NANOVIBRATION CONTROL

by

RUSS GREENALL

B.Sc., Simon Fraser University, 2001

A THESIS SUBMITTED IN PARTIAL FULLFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

In

The Faculty of Graduate Studies

(Department of Physics)

University of British Columbia

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

June 2004

Library Authorization

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Russ Greenall — 19/07/2004

Name of Author *(please print)* — Date (dd/mm/yyyy)

Title of Thesis: Nanovibration Control

Degree: Master's of Physics Year: 2004

Department of Physics

The University of British Columbia

Vancouver, BC Canada

Abstract

This thesis explores techniques to actively control the position of large masses such as focusing magnets with precision on the order of 1 nm against vibrations. The technique applied (labeled as an "optical anchor") is to actively "stiffen" the support structure using an optical interference method to measure distance to a remote reference point. The magnet is modeled as a mass on a spring, with a piezo electric actuator. In this model, proportional and differential control applied to the piezo allows the mass to be critically damped and the spring coefficient to be arbitrarily increased. A digital implementation with finite sampling rate has a finite stable region in control parameter space. If there are more mechanical degrees of freedom, the stable region and the quality of control can be greatly reduced.

An interferometric instrument design for remote distance measurement is discussed and measurement results reflecting an accuracy of 0.2nm RMS are demonstrated. The instrument requires only two light detectors in a Michelson interferometer configuration. The algorithm design is implemented at a 5KHz sample rate using a circa 2000 DSP processor with 4-byte floating point operations running at a 40 MHz clock rate.

Control tests on a one degree-of-freedom experimental platform are performed using proportional and differential control. These tests demonstrate active control which significantly damps fundamental mode excitations but are insufficient to stiffen the system. More sophisticated models and algorithms will be necessary. Nevertheless, some insight is gained into techniques which will allow control on the nanometer scale against "standard" ground vibrations. In particular, a successful implementation of coherent ground disturbance modeling provides a three-fold reduction in RMS vibration of our test system over our simple PID control.

Table of Contents

ABSTRACT ii

TABLE OF CONTENTS iii

TABLE OF FIGURES v

1. INTRODUCTION 1

2. CONTROL THEORY FOR PHYSICISTS 3

2.1. Simple Continuous Model 3
2.2. Continuous Control 5
2.3. Bode Plots and Transfer Functions 9
2.4. Stability in Continuous Feedback Systems 13
2.5. Discrete Time 16
2.6. Discrete Control 19
2.7. Control of Two-Resonance System in Discrete Time 23
2.8. Conclusion 27

3. APPARATUS 29

3.1. Interferometer 29
3.2. Electronics 30
3.3. Data Acquisition 31
3.4. Lock-in Amplifier 32
3.5. Position Control Test Platform 33

4. INTERFEROMETER 37

4.1. Interferometer Signal Model 37
4.1.1. Physics *37*
4.1.2. Direct Parameterization *39*
4.1.3. Calibration *40*
4.1.4. Additional Algorithm Design Criteria *41*
4.2. Geometric Parameterization 43
4.2.1. Geometric Parameterization Validity *44*
4.2.2. Parameter Iterations in Geometric Model *47*
4.2.3. Algorithm Tests *56*
4.2.4. Non-Correlated Noise Test *58*

5. CONTROL TESTS 60

5.1. Platform Response 60
5.2. Ground-Motion Characteristics 62
5.3. PID Control Model 66
5.4. Results 67
5.5. Exploration of Gain Limits 70
5.6. PID Control with Resonance Notching 72
5.7. Coherent Disturbance Correction 73
5.8. Best Composite Performance 76

6. CONCLUSIONS 77

APPENDIX A NLC VIBRATION CONTROL EDITORIAL 79

APPENDIX B DEADBEAT GAIN 82

APPENDIX C NOTCH FILTER ALGORITHM 84

INDEX 88

Table of Figures

Figure 1: Simplified Model of a Single Mass with Position Control 3
Figure 2: Sample Evolution of the Simplified Model 4
Figure 3: Evolution of "Stiffened" System Kp=10 6
Figure 4: Damped System 7
Figure 5: Simple Mass on a Spring Viewed as a Transfer Function 9
Figure 6: Bode Plot for Simple Mass on Spring 10
Figure 7: Mass on a Spring Control Schematic 11
Figure 8: Bode Plot for Differentially Controlled Mass 12
Figure 9: Unstable System 14
Figure 10: Two Resonance System 15
Figure 11: Discrete Evolution Sample of the Simplified Model 18
Figure 12: Gain Stability Contours 20
Figure 13: Simplified Model with PD Gains Set at 50% of Deadbeat 21
Figure 14: Simplified Model with PD Gains Set at 100% of Deadbeat 22
Figure 15: Modified Model Adding a Single Resonance 23
Figure 16: Two-Resonance Model Gain Stability 24
Figure 17: Modified Model at Previously Stable Gains 25
Figure 18: Two-Resonance Model for 4-Dimensional Gain Stability 26
Figure 19: Interferometer Setup 29
Figure 20: Piezo Driver Circuit Schematic 31
Figure 21: Lock-in Amplifier Block Schematic 32
Figure 22: Test Platform Isometric Drawing 33
Figure 23: Test Platform Photograph 34
Figure 24: Actuator Detail 35
Figure 25: Interferometer Setup 37
Figure 26: Characteristic Interferometric Measurement Data 39
Figure 27: Table of System Parameters 40
Figure 28: Lissajous Pattern 43
Figure 29: Symmetric Matrix Operation 48
Figure 30: Circularization 50
Figure 31: First Order Iteration to Find Centre 51
Figure 32: Orthogonal Iteration to Find Centre 52
Figure 33: Centre Finding Operation 54
Figure 34: Interferometer and Algorithm Test 57
Figure 35: Non-correlated Noise Test 58
Figure 36: System Bode Plot 61
Figure 37: Uncontrolled System Vibration Spectrum 62
Figure 38: Measurement Model (No Control) 63
Figure 39: Ground Spectrum Approximation 65

Figure 40: Control System Model 66
Figure 41: Model Response 68
Figure 42: PID Control System Results 69
Figure 43: PID Control System Bode Plots 70
Figure 44: Gain Increase due to Notch Filters 72
Figure 45: Coherent Disturbance Elimination 74
Figure 46: Best Control Achieved 76

1. Introduction

Particle accelerators have been increasing in power since their inception in the 1930's. Since the advent of two-beam particle colliders in the 1950's, beam diameter has been decreasing to increase luminosity. The proposed "Next Linear Collider" project proposes a beam diameter as small as 2 X 200 nanometers[1] or just about twenty times the diameter of an atom in the small dimension. At this diameter, stabilization of final focusing magnets against environmental vibration becomes critical. Acceptable levels of stabilization will be on the order of one nanometer.

Some concern has been raised regarding the magnitude of the ground motion which will be encountered in the NLC environment[2],[3],[4]. These sources indicate an overall expected vibration level on the order of 100nm RMS with no effective vibration above 300Hz. Further, narrow band near-field excitations from rotating machinery will cause narrow band peaks of up to hundreds of nanometers in amplitude at frequencies anywhere from 30 Hz to 200 Hz[4].

J. Frisch et al. [1] discuss stabilization technologies with respect to the NLC project as of 2001. These are broken down into three categories:

1) **Beam Based Systems**. The interaction of the electron and positron beams at the IP (intersection point) causes a beam deflection that is related to the beam offset. This allows the offset to be measured to a fraction of the spot size at the beam rate of 120Hz. This beam deflection provides the only long-term measure of the relative positions of the beams. A variety of feedback algorithms can be used with the beam – beam deflection data, with the selection based on the trade off between low frequency attenuation, and high frequency amplification of noise.

2) **Interferometer Based Systems**. Optical interferometers can be used to measure the distance between the final focus magnets and an external reference which may or may not be rigidly attached to the ground. These are termed "optical anchor" systems.

3) **Inertial Based Systems**. Inertial sensors can be used to measure the motion of the final quadrupoles relative to the "fixed stars". At low frequencies the position noise of an inertial sensor increases and a transition to a beam based system must be made. An inertial vibration stabilization system has been constructed and is being used to test feedback hardware and algorithms [1].

This thesis is dedicated to investigating two aspects of: **2) Interferometer Based Systems**:

1. Instrument Design for an accurate and relatively inexpensive interferometric distance measurement device. This is explored in chapter 4, (page 37).

2. Active control of one dimension of a large mass to nanometer precision using the paradigm of "ground tracking" or "optical anchor". First principles analysis would imply that position control of a large mass should be straightforward (refer to the theory chapter section 2.1: "Simple Continuous Model" on page 3). Actual implementation proved very elusive however.

   The paradigm of the "optical anchor" requires that the large mass must be accelerated to undergo a trajectory which matches the ground-based reference. This is an extremely difficult proposition as accelerations of this order invariably excite dozens of internal modes of the large mass and/or of the support structure which are NOT negligible. This is explored in chapter 5, "Control Tests" on starting page 60.

2. Control Theory for Physicists

2.1. Simple Continuous Model

We shall start with the simplest conceivable model for nano-position control. Reduced to simplest terms of one degree of freedom and a monolithic mass the system is as shown in the figure below.

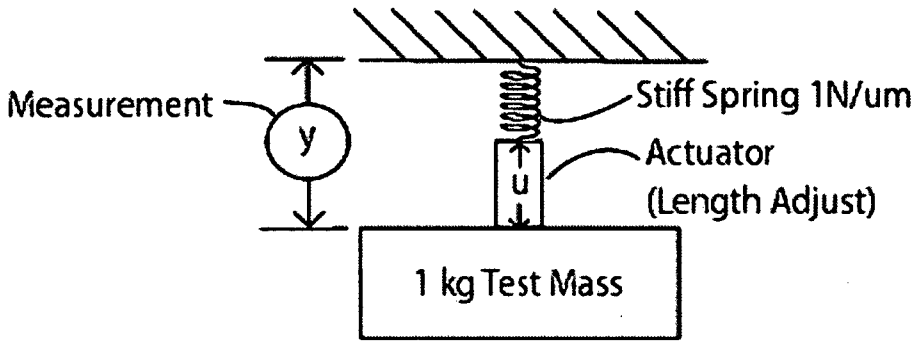

Figure 1: Simplified Model of a Single Mass with Position Control

Let us carry through a numeric example allowing illustration of the mixed magnitudes. That is, perform a numeric example where our mass is considered to be 1kg and yet we are interested in nanometer motions.

We shall ignore damping, because natural damping has a negligible effect on system gain limitations and it adds a term to our equations with which we need not deal. Similarly, gravity is ignored in this model. We just have a mass with one degree of freedom mounted to a bulkhead through a stiff spring and an actuator which can change its length under a control system's command.

The system has no damping (infinite "Q"), thus the differential equation of motion is as follows from a simple force balance (y measured from equilibrium):

$$0 = m\frac{d^2y(t)}{dt^2} + ky(t) - ku(t) \qquad \text{Eqn(1)}$$

Where m is the mass, $y(t)$ is the test mass position from equilibrium, $u(t)$ is the actuator length (measured from null) and k is the spring stiffness.

Note that $u(t)$ is an arbitrary function which we can specify. The resultant position function from the combination of natural evolution and the control function is $y(t)$.

Since we take the damping in this system to be negligible, this system will be on the border of stability without active feedback. Once set oscillating, this system as described will continue to oscillate forever.

Below is a plot of the system evolution generated using numerical methods. The plot shows an initial small oscillation of 0.1 nm amplitude. At time t=0, the actuator is stepped from zero to one nm. The resultant system response is shown.

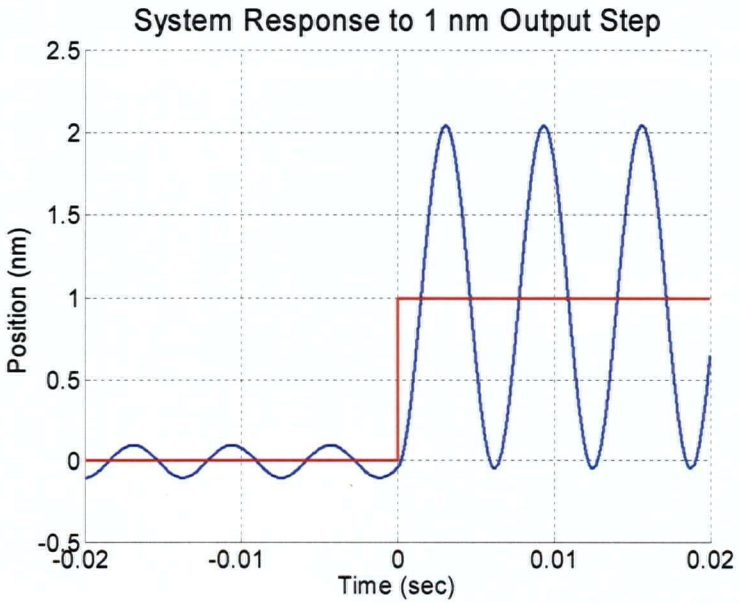

Figure 2: Sample Evolution of the Simplified Model

The actuator is shown in red. The resulting system response is shown in blue.

Overall, the second order response is very simple to understand. One can easily see that adjustment of the actuator length *u(t)* has a strong coupling to the position and should be capable of very effective control of the system.

2.2. Continuous Control

As mentioned, the function *u(t)* is under our control. It is there to influence the behavior of the system. Indeed, in the continuous (and perfectly modeled) world, the output *u(t)* can be set as a function of position and velocity to damp position oscillations arbitrarily fast. There is a host of engineering literature related to control of simple continuous second order systems. We are going to examine continuous control from the point of view of physical equivalence. That is, the control function can be made to artificially "stiffen" and "damp" the system. Manipulation of the system evolution sets this analysis apart from many mechanical analyses where one is merely trying to predict system outcomes. Here, we are purposefully interfering with the natural evolution of our system.

There are two general classes of control: feedforward and feedback. In feedforward, we adjust *u(t)* in some arbitrary fashion without regard to system measurements. This mode might be applied if, for example, we wanted to cause the system to oscillate at some fixed frequency. We would simply apply this pure frequency to the actuator (*u(t)*) and know that the system response would be to respond at exactly the same frequency. We would not offhand know the phase or amplitude.

Feedback control is far more common. In feedback, we examine the system measurements and use these measurements to calculate an appropriate actuator length (*u(t)*). This mode of control is almost universally implied in control theory.

So, what are some obvious feedback strategies? For one, there is the "proportional" control strategy. If the measurement shows the position too low, retract the actuator, which will apply a restoring force to raise the mass. Similarly, if the position is too high, extend the actuator.

In proportional control, the "error" is taken to be the difference between an actual measurement and a setpoint or desired point for that measurement.. The actuator is then set in opposition to the error by an amount proportional to the error.

$$u(t) = -K_p y(t) + r(t) \qquad \text{Eqn(2)}$$

K_p is called the proportional gain. The negative sign is included since by convention, the state of negative feedback is considered the norm and K_p is expected to be positive. The term *r(t)*

is included here as the step function shown in red on all of the plots. It is similar in function to what would normally be referred to as the setpoint or demand function in control literature.

The composite system equation when *u(t)* is set using proportional feedback is shown below. Substituting the proportional control condition equation (2) into the system equation (Eqn 1) yields:

$$0 = m\frac{d^2y(t)}{dt^2} + k(1+K_p)y(t) - kr(t) \qquad \text{Eqn(3)}$$

Examining equation 3 above, we can see that the effective spring constant for the system is now $k(1+K_p)$. With a second order system such as we have modeled, proportional control only serves to "stiffen" the system. That is, the system now has a higher resonant response than the natural system but there is still no damping. This can be seen in figure 3 where *Kp* is stepped from 0 to 10 at time zero.

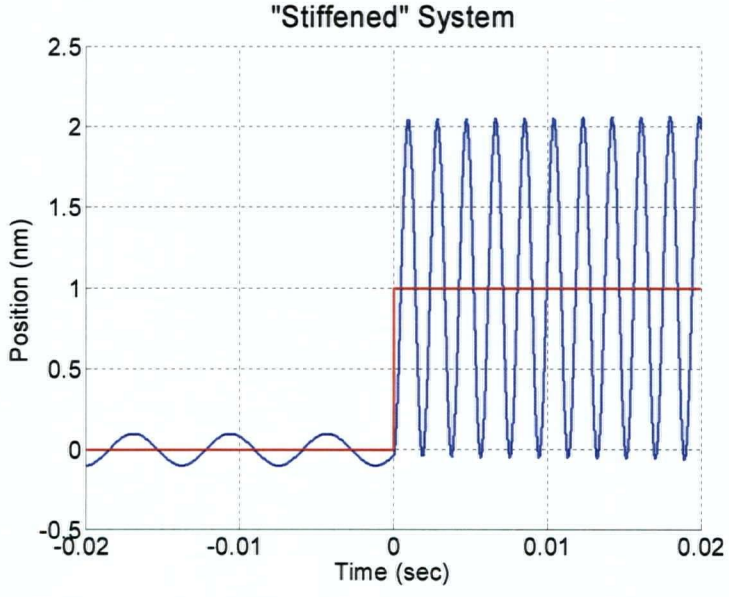

Figure 3: Evolution of "Stiffened" System Kp=10

The red step function shows the demand or setpoint function and the system response is shown in blue.

Note that in this resultant system, what is known as positive feedback would result from a negative K_p (actuator action which does not oppose the error). Positive feedback up to a point may be applied to "soften" the system. As K_p approaches -1 the system formula becomes that

of an unconstrained mass. This positive feedback would be limited to K_p = -1 at which point the natural spring would be overcome by the control and the system would become unstable.

Another feedback strategy is called "derivative" control where the actuator length is set proportionally to the velocity of the mass. When set in a contrary direction to the velocity, the control serves to dampen the system oscillations. This strategy for u(t) is defined as follows

$$u(t) = -K_d \frac{dy(t)}{dt} + r(t) \qquad \text{Eqn(4)}$$

This relation establishes derivative control where K_d is defined as the derivative gain. This leads to a system equation as follows:

$$0 = m\frac{d^2y(t)}{dt^2} + kK_d\frac{dy(t)}{dt} - kr(t) + ky(t) \qquad \text{Eqn(5)}$$

An example of this control is shown below where the derivative gain factor is set to K_d=0.001 sec at t=0. The system now appears damped despite the absence of any natural damping.

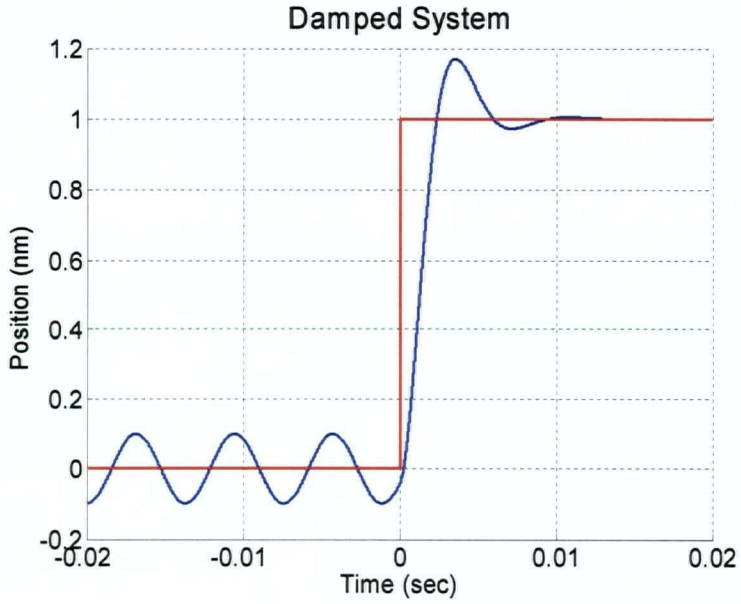

Figure 4: Damped System

If we include a combination of both proportional and derivative control, our new equation of motion becomes:

$$0 = m\frac{d^2y(t)}{dt^2} + K_d\frac{dy(t)}{dt} + k(1+K_p)y(t) - kr(t) \qquad \text{Eqn(6)}$$

This standard form of damped simple harmonic motion has a well known condition for critical damping about the equilibrium point which is:

$$K_d = \sqrt{4k(1+K_p)} \qquad \text{Eqn(7)}$$

In our previous figure (4) showing a damped response, K_d was set to 0.001 with no proportional gain. Critical damping according to the relation above would be at K_d=0.002. This is consistent with figure (4), which shows significant damping but is still somewhat under-damped.

For the mass and spring system described, we could make the proportional gain arbitrarily large and the derivative gain critically damped and achieve arbitrarily good control of the mass despite the finite natural frequency. However, to raise the effective frequency by a factor of 100, we would need K_p=9999.

This is not to say that a real system can be controlled infinitely well by applying infinite feedback gain. We shall see that the system can become unstable at large proportional or derivative gains due to additional system modes, or to finite time step effects in digital implementations.

Another form of feedback called integral control is used to remove all steady-state error from control system output. However our application does not specify any steady-state (DC) requirements and therefore we will not cover this type of control here.

Together, proportional, integral and derivative control are known as PID control and are used extensively both industrially and in the laboratory as an excellent generalized first approach to a control problem.

2.3. Bode Plots and Transfer Functions

It is useful to analyze systems in terms of transfer functions. A transfer function is simply the amplitude and phase relationship between the input and output of a system. For example, in our simple system, we can regard the control signal u as an input and the resultant position measurement y as an output as follows:

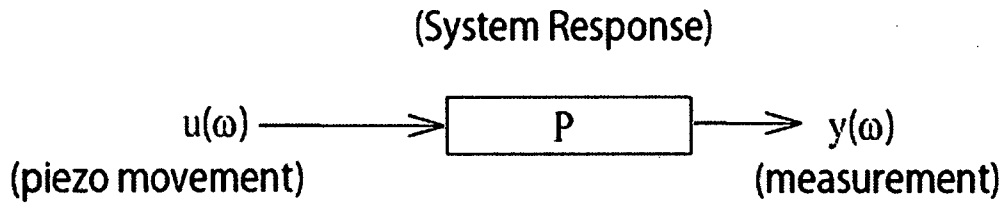

Figure 5: Simple Mass on a Spring Viewed as a Transfer Function

The simple mass on a spring can be regarded as a transfer function P which relates the amplitude and phase relationship between the input u(ω) and the output y(ω)

The inputs and outputs are Fourier transformed to be complex functions of frequency rather than real functions of time. This is why the diagram shows the funtions $u(\omega)$ and $y(\omega)$ as functions of frequency ω. The system response (often referred to as the "Plant" response) is characterized by the transfer function P where

$$y(\omega) = P(\omega)u(\omega) \qquad \text{Eqn(8)}$$

And where the explicit references to the frequencies are invariably omitted. The convention maintained hereafter will be that transfer functions and the associated functions of frequency will be written as non-italic, non-bold typeface. Where this is not definitive, context will hopefully be sufficient for clarity:

$$y=Pu \qquad \text{Eqn(9)}$$

Notwithstanding the explicit omission of a reference to frequency, transfer functions are assumed to apply to Fourier space exclusively. Application of a signal to a transfer function can be explicitly achieved simply through complex multiplication at each frequency.

