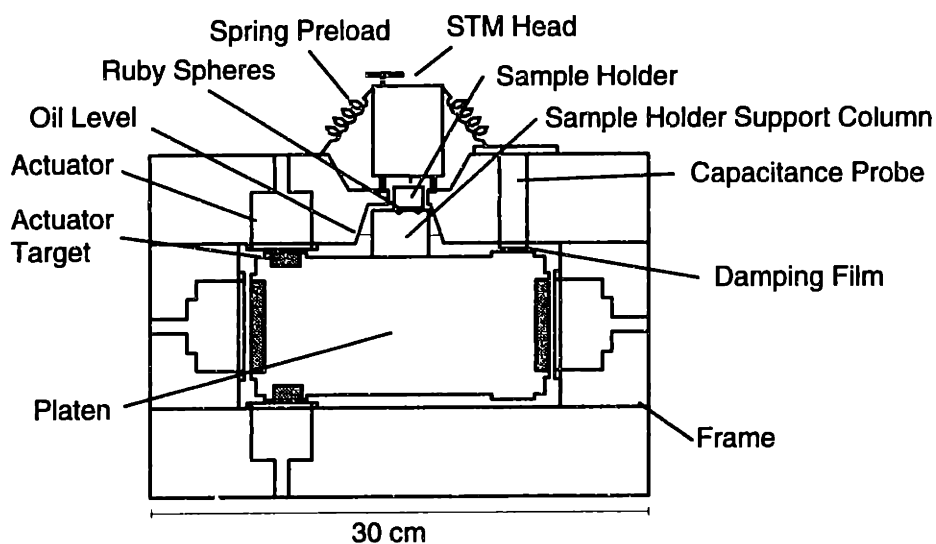


Figure 2-2: Cross sectional view of the Angstrom Stage (from Holmes).

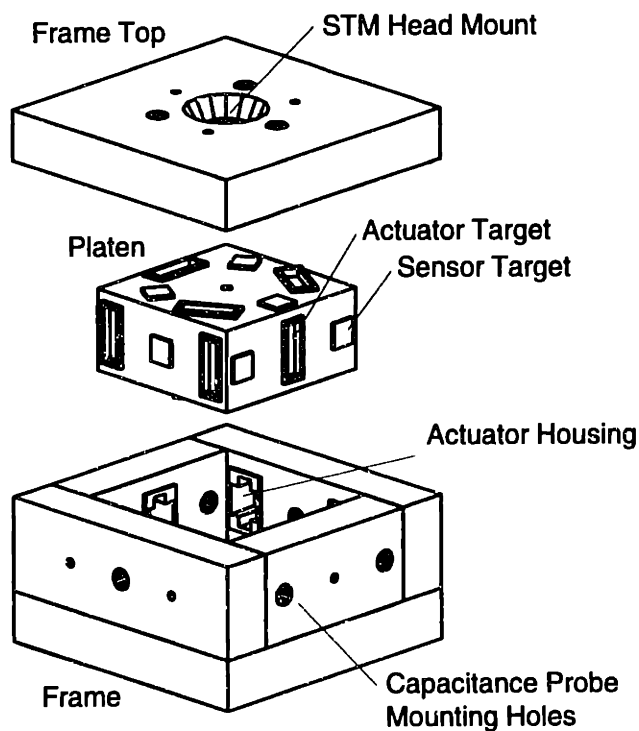


Figure 2-3: Isometric view of the Angstrom Stage (from Holmes).

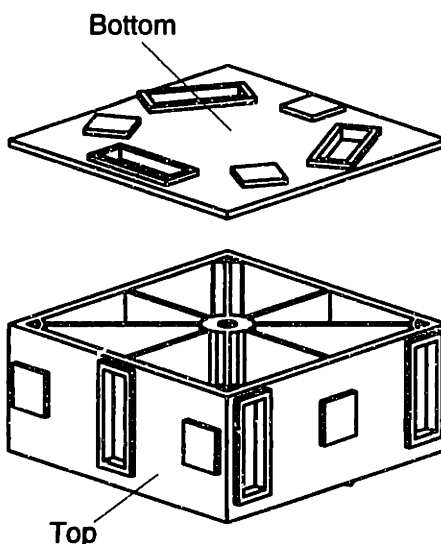


Figure 2-4: Lightweighted platen (from Holmes).

board² residing in a 50 MHz 486DX-based PC. All of the control calculations are performed on the 'C30 due to its superior floating point arithmetic capabilities. The '486 performs the I/O between the 'C30 and the A/D and D/A converters and generates the user interface. The metrology loop that provides position data to the controller, and the force loop that converts controller outputs into physical forces are both described next.

2.2 Metrology Loop

The metrology loop contains all of the components that measure the stage position and transfer the data to the controller. By definition, a change in any of the elements in the metrology loop cannot be distinguished from a change in the actual stage position. Therefore, it is very important that these elements be stable and well characterized. First, the six capacitance probes measure the stage position. Following that, six low pass filters remove high frequency noise from the position signal, and an analog-to-digital converter replaces the physical voltage by a digital word. This number is then available to the controller as an indication of the stage's position.

²Banshee, Atlanta Signal Processors Inc., 770 Spring St. NE, Suite 208, Atlanta, GA, 30308 (404) 892-7265

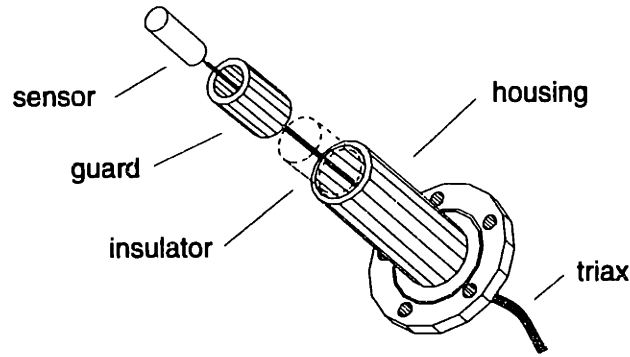


Figure 2-5: Schematic of a capacitance probe (from Holmes).



Figure 2-6: Photograph of one of the capacitance probes (penny shown to provide scale).

2.2.1 Capacitance Probes

Six capacitance probes measure the position and orientation of the platen relative to the frame. A representative capacitance probe is shown in Figure 2-5, and a photograph of one appears in Figure 2-6. Off-the-shelf components did not meet the packaging or noise-performance requirements of the stage, and so were modified to take advantage of the higher dielectric constant of the oil relative to air. The reader is referred to Batchelder [1] for the details of this effort. The probes were calibrated for a 100 micron range with a corresponding output of ± 10 V. For more accuracy however, each probe's response was measured individually and the resulting scaling factors are shown in Table 2.1. The gain in terms of meters per count is based upon a sixteen bit analog-to-digital converter with a range of ± 10 volts.

Probe Number	Gain (m/V)	Gain (m/count)
1	5.1613×10^{-6}	1.5751×10^{-9}
2	4.9820×10^{-6}	1.5204×10^{-9}
3	5.2619×10^{-6}	1.6058×10^{-9}
4	4.9735×10^{-6}	1.5178×10^{-9}
5	4.9922×10^{-6}	1.5235×10^{-9}
6	5.0184×10^{-6}	1.5315×10^{-9}

Table 2.1: Capacitance probe gains.

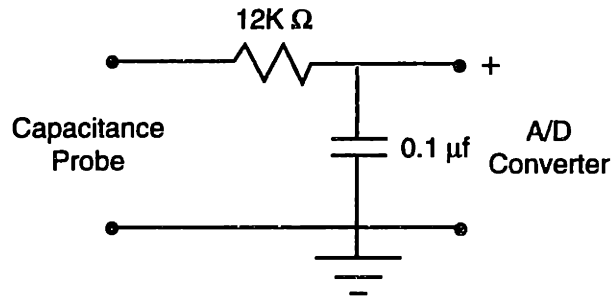


Figure 2-7: Schematic diagram of the anti-aliasing filter.

2.2.2 Anti-Aliasing Filter

A passive first order low pass filter with a cutoff frequency of approximately 130 Hz prevents aliasing by the analog-to-digital converter. The circuit consists of a simple RC filter, which has a transfer function of

$$\frac{V_o(s)}{V_i(s)} = \frac{1/RC}{s + 1/RC}. \quad (2.1)$$

Substituting the component values shown in Figure 2-7 gives a numerical transfer function of

$$\frac{V_o(s)}{V_i(s)} = \frac{833}{s + 833}. \quad (2.2)$$

The measured frequency response of the filter (taken with an HP 35665A Dynamic Signal Analyzer) is shown in Figure 2-8.

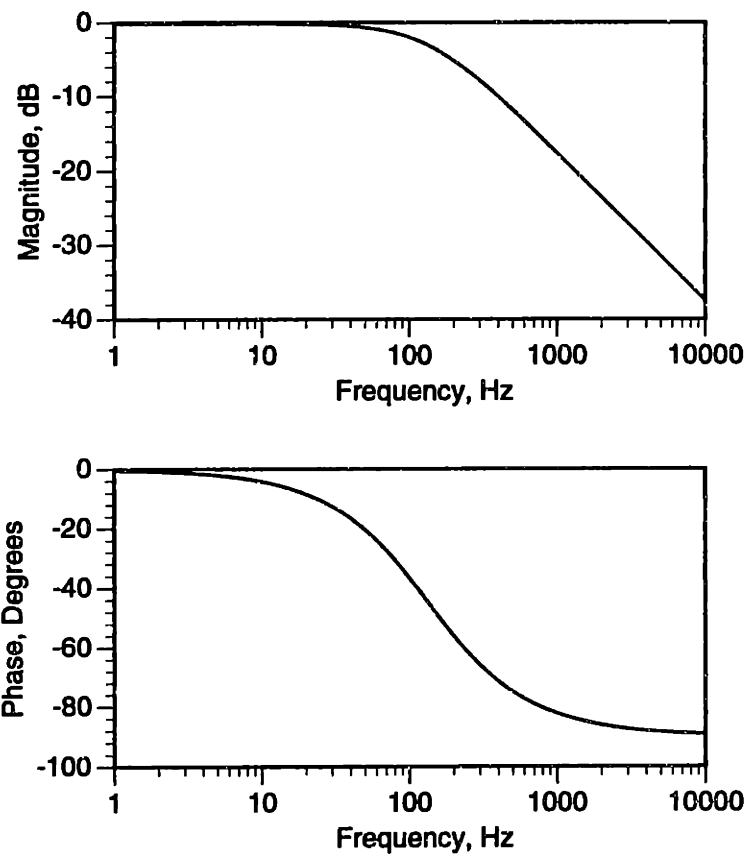


Figure 2-8: Measured frequency response of the anti-aliasing filter.

2.2.3 Analog-to-Digital Conversion Board

A 16-bit analog-to-digital conversion board converts the voltage from the capacitance probes into a digital word for the PC to use in control calculations.³ The board is configured for seven differential inputs with a range of ± 10 V. Six channels measure the capacitance probe voltages, and the seventh measures a signal from the Z-servo of an STM. The maximum sampling rate of the board is 160,000 samples per second, and built-in DMA capabilities allow us to operate at this speed. Noise in the A/D converter electronics constitute the largest error in the position measurement, and therefore weigh heavily in the control system design. Specifics of the noise are subsequently described in greater detail in Chapter 4.

2.3 Force Loop

2.3.1 Digital-to-Analog Conversion Board

The digital-to-analog converter acts as a link between commanded voltages in the control software and an actual physical output voltage. The converter has 12-bit resolution with an output range of 0–10 V.⁴ A useful capability of this board is the simultaneous output option. Using this feature, the DAC output voltages change at essentially the same time rather than sequentially as they are written to. Although seemingly trivial, time skew between individual DAC outputs may lead to a “beating” effect in the platen position at higher bandwidths.

This phenomenon, displayed in the trace of Figure 2-9 was prevalent before I used the simultaneous output feature of the board, and disappeared immediately after. Some time later (after many more changes to the controller) I was unable to return and recreate the effect. Therefore, this section has to rest in my suspicion only.

³DT2838, Data Translation Inc., 100 Locke Dr., Marlboro, MA 01752-1192, (508) 481-3700

⁴PC-422, Datal Inc., 11 Cabot Blvd, Mansfield, MA 02048-9984, (508) 339-3000

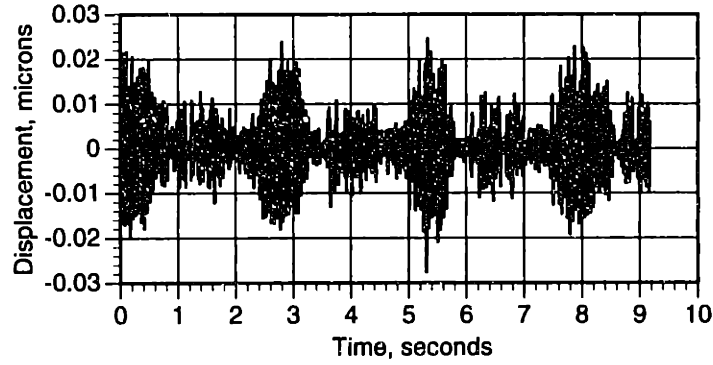


Figure 2-9: Beating effect possibly due to time skew in DAC outputs.

Amplifier Number	Gain (A/V)	Gain (A/Count)
1	0.033344	8.1408×10^{-5}
2	0.034165	8.3410×10^{-5}
3	0.034305	8.3754×10^{-5}
4	0.034388	8.3954×10^{-5}
5	0.034072	8.3182×10^{-5}
6	0.034352	8.3868×10^{-5}
7	0.034235	8.3582×10^{-5}
8	0.033956	8.2900×10^{-5}
9	0.034118	8.3296×10^{-5}
10	0.033979	8.2956×10^{-5}
11	0.034483	8.4186×10^{-5}
12	0.034048	8.3126×10^{-5}

Table 2.2: Amplifier gain factors.

2.3.2 Power Amplifiers

Twelve linear transconductance power amplifiers transduce the voltage from the D/A converters into the respective actuator currents. Of main interest is the scaling factor from “voltage in” to “current out.” The amplifier does not measure the current directly, but instead controls the voltage across a 1Ω current sense resistor. These are 1% resistors, and so the amplifiers cannot do any better than this on the voltage-to-current scaling factor. The measured voltage-to-current scaling factors are given in Table 2.2, and a schematic diagram of the amplifier is shown in Figure 2-10. Gains given in Amps/Count are based upon a 12-bit board operating over a range of 0–10 V.

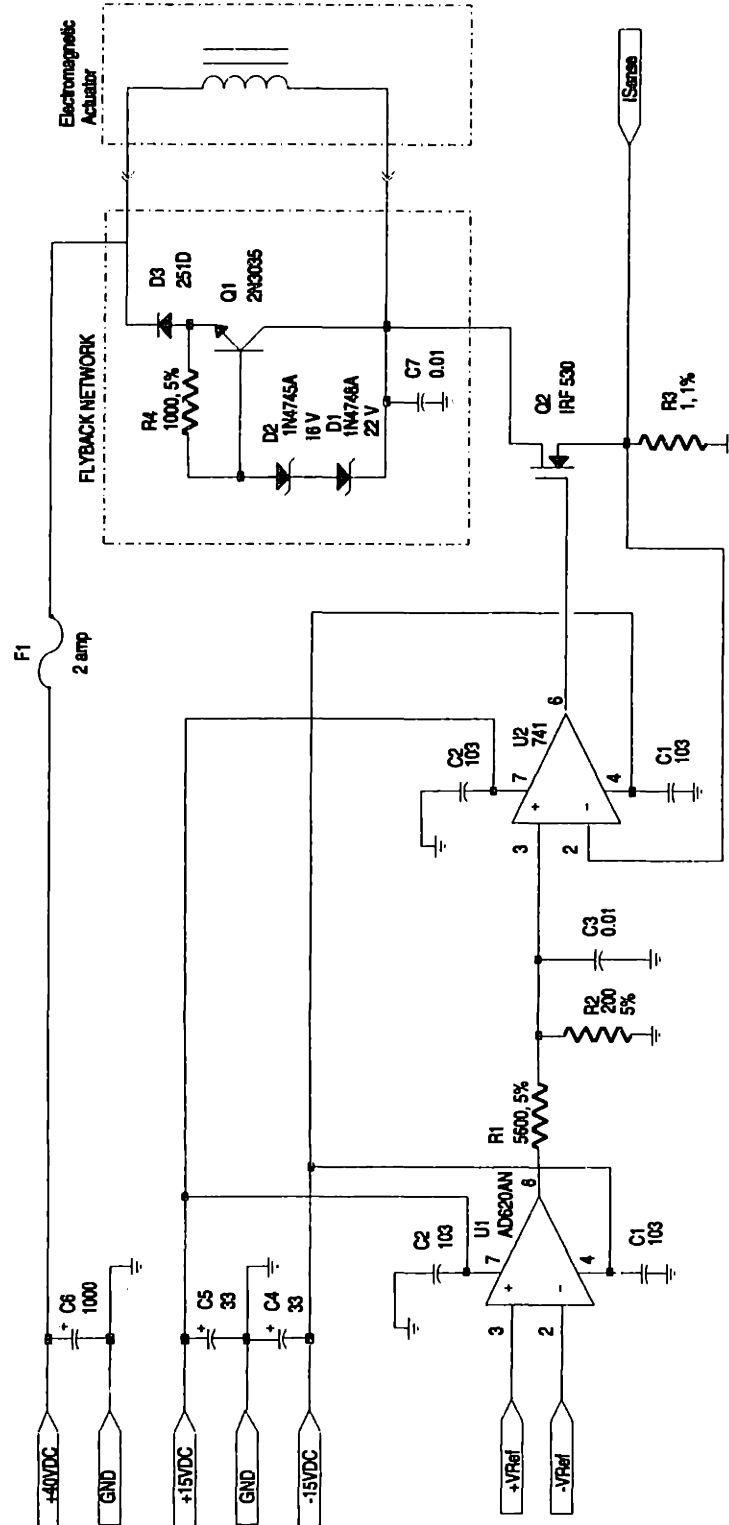


Figure 2-10: Schematic diagram of a power amplifier (from Holmes).

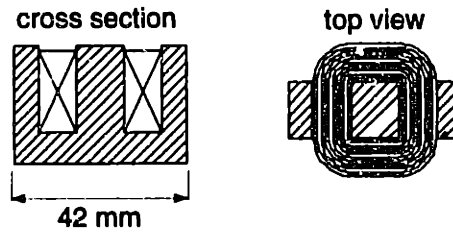


Figure 2-11: Drawing of an unpotted electromagnetic actuator (from Holmes).

2.3.3 Electromagnetic Actuators

The electromagnetic actuators provide the forces necessary to suspend and servo the platen. Each actuator is fabricated with 230 turns of 22 gauge copper wire around 50-50 Ni-Fe E-Core laminations. These actuators are potted directly into the frame, and act upon corresponding 50-50 Ni-Fe I-laminations epoxied into the platen.

The actuators are designed to operate at a nominal gap of $300\ \mu\text{m}$, and are capable of producing in excess of 50 N force at this gap. However, normal operating forces are less than 1 N. The coil has an internal resistance of $1\ \Omega$ and an inductance of 60 mH. and an extra actuator was potted into a mounting cartridge for characterization by a magnetic bearing calibration fixture [46]. Figure 2-11 shows a drawing of one of the unpotted actuators, and Figure 2-12 contains a photograph of the extra actuator that was used for calibration purposes.

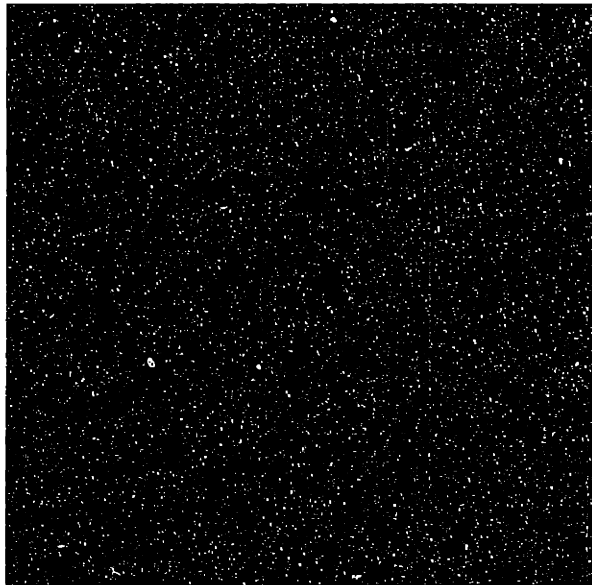


Figure 2-12: Photograph of a potted electromagnetic actuator and one lamination with a penny shown to provide scale.

Chapter 3

Mechanical Modeling

3.1 Introduction

This thesis is mainly devoted to designing a suitable control system for the Angstrom Stage. In doing so, the most critical first step is the development of a model of the physical system. Others have provided guidelines for this process. Nikola Tesla claimed to be able to design motors in his head, set them running, and then think back to them a few weeks later to see how they were performing. He believed that "Engineering, electrical or mechanical, is positive in results. There is scarcely a subject that cannot be mathematically treated and the effects calculated or the results determined beforehand from the available theoretical or practical data" [8]. Along the same lines, Bob Donaldson, the designer of the large optics diamond turning machine (LODTM) at Lawrence Livermore, maintains that "A basic finding from our experience in dealing with machining accuracy is that machine tool errors are deterministic. By this we mean that machine tool errors obey cause-and-effect relationships, and do not vary randomly for no reason." [48]. Finally, a good design guideline is Einstein's recommendation to "make things as simple as possible, but not simpler."

The modeling of the mechanical aspects of the stage begins with the fluid dampers that dominate the stage dynamics. The interactions of these dampers in the six degrees of freedom combine to form the plant transfer function from force to velocity. Because the stage is so highly damped, these viscous forces by far dominate any inertial terms

on time scales shorter than ten's of microseconds. Following this analysis is a modeling of the electromagnetic actuators that obtains the transfer function from current to force, and a theoretical derivation of the effects of angular displacements between the electromagnet and target faces. The chapter concludes with the derivation of the various coordinate transformations needed in the control system design. Because of their complexity, the exact algebraic and numerical values for the damping matrix and coordinate transformation matrices are presented in Appendix A.

3.2 Nomenclature

A	\equiv	Generic area variable
α	\equiv	Platen density matching factor
\mathbf{B}	\equiv	Stage damping matrix
$\hat{\mathbf{B}}$	\equiv	Estimated stage damping matrix
B_a	\equiv	Actuator damping coefficient (function of gap)
B_c	\equiv	Capacitance probe damping coefficient (function of gap)
b	\equiv	Width of an actuator damper: 0.06 m
b_a	\equiv	Damping constant for the actuators: 9.1494×10^{-7} Nsm ²
b_c	\equiv	Damping constant for the capacitance probes: 6.2345×10^{-8} Nsm ²
b_{ij}	\equiv	An equivalent damping term that relates a force or torque in the i direction to a velocity in j
C	\equiv	Actuator force constant: 2.456×10^{-6} Nm ² /A ²
\hat{C}	\equiv	Estimated actuator force constant
F	\equiv	Generic force variable
F_d	\equiv	Generic desired force variable
\mathbf{F}	\equiv	Modal force vector containing the six net forces and torques about the platen center
\mathbf{F}_{em}	\equiv	Electromagnetic actuator force vector
$F_{i,j}$	\equiv	Force applied by the actuator pair i, j

\hat{g}_a	\equiv	Estimated actuator gap
g_{ai}	\equiv	The instantaneous gap of the i th actuator
g_{aio}	\equiv	The null gap of the i th actuator
g_{ci}	\equiv	The instantaneous gap of the i th capacitance probe
g_{cio}	\equiv	The null gap of the i th capacitance probe
\mathbf{H}	\equiv	Measurement sensitivity matrix
i	\equiv	Current supplied to an actuator
i_d	\equiv	Desired current to be supplied to an actuator
K_p	\equiv	Proportional control gain
K_s	\equiv	Shape factor for rectangular squeeze film dampers
ℓ	\equiv	A characteristic length for actuator and capacitance probe spacing: 0.0665 m
l	\equiv	Length of an actuator damper: 0.025 m
m	\equiv	Platen mass: 3.09 kg
μ	\equiv	Viscosity of the fluorosilicone oil: 1.323 kg/ms
Φ_F	\equiv	Modal force transformation matrix
R	\equiv	The radius of the capacitance probe face: 0.01 m
T	\equiv	Time between controller updates: 0.00224 s
T_d	\equiv	Derivative control gain
T_i	\equiv	Integral control gain
\mathbf{V}	\equiv	Modal velocity vector
\mathbf{V}_d	\equiv	Desired modal velocity vector
V	\equiv	Generic velocity variable
\mathbf{X}	\equiv	State vector containing the location of the platen in cartesian coordinates relative to the nominal center location of the platen
X	\equiv	Generic position variable
X_d	\equiv	Generic desired position variable
\mathbf{Y}	\equiv	Measurement vector containing the capacitance probe readings

3.3 Fluid Dampers

The dominant components of the stage are the squeeze film dampers formed by the thin film of oil trapped between the platen and the actuators and capacitance probes. When the platen moves, it must squeeze a viscous fluid from the shrinking gap. The same effect occurs when the platen moves away and fluid must enter the newly formed space. This cannot happen instantaneously and so the amount of damping depends upon the shape of the damping pads and the size of the gap. Both the circular and rectangular cases are examined next.

3.3.1 Circular Squeeze Film Dampers (Capacitance Probes)

The relationship between the velocity of the platen (i.e. the speed with which the gap is opening or closing) and the required force for the circular capacitance probe dampers comes from an application of the Navier-Stokes equation for viscous fluid flow as given in [14], and is presented here without derivation.

$$F = \frac{3\pi\mu R^4}{2g_c^3} V. \quad (3.1)$$

The oil used in the stage has a viscosity of $\mu = 1.323 \text{ kg/ms}$, and the capacitance probes have a radius of $R = 1 \text{ cm}$. The constants combine to form a single term for the capacitance probe damping. This yields

$$F = \frac{b_c}{g_c^3} V, \quad (3.2)$$

where $b_c = 6.2345 \times 10^{-8} \text{ Nsm}^2$. At the nominal film thickness, $g_c = 100 \text{ }\mu\text{m}$; each probe then contributes a damping term to the system of $B_c = 62,345 \text{ Ns/m}$. Because of the changing gap however, this term varies from $498,000 \text{ Ns/m}$ when the platen is close to the probes ($g_c = 50 \text{ }\mu\text{m}$) to $18,500 \text{ Ns/m}$ when it is farthest away ($g_c = 150 \text{ }\mu\text{m}$). Although this is a wide variation in damping over the travel of the stage, the nonlinearity is extremely predictable and well-behaved, and thus can be linearized via computation as described in Section 6.3.2.

3.3.2 Rectangular Squeeze Film Dampers (Actuators)

The damper formed by the gap between an electromagnet and its target is more difficult to model in a closed form because of its rectangular shape. The geometry leads to an infinite series solution for the relationship between force and velocity. Hays [23] details a derivation, and Pinkus and Sternlicht [45] give the form used here,

$$F = \frac{192\mu l^3 V}{\pi^4 g_a^3} \sum_{n=1,3,5}^{\infty} \left(\frac{b}{2n^4} - \frac{l}{n^5 \pi} \tanh \frac{n\pi b}{2l} \right). \quad (3.3)$$

The summation contains only the fixed width ($b = 0.06$ m) and length ($l = 0.025$ m) of the damping pad, and so evaluates to a constant independent of the gap. Thus the damping due to an actuator in our system can be expressed as

$$F = \frac{b_a}{g_a^3} V. \quad (3.4)$$

Numerically evaluating the series in eq.(3.3) to a sufficient number of terms leads to a value of $b_a = 9.1494 \times 10^{-7}$ Nsm². Equation (3.3) is somewhat cumbersome to use routinely, and is of little use in determining a quick estimate of the damping value. As an alternative approach, Fuller [17] tabulated the results for various length-to-width ratios, since the series can be factored to depend only upon the ratio b/l . The simplified expression,

$$F = K_s \frac{\mu b l^3}{g_a^3} V, \quad (3.5)$$

contains a constant K_s that accounts for the different width-to-length ratios. A plot of the constant for different values of b/l is shown in Figure 3-1. As the ratio of b/l becomes very large, K_s approaches a value of one. The discrete points in this figure indicate the values tabulated by Fuller, and the curve fit behind them was performed by Slocum et al [49]. Slocum found that the shape factor could be approximated by the equation

$$K_s = 0.7925 - 1.1005/e^{b/l} + 0.0216 l/b + 0.0153 b/l. \quad (3.6)$$

Note that there is actually a misprint of this equation in [49]. In this paper there are parentheses around the last three terms that should not be present. Equation (3.6) contains the corrected form. The ratio of b/l for the actuator damping pads is 2.4, and so interpolating between the data tabulated by Fuller yields a value of approximately 0.738 for K_s . Calculating the damping term at the nominal actuator oil film thickness ($g_a = 300$ μ m) gives a value of $B_a = 33,869$ Ns/m. Again however, the gap thickness changes with the motion of the stage, and the contribution of each individual actuator can range between 21,300 Ns/m ($g_a = 350$ μ m) and 58,500 Ns/m ($g_a = 250$ μ m). The variation here is not as significant as it is for the capacitance probes since the ± 50

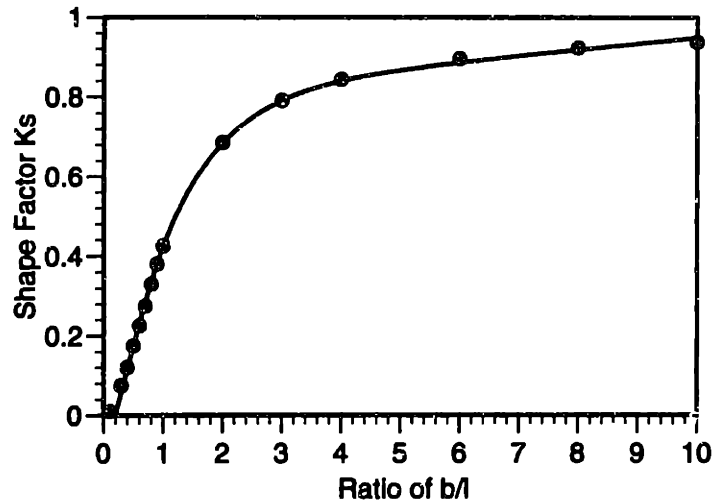


Figure 3-1: Shape factor for the damping between rectangular plates.

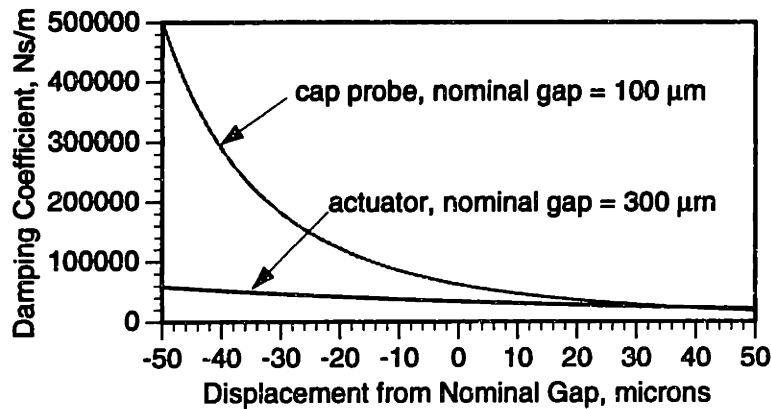


Figure 3-2: Damping coefficients for a single actuator and capacitance probe

micron variation in gap thickness is not as large a percentage of the actuators' 300 micron nominal gap as it is for the capacitance probes' nominal 100 micron gap. To clarify this issue, Figure 3-2 compares the damping factors of a single capacitance probe and electromagnetic actuator as the moving platen changes their gaps by $\pm 50 \mu\text{m}$. Note that although the damping for a capacitance probe is a factor of ten higher at the minimum gap, at the larger gaps the actuator damping factor is actually the greater of the two by a small margin.

3.3.3 Shear Damping

The other potential contribution to the stage's damping comes from the shear forces generated when the platen moves in parallel with the face of a probe or actuator. The

relationship between force and velocity in this case is

$$\frac{F}{A} = \mu \frac{dV}{dg}. \quad (3.7)$$

Assuming a linear velocity profile between the platen and frame, this equation reduces to

$$F = \frac{\mu A}{g} V. \quad (3.8)$$

Substituting numerical values reveals that the maximum shear damping coefficient for both the capacitance probes and the actuators is about 8 Ns/m, and so the shear damping term is insignificant in comparison with the squeeze film effects. This makes intuitive sense if one imagines a flat plate sliding on a film of water on a smooth tabletop. Pulling the plate away from the table can be difficult due to the squeeze film effect, but the plate can slide from side to side very easily. Because of its insignificant value, the shear damping term is ignored in further calculations.

3.3.4 Combined Damping Model in Six Degrees-of-Freedom

The inverse cubed relationship between the capacitance probe and actuator gaps and the damping terms leads to coupling between motions in the six degrees-of-freedom. Referring to Figure 3-3 which shows the positions of the actuators and capacitance probes relative to the platen helps in understanding these relationships. As an example, consider the case where actuator E3 is applying a force on the platen in the X-direction. If the gaps of capacitance probes C2 and C3 are equal, then there will be no torque applied about the Z axis. However, if one of the gaps is larger, it will have a lower damping value. Now as the stage moves in X, it will also rotate about the Z axis leading to an even bigger disparity in the gaps; i.e. this is an unstable process. To this point, we have only calculated the damping for a single isolated actuator or capacitance probe. In practice, multiple dampers act on the stage for every motion. The three lateral degrees of freedom are coupled, as are the three vertical ones, but to first order there is no inherent coupling between lateral and vertical motions.

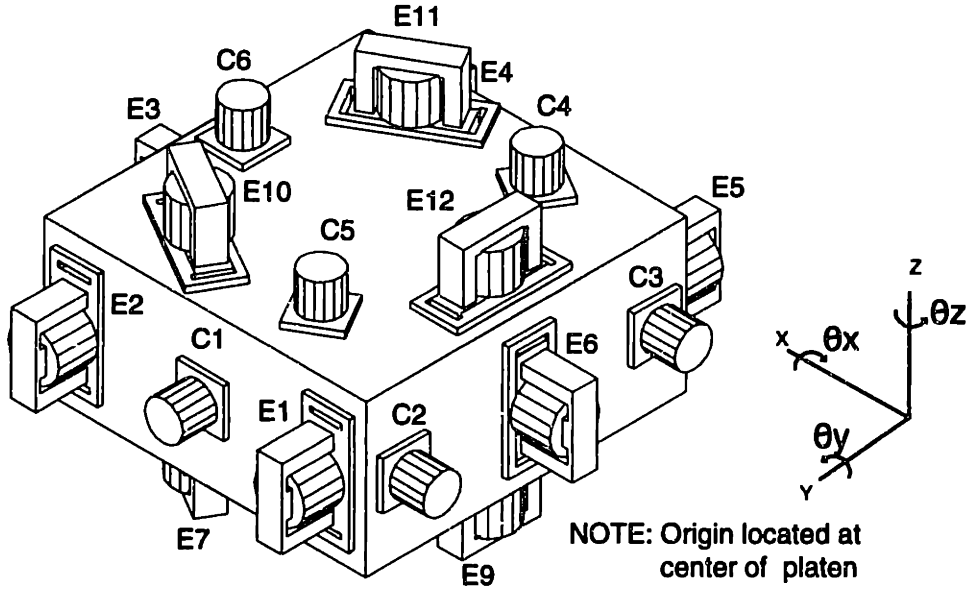


Figure 3-3: Location of probes and actuators relative to the platen (from Holmes).

Since it involves mainly algebra, the derivation of the equations for the elements of the damping matrix is included only in Appendix A. The end result is that the damping in all six degrees of freedom can be represented by a single 6×6 matrix,

$$\mathbf{B} = \begin{bmatrix} b_{xx} & 0 & 0 & 0 & 0 & b_{x\theta_z} \\ 0 & b_{yy} & 0 & 0 & 0 & b_{y\theta_z} \\ 0 & 0 & b_{zz} & b_{z\theta_x} & b_{z\theta_y} & 0 \\ 0 & 0 & b_{\theta_x z} & b_{\theta_x \theta_x} & b_{\theta_x \theta_y} & 0 \\ 0 & 0 & b_{\theta_y z} & b_{\theta_y \theta_x} & b_{\theta_y \theta_y} & 0 \\ b_{\theta_z x} & b_{\theta_z y} & 0 & 0 & 0 & b_{\theta_z \theta_z} \end{bmatrix}, \quad (3.9)$$

where each of the individual b_{ij} terms is a function of the instantaneous stage position with as functional variation as calculated in the previous section. By defining

$$\mathbf{F} = \begin{bmatrix} F_x & F_y & F_z & T_x & T_y & T_z \end{bmatrix}^T, \quad (3.10)$$

and

$$\mathbf{V} = \begin{bmatrix} \dot{x} & \dot{y} & \dot{z} & \dot{\theta}_x & \dot{\theta}_y & \dot{\theta}_z \end{bmatrix}^T, \quad (3.11)$$

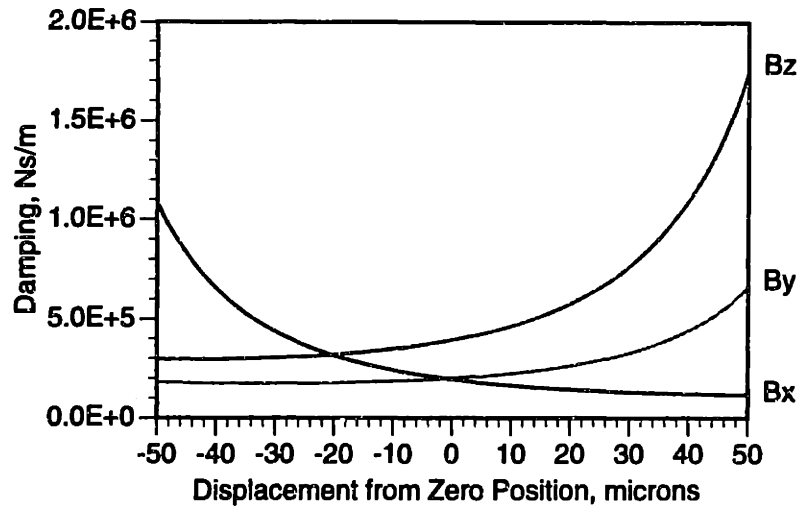


Figure 3-4: Damping coefficients as the stage moves in the translational directions.

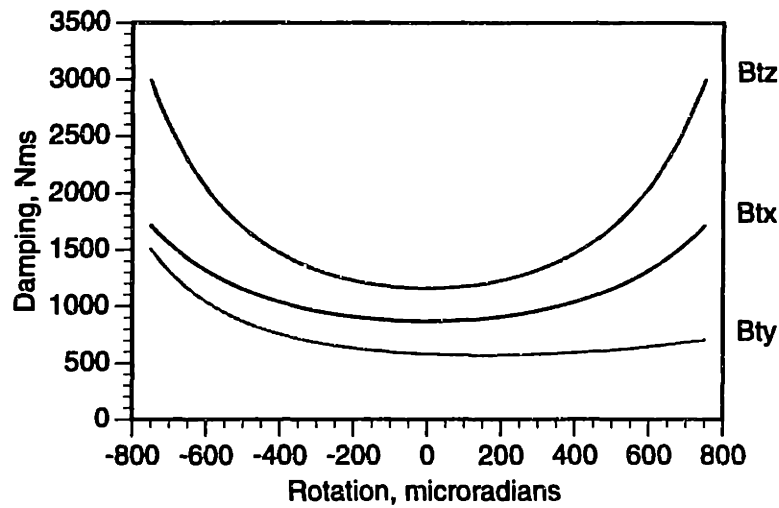


Figure 3-5: Damping coefficients as the stage moves in the rotational directions.

we have the equation of motion of the Angstrom Stage,

$$\mathbf{F} = \mathbf{B}\mathbf{V}. \quad (3.12)$$

For perspective on the magnitude of the damping values in the combined model, Figures 3-4 and 3-5 shows the damping in each degree-of-freedom as the stage moves from plus to minus full scale in that axis.