The presentation of a transfer function can be made through use of a Bode plot which shows amplitude and phase response in two separate graphs which share a frequency axis. The Bode plot for our simple mass on a spring is shown below.

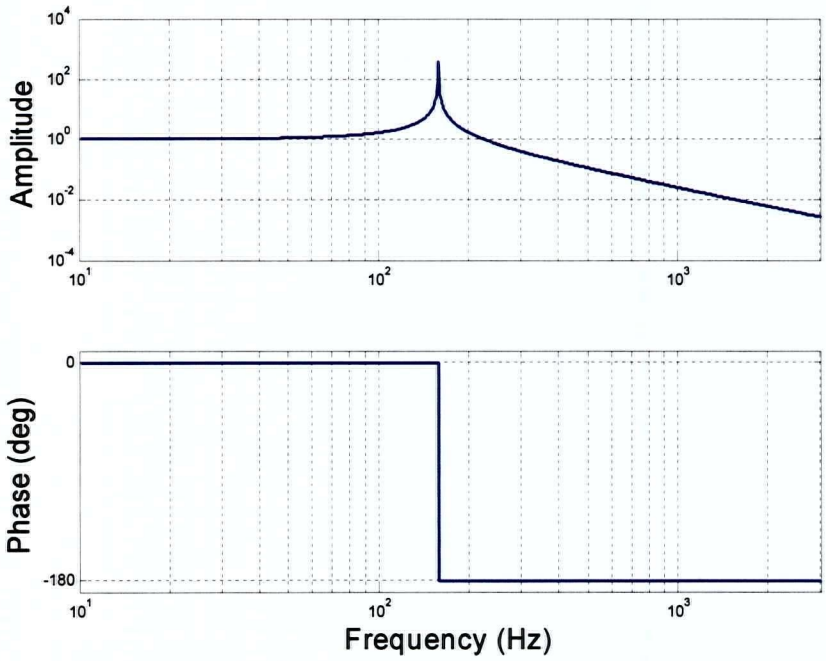

Figure 6: Bode Plot for Simple Mass on Spring

The amplitude actually rises to infinity for the undamped resonance at about 160 Hz but quantization of the plot produces this approximation of the actual response.

At very low frequencies, the mass moves in the same manner as the piezo extension. As the frequency approaches resonance, the response gets larger reaching an infinite result exactly at resonance (for our perfect undamped system). After resonance the response falls off with increasing frequency. The phase is unchanged at zero degrees until after resonance where it instantly (for an undamped resonance) switches to 180 degrees of lag.

Note that the Bode plot does not directly show transient responses but rather shows the steady state output to input relationship only.

When *u(t)* is made to be a function of the measurement *y(t)* that process is said to be feedback. The configuration is shown schematically below.

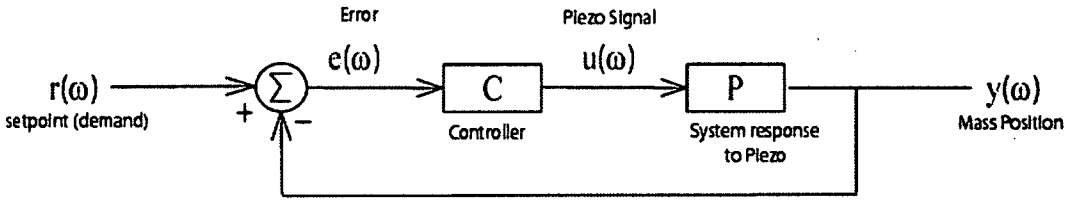

Figure 7: Mass on a Spring Control Schematic

Figure 7 requires the system output y to depend upon the plant transfer function P, the controller transfer function C and implicitly on itself through the difference node output (r-y). The resulting system equation is:

$$y = (r - y)CP \qquad \text{Eqn(10)}$$

Where all of the symbols are complex functions of frequency and the equation holds for any given value of frequency.

This can be manipulated to lead to an explicit expression for the closed-loop transfer function.

$$\frac{y}{r} = \frac{CP}{1+CP} \qquad \text{Eqn(11)}$$

This equation only applies if the system is stable.

An ideal control system has an overall closed-loop transfer function of identically one for all frequencies. As can be seen by equation 11, this ideal can only be approached as the controller-plant combination (CP) approaches infinity. Thus, system designers are always seeking to increase gain to improve performance. However, gain may not be increased arbitrarily as we will explore subsequently.

Let us examine a non ideal application of control using only derivative control to influence our mass on a spring. Again applying only K_d=0.001 (sub critical damping), the following plots reflect the closed loop response of the same system as shown figure 4. The plot is no longer logarithmic in frequency as there is a region of interest near the open-loop resonance to examine in detail.

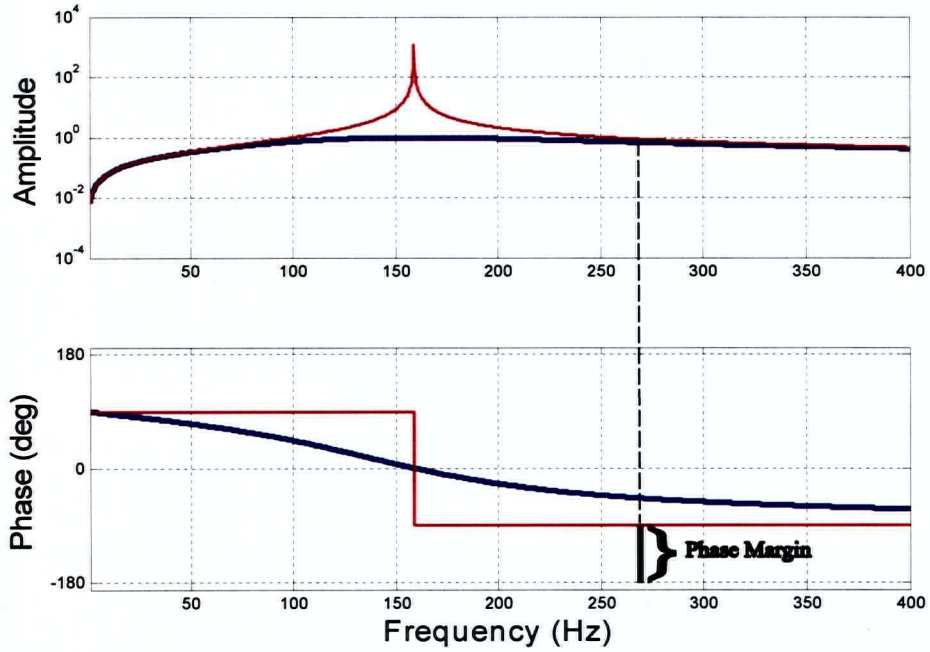

Figure 8: Bode Plot for Differentially Controlled Mass

The blue plot (thick) represents the closed-loop transfer function for the system. The red plot (thin) represents the open-loop transfer function of the control-mass system.

Our system only approaches the ideal of unity at zero phase from about 100 Hz to about 220 Hz. Yet over this region, where the system has an open loop resonance, the control response is good. Differential control alone of a simple second order system performs well near resonance with degraded performance near DC and above resonance.

The open-loop plot is the combination of differential control and the fundamental system. Notice that the fundamental resonance is still present but the shape of the amplitude function has changed. This is the nature of "Bode plot driven design" of control systems where the open-loop Bode plot is adjusted with controller functions to achieve a desired result.

Notice that the closed loop transfer function is very close to unity from 100 Hz to 230 Hz and that the resonance is completely eliminated.

2.4. Stability in Continuous Feedback Systems

The open loop transfer function (CP) can give some indication of stability. Stability criteria for systems can be inferred from the open loop Bode plot. The phase margin is the phase difference from –180 degrees at the frequency where the amplitude plot crosses unity gain. The gain margin is the reciprocal of the amplitude where the phase crosses –180: this must be greater than unity (ie. the amplitude plot must be below unity as phase crosses –180). Stability in a continuous system is explored more fully in the following section.

If the complex phase of CP is -180 degrees at some frequency and the magnitude of CP at that frequency is slightly less than 1, the denominator of eqn 11 will be nearly zero, and the closed-loop system response will be very large. Physically, this is just an oscillator with low damping being driven near its resonant frequency.

If CP is exactly equal to -1 at some frequency, the closed-loop response is infinite. Equivalently, when CP=-1, any value of y whatsoever is consistent with eqn 10. The physical meaning of this is that the system with feedback activated will oscillate indefinitely at constant amplitude at that frequency, even in the absence of any input. This is known as the Barkhausen criterion for oscillations.

If the phase of CP is not exactly -180 degrees at some frequency, the only solution of eqn 10 at that frequency with r=0 is y=0. Self-perpetuating oscillations of a feedback system are not possible at a frequency where the phase of CP is not -180 degrees.

If magnitude of CP is greater than 1 at a frequency where the phase of CP is -180 degrees, there is another solution to eqn 10 besides y=0, namely y=infinity. Physically, the plant output causes the controller to command a still larger plant output. The system with feedback will display growing oscillations at that frequency. (Predicting the actual growth rate requires knowledge about the plant and controller response at other frequencies.)

This is made concrete in figure 9, a Bode plot of the same system we have been studying, except with negative derivative feedback. Physically, this corresponds to anti-damping, and we expect the system to exhibit growing oscillations at the resonant frequency. We see that the open loop phase (CP in red) crosses -180 degrees at the resonant frequency, and the magnitude of the response is greater than one, so the Barkhausen criterion predicts the oscillations that we intuitively expect.

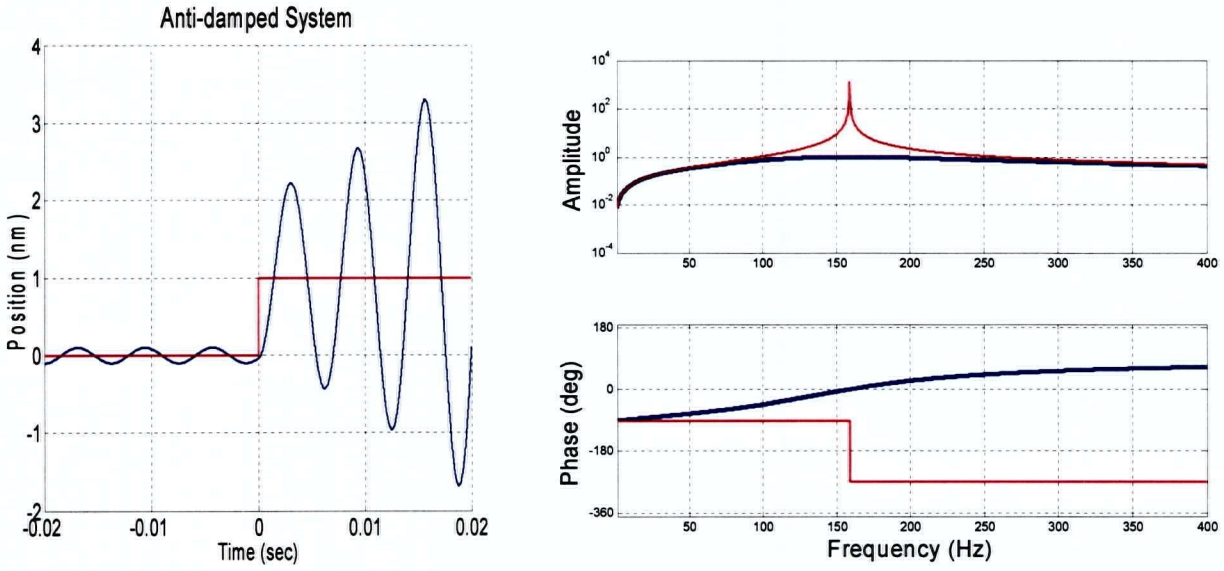

Figure 9: Unstable System

The plot on the left shows a system with slightly negative derivative gain and unstable time-domain response. The figure on the right show a corresponding Bode plot where the Barkhausen criterion is violated at resonance.

Stable feedback control requires that the magnitude of CP be less than 1 at any frequency where the phase of CP is -180 degrees (or more generally, 180 + n x 360 degrees).

Control engineers call the reciprocal of the magnitude of CP at such a frequency the "gain margin" at that frequency. If the gain margin is less than 1, the gain is greater than 1, and the system is unstable. Higher gain margin is better.

Typically, the magnitude of CP falls below 1 at some frequency and continues to fall off at higher frequencies. The "phase margin" is the difference between the actual phase of CP and the nearest unstable phase at the frequency where the magnitude of CP finally falls below 1.

The system shown in figure 7 (with Bode plot shown in figure 8) is stable, with a phase margin of 90 degrees. The gain margin is undefined because the phase never crosses -180 degrees.

The system shown in figure 9 applies negative derivative gain to demonstrate instability. This was a highly artificial example shown for stark simplicity. In actual practice, systems may show an unstable response to simple control strategies whenever there is more than one resonance.

For example, the Bode plot in figure 10 below shows a system with two strong resonant responses at 100 Hz and 300 Hz. The controller has been set to provide a derivative gain appropriate to control the first peak. Note however that the phase will always cross –180 degrees at the second peak and that the system will be unstable with this gain setting. The open loop response crosses –180 degrees at 300 Hz and the amplitude is high at this point since it is a resonant point. Thus the system will oscillate at the second resonant frequency even though the system is well designed to control the first resonance.

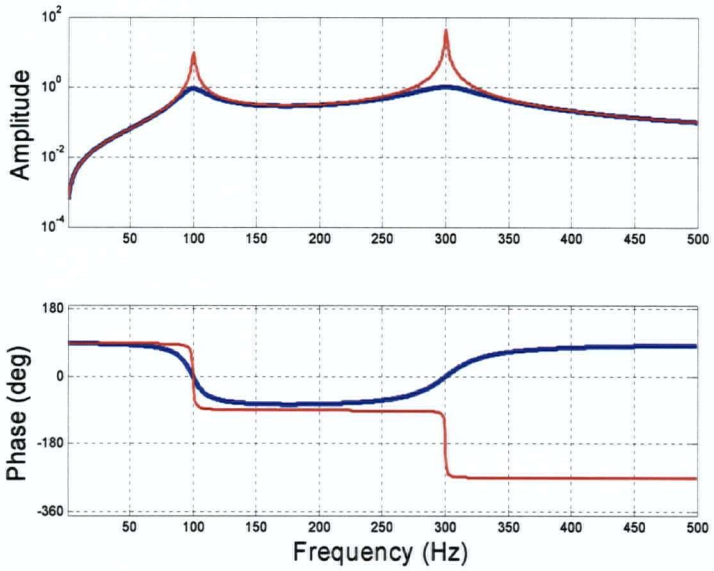

Figure 10: Two Resonance System

This bode plots shows the open loop response (CP) in red and the closed loop response in blue.

Derivative control response increases with frequency and any resonance above the fundamental may cause instability. For this reason, derivative gain is usually applied with a first or second order low-pass filter with the rolloff frequency below any additional plant resonances.

2.5. Discrete Time

A common practice with higher order differential equations is to separate an n-order differential equation into n first order equations. The standard procedure with control system notation is to separate the control term so that natural system evolution and control response are clearly separated and the system is converted into a matrix vector equation. Notation convention used herein is that a bold typeface indicates a matrix whereas simple, non-bold italics constitute a scalar or vector as noted.

As an example, the system as described in figure 1 and in equation 1 can be represented as follows:

$$\dot{x} = \mathbf{a}x + \mathbf{b}u \qquad \text{Eqn(12)}$$

Where **a** is the continuous system propagation operator, **b** is the continuous system response to the control actuator, x is the state vector, and u is the actuator length (measured from null).

The state vector may be written as:

$$x = \begin{pmatrix} y(t) \\ \dfrac{dy(t)}{dt} \end{pmatrix} = \begin{pmatrix} y \\ \dot{y} \end{pmatrix} \qquad \text{Eqn(13)}$$

where y is the test mass position scalar.

Our specific numerical example from figure 1 is defined as follows:

$$\mathbf{a} = \begin{pmatrix} 0 & 1 \\ -k/m & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1\text{x}10^{6} & 0 \end{pmatrix} \qquad \text{Eqn(14)}$$

$$\mathbf{b} = \begin{pmatrix} 0 \\ -k/m \end{pmatrix} = \begin{pmatrix} 0 \\ -1\text{x}10^{6} \end{pmatrix} \qquad \text{Eqn(15)}$$

Where k and m are (as previously noted) the spring constant scalar and the mass scalar.

Loosely based upon the derivation in Dutton et al. [5], a continuous system can be translated into a discrete parameterization as follows:

Firstly equation 12 may be integrated to solve for an evolving state $x(t)$ as follows.

$$x(t) = \exp(\mathbf{a}t)x(0) + \int_0^t \exp(\mathbf{a}(t-\tau))\mathbf{b}u(\tau)d\tau \qquad \text{Eqn(16)}$$

If we assume that *u(t)* is constant for the time interval *h*, this implies::

$$x(h) = \exp(\mathbf{a}h)x(0) + \mathbf{b}u\int_0^h \exp(\mathbf{a}(h-\tau))d\tau \qquad \text{Eqn(17)}$$
$$= \exp(\mathbf{a}h)x(0) + \mathbf{b}u\mathbf{a}^{-1}\exp(\mathbf{a}(h-\tau))\Big|_{\tau=0}^{h}$$
$$= \exp(\mathbf{a}h)x(0) + \mathbf{b}u\mathbf{a}^{-1}(\exp(\mathbf{a}(h)-\mathbf{I}))$$

We now define the free and forced matrices:

$$\mathbf{A} = \exp(\mathbf{a}h) \qquad \text{Eqn(18)}$$

$$\mathbf{B} = \mathbf{a}^{-1}[\mathbf{A}-\mathbf{I}]\mathbf{b} \qquad \text{Eqn(19)}$$

And we can then write

$$x(h) = \mathbf{A}x(0) + \mathbf{B}u \qquad \text{Eqn(20)}$$

If u takes on a different value u_k for each time interval of duration h:

$$x_{k+1} = \mathbf{A}x_k + \mathbf{B}u_k \qquad \text{Eqn(21)}$$

Turning back to our simple numerical example, let us assume a sampling frequency of 1 kHz. This time interval of 0.001 seconds results in the following parameters for our discrete system description:

$$\mathbf{A} = \begin{pmatrix} 0.5403 & 0.0008415 \\ -841.5 & 0.5403 \end{pmatrix} \qquad \text{Eqn(22)}$$

$$\mathbf{B} = \begin{pmatrix} 0.4597 \\ 841.5 \end{pmatrix} \qquad \text{Eqn(23)}$$

Where **A** is the discrete system propagation operator and **B** is the discrete effect of unit output. The solution of **A** was found by eigenvalue decomposition of the system explicitly using the Maple[17] mathematics package. Alternatively, Matlab[18] offers a direct continuous to discrete system conversion command in its "Control System Toolbox" package.

Using the above discrete system description, we can generate discrete system steps. The figure below shows such a series of discrete positions generated using the matrix values

above. These are superimposed upon the continuous system response shown previously in Figure 2.

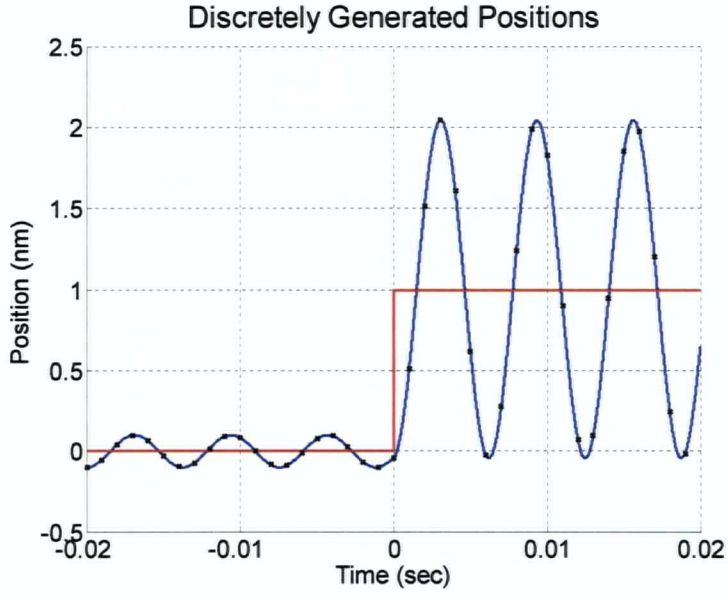

Figure 11: Discrete Evolution Sample of the Simplified Model

The black symbols represent the discrete evolution of the system at 0.001 second intervals. The continuous system response is the blue curve. The actuator position is in red.

In the previous figure, we see a set of generated discrete position values superimposed upon the continuously resolved system plot. The discrete values correctly and exactly superimpose on the plot. This would be true even if the discrete time interval had been taken to be many cycles of the system oscillation.

2.6. Discrete Control

Discrete-time descriptions of dynamical systems are ideally suited to analysis of control using a computer, where measurement and control signals are updated at regular intervals.

Proportional and derivative control of the mass and spring system can be combined from eqn. 2 and eqn. 4 to be:

$$u(t) = -K_p y(t) - K_d \frac{dy}{dt} + r(t) \qquad \text{Eqn(24)}$$

The *r(t)* term allows the control to displace the equilibrium away from the natural y=0.

If the sampling rate is high, we can approximate each function as discrete steps

$$u_k = -K_p y_k - K_d \left(\frac{dy}{dt}\right)_k + r_k \qquad \text{Eqn(25)}$$

We can define a (1x2) gain matrix:

$$\mathbf{K} = [-K_p \quad -K_d] \qquad \text{Eqn(26)}$$

Which allows us to write:

$$u_k = \mathbf{K}x_k + r_k \qquad \text{Eqn(27)}$$

If we only wish to control the dynamics of how the system approaches (or possibly diverges from) the equilibrium, and are not interested in changing the equilibrium position, we can set r_k=0. Then we can write the system evolution with feedback control as:

$$\begin{aligned} x_{k+1} &= \mathbf{A}x_k + \mathbf{B}u_k \\ &= \mathbf{A}x_k + \mathbf{BK}x_k = (\mathbf{A}+\mathbf{BK})x_k \end{aligned} \qquad \text{Eqn(28)}$$

The stability properties of the system are determined by the eigenvalues and eigenvectors of the matrix:

$$\mathbf{P} = \mathbf{A} + \mathbf{BK} \qquad \text{Eqn(29)}$$

If any eigenvalue of **P** is greater than unity, then the system is unstable. An initial state x will grow in magnitude on subsequent iterations (unless it happens to be orthogonal to all the eigenvectors whose eigenvalue is greater than 1).

The figure following plots contours of greatest eigenvalue for the **P** matrix against PD gains for our numerical example.

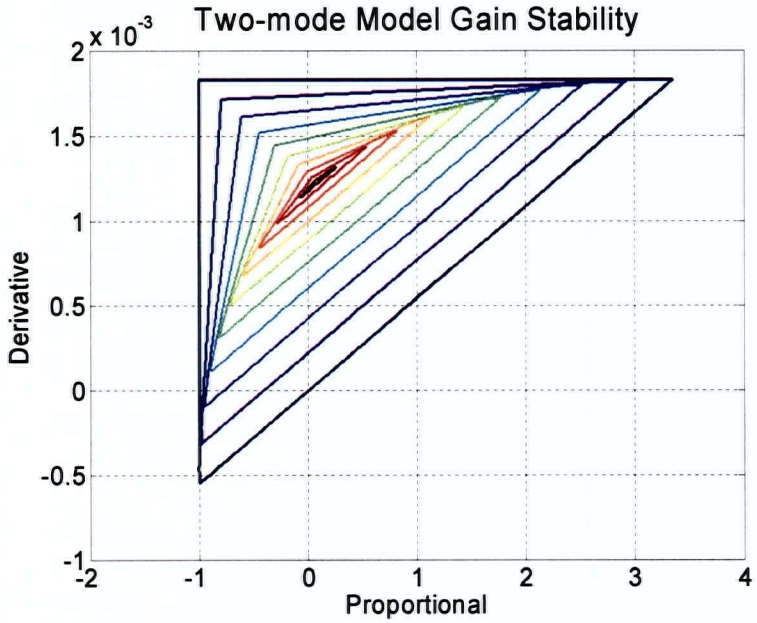

Figure 12: Gain Stability Contours

System propagation eigenvalue contour (contour interval 0.1).

The outer triangle represents the locus of points where the largest eigenvalue (absolute value) of the propagation matrix is exactly unity. Gain values outside of this region are unstable. Gain values inside of this region show decreasing values of maximum eigenvalues.

In the continuous control case, the system was stable for any positive value of derivative gain K_d and any value of proportional gain K_p greater than –1. In the discrete-time case the stable region is different. Interestingly, pure positive proportional gain is unstable. This is because the finite time step size keeps the control from reversing sign at exactly the same time the position reverses sign. That is, pure proportional gain in the discrete model introduces some lag. The time lag due to discrete sampling pushes the phase past -180°.

Additionally, unlike the continuous case, there is a maximum stable value of proportional and derivative gain. Physically, this corresponds to applying a force sufficient to more than reverse the velocity during a single time-step. On the next step, an even larger force in the opposite direction will be applied, causing a divergent oscillation. The discrete time steps introduce a phase lag that makes the system unstable in the absence of natural or artificial damping.

With a reasonable amount of derivative gain applied, there is still a maximum stable proportional gain. Physically, this corresponds to applying a force more than sufficient to reverse the position of the mass during a single time step. The allowable proportional gain increases as the derivative gain is increased because the derivative gain (when lower than its own instability threshold) tends to oppose the excessive proportional gain.

An obvious feature of the stability contour is the position near K_p=0.1 and K_d=1.7x10^{-3}. This position has the interesting property that both eigenvalues are zero for these gains. This is known as the system deadbeat gain and can be explicitly derived (refer to Appendix B **Deadbeat Gain**).

The maximum stable gains, and the point of optimal gains, depend on the time step compared to the natural frequency of the system with zero gains. The time step was chosen to be relatively large for this example, about 1/6 of the period at the natural frequency. This makes the optimal proportional gain relatively small. In other numerical experiments not shown here, if the time step is made smaller, the maximum stable gains increase. Higher proportional gain increases the closed-loop oscillation frequency. The system typically becomes unstable when the closed-loop period is reduced to only a few time steps.

We shall explore the stability region in gain space just a bit. The following plot is the system response with gain set at just 50% of the deadbeat gain activated at t=0 with a setpoint of 1.

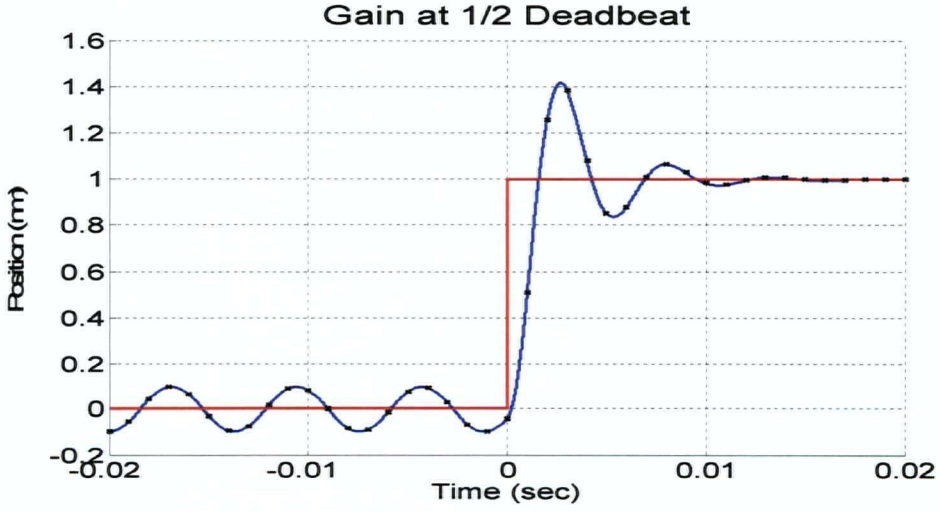

Figure 13: Simplified Model with PD Gains Set at 50% of Deadbeat

The continuous response is shown with the smooth curve (blue) and the discrete positions are shown at the tick marks. The step function (red) is the ideal response.

The response approximates an underdamped second order response. This would be exactly true in the case of continuously controlled output. However, since the output is controlled in steps and remains constant between tick marks, the output actually only approximates an underdamped second order response.

Let us look at system response at the exact deadbeat gain settings. The plot below shows this theoretical response.

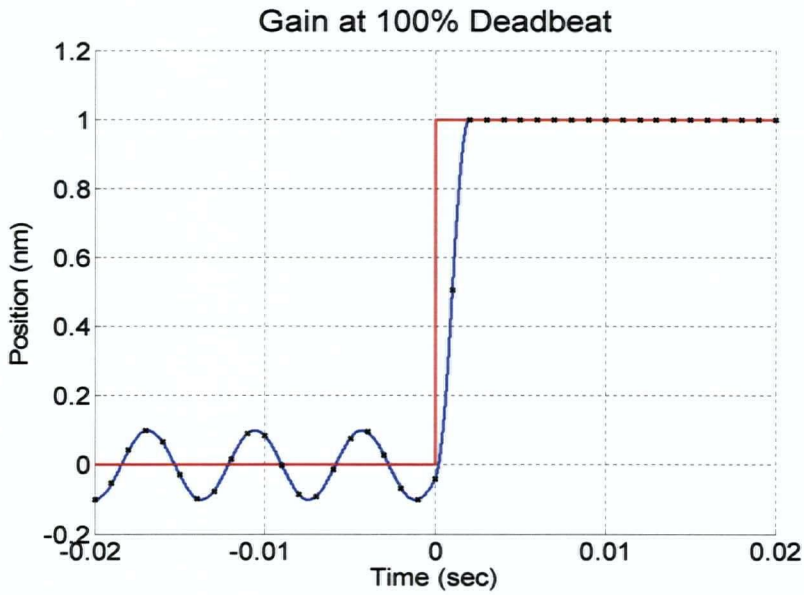

Figure 14: Simplified Model with PD Gains Set at 100% of Deadbeat

This plot shows the system response to activation of the control system at t=0 sec. The setpoint is set at 1 nm to reflect the similarity with previous plots.

The continuous response is shown with the smooth curve (blue) and the discrete positions are shown at the tick marks. The step function (red) is the ideal response.

After only two tick times after start of control, the response is exactly flat. This is not an exponential asymptote but rather is exact.

Superficially, this might appear superior to the exponential damping achieved in the continuous-time analysis of a single-mode system. However, for continuous time there was no limit to how small the exponential damping time constant could be for the corresponding single mass and spring system, while here we are limited by the sampling time.

2.7. Control of Two-Resonance System in Discrete Time

In section 2.4 (Stability in Continuous Feedback Systems) we found that the addition of a second resonance could constrain the range of stable gains, but that discussion was only qualitative. In the previous section, we found that discrete time steps limited the range of stable gains. Now let us combine the constraints of discrete time and of a second resonance, and in a quantitative way.

The one-mass system as previously described is exceedingly easy to control well. The transponder mount to the bulkhead is extremely stiff at 1N/um. There is only one system resonance. However, this simple model of a monolithic mass and infinitely stiff bulkheads is only adequate to explain and control the response of a real system up to perhaps 20Hz.. Beyond this limit, the model becomes inadequate since the mass and bulkheads begin to exhibit non-ideal responses.

Let us simulate one such addition to complexity meant to model a possible internal mode in the mass. Let us add a spring and thus an additional vibration mode to the system.

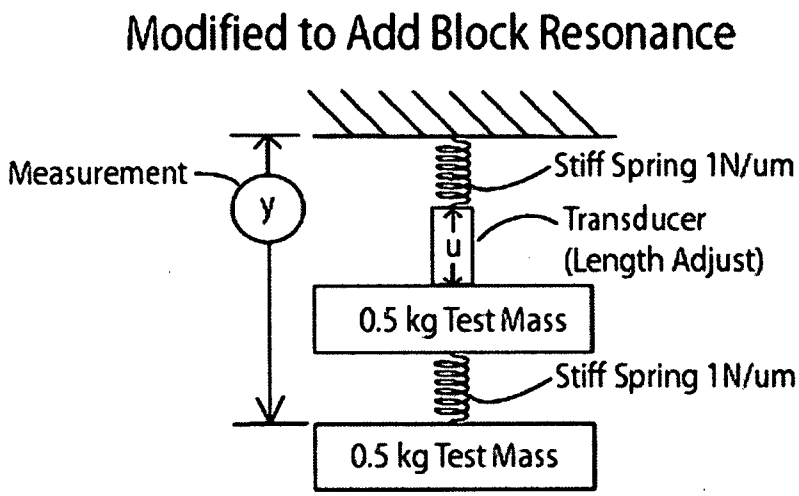

Figure 15: Modified Model Adding a Single Resonance

This model adds a simple resonance without damping to the system.

We have split the mass in half, and added a spring in the middle identical to the original single spring. Now there will be two modes of oscillation. One will have the two masses moving roughly as a single unit with about the original frequency. For the other mode, the lower half-mass will be approximately fixed, and the upper mass will move at about twice the original frequency (since it has half the mass, and sees twice the spring-constant).

The system is now a four state model represented as follows:

$$x = \begin{pmatrix} y_1 \\ \dot{y}_1 \\ y_2 \\ \dot{y}_2 \end{pmatrix} \quad \text{And} \quad \mathbf{a} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -(k_1+k_2)/m_1 & 0 & k_2/m_1 & 0 \\ 0 & 0 & 0 & 1 \\ k_2/m_2 & 0 & -k_2/m_2 & 0 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 0 \\ k_1/m_1 \\ 0 \\ 0 \end{pmatrix} \qquad \text{Eqn(30)}$$

Let us now apply proportional and derivative control using the position and velocity of mass 2 (the lower one in figure 15). The **P** matrix is $\mathbf{P}=\mathbf{A}+\mathbf{B}u$ where $u=(0,0,K_p,K_d)$ and **A** and **B** are exponentiated from **a** and **b**. Figure 16 shows the contours of the largest eigenvalue as a function of K_p and K_d. The contour for the 2-state problem are also shown.

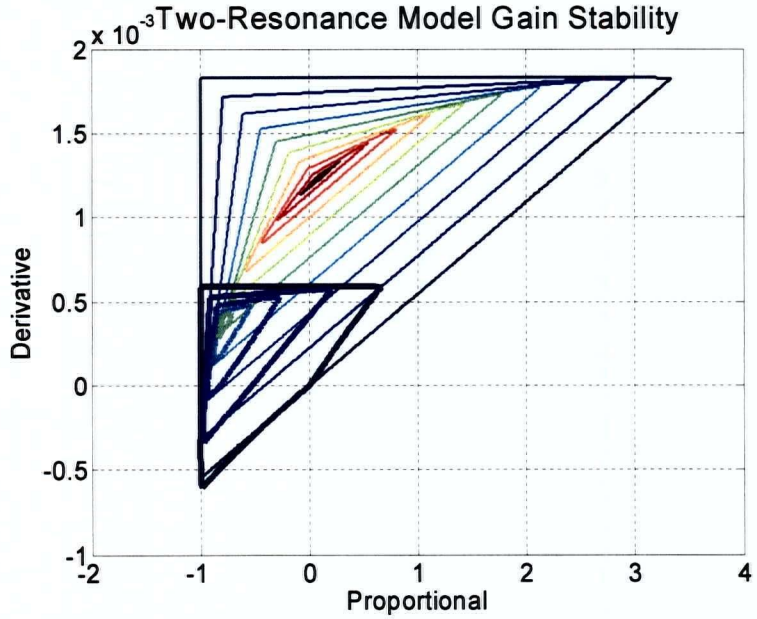

Figure 16: Two-Resonance Model Gain Stability

Locus of maximum eigenvalue of P Vs feedback gains: 1 for outer contour, decreasing by 0.1 per contour. Bold lines are for two-resonance system, finer lines for one-resonance system.

In comparing the two-resonance system with the one-resonance system, the stable region is much smaller. The maximum stable gains are about a factor of 3 smaller. The "deadbeat" solution for the single-resonance system is a bit more than a factor of 2 outside the maximum stable gains of the two-resonance system.

Since the new mode of the system has about twice the frequency, it is not surprising that the maximum stable gains are reduced by a factor of about two.

Also, the maximum eigenvalue of the P matrix is never zero, so there is no "deadbeat" solution that will bring the mass to a dead stop in two time steps.

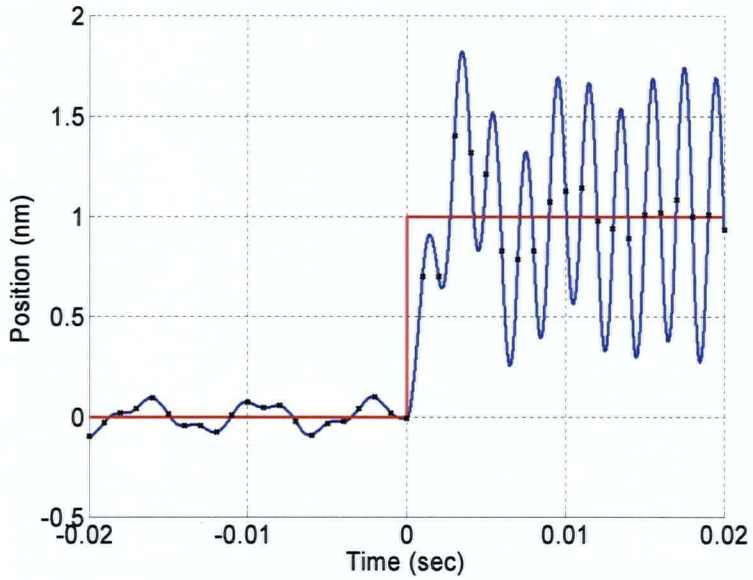

Figure 17: Modified Model at Previously Stable Gains

The black points show the system at discrete time samples, the blue plot is the continuous motion of the system and the red trace is the ideal response.

Figure 17 shows the time response of the lower mass in the two-mode system, using the same gains as in figure 13, which are slightly outside the stable region of figure 17. For t<0, the setpoint is r=0, and the mass shows motion in both modes. At t=0, the feedback is turned on and the setpoint is changed to r=1. The mass then starts oscillating around the new setpoint, but the oscillations are not damped, and in fact are slowly growing.

Interestingly, the discrete time samples (the black crosses) show much less motion than is actually occurring between the samples. It should be noted that the feedback gain in this model is applied to the actual velocity, the slope of the smooth curve, rather than the velocity that might be derived from finite differences of the discrete time samples.

A natural question is, can we do better by allowing the actuator setting to be a function of all of the elements of the state vector x, using more gain coefficients? For this system, we need to search in a 4-dimensional gain space.

We can plot multi-dimensional stability contours by projecting the maximum eigenvalue found in all non-plotted axes onto two axes which can be plotted. Six plots then can map the stability zone onto all possible gain pairs for a four-state system.

Let us now plot the stability region of our four state model allowing "perfect" information from our two masses (position and velocity of each mass) to be used by our control algorithm.

Our controller is still described by equation 29 only now the vector K consists of gains for each of the four system states (mass 1 position, mass 1 velocity, mass 2 position, mass 2 velocity).

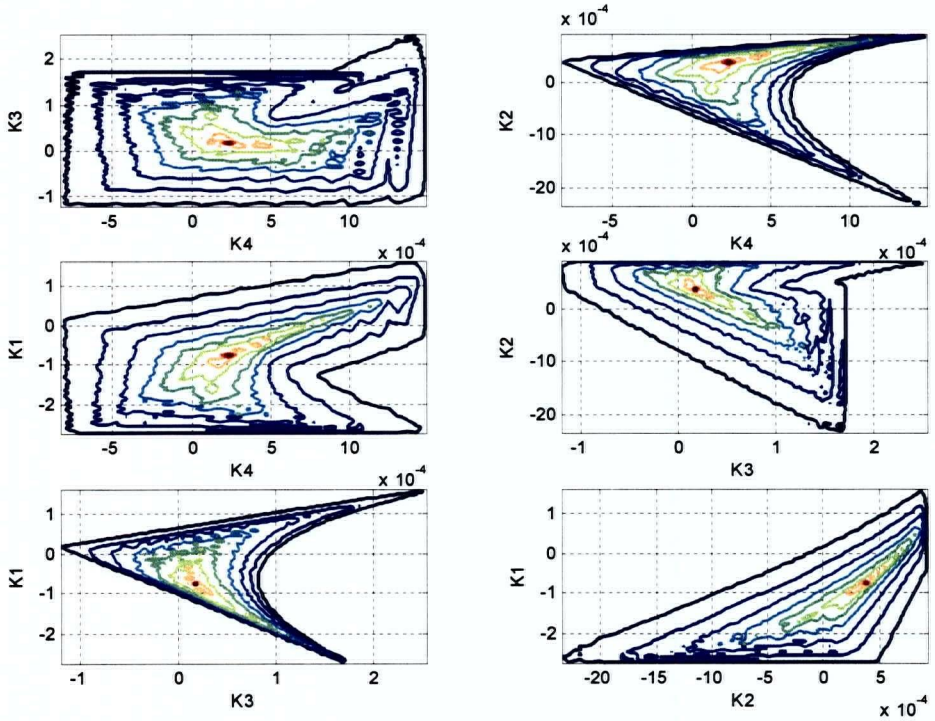

Figure 18: Two-Resonance Model for 4-Dimensional Gain Stability

Locus of maximum eigenvalue of P Vs feedback gains: 1 for outer contour, decreasing by 0.1 per contour. Bold lines are for two-resonance system, finer lines for one-resonance system. Each plot is a projection of the 4 dimensional locus onto two gain axes.

It appears that there is still a deadbeat solution, with all 4 eigenvalues of the **P** matrix being zero. But it requires that the actuator setting be a function of the position and velocity of both masses.

The deadbeat gains are positive damping for both velocities, slightly positive proportional gain for one mass, and nearly -1 proportional gain for the other. The solution does not appear to be fine-tuned in the sense that small changes in any gain parameter still gives reasonably small maximum eigenvalues. However, small changes in the parameters of the mechanical system can cause large changes in the location of the deadbeat gains.

It is shown in Appendix B that there is a deadbeat solution to any order N linear control problem that makes the entire state vector zero. The solution generally requires N time steps to reduce all elements of an arbitrary initial the state vector to zero. However the solution theoretically still exists even when the time-step is reduced to zero.

This would appear to be potentially superior to the continuous-time case, where the introduction of additional modes put a limit on the gain and thus on the exponential time constant. This is not a fair comparison, because the discrete time deadbeat solution uses more gain coefficients than we used for the continuous time two-mode case. It is possible to use more gain coefficients and reduce the exponential time-constant to zero in continuous time as well.

Numerical experiments indicate that in both the continuous and discrete time cases, large gain values are required to achieve solutions with fast control (short time steps). Also, the positions and velocities may take on extreme values during the N time steps of the solution.

2.8. Conclusion

For a simple mass and spring mechanical system, with control applied through a piezoelectric actuator, proportional gain increases the resonant frequency, and derivative gain damps the oscillations. Control gains can be determined to artificially raise the resonant frequency while applying artificial critical damping. In continuous time, there is no limit to the speed of control in such a simple system.

The response of a system to the control actuator, with or without feedback control, can be described through a Bode plot of magnitude and phase versus frequency. If the magnitude of the response of the system and controller (including the gain coefficients) exceeds 1 at a frequency where the phase is 180+n x 360 degrees, the system will be unstable at that frequency if feedback is applied. This can greatly limit the maximum stable gain for systems with multiple resonances, if the control algorithm is limited to use only the value and derivative of the system output.