3.4 Electromagnetic Actuators

The control system for the stage can be designed without any understanding of the underlying electromagnetic theory. All the designer needs to know is that the force applied by one of the actuators is proportional to the current squared divided by the gap squared. However, the derivation of this relationship is not overly complicated, and may help to explain characteristics of the stage's behavior. Some caution concerning notation is required here. Several variables have different meanings in this section than they do elsewhere in this thesis. One option would have been to apply subscripts throughout the paper in order to clearly differentiate them, but this detracts greatly from the readability. Instead, since the electromagnetic terms are used only in section 3.4, I have chosen to simply redefine them here.

3.4.1 Electromagnetics Nomenclature

A	\equiv	Electromagnet pole face area
B	\equiv	Magnetic flux density
B_o	\equiv	Magnetic flux density in a vacuum
C	\equiv	Actuator electromagnet force constant: $2.456 \times 10^{-6} \text{ Nm}^2/\text{A}^2$
d	\equiv	The length of the sides of the square center pole face: 1.27 cm
F	\equiv	Force applied by an actuator
H	\equiv	Magnetic field intensity
i	\equiv	Current flowing in a wire
j_M	\equiv	Magnetization current
ℓ	\equiv	A characteristic length for actuator and capacitance probe spacing: 0.0665 m
M	\equiv	Magnetization density
μ	\equiv	Magnetic permeability of a material
μ_o	\equiv	Permeability of free space: $4\pi \times 10^{-7} \text{ Tm/A}$
λ	\equiv	Magnetic flux linkage
N	\equiv	Total number of turns of wire
n	\equiv	Number of turns of wire per unit length
T	\equiv	Torque applied by an actuator
θ	\equiv	The angle of actuator rotation
x	\equiv	The gap at the actuator's center pole face
χ	\equiv	Magnetic susceptibility
W	\equiv	Magnetic coenergy
w	\equiv	The distance between actuator pole faces: 1.75 cm

3.4.2 Terminology / Basic Electromagnetics

Even a broad overview of electromagnetic field theory is too extensive for this thesis, and really does not belong in it. Most physics texts contain the basics; I have found Marion and Hornyak [37] to be particularly useful. In the interest of completeness though, I include here the minimum necessary to follow the derivation of the force-current-gap relationship.

Begin by accepting that a current carrying wire creates lines of *magnetic flux* in concentric circles around it. The direction of the flux is given by the right hand rule,¹ and it is important to note that lines of flux are continuous, have no beginning or ending, and never intersect. This is stated more analytically by Faraday's Law,

$$\oint_S \mathbf{B} \cdot \mathbf{n} dA = 0. \quad (3.13)$$

Any flux which leaves a closed surface also has to reenter it, and so the net flux passing through it is zero. We will make use of this law later.

The density of these flux lines indicates the strength of the magnetic field. For an infinitely long wire of negligible thickness, the *magnetic flux density* \mathbf{B} is given by

$$\mathbf{B} = \frac{\mu_o i}{2\pi R} \hat{\mathbf{u}}_\phi. \quad (3.14)$$

In the SI system of units, \mathbf{B} is measured in units of *tesla* where $1 \text{ T} = 1 \text{ N/Am}$. As a vector quantity, \mathbf{B} has both a magnitude and direction. In this case, the direction $\hat{\mathbf{u}}_\phi$ is tangent to a circle centered on the wire. The constant μ_o is called the *permeability of free space* and equals $4\pi \times 10^{-7} \text{ Tm/A}$. The other variables are the current in the wire, i , and the distance from it, R .

Wrapping the wire into a coil forms a *solenoid*, and if the length of this coil is much larger than the radius, then the flux density is nearly constant inside the coil at a value of

$$\mathbf{B} = \mu_o n i \hat{\mathbf{k}}. \quad (3.15)$$

¹With the thumb pointing in the direction of the current, the fingers show the direction of the flux.

The unit vector $\hat{\mathbf{k}}$ is oriented along the axis of the cylinder, and n is the number of turns of wire per unit length. The relationship is an intuitive one; more turns yield more flux.

The strength of the magnetic flux changes when the core of the solenoid is no longer a vacuum, since each material has permeability, μ , different than that of free space. A quick explanation is that the individual molecules in a material develop effective magnetic moments, \mathbf{m} . Over the entire material then, the individual magnetic moments sum to form a macroscopic magnetization density vector \mathbf{M} . An older but highly readable and informative article on the subject of magnetization is one written by Becker [2]. His article concerns rare earth magnets specifically, but he supplies a good background on the subject of magnetization in general. While magnetization must ultimately be treated via the concepts of quantum physics, a classical viewpoint suffices here. That is, we think of the individual magnetic moments as setting up elementary current loops within the material. If the magnetization density is uniform, these currents cancel inside of the material, but add on the surface to yield a total magnetization current I_M that circulates around the material over the length ℓ . The total magnetic moment is then defined as this current times the cross sectional area. Dividing by the volume of the material gives the net magnetic moment per unit volume,

$$M = \frac{I_M A}{A\ell} = \frac{I_M}{\ell} = j_M, \quad (3.16)$$

which can also be expressed as a surface magnetization current per unit length, j_M . Inside the solenoid then, the flux density equals the sum of that expected from the applied current plus an additional amount due to the magnetization current,

$$B = \mu_o(ni + j_M). \quad (3.17)$$

This relationship holds for the simple geometry of the solenoid, but not for the more general case. There, it is best to express the difference between the actual flux density B , and that which would be experienced in a vacuum B_o as

$$B - B_o = \mu_o M. \quad (3.18)$$

With that relationship established, it becomes convenient to introduce another quantity, the *magnetic field intensity* \mathbf{H} . It relates to the magnetic flux density through

$$\mathbf{H} = \frac{\mathbf{B}_o}{\mu_o} = \frac{\mathbf{B}}{\mu_o} - \mathbf{M}. \quad (3.19)$$

In a magnetically linear material, the magnetization is proportional to the magnetic field intensity, i.e.

$$\mathbf{M} = \chi \mathbf{H}. \quad (3.20)$$

The constant, χ , is referred to as the *magnetic susceptibility*. Substituting eq.(3.20) back into eq.(3.19) yields

$$\mathbf{B} = \mu_o (1 + \chi) \mathbf{H}. \quad (3.21)$$

Defining the permeability of the material as $\mu = \mu_o(1 + \chi)$ leaves us with the familiar relationship between the flux density and magnetic field intensity,

$$\mathbf{B} = \mu \mathbf{H}. \quad (3.22)$$

With these definitions established, we can state Ampère's circuital law,

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = Ni. \quad (3.23)$$

Ampère found that the net magnetic field intensity around a closed contour equals the total current passing through a surface bounded by that contour. By stating the law in terms of the magnetic field intensity, we are able to work entirely in terms of applied currents, and not concern ourselves with those modeled as induced on the surface of a material.

At this point, we have seen enough basic electromagnetics to proceed with the derivation of the force-torque-gap relationship for a magnetic actuator. However, this relationship has been derived numerous times before, for example in Woodson and Melcher [59], and little would be gained by rederiving it here. To yield a more useful result, we can apply the same principles and derive the relationship for the more general case where the

pole faces and the target are not parallel to each other, but instead are angled so as to form a non-uniform gap, as shown in Figure 3-6.

3.4.3 Derivation of Force and Torque for a Non-Uniform Gap

In general, we have calculated that the force produced by an electromagnetic actuator is proportional to the current squared divided by the gap squared,

$$F = C \frac{i^2}{g_a^2}. \quad (3.24)$$

However, this model contains the implicit assumption that the gap remains uniform across the face of the actuator. The platen can move in six degrees of freedom, and so the planes formed by the actuator and target faces can be at an angle to each other. The calculated maximum angle based upon the allowable motion of the stage is about 1 milliradian (0.057 degrees). (In practice, manufacturing errors can raise the angle even further.) This rotation leads to a variation in gap of 4 microns over the long length of the magnet face. As a first approximation, because the actuator is much wider than it is deep, we will assume that only rotations across this wider face are significant. We thus need to derive the magnetic force equations in terms of both gap x and rotation angle θ . Figure 3-6 shows the notation used in the derivation.

We begin with the assumption that the core is an infinitely permeable material ($\mu = \infty$). In order to have a finite flux density \mathbf{B} , the field intensity \mathbf{H} must thus be zero everywhere within the magnetic core. The only non-zero magnetic field intensity occurs in the gaps where

$$B = \mu_o H. \quad (3.25)$$

Next, we make use of Ampere's law for quasi-static fields,

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = Ni, \quad (3.26)$$

where N is the number of turns of the coil with a terminal current i . Evaluating this

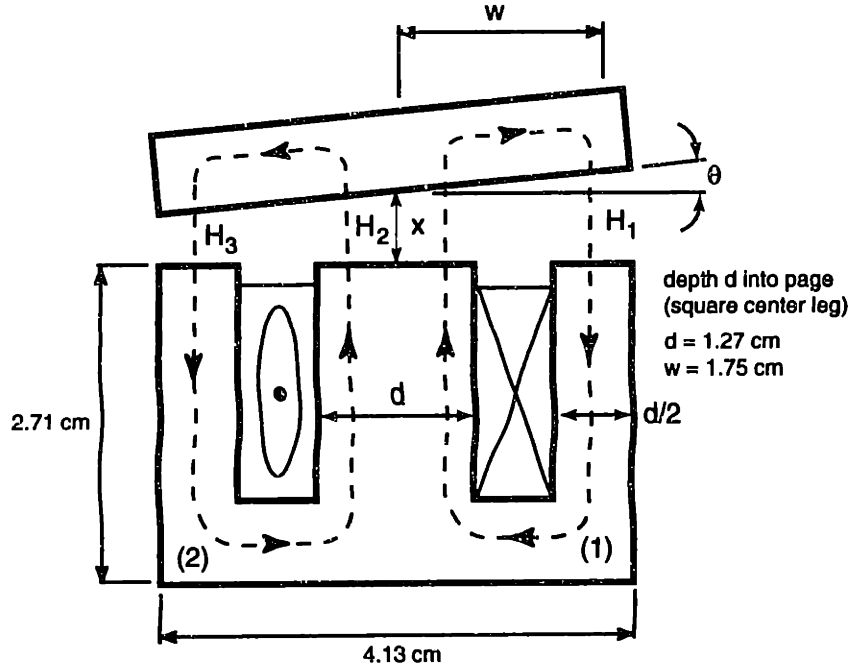


Figure 3-6: Electromagnetic actuator at a non-uniform gap.

integral along contours (1) and (2) in Figure 3-6 leads to the pair of equations

$$H_2x + H_1(x + w\theta) = Ni \quad (3.27)$$

and

$$H_2x + H_3(x - w\theta) = Ni. \quad (3.28)$$

Here, the H_k terms are the magnetic field intensities in each of the three gaps in the direction of the contour. The total current through the area enclosed by the contour is simply the current i passing through the wire multiplied by the number of turns N . In the above calculations we have made the assumption that the flux is uniform across each of the pole faces. In reality, each of the gaps is wedge shaped, but for small angles we can ignore this more complex geometry.

We need a third equation in order to solve for each of the individual magnetic field intensities. This is provided by an application of Faraday's law, which states that the

total magnetic flux through any closed surface must be zero:

$$\oint_S \mathbf{B} \cdot \mathbf{n} dA = 0. \quad (3.29)$$

Integrating over a surface which contains the target and passes through the gaps yields

$$\mu_o H_2 d^2 - \mu_o H_1 \frac{d^2}{2} - \mu_o H_3 \frac{d^2}{2} = 0. \quad (3.30)$$

This simplifies to

$$H_2 - \frac{1}{2}H_1 - \frac{1}{2}H_3 = 0, \quad (3.31)$$

which when combined with eqs.(3.27) and (3.28) creates a set of three equations in three unknowns (H_1 , H_2 , and H_3). After some algebraic manipulation, the three magnetic field intensities are found to be

$$H_1 = \frac{Ni(x - w\theta)}{2(2x^2 - w^2\theta^2)} \quad (3.32)$$

$$H_2 = \frac{Nix}{2x^2 - w^2\theta^2} \quad (3.33)$$

$$H_3 = \frac{Ni(x + w\theta)}{2(2x^2 - w^2\theta^2)}. \quad (3.34)$$

The magnetic flux linkage through the N turn coil is defined as $\lambda_k = N\mu_o A_k H_k$, and so in solving for them we arrive at

$$\lambda_1 = \frac{\mu_o d^2 N^2 (x - w\theta)}{2(2x^2 - w^2\theta^2)} i \quad (3.35)$$

$$\lambda_2 = \frac{\mu_o d^2 N^2 x}{2x^2 - w^2\theta^2} i \quad (3.36)$$

$$\lambda_3 = \frac{\mu_o d^2 N^2 (x + w\theta)}{2(2x^2 - w^2\theta^2)} i. \quad (3.37)$$

The magnetic coenergy for each of the individual pole faces is the integral of this flux linkage

$$W'_m = \int_0^i \lambda(i', x, \theta) di', \quad (3.38)$$

where $\lambda \triangleq \lambda_1 + \lambda_2 + \lambda_3$. The integration is simple since the flux linkage is a linear function of current, and so after some algebraic simplification we arrive at

$$W'_m = \frac{\mu_o d^2 N^2 x}{2x^2 - w^2 \theta^2} i^2 \quad (3.39)$$

for the total magnetic coenergy. Finally, it is just a matter of taking partial derivatives of the magnetic coenergy with respect to x and θ in order to obtain the applied force and torque,

$$F = -\frac{\partial W'_m}{\partial x} = \frac{\mu_o d^2 N^2 (x^2 + w^2 \theta^2 / 2)}{2 (2x^2 - w^2 \theta^2 / 2)^2} i^2, \quad (3.40)$$

and

$$T = -\frac{\partial W'_m}{\partial \theta} = -\frac{\mu_o d^2 N^2 w^2 x \theta}{2 (2x^2 - w^2 \theta^2 / 2)^2} i^2. \quad (3.41)$$

Note that if θ is zero, the equation for the force reduces to the expected form of current squared over gap squared, and the torque term disappears. We can combine the multiple constants into a single value $C = \mu_o d^2 N^2 / 2$, and so arrive at the final equations for magnetic force and torque produced by the electromagnetic actuators,

$$F = C \frac{(x^2 + w^2 \theta^2 / 2)}{(x^2 - w^2 \theta^2 / 4)^2} i^2, \quad (3.42)$$

and

$$T = -C w^2 \frac{x \theta}{(x^2 - w^2 \theta^2 / 4)^2} i^2. \quad (3.43)$$

With these relationships, we can look at the effect that a non-uniform gap has on our actuators. Figure 3-7 shows the percent error in the applied force (relative to a parallel gap) that the rotation causes. The curves are parametric on gap, since the effects of the angular misalignment will be more pronounced at the smaller gaps. In the case of the Angstrom Stage actuators, even in a worst case scenario of a small gap and large rotation, the difference between the expected and actual forces will be less than one part in ten thousand.

Finally, we want to examine the effect of the actuator torque. An actuator with a parallel gap cannot apply a torque to a target, but a rotated one can. Figure 3-

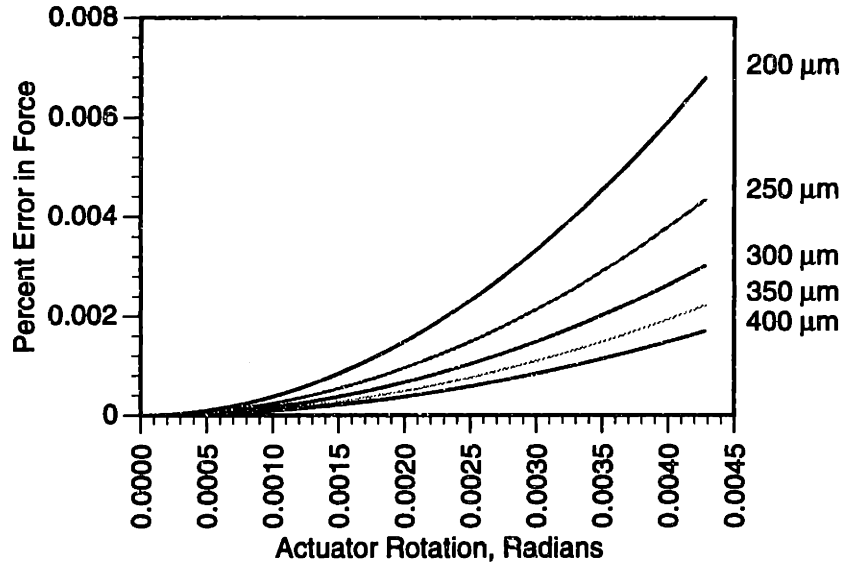


Figure 3-7: Percent error in the applied force due to actuator rotation.

8 shows the expected levels of torque at a nominal applied current of 100 mA. An order of magnitude set of calculations can show whether this amount of torque could cause unwanted stage motions. In the worst case scenario (manufacturing errors and platen position combining to form a large angular misalignment at a small gap), the maximum torque of $80 \mu\text{Nm}$ would act at the minimum damping of about 500 Ns/m . This leads to a rotational velocity of $0.16 \mu\text{rad/s}$. Dividing this by the characteristic actuator spacing ($\ell = 0.0665 \text{ m}$) transforms it into an expected velocity at the actuators of about 10 nm/s . This number is about two orders of magnitude smaller than typical stage velocities. However, this does appear to be a significant effort that could lead to undesired coupling between the lateral and vertical axes. Section 7.2 presents an example and expands upon this possibility.

3.5 Kinematics

The stage kinematics describe the coordinate transformations between the locations of the electromagnetic actuators and capacitance probes and the Cartesian axes. The stage appears as a rectangular prism that can translate in three orthogonal directions and rotate about each of the three axes. An important simplification is the assumption of small angles, meaning that $\tan \theta \cong \theta$. Further, the assumption of small angles, linearizes

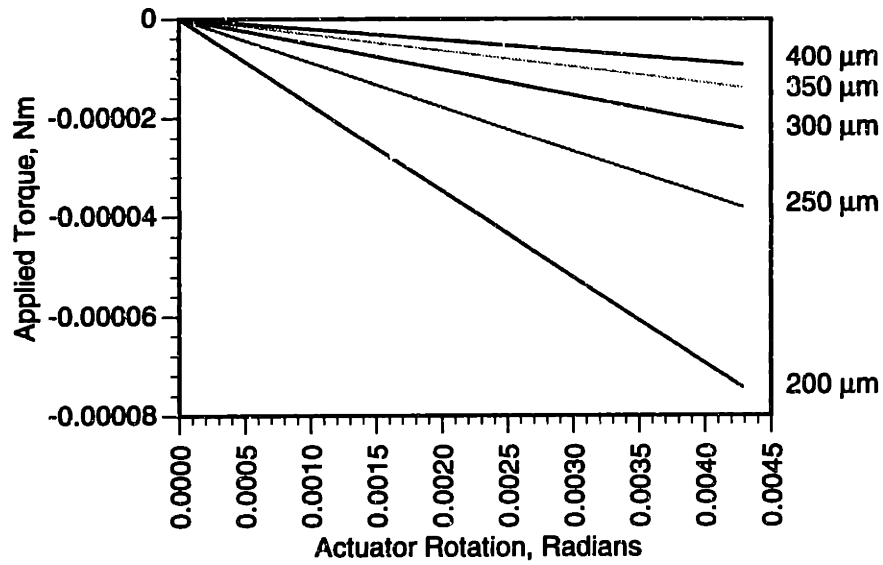


Figure 3-8: Torque applied by an actuator due to actuator rotation.

the kinematics in each of the three translations and rotations. Because they are based entirely on the geometry of the platen, the coordinate transformations are given simply as a matrix variable. The algebraic and numerical components appear in Appendix B.

3.5.1 Measurement Sensitivity Matrix

The 6×6 measurement sensitivity matrix \mathbf{H} converts position and orientation of the platen in the Cartesian reference frame into measurements taken at the capacitance probe locations. The capacitance probe vector \mathbf{Y} , contains the displacement (in meters) of each capacitance probe from its nominal gap, and so from the stage geometry,

$$\mathbf{Y} = \mathbf{H}\mathbf{X}, \quad (3.44)$$

where \mathbf{X} is the vector of platen translations and rotations. The symbolic and numerical values of \mathbf{H} are given in the appendix, Section B.1.

3.5.2 Modal Force Transformation

The electromagnets are arranged in opposing pairs, and apply forces at their individual locations, rather than at the center of the platen. Therefore, a coordinate transformation

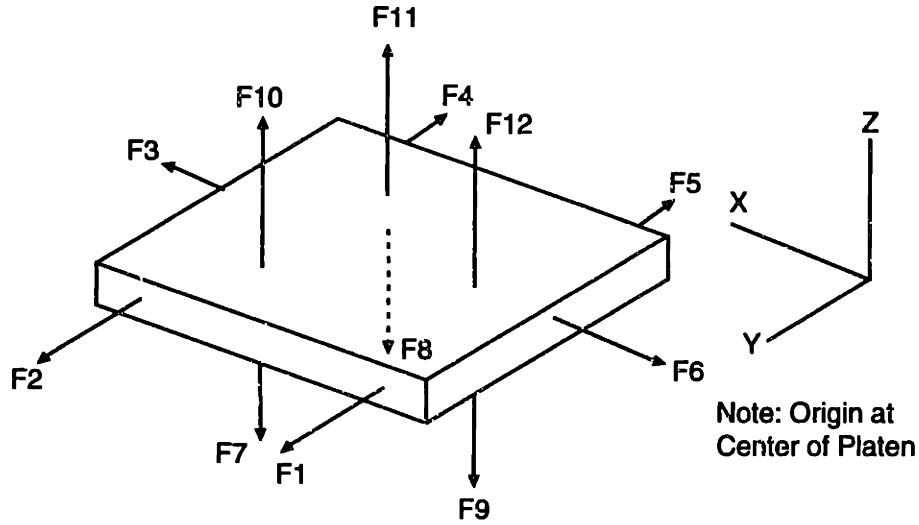


Figure 3-9: Positions of the force vectors located about the platen (from Holmes).

is necessary in order to convert the forces at the actuator locations \mathbf{F}_{em} into forces and torques at the center of the platen \mathbf{F} . These force vectors can be seen in Figure 3-9. It is assumed in calculating the vector \mathbf{F}_{em} that opposing actuator pairs operate together in order to yield a net force at that location. Therefore, the electromagnet vector is defined as

$$\mathbf{F}_{em} = \begin{bmatrix} F_{1,5} & F_{2,4} & F_{3,6} & F_{10,7} & F_{11,8} & F_{12,9} \end{bmatrix}^T. \quad (3.45)$$

Here, the force $F_{i,j}$ represents the net force applied in the i direction by the actuator pair i,j . This can be transformed into forces and torques referenced to the center of the platen,

$$\mathbf{F} = \begin{bmatrix} F_x & F_y & F_z & T_x & T_y & T_z \end{bmatrix}^T, \quad (3.46)$$

through the relationship

$$\mathbf{F}_{em} = \Phi_F \mathbf{F}. \quad (3.47)$$

The modal force transformation matrix, Φ_F is a function of the stage geometry, and is given in the appendix, Section B.2.

3.6 Gap Estimation

The plant model and the control system design both rely heavily on the actuator and capacitance probe gaps. These are nominally 300 microns and 100 microns respectively, but manufacturing errors can cause them to differ from those values. In addition, the setting of the offset on the capacitance probe electronics (so that zero volts corresponds to 100 microns) was an inexact science. Measurements taken with a coordinate measuring machine gave the inside dimension of the frame as well as the size of the platen. However, these measurements could only give the sum of the gaps of opposing actuator pairs in the horizontal degrees of freedom. It could not give their exact values. In addition, CMM data could not provide any information about the vertical gaps, since the top of the frame had to be removed in order to take the measurements. However, some methods were useful in bringing the approximations of the gaps closer to their true values.

A software simulation of the plant was crucial in estimating the gaps. We compared experimental stage responses to the simulated ones, and adjusted the gaps in the simulation in order to bring the two closer together. However, the large number of gaps (eighteen), and the interactions between them made tuning the gaps by hand a time consuming exercise. Two different methods were instead used to estimate the horizontal and vertical gaps.

There are six actuator gaps and three capacitance probe gaps in the horizontal degrees of freedom. However, by exploiting the CMM data and the physically required relationships between the gaps (i.e. for the gap of E3 to shrink, the gap of E6 has to grow), the number of independent variables could be reduced from nine to three. The optimization routines available in MATLAB and Microsoft Excel could then be used to calculate gaps that minimized the error between the actual and simulated stage responses in the horizontal degrees of freedom.

Calculating the gaps in the vertical degrees of freedom was more difficult since there was not CMM data capable of reducing the number of independent gaps as in the case of the horizontal axes. However, the three upper actuator gaps could be estimated by using the fact that the density matching of the platen was not exact. Under closed loop

Actuator	Revised Gap Estimate	Cap Probe	Revised Gap Estimate
E1	266 μm	C1	80 μm
E2	308 μm	C2	94 μm
E3	318 μm	C3	96 μm
E4	279 μm	C4	100 μm
E5	299 μm	C5	100 μm
E6	305 μm	C6	100 μm
E7	300 μm		
E8	344 μm		
E9	300 μm		
E10	278 μm		
E11	300 μm		
E12	294 μm		

Table 3.1: Revised estimates of the null gaps.

control then, a certain amount of current is required to keep the platen suspended and this current changes as the setpoint in Z is varied. However, regardless of the setpoint, the disturbance force in Z has to remain the same. Measuring the current at different setpoints and again using optimization software led to the setting of the three upper actuator gaps. The others remain at their nominal settings in the controller.

The gap adjustments were not made by any exact method, but did yield a model which more closely followed the plant. The errors in the actuator gaps were as high as 44 μm , but the main error uncovered was that the null gap of cap probe 1 was 80 μm rather than 100 μm . At the limit of the stage travel when this gap is smallest the error results in a damping value that is over four times higher than was originally anticipated. The main reason for refining the gap estimates was to be able to command constant velocity scans that relied on the plant model rather than on feedback control of a differentiated position signal. To that end, refining the gaps in the horizontal degrees of freedom was the most important since these are the typical scanning directions. Table 3.1 contains the updated estimates of the true actuator and capacitance probe gaps when all of the capacitance probes are reading zero volts. Those for which the estimates remain at 100 μm and 300 μm are those for which I could obtain no better estimates of the null gaps. The estimates could perhaps be refined via further experimentation. However, such efforts have not been undertaken in this thesis.

Chapter 4

Noise Propagation and Modeling

4.1 Introduction

If positioning resolution only on the order of nanometers were the goal of this research, controlling the stage would be reasonably straightforward. Any errors between the position measurements and the actual state of the plant are small enough to be ignorable at this scale. However, the goal of this research is to obtain positioning capabilities on the order of tenths of nanometers. This resolution is significantly finer than the level of noise in the sensor system. In addition, the floor-borne vibrations may be large enough to degrade the positioning performance of the stage. Because the above are random disturbances, a basic understanding of stochastics is needed in order to model the propagation of these signals through the system, and thereby predict their effects. I was unfortunately not able to find a single outstanding reference on the subject. The two that I found to be the most useful were Mortensen [41] and Gray [20]. To a lesser extent, Chapter 19 of Seibert [47] addresses random signals as do appendices in both Oppenheim and Schaeffer [44] and Franklin, Powell, and Workman [15].

This section begins with an introduction to stochastics which is aimed at showing how a random waveform is altered when passed through a linear system. The basic result is used in a single degree of freedom model of the stage that determines the noise-optimal gain, and predicts what the final position error variance will be. This noise transmission model is developed symbolically in continuous time. Finally, the remainder

of this chapter is devoted towards modeling each of the noise sources present in the stage, and incorporating them into the single degree of freedom model. The results here could readily be extended to the full six stage degrees of freedom, although that issue is not addressed further in this thesis.

4.2 Basic Stochastics

One distinctive characteristic of random signals is an obvious one: they look random. The waveform is so complex and detailed that one cannot predict exactly where the next point will be based upon all of the previous points. Instead, the signal has to be described in terms of probabilities. Where is the next point most likely to be? With that goal, several parameters can be defined to help describe the random waveform. These are presented here in terms of their continuous time formulations since that will eventually allow a discussion of the noise properties of this stage design without adding the additional complexity of finite sampling time. However, I feel that the derivation of the parallel discrete time relationships is easier to understand, and for that reason I have included the discrete time derivations in Appendix E.

The first property of a waveform that we have to define is the expected value of a waveform,

$$\langle x(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) dt. \quad (4.1)$$

The value of a waveform is also known as its *mean*, and it is given the value μ_x . Another quantity that we would like to know about a waveform is its spread about a mean value. One such measure is the *variance*, defined as

$$\sigma_x^2 = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T (x(t) - \mu_x)^2 dt. \quad (4.2)$$

A related measure is the mean squared value of a waveform, which is simply

$$\langle x^2(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x^2(t) dt. \quad (4.3)$$

Notice that if the mean of the waveform is zero (as is often the case in control systems) then the mean square value and the variance are the same. This fact will come in very useful later.

A more general measure of the “spread” of a waveform is the *autocorrelation* function,

$$R_x(\tau) \equiv \langle x(t) x(t - \tau) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) x(t - \tau) dt. \quad (4.4)$$

Notice that the value of the autocorrelation function evaluated at $\tau = 0$ is the mean squared value of the waveform, and that if it is a zero mean signal, the mean squared value also equals the variance.

The Fourier transform of the autocorrelation function is known as the *spectral density* of the waveform,

$$S_x(\omega) = \int_{-\infty}^{\infty} R_x(\tau) e^{-j\omega\tau} d\tau. \quad (4.5)$$

It is the spectral density of a waveform that is modified when it passes through a linear system. If the random waveform $x(t)$ passes through the linear system with an impulse response $h(t)$, then it yields a waveform $y(t)$ where the spectral density is given by

$$S_y(\omega) = |H(\omega)|^2 S_x(\omega). \quad (4.6)$$

Given the spectral density of an input waveform $S_x(\omega)$ (which can be measured experimentally), we can predict the spectral density of the output waveform. It may be more practically useful to know the variance of the output waveform however. Recall that if we are dealing with a zero mean waveform, then the variance is simply the value of the autocorrelation function at $\tau = 0$. Returning from the spectral density to autocorrelation requires taking the inverse Fourier transform, but since only the value at zero is important here, the inverse transform reduces to the area under the spectral density curve,

$$\sigma_y^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_y(\omega) d\omega. \quad (4.7)$$

These are important relationships which will be used in the next sections in order to

model the propagation of the sensor and disturbance forces through the single degree of freedom model.

4.3 Noise Propagation in One Degree-of-Freedom

The stage is a six degree-of-freedom device, and a detailed model of the noise propagation will reflect this. However, the general trends and design guidelines can be more clearly seen in the context of a single degree of freedom model, as shown in Figure 4-1. These results can be expanded to the other axes, but that work is not pursued in this thesis. The following assumptions were made in creating this simplified model.

- The platen is massless. The damping of the stage is so high that inertial effects are only significant at time scales on the microsecond level. At these time scales, the finite dimensional linear model is already inadequate, and so including the mass does nothing to support this analysis.
- The damping of the stage is constant, rather changing according to $1/\text{gap}^3$. This analysis is targeted towards addressing steady-state error motions; in that case the nominal capacitance probe gaps of $100\text{ }\mu\text{m}$ are only changing by nanometers. This gap change does not significantly affect the damping value.
- W_d is a disturbance force due to floor vibrations, D/A converter quantization, and other environmental disturbances.
- V_n is sensor noise that includes actual probe electrical noise, A/D converter quantization, and A/D converter electrical noise.
- The controller uses pure proportional gain. The use of integral control will not significantly alter this discussion, but we wanted to keep the resulting expressions as simple as possible and therefore use a pure proportional controller.

In this analysis, there are only two noise sources acting upon the platen, sensor noise and disturbance forces. They are described in general terms in this section, and later in Sections 4.4 and 4.5 they are specifically defined for this hardware. The goal of this

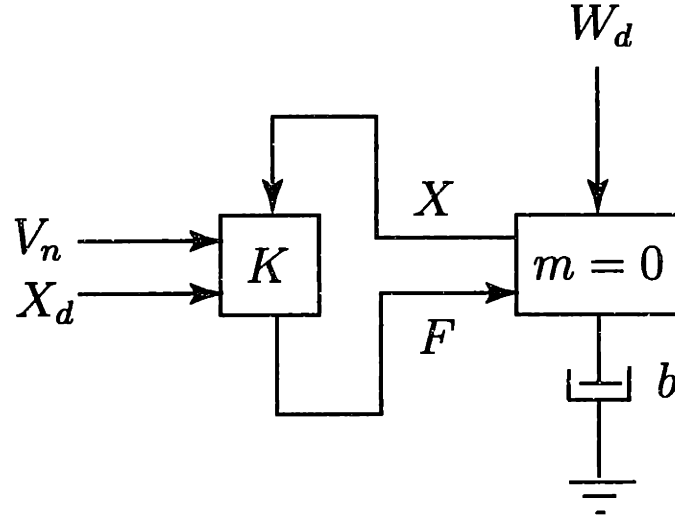


Figure 4-1: Free body diagram of a single degree-of-freedom model under proportional control.

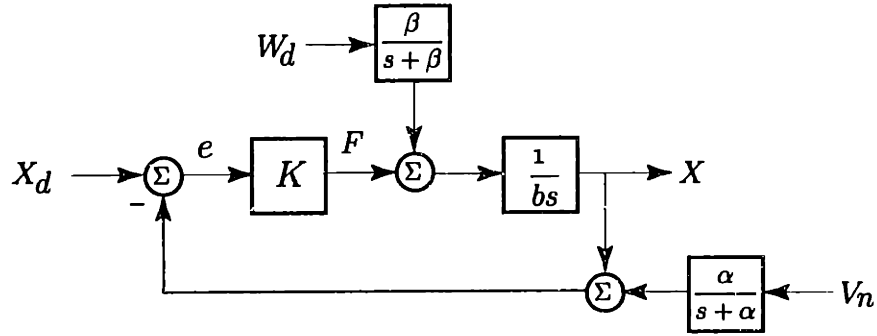


Figure 4-2: Linear block diagram of a single degree-of-freedom model.

section is to be able to predict the variance of the stage's position given the details of these disturbances.

4.3.1 Propagation of Sensor Noise

We know qualitatively that higher levels of sensor noise will lead to an increase in the stage's positioning variance. The results of Section 4.2 allow the relationship to be developed quantitatively. We now apply these results to a single degree of freedom, as shown in Figure 4-2.

The analysis begins with the transfer function from sensor noise to stage position,

$$\frac{X(s)}{V_n(s)} = \frac{-\frac{\alpha K}{bs(s+\alpha)}}{1 + \frac{K}{bs}}, \quad (4.8)$$

which can be simplified to

$$\frac{X(s)}{V_n(s)} = \frac{-\alpha K/b}{(s + \alpha)(s + K/b)}. \quad (4.9)$$

The sensor noise and position error are stochastic signals, and so the transfer function operates upon their respective spectral densities. Applying eq.(4.6) yields

$$S_x(\omega) = \left| \frac{-\alpha K/b}{(j\omega + \alpha)(j\omega + K/b)} \right|^2 S_n(\omega), \quad (4.10)$$

which then reduces to

$$S_x(\omega) = \frac{\alpha^2 K^2 / b^2}{(\alpha^2 + \omega^2)(K^2/b^2 + \omega^2)} S_n(\omega). \quad (4.11)$$

Recall that the spectral density is the Fourier transform of the autocorrelation function, $R_x(\tau)$. Therefore, taking the inverse transform of the expression for $S_x(\omega)$ will return the autocorrelation of the position error. More interesting for this application though is the value of the autocorrelation at $\tau = 0$, namely the mean squared value of the waveform. Because we are making the assumption of zero-mean waveforms, the mean squared value and the variance of the position are identical. Also, since we are only interested in the value of the autocorrelation at $\tau = 0$, the inverse Fourier transform reduces to evaluating the area underneath the spectral density curve.

Now, in order to transform the spectral density of the position error into the variance, we need to integrate $S_x(\omega)$ over all frequencies. This is not possible for an arbitrary noise spectral density $S_n(\omega)$. However, we have modeled the sensor noise as white noise with a constant spectral density S_n which then enters the low pass filter $\alpha/(s + \alpha)$. The position error variance due to the sensor noise is then determined by evaluating the integral

$$\sigma_{x,n}^2 = \frac{S_n}{2\pi} \int_{-\infty}^{\infty} \frac{\alpha^2 K^2 / b^2}{(\alpha^2 + \omega^2)(K^2/b^2 + \omega^2)} d\omega. \quad (4.12)$$

The best way to solve this integral is through a partial fraction expansion. That reduces

the equation to

$$\sigma_{x,n}^2 = \frac{S_n}{2\pi} \int_{-\infty}^{\infty} \frac{\frac{\alpha^2 K^2/b^2}{K^2/b^2 - \alpha^2}}{\alpha^2 + \omega^2} d\omega - \frac{S_n}{2\pi} \int_{-\infty}^{\infty} \frac{\frac{\alpha^2 K^2/b^2}{K^2/b^2 - \alpha^2}}{K^2/b^2 + \omega^2} d\omega, \quad (4.13)$$

which can now be solved analytically. Evaluating the integrals brings us to

$$\sigma_{x,n}^2 = \left(\frac{S_n}{2\pi} \right) \left(\frac{\alpha^2 K^2/b^2}{K^2/b^2 - \alpha^2} \right) \left[\frac{1}{\alpha} \tan^{-1} \left(\frac{x}{\alpha} \right) \Big|_{-\infty}^{\infty} - \frac{b}{K} \tan^{-1} \left(\frac{b}{K} x \right) \Big|_{-\infty}^{\infty} \right], \quad (4.14)$$

and then to simply

$$\sigma_{x,n}^2 = \frac{S_n}{2} \frac{K/b}{1 + K/\alpha b}. \quad (4.15)$$

Figure 4-3 shows the transmission of the sensor noise at different gains with the curves being parametric on the bandwidth of the incoming sensor noise signal. Note that if the incoming sensor noise is at a very low bandwidth, then the stage can operate at higher gains for the same amplification of sensor noise. Also, for the typical values of K/b , above sensor bandwidths of about 50 Hz, the curve closely approximates the linear case of infinite sensor bandwidth. Significant simplification can therefore be had by reducing eq.(4.15) to

$$\sigma_{x,n}^2 = \frac{K}{2b} S_n. \quad (4.16)$$

In making the assumption that the sensor noise has a very high bandwidth, we arrive at a simple linear relationship between the spectral density of the sensor noise and the variance of the resulting position error. Note that increasing the gain K increases the position error variance, and that increasing the damping b reduces it. However, with reference to Figure 4-2, we can see that the loop crossover frequency occurs at K/b . Therefore, for a given positioning bandwidth, K/b will be a constant, and so too will be the position variance. With only that point of view, it would appear that the actual damping in the stage is insignificant, since any value of b can be compensated for with the proportional gain to set a desired K/b ratio. We have to look at the transmission of disturbance forces in order to see the clear advantages of using a highly overdamped

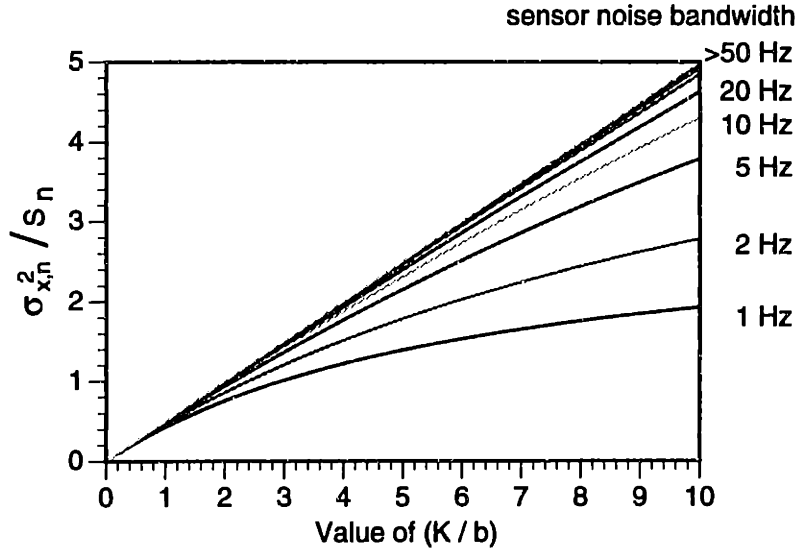


Figure 4-3: Propagation of the sensor noise according to the controller gain setting.

system.

4.3.2 Propagation of Disturbance Forces

The other input that affects the variance of the position of the stage is the disturbance force. As we will see in Section 4.5, this force disturbance can model the effect of base vibrations as well as those forces applied directly to the stage. Once again the framework developed in Section 4.2 can be used to predict specifics of the disturbance force propagation. Referring back to Figure 4-2, the transfer function between disturbance force and position becomes

$$\frac{X}{W_d}(s) = \frac{\beta/b}{(s + K/b)(s + \beta)}, \quad (4.17)$$

with the corresponding relationship between their spectral densities

$$S_x(\omega) = \frac{\beta^2/b^2}{(\beta^2 + \omega^2)(K^2/b^2 + \omega^2)} S_w(\omega). \quad (4.18)$$

Assuming that the disturbance force entering the low pass filter has a constant spectral density of S_w , the variance of the position error is found through

$$\sigma_{x,w}^2 = \frac{S_w}{2\pi} \int_{-\infty}^{\infty} \frac{\beta^2/b^2}{(\beta^2 + \omega^2)(K^2/b^2 + \omega^2)} d\omega. \quad (4.19)$$

The function can be most easily integrated by taking a partial fraction expansion of the integrand. This leads to

$$\sigma_{x,w}^2 = \frac{S_w}{2\pi} \int_{-\infty}^{\infty} \frac{\frac{\beta^2/b^2}{K^2/b^2 - \beta^2}}{(\beta^2 + \omega^2)} d\omega - \frac{S_w}{2\pi} \int_{-\infty}^{\infty} \frac{\frac{\beta^2/b^2}{K^2/b^2 - \beta^2}}{(K^2/b^2 + \omega^2)} d\omega, \quad (4.20)$$

and ultimately to

$$\sigma_{x,w}^2 = \frac{S_w}{2} \frac{1}{Kb + K^2/\beta^2}. \quad (4.21)$$

Rewriting this equation as a function of K/b more clearly shows the dependence of the transmission ration on the damping value. Dividing through by b^2 transforms the equation into

$$\sigma_{x,w}^2 = \frac{S_w}{2} \frac{1/b^2}{K/b + (K/b)^2 / \beta^2}. \quad (4.22)$$

Now it becomes even more apparent that higher loop crossover frequencies (K/b) reduce the transmission of the force disturbances. In addition, there is a $1/b^2$ relationship between the damping and the force disturbance transmission, so high damping will quickly reduce the position variance. As Figure 4-4 shows, we can again simplify the analysis by assuming a large bandwidth on the incoming disturbance forces. With this, eq.(4.21) reduces to

$$\sigma_{x,w}^2 = \frac{S_w}{2Kb} = \frac{S_w}{2b^2} \frac{1}{K/b}, \quad (4.23)$$

and the position variance is just inversely proportional to the loop crossover frequency K/b , and inverse square proportional to the stage damping b .

4.3.3 Combined Noise Propagation Model

The preceding sections show the fundamental tradeoff involved in choosing the controller gain. Higher proportional gain decreases the propagation of disturbance forces, but increases the propagation of sensor noise. Figure 4-5 shows this graphically. We need to choose the value of the gain which minimizes the variance of the position error. This

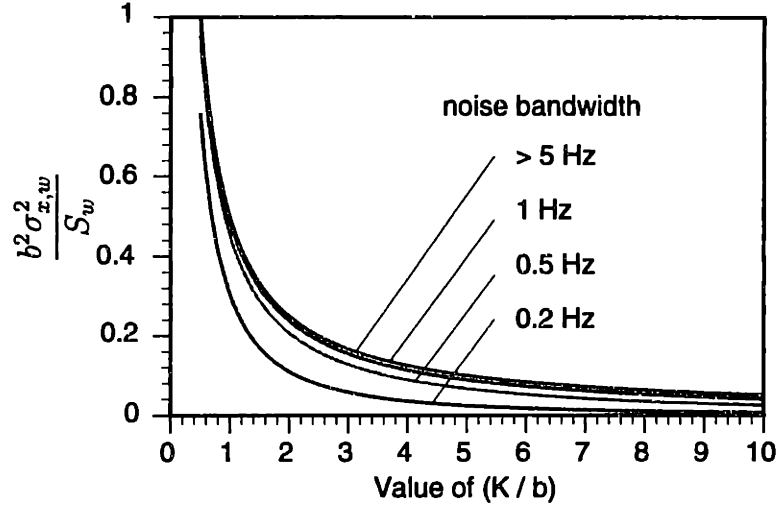


Figure 4-4: Propagation of the disturbance forces according to the controller gain setting.

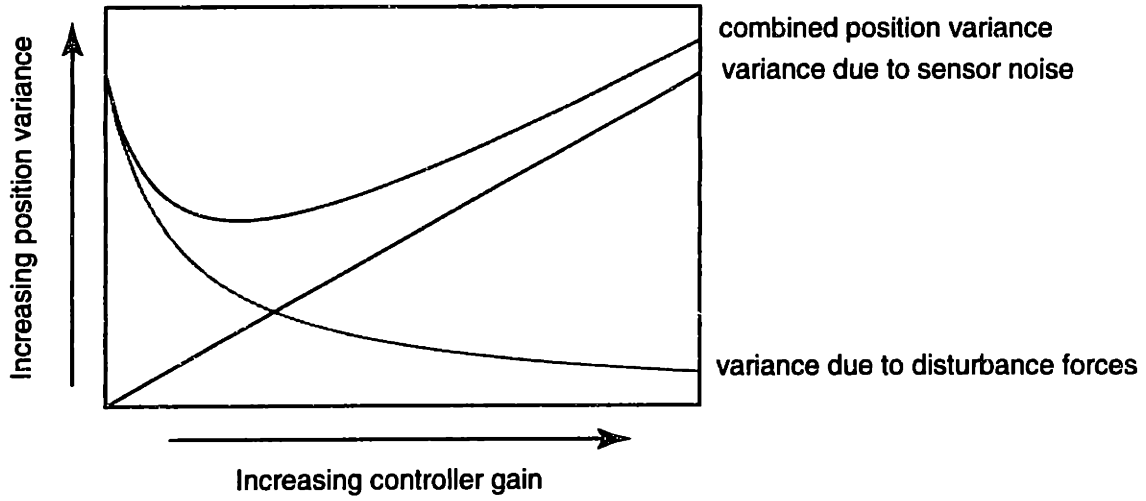


Figure 4-5: Net position variance as a sum of sensor noise and disturbance force transmissions.

position error is a linear combination of the two disturbances,

$$\sigma_x^2 = \frac{S_n}{2} \frac{K/b}{1 + K/\alpha b} + \frac{S_w}{2} \frac{1}{Kb + K^2/\beta^2}. \quad (4.24)$$

The proportional gain that minimizes the position error variance can be found by setting the derivative of eq.(4.24) equal to zero, to yield

$$\frac{d(\sigma_x^2)}{dK} = \frac{S_n}{2} \frac{b + 2K/\beta^2}{(Kb + K^2/\beta^2)^2} - \frac{S_w}{2} \frac{1/b}{(1 + K/\alpha b)^2} = 0. \quad (4.25)$$

Solving for the proportional gain K leads to a 4th order equation,

$$\frac{S_n}{b\beta^4}K^4 + \left(\frac{2S_n}{b^2\beta^2} - \frac{2S_w}{\alpha^2\beta^2b^2}\right)K^3 + \left(S_nb - \frac{S_w}{\alpha^2b} - \frac{4S_w}{\alpha\beta^2b}\right)K^2 - \left(\frac{2S_w}{\alpha} + \frac{2S_w}{\beta^2}\right)K - bS_w = 0 \quad (4.26)$$

If the constants are known, this equation can readily be solved analytically. However, recall that we could simplify the equations for the transmission of sensor noise and disturbance forces greatly if we made the assumption that they entered the system with a very high bandwidth. As was seen in Figures 4-3 and 4-4, even finite bandwidths can be modeled as infinite with little difference in the transmission ratio. Making the assumption that the disturbance bandwidths are very high, and thus that the variance is represented by eq.(4.16) and eq.(4.23), reduces eq.(4.26) to

$$S_nbK^2 - bS_w = 0, \quad (4.27)$$

which can then be solved for the noise-optimal gain K ;

$$K = \sqrt{\frac{S_w}{S_n}}. \quad (4.28)$$

The noise-optimal gain is thus the square root of the ratio of the disturbance force and sensor noise spectral densities. This gain value follows our intuition. Higher disturbance force spectral densities S_w require a higher controller gain to counteract the true forces acting on the plant. Larger sensor noise S_n requires a lower controller gain to prevent the controller from moving the stage to correct for a non-existent error motion. Using this value for the controller gain, we can predict what the final position error covariance will be. Inserting the optimal gain of eq.(4.28) into the net position variance calculation of eq.(4.24) (and again making the assumption of high disturbance bandwidths) yields,

$$\sigma_x^2 = \frac{\sqrt{S_w S_n}}{b}. \quad (4.29)$$

Note the simple form of the variance after all of the derivation required to produce it. As the sensor and disturbance force spectral densities rise, the position variance rises as

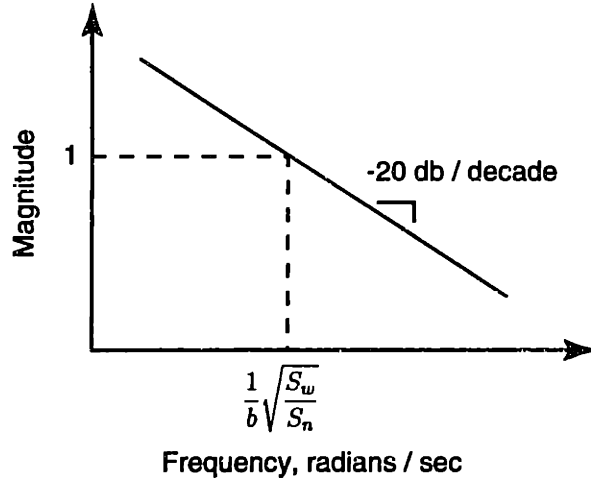


Figure 4-6: Loop transmission magnitude at the noise-optimal gain.

well. In addition, the damping term in the denominator indicates that higher damping does reduce the position variance as we would expect. It is also worth noting that the relationship between damping and achievable position variance is linear, and so doubling the damping will reduce the position variance by a factor of two. In comparison, reducing the spectral density of either the sensor noise or the disturbance force by a factor of two will only reduce the position variance by a factor of $\sqrt{2}$.

A look at the loop transmission reveals that we do pay a price for the damping-supplied position stability. The loop crossover frequency occurs at K/b , as shown in Figure 4-6. Substituting the noise optimal proportional gain value of $\sqrt{S_w/S_n}$ places the crossover frequency at

$$\omega_c = \frac{1}{b} \sqrt{\frac{S_w}{S_n}}. \quad (4.30)$$

Higher damping reduces the noise-optimal crossover frequency, and so although additional damping reduces the position variance, it also reduces the bandwidth. This is initially disturbing, since throughout this analysis additional damping has always improved the stage performance. However, consider a slightly different viewpoint. Could the stage really be expected to perform faster with additional damping? The damping allows for a low position variance, but to achieve this requires operating the stage more slowly.

Another way to view this set of tradeoffs is that is that for a desired stage bandwidth (K/b) we can increase b to the point that the stage position variance is dominated by sensor noise, and base vibration is insignificant. In this regime, the position variance is given by eq.(4.16), i.e. $\sigma_x^2 \cong \frac{K}{2b} S_n$. The use of damping thus allows the system to operate at the sensor noise limits. To reduce the noise further requires a better sensor or lower bandwidth.

Up to this point, we have developed all of the relationships in symbolic notation. This illuminates the system design tradeoffs. To apply the results to a specific system requires developing numeric values for the noise parameters. In this area that I was not able to find much help in the literature. Most presentation begin at the point where the spectral density of the noise is known, and do not explicitly detail the jump from an experimental time trace to a bandlimited spectral density. The remainder of this chapter develops such a model for the oil-floated stage.

4.4 Sensor Noise

The results of Section 4.3.1 show that higher levels of sensor noise lead to higher levels of position variance. The origin of the sensor noise itself was not expanded upon. There are three main sources of sensor noise in the Angstrom Stage, the capacitance probe electronics, quantization in the A/D converter, and electrical noise in the A/D converter. Each of these sensor noise sources is described in detail in the following sections.

4.4.1 Capacitance Probe Noise

The capacitance probes themselves are not ideal, and some amount of electrical noise is present in their position reading. Batchelder [1] experimentally measured the performance of the probes, and found that they had a peak-to-peak noise of 0.1 mV when measured through a bandwidth of 100 Hz. Assuming that the noise follows a Gaussian distribution, the peak-to-peak value corresponds to about six standard deviations, and so the standard deviation of the probe electrical noise is 0.0166 mV, corresponding to a position standard deviation of 0.0833 nm, and a position variance of 0.00694 nm². Draw-

ing on the examples in the previous section, we know that the spectral densities of the probe noise and the *measured* probe noise are related through

$$S_{probe}(\omega) = \frac{(200\pi)^2}{\omega^2 + (200\pi)^2} S_{probe, meas}(\omega). \quad (4.31)$$

We would like to know the underlying spectral density of the probe noise, but only know the variance of the filtered waveform. If we assume that the probe noise entering the filter is essentially white, we know that its spectral density and the measured variance are related through

$$\sigma_{probe}^2 = \frac{S_{probe}}{2\pi} \int_{-\infty}^{\infty} \frac{(200\pi)^2}{\omega^2 + (200\pi)^2} d\omega \quad (4.32)$$

which evaluates to

$$S_{probe} = \frac{\sigma_{probe}^2}{100\pi}. \quad (4.33)$$

Using the above numbers, the spectral density of the probe noise S_{probe} is found to be $2.21 \times 10^{-23} \text{ m}^2/\text{Hz}$. The analysis of the probe noise was done for the sake of completeness. In practice, this noise level is so far below the noise levels of the other sources (the electrical noise in the A/D converter specifically) that the additional noise added by the probes can be safely neglected. If the other system noise sources can be reduced, this represents the limit of system performance with the present probes.

4.4.2 A/D Converter Quantization Noise

Quantization effects in the analog-to-digital converter cannot be neglected in the development of the stage control system. In our A/D, only 16 bits are available to represent the position of the stage over 100 microns. This leads to a least significant bit value of 1.5259 nm, and so the true stage position will always be somewhere within ± 0.7689 nm of the indicated position. The simplest model of this error treats it as uniformly distributed zero mean white noise [44]. The uniform distribution assumes that the true position of the stage is equally likely to be anywhere within a single quantization level, with samples uncorrelated from each other. Gray [19] points out that this model is not entirely accurate since prior conversion values do influence the value of the error, and

proceeds to develop a more rigorous model of the quantization noise. However, little would be gained here by applying a more detailed model to our A/D board. The noise in the A/D converter electronics completely overshadows quantization noise, and so the marginal benefit from a more precise model does not justify its application.

Therefore, the quantization noise is modeled as being uniformly distributed between any two consecutive discrete values. Defining Δ to be the minimum resolution of the quantizer, the variance of the noise is given by

$$\sigma_{A/D \text{ quant}}^2 = \frac{\Delta^2}{12}. \quad (4.34)$$

Plugging in $\Delta = 1.5259 \text{ nm}$ for the stage leads to a quantization variance of $\sigma_{A/D \text{ quant}}^2 = 0.194 \text{ nm}^2$.

Now comes an issue that requires some care. We need to convert the variance of a discrete time sampling of points into an equivalent continuous time spectral density. Finding the discrete time spectral density is straightforward since we have again assumed white noise. The discrete time spectral density is simply $0.194 \text{ nm}^2/(\text{rad/sample})$. To convince oneself of this, remember that the variance of a waveform has to equal the area under the spectral density over the bounds of $\pm\pi \text{ rad/sample}$, divided by 2π . Therefore, a units change is all that is required. However, we need the spectral density to be in units of nm^2/Hz rather than $\text{nm}^2/(\text{rad/sample})$. One way to go about this is through a direct unit conversion,

$$\left(\frac{0.194 \text{ nm}^2}{\text{rad/sample}} \right) \left(\frac{0.0000448 \text{ sec}}{\text{sample}} \right) \left(\frac{2\pi \text{ rad/sec}}{\text{Hz}} \right) \left(\frac{1 \text{ m}}{10^9 \text{ nm}} \right)^2 = 5.46 \times 10^{-23} \text{ m}^2/\text{Hz}. \quad (4.35)$$

This method gives a spectral density for the A/D converter quantization noise of $S_{A/D \text{ quant}} = 5.46 \times 10^{-23} \text{ m}^2/\text{Hz}$.

4.4.3 A/D Converter Electrical Noise

The analog-to-digital converter adds the most uncertainty to the position measurement of any element in the feedback loop. Attaching all of the input channels to a common ground

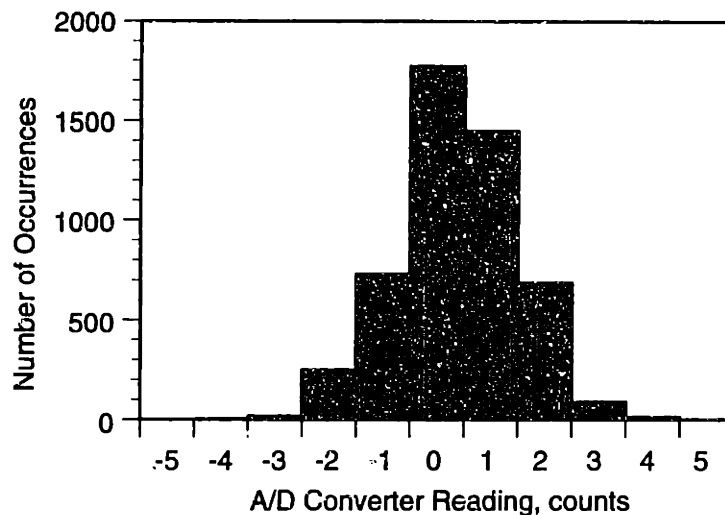


Figure 4-7: Histogram of the reading from a grounded A/D converter channel.

yields the best possible performance that can be expected from the board. The maximum throughput of the board is about 160 kHz, and sampling seven channels reduces the per-channel rate to about 22.3 kHz. In this configuration, the board should be reading exactly zero counts on every channel. Instead, the readings fluctuate by approximately six counts peak-to-peak. Figure 4-7 shows a histogram of this data. The converter noise shown here is normally distributed with a standard deviation of about 1.2 counts. At the conversion factor of 1.5 nm/count this corresponds to a 6σ width of about 11 nm in the position measurement.

We can look for structure in the noise by viewing its autocorrelation and power spectral density plots. Figure 4-8 shows the autocorrelation of a sample of the A/D converter noise. For truly uncorrelated white noise, the plot should show an impulse at zero with a magnitude equal to the variance of the waveform. The waveform of Figure 4-8 is close enough to the impulse function to model it as such, allowing the simplification of a white noise assumption.

Figure 4-9 shows the power spectral density plot of the readings from a grounded capacitance A/D converter channel, sampled at 22.3 kHz. The spectral density of the waveform is essentially constant, with a value in discrete time equal to the variance of the waveform, $3.24 \text{ nm}^2/(\text{rad/sample})$. Again, we have to convert this discrete time

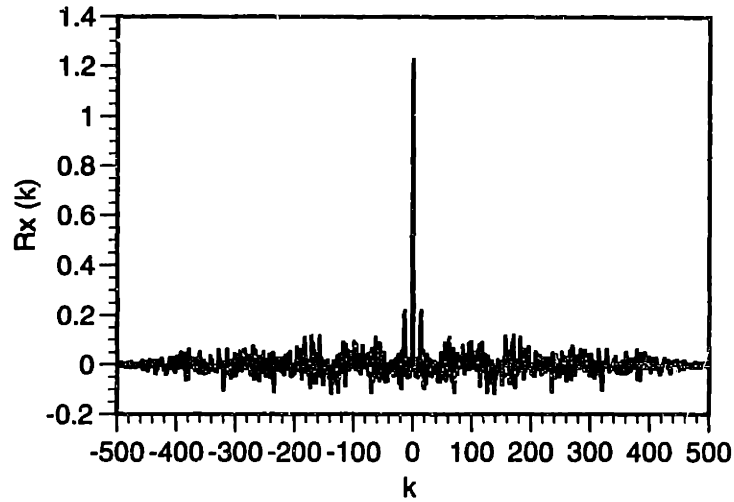


Figure 4-8: Autocorrelation of the A/D converter noise data.

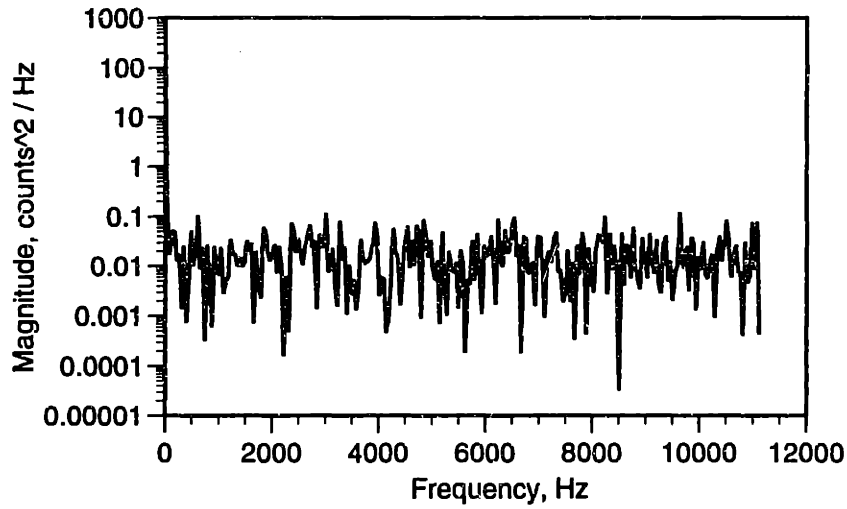


Figure 4-9: Power spectral density plot of the A/D converter noise data.

spectral density to it's continuous time equivalent. In doing so, we find that

$$S_{A/D \text{ elec}} = \frac{3.24 \text{ nm}^2}{\text{rad/sample}} \frac{2\pi \text{ rad/sec}}{\text{Hz}} \frac{0.0000448 \text{ sec}}{\text{sample}} \frac{1 \text{ m}^2}{10^{18} \text{ nm}^2} = 9.12 \times 10^{-22} \text{ m}^2/\text{Hz}. \quad (4.36)$$

Note that the spectral density of the electrical noise in the A/D converter is more than an order of magnitude higher than either the quantization or probe spectral densities. As such, it represents the largest noise source in the feedback loop.

4.4.4 Bandlimiting of Sensor Noise

In the previous sections, we found equivalent continuous time white noise spectral densities for each of the three sensor noise sources. However, these are actually bandlimited before entering the control system. The dominant source of this bandlimiting takes place in the filtering algorithm of the control code. For simplicity, we will assume that this is a simple moving average filter of the previous 100 points. In this case, the magnitude of the main lobe of the discrete time frequency response of the filter is at $2\pi \cdot 100$ rad/sample, which corresponds to a continuous time frequency of 223 Hz. This is the frequency at which the magnitude of the response first falls to zero. By inspection, the equivalent continuous time frequency at which the magnitude has fallen to -3 dB is about 100 Hz. Therefore, in the framework of Figure 4-2, the averaging routine on the control code is modeled as a first order low pass filter in which $\alpha = (200\pi)$ rad/sec. As a final check to see the validity of our numbers, we can sum all of the spectral densities, pass them through our low pass filter, and then calculate the resulting variance. The net spectral density is the sum of the three sources,

$$S_n = S_{probe} + S_{A/D \text{ quant}} + S_{A/D \text{ elec}} = 9.89 \times 10^{-22} \text{ m}^2/\text{Hz}. \quad (4.37)$$

Note again that this value is not significantly different than the spectral density of the A/D converter electrical noise itself. This again points to the A/D converter as the dominant source of measurement error.

Now, we pass the white noise with spectral density S_n through a low pass filter with a cutoff frequency of 100 Hz. Using the same integration as before,

$$\sigma_n^2 = \frac{S_n}{2\pi} \int_{-\infty}^{\infty} \frac{\alpha^2}{\omega^2 + \alpha^2} d\omega, \quad (4.38)$$

and being careful careful with the units, we arrive as

$$\sigma_n^2 = \frac{9.89 \times 10^{-22} \text{ m}^2/\text{Hz}}{2} (100 \text{ Hz}) = 4.95 \times 10^{-20} \text{ m}^2. \quad (4.39)$$

This corresponds to a standard deviation of 0.22 nm, and a 6σ peak-to-peak value of

1.3 nm which is in agreement with the experimentally observed value. To recap then, the sensor noises are large enough that after all of our filtering, each capacitance probe reading sent to the controller still occurs within a peak-to-peak range of 1.3 nm, making position control to the sub-nanometer level difficult. The method used for accomplishing this is an estimator-based controller; this is described in more detail in section 6.6.

4.5 Plant Noise

There are three main sources of plant noise: floor vibrations, D/A quantization, and a disturbance force applied directly to the platen. Vibrations in the table enter the stage model of Figure 4-2 as a disturbance force. Note that the magnitude of the table motions in themselves are not directly important, but it is the relative motion between the platen and the frame that concerns us. In other words, the table can move inches, and as long as the platen moves along with it, the accuracy of the stage will not be compromised. Quantization in the D/A converter enters the stage as a disturbance force since the actual applied force will not equal the commanded force. Finally, there is always the possibility that a disturbance force will be applied directly to the platen. This disturbance is not typically present in the current experimental setup, but could always be incorporated into the model by simply adding another variance to the previous two.

4.5.1 Disturbance Forces Due to Base Vibrations

We cannot measure the disturbance forces directly, but can experimentally measure the base accelerations. Therefore, we have to develop a model of the open loop system that will convert the base accelerations into plant disturbance forces. In its most basic open-loop form, the plant looks like a mass and a damper. With respect to base vibrations, we can view the mass as having an effective value of αm , where $\alpha = (\rho_{\text{platen}} - \rho_{\text{oil}}) / \rho_{\text{platen}}$ measures the relative success of density matching in the system design. The constant ρ_{platen} is the average density of the platen. Recall that the platen is designed with a hollow structure, as was shown in Figure 2-4 in order to set ρ_{platen} close to ρ_{oil} .

If the density matching were exact, then $\alpha = 0$ and the platen would always move

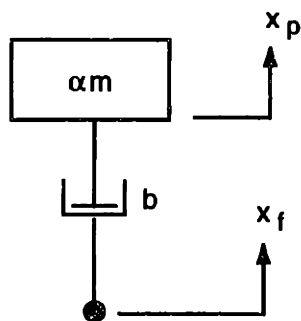


Figure 4-10: Free body diagram of the platen-frame system as by base vibrations.

along with the oil. There could then be no relative motion between the platen and the frame resulting from translational base vibrations. (The case for rotational vibration is more complex and is not presented here.) However, we know that density matching cannot be perfect, and so the value for α has to be experimentally determined. The other element of the model is simply a damper, b . For the purposes of this simple model, we assume that the dampers are linear, rather than their true inverse cube relationship. This assumption is justified due to the extremely small changes in the damper gap resulting from base vibrations. The basic two-element system is shown in Figure 4-10.

With this two-element model established, we can now show how the base accelerations are transformed into disturbance forces. This begins with a free body diagram of the system that leads to the equation of motion,

$$\alpha m \ddot{x}_p = b(\dot{x}_f - \dot{x}_p). \quad (4.40)$$

We also know that the disturbance force applied to the platen can only come from the damper, and so

$$f_d = b(\dot{x}_f - \dot{x}_p). \quad (4.41)$$

The absolute position of the platen relative to ground (x_p) cannot easily be experimentally measured, i.e. our probes measure $x_f - x_p$, and so we would like to eliminate this variable from the relationships. To do so, first take the time derivative of eq.(4.41),

$$\dot{f}_d = b(\ddot{x}_f - \ddot{x}_p), \quad (4.42)$$

solve for \ddot{x}_p , and substitute both it and eq.(4.41) back into eq.(4.40). This eliminates the dependence on the absolute platen position, and leaves us with

$$\alpha m \left(\ddot{x}_f - \frac{1}{b} \dot{f}_d \right) = f_d. \quad (4.43)$$

Now taking the Laplace transform of the above equation gives the open-loop transfer function from base accelerations to disturbance forces,

$$\frac{F_d(s)}{\ddot{X}_f(s)} = \alpha m \left(\frac{b/\alpha m}{s + \frac{b}{\alpha m}} \right). \quad (4.44)$$

At high frequencies (above $b/\alpha m$), the disturbance force transmission rolls off and the platen and frame are effectively “locked” together with respect to base accelerations. The mechanical design itself eliminates the transmission of high frequency noise, but the lower frequencies are transmitted with a magnitude proportional to αm , the effective mass of the platen.

An estimate of the platen mass relative to the density of the surrounding fluid is needed in order to use the results of eq.(4.44). In order to obtain this, I modified the control code to to display the approximate force output in each of the six directions. These are not the true forces because of uncertainties in the gaps and force constants, but they represent a reasonable approximation. For repeatability in this case, the null gaps are assumed to be 300 μm and the actuator force constant is $2.5 \times 10^{-6} \text{ Nm}^2/\text{A}^2$. The platen is suspended at different Z-positions and the force required gives an indication of the effective mass of the platen. Using this method, the effective mass αm ranges from 71.2 g to 72.8 g. For comparison, the mass is measured again with the sample holder removed. This time the forces range from 47.7 g to 52.0 g. The difference between these numbers correlates well with the true sample holder mass of 22.5 g. Therefore, we can call the effective mass of the combined platen and sample holder 72 g, meaning that the density matching of the platen offloads about 98% of the true platen mass, and thus $\alpha = 0.02$.

The transfer function between base acceleration and position error really only becomes

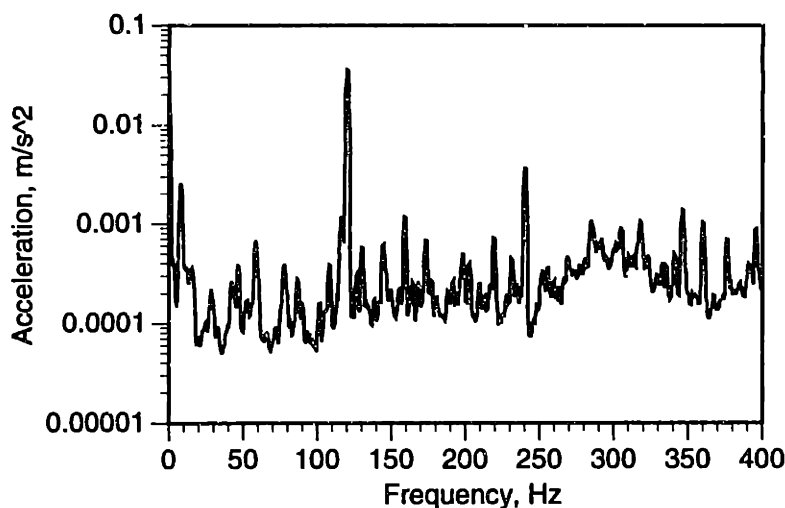


Figure 4-11: DFT of the floor accelerations from 0-400 Hz.

useful with some measure of amplitudes and frequencies of the base vibration. For this experiment, I used a PCB¹ seismic accelerometer² along with an HP 35665A Dynamic Signal Analyzer for measuring and recording the data.

I first measured the vibrations of the floor, and then moved on to the accelerations that are transmitted through the vibration isolation table to the stage frame itself. Figure 4-11 shows the frequencies that are contained in a typical sample of floor accelerations. This trace (as are all of the Fourier transform plots in this section) is the average of 25 individual traces as measured by the dynamic signal analyzer. Its most obvious feature is the large peak occurring at 120 Hz. This vibration can be felt by hand in the lab, and is believed to originate in some of the building HVAC equipment located in the adjoining room. However, the origin of this noise has not been rigorously pursued.

The time domain plot of Figure 4-12 further emphasizes the size of the floor accelerations at this frequency. As a cross check between the time and frequency domain plots, note that the amplitude in the time domain plot is about 0.05 m/s², which agrees with the amplitude of the peak in the frequency domain plot. The time domain plot also contains about 15 cycles of the waveform, which is what we expect for a 0.125 sec sample of a 120 Hz signal. The next highest peak in Figure 4-11 occurs at 240 Hz which is a harmonic of the 120 Hz signal, and there is also the suggestion of a low frequency peak.

¹PCB Piezotronics, Inc. 3425 Walden Ave., Depew, NY 14043-2495

²Model No. 393C, Serial No. 3536, Sensitivity = 1104 mV/g

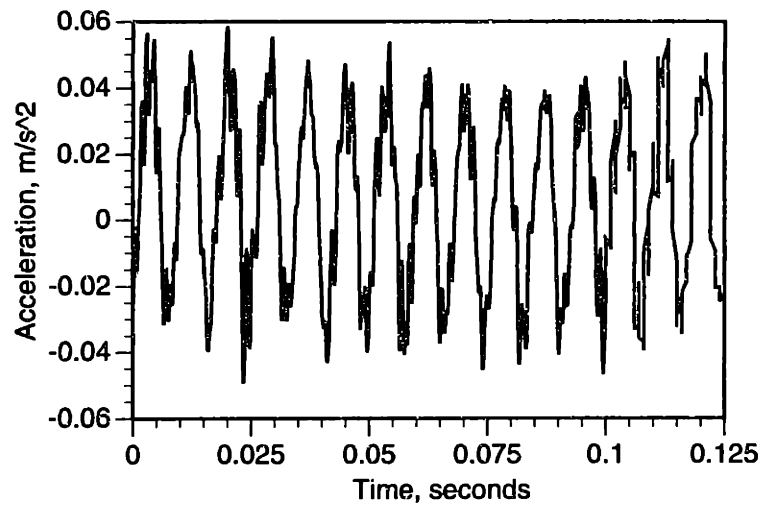


Figure 4-12: Time plot of the floor accelerations.

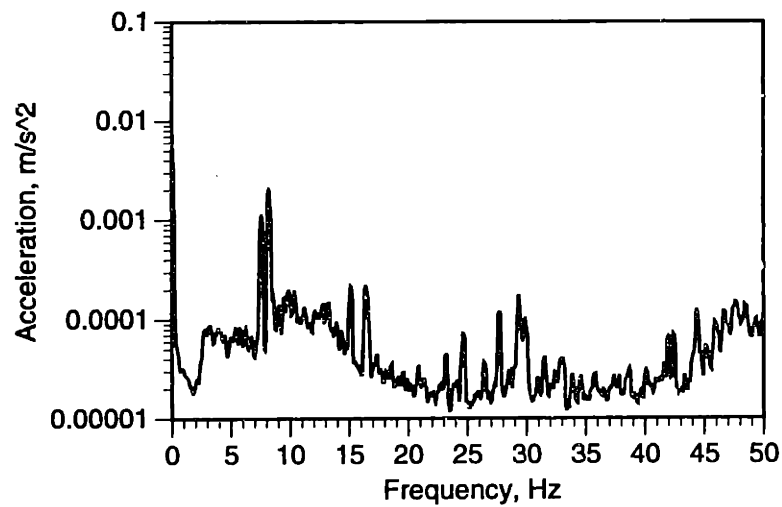


Figure 4-13: DFT of the floor accelerations from 0–50 Hz.

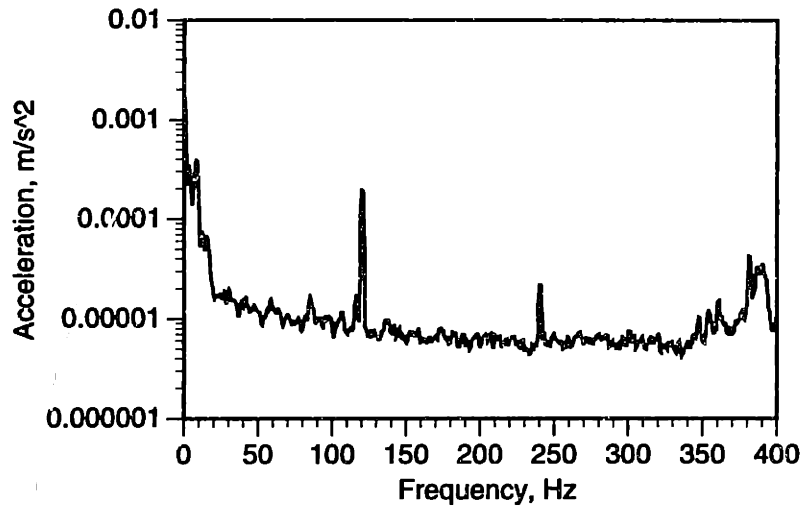


Figure 4-14: DFT of the table accelerations from 0-400 Hz.

The trace in Figure 4-13 displays the frequency content of the floor accelerations at frequencies below 50 Hz. There are two significant peaks present at frequencies of about 7.5 Hz and 8.125 Hz. At this scale, the resolution is limited to 0.0625 Hz. There are other peaks present in the plot, but these are the dominant ones.

The previous plots all looked at the frequencies contained in the floor vibrations. However, the stage rests on an air isolation table which itself rests on top of a workbench. Therefore, measurements of the table itself have to be taken directly in order to characterize the accelerations that the stage frame undergoes. Figure 4-14 shows the frequencies of the floor vibrations that pass through to the air table. The 120 Hz peak is still very apparent, as is its harmonic at 240 Hz. However, due to table filtering, the amplitudes of the peaks are attenuated by better than two orders of magnitude. Now the 120 Hz peak is at roughly the same magnitude as the low frequency (8 Hz) peak. The plot in Figure 4-15 shows more clearly the vibrations contained in the range of interest below 50 Hz.