It is possible to describe exactly a continuous time system with actuator settings that are updated at discrete time intervals using a matrix formalism. The maximum stable gain is limited even for a single-resonance system. But there is a gain setting that will bring a single-resonance system to a stop in 2 time steps, no matter how short the time steps are.

If there are multiple masses and springs, but the feedback control responds to only a single position and velocity, the maximum stable gain can be greatly reduced. If the feedback control responds to all the positions and velocities, there is again a solution that brings all M masses to a stop in 2M time steps. However if the time steps are short, the required gains can become very large.

3. Apparatus

3.1. Interferometer

The interferometer for nanometer position measurements was assembled from standard optical laboratory equipment and is shown in Figure 19 below. It is in Michelson configuration, except the arms are made parallel by a 45 degree mirror. This allows the arm lengths to be extended easily, and keeps the light paths close together so they traverse the same air density, reducing some sources of error.

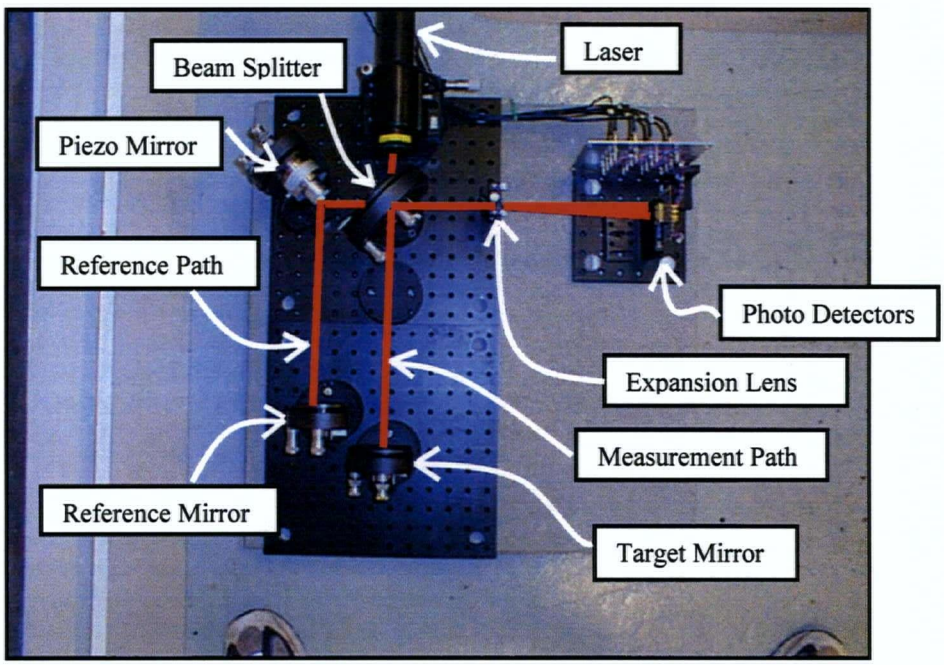

Figure 19: Interferometer Setup

Photo of Michelson interferometer setup. The reference arm optical path has been reflected to be in the same direction as the measurement path. The red laser has been added with a paint package for illustration since the beam itself does not show in the photo.

In actual operation, one optical baseplate has the laser, the beam splitter, and the 45 degree mirror, which is mounted on a Physik-Instrumente[19] piezoelectric actuator to allow the optical path length of one interferometer arm to be varied for calibration and tests. This piezo is far larger than required to move the small mirror, but the large size is convenient for attaching the mirror and for translating it without tilting it.

The two end mirrors are located asymmetrically to compensate for the path-length difference introduced by the 45 degree mirror.

The light source is a Melles-Griot model STP-901 intensity/frequency-stabilized HeNe laser with power at 1 milliwatt. It is possible to adjust the mirrors so light does not go back into the laser aperture, so an optical isolator is not required. For arm lengths of less than 10 meters, the divergence of the beam is small enough that a collimator is not required either.

The interference pattern is expanded by a cylindrical lens onto a linear silicon photodiode array. The orientation and spatial period of the fringes can be controlled by adjusting the mirror angles. Intensity control is done by vertical steering of the laser partially off of the sensitive area of the photodiode array.

3.2. Electronics

The photodiode detectors were operated in unbiased mode, with current signals travelling in coaxial cables to a multichannel amplifier unit. Each channel had an OP-27 amplifier configured to produce a DC-coupled current to voltage output capable of producing a saturated 10 Volt output with moderate laser light incidence upon the detector.

For most of the test-platform control experiments described later, it was convenient to use a simple capacitive position sensor instead of the interferometer. This was a Physik-Instrumente model D100 sensor head pair and model E610-C0 electronics module. This combination has a noise figure of 8 pm per root Hertz, for a position resolution of 0.6 nm at our 5 kHz sampling frequency.

For early experiments we used a commercial piezo driver to buffer our digital output to the piezo. Specifically, the model used was Physik-Instrumente model E610-00 with a voltage gain of 10 and maximum output of 100 volts and 60 milliamps (RMS). This created a slew-rate issue with respect to our control system. Over the full scale of 100V, with a 10uF piezo, slew time at 60mA is 17mS. Our digital control system data-acquisition/control loop time is 0.2mS. Even a 1% of full scale adjustment would require the full loop time to slew. This proved unacceptably slow.

To insure that control effects were linear and related to the system and not to slew rate, we needed a much higher current source. Since we didn't actually need the full voltage range of our piezo, we designed a 24V push-pull output amplifier which was capable of delivering 2A. Schematic for this amplifier is seen in the figure 20 following:

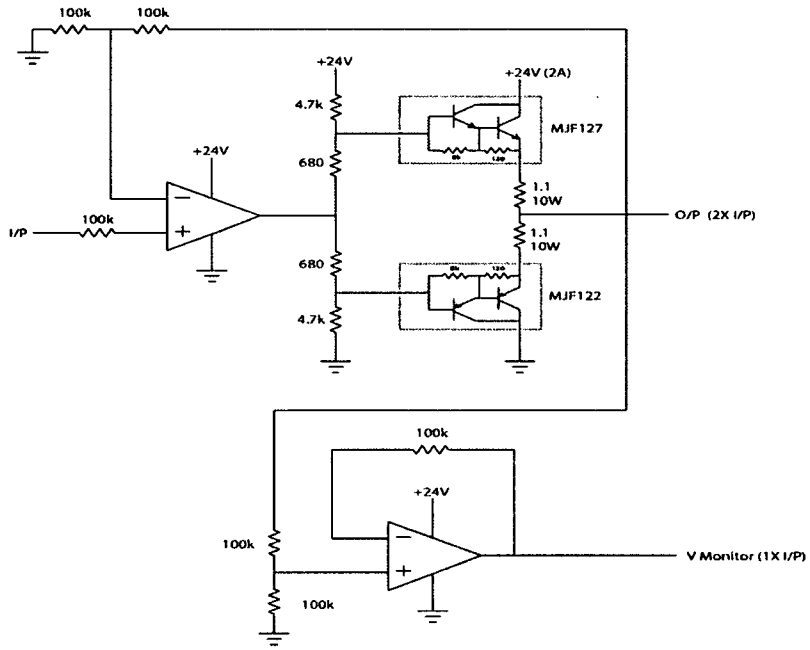

Figure 20: Piezo Driver Circuit Schematic

This simple push-pull current driver also provides a X2 voltage amplification. Cross-over distortion is negligible since the O/P load is capacitive.

V monitor can be used to verify actual output voltage and test for slew rate difficulties.

3.3. Data Acquisition

Data acquisition and control used an Adwin-Gold real-time computer system[14]. The on-board DSP is programmable in BASIC and has a 25nS clock cycle with BASIC statements taking as little as four clock cycles and as great as 400 clock cycles for floating point math functions such as log.

Data presentation is made using the "Igor"[15] data manipulation package from Wavemetrics. Interface software between the Adwin-Gold and IGOR is provided by RTS Consulting Ltd[16].

Real-time oscilloscope functions as well as control algorithm communications using window tools such as sliders is made relatively straightforward. Real-time determinacy is guaranteed since the Adwin-Gold is a dedicated real-time module and the communications to and from the module are relegated to background process within the Adwin.

Using the Adwin-Gold in conjunction with the IGOR software, it is quite straightforward to create hybrid instruments, to test control strategies and to perform data acquisition.

3.4. Lock-in Amplifier

One notable hybrid instrument is the lock-in amplifier used to take spectral system response data. The instrument consists of a sine-wave frequency generator with a phase-locked measurement correlation. An executive control can sweep the frequency and archive the individual frequency responses.

The software architecture of this instrument is laid out in figure 21 following.

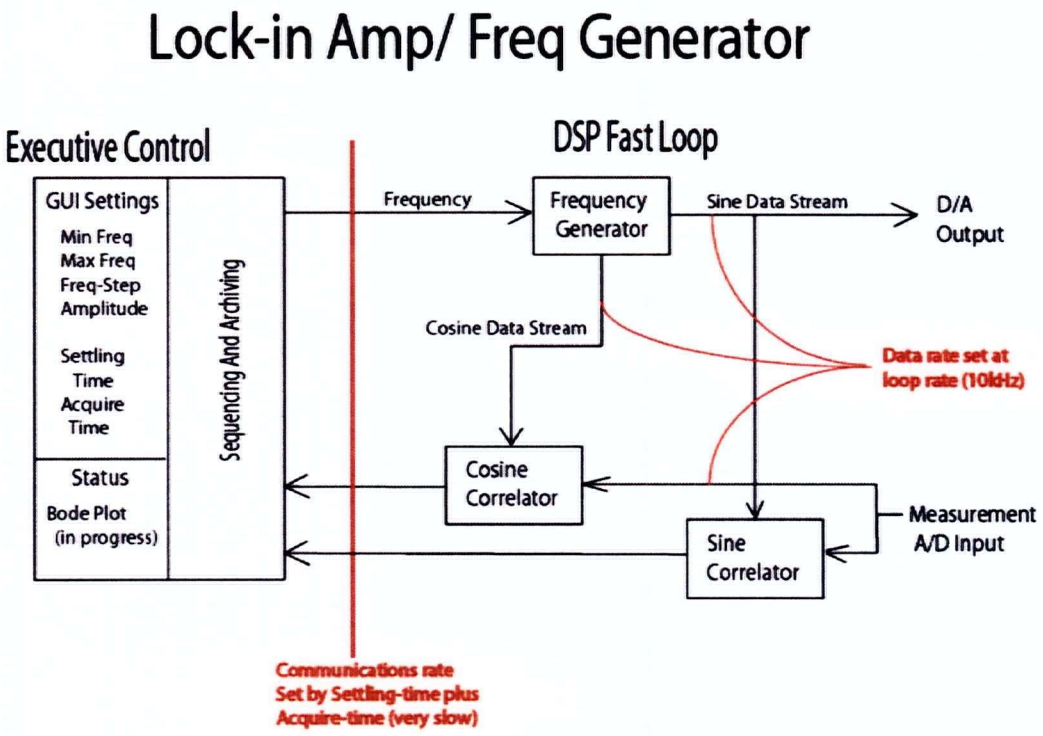

Figure 21: Lock-in Amplifier Block Schematic

Executive Control: The executive control is required to sequence through the frequency limits and apply a settling time and acquire time for each individual test frequency. This process is very slow and does not need to be deterministic. However an additional signal (not shown) for end-of-acquisition and change-frequency is required for the DSP fast loop since correlations are only made using complete sine-wave cycles and zero crossing must thus be used as triggers.

Frequency Generator: Frequency generation is performed using floating point calculations for both the phase operand of the trig functions and the trig functions themselves. This allows the system to operate at frequencies without an algorithmic lower bound. Practically, the lower bound is limited to 0.001Hz or so simply due to realistic acquisition-time requirements. Similarly, the floating point phase operand allows frequencies of any non-harmonic (of the loop time) and effectively allows operation right up to the Nyquist frequency of the loop time (in this case fn=5kHz).

Correlators: Correlation is performed at full loop rate and is simply the accumulated sum of the measurement signal multiplied with the sine (or cosine) data stream. Correlation values are initiated and archived at zero crossings resulting in accurate integrations even over very few cycles. This allows very low frequencies to be analyzed.

3.5. Position Control Test Platform

The test platform assembly is intended to allow a 10-100 kg mass to move in one dimension, with force applied passively by an adjustable spring constant, and controlled actively by a piezoelectric actuator, but constrained from moving in the other 5 degrees of freedom. The concept is illustrated schematically in figure 22 below.

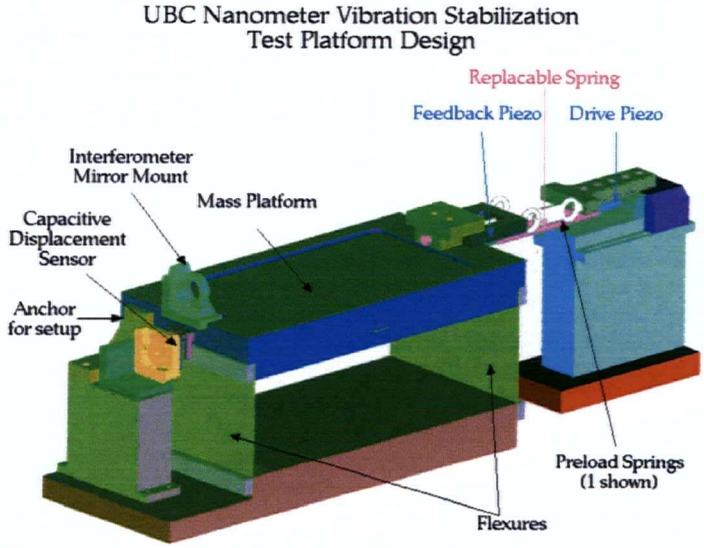

Figure 22: Test Platform Isometric Drawing

The design allows for various spring constants to be used for platform connection to the bulkhead through the piezo actuator. The elimination of all but one degree of freedom is achieved from the flexure mounts. Provision is made for both capacitive position sensing and interferometric position sensing.

The actual platform was constructed out of aluminum and shown in the photograph below:

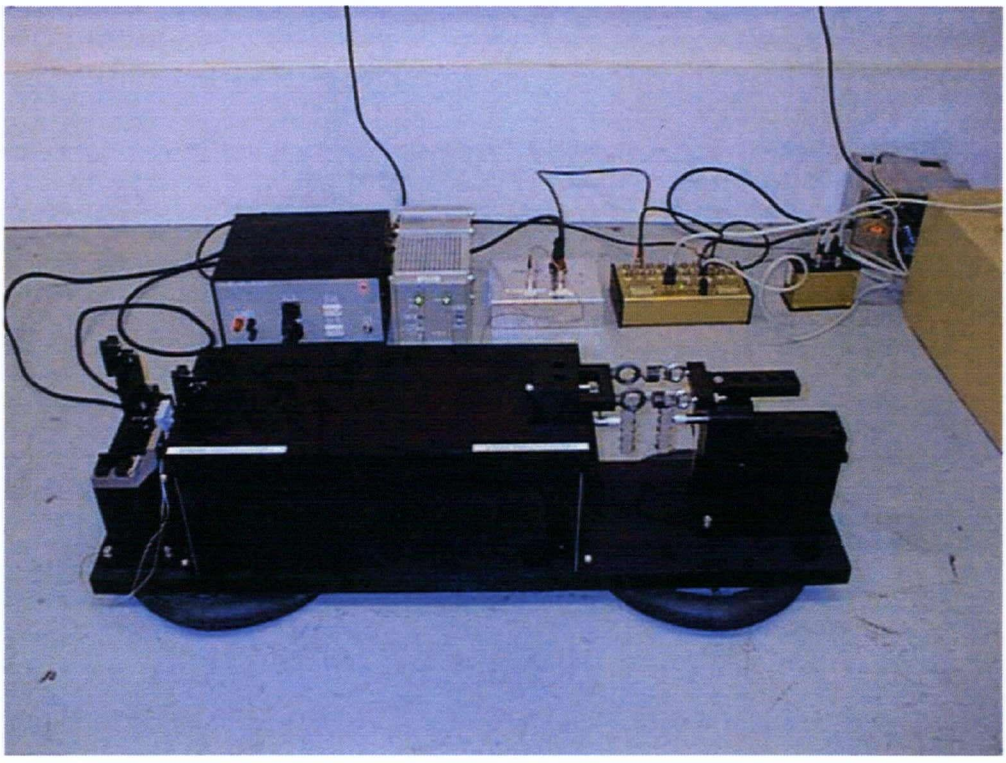

Figure 23: Test Platform Photograph

The baseplate assembly has mounts for a pair of aluminum or steel flexures to hold the test platform. It also has end posts for controlling and measuring the platform position. The entire assembly has provisions for bolting to the floor, although it was simply resting on the floor or isolated by innertubes for the data presented here.

At high system control gains, it was excitations of mechanical modes within the base structure which provided some of the most intractable resonance responses. Future test platform designs anticipate deliberate construction of a reaction mass rather than a bulkhead for the control actuator (piezo element) to act against.

The left baseplate end post has an optical mount used to adjust the gap and parallelism of a Physik-Instrument model D100 capacitive position sensor. It also has a fixed mirror for the reference arm of the interferometer, and an anchor point to fix the moving platform (relatively) rigidly in place.

The test platform to be controlled is a solid aluminum block. At the right end there is a Physik-Instrumente model P840-10 piezoelectric actuator, and on the upper right surface there is an anchor for a fixture to adjust the tension of piezo preload springs. The right baseplate end post has a slider with another PI model P840-10 piezo, and eyebolts for the preload springs. The detail of the actuator can be seen in figure 24 below.

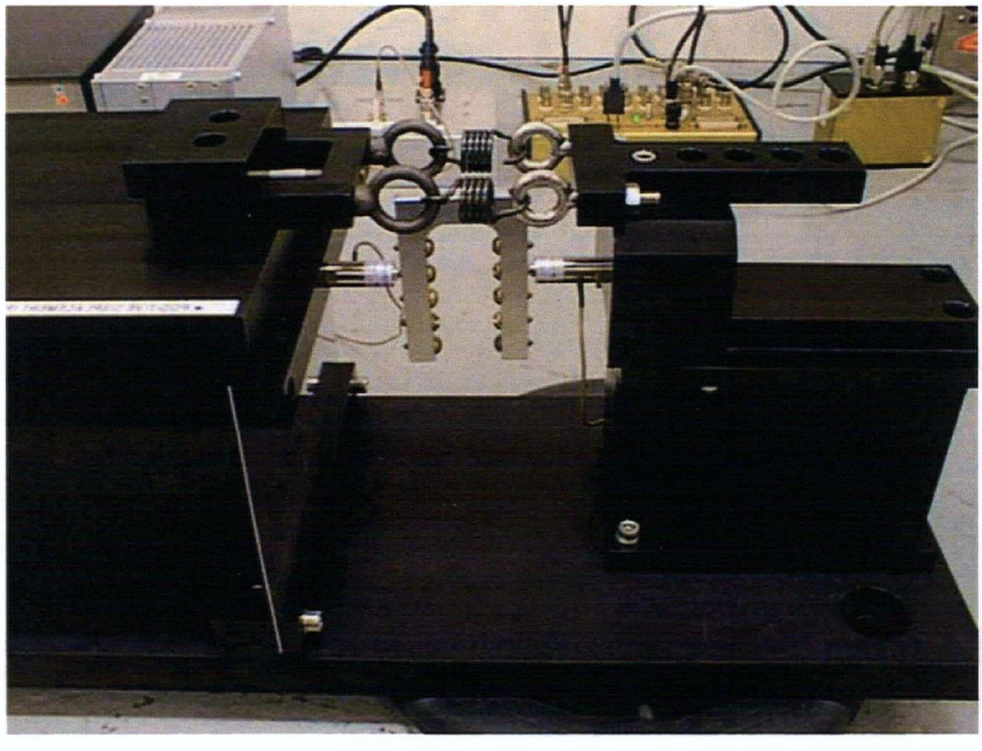

Figure 24: Actuator Detail

The 'U' shaped part is the main reaction spring with the pretension for this spring supplied by the top pair of tensioning springs.

Between the two piezos is a U-shaped aluminum "spring" with multiple attachment points. Changing which attachments are used changes the spring constant for the test platform's motion relative to the baseplate.

It was intended to be possible to change the mass of the test platform by simply putting more objects on top. As will be seen, it proved difficult enough to control the platform at minimum mass, and such high-mass experiments have yet to be performed.

4. Interferometer

4.1. Interferometer Signal Model

4.1.1. Physics

The optical system is a folded Michelson interferometer that was described in section 3.1: Interferometer. Figure 25 shows the classic Michelson configuration The mirrors are adjusted so the two laser beams are not quite parallel at the detector plane, so the beams interfere constructively at some points and destructively at others, forming interference fringes. The fringe pattern migrates across the detector plane if the length of either interferometer arm changes. The change in path length is twice the amount that the target mirror moves. For a He-Ne laser with 632.8 nm wavelength, 50.35 nm of mirror motion corresponds to a beam phase change of 1 radian.

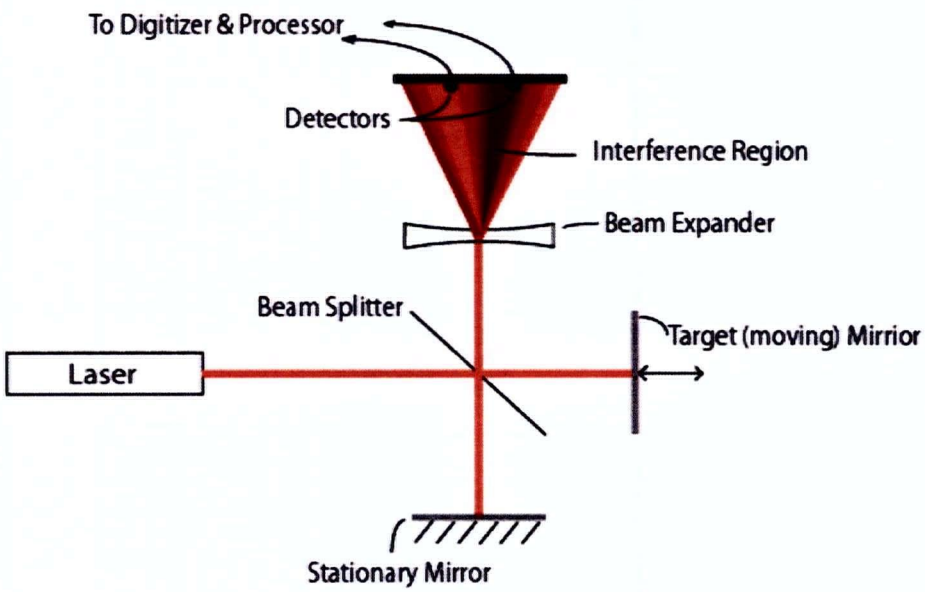

Figure 25: Interferometer Setup

The electric field at photodiode i may be written

$$\vec{E}_i = \vec{F}_i \exp\left[i(\vec{k} \cdot \vec{x}_i - \omega t + \varphi) \right] \qquad \text{Eqn(31)}$$

where F_i is the transverse intensity and polarization vector of the beam at photodiode i, and the subscript allows for the fact that the intensity envelopes of the beams vary across the detector plane. The beam wavevector is k, x_i is the coordinate of photodiode i, ω is the frequency, and φ is a phase offset.

We want to measure the phase difference between two beams after they have traversed the two interferometer arms. The beams have the same frequency and thus the same magnitude $|k|$, but slightly different directions. The polarization vectors are essentially parallel but the intensity envelopes may not be identical. The sum of beams A and B may be written

$$\vec{E}_i = \left\{ \vec{F}_{iA} \exp\left[i(\vec{k}_A \cdot \vec{x}_i + \varphi_A) \right] + \vec{F}_{iB} \exp\left[i(\vec{k}_B \cdot \vec{x}_i + \varphi_B) \right] \right\} \exp(-i\omega t) \qquad \text{Eqn(32)}$$

The measured light intensity is proportional to this dotted with its complex conjugate

$$I_i \propto \vec{F}_{iA} \cdot \vec{F}_{iA} + \vec{F}_{iB} \cdot \vec{F}_{iB} + \vec{F}_{iA} \cdot \vec{F}_{iB} \left\{ \begin{array}{l} \exp\left[i\left((\vec{k}_A - \vec{k}_B) \cdot \vec{x}_i + (\varphi_A - \varphi_B) \right) \right] \\ + \exp\left[-i\left((\vec{k}_A - \vec{k}_B) \cdot \vec{x}_i + (\varphi_A - \varphi_B) \right) \right] \end{array} \right\} \qquad \text{Eqn(33)}$$

$$= \left| \vec{F}_{iA} \right|^2 + \left| \vec{F}_{iB} \right|^2 + (\vec{F}_{iA} \cdot \vec{F}_{iB}) 2 \cos\left[(\vec{k}_A - \vec{k}_B) \cdot \vec{x}_i + (\varphi_A - \varphi_B) \right]$$

The intensity at a given diode is the sum of the two beam intensities plus an interference term. The interference term has a phase contribution $(\vec{k}_a - \vec{k}_b) \bullet \vec{x}_i$ that is different for each diode and a contribution that is the same for all diodes that depends on the phase difference between the two beams φ_A-φ_B.

The photodiode current is proportional to light intensity, and is converted to a voltage by the linear analog electronics. The digitized voltage for photodiode i can thus be parameterized as

$$V_i = c_i + b_i \sin(\varphi + \delta_i) \qquad \text{Eqn(34)}$$

where c_i is due to $|F_{iA}|^2 + |F_{iB}|^2$ plus electronic offset and any background light, b_i is due to $2(\vec{F}_{iA} \cdot \vec{F}_{iB})$ and δ_i is $(\vec{k}_A - \vec{k}_B) \cdot \vec{x}_i - \pi/2$

If we know the constants c_i, b_i, and δ_i, we can take a single voltage measurement V_i and solve for φ_i. There are multiple solutions, differing by multiples of 2π, and also solutions with the same value but opposite slope for the sine function. If we have a second voltage measured by a detector with a different δ_i, we can eliminate the opposite-slope ambiguity. If we take a

sequence of measurements over time, frequent enough that the phase change between measurements is much less than 2π, we can track motions through many wavelengths. In our application, the remaining arbitrary overall phase offset is acceptable.

Thus with two light intensity detectors, we should be able to reconstruct the motion of the interferometer end mirror.

4.1.2. Direct Parameterization

Signals from such a system have an appearance characteristically shown in Figure 26 (below). This figure shows the signals expected in two detectors for a bit more than a quarter cycle of sinusoidal mirror motion with an amplitude of a few microns . The intensity measurements vary sinusoidally with target position. But the amplitude, offset and phase shift are arbitrarily set by system optics and detector sensitivity. Individual detectors will have different sinusoid envelope parameters. A model can be constructed which assumes that all angles in the system remain constant and the only variation with target position is the phase of the measurement arm.

The explicit model then is given below where for our purposes we will restrict ourselves to two detectors and thus $i \in \{1,2\}$.

$$v_i(\varphi) = c_i + b_i \sin(\varphi + \delta_i) \qquad \text{Eqn(35)}$$

This is simply a restatement of equation 34. Where v_i is the measured voltage, c_i, b_i are arbitrary gain coefficients, φ is the absolute phase due to mirror position, δ_i is the relative phase for different detectors (δ_1 is arbitrarily set to zero) and where i is the detector index.

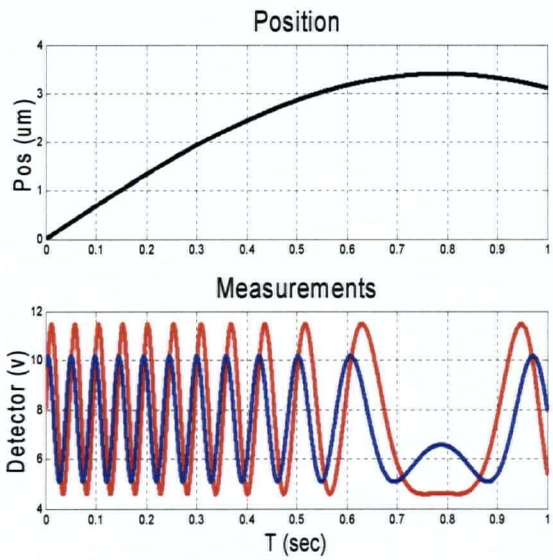

Figure 26: Characteristic Interferometric Measurement Data
Top is slow sinusoidal position vs. time. Bottom is the photodetector signals vs. time for typical parameters.

With only two detectors, it is desirable for accuracy that the interference phase difference between the detectors should be set close to 90°. This is the optimum phase difference for standard "quadrature" phase detection. The phase difference can be easily adjusted through various mechanical manipulations such as mirror rotation and detector distance.

4.1.3. Calibration

For two detectors, there are five calibration parameters in the model from equation 35. These are summarized below.

1	c_1	Detector 1 voltage offset (v)
2	b_1	Detector 1 amplitude (v)
3	c_2	Detector 2 voltage offset (v)
4	b_2	Detector 2 amplitude (v)
5	δ	Detector 2 phase compared to detector 1

Figure 27: Table of System Parameters

Once the five calibration parameters are known, it is quite straightforward to solve for a unique position given the two measurements. That is, there are two equations with one unknown. The redundant information is sufficient to resolve ambiguities which arise when solving individual inverse sine equations.

However, the five calibration parameters are NOT known. This causes difficulty in designing a detection algorithm.

The principle design requirement is to calculate a continuing series of position measurements from a continuing series of light-intensity detector-pair readings. This sounds very straightforward. As a static analysis of lab data, it would be quite simple.

Given a measurement data set, it is straightforward to calculate, a-posteri, a position data set. Firstly, the measurement data set could be plotted to establish regions of time where the position is exploring a large-scale trajectory. If such a time span exists where the position is traversing a region of the order of a half wavelength, then a first estimate of parameter values can be made simply by extraction of detector maximums and minimums. These first estimates can be refined using a linearized model centered on the initial estimates and then very finely resolved with a least-squares fit.

Such a system has limitations if the parameters are not static but slowly varying. Standard analysis practice would then be to sub-divide the sample into time-slices small enough to

regard the parameters as constant. Within these slices, parameters could be independently fit. Statistics relating parameter change rates could be determined and a position data set could be generated for the entire time period. Of course, again, each time slice must be analyzed using data which explores a sufficient region of the sinusoid to enable a fit; and this is just not certain.

Other design teams have had excellent success with this problem using multiple detectors (>2) and an explicit calibration procedure[7]. These algorithms are composite in that the position is actually controlled and thus manipulated to insure a non-stalled state during calibration. This relaxed constraint allows a statistically sound calibration phase to operate with statistically predictable error.

4.1.4. Additional Algorithm Design Criteria

I wanted to explore a different approach which allows for some improved operational criteria.

Additional criteria are:

- No explicit calibration procedure.
- Minimal assumptions about calibration parameter values at startup.
- The calibration parameter estimates should monotonically improve in accuracy from startup.
- The algorithm should be capable of implementation on a moderately priced DSP processor as of 2003 vintage with an algorithm iteration / data rate of about 5kHz. This precludes large dimensional array inversion procedures.
- The calibration parameter estimates must not degrade even if the position remains essentially constant for long and unpredictable times.

The algorithm must be capable of achieving a good estimate from an initial extreme condition of ignorance. The physical limits are:

- The amplitudes and offsets are less than the saturation voltage and more than zero.
- The phase angle between 0 and 180 degrees since another valid selection between 180 and 360 degrees will simply result in a velocity inversion of any calculated position.

The final design criteria expressly allows a stalled trajectory. That is, the algorithm has no control nor a-priori knowledge of whether the trajectory is stalled. Nor does it have any knowledge about the length of the stalled time interval. Clearly, the algorithm can never resolve a position if the trajectory is permanently stalled. Therefore, the assumption is that at some time the trajectory WILL explore a reasonable expanse. The algorithm must be prepared to wait an indefinitely long period of time for this to occur.

4.2. Geometric Parameterization

In finding a solution to the dilemma presented by startup combined with a stalled trajectory, the principle issue is to find a system which will converge correctly from any starting location. For non-linear systems in general, this entails understanding local minima to which a system might become attracted. Alternatively, if a system has a simple minimum then the other aspects of the design become tractable.

In this design there is an easy-to-visualize system model which is geometric. The origin for this idea came about from observing a plot of the two detector signals on an oscilloscope plotted against each other in X-Y mode. The position was free moving at the time and definitely not stalled. The oscilloscope was presenting a Lissajous ellipse pattern. Every half-wavelength motion of the position resulted in one traverse of the ellipse in the pattern. For a phase difference of 90 degrees and equal amplitude parameters between the two detectors, such a Lissajous pattern traces a circle and the position is simply proportional to the generating vector angle.

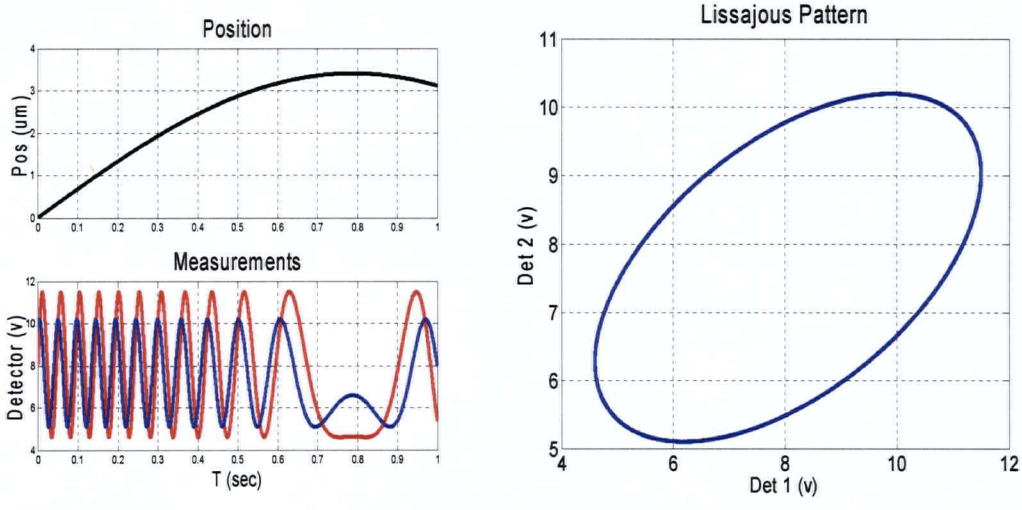

Figure 28: Lissajous Pattern

Graph above on the right shows the Lissajous pattern for the same characteristic data as figure 26. The pattern is actually traced once for every half-wavelength position change (this is not evident in the plot). Data from figure 26 is replicated on the left for reference.

The main idea of the new algorithm can be simply stated:

- The geometrical model plots the two measurements against each other; then normalizes the measurements into a circle. The position is then taken as the arc-tan of the normalized measurements.

At this point it may not be clear why this has any advantage to the more straightforward model given by the direct parameterization of equation 35 (pg 39). The answer to this lies in

the fact that the parameters are easy to visualize (illustrated subsequently) and furthermore, iteration techniques can be designed which always converge for an easily defined starting condition.

4.2.1. Geometric Parameterization Validity

There are three questions which must be answered to verify the geometric parameterization's validity:

1. How are the measurements normalized into a circular Lissajous pattern?
2. What are the parameters of the geometric parameterization?
3. Does the normalization procedure map position to angle proportionately for all positions and angles?

(1) How measurements are normalized:

A linear mapping can be easily achieved by first operating on the measurements to remove the offsets so that the ellipse pattern is centred at the origin. Subsequently, a general matrix operation can be applied to circularize the pattern. Any anti-symmetric component to this matrix would only serve to rotate our measurement vector and is therefore not wanted or required. Hence the matrix can be assumed to be a symmetric operator.

(2) Parameters of the geometric model:

The geometric parameterization is best illustrated in matrix form as follows:

$$p(\phi+\psi) = \mathbf{S}(v(\varphi) - c) = \mathbf{S}r(\varphi) \qquad \text{Eqn(36)}$$

where we define

$$r(\varphi+\psi) = v(\varphi) - c \qquad \text{Eqn(37)}$$

Where p is the voltage vector (1x2) mapped onto a unit circle, v is the measured voltage vector (1x2) in volts, $\mathbf{S}$ is a symmetric matrix operator used to circularize the Lissajous pattern, φ is the absolute phase due to mirror position, ψ is a phase matching constant, c is the centre of the Lissajous pattern (1x2) in volts. Also, where we introduce r as a centre adjusted voltage measurement.

The phase parameter (our inferred distance indicator) is explicitly related to p as follows:

$$p(\theta) = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \qquad \text{Eqn(38)}$$

(3) Equivalence of Geometric Angle & Physical Light Phase:

The geometrical fact that a symmetric matrix operation can be used to circularize an ellipse does not guarantee that the geometric angle and the physical light phase are identical. This needs to be demonstrated.

There is an analogy argument to intuitively explain the validity of the geometric parameterization. This argument relies upon a familiarity with light polarization. I must credit discussions with Ali Mohazab[13], who identified this argument in discussions held in April of 2003 at the University of British Columbia.

Elliptically polarized light may be defined to have the same parametric representation of electric field as our intensity functions have of voltage given by equation 35 (pg 39) with two exceptions:

- Elliptically polarized light does not have an offset parameterization (c_i in the equations).
- The fundamental phase of light is driven by time whereas our system phase is driven by position.

We can neglect the first concern because we will subtract the offset from our system before we take any arc-tan position measurements. The second point is critical to the argument as follows.

> Elliptically polarized light can be considered the sum of left and right-polarized light. In circularizing the ellipse, we are isolating one of the left or right polarizations. That is, in applying a linear operation to circularize the ellipse, you are simply taking the projection of your polarization onto either pure left or pure right circular polarizations. A circular polarization guarantees constant angular speed because that is part of its fundamental definition. Constant angular speed for the polarized light implies proportional position versus angle for our problem.

Alternatively, and less abstractly, we can show explicitly that the geometric parameterization maps onto the direct parameterization as follows:

-Expand **S** into explicit parameters A,B,D
-Substitute the initial model from equation 35 (pg 39) into r.
-Expand p into components.

$$\begin{pmatrix} A & B \\ B & D \end{pmatrix}\begin{pmatrix} b_1 \sin(\varphi) \\ b_2 \sin(\varphi+\delta) \end{pmatrix} = \begin{pmatrix} \sin(\varphi+\psi) \\ \cos(\varphi+\psi) \end{pmatrix} \qquad \text{Eqn(39)}$$

Where b_I are the measurement amplitudes, A,B,D are the symmetric tensor (**S**) parameters, φ is the system phase, δ is the detector phase difference and ψ is the arbitrary phase match.

Since this equation can be solved in general, the parameters which circularize the pattern are ALSO shown to meet the mapping rate criterion of question (3).

The general solution is:

$$A = \frac{b_2 + b_1 \sin(\delta)}{b_1 \sin(\delta)\sqrt{b_1^2 + 2b_1 b_2 \sin(\delta) + b_2^2}} \qquad \text{Eqn(40)}$$

$$B = \frac{-\cot(\delta)}{\sqrt{b_1^2 + 2b_1 b_2 \sin(\delta) + b_2^2}} \qquad \text{Eqn(41)}$$

$$D = \frac{b_1 + b_2 \sin(\delta)}{b_2 \sin(\delta)\sqrt{b_2^2 + 2b_2 b_1 \sin(\delta) + b_1^2}} \qquad \text{Eqn(42)}$$

$$\cot(\psi) = \frac{b_1 + b_2 \sin(\delta)}{b_2 \cos(\delta)} \qquad \text{Eqn(43)}$$

Incidentally, the solutions above can be used to verify values for the symmetric matrix S if parameters from equation 29 (pg 32) are known.

4.2.2. Parameter Iterations in Geometric Model

From the previous section we know explicitly how to convert from a direct parameterization to the geometric parameterization. The geometric model allows for geometric visualization of parameter iteration. However, the motivation for designing the geometric model of the system is that it must be directly derivable from the measurement stream.

There are two conceptually separate stages to obtaining the geometric parameterization from the measurement stream. The first involves "circularization" of the Lissajous pattern (derivation of the **S** matrix). The second involves finding the centre of the Lissajous pattern (derivation of the *c* vector).

Operationally, we need a method which generates our measurement without too great a computational load since we would like to implement this algorithm with a moderate power DSP (digital signal processor). Fewer number of computations per measurement cycle will result in a better algorithm. Certainly, we would like the number of computations to be on the order of less than 100 floating-point additions and multiplications.

The algorithm will have a relatively high measurement cycle frequency (~ 10 kHz) and thus an iterative scheme with only a small improvement per cycle could still provide a highly effective estimate of the geometric parameters. Functionally, an iteration method which provides some small adjustment to parameters will perform well as long as the iteration adjustment can be shown to be in the "correct" direction.

Symmetric Matrix Operation:

If we assume the centre of the Lissajous pattern is exactly correct then we have previously shown that there is some matrix **S** which correctly converts the elliptical Lissajous pattern into a unit circle.

If we are given just one measurement (pair of detector voltages), what could we do to improve our estimate of **S**? If our single measurement is outside the unit circle, we know that our correction **S** must decrease signals in our measurement's direction. If our measurement is inside the unit circle, we know that **S** should apply a correction to increase signals in our measurement direction. For the direction orthogonal to our measurement, we haven't gained any information and we shouldn't adjust the matrix **S** in that direction.

An operation to achieve this result from a single measurement is shown below:

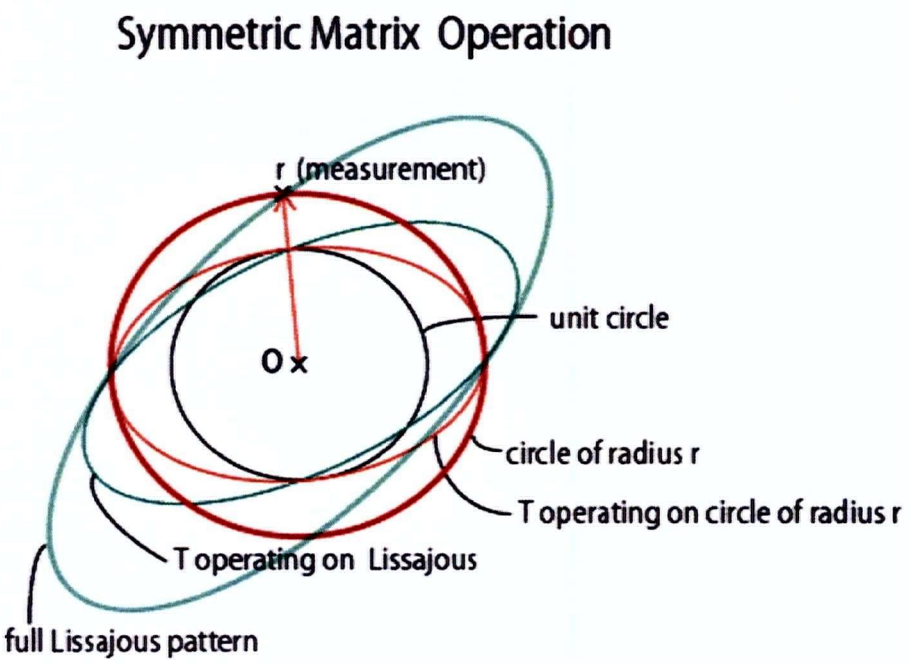

Figure 29: Symmetric Matrix Operation

The operation of the operation **T** as calculated from a single measurement r is shown on two different patterns. The red ellipse shows how **T** operates to force the circle of radius r onto the unit circle in the direction of r. The large green ellipse is a sample Lissajous pattern. The smaller green ellipse shows the operation of **T** operating on the lissajous pattern. Again the new pattern is constrained to intersect the unit circle in the direction of r.

In the above diagram, the symmetric operator **T** has an eigenvector in the direction of r with an eigenvalue of $1/|r|$. The orthogonal eigenvalue is unity. That is, we have no information in that direction. Clearly to gain complete information, measurements are required in more than a single direction.