The information suggested by the frequency domain analysis is again confirmed by a look at the time domain plots. Figure 4-16 contains a trace made to the same time scale as was Figure 4-12. The 120 Hz component is still somewhat visible, but is much more subdued. Now the 8 Hz component of the vibrations becomes more visible; the plot of Figure 4-17 shows this component clearly. It passes through the table attenuated only

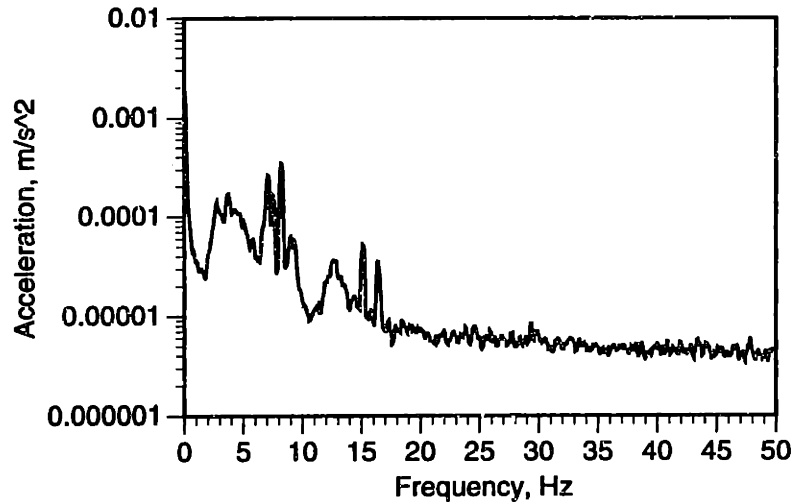


Figure 4-15: DFT of the table accelerations from 0-50 Hz.

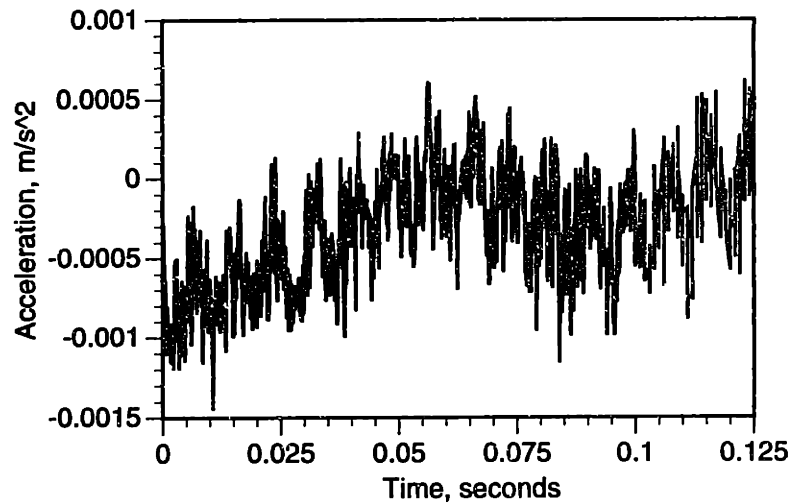


Figure 4-16: Time plot of the table accelerations showing high frequency component.

by about an order of magnitude.

Very late in this project, we purchased a new optical table for use with another project. However, there was some time for me to try it out with this hardware in order to see whether it was any better at attenuating the base vibrations. Because of its larger size (1.2 m × 1.2 m), I had expected it to perform better than the benchtop vibration isolation system. It showed a marginal improvement over the smaller table, but not significantly so. What became more clear was the fact the the floor vibrations are not constant, but that frequencies in discrete components come and go. Figure 4-18 shows the frequency components of two traces of the optical table vibrations taken within fifteen minutes of

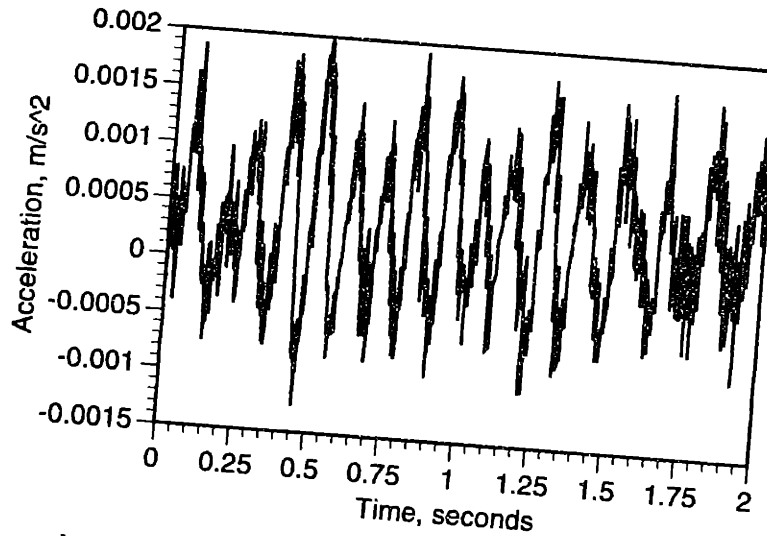


Figure 4-17: Time plot of the table accelerations showing low frequency component.

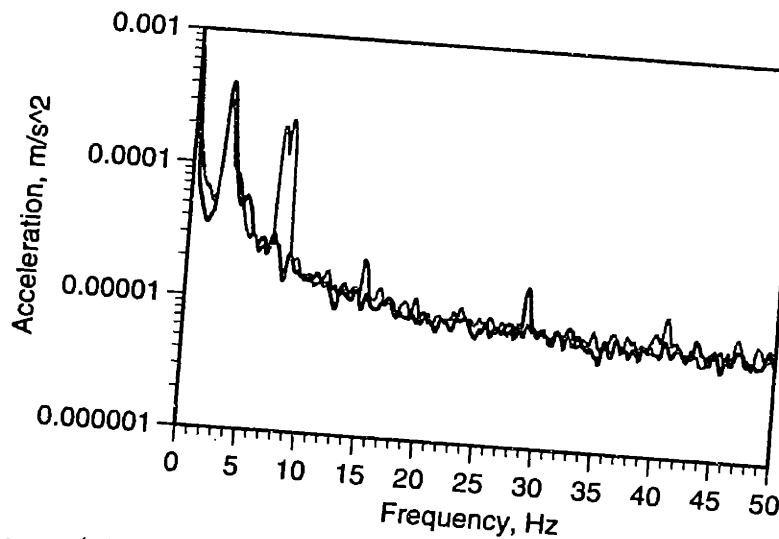


Figure 4-18: Two DFT's of the optical table accelerations.

each other. The first trace does not contain the 8 Hz component, but this frequency is highly prevalent in the second. At various other times, components at about 15 Hz and 28 Hz have also come and gone. The transitory nature of some of the components of the floor accelerations helps to explain the occasional difficulty in obtaining a good quality trace from the stage.

Overall, the floor vibrations in the present laboratory are far too “colorful” to ever actually be referred to as white noise. However, for the sake of the present analysis, I will come close to doing that. We will model the floor accelerations as white noise that has passed through a first order low pass filter. The spectral density and cutoff

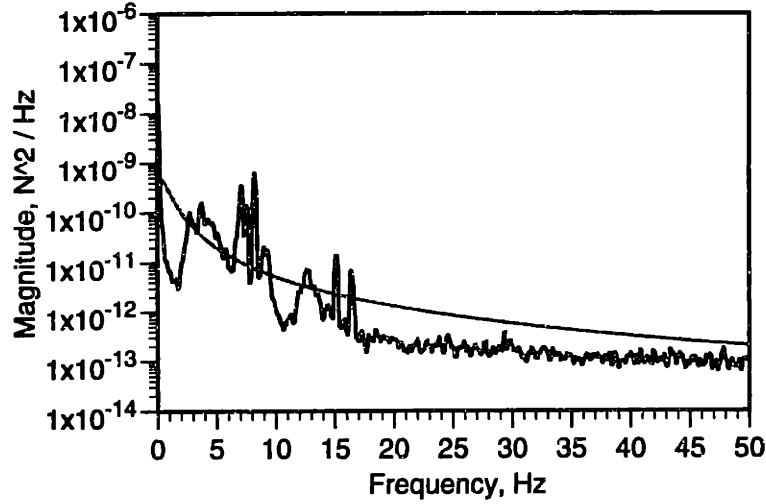


Figure 4-19: Measured and fitted power spectral density plots of disturbance forces.

frequency of this noise is determined empirically, and Figure 4-19 shows a comparison between the actual and modeled power spectral density plots of the floor vibrations. The measured spectral density of the accelerations was scaled by the squared effective mass of the platen (72 g) in order to transform the accelerations into disturbance forces. Therefore, by empirically fitting a model to the floor vibrations, we find that they can be represented as white noise with spectral density of $5 \times 10^{-10} \text{ N}^2$ that is bandlimited to 1 Hz.

4.5.2 D/A Converter Quantization Noise

There will also be a disturbance force associated with the quantization in the D/A converter. This is slightly less straightforward to analyze than was the case of A/D quantization since there is a *squared* rather than linear relationship between the current and force,

$$F = C \frac{i^2}{g_a^2}. \quad (4.45)$$

Performing a Taylor series linearization about a nominal current i_o leads to the approximation that

$$\Delta F = \frac{2Ci_o}{g_a^2} \Delta i. \quad (4.46)$$

The size of the force quanta depends upon the nominal current being applied. The variance of the disturbance force is therefore

$$\sigma_{D/A}^2 = \frac{4C^2 i_o^2}{12g_a^4} \Delta i^2. \quad (4.47)$$

The 12-bit DAC is set for an output range of 0–10 V, which leads to a voltage quantization level of 2.4414 mV. The approximate gain of the amplifiers is 34.5 mA / V leading to a current quantization value of $\Delta i = 84 \mu\text{A}/\text{count}$. The electromagnet force constant C is approximately $2.5 \times 10^{-6} \text{ Nm}^2/\text{A}^2$, and the nominal gap is 300 microns. Using these numbers, the variance due to D/A quantization becomes

$$\sigma_{D/A}^2 = \left(1.75 \times 10^{-6} \text{ N}^2/\text{A}^2\right) i_o^2. \quad (4.48)$$

The largest force quanta will occur at the full scale current of $i_o = 0.344 \text{ A}$, leading to a D/A quantization force variance of $\sigma_{D/A}^2 = 2.07 \times 10^{-7} \text{ N}^2$ in this extreme case. At a more reasonable nominal current of 0.150 A (roughly half full scale), the force variance due to the D/A converter reduces to $\sigma_{D/A}^2 = 3.94 \times 10^{-8} \text{ N}^2$. At the typical operating levels of 50 to 100 mA, these quanta become even less significant. In this sense, the current squared actuator characteristic is an advantage if the actuators are operated at low current.

The discrete-time spectral density of the D/A quantization force error is numerically the same as the variance, i.e $S_{D/A} = 3.94 \times 10^{-8} \text{ N}^2/(\text{rad}/\text{sample})$. Converting this into the equivalent continuous time value involves multiplying it by the sample period and then dividing by 2π in order to convert radians to Hz. The effective spectral density of the disturbance forces in continuous time is thus $2.81 \times 10^{-11} \text{ N}^2/\text{Hz}$.

4.6 Final Predicted Position Variance

The models of the noise propagation and measurements of the disturbances presented in this section can be used to predict the position error variance. Recall that

$$\sigma_x^2 = \frac{K}{2b} S_n + \frac{1}{2b^2} \frac{1}{K/b} S_w \quad (4.49)$$

and that the optimal controller gain to minimize noise is

$$K = \sqrt{\frac{S_w}{S_n}}. \quad (4.50)$$

From the results of the previous noise modeling sections, we know that the stage is subject to a combined sensor noise spectral density of $S_n = 9.89 \times 10^{-22} \text{ m}^2/\text{Hz}$ and disturbance force spectral density of $S_w = 5.3 \times 10^{-10} \text{ N}^2/\text{Hz}$. This leads to a noise optimal controller gain of $K = 0.73 \text{ N}/\mu\text{m}$. At a nominal stage damping coefficient of $b = 200000$, this leads to a closed loop system bandwidth of about 0.6 Hz, and a steady-state position standard deviation of 0.06 nm. The peak-to-peak (6σ) value of 0.36 nm is in good agreement with our experimental results shown in Chapter 7. Figure 4-20 shows the contribution of each of the five major sources of error to the position variance. Clearly, the A/D converter electrical noise and the floor vibrations are the dominant sources, and both significantly affect the final position variance. A quieter A/D converter would give more confidence in the position readings, allowing for a higher system bandwidth and rejection of disturbance forces. On the other hand, a quieter floor would allow us to reduce the stage bandwidth since there is no longer a need to compensate for as great a force disturbance.

Finally, we can review what has been learned through this section. First, we detailed how a noise signal is modeled, and how it is acted upon by LTI systems. Next, we symbolically derived the transmission of probe and sensor noise through a damper-dominated single degree of freedom model of the Angstrom stage. Higher controller gains raise the transmission of sensor noise while attenuating the transmission of base vibrations. The strongest advantage of the damper dominated design is that for a given positioning bandwidth, the transmission of disturbance forces attenuates according to $1/b^2$. We then

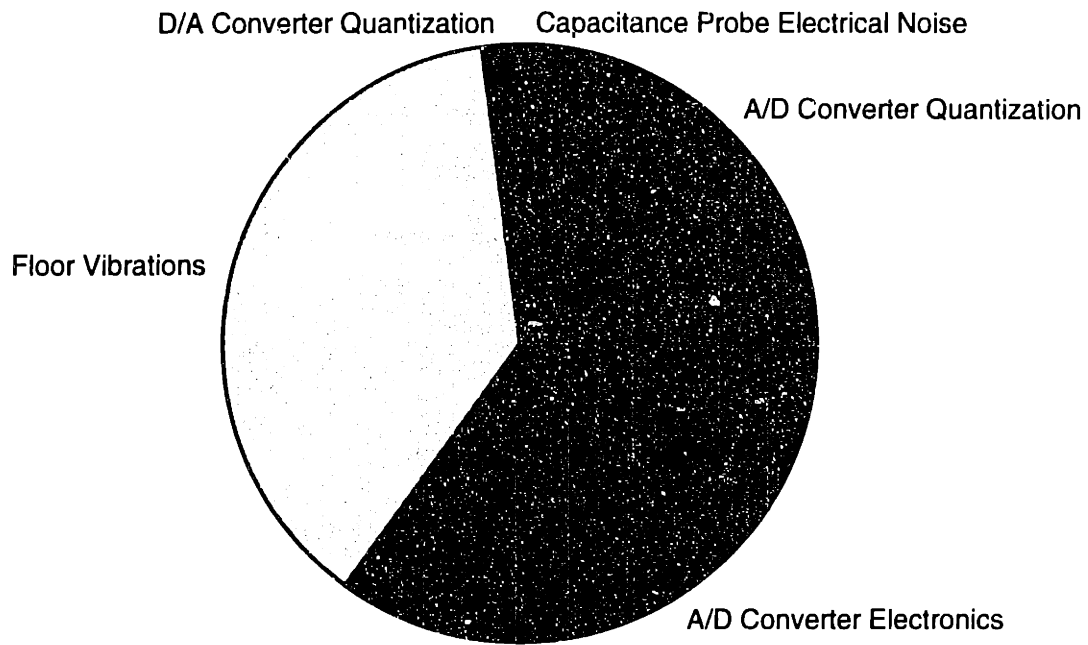


Figure 4-20: Pie chart showing contribution of each noise source to the position error.

modeled the actual sensor noise and disturbance forces present in the stage, and used our model to establish the controller gain which minimizes the positioning variance.

Chapter 5

Filtering

The biggest problem faced in designing the control system is the fact that the sensor system does not directly provide a high-enough level of resolution. Presently, the available analog-to-digital conversion hardware limits the capacitance probe readings to 16-bit resolution. Over the range of 100 microns, this results in position quanta of about 1.53 nm. This is a factor of 15 higher than the desired position resolution in and of itself; however noise in the A/D converter electronics raises this by another factor of six or seven. The end result is that the controller is trying to move the stage with 0.1 nanometer resolution when each reading occurs in a 10 nm fuzzball surrounding the true position. An appropriate digital filter can significantly attenuate these sources, since they contain significant energy outside the system bandwidth of approximately 5 Hz. Three possible filters are a simple arithmetic averaging of the points, a linear regression, or a finite impulse response (FIR) filter. Each of these is described in more detail below.

5.1 Arithmetic Average

This method is the sum of N readings divided by N . Since the noise is approximately distributed through a Gaussian distribution, the use of this method reduces the standard deviation of the position reading by a factor of $1/\sqrt{N}$. For example, averaging 100 readings reduces the noise level by a factor of 10. An unfortunate side effect of this method is that it adds an equivalent additional $T/2$ units of time delay to the feedback loop, where

T is the time over which the signal is averaged. The average thereby essentially computes the capacitance probe reading at the midway point between DAC outputs. In addition, there is an issue of aliasing in the frequency domain which requires that the controller not downsample by any more than half the number of points it averages. This concept will be described in more detail in the section on FIR design. Before realizing that though, in an attempt to eliminate this time delay, I experimented with using linear regression to fit a straight line to the incoming position data points in between DAC outputs.

5.2 Linear Regression

A simple average is useful if we know that the data are stationary with a fixed mean. Averaging more points keeps pulls the sample average closer to this true mean. However, the position readings from the capacitance probes are not stationary, since the stage is moving. Over the distance traveled by the stage in one sample period, the percentage change in the gaps is minimal, and so the force and damping are essentially constant in this interval. Thus the stage can be modeled as moving at constant velocity between sampling instants. The readings from the capacitance probes have an underlying structure, and so the controller can exploit it. If the stage moves with a constant velocity during each sample period, then the true position looks like a series of straight lines.

Performing a linear regression yields two benefits over a simple average. First, it eliminates the the $T/2$ time delay inherent in the averaging. The controller can use the endpoint of the line as the most recent estimate of the probe readings. Second, it gives an estimate of the stage velocity. The slope of the best fit line from the regression is the velocity of the stage during that sample period. Most any statistics textbook [25] or book on numerical methods [24] details the derivation of a linear regression, and so the derivation will not appear here. Basically, a least squares curve fit looks to minimize the sum of the squared difference between the data points and the best fit curve to them. Using this method, it is a simple matter to fit any polynomial to a data set, and a linear regression is just a simple form of that. In practice, this method reduced the variance of the position error slightly as compared with the simple averaging method, but the long

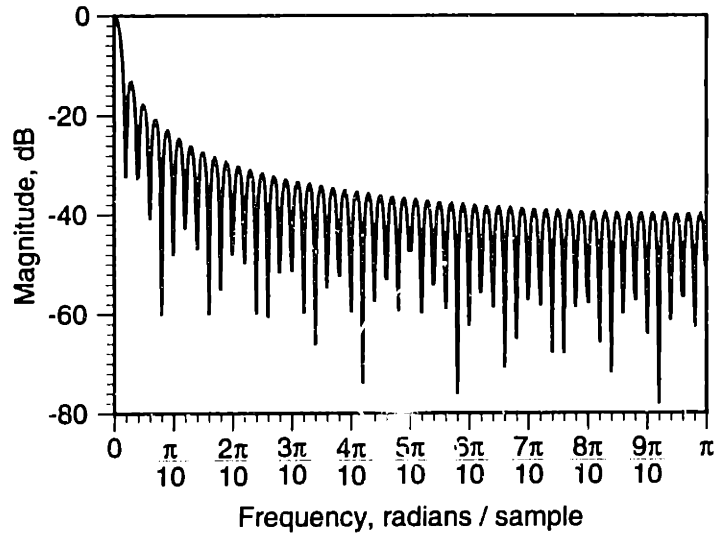


Figure 5-1: Frequency response of a moving average filter ($N = 100$).

computation time for the regression caused me to abandon it in favor of an FIR filtering technique. If more computational capability were available, this could be a promising filtering method.

5.3 FIR Filter

The final digital filter design is a finite impulse response filter. Actually, the averaging method discussed earlier is simply one design for a FIR filter. Averaging N points at a time is the equivalent of continuously filtering the waveform with an N -point filter, and then downsampling the resulting waveform by N points. An example of the frequency response for a 100-point rectangular window is given in Figure 5-1. Note that the main lobe is $2\pi/N$ wide. The width of this main lobe is important because downsampling a waveform has the effect of spreading out its frequency spectrum [44]. Specifically, downsampling a waveform by 100 points stretches the frequency content by a factor of 100. Therefore, the components that used to appear with a frequency of $2\pi/100$ will now be at 2π . The biggest problem is that the repeated copies of the frequency response which are centered at multiples of 2π will now alias into the region between π and $-\pi$. In order to eliminate aliasing, we would like to implement an ideal low pass filter with a cutoff of $2\pi/200$ so that when we later downsample by 100, there is still no aliasing.

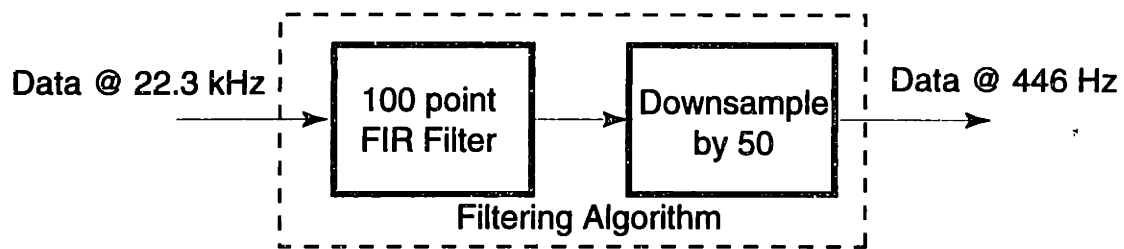


Figure 5-2: FIR filtering and downsampling (each of 6 channels).

Unfortunately, it is not possible to design a filter with such a low cutoff frequency and yet still have a reasonable level of group delay. Achieving lower cutoff frequencies involves calculating over longer lengths of time, and therefore the filter length has to increase. The group delay of an N point minimum phase filter is $(N - 1)/2$, and so the group delay increases linearly with the filter length. Rather than trying to design for a very low cutoff frequency, we instead will downsample by fewer points and allow for a wider main lobe in the frequency response.

The available processing speed also has an impact on what algorithms we can use. Within the original limits imposed by the computation time, using a 100 point FIR filter, updated every 50 points would have required slowing down the sampling rate, and thereby adding to the problem of inadequate position resolution. To get around this problem, I re-wrote the FIR filtering portion of the controls code in 'C30 assembly language. This increased the filter implementation speed by better than a factor of five. With this change I am able to implement a 100 point filter, but only downsample by 50. Each time the A/D converter samples 50 new points per channel, the code adds these to the end of a string of 50 previously existing points. After applying a 100 point FIR filter to the entire string, the controller gets new position data to work with and correspondingly computes a new D/A command. This method increases the effective controller update rate from 223 Hz to about 446 Hz. Just as importantly though, it lessens the problem with aliasing as described below.

The impulse response of the redesigned FIR position filter is generated using the MATLAB signal processing toolbox. This toolbox contains an M-file for calculating the coefficients of a Parks-McClellan filter using the Remez algorithm. The actual numerical coefficients are given in Appendix C, but the form of the impulse response is shown in

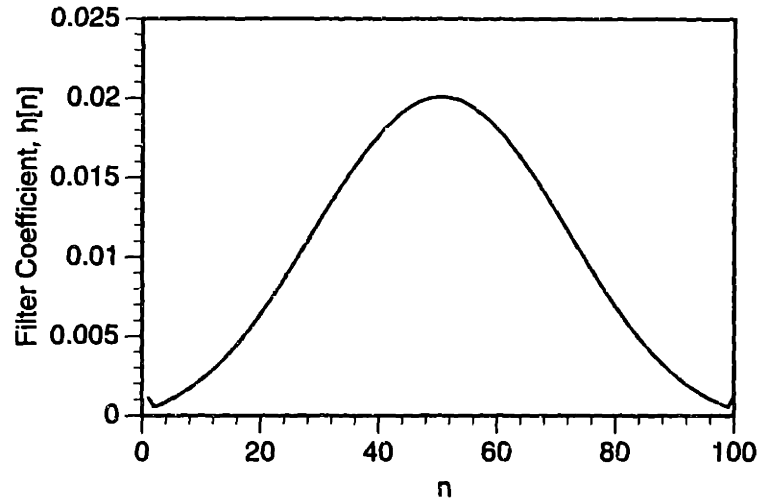


Figure 5-3: Impulse response of the input FIR filter ($N = 100$).

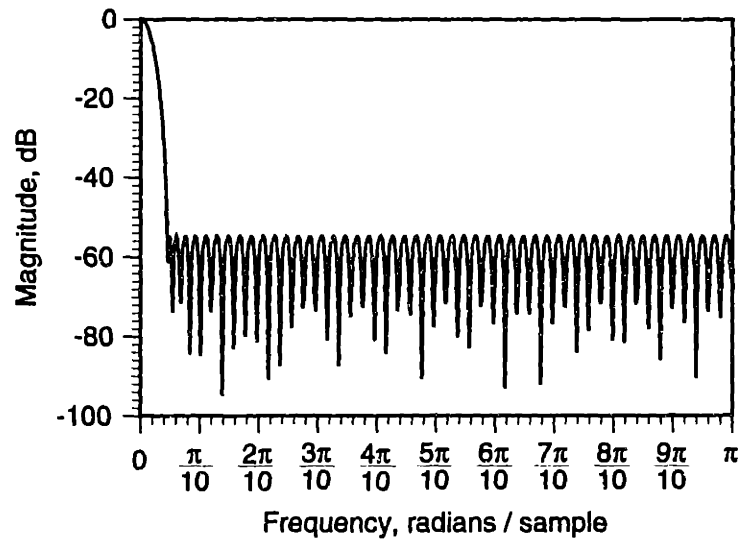


Figure 5-4: Frequency response of the redesigned filter ($N = 100$).

Figure 5-3. In addition, Figure 5-4 shows the magnitude of the frequency response. The most important characteristic to note is that the sidelobes are at much lower magnitudes, and therefore present a lesser problem with aliasing under downsampling when compared to the simple averaging filter.

To this point, we have discussed the filtering in general terms. We can also see the effects of the filter in a concrete example. Figure 5-5 contains 500 points taken from the center of a longer 5000 point waveform of a grounded A/D converter channel. These points were sampled at about 22.3 kHz, which is the normal operating rate of the controller. This trace represents the best performance that can be expected from the con-

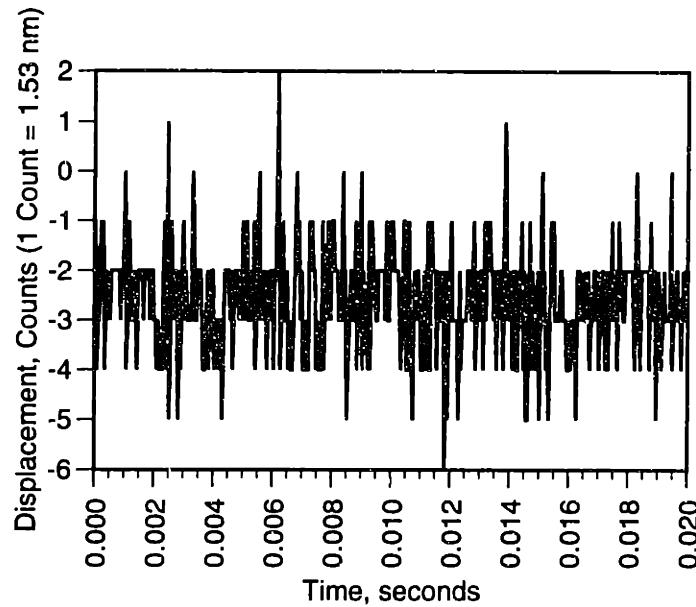


Figure 5-5: Unfiltered readings from a grounded A/D converter channel.

verter, and clearly is not immediately capable of revealing position to the sub-nanometer level. The next plot (Figure 5-6) shows the waveform after it has been convolved with a 100 point rectangular window. Note that while the noise level has been significantly reduced, a high frequency component remains. In practice, this waveform would then be downsampled by 50 before being sent to the controller.

The effect of using the Parks-McClellan filter (the impulse response of Figure 5-3) is shown in Figure 5-7. In this case the high frequency component of the noise has been largely removed, and only the lower frequency components remain. This is the filter used with the stage in practice. As a final check of the effectiveness of the new filter, we can look at the frequency content of the original and filtered waveforms. Figure 5-8 contains 256-pt DFT's of each of the 4096-pt waveforms. As expected, the redesigned filter is better at removing the high frequency components of the noise than was the simple averaging method.

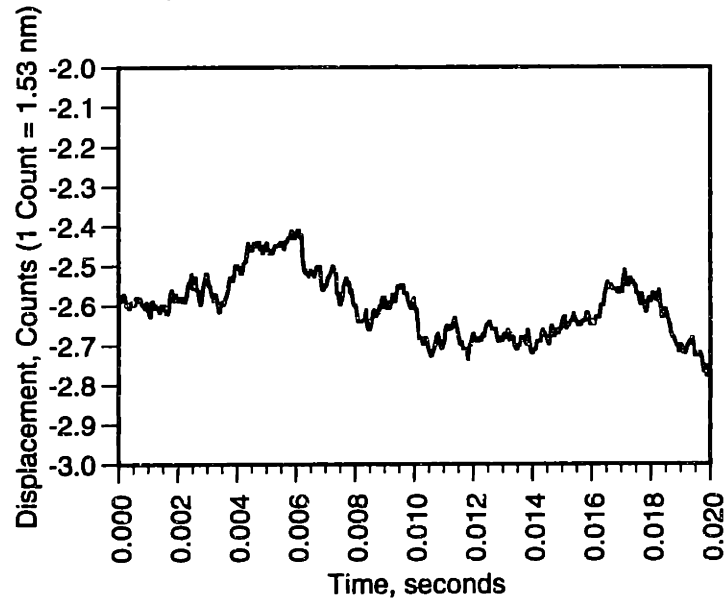


Figure 5-6: A/D converter readings filtered with a 100 point moving average.

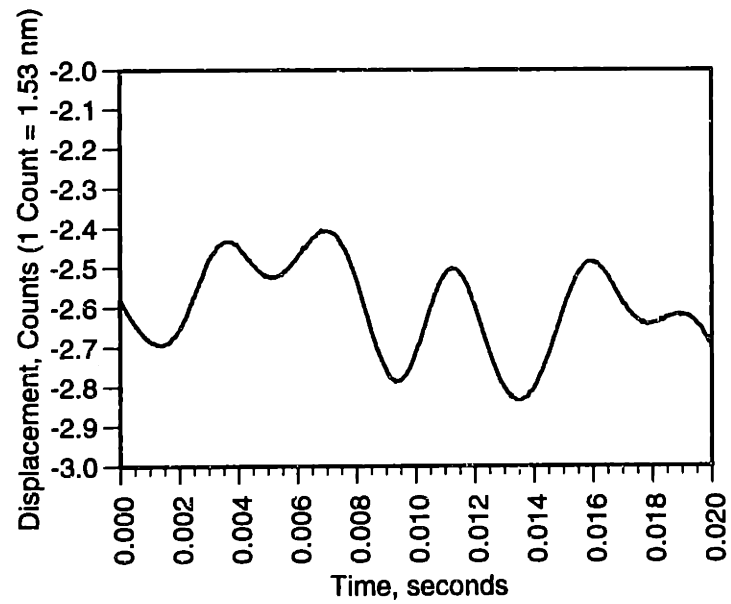


Figure 5-7: A/D converter readings filtered with the redesigned FIR filter.

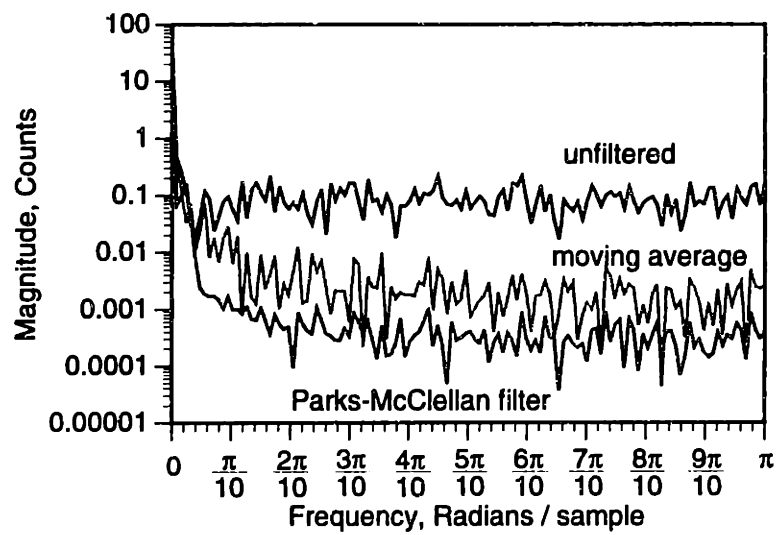


Figure 5-8: 256-pt DFT's of unfiltered, averaged, and filtered A/D noise waveforms.

Chapter 6

Control System Design

The main goal chosen for the control system design is to obtain high resolution even at the expense of bandwidth. Along with that priority comes a goal to make the response position independent, and to eliminate coupling between the axes of motion. The simplest option is to use a simple linear (PI) control system. Pursuing this approach shows some advantages of the damper-dominated design, but fails to yield an acceptable controller because of the position dependence of the response due to actuator and damper nonlinearities. A feedback linearization technique, applied to both the actuators and dampers, yields an effectively linear plant to which standard linear feedback can be applied. The controller also makes use of an estimator to reduce the noise level in the position readings. Finally, closed loop velocity control is considered as a means to improve the stage's trajectory following capabilities. Each of these methods will now be described in greater detail.

6.1 Open Loop Plant

Chapter 3 contains the derivation of a detailed nonlinear, six degree of freedom plant model. Although useful and necessary for the final control system design, it is of little value in deciding upon the initial design. Instead, a simplified linear model works best for the first cut at the control system. In this case, the model for the stage consists of a

mass and a damper. The transfer function between force and displacement,

$$\frac{X(s)}{F(s)} = \frac{1}{s(ms + b)}, \quad (6.1)$$

contains two poles. The pole at the origin represents the damping, and the pole at $s = b/m$ shows the influence of the platen's mass. If only a single damper were acting on the platen at its maximum gap (minimum damping), the damping value would be 18,500 Ns/m. This is a more lightly damped case than the platen would ever experience since multiple dampers act on the platen in every degree of freedom. Since the platen's mass is 3.09 kg, the inertial pole is located at $s = -5790 \text{ sec}^{-1}$, yielding a largest possible time constant of $\tau = 0.167 \text{ ms}$. Typical values are ten times faster, and this simple linear model is certainly not accurate on the microsecond time scale. Sensor and actuator bandwidth as well as discrete time effects would have to be included in a model at the microsecond time scale. To put this time constant into perspective, if a step in force is applied to the platen, the platen reaches a constant velocity in well under a millisecond. This response time is significantly faster than the dynamics of interest in the stage, and thus the effect of mass is ignorable. That is, viscous forces dominate for all frequencies of interest. A satisfactory model is thus simply a first-order damper driven by the actuator forces.

6.2 Continuous Time PI Control

The next step in the control system design is to investigate closing the loop with a linear PI controller design. Figure 6-1 contains a block diagram of this controller scheme. Notice that at this point we are not concerned with the nonlinearity of the actuators, and are just assuming that we can in some way command that a given force be applied to the platen. The closed loop transfer function between commanded and actual position now becomes

$$\frac{X(s)}{X_d(s)} = \frac{\frac{K_p}{b} \left(s + \frac{1}{T_i} \right)}{s^2 + \frac{K_p}{b}s + \frac{K_p}{bT_i}}, \quad (6.2)$$

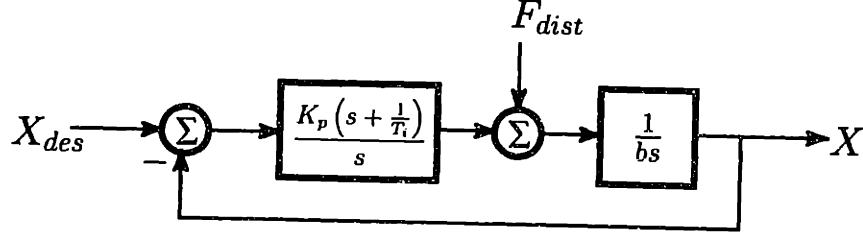


Figure 6-1: Block diagram of a linear PI control scheme.

and we only need to determine values of K_p and T_i that will give the desired dynamic response. One thing interesting to note from this relationship is that the damping (b) and the proportional gain (K_p) always appear as a pair. In a loop transmission sense, raising the proportional gain has the identical response to lowering the damping, and vice versa. Section 6.3.2 makes explicit use of this fact in the feedback linearization of the squeeze film dampers.

6.2.1 Selection of Gains

We can select values for the proportional and integral gains based upon the desired time response of the closed loop plant. Figure 6-2 shows a standard root locus plot of the system, and the effect of changing the gains. Suppose that we are looking for a settling time of about 200 ms, which would give a closed loop bandwidth close to 5 Hz. Assuming that this is the standard second order plant,

$$G(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}, \quad (6.3)$$

(neglecting the the influence of the zero for the time being), we know to set

$$\frac{4.6}{\zeta\omega_n} = 0.2 \text{ sec.} \quad (6.4)$$

The proportional gain is now found from this relationship through

$$K_p = b(2\zeta\omega_n). \quad (6.5)$$

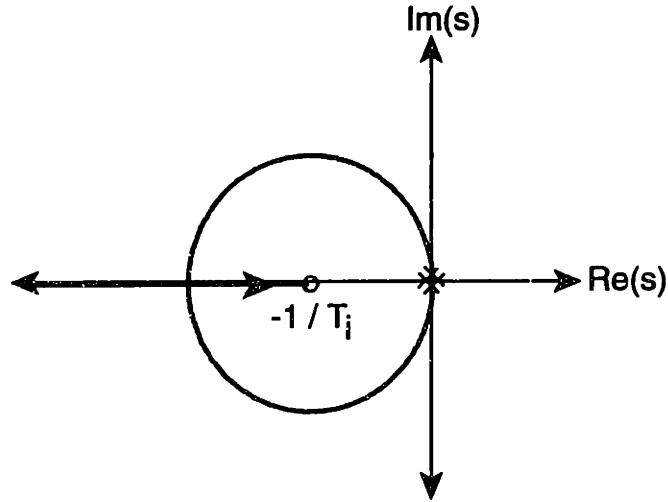


Figure 6-2: Continuous time root locus of plant under PI control.

In the actual plant, the translational damping in both the X and Y directions is about 200,000 Ns/m when the platen is at its central position. Using this value in order to numerically evaluate the different control designs gives a proportional gain of $K_p = 9.2 \times 10^6$ N/m. This sounds like a ridiculously high value for the gain. However, recall that the errors will typically be on the order of nanometers, and this will reduce the applied force to fractions of a Newton, which becomes much more realistic.

The value of the integral gain comes from setting the desired damping ratio of the closed loop plant. Knowing that the zero in the origin will increase the level of overshoot, we will set this to $\zeta = 1.2$. This gives $\omega_n = 19.2$ rad/s, and a value of $T_i = 8$.

6.2.2 Control Results

The first attempt at the control system design is now finished, with a closed loop position transfer function of

$$\frac{X(s)}{X_d(s)} = \frac{46(s + 8)}{(s + 10.3)(s + 35.7)}. \quad (6.6)$$

The step response to a commanded change in position for this plant is shown in Figure 6-3. As expected, the zero located at -8 sec^{-1} has added some amount of overshoot. Bode plots for the closed loop plant are shown in Figure 6-4. The zero at $s = -8$ causes a small resonant peak in the area of 1 to 2 Hz, but its effect is quickly canceled by the pole at $s = -10.3$. From there, the effect of the pole at $s = -35.7$ dominates, and the

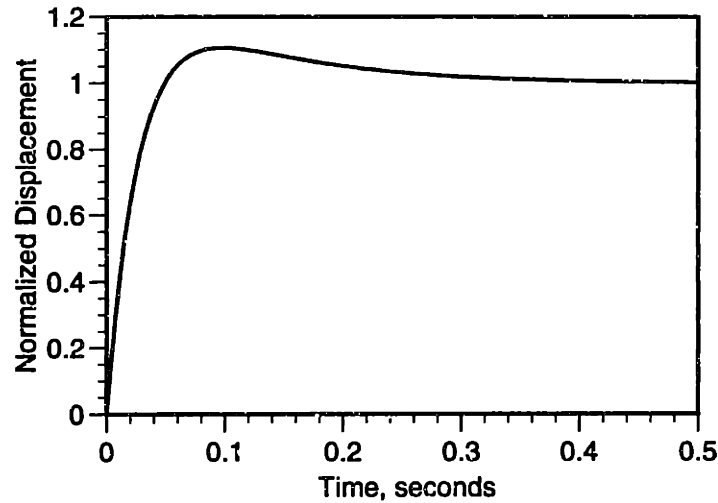


Figure 6-3: Simulated position step response with the PI controller.

response essentially looks like that of a first order system.

The situation becomes a little more interesting in the case of the system rejection of disturbance forces. The block diagram of Figure 6-1 gives the transfer function from disturbance force to position error,

$$\frac{X(s)}{F_d(s)} = \frac{\frac{1}{b}s}{s^2 + \frac{K_p}{b}s + \frac{K_p}{T_i b}}. \quad (6.7)$$

Figure 6-5 shows the Bode plots of this system. At low frequencies, the controller can provide the disturbance rejection. At high frequencies, however, the disturbance rejection is provided by the plant itself. As the control system bandwidth rises, the magnitude of the transition region between the two states falls. One design goal will then be to raise the controller bandwidth sufficiently high that any disturbance forces which are transmitted through to position errors will be sufficiently attenuated. Unfortunately though, as was described in more detail in Section 4.3.1, increasing the gain also increases the transmission of sensor noise. The response to sensor noise is the same as it is for setpoint, therefore higher bandwidths lead to a greater transmission of sensor noise and the optimal gain setting again becomes a compromise.

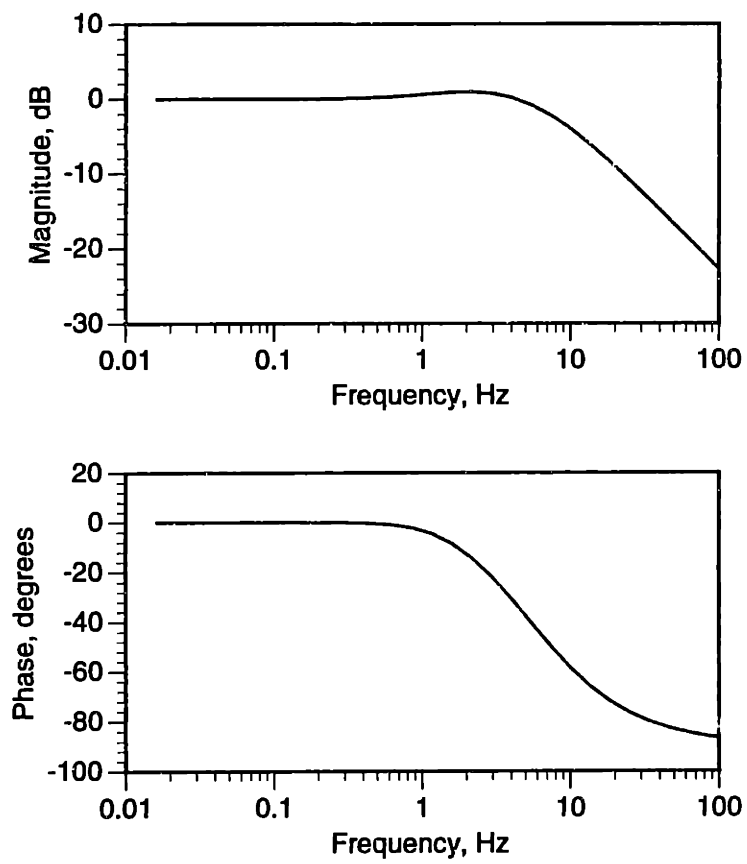


Figure 6-4: Frequency response of the plant under PI control.

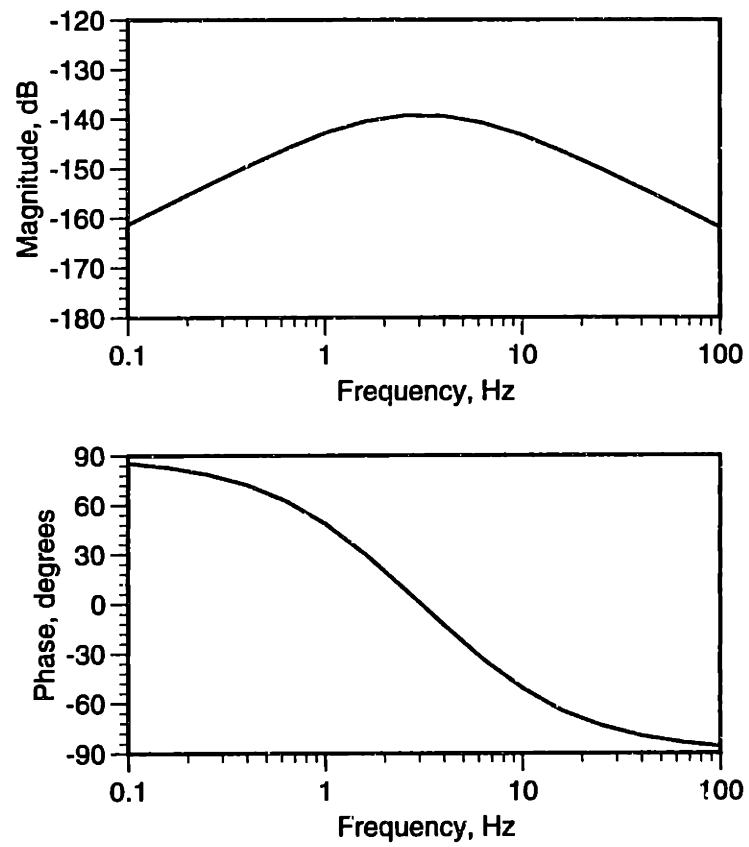


Figure 6-5: Disturbance force transmission of the PI controlled system.

6.2.3 Effect of the Changes in Damping

The linear PI control system above was designed under the assumption that the damping does not appreciably change over the travel of the stage. However, as was shown in Section 3.3.4, the damping can change by as much as an order of magnitude as the position changes. This changing damping will also change the closed loop pole locations. Rewriting the closed loop plant transfer function as

$$\frac{X(s)}{X_d(s)} = \frac{\frac{K_p \left(s + \frac{1}{T_i} \right)}{bs^2}}{1 + \frac{1}{b} \left(\frac{K_p \left(s + \frac{1}{T_i} \right)}{s^2} \right)} \quad (6.8)$$

shows that the effects of changes in damping (or more specifically, inverse damping) can be examined with the same methods used to understand changes in proportional gain.

A particularly useful tool for examining the effects of a changing parameter on closed loop pole locations is a variation on the root locus plot referred to as a *gain plot* [33] [13]. This method was designed by Dr. Tom Kurfess and Dr. Mark Nagurka, and explicitly shows the closed loop eigenvalue magnitudes and angles as a function of gain. Figure 6-6 shows how the pole locations change in our single degree of freedom system as the damping increases and decreases one order of magnitude.

Some amount of explanation is in order if one has never seen this type of a plot before. Starting at a value of $b/b_o = 1$, the magnitude plot shows two poles, and the angle plot shows them both at 180° . Therefore, at the nominal damping value there are two poles on the negative real axis. These are the overdamped poles that the controller was designed to produce. As the damping increases, the magnitudes come together, and the angles diverge. Recall that increasing the damping has the same effect as decreasing the proportional gain, therefore this situation represents the case where the poles are coming back together on the real axis, and are then leaving it as a complex conjugate pair. Their magnitudes are the same, but as the damping increases the angles are approaching values of 90° and 270° . Going in the other direction (decreasing damping), the angles remain at 180° (poles on the real axis). However, one of the poles is approaching a finite

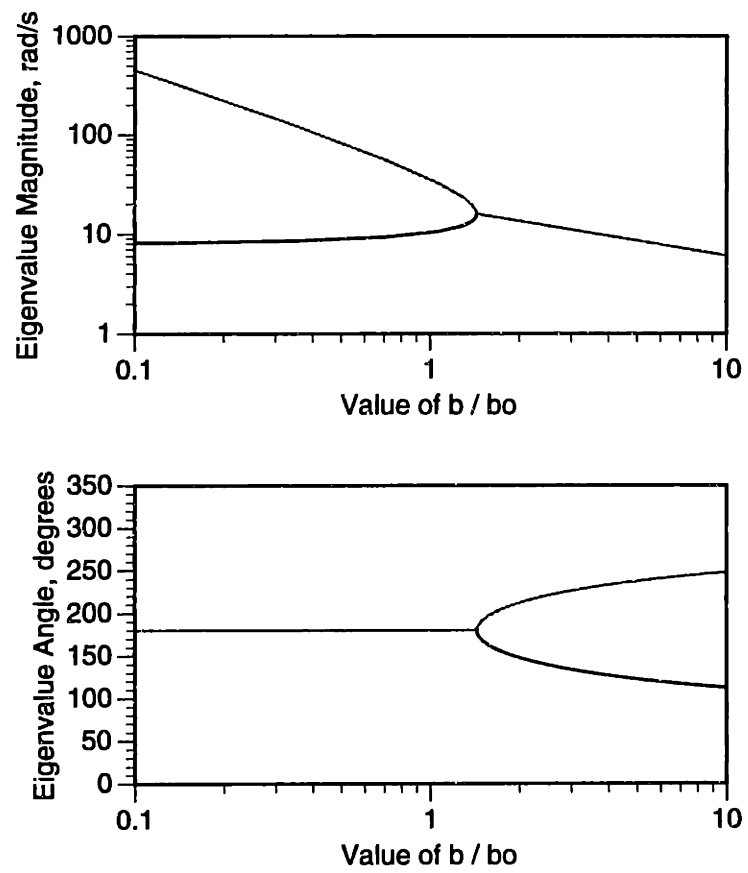


Figure 6-6: Gain plots of the pole locations as the damping changes.

limit at a zero while the other is continuing off to infinity. A look back at the root locus plot of Figure 6-2 may help to show why increasing damping in this case leads to the counter-intuitive result of an underdamped system, and will also clarify the relationship between gain plots and the root locus plot. From these diagrams, we can see that changing damping over the travel of the stage will significantly change the pole locations of the closed loop system, and therefore the dynamic response. Since this is definitely an undesirable characteristic, we will next examine *feedback linearization*, a method by which we can make the stage's operating characteristics essentially independent of position.

6.3 Feedback Linearization

Feedback linearization is a technique by which nonlinear behavior of a plant is compensated for through an inner loop of the control algorithm. The overall effect is that the linearization block operates transparently, and the outer loop can be designed as though the plant truly were linear. Linearizing the electromagnetic actuators with this technique is a standard procedure in our group's design process [53] [43] [7] [58]. In my work, I expand the technique to include reducing the nonlinear damping of the plant itself to six decoupled integrators with transfer functions of $1/s$ each.

6.3.1 Electromagnet Linearization

A basic model of the electromagnetic actuators states that the force applied by the actuator is proportional to the current squared divided by the gap squared. This model works well up until the core of the magnetic material begins to saturate, but its nonlinearity makes it difficult to apply classical linear control theory. A Taylor series linearization can be applied in cases where the electromagnets are being used to hold an object about a constant setpoint, but this loses effectiveness when the setpoint changes significantly [43]. Alternatively, an inverse model of the actuator can be used in a feedback linearization scheme to eliminate the dependence on the setpoint.

The electromagnet linearization portion of the controller takes a desired force and an instantaneous gap as its inputs, and returns a DAC voltage for its output. As was

detailed in Section 3.4, the force output of an electromagnet is

$$F = C \frac{i^2}{g_a^2}. \quad (6.9)$$

Therefore, the inverse model containing the desired force F_d and estimates of the gap \hat{g}_a and force constant \hat{C} becomes

$$i_d = \hat{g}_a \sqrt{\frac{F_d}{\hat{C}}}. \quad (6.10)$$

This calculation is performed by the 'C30 for each of the twelve actuators to return a desired current i_d for each. Next, this desired current is multiplied by the current-to-voltage factor of the power amplifiers to yield a desired DAC voltage. This scaling factor is nominally 29 A/V, but Section 2.3.2 contains the detailed values. Finally, the voltage is limited to the 10 V maximum output of the DAC, and scaled to an integer value between 0 and 4095. To recap, this integer is a number that when written to the D/A converter will yield a physical voltage which when applied to the current amplifier produces the desired current. If the estimates of the gap and force constant used in the electromagnet model are correct, this current creates the desired electromagnet force. The operation of the electromagnet linearization is transparent to the outer loop of the controller, and so the controller can merely specify a desired force and the linearization block will output the voltage that produces it.

6.3.2 Damper Linearization

The squeeze film dampers that dominate the stage dynamics can also be accommodated in the controller through a feedback linearization scheme. In this case, the problems are a damping coefficient that changes according to the inverse cube of the gap as well as coupling between the different axes. Rather than linearizing a single squeeze film damper independently, as was the case for the electromagnetic actuators, the action of all of the dampers are linearized at once.

The damper linearization starts with the instantaneous model of the squeeze film damping that was detailed in Section 3.3.4. In the true plant, the applied modal forces

and velocities are related through

$$\mathbf{F} = \mathbf{B}\mathbf{V}. \quad (6.11)$$

The 6×1 matrix \mathbf{F} contains the applied forces in the X, Y, and Z directions as well as the torques about each axis. Similarly, the 6×1 matrix \mathbf{V} contains the velocities in each of those directions. The 6×6 damping matrix \mathbf{B} is the gap dependent relationship between them. Inverting the dynamics yields

$$\mathbf{V} = \mathbf{B}^{-1}\mathbf{F}, \quad (6.12)$$

and this inversion shows the velocities that can be expected from an application of a set of forces. If the input to this linearization scheme is taken as a desired velocity then the output is the vector of forces that will produce it. Using a model of the plant $\hat{\mathbf{B}}$, the required forces are

$$\mathbf{F} = \hat{\mathbf{B}}\mathbf{V}_d. \quad (6.13)$$

Inserting this model into the plant dynamics of eq.(6.12) yields

$$\mathbf{V} = \mathbf{B}^{-1}\hat{\mathbf{B}}\mathbf{V}_d. \quad (6.14)$$

Now, if the model of the plant is accurate, the dynamics will cancel ($\mathbf{B}^{-1}\hat{\mathbf{B}} = \mathbf{I}$) and the actual velocity of the stage will be equal to the desired velocity. Note the implicit use of the electromagnet linearization. Forces are used as plant inputs in this design even though the true control outputs are integers written to DAC registers commanding current.

6.3.3 Combined Linearized Plant

The use of both the electromagnet and damper linearization schemes greatly simplifies the control design for the outer loop. The block diagram for the closed loop system under proportional control is shown in Figure 6-7. This reduces to Figure 6-8 if the linearization is successful. The effective plant has been reduced to six decoupled integrators with

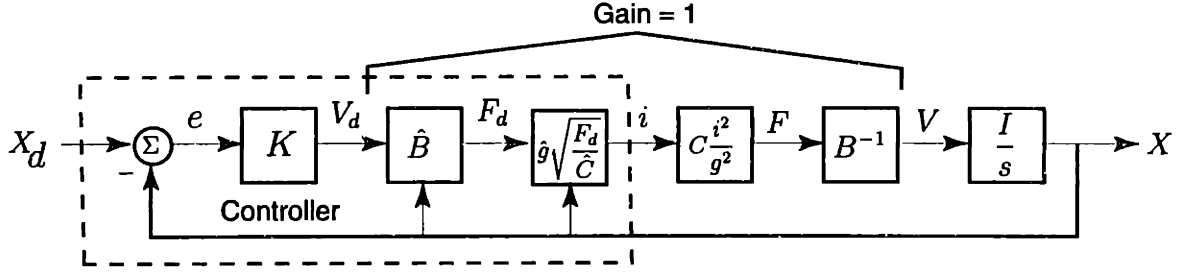


Figure 6-7: Block diagram of the feedback linearization scheme.

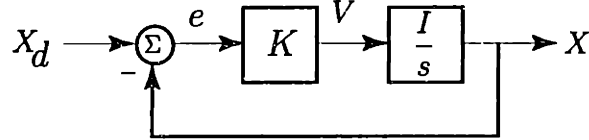


Figure 6-8: Simplified block diagram after feedback linearization.

transfer functions of $1/s$ for each (the output of the controller thus has the somewhat unusual units of velocity).

If high performance is not required from the stage then the control system can again be designed using continuous-time methods. As a result of the feedback linearization, each degree-of-freedom looks like a decoupled plant with a transfer function of $1/s$. Under proportional control, the closed loop transfer function of each degree of freedom is

$$\frac{X(s)}{X_d(s)} = \frac{K}{s + K}, \quad (6.15)$$

and the designer simply chooses the gain to set the desired time constant. For example, a gain of 5 will yield a time constant of 0.2 s and therefore an approximate settling time of 1 second. Figure 6-9 compares the predicted and measured responses with this gain for a commanded 10 nm step in the Y-direction.

6.4 Velocity Control

An additional problem posed for the Angstrom Stage is the generation of constant velocity scans. Originally, this was achieved with the open loop model of the plant used for the feedback linearization. With it, the controller commanded a velocity and relied on the model to determine the DAC output that would produce it. Because of the detail of the

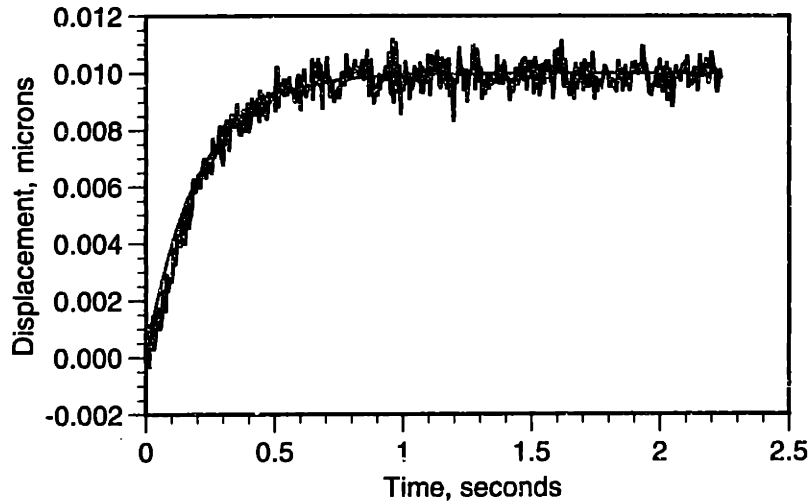


Figure 6-9: Predicted and measured step responses with a proportional gain of 5 s^{-1} .

model, this method worked quite well (as will be described in Section 7.3), but we then decided to close the velocity loop with a feedback control system in order to see how much better the controller could maintain a constant velocity.

6.4.1 Measuring Velocity

The main requirement for closed loop velocity control is a measure of the plant velocity. This is not so simple, since the stage does not have a sensor that measures velocity directly. Instead, the software must differentiate the position readings, and I tried several different methods of generating the velocity. First, I planned to use another 100 point FIR filter, similar to that already in use on the position readings. Following an example in Oppenheim and Schaeffer [44], one can simply window the ideal differentiating filter. Even better is to multiply the frequency response of the ideal differentiating filter with the ideal low pass filter in order to prevent the amplification of high frequency noise. With a single 100 point filter, this method worked wonderfully in MATLAB simulations. However, it generated a very noisy velocity signal when implemented on the stage. The reason for this is the quantization of the capacitance probe readings.

The smallest velocity that we can recognize during a single control period occurs in the case where a single sample is one count higher than all of the others. This represents a change of about 1.53 nm over a time of 4.48 ms, or a velocity of $0.341 \mu\text{m/s}$. Since a

comfortable speed for the stage is a couple of microns per second, these velocity increments are too large to provide adequate control. An additional possibility is to include the calculation of velocity in the estimator, but this option is not pursued in this work.

A single block of one hundred capacitance probe readings is thus not long enough to generate a clean velocity signal. Therefore, it is necessary to look over longer time periods. Rather than storing many hundreds of past capacitance probe readings, it is easier to instead store past filtered position values. Each of these values effectively averages about 50 new probe readings, and represents a significant reduction in the amount of data required to be saved from one sample period to the next. A backward difference equation generates the velocity from these position readings [24],

$$V(x_0) = \frac{1}{h} \sum_{k=1}^{\infty} \frac{1}{k} \nabla^k x_0. \quad (6.16)$$

With this notation, $\nabla x_0 = x_0 - x_{-1}$, $\nabla^2 x_0 = \nabla x_0 - \nabla x_{-1}$, and so forth. The constant h is the sampling time which equals 4.48 ms in this case. I settled on using a fifth order difference equation, which has the numerical coefficients of

$$V = \frac{1}{(60)(0.00448)} (137x_0 - 300x_{-1} + 300x_{-2} - 200x_{-3} + 75x_{-4} - 12x_{-5}). \quad (6.17)$$

The numerical coefficients for this difference equation are derived from eq.(6.16), which is given in [24] as a standard form for a backward difference equation. Each x_i in this case represents the position during one *control* period (i.e. one application of the FIR filter output). The frequency response of this filter is shown in Figure 6-10. The difference equation alone yielded better results than did the FIR filter, but was not yet suitable to control velocity based on it. It is likely that a better filter than this exists, but the possibility is not explored further.

The final step required in smoothing the velocity signal was to pass it through a second order digital Butterworth filter with a cutoff frequency of 10 Hz. This is a simple

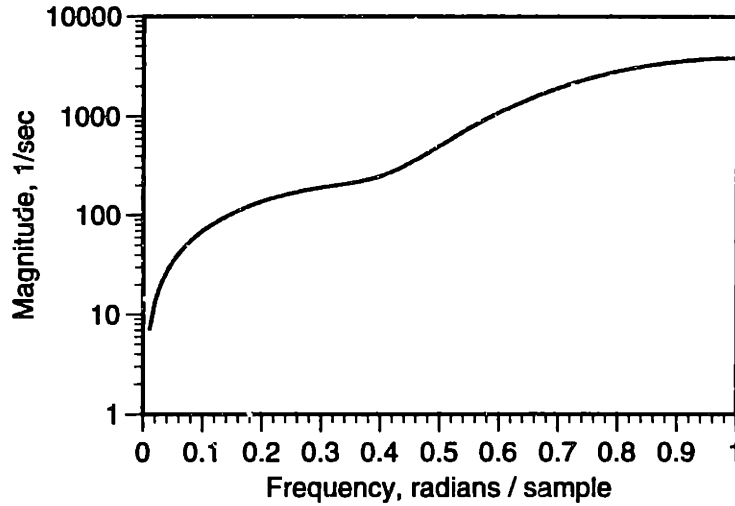


Figure 6-10: Frequency response of the velocity filter.

derivation with MATLAB, and yields a net transfer function of

$$\frac{V_{out}(z)}{V_{in}(z)} = \frac{0.00449785 + 0.00899570z^{-1} + 0.00449785z^{-2}}{1 - 1.80152637z^{-1} + 0.81951777z^{-2}}. \quad (6.18)$$

This IIR filter, when cascaded with the backward difference equation, yielded a velocity signal clean enough to feed back for closed loop control.

6.4.2 Control Algorithm

The model of the plant is not exact, and so the commanded velocity will not equal the resulting velocity from the true plant. Therefore, closing the loop on velocity in this case consists of adding a signal to the commanded velocity that is related to the previous velocity error. A block diagram of this procedure is shown in Figure 6-11. The desired velocity is V_d , the commanded velocity is V_c , and the measured velocity of the plant is V_a . The block labeled G_p is the net gain of the plant from commanded to actual velocity. Much is contained inside it, but in this model it will simply look like a gain (ideally equal to 1). The block labeled G_c contains the transfer function of the velocity controller.

We can examine the system in these general terms before deciding on a particular controller structure. Reducing the block diagram of Figure 6-11 to an algebraic transfer

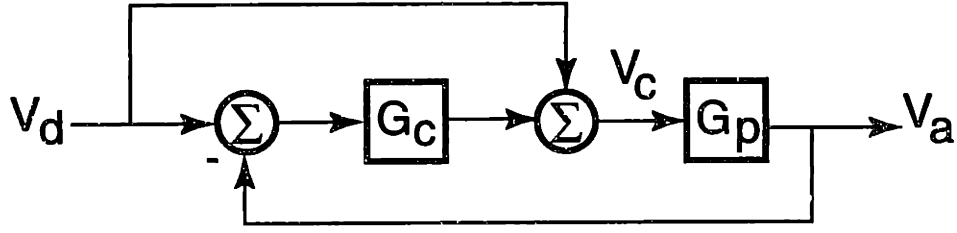


Figure 6-11: Block diagram of the velocity control structure.

function from desired to actual velocity yields

$$\frac{V_a}{V_d} = \frac{G_p + G_p G_c}{1 + G_p G_c}. \quad (6.19)$$

If the gain of the plant (G_p) equalled 1, the actual and desired transfer functions would identically equal each other. This is not the case, and so we do need to use the closed loop control. If the controller contained simply proportional gain, then we could drive the transfer function close to unity, since as G_p gets very large, the numerator and denominator approximately equal each other.

Adding integral control improves the response even more. Replacing G_c by $K_{pv}(1 + \frac{1}{T_{ivs}})$ brings the transfer function to

$$\frac{V_a}{V_d} = \frac{G_p T_{ivs} + G_p K_{pv} T_{ivs} + G_p K_{pv}}{T_{ivs} + G_p K_{pv} T_{ivs} + G_p K_{pv}}. \quad (6.20)$$

Now, by the Final Value Theorem, the actual velocity will eventually equal any desired constant velocity. This is the form of the controller that was implemented on the stage. The best results were obtained with an empirically determined velocity controller bandwidth of 1 Hz, and these are given in Section 7.3.

6.4.3 Trajectory Following

A second method that can be used to control velocity, other than an explicit feedback of a measured velocity signal, is a trajectory generation. This is actually a more general approach, since position contours other than a straight line can be programmed with similar ease. The position setpoint is incremented by some amount each control cycle in order to generate a constant velocity motion. In practice, it is possible to command very

small velocities this way. However, the explicit feedback of a measured velocity proved to be more accurate for practical stage speeds of about $2 \mu\text{m/s}$. The main source of velocity error is inaccuracy of the plant model in the feedback linearization algorithm. The trajectory following method does not modify the velocity that this model commands, while the explicit feedback of velocity does so. Using integral feedback control in combination with the position estimator would probably increase the accuracy with which the plant can track a ramp in position, but velocity control was not prioritized enough to pursue it. A further comparison of the two velocity control methods is made in Section 7.3.2 of the Results chapter.

6.5 Discrete Time Plant Model

Developing a model that accurately describes stage dynamics at higher gains requires the addition of discrete time effects. A look at the sequence of steps performed by the controller helps in clarifying some of these effects.

- The A/D converter generates an interrupt to the '486 signalling a full buffer of new capacitance probe data $\mathbf{y}(k)$.
- The '486 writes the latest control outputs to the D/A converter $\mathbf{u}(k)$.
- The '486 transfers the new capacitance probe data $\mathbf{y}(k)$ from its location in PC memory to the dual port RAM.
- The 'C30 reads the new capacitance probe data from the dual port memory, performs the necessary calculations, and returns a new set of control outputs $\mathbf{u}(k+1)$ to the dual port memory.
- The A/D converter generates another interrupt signalling a full buffer of new capacitance probe data $\mathbf{y}(k+1)$.
- The '486 writes the latest control outputs to the D/A converter $\mathbf{u}(k+1)$.

In this special case of a perfectly linear plant, the action of a single degree-of-freedom in the stage can be modeled through the difference equation

$$x(k+1) = x(k) + Tu(k). \quad (6.21)$$

Recall that because of the feedback linearization, the input to the plant $u(k)$ is simply a velocity. Multiplying it by the sample time T yields a change in position that can be directly added to the previous position in order to generate the new position. Taking the Z-transform of the above difference equation leads to

$$\frac{X(z)}{U(z)} = \frac{T}{z-1}. \quad (6.22)$$

It is important to remember that in this case, with the assumption of a perfectly linearized plant, the Z-transform is not an approximation to the continuous system, but instead represents it exactly. With the plant model, we can proceed to design and implement estimation schemes that achieve sub-nanometer positioning capabilities.

6.6 Estimators

Plants with multiple inputs and outputs are usually described by the state space formulation of their dynamic equations. In this context, the closed loop control systems are designed around the idea of full state feedback. This method assumes that measurements of all of the states are available. However, this is not generally the case. The number and cost of sensors required to achieve such feedback would be prohibitively high. Instead, an *estimator* is used to predict the missing states from the measured ones and in the process, filter the random errors from the sensor readings. Another term for an estimator is an *observer*, and the literature uses the two terms interchangeably. I have chosen to follow the conventions in Franklin, Powell, and Workman [15] most closely. This section begins with a general background on estimators, and then continues with the design of a position estimator for the Angstrom Stage. Next, we design a bias estimator which can eliminate the steady state errors that are due to constant disturbance forces. Finally, we

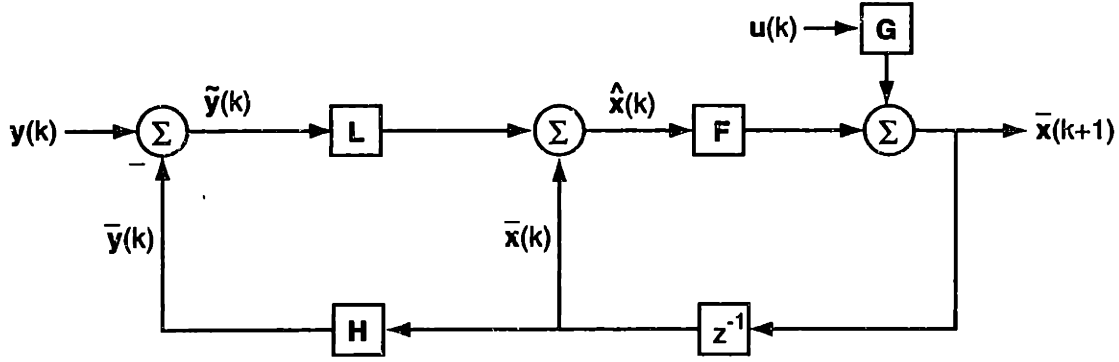


Figure 6-12: Block diagram of a basic estimator.

design an estimator which makes a reasonable tradeoff between sensor noise and plant noise, drawing on the theory of the Kalman filter.

6.6.1 Background

Central to an estimator is a model of the plant that creates predicted measurements which are compared against the actual sensor outputs. To begin, suppose that the open loop plant is described in state space form as

$$\mathbf{x}(k+1) = \mathbf{F}\mathbf{x}(k) + \mathbf{G}\mathbf{u}(k). \quad (6.23)$$

The sensor readings are then defined as

$$\mathbf{y}(k) = \mathbf{H}\mathbf{x}(k). \quad (6.24)$$

There can also be sensor noise associated with the measurements, but this is not important in the preliminary development of an estimator. Both plant and sensor noise will be handled as the estimator design increases in complexity later in this chapter. Utilizing the model above we can continue with the estimator derivation. Following the block diagram in Figure 6-12 should be helpful.

In Figure 6-12, a previous estimate of the state $\bar{\mathbf{x}}(k)$ is multiplied by the measurement sensitivity matrix \mathbf{H} to generate a prediction of the measurement vector $\bar{\mathbf{y}}(k)$. This prediction is subtracted from the actual measurements $\mathbf{y}(k)$ in order to produce a mea-

surement error vector $\tilde{\mathbf{y}}(k)$. The error vector is multiplied by the estimator gain matrix \mathbf{L} , and the result is added to the prior estimate of the state $\bar{\mathbf{x}}(k)$. In general terms, the estimator gain matrix \mathbf{L} determines the extent to which the estimator “believes” the actual measurements relative to the model based predictions. This computation yields an estimate of the state $\hat{\mathbf{x}}(k)$ at the time the measurements were taken,

$$\hat{\mathbf{x}}(k) = \bar{\mathbf{x}}(k) + \mathbf{L} [\mathbf{y}(k) - \mathbf{H}\bar{\mathbf{x}}(k)]. \quad (6.25)$$

The next step involves updating the state estimate for the next set of measurements. Here the model of the plant comes into use. The known controller outputs $\mathbf{u}(k)$ and the prior estimate are combined to produce the new state estimate,

$$\bar{\mathbf{x}}(k+1) = \mathbf{F}\hat{\mathbf{x}}(k) + \mathbf{G}\mathbf{u}(k). \quad (6.26)$$

The estimator produces two outputs that we can use as the states for the controller. The first is the *current* estimate $\hat{\mathbf{x}}(k)$ which represents the state at the time the measurements were taken. If the computation time is small in respect to the sample period, this would be the suggested estimate to use. However, as the computation time increases the *predictor* estimate $\bar{\mathbf{x}}(k+1)$ may be more accurate since it represents the expected state at next sampling interval.

In this structure, the only estimator design choice is the estimator gain \mathbf{L} . This matrix determines the rate at which the estimate of the state converges to the true state. In other words, it determines the rate at which the error in the estimate converges to zero.

6.6.2 Position Estimator

The title “position estimator” may be somewhat confusing since the stage position does not have to be estimated, but is instead read directly from the capacitance probes. In this case though, we are interested in the smoothing action of the estimator to reduce the random errors present in the measurements. The previous section describes the general form of an estimator, and now the plant model takes the simple form of eq.(6.22) via

the feedback linearization. The state $\mathbf{x}(k)$ consists only of the position and orientation of the platen (relative to the frame) in cartesian coordinates. The input $\mathbf{u}(k)$ is velocity, and thus multiplying by the sample period T produces the position change during that period,

$$\mathbf{x}(k+1) = \mathbf{x}(k) + T\mathbf{I}\mathbf{u}(k). \quad (6.27)$$

Using the notation of the previous section, $\mathbf{F} = \mathbf{I}$, and $\mathbf{G} = T\mathbf{I}$. The main complication, and reason for using an estimator, is that the capacitance probe readings are corrupted by the A/D converter noise $\mathbf{v}(k)$, and so the measurements become

$$\mathbf{y}(k) = \mathbf{H}\mathbf{x}(k) + \mathbf{v}(k). \quad (6.28)$$

The estimator calculations proceed in two steps in the controller. In the first step, the measurements and the prior state estimate combine to form the *current* state estimate,

$$\hat{\mathbf{x}}(k) = \bar{\mathbf{x}}(k) + \mathbf{L}[\mathbf{y}(k) - \mathbf{H}\bar{\mathbf{x}}(k)]. \quad (6.29)$$

Using these, the predictor estimates are calculated from the plant model according to

$$\bar{\mathbf{x}}(k+1) = \hat{\mathbf{x}}(k) + T\mathbf{I}\mathbf{u}(k). \quad (6.30)$$

Because of the time delay associated with the extensive calculations taking place in the control algorithms, these predictor estimates provide a more accurate assessment of the plant state at the time that control is applied to the plant.

The only design variable is the estimator gain matrix \mathbf{L} which is chosen to make the error in the state estimate converge to zero. Subtracting eq.(6.30) from eq.(6.27), and substituting in the equation for the current estimate, eq.(6.29), leads to the dynamic equation for the error,

$$\tilde{\mathbf{x}}(k+1) = [\mathbf{I} - \mathbf{L}\mathbf{H}]\tilde{\mathbf{x}}(k). \quad (6.31)$$

The eigenvalues of $[\mathbf{I} - \mathbf{L}\mathbf{H}]$ determine the dynamics of the state convergence, and in theory \mathbf{L} could be chosen to make the estimate converge arbitrarily fast. In practice

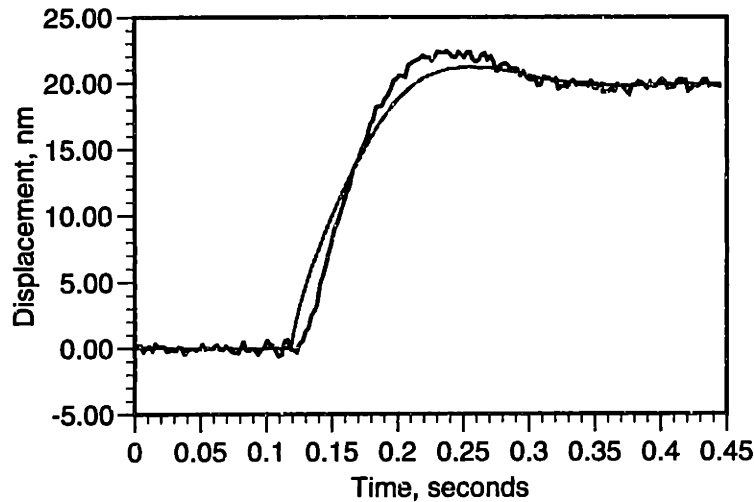


Figure 6-13: Example of estimator convergence.

though, modeling errors and unmodeled plant disturbances require some finite settling time. One rule of thumb is to select well damped estimator roots that converge two to six times faster than the control roots. In this way, the dynamic response of the plant is dominated by the control roots [15]. However, in the presence of significant sensor noise, we may choose slower estimator convergence in order to ignore the high frequency sensor noise. The Kalman filter design yields the optimal estimate, and this will be discussed in a later section.

As a design example, suppose that we want a 5 Hz bandwidth in the closed loop plant. The estimator bandwidth should then be 10 Hz, which places the poles at $-20\pi \text{ sec}^{-1}$ in the s -plane. Using $z = e^{sT}$ for the z -plane mapping, the poles move to $z = 0.8687$ rad/sample. After this, it is a matter of using Ackerman's formula ¹ to select the estimator gain matrix that will yield these roots. The plot of Figure 6-13 shows the experimental data for the estimate of the Y -position of the stage converging to the true position for a 20 nm step.

6.6.3 Bias Estimator

An estimator by itself cannot eliminate the problem of steady state error. Typically, this would still require the addition of integral control and therefore the introduction of another set of states \mathbf{x}_I to the state space formulation of the plant. However, in the case

¹This method is performed with the MATLAB routine PLACE.M.

of the Angstrom Stage, we know that most of the steady state position error is caused by constant disturbance forces. More specifically, the density matching of the platen with the oil is not exact and so gravity applies a constant downward force to the platen. Some of this weight then shows up as a force in the X and Y directions because the air table will never be perfectly level. Even the mass of the sample on the sample holder will add a disturbance force. These forces are genuine physical quantities, and so can be predicted through the use of a *bias estimator* [15]. Once the bias forces are known, the controller can add an additional force component to the control output that will cancel them and thus eliminate the steady state errors they cause.