A symmetric operator with an eigenvector of value α in the direction of $r(x,y)$ and an orthogonal eigenvector of unity is given by:

$$\mathbf{T} = \frac{1}{x^2+y^2}\begin{pmatrix} \alpha x^2 + y^2 & (\alpha-1)xy \\ (\alpha-1)xy & x^2 + \alpha y^2 \end{pmatrix} \qquad \text{Eqn(44)}$$

Where **T** is the symmetric operator, α is the eigenvalue of operator in direction of $r(x,y)$ and is set in the diagram to be $1/|r|$, x is the x value of point r onto the unit circle and y is the y value of $r(x,y)$. The formula above can be simply demonstrated by showing $\mathbf{T}\begin{pmatrix} x \\ y \end{pmatrix} = \alpha\begin{pmatrix} x \\ y \end{pmatrix}$ and $\mathbf{T}\begin{pmatrix} -y \\ x \end{pmatrix} = \begin{pmatrix} -y \\ x \end{pmatrix}$.

Explicit S Iteration:

In operation, the full correction **T** should not be applied because the Lissajous pattern is simultaneously being adjusted to find its centre and this can cause some instability. Additionally, a smaller correction per cycle will reduce the effect of measurement errors.

A description of the iteration with a small gain factor is as follows:

$$\mathbf{S}_0 = \mathbf{I} \qquad \text{Eqn(45)}$$

$$\alpha_i = \left(\frac{1}{|r_i|}\right)^{1/n} \approx 1 + \frac{r_i^{-2} - 1}{2n} \qquad \text{Eqn(46)}$$

$$\mathbf{t}_i = \frac{1}{r_i^2}\begin{pmatrix} \alpha_i x_i^2 + y_i^2 & (\alpha_i - 1)x_i y_i \\ (\alpha_i - 1)x_i y_i & x_i^2 + \alpha_i y_i^2 \end{pmatrix} \qquad \text{Eqn(47)}$$

$$\mathbf{S}_i = \mathbf{t}_i \mathbf{S}_{i-1} \qquad \text{Eqn(48)}$$

Where **S** is the symmetric operator, α is the eigenvalue of operator in direction of $r(x,y)$, r is the measurement vector with components (x,y), **t** is the symmetric iterant. The symmetric operator **S** is initiated with the identity operator. Each iteration multiplies **S** with the iterant operator **t**. The convergence rate is set by n where as n approaches infinity, **t** approaches the identity matrix and as n approaches 1, **t** applies a full circularization to each measurement forcing r onto the unit circle in one iteration.

Note that if r is on the unit circle, **t** is the identity matrix and thus **S** is stationary. Thus if **S** is the correct value to "circularize" the Lissajous pattern, the value of **S** will remain stationary. This is a necessary condition for convergence.

Since this iteration requires just one measurement, adjustments to **S** take place even when the trajectory is stalled. However, overall system performance is not compromised.

A sample of the symmetric matrix iteration scheme operation is shown in the figure below.

The non-stalled trajectory is causing the raw Lissajous pattern to traced as a track clockwise and counter clockwise around the ellipse (outer ellipse in red). Subsequent iterations to the symmetric operator S cause the modified trace to spiral inwards towards a unit circle. After only a few full oscillations of the Lissajous pattern, the modified pattern is confined to the unit circle with negligible error. Convergence is monotonic and trend is exponentially asymptotic. Since the pattern is converging to a unit circle, you can see the actual trajectory reverse direction as the platform velocity reverses direction.

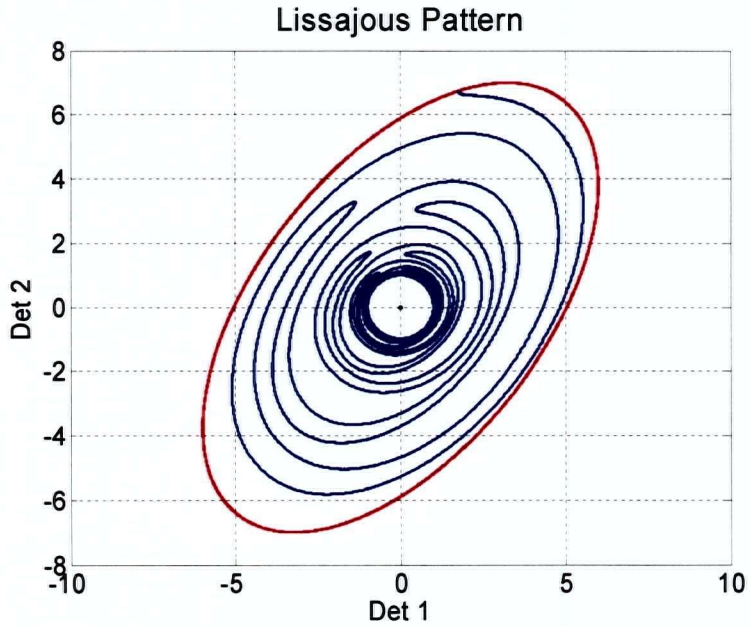

Figure 30: Circularization

The direct detector measurements are shown in red and describe the usual Lissajous pattern. The blue trace represents the modified trace as S iterates such that the pattern becomes a unit circle.

Finding the Lissajous pattern centre:

A first order iteration to converge towards the centre of a circle from a single new measurement is as follows: pull the existing estimate of the centre along the vector between the estimate and the measurement to the distance of the known radius. This is illustrated in Figure 31 (below).

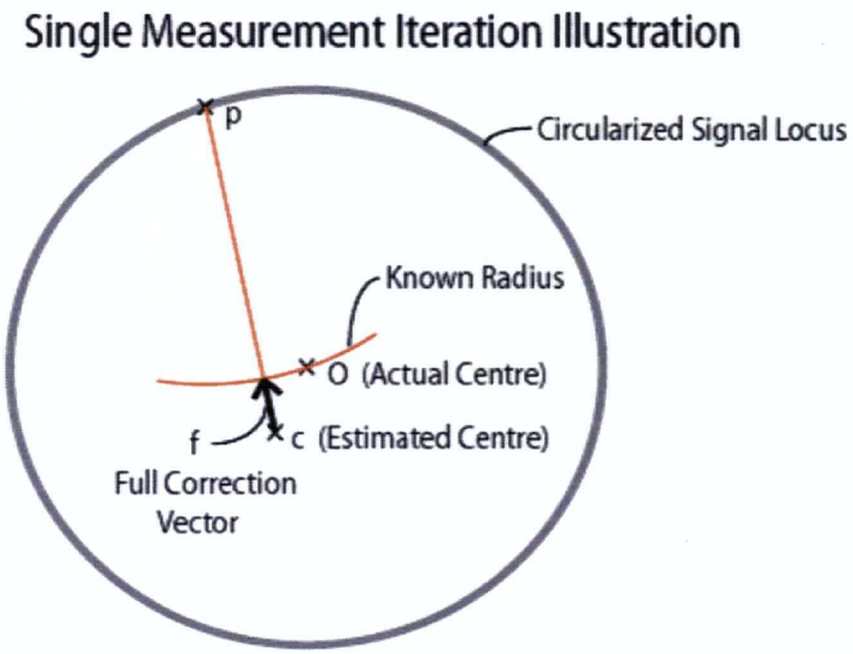

Figure 31: First Order Iteration to Find Centre

This figure illustrates graphically a construction that might be used to calculate a correction to an estimated centre given a single measurement. Only a small percentage of the correction vector is applied for each measurement. This scheme converges rapidly for estimates within the pattern circle.

A fatal problem with the scheme is if the estimated centre is in great error and OUTSIDE the Lissajous pattern, then the scheme does not converge at all. Furthermore, the scheme adjusts the centre in the direction of the measurement. This is the same direction of adjustment as the symmetric matrix iteration. Since both algorithms have to operate simultaneously, they must behave as independently as possible. They should not operate in the same direction. We have to find an iteration which adjusts the centre estimate in a direction orthogonal to the measurement direction.

Orthogonal Iteration to find pattern centre:

An iteration involving two subsequent measurements can be derived which makes adjustment to the estimated centre in a direction orthogonal to the single measurement iteration. This is illustrated in Figure 32 (below).

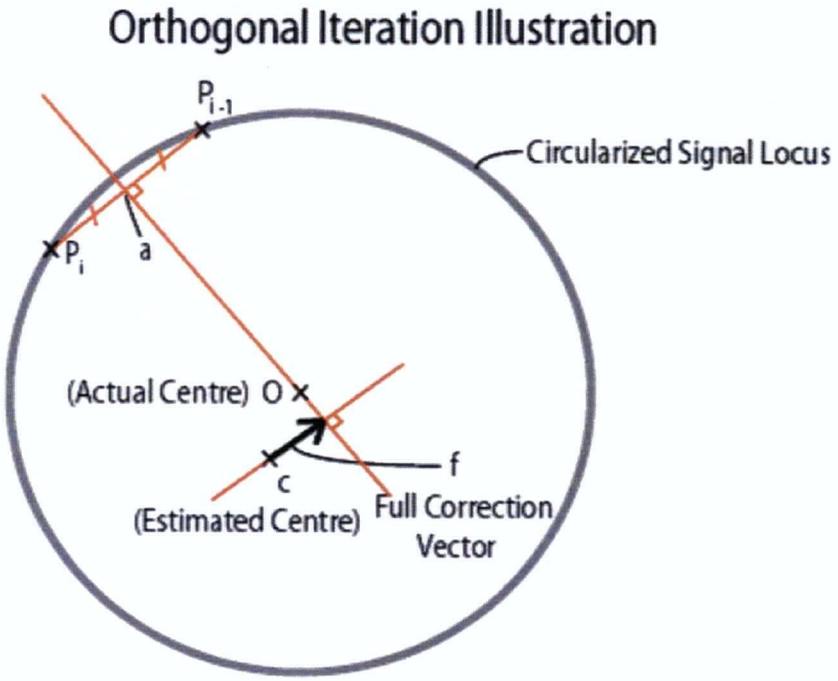

Figure 32: Orthogonal Iteration to Find Centre

The illustration for the orthogonal iteration scheme illustrates graphically the construction used to calculate a correction to an estimated centre given a two measurements. Only a small percentage of the correction vector is applied for each measurement. This scheme converges correctly for all initial estimates inside or OUTSIDE (or actually on) the pattern circle. Furthermore, the pattern radius is not required.

The correction vector can be found with vector arithmetic to be:

$$d = p_i - p_{i-1} \qquad \text{Eqn(49)}$$

$$a = \frac{p_i + p_{i-1}}{2} \qquad \text{Eqn(50)}$$

$$f = \frac{((a-c)\cdot d)d}{d^2}$$ Eqn(51)

Where d is the difference vector, a is the mid-point between measurements, c is the current estimated centre and f is the full adjustment vector.

Real data will have errors. When $|d|$ is small the direction of d is poorly defined. This is a common condition which will happen whenever the mirror position is nearly stationary. The error in the correction vector f is proportional to the measurement error and inversely proportional to the distance between measurements (length of d). Thus we should weight the correction proportional to the inverse of the error squared which is a measure of the inverse variance of the adjustment. Thus the weighting for the correction can be achieved by multiplying by d^2. This makes the algorithm even more efficient since the derivation above involves division by d^2. Elegantly, we simply do not divide by d^2 in equation 51.

Additionally, as in the case of the symmetric matrix iteration, only a small portion of the weighted correction should be used to aid in accuracy against measurement noise and to aid in stability against algorithm interaction as both algorithms act simultaneously.

This orthogonal iteration scheme meets all of the criteria required for our algorithm:

- The correction vector estimate monotonically improves the centre estimate no matter where the initial estimate is located.
- The radius of the Lissajous pattern is not required for the iteration scheme.
- If the Lissajous pattern is non-circular, the centre estimate will not converge correctly to the pattern centre however it will not disastrously diverge either.
- When weighted the adjustments are very small for stalled trajectories.

The operation of the orthogonal iteration can be seen in the figure below where the measurement stream reflects a circular Lissajous pattern not centred at the origin.

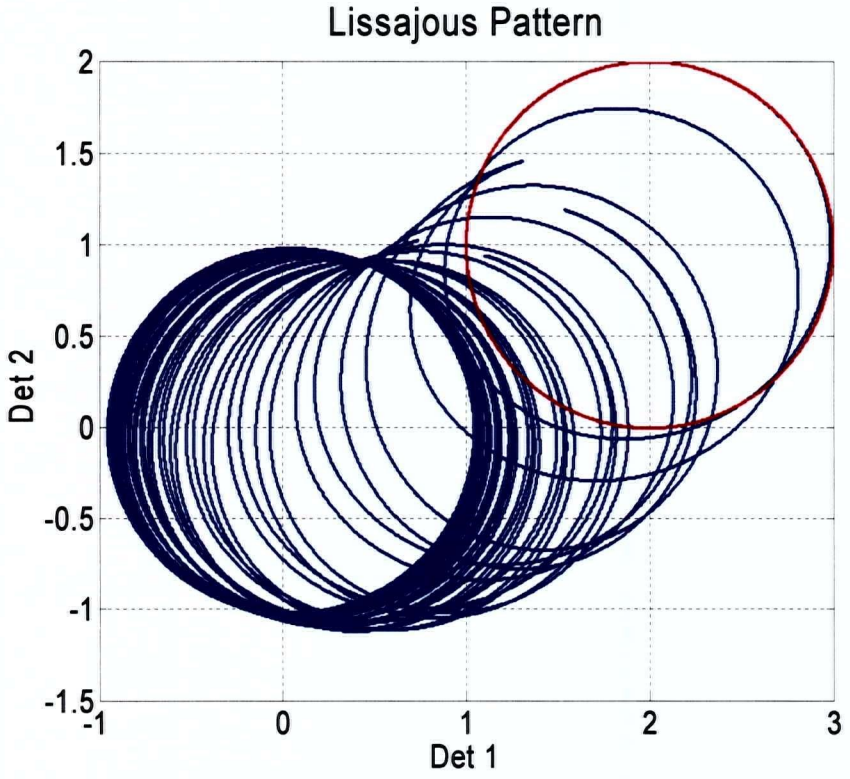

Figure 33: Centre Finding Operation

The direct detector measurements are shown in red and describe a circular Lissajous pattern which is NOT centred at the origin. The blue trace represents the modified trace as the centre finding iteration operates such that the pattern becomes centred at the origin

Assembled Algorithm:

The overall operation of the system requires that each algorithm iteration performs both operations: an adjustment to the symmetric matrix S and an adjustment to the centre.

Note that when the symmetric algorithm and the centre finding algorithm are operated sequentially in a single measurement iteration, the centre adjustment algorithm is NOT receiving data guaranteed to lie on a circle. Any error in **S** will allow some eccentricity to the Lissajous pattern. Simulations support this adulteration in two ways. Firstly, the centre finding algorithm actually works quite well on elliptical Lissajous patterns as long as the domain is well explored. Further, the centre finding algorithm is well behaved (does not diverge) even when the domain is not well explored (stalled position). Secondly, under normal operational limits of a domain from 0 to 10 volts with an eccentricity greater than 0.1, the symmetric operator will almost always reduce eccentricity of the ellipse even when the centre is poorly estimated (or even outside the locus of the pattern).

Thus simulations and operational tests which demonstrate convergence indicate a strong degree of independence between the symmetric iteration and the centre-finding iteration even when coupled as above.

Effective operation of the algorithm require that the gain factors be set to iterate only very small steps per iteration. Correct operation of actual simulations indicate gain factors on the order of one part per 100000 provide a good margin against instability between the two interacting algorithms while still providing a reasonable convergence rate. Gains on the order of one part in 10000 did on occasion demonstrate instability.

The final calculation of the position involves taking a four quadrant arctan of the p vector. This operation can be performed simply if your system has a 4-quadrant arctan function. In our case, our arctan function required logic to identify quadrant. In either case, the number of rotations must be accumulated to integrate a total distance result.

4.2.3. Algorithm Tests

Unfortunately, there wasn't any independent position measurement of sufficient accuracy available to measure the accuracy of the algorithm-interferometer setup. Initial tests and expectations from previous work Gray et al [6] indicated we could expect errors on the order of less than one nanometer. Taking a tip from Gray et al, we decided to use duplicate detector sets to determine accuracy. That is, we ran two independent detector pairs/algorithms and compared the measurements between the two systems to determine accuracy.

Two independent pairs of detectors and algorithms were then used to infer position information with respect to the measurement arm. In comparison of these two measurements, some inference can be made as to the accuracy and precision of the detector system and the algorithm.

For our tests, we acknowledge that systematic errors in common between our two detector pairs/algorithms will not be illuminated. However, the interferometric phase difference between the two systems was significant as the detector pairs were physically separated in the interference pattern. This should ensure that systematic errors due to geometry (geometry in the sense of the geometrical interpretation of the algorithm) will not be in-phase correlated and should manifest themselves in the measurements.

The end reflection mirror in the reference arm was fitted with a piezo which would vary the reference arm distance of our Michelson interferometer. A triangular waveform was fed to this piezo which provided a dynamic measurement for our detector system. Optimal tuning included adjusting interferometer geometry such that the intensity amplitude for all four detectors exceeded 50% of full scale. Further, 90 degree phase separation between detectors in each pair was tuned as well as possible by eye with only the Lissajous pattern as a guide.

The algorithm/data loop rate was 5KHz. Detector digitization was conducted for each pair using two D/A converters allowing simultaneous readings from both detectors. D/A digitization was 16 bit over a -10V to 10V range. Mechanical tuning was specified to provide a maximum to minimum swing of greater than 4V and a maximum of greater than 6V. For the most part this was achieved; although obtaining a setup which most closely met these specifications for all four required detectors required some signals to be marginally below these levels.

The position measurements are shown below. Only one trace is shown because the two systems/algorithms matched to such a degree that the difference could not be seen on the full scale plot. Also shown below is the difference between the two systems/algorithms.

Overall, the RMS difference between the two systems was 0.22nm over a 120 second interval test. For random correlations, this would imply an RMS error for a single system of 0.16 nm. Clearly there is an element of the difference-plot which is harmonically related to the position phase. This is indicative of the magnitude of the systematic error for the system/algorithm. It is larger than the random fluctuations visible on the same plot and thus dominates the RMS result.

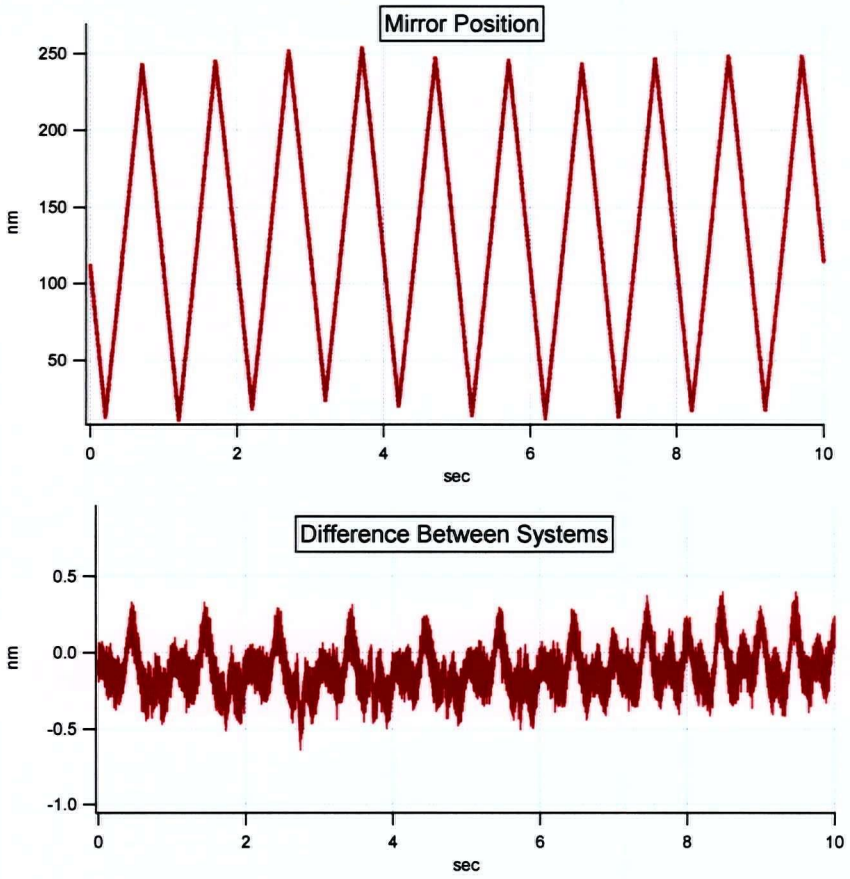

Figure 34: Interferometer and Algorithm Test

The top figure shows the position result from a 2Hz triangle test signal applied to a test mirror. Both systems plot exactly the same on the scale shown here. The bottom figure shows the signal difference between two separate systems.

However, the dominant error is clearly **not** randomly correlated. The actual error of a single system will depend upon the correlation between systematic errors on the two systems. This is unknown. However, the physical separation between the two systems' detector pairs should have ensured a phase shift (phase now referring to the test mirror position triangle

wave) between the two systems and thus lead us to believe that the overall error for a single system is certainly on the order of 0.22nm. That is, we do not believe that there is a strong in-phase correlation between the systematic errors (in phase with respect to the triangle test signal).

When the systems were not tuned optimally, the RMS difference between the signals varied between 0.3nm and 1.5nm. This non optimal tuning included voltage amplitudes of less than 2 volts, phase differences of as little as 40 degrees between detectors and low signals of less than 30% of maximum.

It should also be noted that RMS results were uniformly above 1nm for **any** mechanical tuning unless a shroud was placed over the assembly to block air currents.

4.2.4. Non-Correlated Noise Test

An estimate of the white noise level of the system was generated by analyzing the magnitude of the difference between subsequent positions as output by the algorithm when the mirror piezo was held at a fixed voltage. A sample of this data is shown below:

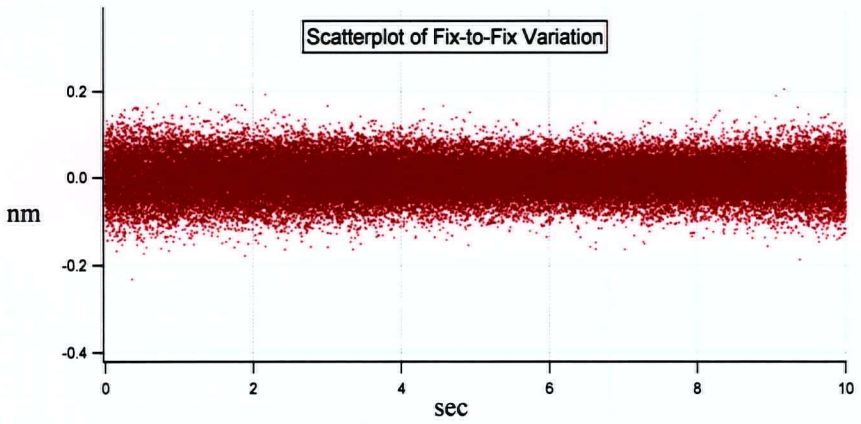

Figure 35: Non-correlated Noise Test

A measurement of the non-correlated noise was taken by subtracting the position fix from its subsequent value. This particular sample was taken from a static system test with no applied mirror motion. However, plot appearance is identical to within expected sample statistical variance for dynamic mirror motion.