This sounds like an overly complex way to eliminate steady state errors, vis-a-vis integral control. In some instances however, bias estimation has an advantage over integral control. In a typical design, a step change in the reference position will result in a position error which is continually summed by the integrator. The plant then has to overshoot the reference position in order to return the integrated error to zero. This is the reason for the additional overshoot and decreased stability of plants using integral control. Estimates of the bias forces are not affected by changes on the reference position of the plant. In theory then, a plant with bias estimation followed by proportional control will respond exactly the same as the plant under just proportional control except for the fact that steady state errors due the bias forces are eliminated.

In the end, both bias estimation and integral control will eliminate the steady state errors caused by disturbance forces. Integral control is even more general and will eliminate steady state errors regardless of their source. However, I chose to pursue bias estimation for other reasons. First, steady state errors represent more of a nagging irritation than any significant limiting flaw in the present application of the stage. Therefore totally eliminating them is not an overriding concern. Second, we know that most of the steady state error is caused by bias forces, and it is generally good practice to incorporate as much model knowledge into the controller as is feasible. Third, a properly functioning bias estimator will operate transparently to the main control loop. Much of the control system design (the feedback linearization algorithm specifically) has been devoted to keeping this main loop simple, and the bias estimator continues this trend. Thinking

back to the continuous-time, single degree of freedom case, the plant with bias estimation can be represented by a pole at the origin. To the limits of the linear model, no amount of gain can force the closed loop pole to leave the negative real axis. Integral control, on the other hand, adds an additional pole at the origin as well as a zero on the negative real axis. The closed loop system is likely to be underdamped and the zero limits the ultimate bandwidth of the stage. Finally, bias estimation is a control strategy that I had not known of previously, and the stage provided an excellent system on which to experiment with it.

Adding bias estimation to the Angstrom Stage requires augmenting the plant model with a 6×1 vector $\mathbf{w}(k)$ of the bias forces,

$$\mathbf{x}(k+1) = \mathbf{x}(k) + T\mathbf{u}(k) + T\mathbf{B}^{-1}\mathbf{w}(k), \quad (6.32)$$

$$\mathbf{y}(k) = \mathbf{H}\mathbf{x}(k) + \mathbf{v}(k). \quad (6.33)$$

Since the bias forces are essentially constant values, the new state equations become

$$\begin{bmatrix} \mathbf{x}(k+1) \\ \mathbf{w}(k+1) \end{bmatrix} = \begin{bmatrix} \mathbf{I} & T\mathbf{B}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x}(k) \\ \mathbf{w}(k) \end{bmatrix} + \begin{bmatrix} T\mathbf{I} \\ \mathbf{0} \end{bmatrix} \mathbf{u}(k). \quad (6.34)$$

The assumption of a constant bias force allowed the simplification of setting $\mathbf{w}(k+1) = \mathbf{w}(k)$. However, this estimator arrangement can also model bias forces that behave otherwise. Disturbances such as sinusoids can be modeled by simply adding additional states. With these equations, we can redefine the state matrices as

$$\mathbf{F}_w = \begin{bmatrix} \mathbf{I} & T\mathbf{B}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \quad (6.35)$$

$$\mathbf{G}_w = \begin{bmatrix} T\mathbf{I} \\ \mathbf{0} \end{bmatrix}, \quad (6.36)$$

$$\mathbf{H}_w = \begin{bmatrix} \mathbf{H} & \mathbf{0} \end{bmatrix}. \quad (6.37)$$

and the problem returns to the previous form of the estimator in Section 6.6.1. Figure

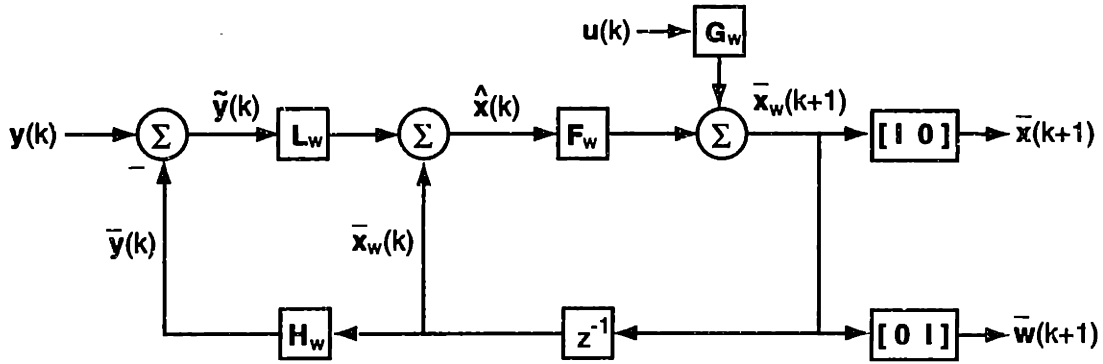


Figure 6-14: Block diagram of a bias estimator.

6-14 shows a block diagram of a bias estimator structure. Because the bias forces are not expected to change quickly (except when someone leans on the air table, places or removes a sample, or pokes at the sample holder with a pen), the bias estimator poles are set much slower than are the poles for the position estimates. In this case, I set them to converge in ten seconds. This sounds very slow, but remember that these forces are expected to be constant and so the rate at which changes in them can be tracked is not a concern. Setting the bandwidth of the bias estimator very low keeps it from affecting the main stage dynamics at the cost of a slow settling time to the *very* occasional changes in the disturbance forces.

Setting the gain matrix \mathbf{L}_w is done in a manner similar to the position estimator case. Requiring that the position estimate converge at 10 Hz and the bias force estimate converge at 0.1 Hz means placing their poles at $-20\pi \text{ sec}^{-1}$ and $-0.2\pi \text{ sec}^{-1}$ in the s-plane. Again using $z = e^{sT}$ to map these to the z-plane, the poles move to $z = 0.8687$ rad/sample and 0.9986 rad/sample respectively. Ackerman's formula can then be used to determine the gain matrix that will achieve these closed loop poles. The MATLAB code that accomplishes this is included in Appendix D.

6.7 Kalman Filter

In looking at the position readings from the capacitance probes, we are faced with the problem of separating the useful signal from the noise. One way is to design a digital filter with a cutoff frequency below the frequency of the noise component. Note that

an analog filter is of little use in this case, since most of the noise is added after the signal has entered the A/D converter. A digital filter, even one as simple as an average, limits the bandwidth of the controller and adds time delay to the system which will again limit achievable response times. The problem is that most filters do not utilize known information about the expected response from plant inputs, but instead merely remove certain frequencies from a waveform.

An estimator does use known information about the plant when filtering noise from a sensor signal. In addition to using a plant model to generate the unknown states, an estimator smooths the measured signal and lessens the disturbances caused by measurement noise. However, we do not know where exactly to set the estimator poles. If the measurements are noisy and the plant has small disturbances, slow poles can be used to force the state estimate to follow the model. In the opposite case, clean measurements of a plant with large disturbances are best approached with fast poles that can track the true motion in the plant. The Kalman filter provides a quantitative method of setting the estimator gain matrix to account for these tradeoffs. By utilizing information about the relative levels of the plant and measurement noise, it generates the *noise-optimal* estimate of the state.

Much has been written about the Kalman filter since its introduction in 1960 [31] [32]. Unfortunately though, much of this literature is concerned more with the mathematics of the broader field of stochastic control theory, and is not particularly readable. Franklin, Powell, and Workman [15] seem to move a little too far in the other direction, and provide the result with a minimum of derivation and a high reliance on MATLAB. The books that I have found most useful are by Grewal and Andrews [21] and Lewis [35]. In particular, Grewal and Andrews give an excellent background on estimation theory beginning with the work of Gauss in 1795 that helps to provide the context for the later work. However, Lewis contains the derivations that I found easiest to understand and follow most closely. I have chosen to include a derivation of the Kalman filter in this thesis mainly to de-mystify it for myself. For the most part, this derivation is merely an exercise in matrix algebra that only becomes confusing when steps are omitted in the interest of conserving space. Words are cheap in this instance, and so I have included

much of the algebraic detail.

6.7.1 Minimizing the Squared Error

The main idea behind the derivation of the Kalman filter is that it minimizes the error between the true and estimated state in a least squares sense. To begin, look at the equation for a very general estimator,

$$\hat{X}(Y) = AY + b. \quad (6.38)$$

The estimate of the state \hat{X} is a linear combination of the measurements Y and the constants A and b . Defining J as the expected value of the sum of the squared error between the true and estimated states,

$$J = \overline{(X - \hat{X})^T(X - \hat{X})}, \quad (6.39)$$

we can proceed to derive the least squared error solution. The *trace* of a matrix is the sum of the elements of its main diagonal, and so

$$J = \overline{\text{tr}(X - \hat{X})(X - \hat{X})^T}. \quad (6.40)$$

Substituting eq.(6.38) yields,

$$J = \overline{\text{tr}(X - AY - b)(X - AY - b)^T}, \quad (6.41)$$

which can then be refactored as

$$J = \overline{\text{tr}[(X - \bar{X}) - (AY + b - \bar{X})] [(X - \bar{X}) - (AY + b - \bar{X})]^T}. \quad (6.42)$$

The variable \bar{X} is a previously existing estimate of the state vector. Expanding the expression takes us to

$$J = \overline{\text{tr}(X - \bar{X})(X - \bar{X})^T - 2(AY + b - \bar{X})(X - \bar{X})^T + (AY + b - \bar{X})(AY + b - \bar{X})^T} \quad (6.43)$$

in which we recognize the state error covariance matrix,

$$P_X = \overline{(X - \bar{X})(X - \bar{X})^T}. \quad (6.44)$$

With this simplification, the expected sum of the squared errors becomes

$$J = \text{tr} \left[P_X - \overline{2(AY + b - \bar{X})(X - \bar{X})^T} + \overline{(AY + b - \bar{X})(AY + b - \bar{X})^T} \right]. \quad (6.45)$$

Next, we can expand the last term in the expression. This yields

$$J = \text{tr} \left[P_X - \overline{2(AY + b - \bar{X})(X - \bar{X})^T} + \overline{AYY^T A^T} + \overline{(b - \bar{X}) Y^T A^T} + \overline{AY (b - \bar{X})^T} + \overline{(b - \bar{X})(b - \bar{X})^T} \right]. \quad (6.46)$$

Now, since b and \bar{X} are known, we can remove the expected value overline from them.

Also, recognize that

$$\text{tr} \left[(b - \bar{X}) Y^T A^T \right] = \text{tr} \left[(b - \bar{X}) Y^T A^T \right]^T = \text{tr} \left[AY (b - \bar{X}) \right], \quad (6.47)$$

and that

$$\begin{aligned} \overline{AYY^T A^T} &= \overline{A(Y - \bar{Y} + \bar{Y})(Y - \bar{Y} + \bar{Y})^T A^T} \\ &= \overline{A(Y - \bar{Y})(Y - \bar{Y})^T A^T} + \overline{AY(Y - \bar{Y})^T A^T} + \overline{A(Y - \bar{Y})\bar{Y}^T A^T} + \overline{AY\bar{Y}^T A^T} \\ &= AP_Y A^T + A\bar{Y}\bar{Y}^T A^T. \end{aligned}$$

After making those substitutions, the expression for J becomes

$$J = \text{tr} \left[P_X + A(P_Y + \bar{Y}\bar{Y}^T)A^T + 2A\bar{Y}(b - \bar{X})^T + (b - \bar{X})(b - \bar{X})^T - \overline{2(A\bar{Y} + b - \bar{X})(X - \bar{X})^T} \right]. \quad (6.48)$$

Because A and b are constants, we recognize that $\bar{X} = A\bar{Y} + b$. Making this substitution into the last expression in eq.(6.49) leads to a cancellation of the b terms and also to an expression for the cross covariance of Y and X . This produces the final form for the expected value of the sum of the squared errors,

$$J = \text{tr} \left[P_X + A(P_Y + \bar{Y}\bar{Y}^T)A^T + 2A\bar{Y}(b - \bar{X})^T + (b - \bar{X})(b - \bar{X})^T - 2AP_{YX} \right]. \quad (6.49)$$

Arriving at values for A and b that minimize J involves taking the partial derivatives with respect to each term and then setting these equal to zero. Taking these derivatives and using some linear algebra identities² yields

$$\frac{\partial J}{\partial A} = 2A(P_Y + \bar{Y}\bar{Y}^T) - 2P_{XY} + 2(b - \bar{X})\bar{Y}^T = 0, \quad (6.50)$$

$$\frac{\partial J}{\partial b} = 2(b - \bar{X}) + 2A\bar{Y} = 0. \quad (6.51)$$

Solving for A and b gives us

$$\begin{aligned} A &= P_{XY}P_Y^{-1} \\ b &= \bar{X} - P_{XY}P_Y^{-1}\bar{Y}, \end{aligned}$$

and substituting these values back into eq.(6.38) reveal that the solution which minimizes the mean square error in the state estimate is simply

$$\hat{X}_{LMS} = \bar{X} + P_{XY}P_Y^{-1}(Y - \bar{Y}). \quad (6.52)$$

The term $P_{XY}P_Y^{-1}$ determines the extent to which we base our estimated state on the

²The required identities are $(d/dA)\text{tr}(ABA^T) = 2AB$ and $(d/dA)\text{tr}(BAC) = B^TC^T$.

model or on the measurements. We already have a previously existing estimate of the state \bar{X} , and make a correction factor based on the error between the expected and actual measurements. If we have a high degree of confidence in the measurements, P_Y is small, making the correction factor for the state estimate large. Otherwise, P_Y is large and we attenuate the effects of a measurement error.

We would also like to see how the how this update of the state estimate has affected our estimate of the error covariance matrix. Start with the definition of the new error covariance matrix,

$$P_{\hat{X}} = \overline{(X - \hat{X})(X - \hat{X})^T}. \quad (6.53)$$

Now substitute the value for \hat{X} that we obtained in eq.(6.52). This leads to

$$P_{\hat{X}} = \overline{[(X - \bar{X}) - P_{XY}P_Y^{-1}(Y - \bar{Y})] [(X - \bar{X}) - P_{XY}P_Y^{-1}(Y - \bar{Y})]^T}, \quad (6.54)$$

which we can then multiply out into

$$P_{\hat{X}} = \overline{(X - \bar{X})(X - \bar{X})^T} - P_{XY}P_Y^{-1}\overline{(Y - \bar{Y})(X - \bar{X})^T} \\ - \overline{(X - \bar{X})(Y - \bar{Y})^T}P_Y^{-1}P_{YX} + P_{XY}P_Y^{-1}\overline{(Y - \bar{Y})(Y - \bar{Y})^T}P_Y^{-1}P_{YX}. \quad (6.55)$$

This reduces to

$$P_{\hat{X}} = P_X - P_{XY}P_Y^{-1}P_{YX} - P_{XY}P_Y^{-1}P_{YX} + P_{XY}P_Y^{-1}P_YP_Y^{-1}P_{YX}, \quad (6.56)$$

which then reduces one step further to

$$P_{\hat{X}} = P_X - P_{XY}P_Y^{-1}P_{YX}. \quad (6.57)$$

The new error covariance $P_{\hat{X}}$ after taking the measurements is the previous covariance P_X minus an adjustment.

6.7.2 Calculation of the State Error Covariance Matrices

In the previous section we found the optimal estimator gain, but did so in terms of the two unknown covariance matrices, P_Y and P_{XY} . We also found the error covariance of this optimal estimate of the state, but it was again in terms of unknown covariance matrices. In this section, we will derive their values. Begin with the linear model of the measurements,

$$Y = HX + V. \quad (6.58)$$

The variable X is the true unknown state, H is the known constant relationship between them, and V is additive measurement noise that has a mean of zero and a covariance of R . Also note that we are assuming the measurement noise is not affected by the value of the state. This makes X and V orthogonal so that their cross covariance $\overline{XV^T}$ equals zero. The mean of the measurements is not tremendously difficult to determine. Through linearity and the fact that the noise has a mean of zero, the mean of the measurements is just

$$\bar{Y} = H\bar{X}. \quad (6.59)$$

Deriving the covariances takes slightly more work. Start by deriving the covariance of the measurements. This is defined as

$$P_Y = \overline{(Y - \bar{Y})(Y - \bar{Y})^T}. \quad (6.60)$$

Next, substitute eq.(6.58) and eq.(6.59) to get

$$P_Y = \overline{(HX + V - H\bar{X})(HX + V - H\bar{X})^T}, \quad (6.61)$$

which can be expanded into

$$P_Y = \overline{H(X - \bar{X})(X - \bar{X})^T H^T} + \overline{H(X - \bar{X})V^T} + \overline{V(X - \bar{X})^T H^T} + \overline{VV^T}. \quad (6.62)$$

The first term contains the covariance of X and the last is the measurement noise covariance R . Making these two substitutions and expanding the remaining two terms

model or on the measurements. We already have a previously existing estimate of the state \bar{X} , and make a correction factor based on the error between the expected and actual measurements. If we have a high degree of confidence in the measurements, P_Y is small, making the correction factor for the state estimate large. Otherwise, P_Y is large and we attenuate the effects of a measurement error.

We would also like to see how the how this update of the state estimate has affected our estimate of the error covariance matrix. Start with the definition of the new error covariance matrix,

$$P_{\hat{X}} = \overline{(X - \hat{X})(X - \hat{X})^T}. \quad (6.53)$$

Now substitute the value for \hat{X} that we obtained in eq.(6.52). This leads to

$$P_{\hat{X}} = \overline{\left[(X - \bar{X}) - P_{XY}P_Y^{-1}(Y - \bar{Y}) \right] \left[(X - \bar{X}) - P_{XY}P_Y^{-1}(Y - \bar{Y}) \right]^T}, \quad (6.54)$$

which we can then multiply out into

$$\begin{aligned} P_{\hat{X}} = & \overline{(X - \bar{X})(X - \bar{X})^T} - P_{XY}P_Y^{-1}\overline{(Y - \bar{Y})(X - \bar{X})^T} \\ & - \overline{(X - \bar{X})(Y - \bar{Y})^T}P_Y^{-1}P_{YX} + P_{XY}P_Y^{-1}\overline{(Y - \bar{Y})(Y - \bar{Y})^T}P_Y^{-1}P_{YX}. \end{aligned} \quad (6.55)$$

This reduces to

$$P_{\hat{X}} = P_X - P_{XY}P_Y^{-1}P_{YX} - P_{XY}P_Y^{-1}P_{YX} + P_{XY}P_Y^{-1}P_YP_Y^{-1}P_{YX}, \quad (6.56)$$

which then reduces one step further to

$$P_{\hat{X}} = P_X - P_{XY}P_Y^{-1}P_{YX}. \quad (6.57)$$

The new error covariance $P_{\hat{X}}$ after taking the measurements is the previous covariance P_X minus an adjustment.

6.7.2 Calculation of the State Error Covariance Matrices

In the previous section we found the optimal estimator gain, but did so in terms of the two unknown covariance matrices, P_Y and P_{XY} . We also found the error covariance of this optimal estimate of the state, but it was again in terms of unknown covariance matrices. In this section, we will derive their values. Begin with the linear model of the measurements,

$$Y = HX + V. \quad (6.58)$$

The variable X is the true unknown state, H is the known constant relationship between them, and V is additive measurement noise that has a mean of zero and a covariance of R . Also note that we are assuming the measurement noise is not affected by the value of the state. This makes X and V orthogonal so that their cross covariance $\overline{XV^T}$ equals zero. The mean of the measurements is not tremendously difficult to determine. Through linearity and the fact that the noise has a mean of zero, the mean of the measurements is just

$$\bar{Y} = H\bar{X}. \quad (6.59)$$

Deriving the covariances takes slightly more work. Start by deriving the covariance of the measurements. This is defined as

$$P_Y = \overline{(Y - \bar{Y})(Y - \bar{Y})^T}. \quad (6.60)$$

Next, substitute eq.(6.58) and eq.(6.59) to get

$$P_Y = \overline{(HX + V - H\bar{X})(HX + V - H\bar{X})^T}, \quad (6.61)$$

which can be expanded into

$$P_Y = \overline{H(X - \bar{X})(X - \bar{X})^T H^T} + \overline{H(X - \bar{X})V^T} + \overline{V(X - \bar{X})^T H^T} + \overline{VV^T}. \quad (6.62)$$

The first term contains the covariance of X and the last is the measurement noise covariance R . Making these two substitutions and expanding the remaining two terms

produces

$$P_Y = HP_XH^T + \overline{HXV^T} - H\overline{X}\overline{V}^T + \overline{VX^TH^T} - \overline{V}\overline{X}^TH^T + R. \quad (6.63)$$

We have assumed that V is a zero mean vector, and so the two terms containing \overline{V} drop out. In addition, the assumption of orthogonality removes the terms containing $\overline{XV^T}$ and $\overline{VX^T}$. This reduces the measurement covariance matrix to

$$P_Y = HP_XH^T + R. \quad (6.64)$$

Replacing the unknown measurement covariance matrix P_Y by the unknown state covariance matrix P_X may seem pointless right now, but it will be beneficial later.

Next, we want to calculate the cross covariance matrix of X and Y . Once again, start with the definition of the covariance matrix,

$$P_{XY} = \overline{(X - \overline{X})(Y - \overline{Y})^T}, \quad (6.65)$$

and substitute eqs.(6.58) and (6.59) producing

$$P_{XY} = \overline{(X - \overline{X})(HX + V - H\overline{X})^T}. \quad (6.66)$$

Expanding the relationship yields

$$P_{XY} = \overline{(X - \overline{X})(X - \overline{X})^TH^T} + \overline{XV^T} - \overline{X}\overline{V}^T, \quad (6.67)$$

which then just simplifies to

$$P_{XY} = P_XH^T. \quad (6.68)$$

Now we can begin to see the reason for these derivations. We have replaced the measurement covariance P_Y and the cross covariance between the state and the measurement P_{XY} by equations in terms of the single unknown P_X . If we go back to eq.(6.52) which gave the least squares estimate of the state, we can now substitute into it eq.(6.59),

eq.(6.64), and eq.(6.68), transforming it into

$$\hat{X}_{LMS} = \bar{X} + P_X H^T (H P_X H^T + R)^{-1} (Y - H \bar{X}). \quad (6.69)$$

In this relationship we have a formula for the best estimate of the state (in the least squares sense) that can be obtained given a set of measurements Y . All that we need is a previous knowledge of the mean of the state \bar{X} and its covariance P_X .

6.7.3 Time and Measurement Updates

Up until this point, we have been fairly loose with the concept of time. However, since most Kalman filters are implemented in discrete time, it is important to clarify at what times the measurements and state estimates are available to what parts of the system. This concept of “current and predictor” or “*a priori* and *a posteriori*” estimates seems to be approached in a different notation in each reference that I examine. Simply put, *a priori* estimates are those that exist before a measurement is taken. They are equivalent to the predictor estimates of the previous section. After the measurement, the state estimate is modified accordingly to form the *a posteriori* estimate, or the equivalent of the current estimate. Since I have followed Franklin, Powell, and Workman [15] most closely for the notation throughout the control system design, I will continue with it here. An overbar \bar{X} represents the predictor estimate, and a hat \hat{X} represents the current measurement. Two changes that I will make for readability in this section are to move the time index to a subscript, thus replacing $Y(k)$ by Y_k , and to replace P_X by M and $P_{\hat{X}}$ by P . The variable M represents the state error covariance matrix before the latest set of measurements, and P is the error covariance matrix after the measurement update.

We can start with the *measurement update* relationships. These use the measurements to adjust previously existing estimates in order to form the current estimates. The form for the estimate update should be familiar from the previous section,

$$\hat{X}_k = \bar{X}_k + M_k H^T (H M_k H^T + R)^{-1} (Y_k - H \bar{X}_k). \quad (6.70)$$

This is exactly the same as the estimator in eq.(6.25), except that now the gain matrix L is replaced by $M_k H^T (H M_k H^T + R)^{-1}$.

The difference between this and a typical estimator is that we also need to generate a current estimate of the error covariance matrix. Taking eq.(6.57), substituting into it the covariance matrices in eq.(6.64) and eq.(6.68), and noting that $P_{XY} = P_Y^T X$, we arrive at the measurement update of the error covariance,

$$\hat{P}_k = M_k - M_k H^T (H M_k H^T + R)^{-1} H M_k. \quad (6.71)$$

As was said, these relationships give the current estimates of the state and the error covariance matrix. The next step is to use the model of the plant dynamics to predict their values one sample period into the future. This time, in addition to the expected plant dynamics, the system contains process noise w that creates a real, measurable affect on the system. Taking the system model as

$$X_{k+1} = F X_k + G u_k + G_1 w_k, \quad (6.72)$$

we can assume the process noise to have zero mean, a covariance of Q , and to be uncorrelated with the state. Note that these are the same standard assumptions made about the sensor noise. Now, if we want to see how the expected value of the state propagates, we simply write

$$\overline{X}_{k+1} = \overline{F \hat{X}_k + G u_k + G_1 w_k}. \quad (6.73)$$

Again invoking linearity, this reduces to

$$\overline{X}_{k+1} = F \overline{\hat{X}_k} + G \overline{u_k} + G_1 \overline{w_k}. \quad (6.74)$$

We know the current value of the state estimate \hat{X}_k , as well as the value of the deterministic input u_k and the mean of the process noise. Because of that, we can rewrite the time update of the state estimate as

$$\overline{X}_{k+1} = F \overline{X}_k + G u_k. \quad (6.75)$$

This should come as no surprise, since it is the identical form used back in eq.(6.30).

The difference with the Kalman filter is that we now need to predict how the error covariance matrix will propagate through the system dynamics. Start with the definition of the error covariance matrix for the new predicted estimate of the state,

$$M_{k+1} = \overline{(X_{k+1} - \bar{X}_{k+1})(X_{k+1} - \bar{X}_{k+1})^T}. \quad (6.76)$$

Substituting in the models of the system dynamics from eqs.(6.72) and (6.75), we arrive at

$$M_{k+1} = \overline{[F(X_k - \hat{X}_k) + G_1 w_k] [F(X_k - \hat{X}_k) + G_1 w_k]^T}. \quad (6.77)$$

The input to the plant is deterministic, so it cancelled in the above equation. Following the routine procedure at this point, we expand the expression into

$$\begin{aligned} M_{k+1} = & \overline{F(X_k - \hat{X}_k)(X_k - \hat{X}_k)^T F^T} + \overline{G_1 w_k (X_k - \hat{X}_k)^T F^T} \\ & + \overline{F(X_k - \hat{X}_k) w_k^T G_1^T} + \overline{G_1 w_k w_k^T G_1^T}. \end{aligned} \quad (6.78)$$

With the standard assumption of the state and process noise being uncorrelated, the time update of the error covariance matrix becomes

$$M_{k+1} = F P_k F^T + G_1 Q G_1^T. \quad (6.79)$$

This form of a linear matrix equation occurs routinely enough to be called a *Lyapunov equation* for M_k .

At this point, we have completed the derivation of the Kalman filter. As a summary, we have a plant modeled by

$$X_{k+1} = F X_k + G u_k + G_1 w_k \quad (6.80)$$

$$Y_k = H X_k + V_k \quad (6.81)$$

where w_k and V_k follow Gaussian distributions with means of zero and covariance matrices of R and Q respectively. The Kalman filter generates the optimal current estimate of the

state from the measurements using the measurement update equations

$$\hat{X}_k = \bar{X}_k + M_k H^T (H M_k H^T + R)^{-1} (Y_k - H \bar{X}_k) \quad (6.82)$$

$$P_k = M_k - M_k H^T (H M_k H^T + R)^{-1} H M_k. \quad (6.83)$$

Then, we can use the model of the plant dynamics to update these into predicted measurements at the next sample period,

$$\bar{X}_{k+1} = F \hat{X}_k + G u_k \quad (6.84)$$

$$M_{k+1} = F P_k F^T + G_1 Q G_1^T. \quad (6.85)$$

6.7.4 Steady State Solution

At a first glance, the state and measurement update equations in the previous section seem simple enough. In a MATLAB simulation, they work wonderfully. However, they become tricky when they have to be implemented in C code. The problem is that calculating a matrix inverse, especially of a 6×6 or a 12×12 , is not a trivial task in a real time application. Fortunately, there is a way around this problem. After a certain amount of time, the error covariance matrix approaches a steady state value. As long as we are interested mainly in longer term trends, we can calculate this steady state value beforehand and then just use it in determining the weighing factor for the new measurements.

The key idea in this derivation is that for a steady state solution, the predicted error covariance is no longer changing, meaning that $M_{k+1} = M_k$. Substituting eq.(6.83) into eq.(6.85), we arrive at

$$M_{k+1} = F \left[M_k - M_k H^T (H M_k H^T + R)^{-1} H M_k \right] F^T + G_1 Q G_1^T. \quad (6.86)$$

Now, replacing both M_{k+1} and M_k by the steady state covariance P , we have the *algebraic*

Ricatti equation

$$\boxed{P = F \left[P - PH^T(HPH^T + R)^{-1}HP \right] F^T + G_1QG_1^T.} \quad (6.87)$$

This is again a non-trivial equation to solve for P , and because it is nonlinear, there can be multiple solutions. Rather than becoming involved with the details of when it converges to our desired unique, positive definite solution, we will simply assume that it can be solved. Therefore, at this point, we have the steady state error covariance matrix P , and substituting this value into eq.(6.82) yields

$$\boxed{\hat{X}_k = \bar{X}_k + L_K(Y_k - H\bar{X}_k).} \quad (6.88)$$

The matrix L_K is then the steady state Kalman gain matrix defined as

$$L_K = PH^T(HPH^T + R)^{-1}, \quad (6.89)$$

and the Kalman filter functions as a simple estimator.

The final problem is that we have still not solved the algebraic Ricatti equation in eq.(6.87) for the steady state value of P that we need in eq.(6.89). For this, we can head to MATLAB. The function “DLQE” takes the plant model and the noise covariances, and produces both the steady state values of the error covariance and Kalman filter estimator gain. We could have done this from the very start, and skipped the last several pages of derivations, but what fun would that have been?

6.7.5 Application to the Angstrom Stage

We have seen that the Kalman filter provides insight into the tradeoff between bandwidth and position error. However, at present we have not formally used it to set the estimator gains on the basis of the sensor noise and disturbance force statistics. The reason for this is that while the sensor noise is well documented, the disturbance forces are only known for a single degree of freedom. Formally applying a Kalman filter requires knowing the characteristics of the disturbance forces in all six degrees-of-freedom. Instead, I have

set the estimator gain empirically. I knew that the disturbance forces were of a much lower magnitude than was the sensor noise. Therefore, the sensor noise would dominate the system error, and the closed loop poles should be set as slow as possible. It turned out that position estimates converging at 10 Hz kept the position errors reasonably low while still allowing the stage to move quickly enough to be interesting.

6.8 Final Control System Design

The final control system design shown in Figure 6-15 is a mix of several of the elements previously described. The FIR filter of Chapter 5 comes first. Following that are the feedback linearization algorithms. Although these take some time to develop, and require the prior verification of a model, they operate transparently and really simplify the remainder of the control system design. Beyond this, the control system is tailored to meet the specific needs of a particular experiment. The baseline controller is simply proportional control, with the bandwidth set at about 5 Hz (proportional gain of 30 s^{-1}). If obtaining the highest resolution is the goal, the estimator is turned on. However, this adds another level of complexity to the controller. Normally, constant velocity scans are performed using an open loop model of the plant. This model is sufficiently accurate for normal operation, but closing the loop on velocity has the expected result of reducing the errors from the desired straight line trajectory. This comes at the expense of decreased position resolution, since the DAC outputs are based upon a noisy velocity signal. The noise level of the velocity measurement could be reduced through a longer filter, or through the use of an estimator, but this approach is not pursued here.

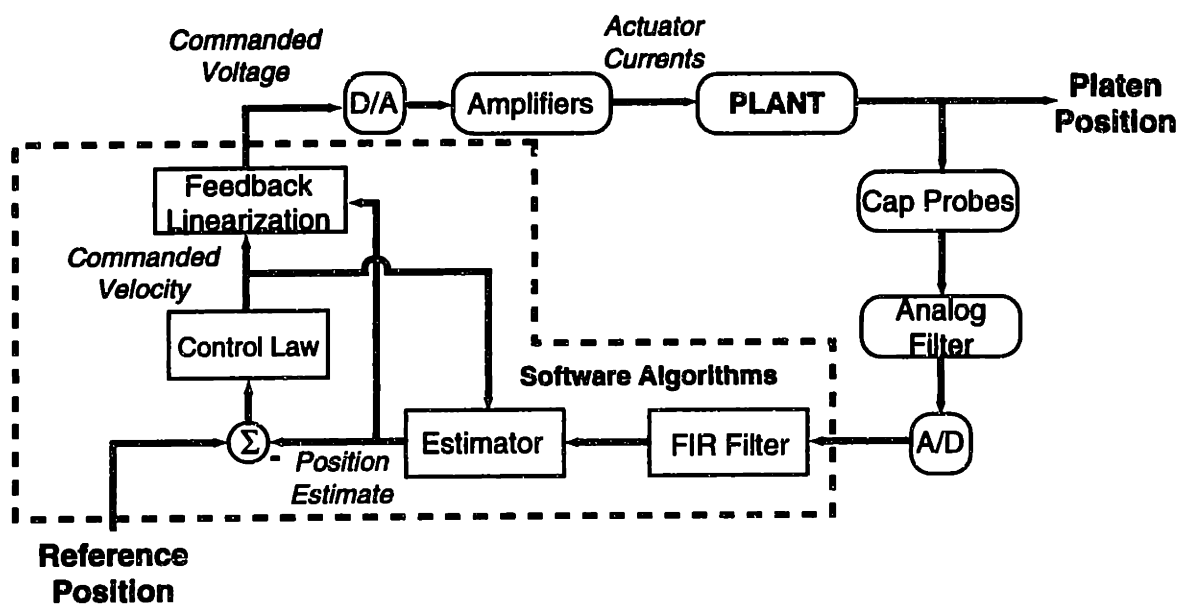


Figure 6-15: Block diagram of the overall control structure.

Chapter 7

Results

This chapter contains examples of the positioning performance of the stage. Rather than interspersing them throughout the thesis, most of the results are collected here. There are two methods generally used for gathering the data shown in these plots. The first is to record the capacitance probe readings through the A/D converter. For scales on the order of tens of nanometers and larger, this is a very efficient approach, since the numerical values of the positions can be stored directly into a text file. The data collection rate is set by the length of time that I wanted to look at while still keeping to a reasonable file size. It is a simple matter to save every capacitance probe reading, every other, every tenth, etc. The downside to this method is that the position readings are corrupted with the A/D converter noise. Since this noise level is greater than the position resolution of the stage, we cannot determine the position of the stage to the sub-nanometer level by this means.

The second method is to use a low noise differential amplifier¹ in order to obtain the highest resolution indication of the stage's position. This is only effective when the output of the capacitance probe is within ± 1 V, since this is the limit of the offset adjustment on the differential amplifier. Within this range, deviations about a constant voltage can be amplified to the level at which a standard oscilloscope or a HP Dynamic Signal Analyzer were capable of recording them. A movement of 1 nm only changes the

¹Tektronix model AM 502

voltage coming from the capacitance probe by $200\ \mu\text{V}$. This is readily detected by the differential amplifier, but is below the noise floor of the other instruments. The downside to this method is that a maximum of two capacitance probe signals can be recorded at once (since we only have two differential amplifiers), and that the waveform is recorded as a voltage. This then has to be multiplied by a scale factor in order to transform it into a position reading. There will be an indication for each plot in this chapter as to which method is being used to record the position data.

A further variable that can change the results is the distance of the platen from the capacitance probes. In other words, the traces can be taken with the platen at its nominal zero position, or at a position closer to the $\pm 50\ \mu\text{m}$ extremes of the scan. Recall that the true physical damping rises quickly as the oil film thickness decreases, therefore even with the feedback linearization algorithm in place, the positioning resolution will change somewhat. Holmes [27] achieved the highest quality images by pulling the platen close to the capacitance probes, and thus damping the stage to the fullest extent possible. In the work presented in this thesis, I have kept the platen located at about the nominal zero position unless otherwise noted.

7.1 Positioning Capabilities

I first present the steady state position noise. For this experiment I measured the capacitance probe data through the differential amplifiers, and recorded it with the HP Dynamic Signal Analyzer. Figure 7-1 shows a trace of the Y-axis position measured through a 1 Hz analog lowpass filter. The lowpass filter is required to attenuate the 0.5 mV of noise present in the capacitance probe electronics within a bandwidth of 100 Hz. In this case, the bandwidth of the Y-axis was set at 5 Hz, and a total of 1024 points were recorded over 128 seconds. The resulting trace of Figure 7-1 has a standard deviation of 0.05 nm. This corresponds to a 6σ peak-to-peak position noise of 0.3 nm.

We next consider small displacement step responses. The control bandwidth is again set at 5 Hz, and the measurement is taken as before through a differential amplifier with a bandwidth of 1 Hz. Figure 7-2 displays repeated 0.5 nm step responses. Included is a

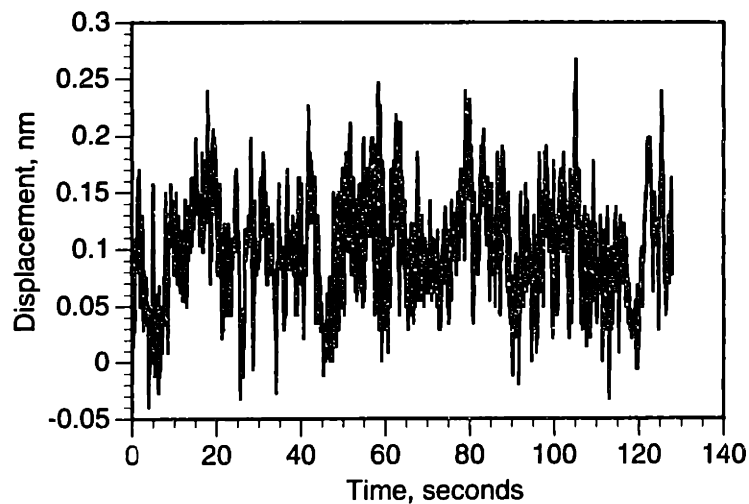


Figure 7-1: Steady state position noise measured through a 1 Hz filter.