This figure is taken from a test run with no applied mirror motion. However, analysis of the dynamic data (from previous test) show the same RMS result of 0.07nm.

Under the assumption that fix-to-fix variations are uncorrelated, the single fix RMS noise can be inferred to be 0.05nm ($0.07/\sqrt{2}$).

5. Control Tests

5.1. Platform Response

As discussed in the apparatus section 3.5, Position Control Test Platform, we undertake to control a macroscopic mass (approximately 10 Kg) of aluminum along one degree of freedom in an attempt to "lock" the mass to ground motion. To accomplish this, we need to understand the response of the platform to our control piezo, and to understand the ground motion to which we are trying to respond.

To obtain the system response in isolation, we must ensure that the vibration signals from ground motion do not influence our measurement. We achieved this in two separate tests.

In one test, we excite our piezo actuator with a 50 nm sinusoidal signal. We use a lockin-amplifier setup and sweep the frequencies from 1 Hz to 1000 Hz in 0.5 Hz steps. Each individual frequency test was held for 30 cycles or 5 seconds (whichever is greater) before sampling begins to allow any transients to damp away. Then a 10 second average response is measured providing an explicit amplitude and phase response (with an approximate 0.1 Hz frequency discrimination). The lock-in nature of the measurement ensures that only the frequency component exactly matching the excitation frequency will be measured. This reduces measurement effects from ground vibration since ground vibrations contributions are confined to bandwidths of the order of 0.1Hz. This system measures response with an error independent of frequency.

Another test involves taking an average response to a step (square-wave) excitation. In this case the average was taken over 5 hours. Then a Fourier transform is taken of this average response. Taking odd harmonics only (since a square wave has no even harmonics) and dividing by the theoretical square wave transform results in the Bode plot shown in the (thin) red trace in the figure. This method has an error proportional to the frequency squared. The error at the high frequency end is clearly visible in the fluctuations of the red trace.

Both methods provide essentially the same measurement of system response shown in figure 36 following.

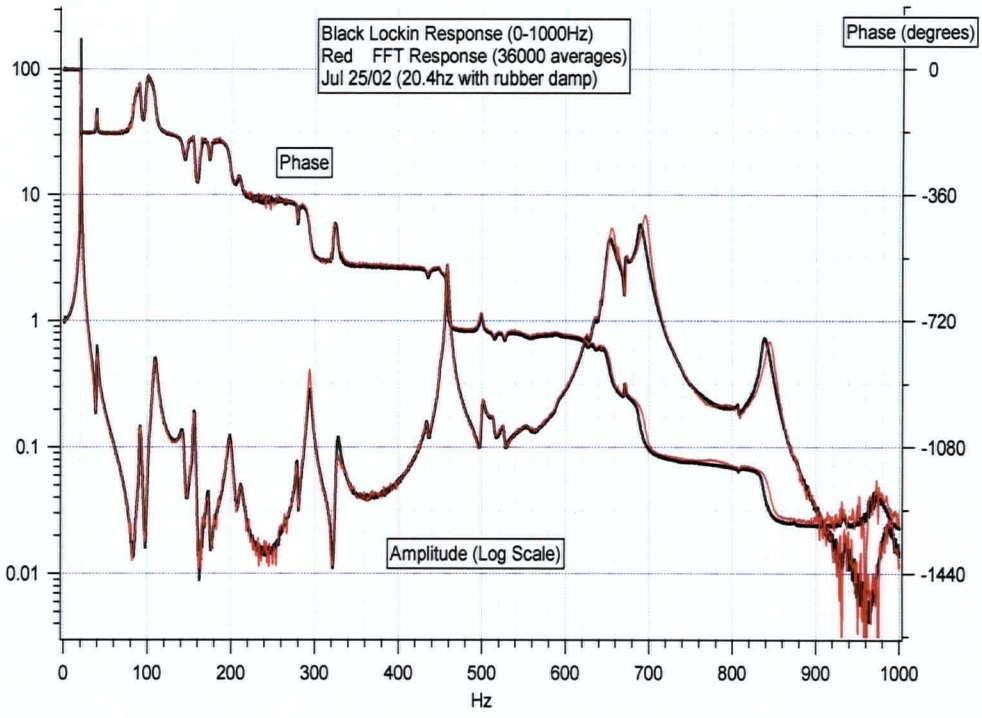

Figure 36: System Bode Plot

The Bode plot for the horizontal position control system is shown in this figure. Thick (black) traces represent a lock-in data acquisition and the thin (red) line represents a Fourier transform of an average square-wave response. These two independent methods conform very nicely.

The system as designed presents a control-system challenge since there are obviously many resonant modes in the system. For example, notice the greater-than-unity amplitude response not only at the slow fundamental resonance of around 20 Hz, but also other resonances at 460 Hz and 650 to 700 Hz. These higher frequency resonances are easily excited.

The complexity of the platform response and in particular the highly excitable resonances could be resolved with a modified mechanical design (refer to Appendix A, NLC Vibration Control Editorial). Notwithstanding, actual NLC design might easily encounter difficult to control mechanical systems and it is instructive to attempt to control such systems. As such we continue in this chapter to examine some rudimentary control strategies for this test platform.

5.2. Ground-Motion Characteristics

The frame and bulkheads of our test system rest in direct contact with the ground and vibrate in response to vibrations in the ground itself. Without control, our test system measurement indicates a vibration level averaging about 50 nm RMS. This is shown in the figure below.

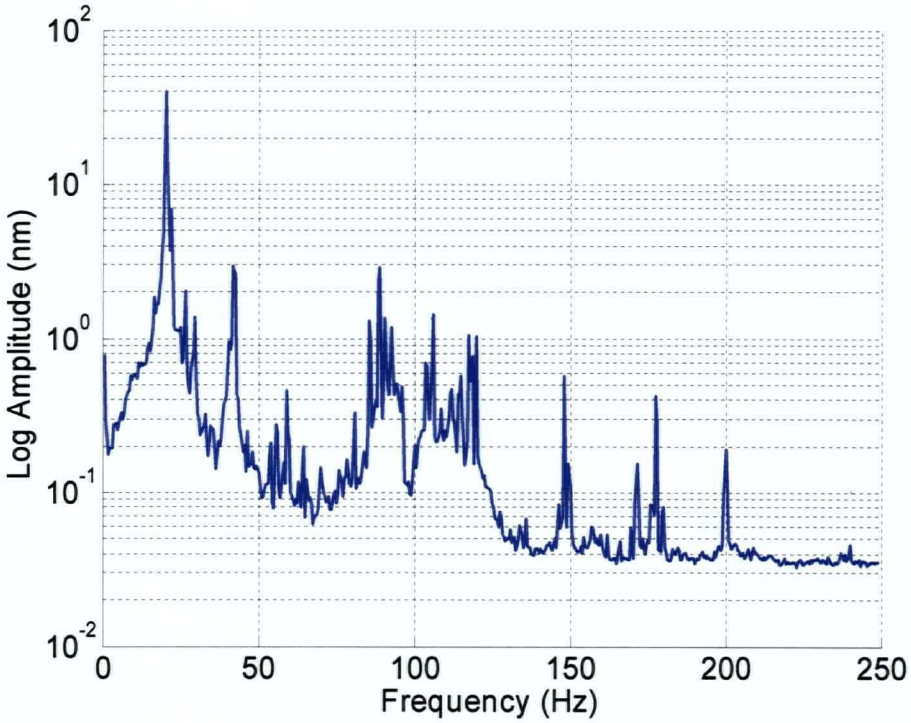

Figure 37: Uncontrolled System Vibration Spectrum

The vibration spectrum is calculated from a time data series of 1,000,000 samples over 300 seconds. DFT's over four second intervals of the time series are calculated. This results in frequency values spaced at 0.25 Hz intervals. Multiple DFT's are calculated starting in the time series at intervals of every 0.15 seconds. This results in 1974 overlapping DFT regions which are subsequently averaged as to their power spectra. The individual DFT's are calculated using a Hanning window and thus have an approximately .5 Hz bandwidth (2 times 0.25 due to Hanning). Once combined into an overall average power spectrum, alternate values are again averaged to produce a power spectrum with values spaced at 0.5 Hz with a bandwidth also of approximately 0.5 Hz.

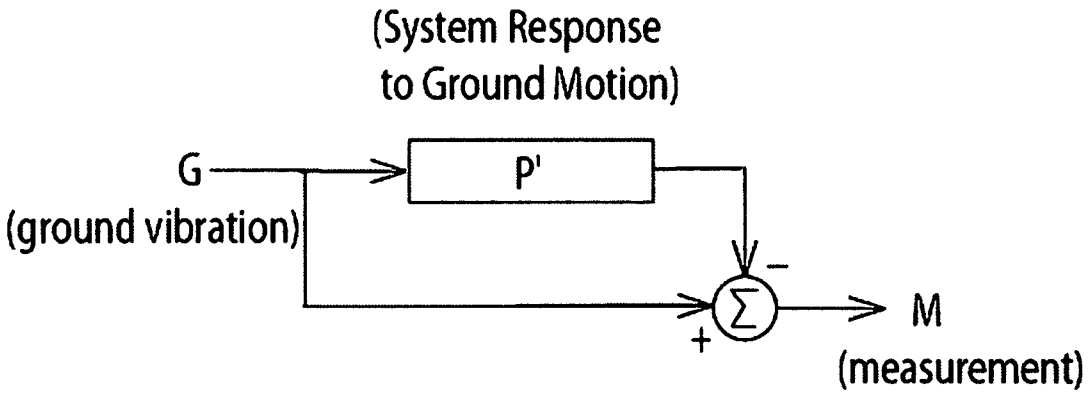

Figure 38: Measurement Model (No Control)

Our system measurement measures the distance between the bulkhead and the test mass providing the difference function..

The system vibration spectrum shows two effects of interest. Firstly, the spectrum shows some very steep peaks which represent highly coherent man-made vibrations from rotating equipment. In particular, 60 Hz, 90 Hz and 120 Hz vibrations are apparent which clearly indicate synchronous vibrations to the North American power standard of 60 Hz. Secondly, the spectrum shows its most pronounced vibration at about 20 Hz. This is the fundamental resonance of our system. Thus this spectrum shows a combination of actual ground motion and our system response.

The power spectrum of the actual ground motion exciting the system is unknown. However, we can arrive at a good approximation of a ground motion component in the direction of our single degree of freedom. To do this we appeal to the system diagram below.

Ground vibration G moves the measurement bulkhead directly and also acts through the system response P' to move the test mass. We measure the difference in position between the bulkhead and the test mass. Assuming that the ground vibration couples perfectly to the bulkhead, we see that our measurement implies

$$M = G - P'G \qquad \text{Eqn(52)}$$

Where M is our measurement, G is our ground motion and P' is the system transfer function. If the test mass was perfectly attached to the bulkhead, P' would equal unity and the measurement M=G-P'G would be zero. If the mass were inertially floating, P' would be zero and M=G. Both these limits bear scrutiny to their logical extremes.

We trivially manipulate the equation to obtain an equation for the ground motion.

$$G = \frac{M}{1 - P'} \qquad \text{Eqn(53)}$$

But P' is the system transfer function from ground motion to test mass motion. We have no direct method of obtaining this transfer function. However, we can make the assumption that the piezo excitation transfer function (see figure 36) is very similar to P'. The piezo transfer function from figure 36 provides an excitation to the test mass identical to that of displacement of the reaction bulkhead. On the other hand, P' is the transfer function of the *entire frame* including the reaction bulkhead to the test mass. We expect that the system transfer function and P' should be identical to figure 36 at least at low frequencies where modes of the frame and bulkhead will not yet appear. Thus a reasonable assumption to obtain an approximation of the ground motion is that P' is the same as the transfer function calculated in section 5.1, Platform Response.

Thus we have a method to estimate the ground motion component in the direction of our degree of freedom through direct manipulation of system response data and uncontrolled time-sequence samples. Transfer functions are complex functions in frequency space and as such each, and every raw DFT from our time sample must be treated against the complex transfer function P before being averaged into a power spectrum for ground motion.

The following plot indicates an approximate ground motion spectrum deduced using this method.

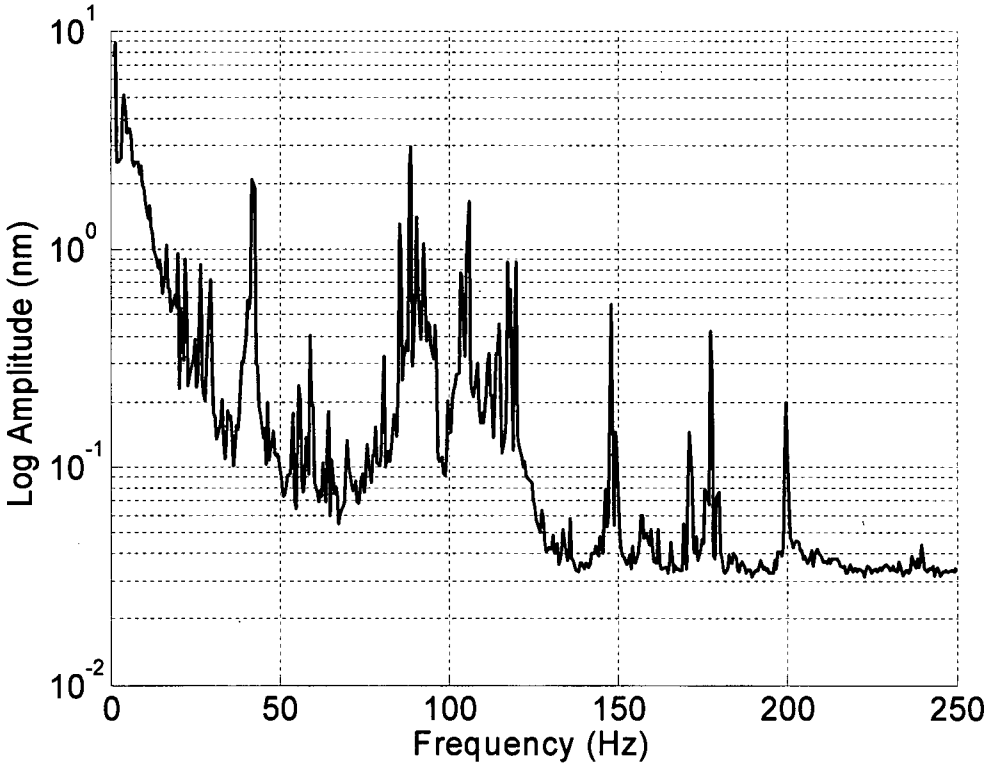

Figure 39: Ground Spectrum Approximation

This ground spectrum shows the frequency distribution for an overall 23.5 nm RMS vibration.

This procedure essentially eliminates the 20 hz peak due to the platform resonance. Additionally, there is a strong vibration indicated near 0 Hz attenuating as frequency increases which is a common indicator of natural ground vibration. These two observations support the validity of using the system response transfer function for P' in equation 53.

The approximated ground motion has an overall amplitude of 23.5 nm RMS. This is less than half of our uncontrolled system amplitude. This effect will occur for any suspended structure employed in the actual NLC project where the structure will resonate from ground vibration.

For a control strategy attempting to harden support structures, ground vibration represents the demand function or setpoint for the control system. Our test mass must be accelerated to match this vibration.

5.3. PID Control Model

The physics of our system are different from a standard application of control. In a standard application there is a setpoint signal (which is often a constant) and an output value signal. Both these signals are generally known directly. These two signals are subtracted to produce an error signal.

In our system, the setpoint and the output from P are not measured nor known directly. Rather, only the error is known (labeled M for measurement below). The setpoint can be considered to be the difference between the ground vibration and the natural mass response to the ground vibration. The desired result for our system design is to obtain a transfer function from G to M equal to zero. Normal control system design discusses the transfer function at the output of P as driven by the setpoint and strives for an ideal value of unity. These performance criteria are actually the same. That is, we are still striving for a closed loop transfer function of unity.

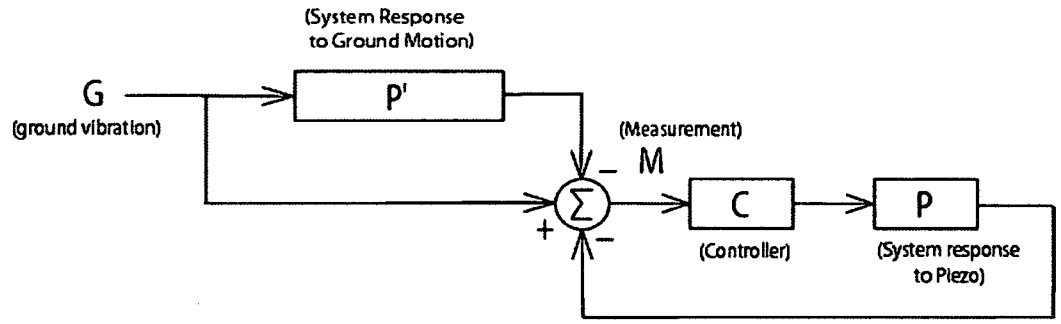

Figure 40: Control System Model

From the diagram, the measurement M can be seen to be:

$$M = G - P'G - MCP \qquad \text{Eqn(54)}$$

Where C is the controller transfer function and P is the piezo to measurement transfer function (see figure 36),

This can be manipulated to obtain the attenuation factor for the transfer function from G to M

$$\frac{M}{G} = \frac{1-P'}{1+CP} \qquad \text{Eqn(55)}$$

This factor allows us to calculate expected attenuation of ground motion to measurement.

A design criteria which sets our problem apart from most control applications is that the setpoint signal is so complicated. The setpoint is not known a priori, rather, the system is expected to track the ground motion vibrations which can be very intricate.

Initially, we coded the controller with classical PID capability. The differential signal is provided with a second order low pass filter with a configurable rolloff frequency and the proportional control is applied without a filter. Our system goal of vibration elimination has no requirement for absolute positioning and as such integral control is completely unnecessary. Although included in our control algorithms, we never use integral control.

Our system loop time is set at 5000 Hz giving a control Nyquist frequency of 2500 Hz. An examination of our system Bode plot shows that this loop time is fast enough to consider our control system from a classically continuous perspective. In support of fast loop times, steps were taken to ensure that the piezo would respond quickly to output signal variations (refer to section 3.3, Data Acquisition).

5.4. Results

Initial attempts at simple proportional control were dismal. Effectively no proportional gain at all could be applied. Turning to the system Bode plot (see figure 36) we see that the open loop system response is above unity at 450 Hz and also between 650 and 750 Hz. Phase for these regions is broad and in combination covers over 270 degrees. Proportional gains of as little as 0.2 nm/nm are enough to cause oscillation under these conditions and low gains like this of very little utility for control.

Subsequent attempts using only differential control gave moderately better performance.

With the low-pass filter rolloff set at 65 Hz, a differential gain of 9.7 ms was reached before oscillation occurred. This provided stable performance and complete elimination of the fundamental system resonance. The expected attenuation for this configuration is calculated using equation 55 (again assuming P'=P). We are striving for a response of zero. The response we obtain of greater than unity at frequencies above 25 Hz is not particularly impressive.

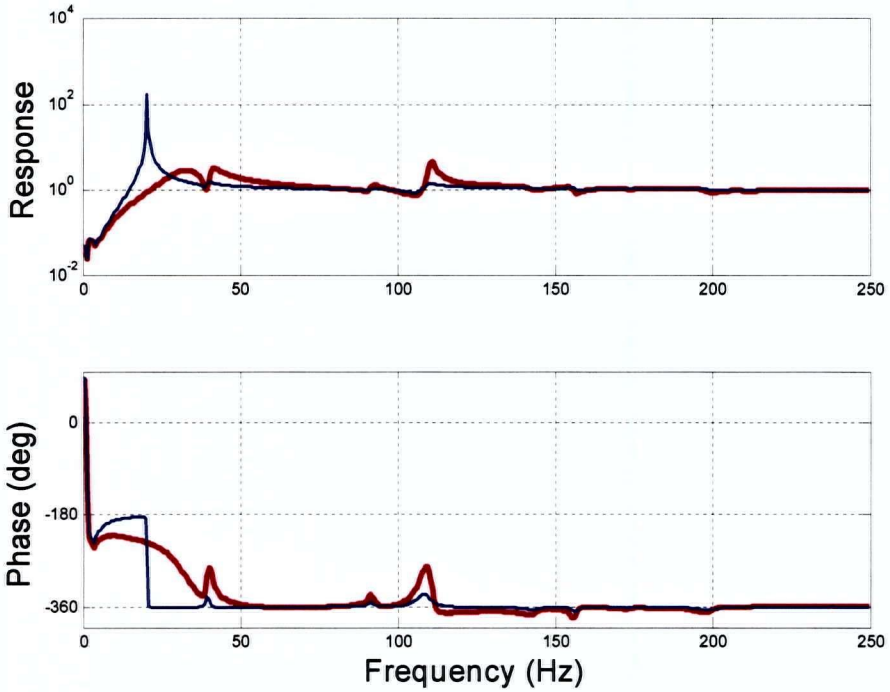

Figure 41: Model Response

The blue plot (thin) represents the system measurement response to ground vibration without any control. The red plot (thick) represents the system measurement response to ground vibration with differential control set at 97ms and a low-pass filter set at 65 Hz rolloff.

On the other hand, the application of the control system does succeed in eliminating the fundamental resonant response. In the logarithmic plot, the area under that resonant peak is significant. Imagine our system exposed to ground vibrations consisting of a white noise source with complete cutoff at 250 Hz and an RMS value of 1nm. The uncontrolled system would average 7.9 nm RMS whereas the controlled system would average 1.13 nm RMS. This would be a seven fold improvement in RMS if the ground vibrations were white noise. However, as we have seen they are not white noise.