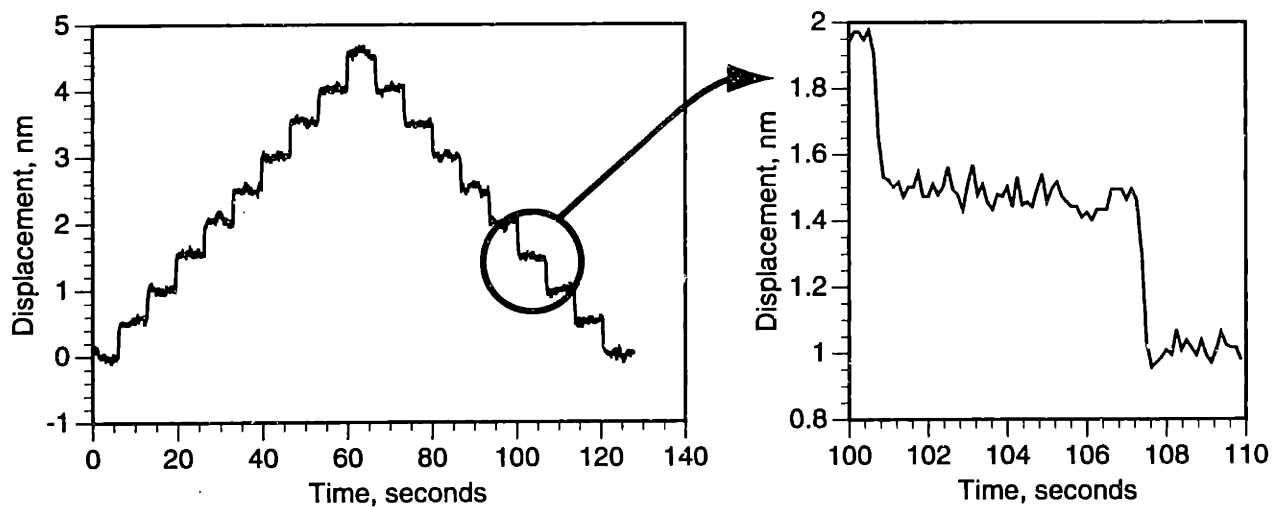


Figure 7-2: Repeated 0.5 nm Step Responses (5 Hz Controller Bandwidth).

detail of one of the individual steps. This is the type of motion required when the stage scans in one direction, and then steps in the perpendicular direction for the next trace.

The final positioning question asks what the best possible resolution of the stage is. With some amount of patience involved in waiting for a lull in the floor vibrations and a decent looking step, I was able to obtain the 0.2 nm step response of Figure 7-3. The controller bandwidth in this case was set to 0.16 Hz. This thus represents a best-case trace with the platen at the center position. Still lower noise could probably be achieved by moving closer to the capacitance probes.

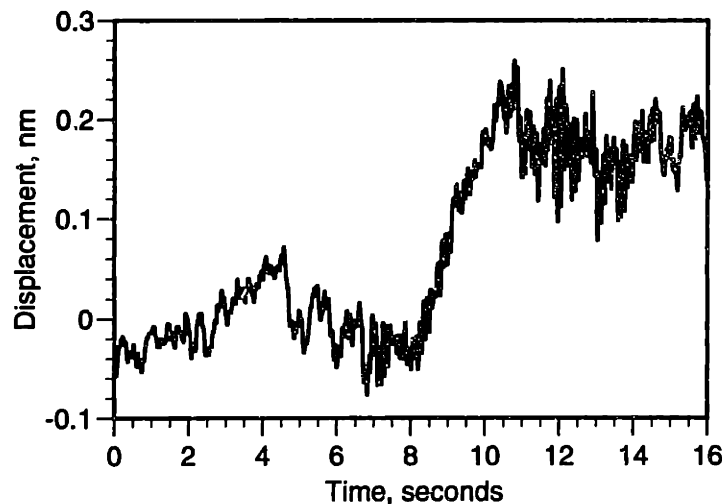


Figure 7-3: An example of a 0.2 nm step response (0.16 Hz controller bandwidth).

7.2 Decoupling of the Six Degrees of Freedom

There are two main reasons for using a feedback linearization procedure on the entire plant. The first is to make the response nearly position independent. The other, which is demonstrated in this section, is to decouple the motions in the axes from each other. Figure 7-4 shows an example of a 50 nm step in the X-direction, as well as the accompanying motions in the other five axes. Ideally, the other five will remain motionless. This data was gathered through the A/D converter (there is no way to collect six channels at once on the HP), and so remember that the readings contain electrical noise on top of the true position indication. The control bandwidth in each of the axes was set at about 3 Hz, and samples from all six channels were taken every 4.46 ms.

Note that in most cases the error motions are indistinguishable from the electrical noise. However, the notable exception is the case of rotations about the Y-axis. In order to generate this scan, actuator E3 applied the X-direction force to the platen. Based upon our plant model, this should not be able to generate a rotation about the Y-axis. However, it is definitely there. It may be that the torque required to generate this motion is caused by a non-uniform gap between actuator E3 and its target. Section 3.4 derived how this geometry causes an actuator to apply a torque as well as a force. Another possibility is the additional shear drag applied to the platen by the top three capacitance probes. At the nominal gap of 100 μm , and a nominal speed of 1 $\mu\text{m/s}$, the probes apply

a torque of about $0.7 \mu\text{Nm}$ to the platen. This is on the same order of magnitude as the actuator torque in the case of small misalignment angles, but is two orders of magnitude smaller when the misalignment angle is large. This is a repeatable error, and so could be modeled with the correction incorporated into the controller software. However, this option was not investigated in the present work.

7.3 Velocity Control

There were two different methods described in Section 6.4. The first relied on an explicit feedback of a measured velocity. This measured velocity was obtained by differentiating the position readings, and the noise inherent in these readings becomes a source of positioning error at the nanometer level. However, this method proved to be more effective than the second, simpler method of trajectory planning.

7.3.1 Feedback of Measured Velocity

At this point we can evaluate the effectiveness of the velocity control strategy. To begin with, Figure 7-5 compares constant velocity scans over nearly the entire X-axis travel, $+45 \mu\text{m}$ to $-45 \mu\text{m}$. The each trace contains 800 data points collected through the A/D converter at a sampling rate of 22.3 Hz. In this case, the commanded velocity was $2.5 \mu\text{m/s}$ which did not saturate the DAC's at any point over the travel. At this scale the traces are nearly coincident and only begin to diverge at the extreme right when the platen is getting very close to two capacitance probes. Here, the damping value is rising very quickly because of the $1/\text{gap}^3$ relationship.

Fitting a line to the traces more clearly shows how constant the velocity is being held. A linear regression on each indicated a velocity of $2.493 \mu\text{m/s}$ in the closed loop case, and $2.466 \mu\text{m/s}$ open loop. Figure 7-6 plots the residuals for both cases. The maximum deviation from a straight line in the open loop case was about 800 nm, while in the in the closed loop case it was reduced to about 400 nm. However, note that most of the error again came at the extreme right of the scan when the damping was changing the fastest.

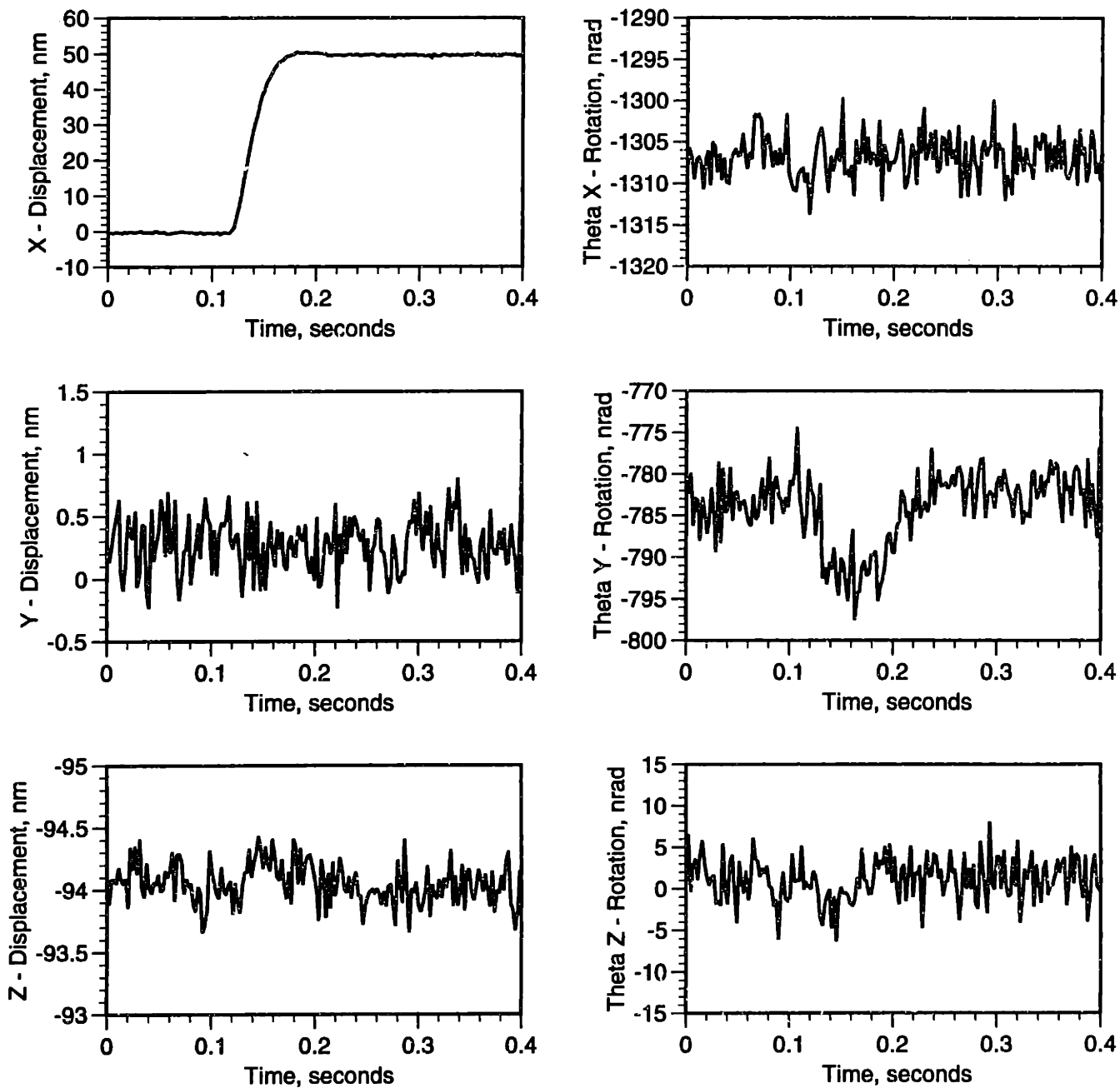


Figure 7-4: Error motions in the other axes after a 50 nm X-direction step.

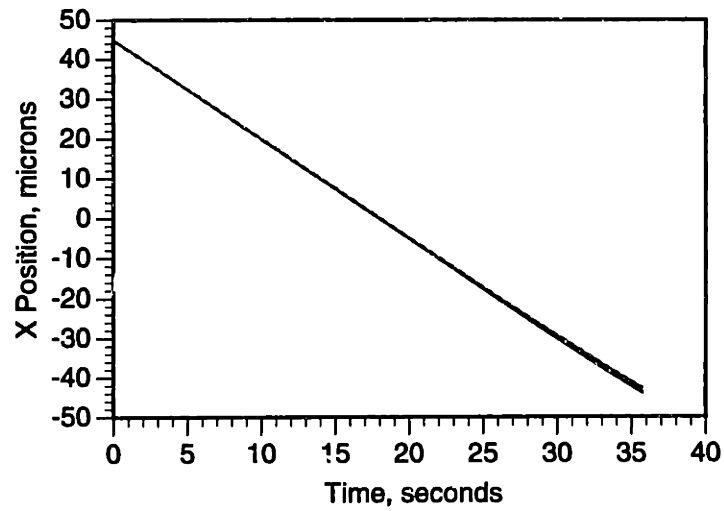


Figure 7-5: A comparison of scans with both open and closed loop control.

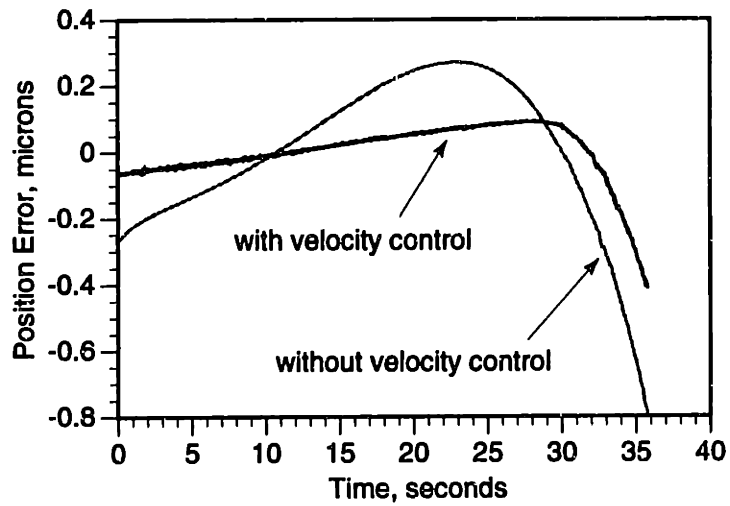


Figure 7-6: Residual errors from a best fit straight line to the entire scan.

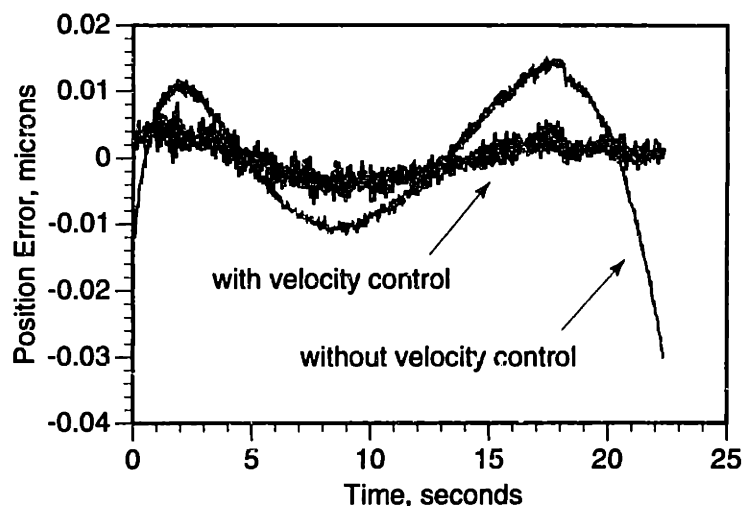


Figure 7-7: Residual errors from a best fit straight line to a partial scan.

We can also try fitting a line to only the first 500 points of this 800 point sequence, thereby ignoring the worst data at the end of the scan. Over the range from about $45\text{ }\mu\text{m}$ to $-10\text{ }\mu\text{m}$, the velocity was $2.491\text{ }\mu\text{m/s}$ in the open loop case, and $2.499\text{ }\mu\text{m/s}$ with closed loop control. More interesting is again the plot of the residuals in Figure 7-7. Now, the deviation from the best fit straight line is less than 5 nm with closed loop control.

A remaining question is why the closed loop control still failed at the extreme of the scan. It was definitely an improvement over the open loop case, but did not hold the velocity as constant as we would like. After some trial and error, I believe this to be caused by the low bandwidth (below 1 Hz) of the velocity controller. With higher bandwidths, I was able to keep the mean of the velocity constant over the entire range, but this came at a great cost in position stability. Further improvements to velocity control will have to start with improvements to our method of estimating velocity.

7.3.2 Updating Commanded Position (Trajectory Following)

The other method by which we could attempt to generate a constant velocity is by incrementing the commanded position. Figure 7-8 shows a trace in which the commanded position in the Y-direction was incremented by 0.1 nm each clock cycle of the 'C30 (2.24 ms), yielding a net velocity of 44.6 nm/s . This data was recorded digitally

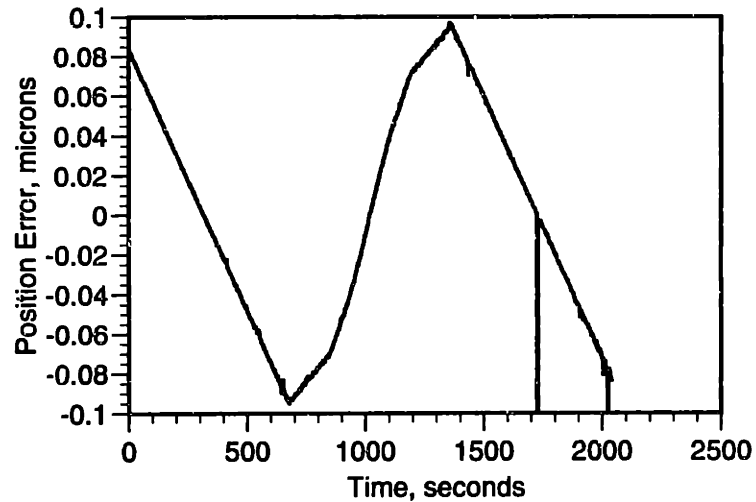


Figure 7-8: Residual errors from a best fit straight line to a trajectory generated by updating the position setpoint.

through the analog to digital converter. Several characteristics of Figure 7-8 deserve a comment. First, the magnitude of the errors are much higher than they were in the case of explicit velocity feedback. Second, there is one large single point error. Because it is limited to a single data point over the 2000 seconds (about 33 minutes), I believe it to be caused by another I/O problem that I have not been able to successfully track down. The error curve also has an odd shape with noticable discontinuities in its slope. This is due to an intentional discontinuity on the controller. Whenever the position error is greater than 100 nm (a somewhat arbitrarily chosen value), the commanded velocity from the controller is limited. This prevents the D/A converter from saturating for large commanded step sizes. Clearly, much can be done to improve the ability of the stage to track a commanded position. However, I found that explicit feedback of a velocity measurement was the quickest and simplest method by which the constant velocity scanning capabilities of the stage could be improved.

Chapter 8

Conclusions

The conclusion of this thesis contains some thoughts on what worked well with this design, what did not work so well, and what areas merit further work.

8.1 Design Strong Points

The key element of this design is the oil flotation that not only leads to a highly overdamped system, but is also responsible for the attenuation of the base accelerations. With these tremendously high levels of damping, the stage is very tolerant of different controller designs, and shows no tendency to go unstable as do most designs in air.

A further strong point of this design is that the hardware is simple enough to model analytically. The squeeze film dampers are well modeled by the Navier–Stokes equation, and our magnetic calibration test fixture has given an actuator model that works under the usual operating conditions. This highly analytical model of the plant is crucial to the success of the feedback linearization scheme because it allows us to implement real-time linearization and to base the controller on an estimator utilizing an accurate model.

8.2 Design Weak Points

The clearest limitation of this present hardware is the poor performance of the analog-to-digital converter. However, this does not represent a limitation of the design itself,

only of the components available at the present time. These can easily be replaced with quieter units if sufficient funds become available.

The primary downfall of this design is its cost and complexity.

- Six capacitance probe channels: \$ 12,000
- A/D converter: \$ 2000
- D/A converter: \$ 1000
- C30-DSP board: \$ 2000
- Host computer: \$ 2000

Added to this is the cost of the twelve amplifiers, the power supplies, the actuators, and all of the materials and fabrication costs. This leads to a motion control stage costing close to half of the typical selling price of an Atomic Force Microscope system. As was described in the Introduction, this stage represents a fundamentally better way to position a sample under a scanned probe microscope, but at present the cost remains too high for the widespread use of the stage. However, specialized machines may sufficiently benefit from the performance advances to justify the cost.

8.3 Suggestion for Further Work

Currently, the stage has resolution, but little is known of its accuracy. Therefore, if any additional work is to be done to the stage, it should involve systematically measuring a sample of known geometry (for example, a diffraction grating or carbon atoms on the surface of graphite). In this way, the the true resolution and longer term accuray of the stage can be established with a gage that is independent of the capacitance probes. Thermal effects will become very significant in this study, and some effort will have to be made to stabilize the temperature of the stage. The temperature of the lab varies greatly, so this study may involve building a thermally insulated box, or even placing the entire assembly in a styrofoam cooler with built in temperature controls.

8.4 Acknowledgements

This work was supported by the National Science Foundation under grants DDM-9396605, DMI-9414778, through David Trumper's Presidential Young Investigator Award, DDM-9496102, and through Stephen Ludwick's National Science Foundation Graduate Student Fellowship. We gratefully acknowledge the assistance provided by the ADE Corporation in the development of the capacitance gaging system.

Chapter 9

Bibliography

- [1] David Batchelder. Analysis and design of high-precision capacitance probes for use in a precision motion control stage. M.S. Thesis, Department of Electrical Engineering, University of North Carolina at Charlotte, Charlotte, NC, Jun 1994.
- [2] Joseph J. Becker. Permanent magnets. *Scientific American*, pages 92–100, Dec 1970.
- [3] Roxana Behrouzjou. Analysis and design of a scanning tunneling microscope stage. M.S. Thesis, Department of Electrical Engineering, University of North Carolina at Charlotte, Charlotte, NC, Jun 1994.
- [4] G. Binnig, C. F. Quate, and Ch. Gerber. Atomic Force Microscope. *Physical Review Letters*, 56(9):930–933, Mar 1986.
- [5] G. Binnig and H. Rohrer. Scanning tunneling microscopy. *Helvetica Physica Acta*, 55:726–735, 1982.
- [6] Roger Carr and Ruth Ellen Thomson. Device for high resolution positioning. *Ultra-microscopy*, 42–44:1606–1609, 1992.
- [7] Kuo Shen Chen. A spring-dominated regime design of a high load capacity, electromagnetically driven x–y– θ stage. M.S. Thesis, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, Jun 1995.
- [8] Margaret Cheney. *Tesla, man out of time*. Prentice Hall, Englewood Cliffs, NJ, 1981.
- [9] Derek G. Chetwynd and Stuart T. Smith. High precision surface profilometry: From stylus to STM. In J. W. Gardner and H. T. Hingle, editors, *From Instrumentation to Nanotechnology*, pages 273–297. Gordon and Breach Science Publishers, Philadelphia, 1991.
- [10] John W. Clark. An electronic analytical balance. *Review of Scientific Instruments*, 18(12):915–918, Dec 1947.

- [11] B. W. Corb, M. Ringger, and H.-J. Guntherodt. An electromagnetic microscopic positioning device for the scanning tunneling microscope. *Journal of Applied Physics*, 58(11):3947–3953, Dec 1985.
- [12] D. N. Davydov, R. Deltour, N. Horii, V. A. Timofeev, and A. S. Grokholski. Cryogenic scanning tunneling microscope with a magnetic coarse approach. *Review of Scientific Instruments*, 64(11):3153–3156, Nov 1993.
- [13] Richard C. Dorf. *Modern Control Systems*. Addison Wesley, New Jersey, seventh edition, 1995.
- [14] James A. Fay. *Introduction to Fluid Mechanics*. MIT Press, Cambridge, MA, 1994.
- [15] Gene F. Franklin, J. David Powell, and Michael W. Workman. *Digital Control of Dynamic Systems*. Addison Wesley, Reading, MA, second edition, 1990.
- [16] Richard H. Frazier, Philip J. Gilinson Jr., and George A. Oberbeck. *Magnetic and Electric Suspensions*. The MIT Press, Cambridge, MA, 1974.
- [17] Dudley D. Fuller. *Theory and Practice of Lubrication for Engineers*. John Wiley & Sons, New York, second edition, 1984.
- [18] Larry Germann and Janet Braccio. Fine-steering mirror technology supports 10 nanoradian systems. *Optical Engineering*, 29(11):1351–1359, Nov 1990.
- [19] Robert M. Gray. Quantization noise spectra. *IEEE Transactions on Information Theory*, 36(6):1220–1244, Nov 1990.
- [20] Robert M. Gray and Lee D. Davidson. *Random Processes: A Mathematical Approach for Engineers*. Prentice-Hall Inc., New Jersey, 1986.
- [21] Mohinder S. Grewal and Angus P. Andrews. *Kalman Filtering: Theory and Practice*. Prentice-Hall, Englewood Cliffs, NJ, 1993.
- [22] J. E. Griffith, G.L. Miller, C.A. Green, D.A. Grigg, and P.E. Russell. A scanning tunneling microscope with a capacitive-based position monitor. *Journal of Vacuum Science and Technology B*, 8(6):2023–2027, Nov/Dec 1990.
- [23] Donald F. Hays. Squeeze films for rectangular plates. *Journal of Basic Engineering, Trans. ASME*, pages 243–246, Jun 1963.
- [24] Joe. D. Hoffman. *Numerical Methods for Engineers and Scientists*. McGraw-Hill, New York, 1992.
- [25] Robert V. Hogg and Johannes Ledolter. *Applied Statistics for Engineers and Physical Scientists*. Macmillan Publishing Company, New York, 1992.
- [26] Michael L. Holmes. Analysis and design of a magnetically suspended precision motion control stage. M.S. Thesis, Department of Electrical Engineering, University of North Carolina at Charlotte, Charlotte, NC, Jun 1994.

- [27] Michael L. Holmes, David L. Trumper, and Robert Hocken. Atomic-scale precision motion control stage (The Angstrom Stage). *CIRP Annals*, 44(1):455–460, 1995.
- [28] W. R. Horsfield. Ruling engine with hydraulic drive. *Applied Optics*, 4(2):189–193, Feb 1965.
- [29] Hiroshi Kaizuka. Application of capacitor insertion method to scanning tunneling microscopes. *Review of Scientific Instruments*, 60(10):3119–3122, Oct 1989.
- [30] Hiroshi Kaizuka and Byron Sui. A simple way to reduce hysteresis and creep when using piezoelectric actuators. *Japanese Journal of Applied Physics*, 27(5):L773–L776, May 1988.
- [31] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering: Series D, Trans. ASME*, pages 35–45, Mar 1960.
- [32] R. E. Kalman. New results in linear filtering and prediction theory. *Journal of Basic Engineering: Series D, Trans. ASME*, pages 95–108, Mar 1961.
- [33] Thomas R. Kurfess and Mark L. Nagurka. Understanding the root locus using gain plots. *IEEE Control Systems*, 2(1):48–52, Jul 1991.
- [34] Alexander V. Kuzin, Michael L. Holmes, Roxana Behrouzjou, and David L. Trumper. Analysis of achievable disturbance attenuation in a precision magnetically-suspended motion control system. In *Proceedings of the Second International Symposium on Magnetic Suspension Technology*, pages 653–665, Seattle, WA, Aug 11-13 1993. NASA Conference Publication No. 3247, Part 2.
- [35] Frank L. Lewis. *Optimal Estimation with an Introduction to Stochastic Control Theory*. John Wiley & Sons, New York, 1986.
- [36] X. Liu, S.T. Smith, and D. G. Chetwynd. Tunneling probes for metrological applications using a long-range profilometer. *Review of Scientific Instruments*, 64(11):3161–3168, Nov 1993.
- [37] Jerry B. Marion and William F. Hornyak. *Physics for Science and Engineering: Part 2*. Saunders College Publishing, Philadelphia, 1982.
- [38] J. R. Mately, R. S. Crandall, and B. Brycki. Bimorph-driven x-y-z translation stage for scanned image microscopy. *Review of Scientific Instruments*, 58(4):567–570, Apr 1987.
- [39] James D. Medbery and Lawrence M. Germann. The 6 degree-of-freedom (DOF) magnetically-suspended fine-steering mirror. In *Preprint from Acquisition, Tracking, and Pointing V*, volume 1482. SPIE - The International Society for Optical Engineering, Apr 1991.
- [40] Hiroshi Mizumoto, Makoto Yabuya, Tatsuhito Shimuzi, and Yoshihiro Kami. An angstrom-positioning system using a twist-roller friction drive. *Precision Engineering*, 17(1):57–62, Jan 1995.

- [41] Richard E. Mortensen. *Random Signals and Systems*. John Wiley & Sons, New York, 1987.
- [42] C. V. Newcomb and I. Flinn. Improving the linearity of piezoelectric ceramic actuators. *Electronics Letters*, 18(11):442–443, May 1982.
- [43] Sean M. Olson. Nonlinear compensation of a single degree of freedom magnetic suspension system. M.S. Thesis, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, Jun 1994.
- [44] Alan V. Oppenheim and Ronald W. Shafer. *Discrete Time Signal Processing*. Prentice-Hall Inc., New Jersey, 1989.
- [45] Oscar Pinkus and Beno Sternlicht. *Theory of Hydrodynamic Lubrication*. McGraw Hill, New York, 1961.
- [46] Tony Poovey, Michael L. Holmes, and David L. Trumper. A kinematically coupled magnetic calibration fixture. *Precision Engineering*, 16(2), Apr 1994.
- [47] William M. Seibert. *Circuits, Signals, and Systems*. The MIT Press, Cambridge, MA, 1986.
- [48] Alexander H. Slocum. *Precision Machine Design*. Prentice-Hall, New Jersey, 1992.
- [49] Alexander H. Slocum, Paul A. Scagnetti, Nathan R. Kane, and Christoph Brunner. Design of self-compensated, water-hydrostatic bearings. *Precision Engineering*, 17(3):175–185, Jul 1995.
- [50] A. R. Smith, S. Gwo, and C. K. Shih. A new high-resolution two dimensional micropositioning device for scanning probe microscopy applications. *Review of Scientific Instruments*, 65(10):3216–3219, Oct 1994.
- [51] Douglas P. E. Smith and Scott A. Elrod. Magnetically driven micropositioners. *Review of Scientific Instruments*, 56(10):1970–1971, Oct 1985.
- [52] Stuart T. Smith and Derek G. Chetwynd. *Foundations of Ultraprecision Mechanism Design*. Gordon and Breach Science Publishers, Switzerland, 1992.
- [53] David L. Trumper. *Magnetic Suspension Techniques for Precision Motion Control*. Ph.D. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, Sep 1990.
- [54] David L. Trumper. Analysis and design of a viscously damped magnetic suspension system. In *Proceedings of the 1993 NSF Design and Manufacturing Grantees Conference*, pages 1683–1687, Jan 1993.
- [55] H. Kumar Wickramasinghe. Scanned probe microscopes. *Scientific American*, pages 98–105, Oct 1989.

- [56] Herman W. P. Willemsen and Grantley O. Este. Positioning device for optical system element, Oct 1984. United States Patent No. 4,475,033.
- [57] Mark E. Williams and David L. Trumper. Materials for efficient high-flux magnetic bearing actuators. In *Proceedings of the Second International Symposium on Magnetic Suspension Technology*, pages 135–145, Seattle, WA, Aug 11-13 1993. NASA Conference Publication No. 3247, Part 1.
- [58] Mark E. Williams, David L. Trumper, and Robert Hocken. Magnetic bearing stage for photolithography. *CIRP Annals*, 42(1):607–610, 1993.
- [59] Herbert H. Woodson and James R. Melcher. *Electromechanical Dynamics Part I: Discrete Systems*. Krieger Publishing Company, Malabar, FL, 1968.
- [60] Russell D. Young. Field emission ultramicrometer. *Review of Scientific Instruments*, 37(3):275–278, Mar 1966.
- [61] Russell D. Young, John Ward, and Frederic Scire. The Topografner: An instrument for measuring surface topology. *Review of Scientific Instruments*, 43(7):999–1011, Jul 1972.

Appendix A

Derivation of the Damping Matrix

A.1 Lateral Motions

We begin modeling the axis coupling by drawing a free body diagram of the stage in the lateral directions with the six actuators and three capacitance probes around it, as is shown in Figure A-1. The mass of the stage is much less significant than the damping in the stage dynamics, and so the only responses to externally applied forces come from the dampers. That is, inertial effects are insignificant at the bandwidths we will be considering. Equations of motion in the two lateral and one rotational direction are as follows.

$$\begin{aligned} F_x &= \frac{b_a \dot{x}}{g_{a3}^3} + \frac{b_a \dot{x}}{g_{a6}^3} + \frac{b_c(\dot{x} - \ell \dot{\theta}_z)}{g_{c2}^3} + \frac{b_c(\dot{x} + \ell \dot{\theta}_z)}{g_{c3}^3} \\ F_y &= \frac{b_a(\dot{y} - \ell \dot{\theta}_z)}{g_{a1}^3} + \frac{b_a(\dot{y} + \ell \dot{\theta}_z)}{g_{a2}^3} + \frac{b_a(\dot{y} + \ell \dot{\theta}_z)}{g_{a4}^3} + \frac{b_a(\dot{y} - \ell \dot{\theta}_z)}{g_{a5}^3} + \frac{b_c \dot{y}}{g_{c1}^3} \\ T_z &= \frac{b_a \ell(\ell \dot{\theta}_z - \dot{y})}{g_{a1}^3} + \frac{b_a \ell(\ell \dot{\theta}_z + \dot{y})}{g_{a2}^3} + \frac{b_a \ell(\ell \dot{\theta}_z + \dot{y})}{g_{a4}^3} \\ &\quad + \frac{b_a \ell(\ell \dot{\theta}_z - \dot{y})}{g_{a5}^3} + \frac{b_c \ell(\ell \dot{\theta}_z - \dot{x})}{g_{c2}^3} + \frac{b_c \ell(\ell \dot{\theta}_z + \dot{x})}{g_{c3}^3} \end{aligned}$$

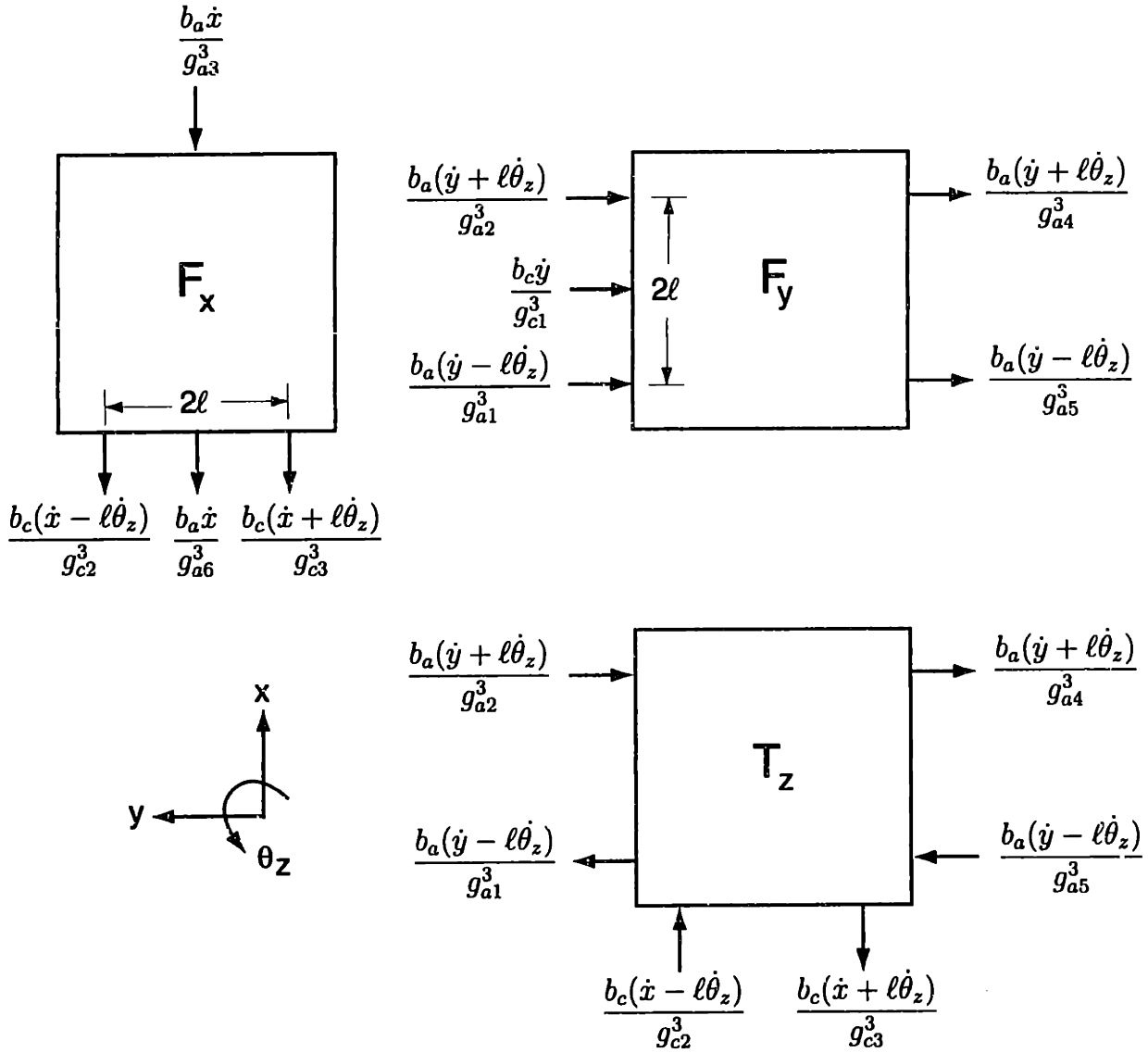


Figure A-1: Free body diagram of the lateral degrees of freedom of the platen.

Algebraically manipulating the above equations so that like velocities are grouped better displays the coupling action, and leads to

$$\begin{aligned} F_z &= b_{xx}\dot{x} + b_{x\theta_z}\dot{\theta}_z \\ F_y &= b_{yy}\dot{y} + b_{y\theta_z}\dot{\theta}_z \\ T_z &= b_{\theta_z x}\dot{x} + b_{\theta_z y}\dot{y} + b_{\theta_z \theta_z}\dot{\theta}_z \end{aligned}$$

The individual damping components are defined as

$$\begin{aligned} b_{xx} &= b_a \left(\frac{1}{g_{a3}^3} + \frac{1}{g_{a6}^3} \right) + b_c \left(\frac{1}{g_{c2}^3} + \frac{1}{g_{c3}^3} \right) \\ b_{x\theta_z} &= b_c \ell \left(\frac{1}{g_{c3}^3} - \frac{1}{g_{c2}^3} \right) \\ b_{yy} &= b_a \left(\frac{1}{g_{a1}^3} + \frac{1}{g_{a2}^3} + \frac{1}{g_{a4}^3} + \frac{1}{g_{a5}^3} \right) + b_c \left(\frac{1}{g_{c1}^3} \right) \\ b_{y\theta_z} &= b_a \ell \left(-\frac{1}{g_{a1}^3} + \frac{1}{g_{a2}^3} + \frac{1}{g_{a4}^3} - \frac{1}{g_{a5}^3} \right) \\ b_{\theta_z x} &= b_c \ell \left(-\frac{1}{g_{c2}^3} + \frac{1}{g_{c3}^3} \right) \\ b_{\theta_z y} &= b_a \ell \left(-\frac{1}{g_{a1}^3} + \frac{1}{g_{a2}^3} - \frac{1}{g_{a4}^3} + \frac{1}{g_{a5}^3} \right) \\ b_{\theta_z \theta_z} &= b_a \ell^2 \left(\frac{1}{g_{a1}^3} + \frac{1}{g_{a2}^3} + \frac{1}{g_{a4}^3} + \frac{1}{g_{a5}^3} \right) + b_c \ell^2 \left(\frac{1}{g_{c2}^3} + \frac{1}{g_{c3}^3} \right) \end{aligned}$$

A.2 Vertical Motions

Similarly, we can calculate the effective damping for the vertical degrees of freedom, i.e. F_z , T_x , and T_y . The free body diagram of this system is shown in Figure A-2. Since the inertial effects are again considered to be insignificant, the equations of motion in each of the directions are

$$\begin{aligned} F_z &= \frac{b_c}{g_{c4}^3} \left(\dot{z} - \frac{\sqrt{3}}{2} \ell \dot{\theta}_x + \frac{1}{2} \ell \dot{\theta}_y \right) + \frac{b_c}{g_{c5}^3} \left(\dot{z} + \frac{\sqrt{3}}{2} \ell \dot{\theta}_x + \frac{1}{2} \ell \dot{\theta}_y \right) + \frac{b_c}{g_{c6}^3} \left(\dot{z} - \ell \dot{\theta}_y \right) \\ &\quad + \frac{b_a}{g_{a7}^3} \left(\dot{z} + \frac{\sqrt{3}}{2} \ell \dot{\theta}_x - \frac{1}{2} \ell \dot{\theta}_y \right) + \frac{b_a}{g_{a8}^3} \left(\dot{z} - \frac{\sqrt{3}}{2} \ell \dot{\theta}_x - \frac{1}{2} \ell \dot{\theta}_y \right) \end{aligned}$$

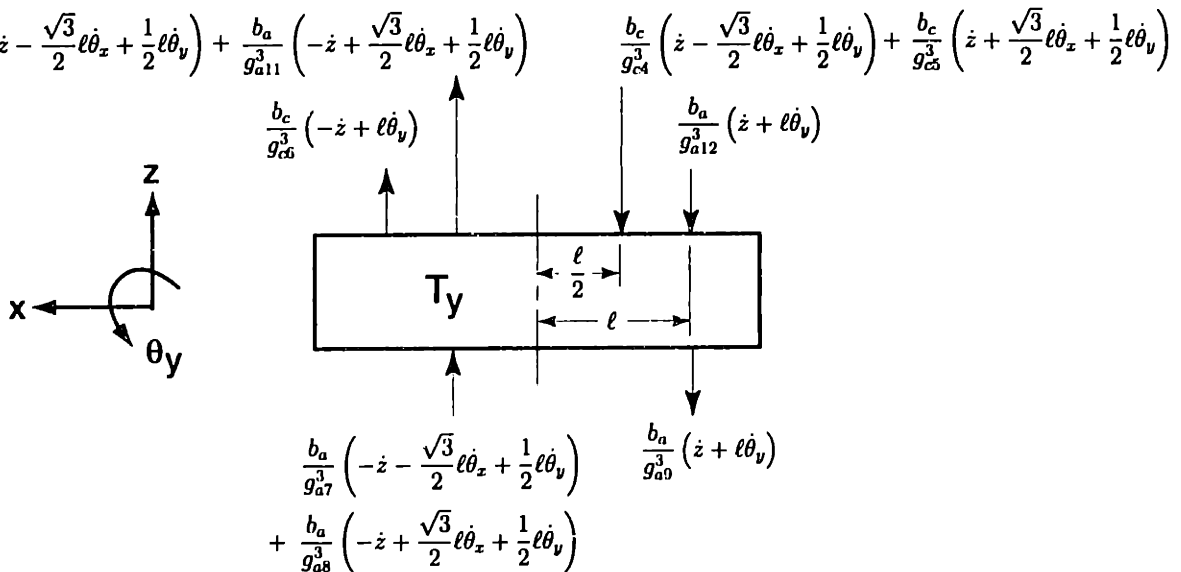
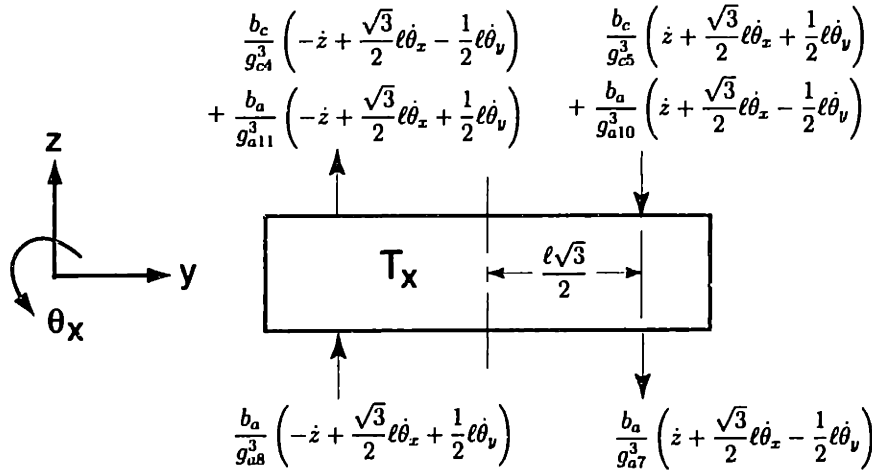
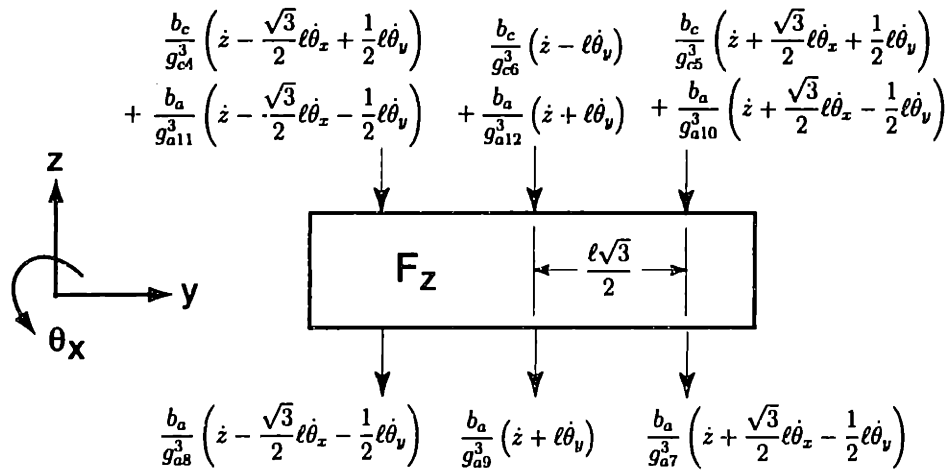


Figure A-2: Free body diagram of the vertical degrees of freedom of the platen.

$$\begin{aligned}
& + \frac{b_a}{g_{a9}^3} \left(\dot{z} + \ell \dot{\theta}_y \right) + \frac{b_a}{g_{a10}^3} \left(\dot{z} + \frac{\sqrt{3}}{2} \ell \dot{\theta}_x - \frac{1}{2} \ell \dot{\theta}_y \right) \\
& + \frac{b_a}{g_{a11}^3} \left(\dot{z} - \frac{\sqrt{3}}{2} \ell \dot{\theta}_x - \frac{1}{2} \ell \dot{\theta}_y \right) + \frac{b_a}{g_{a12}^3} \left(\dot{z} + \ell \dot{\theta}_y \right) \\
T_x = & \frac{\ell \sqrt{3}}{2} \frac{b_c}{g_{c4}^3} \left(-\dot{z} + \frac{\sqrt{3}}{2} \ell \dot{\theta}_x - \frac{1}{2} \ell \dot{\theta}_y \right) + \frac{\ell \sqrt{3}}{2} \frac{b_c}{g_{c5}^3} \left(\dot{z} + \frac{\sqrt{3}}{2} \ell \dot{\theta}_x - \frac{1}{2} \ell \dot{\theta}_y \right) \\
& + \frac{\ell \sqrt{3}}{2} \frac{b_a}{g_{a7}^3} \left(\dot{z} + \frac{\sqrt{3}}{2} \ell \dot{\theta}_x - \frac{1}{2} \ell \dot{\theta}_y \right) + \frac{\ell \sqrt{3}}{2} \frac{b_a}{g_{a8}^3} \left(-\dot{z} + \frac{\sqrt{3}}{2} \ell \dot{\theta}_x - \frac{1}{2} \ell \dot{\theta}_y \right) \\
& + \frac{\ell \sqrt{3}}{2} \frac{b_a}{g_{a10}^3} \left(\dot{z} + \frac{\sqrt{3}}{2} \ell \dot{\theta}_x - \frac{1}{2} \ell \dot{\theta}_y \right) + \frac{\ell \sqrt{3}}{2} \frac{b_a}{g_{a11}^3} \left(-\dot{z} + \frac{\sqrt{3}}{2} \ell \dot{\theta}_x + \frac{1}{2} \ell \dot{\theta}_y \right) \\
T_y = & \frac{\ell}{2} \frac{b_c}{g_{c4}^3} \left(\dot{z} - \frac{\sqrt{3}}{2} \ell \dot{\theta}_x + \frac{1}{2} \ell \dot{\theta}_y \right) + \frac{\ell}{2} \frac{b_c}{g_{c5}^3} \left(\dot{z} + \frac{\sqrt{3}}{2} \ell \dot{\theta}_x + \frac{1}{2} \ell \dot{\theta}_y \right) \\
& + \ell \frac{b_c}{g_{c6}^3} \left(-\dot{z} + \ell \dot{\theta}_y \right) + \frac{\ell}{2} \frac{b_a}{g_{a7}^3} \left(-\dot{z} - \frac{\sqrt{3}}{2} \ell \dot{\theta}_x + \frac{1}{2} \ell \dot{\theta}_y \right) \\
& + \frac{\ell}{2} \frac{b_a}{g_{a8}^3} \left(-\dot{z} + \frac{\sqrt{3}}{2} \ell \dot{\theta}_x + \frac{1}{2} \ell \dot{\theta}_y \right) + \ell \frac{b_a}{g_{a9}^3} \left(\dot{z} + \ell \dot{\theta}_y \right) \\
& + \frac{\ell}{2} \frac{b_a}{g_{a10}^3} \left(-\dot{z} - \frac{\sqrt{3}}{2} \ell \dot{\theta}_x + \frac{1}{2} \ell \dot{\theta}_y \right) + \frac{\ell}{2} \frac{b_a}{g_{a11}^3} \left(-\dot{z} + \frac{\sqrt{3}}{2} \ell \dot{\theta}_x + \frac{1}{2} \ell \dot{\theta}_y \right) \\
& + \ell \frac{b_a}{g_{a12}^3} \left(\dot{z} + \ell \dot{\theta}_y \right)
\end{aligned}$$

As in the case of the lateral motions, the equations become more visual when grouped as a equivalent damping function times the velocity in each of the directions. This leads to the set of equations

$$\begin{aligned}
F_z &= b_{zz} \dot{z} + b_{z\theta_x} \dot{\theta}_x + b_{z\theta_y} \dot{\theta}_y \\
T_x &= b_{\theta_x z} \dot{z} + b_{\theta_x \theta_x} \dot{\theta}_x + b_{\theta_x \theta_y} \dot{\theta}_y \\
T_y &= b_{\theta_y z} \dot{z} + b_{\theta_y \theta_x} \dot{\theta}_x + b_{\theta_y \theta_y} \dot{\theta}_y
\end{aligned}$$

where the equivalent damping functions are defined as

$$\begin{aligned}
b_{zz} &= b_a \left(\frac{1}{g_{a7}^3} + \frac{1}{g_{a8}^3} + \frac{1}{g_{a9}^3} + \frac{1}{g_{a10}^3} + \frac{1}{g_{a11}^3} + \frac{1}{g_{a12}^3} \right) + b_c \left(\frac{1}{g_{c4}^3} + \frac{1}{g_{c5}^3} + \frac{1}{g_{c6}^3} \right) \\
b_{z\theta_x} &= \frac{b_a \ell \sqrt{3}}{2} \left(\frac{1}{g_{a7}^3} - \frac{1}{g_{a8}^3} + \frac{1}{g_{a10}^3} - \frac{1}{g_{a11}^3} \right) + \frac{b_c \ell \sqrt{3}}{2} \left(-\frac{1}{g_{c4}^3} + \frac{1}{g_{c5}^3} \right)
\end{aligned}$$

$$\begin{aligned}
b_{x\theta_y} &= \frac{b_a \ell}{2} \left(-\frac{1}{g_{a7}^3} - \frac{1}{g_{a8}^3} + \frac{2}{g_{a9}^3} - \frac{1}{g_{a10}^3} - \frac{1}{g_{a11}^3} + \frac{2}{g_{a12}^3} \right) + \frac{b_c \ell}{2} \left(\frac{1}{g_{c4}^3} + \frac{1}{g_{c5}^3} - \frac{2}{g_{c6}^3} \right) \\
b_{\theta_x z} &= \frac{b_a \ell \sqrt{3}}{2} \left(\frac{1}{g_{a7}^3} - \frac{1}{g_{a8}^3} + \frac{1}{g_{a10}^3} - \frac{1}{g_{a11}^3} \right) + \frac{b_c \ell \sqrt{3}}{2} \left(-\frac{1}{g_{c4}^3} + \frac{1}{g_{c5}^3} \right) \\
b_{\theta_x \theta_x} &= \frac{3b_a \ell^2}{4} \left(\frac{1}{g_{a7}^3} + \frac{1}{g_{a8}^3} + \frac{1}{g_{a10}^3} + \frac{1}{g_{a11}^3} \right) + \frac{3b_c \ell^2}{4} \left(\frac{1}{g_{c4}^3} + \frac{1}{g_{c5}^3} \right) \\
b_{\theta_x \theta_y} &= \frac{b_a \ell^2 \sqrt{3}}{4} \left(-\frac{1}{g_{a7}^3} + \frac{1}{g_{a8}^3} - \frac{1}{g_{a10}^3} + \frac{1}{g_{a11}^3} \right) + \frac{b_c \ell^2 \sqrt{3}}{4} \left(-\frac{1}{g_{c4}^3} + \frac{1}{g_{c5}^3} \right) \\
b_{\theta_y z} &= \frac{b_a \ell}{2} \left(-\frac{1}{g_{a7}^3} - \frac{1}{g_{a8}^3} + \frac{2}{g_{a9}^3} - \frac{1}{g_{a10}^3} - \frac{1}{g_{a11}^3} + \frac{2}{g_{a12}^3} \right) + \frac{b_c \ell}{2} \left(\frac{1}{g_{c4}^3} + \frac{1}{g_{c5}^3} - \frac{2}{g_{c6}^3} \right) \\
b_{\theta_y \theta_x} &= \frac{b_a \ell^2 \sqrt{3}}{4} \left(-\frac{1}{g_{a7}^3} + \frac{1}{g_{a8}^3} - \frac{1}{g_{a10}^3} + \frac{1}{g_{a11}^3} \right) + \frac{b_c \ell^2 \sqrt{3}}{4} \left(-\frac{1}{g_{c4}^3} + \frac{1}{g_{c5}^3} \right) \\
b_{\theta_y \theta_y} &= \frac{b_a \ell^2}{4} \left(\frac{1}{g_{a7}^3} + \frac{1}{g_{a8}^3} + \frac{2}{g_{a9}^3} + \frac{1}{g_{a10}^3} + \frac{1}{g_{a11}^3} + \frac{2}{g_{a12}^3} \right) + \frac{b_c \ell^2}{4} \left(\frac{1}{g_{c4}^3} + \frac{1}{g_{c5}^3} + \frac{2}{g_{c6}^3} \right)
\end{aligned}$$

A.3 Combined Model

Finally, since the control system is designed for the entire stage at once, it is most convenient to combine the lateral and vertical damping terms into a single 6×6 matrix,

$$\mathbf{B} = \begin{bmatrix} b_{xx} & 0 & 0 & 0 & 0 & b_{x\theta_x} \\ 0 & b_{yy} & 0 & 0 & 0 & b_{y\theta_y} \\ 0 & 0 & b_{zz} & b_{z\theta_x} & b_{z\theta_y} & 0 \\ 0 & 0 & b_{\theta_x z} & b_{\theta_x \theta_x} & b_{\theta_x \theta_y} & 0 \\ 0 & 0 & b_{\theta_y z} & b_{\theta_y \theta_x} & b_{\theta_y \theta_y} & 0 \\ b_{\theta_z x} & b_{\theta_z y} & 0 & 0 & 0 & b_{\theta_z \theta_z} \end{bmatrix}, \quad (\text{A.1})$$

where each of the individual b_{ij} terms is a function of the instantaneous stage position.

Appendix B

Kinematics

B.1 Measurement Sensitivity Matrix

The measurement sensitivity matrix \mathbf{H} contains the coordinate transformation between the location and orientation of the platen in Cartesian coordinates \mathbf{X} and the measurements taken at the capacitance probe locations \mathbf{Y} . Therefore,

$$\mathbf{Y} = \mathbf{H}\mathbf{X} \tag{B.1}$$

where \mathbf{H} is defined as

$$\mathbf{H} = \begin{bmatrix} 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & -\ell \\ 1 & 0 & 0 & 0 & 0 & \ell \\ 0 & 0 & -1 & \frac{\ell\sqrt{3}}{2} & -\frac{\ell}{2} & 0 \\ 0 & 0 & -1 & -\frac{\ell\sqrt{3}}{2} & -\frac{\ell}{2} & 0 \\ 0 & 0 & -1 & 0 & L & 0 \end{bmatrix}. \tag{B.2}$$

Replacing ℓ by its numerical value of 0.0665 m yields the matrix

$$\mathbf{H} = \begin{bmatrix} 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & -0.0665 \\ 1 & 0 & 0 & 0 & 0 & 0.0665 \\ 0 & 0 & -1 & 0.0575907 & -0.03325 & 0 \\ 0 & 0 & -1 & -0.0575907 & -0.03325 & 0 \\ 0 & 0 & -1 & 0 & 0.0665 & 0 \end{bmatrix}.$$

For later reference, it is also helpful to define the inverse matrix which converts the capacitance probe readings into Cartesian coordinates. This is defined as

$$\mathbf{H}^{-1} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \\ 0 & 0 & 0 & \frac{\sqrt{3}}{3\ell} & -\frac{\sqrt{3}}{3\ell} & 0 \\ 0 & 0 & -\frac{1}{3\ell} & -\frac{1}{3\ell} & \frac{2}{3\ell} & \\ 0 & -\frac{1}{2\ell} & \frac{1}{2\ell} & 0 & 0 & 0 \end{bmatrix}, \quad (\text{B.3})$$

and again substituting 0.0665 m for ℓ yields

$$\mathbf{H}^{-1} = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.3333 & -0.3333 & -0.3333 \\ 0 & 0 & 0 & 8.6819589 & -8.6819589 & 0 \\ 0 & 0 & 0 & -5.0125313 & -5.0125313 & 10.025063 \\ 0 & -7.518797 & 7.518797 & 0 & 0 & 0 \end{bmatrix}.$$

for the value of the matrix inverse.

B.2 Modal Force Transformation

The modal force transformation converts forces and torques relative to the center of the platen into forces applied at each of the actuator pair locations:

$$\mathbf{F}_{em} = \Phi_F \mathbf{F}. \quad (\text{B.4})$$

In symbolic notation, the force transformation matrix is defined as

$$\Phi_F = \begin{bmatrix} 0 & 1/2 & 0 & 0 & 0 & -1/2\ell \\ 0 & 1/2 & 0 & 0 & 0 & 1/2\ell \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 1/\sqrt{3}\ell & -1/3\ell & 0 \\ 0 & 0 & 1/3 & -1/\sqrt{3}\ell & -1/3\ell & 0 \\ 0 & 0 & 1/3 & 0 & 2/3\ell & 0 \end{bmatrix}, \quad (\text{B.5})$$

where the modal force is

$$\mathbf{F} = \begin{bmatrix} F_x & F_y & F_z & T_x & T_y & T_z \end{bmatrix}^T, \quad (\text{B.6})$$

and the actuator forces are

$$\mathbf{F}_{em} = \begin{bmatrix} F_{1,5} & F_{2,4} & F_{3,6} & F_{10,7} & F_{11,8} & F_{12,9} \end{bmatrix}^T. \quad (\text{B.7})$$

Once again, it will be useful to obtain the numerical expression for this matrix. After substituting $\ell = 0.0665$ m, it evaluates to

$$\Phi_F = \begin{bmatrix} 0 & 0.5 & 0 & 0 & 0 & -7.518797 \\ 0 & 0.5 & 0 & 0 & 0 & 7.518797 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.3333 & 8.681960 & -5.012531 & 0 \\ 0 & 0 & 0.3333 & -8.681960 & -5.012531 & 0 \\ 0 & 0 & 0.3333 & 0 & 10.025062 & 0 \end{bmatrix},$$

It will sometimes be necessary to use the inverse matrix Φ_F^{-1} for calculating the modal forces from the actuator forces. Symbolically, the inverse matrix is

$$\Phi_F^{-1} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & \frac{\sqrt{3}}{2}\ell & -\frac{\sqrt{3}}{2}\ell & 0 \\ 0 & 0 & 0 & -\ell/2 & -\ell/2 & \ell \\ -\ell & \ell & 0 & 0 & 0 & 0 \end{bmatrix},$$

and again making the substitution for ℓ yields

$$\Phi_F^{-1} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0.0575907 & -0.0575907 & 0 \\ 0 & 0 & 0 & -0.03325 & -0.03325 & 0.0665 \\ -0.0665 & 0.0665 & 0 & 0 & 0 & 0 \end{bmatrix},$$

B.3 Gaps

B.3.1 Actuator Gaps

The value of the actuator gaps is required in order to estimate the damping of the plant and the force from an actuator. In this case, the relationships are easier to visualize by using the Cartesian coordinates that we have already calculated from the capacitance probe readings, rather than the cap probe readings themselves. From the geometry of the platen:

$$g_{a1} = g_{a1o} - y + \ell\theta_z$$

$$g_{a2} = g_{a2o} - y - \ell\theta_z$$

$$\begin{aligned}
g_{a3} &= g_{a3o} - x \\
g_{a4} &= g_{a4o} + y + \ell\theta_z \\
g_{a5} &= g_{a5o} + y - \ell\theta_z \\
g_{a6} &= g_{a6o} + x \\
g_{a7} &= g_{a7o} + z + \frac{\ell\sqrt{3}}{2}\theta_x - \frac{\ell}{2}\theta_y \\
g_{a8} &= g_{a8o} + z - \frac{\ell\sqrt{3}}{2}\theta_x - \frac{\ell}{2}\theta_y \\
g_{a9} &= g_{a9o} + z + \ell\theta_y \\
g_{a10} &= g_{a10o} - z - \frac{\ell\sqrt{3}}{2}\theta_x + \frac{\ell}{2}\theta_y \\
g_{a11} &= g_{a11o} - z + \frac{\ell\sqrt{3}}{2}\theta_x + \frac{\ell}{2}\theta_y \\
g_{a12} &= g_{a12o} - z - \ell\theta_y
\end{aligned}$$

The g_{aio} terms represent the null gaps for each actuator. Nominally, each of them equals $300 \mu\text{m}$

B.3.2 Capacitance Probe Gaps

The gaps of the capacitance probes are simply their measured values, plus the null offset, which nominally equals $100 \mu\text{m}$. If necessary, the measurement sensitivity matrix can be used to convert the location of the platen in Cartesian coordinates back into capacitance probe gaps.

Appendix C

Low Pass FIR Filter Coefficients

This section contains the numerical values for the FIR filter described in Section 5.3. A simpler way to reproduce these numbers is to use the MATLAB command

```
h = REMEZ(99,[0 0.0005 0.045 1],[1 1 0 0]);
```

rather than attempting to retype all of them in. The command yields an approximation to an ideal filter with a magnitude of 1 between the frequencies of 0 and 0.0005π radians/sample, and a magnitude of 0 between the frequencies of 0.045π radians/sample and π radians/sample. Since there is a “don’t care” band between the center frequencies, it is important to check the frequency response of the filter that this command produces. Since h is an FIR filter, it is simply implemented in the controller through

$$x_{filtered}(k) = \sum_{j=0}^{j=99} h_{lp}(j)x(k-j). \quad (C.1)$$

$h_{lp}[0] = 0.0011549078$	$h_{lp}[25] = 0.0098169719$
$h_{lp}[1] = 0.0005716288$	$h_{lp}[26] = 0.0104252048$
$h_{lp}[2] = 0.0007093250$	$h_{lp}[27] = 0.0110376895$
$h_{lp}[3] = 0.0008653610$	$h_{lp}[28] = 0.0116519388$
$h_{lp}[4] = 0.0010403220$	$h_{lp}[29] = 0.0122652871$
$h_{lp}[5] = 0.0012353583$	$h_{lp}[30] = 0.0128746322$
$h_{lp}[6] = 0.0014519558$	$h_{lp}[31] = 0.0134771824$
$h_{lp}[7] = 0.0016912762$	$h_{lp}[32] = 0.0140702285$
$h_{lp}[8] = 0.0019533010$	$h_{lp}[33] = 0.0146505806$
$h_{lp}[9] = 0.0022378499$	$h_{lp}[34] = 0.0152153307$
$h_{lp}[10] = 0.0025467457$	$h_{lp}[35] = 0.0157618816$
$h_{lp}[11] = 0.0028810899$	$h_{lp}[36] = 0.0162869353$
$h_{lp}[12] = 0.0032368910$	$h_{lp}[37] = 0.0167880893$
$h_{lp}[13] = 0.0036190379$	$h_{lp}[38] = 0.0172622703$
$h_{lp}[14] = 0.0040241903$	$h_{lp}[39] = 0.0177070447$
$h_{lp}[15] = 0.0044532577$	$h_{lp}[40] = 0.0181199716$
$h_{lp}[16] = 0.0049052337$	$h_{lp}[41] = 0.0184984974$
$h_{lp}[17] = 0.0053795055$	$h_{lp}[42] = 0.0188406906$
$h_{lp}[18] = 0.0058752510$	$h_{lp}[43] = 0.0191445945$
$h_{lp}[19] = 0.0063909357$	$h_{lp}[44] = 0.0194083619$
$h_{lp}[20] = 0.0069250658$	$h_{lp}[45] = 0.0196305954$
$h_{lp}[21] = 0.0074766352$	$h_{lp}[46] = 0.0198099982$
$h_{lp}[22] = 0.0080436699$	$h_{lp}[47] = 0.0199454215$
$h_{lp}[23] = 0.0086237258$	$h_{lp}[48] = 0.0200362737$
$h_{lp}[24] = 0.0092159149$	$h_{lp}[49] = 0.0200817318$

$h_{lp}[50]$	$=$	0.0200817318	$h_{lp}[75]$	$=$	0.0092159149
$h_{lp}[51]$	$=$	0.0200362737	$h_{lp}[76]$	$=$	0.0086237258
$h_{lp}[52]$	$=$	0.0199454215	$h_{lp}[77]$	$=$	0.0080436699
$h_{lp}[53]$	$=$	0.0198099982	$h_{lp}[78]$	$=$	0.0074766352
$h_{lp}[54]$	$=$	0.0196305954	$h_{lp}[79]$	$=$	0.0069250658
$h_{lp}[55]$	$=$	0.0194083619	$h_{lp}[80]$	$=$	0.0063909357
$h_{lp}[56]$	$=$	0.0191445945	$h_{lp}[81]$	$=$	0.0058752510
$h_{lp}[57]$	$=$	0.0188406906	$h_{lp}[82]$	$=$	0.0053795055
$h_{lp}[58]$	$=$	0.0184984974	$h_{lp}[83]$	$=$	0.0049052337
$h_{lp}[59]$	$=$	0.0181199716	$h_{lp}[84]$	$=$	0.0044532577
$h_{lp}[60]$	$=$	0.0177070447	$h_{lp}[85]$	$=$	0.0040241903
$h_{lp}[61]$	$=$	0.0172622703	$h_{lp}[86]$	$=$	0.0036190379
$h_{lp}[62]$	$=$	0.0167880893	$h_{lp}[87]$	$=$	0.0032368910
$h_{lp}[63]$	$=$	0.0162869353	$h_{lp}[88]$	$=$	0.0028810899
$h_{lp}[64]$	$=$	0.0157618816	$h_{lp}[89]$	$=$	0.0025467457
$h_{lp}[65]$	$=$	0.0152153307	$h_{lp}[90]$	$=$	0.0022378499
$h_{lp}[66]$	$=$	0.0146505806	$h_{lp}[91]$	$=$	0.0019533010
$h_{lp}[67]$	$=$	0.0140702285	$h_{lp}[92]$	$=$	0.0016912762
$h_{lp}[68]$	$=$	0.0134771824	$h_{lp}[93]$	$=$	0.0014519558
$h_{lp}[69]$	$=$	0.0128746322	$h_{lp}[94]$	$=$	0.0012353583
$h_{lp}[70]$	$=$	0.0122652871	$h_{lp}[95]$	$=$	0.0010403220
$h_{lp}[71]$	$=$	0.0116519388	$h_{lp}[96]$	$=$	0.0008653610
$h_{lp}[72]$	$=$	0.0110376895	$h_{lp}[97]$	$=$	0.0007093250
$h_{lp}[73]$	$=$	0.0104252048	$h_{lp}[98]$	$=$	0.0005716288
$h_{lp}[74]$	$=$	0.0098169719	$h_{lp}[99]$	$=$	0.0011549078

Appendix D

Estimator Gain Setting

The following MATLAB program is used to set the value of the estimator gain matrix. It places a call to a procedure `Damping.m` that returns the value of the damping matrix for a given set of actuator and capacitance probe gaps.

```
% BIAS_ESTIM.M

% set the sampling time and the gaps

T = 350/156250;

ga = [0.000266; 0.000308; 0.000318; 0.000279; 0.000299; 0.000305; ...
      0.0003; 0.000344; 0.0003; 0.000278; 0.0003; 0.000294];

gc = [0.000080; 0.000094; 0.000096; 0.0001; 0.0001; 0.0001];

% Compute the required matrices

B = Damping(ga,gc);
Fw = zeros(12,12);
```

```

Fw(1:6,1:6) = eye(6,6);
Fw(7:12,7:12) = eye(6,6);
Fw(1:6,7:12) = 0.00448.*inv(B);
H = [0 -1 0 0 0 0; 1 0 0 0 0 -0.0665; 1 0 0 0 0 0.0665; ...
      0 0 -1 0.057590689 -0.03325 0; 0 0 -1 -0.057590689 -0.03325 0; ...
      0 0 -1 0 0.0665 0];
Hw = zeros(6,12);
Hw(1:6,1:6) = H;

% Set the desired convergence rates

PB = 10;    % Position Estimate Convergence in Hz
AB = 10;    % Angle Convergence in Hz
FB = 0.1;   % Bias Force Estimate Convergence in Hz

PB=-2*pi*PB; FB=-2*pi*FB; AB=-2*pi*AB; % Convert to radians in the s-plane
PB = exp(T*PB); FB = exp(T*FB); AB = exp(T*AB); % Map to the z-plane

P = PB*ones(1,12);
P(4:6) = AB*ones(1,3);
P(7:12) = FB*ones(1,6);

K = place(Fw', Hw', P);
K'
```


Appendix E

Derivation of Discrete Time Noise Propagation Model

This chapter of the appendix contains a derivation of noise propagation through a discrete time linear system. The first parameter used to describe a discrete time random waveform x is its *expected value* $\langle x(k) \rangle$ defined as

$$\langle x(k) \rangle \equiv \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x(k+n). \quad (\text{E.1})$$

The expected value of a waveform x is its statistical mean μ_x . The concept of the expected value is crucial to describing stochastic processes. Two of its properties that we will make use of later are *time invariance* and *linearity*. Time invariance simply states that

$$\langle x(k+T) \rangle = \langle x(k) \rangle, \quad (\text{E.2})$$

which comes from the fact that the expected value is defined as a sum over infinite limits. Therefore, no starting position is different than any other. Linearity allows us to multiply a waveform by a constant and add it to another waveform according to

$$\langle Ax(k) + By(k) \rangle = A \langle x(k) \rangle + B \langle y(k) \rangle. \quad (\text{E.3})$$

Put into words, the expected value of a constant times a waveform is that constant times the expected value of the waveform, and the expected value of the sum of two waveforms is simply the sum of their expected values.

The rules of linearity and time invariance are a necessary beginning, but they limit us to extremely simple transformations of the waveform. Simply knowing $\langle x(k) \rangle$ is not enough even to calculate $\langle x^2(k) \rangle$. Thinking back to statistics, the next piece of information that we might want to know about a waveform is a measure of its spread about the mean value. This is the *variance* of a waveform, defined as

$$\sigma_x^2 = \langle (x - \mu_x)^2 \rangle. \quad (\text{E.4})$$

However, it becomes useful to define a more general measure of deviation when discussing random waveforms. If we take a waveform and convolve it with itself, we generate the *autocorrelation* sequence, defined as

$$R_x(\tau) \equiv \langle x(k)x(k + \tau) \rangle. \quad (\text{E.5})$$

The value of the autocorrelation sequence at $\tau = 0$ is the *mean square value* of the sequence, $\langle x^2(k) \rangle$. If $x(k)$ is a zero mean sequence, as is often the case in control system design, then the mean square value is exactly the variance of the sequence.

One of the useful characteristics of the autocorrelation sequence is its ability to display hidden periodicities in a data set. If a data set has a component that is periodic on n points, then a plot of its autocorrelation will show a peak at every n points. More interesting for our application is defining the way that autocorrelation functions are altered when the waveform is passed through an LTI system. The derivation given in Franklin, Powell, and Workman [15] is most closely followed here.

Suppose that $y(k)$ is the sequence that results when $x(k)$ is passed through the LTI system having an impulse response $h(k)$. The correlation sum yields the output as

$$y(j) = \sum_{k=-\infty}^{\infty} h(k)x(j - k). \quad (\text{E.6})$$

Now substituting this relationship into the definition of the autocorrelation of $y(k)$,

$$R_y(\ell) \equiv \langle x(k)x(k+\ell) \rangle, \quad (\text{E.7})$$

yields

$$R_y(\ell) = \left\langle \left(\sum_{n=-\infty}^{\infty} h(n)x(j-n) \right) \left(\sum_{k=-\infty}^{\infty} h(k)x(j+\ell-k) \right) \right\rangle. \quad (\text{E.8})$$

Next, note that the impulse response of the system is known, and so can be removed from the expected value calculation,

$$R_y(\ell) = \sum_{n=-\infty}^{\infty} h(n) \sum_{k=-\infty}^{\infty} h(k) \langle x(j-n)x(j+\ell-k) \rangle. \quad (\text{E.9})$$

The critical step in the derivation requires the insight that by using the property of time invariance, the expected value is equivalent to an autocorrelation of x : $R_x(\ell - k + n)$.

At this point we have a relationship between the autocorrelations of the input and output sequences,

$$R_y(\ell) = \sum_{n=-\infty}^{\infty} h(n) \sum_{k=-\infty}^{\infty} h(k) R_x(\ell - k + n). \quad (\text{E.10})$$

The real usefulness can only be seen by taking its z -transform,

$$Z \{R_y(\ell)\} = \sum_{\ell=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} h(n) \sum_{k=-\infty}^{\infty} h(k) R_x(\ell - k + n) z^{-\ell}. \quad (\text{E.11})$$

Defining a new variable, $m = \ell - k + n$, and rearranging the summations yields,

$$Z \{R_y(\ell)\} = \sum_{n=-\infty}^{\infty} h(n) \sum_{k=-\infty}^{\infty} h(k) \sum_{m=-\infty}^{\infty} R_x(m) z^{-(m+k-n)}. \quad (\text{E.12})$$

Now, we can distribute the exponentials in z to the corresponding sums,

$$Z \{R_y(\ell)\} = \sum_{n=-\infty}^{\infty} h(n) z^n \sum_{k=-\infty}^{\infty} h(k) z^{-k} \sum_{m=-\infty}^{\infty} R_x(m) z^{-m}. \quad (\text{E.13})$$

By definition of the z -transform then, the equation reduces to

$$Z\{R_y(\ell)\} = H(z^{-1})H(z)Z\{R_x(\ell)\}, \quad (\text{E.14})$$

which gives us a clean relationship between the z -transforms of the autocorrelation functions of the input and output sequences. The z -transform of the autocorrelation sequence is so important that it is redefined as the *spectrum* of the sequence, and given the symbol S . With this, eq.(E.14) becomes

$$S_y(z) = H(z)H(z^{-1})S_x(z). \quad (\text{E.15})$$

Finally, we can generate the discrete time Fourier transform of the relationship by evaluating z on the unit circle and thereby replacing z by $e^{j\omega}$. Assuming that $H(z)$ has a rational system function, $H(e^{j\omega})$ is symmetric about $\omega = 0$, and the relationship reduces to its final form of

$$\boxed{S_y(e^{j\omega}) = |H(e^{j\omega})|^2 S_x(e^{j\omega})}. \quad (\text{E.16})$$

This is a tremendously important relationship. The Fourier transform of the autocorrelation sequence is called the *power spectral density* of the waveform. The area under the spectral density between any two frequencies represents the power that would remain in the signal if it were passed through an ideal bandpass filter with those cutoff frequencies. Most importantly for applications, we can measure the power spectral density of a noise signal. If we then model the frequency response of a system, we can predict the spectral density of the output waveform and show how noise is amplified or attenuated.

The spectral density of a waveform conveys a great deal of information about its frequency content. However, the final effectiveness of a filter or control system is likely to be judged in the time domain. Therefore, we would like to be able predict the variance of a waveform from its spectral density. Recall that the spectral density is the Fourier transform of the autocorrelation function. Taking the inverse Fourier transform of our

final spectral density will then return us to the time domain,

$$R_y(\ell) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_y(e^{j\omega}) e^{j\omega\ell} d\omega. \quad (\text{E.17})$$

The entire autocorrelation function is not necessarily useful though, only $R_y(0)$ which gives the mean squared value of a waveform. This reduces the inverse Fourier transform to a simple integral of the spectral density,

$$R_y(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_y(e^{j\omega}) d\omega. \quad (\text{E.18})$$

The mean squared value of a waveform is the sum of its variance and squared mean,

$$R_y(0) = \langle y^2(k) \rangle = \sigma_y^2 + \mu_y^2, \quad (\text{E.19})$$

and so when dealing with zero mean signals, which is generally the case in control systems, the variance of a waveform will simply be the area underneath the power spectral density curve. Using these relationships, we will now examine how the different noise sources propagate through the Angstrom Stage, look at the effects of changes in gain (assuming a proportional controller), and predict the final position variance.

Appendix F

Closing Thoughts

This chapter contains some additional thoughts that really do not fit anywhere else. Knowing them may save some time for the next individual working with this hardware.

- One of the slowest segments of the controls code is the portion of `cim.c` that transfers the capacitance probe readings from the PC memory into dual port memory for access by the C30. Two things helped to speed this code segment up significantly. First was writing this section in assembly language. The other was changing the output of the A/D from two's complement notation to offset binary. In other words, the A/D has an option of outputting numbers ranging from either -32768 to 32767 or from 0 to 65535. The default when measuring bipolar signals is to use twos complement notation. In that way, a single constant multiplying the digital word will return the input voltage. However, the C30 only recognizes 32 bit words. Therefore, whenever negative sixteen bit words are written to 32 bit memory locations in the dual port memory the upper sixteen bits have to be filled with a sign extension of the lower sixteen bits. This effectively doubles the number of words that have to be written by the 486. If the A/D converter instead returns all positive values, we can initialize all of the dual port memory locations to zero, and then only write the to the lower sixteen bits. After filtering on the C30, we can just subtract 32768 from the six remaining capacitance probe readings.

- One of the actuators (actuator 11) is shorted to the frame of the stage. It operates the same as all of the other actuators, but requires a more specialized amplifier.
- Use caution when writing inline assembly code for the C30. The C compiler completely skips any inline segments, and inserts them verbatim into the compiled assembly code. Usually, this will not pose a problem. However, I ran into the situation I set $x = 0$, operated on x in an inline assembly routine, and then set $y = x$. Logically, this works fine. However, the compiler only saw $x = 0$, $y = x$, and so wrote the code as $x = 0$, $y = 0$. I got around the problem by initializing x in the assembly routine, but the best option is to skip the inline assembly completely, and instead write the assembly portions as separate functions that are called by the main C routine.
- The dual port memory is truly “dual port” in name only. Reading the fine print in the users manual reveals that the identical memory location cannot be read by both sides simultaneously. Originally, the code was structured such that this was possible. When the 486 was writing new cap probe readings to the dual port memory, the C30 could also be reading them. If they both accessed the same location at the same time, that location was filled with a large number. Therefore, if the stage was steady near a zero position, then a large number thrown into the readings led to a fictitious “spike” in position, and a resulting DAC output that did actually change the position. The best policy was to treat the dual port memory as single port, and to be absolutely certain that code was structured such that only one processor was accessing it at a time.