We are now presented with an interesting opportunity to test our approximations in modeling the control system and the ground vibrations. By applying the attenuation as calculated (refer figure 41) to the ground motion as inferred (refer figure 39) we can obtain a predicted power spectrum for the system under control. We can compare this with an actual power spectrum taken from time sequence data of the system while actually controlling. These two spectra are overlaid in the plot below.

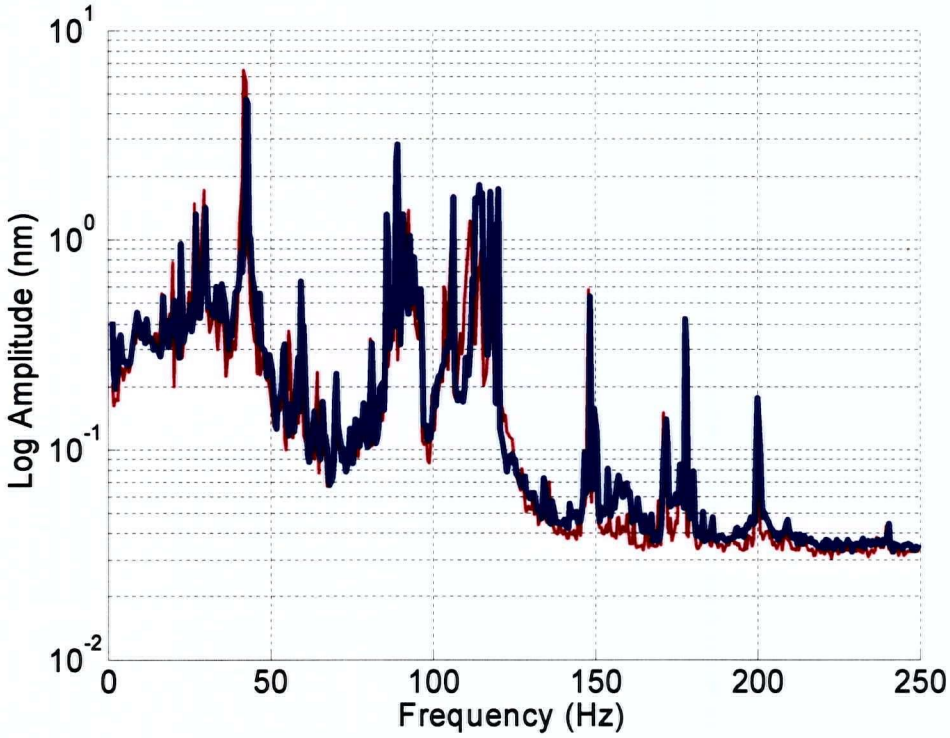

Figure 42: PID Control System Results

The red plot (thin) represents a prediction of the measurement spectrum. The blue (thick) plot represents an actual 300 second sample.

The similarities in both magnitude and spectrum artifacts is very good. Notable are certain lines such as at around 115 Hz which are in one plot but not the other. This represents rotating equipment which was on or off at the two different 300 second time samples. (The prediction is still based upon an actual time sample but of the system not under control).

Overall, even a marginal implementation of PID control has predicts the system measurement will reduce from from 48 nm RMS to 12.1 nm RMS. Simple PID implementation has reduced vibration by a factor of four from an uncontrolled system. Actual tests resulted in an RMS measurement of 11.6 nm.

5.5. Exploration of Gain Limits

But why is the gain so limited? Why can we not get a better performance with our standard controller? The following enlarged Bode plot shows the difficulty.

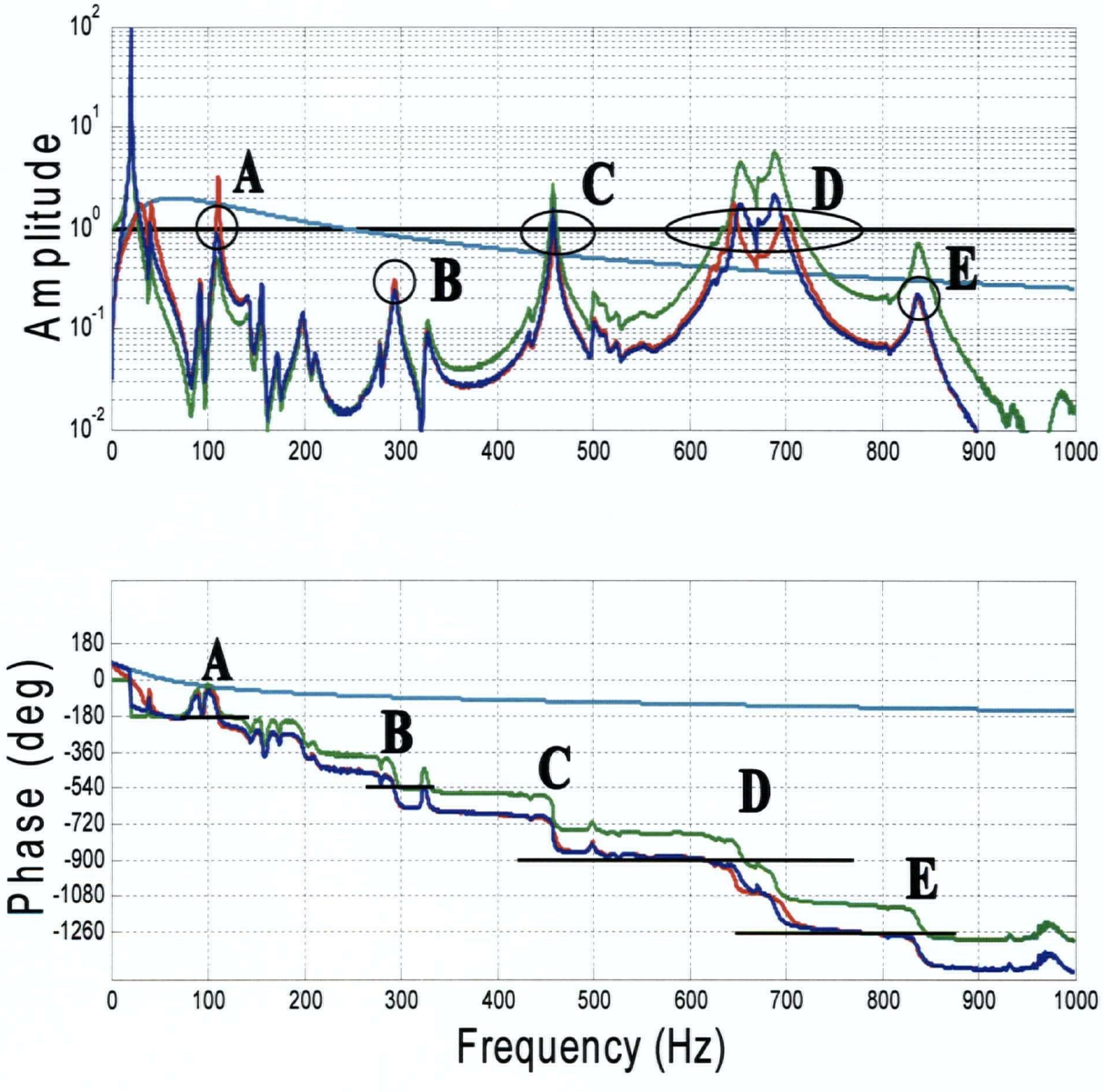

Figure 43: PID Control System Bode Plots

-Green: Mass response to piezo..
-Cyan: Controller at 9.7 ms gain and 65 Hz 2nd order low-pass filter.
-Blue: Open-loop system (product of mass response and controller)
-Red: Closed loop response.

In region A, around 100 Hz, the phase of the product of the gain and system response (the open-loop response) crosses through -180 degrees. However the magnitude of the open-loop response is below unity, so instability is not expected here. The phase crosses -540 degrees in region B (290 Hz), but the magnitude is far below unity, so this is also predicted to be stable. In region C (450 Hz) the magnitude is greater than unity but the phase has some margin from the unstable value of -900 degrees.

In region D (650-700 Hz) there are two strong resonances, with platform response greater than unity, and a phase swing of 360 degrees. The second-order low-pass filter on the velocity signal makes the controller gain less than unity here, but the product of controller and platform gain is still greater than unity. Careful examination of the phase shows that the dangerous values of -900 and -1260 degrees occur where the open-loop magnitude is less than 1. Region E (840 Hz) is not a problem, even though the phase remains near -1260, because the magnitude is much less than unity.

Since the gain used for this plot is one that was just at the edge of stability, we expect that the Bode plot will predict that the stability margins are low. This appears to be the case in region A, where a small increase in derivative gain would cause oscillation at 110 Hz. A larger derivative gain increase would cause problems at B, D, and E. Derivative gain would not necessarily cause problems at Point C because the phase is not dangerous.

The system shows some significant challenges to control using standard PID techniques. Notwithstanding, proportional gain could be applied with a low-pass filter and thus be of greater impact below 100 Hz. Also, the response at the trouble points can be controlled with notch filters. These approaches are discussed in the following section.

5.6. PID Control with Resonance Notching

Resonance notching is a simple concept to implement. The recipe is as follows: Increase gains of the underlying PID controller until oscillation is generated. Then tune a notch filter at that frequency and remove the frequency from the measurement stream thus preventing oscillation. That is, apply notch filters to the measurement stream with adjustable frequencies and bandwidths at the appropriate frequencies to eliminate oscillations. Appendix C details the algorithm used to digitally implement a notch filter and it is a straightforward algorithm to implement.

However, in our case, the results only allow a marginal increase to the PID gains. The differential gain increases from 9.7 ms to 10.3 ms and we manage to apply a proportional gain of 0.48 whereas no appreciable proportional gain at all could be applied previously without resonance notching. Four notches were applied at 600 Hz (width 100 Hz), 298 Hz (width 10 Hz), 840 Hz (width 10 Hz) and 113 Hz (width 8 Hz). The improved attenuation envelope is shown in the figure below. Only the frequencies below 100 Hz are shown to highlight the area where there is a difference in attenuation.

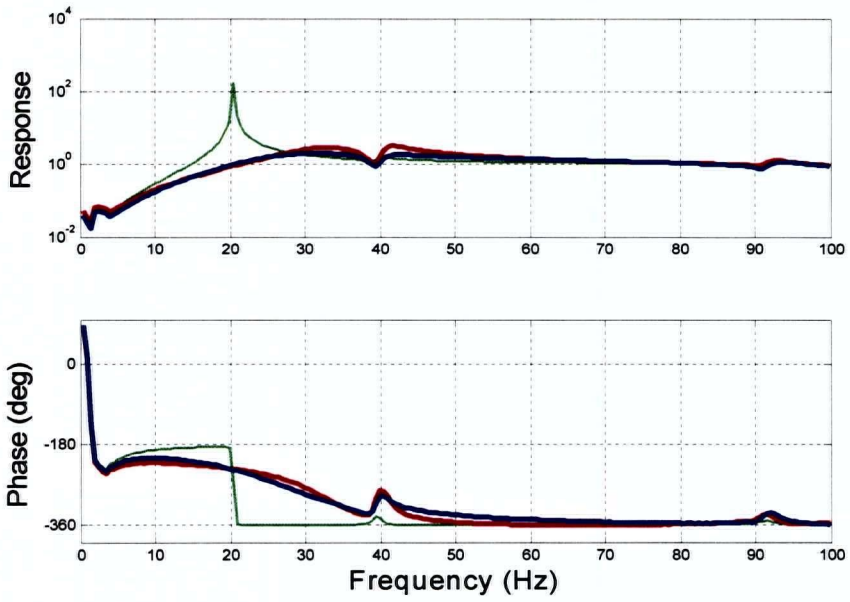

Figure 44: Gain Increase due to Notch Filters

The blue plot shows the maginally improved performance due to an increase in gain due to the notch filters. The red plot represents results from derivative alone. The thin green plot shows the totally uncontrolled response.

Measurement vibrations are now reduced to 9.3 nm RMS which now reflects a 5-fold reduction in RMS vibration from a totally uncontrolled system.

5.7. Coherent Disturbance Correction

The internal details of the numerical notch filters inspired us to perform an interesting technique for vibration elimination. Specifically, after PID application, the largest components of the ground vibration are now observed to consist of narrow spectral spikes. Tracking of these spikes with resonators provide a real-time estimate of their amplitudes. This information proved useful.

To improve the system performance now, we model the vibrations themselves rather than trying to improve feedback performance.

Imagine we want to eliminate a coherent vibration with exactly known frequency, amplitude and phase. If we added a signal of this exact frequency to the piezo independently of any feedback, the system would respond exactly as given by the detailed system Bode plot (refer figure 36). Thus we could adjust the phase and amplitude of our signal using the prescription of the Bode plot to cancel the vibration from the measurement entirely. This is one approach to what is known as feedforward control.

Unfortunately we do not know the exact frequency, amplitude and phase of our fairly coherent ground vibrations. However, we do know from measurements the narrow frequency bands which be contribute extensively to our spectrum.

Using only the frequency and width of our suspected noise peaks we can extract the amplitude and phase of our vibrations by applying the measurement to a resonator tuned to that frequency. Digital signal generators can then feedforward these signals to the piezo and individual sharp noise spikes can be eliminated from the measurement result.

The amplitude and phase of individual noise spikes can be ascertained and tracked by what I call accumulating resonators which are simply resonating filters without damping that increase in amplitude at a constant rate (and indefinitely) when fed with a signal with a resonant frequency component of constant amplitude. This is required since the iteration by iteration signal fed to the piezo may be considered to be feedforward, the amplitude and phase of the actual vibration level (at a given resonator frequency) must be ascertained by slow tracking with integral feedback. The integrating version of a resonator can be created by eliminating the estimation from equation 72 for the resonator algorithm in appendix C as follows:

$$\Delta_k = m_k \qquad \text{Eqn(56)}$$

This alteration eliminates the self-limiting amplitude of the resonator. The residual is simply the measurement. When fed with a signal with any component of resonant signal, the resonator amplitude will grow without limit.

The application of the resonator signal (X_{real}) to piezo output only requires a phase as appropriate as shown in figure 36. The resonator will grow in amplitude until the applied correction to the piezo totally eliminates the spectral contribution. The phase extracted from the Bode plot does not have to be exact to achieve this result and indeed anywhere within 90 degrees of the correct phase will provide some vibration elimination.

Several resonators may be applied simultaneously to different spectral lines of the normal vibration spectrum and the integral tracking means that variation in the excitation amplitudes is quickly tracked by the system. Indeed, for man-made coherent vibration from equipment such as motors, the vibration may be completely absent at times when the equipment is off; this does not present any difficulty to our system except during the fast change in amplitude where the tracking of amplitude lags behind any actual change in amplitude.

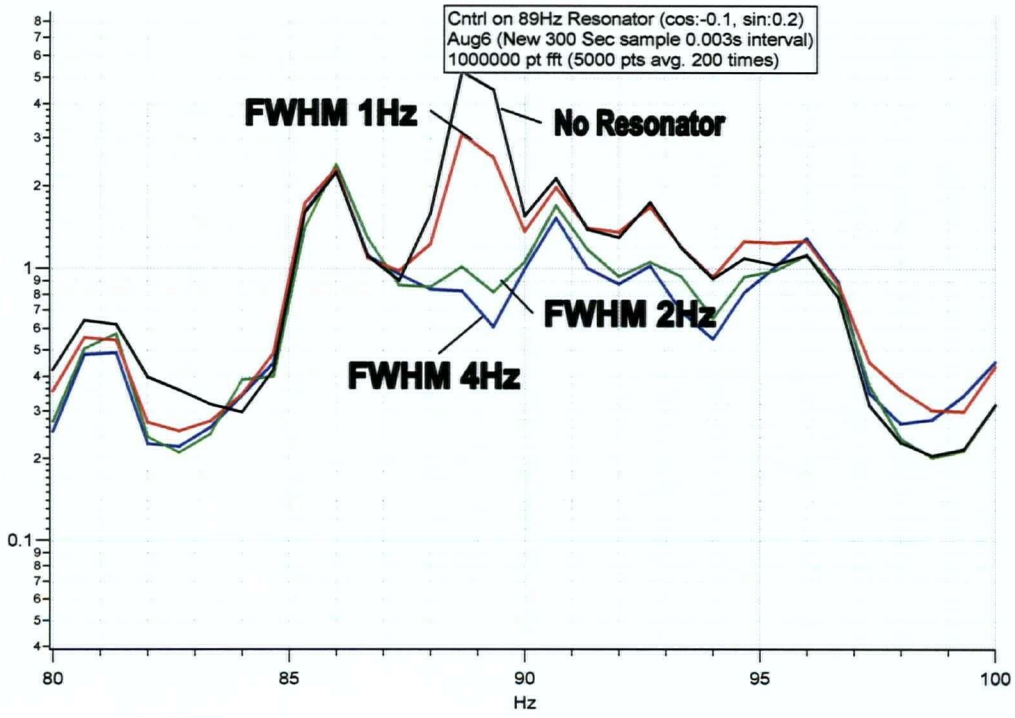

Figure 45: Coherent Disturbance Elimination

This figure shows a detail spectrum near the 89Hz coherent excitation. The different traces represent different line widths applies to the resonator. With the width of the line set at 2Hz or 4Hz, the line is totally removed from the spectrum. In this particular case, this represents a six-fold improvement at 89 Hz. Greater improvement is possible for higher amplitude lines.

Among the most successful spectral line eliminations using this new technique was a line at 89 Hertz where we achieved a fivefold amplitude reduction at this frequency. Effective elimination of an individual noise peak at 89 Hz is detailed previously in figure 45.

Note that the resonator improves response from 90 to 95 Hz and yet does not affect the peak at 86 Hz at all. This can be understood by referring to the system Bode plot which shows a rapid phase shift between 85Hz and 89Hz. The corrective phase prescription is simply incorrect for the 86 Hz peak.

This is an exciting improvement in that it allows us to control and eliminate vibrations at frequencies much higher than could be approached with our PID feedback. Notwithstanding that there are more advanced feedback techniques available to control than our PID setup, the method proved very robust and effective for sharp spectrally narrow disturbances. Any system where disturbances are highly coherent with strong spectral lines may benefit from coherent disturbance modeling.

Indeed, I feel that any non-flat structure to a vibration spectrum could be reduced or eliminated using analogous techniques. Certainly for our sharply defined spectral lines of vibration, the coherent disturbance modeling provided good improvement.

5.8. Best Composite Performance

Using best result empirical tuning for PID gains, notch resonators and disturbance resonators, the overall RMS residual is reduced to 4.1 nm. This represents our best performance of a factor of 12 improvement over an uncontrolled system. The spectrum for this best performance is shown below:

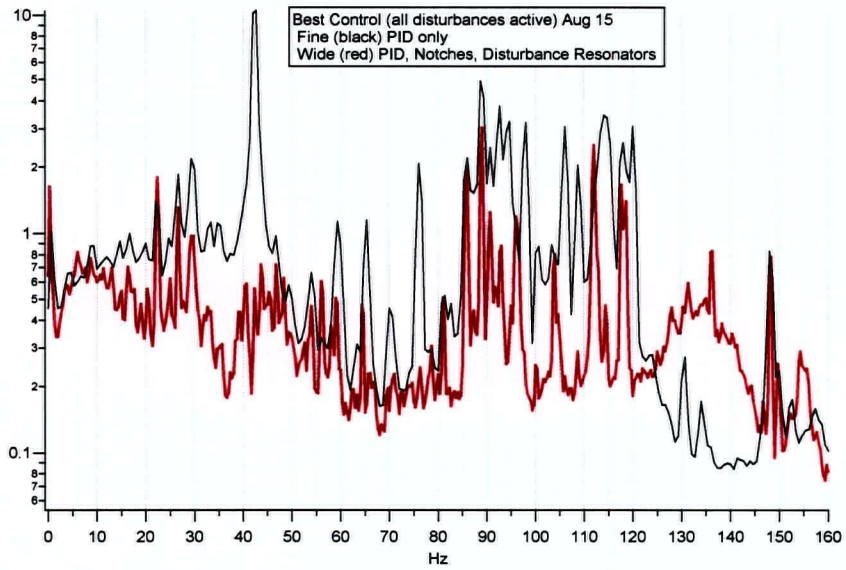

Figure 46: Best Control Achieved

Best performance shown in thick trace (red) and best "PID only" performance shown in thin trace (black).

Notice the excellent elimination of the 40 Hz noise spike using an accumulating resonator. We had 10 um of control authority which was 80% utilized in the above control tests.

Ground disturbance resonators were applied at 42, 89, 120, 105.5 and 59.5 Hz. Note large spectral disturbance eliminations at all of these frequencies. Indeed, the best performance spectrum lies completely within the envelope of the "PID only" below 120 Hz. The poor perfromance above 120Hz is related to larger PID gains without completely compensating notches. Notwithstanding, overall contribution to residual is negligible above 120 Hz for both spectrums. The measured best performance was 4.1nm RMS

6. Conclusions

Two separate aspects of vibration control have been investigated: an alternate algorithm for the "optical anchor" interferometer distance reference and direct control to nanometer precision of a large mass with a piezo-electric actuator.

The "optical anchor" work consisted of algorithmic design to conform to a slightly more constrained design criteria than had been previously achieved. Previous work had achieved excellent position measurement performance. However, the previous work maintained a separate calibration and measurement procedure which might be cumbersome in some applications. The new algorithm was designed to operate without any explicit calibration procedure.

The new algorithm takes a geometrical interpretation of the intensity information from two detectors within the interferometric pattern. The geometrical interpretation prevents non-linearities in the parameterization from forming local minima at incorrect values of calibration parameters. Thus the calibration proceeds monotonically as data is received. If data explores the parameter space, then the parameters converge rapidly.

For fully converged calibration parameters, the random uncorrelated error from measurement to measurement is 0.05 nm with a systematic error related to modeling imperfection of 0.22 nm (reported as rms but actually correlated to the interference phase). This results in an overall accuracy of about 0.23 nm.

This accuracy is sufficient to provide the principle measurement to a vibration control system with a target accuracy of 1 nm. However, measurement arms of the interferometer were short and protected from air currents. Offhand observation of longer measurement arms shows that for long-arm interferometric measurements, air currents must be eliminated and the measurement legs will have to be enclosed in piping.

Control tests on a one degree-of-freedom experimental platform are performed using proportional and differential control. Performance did not meet the 1nm RMS goal for magnet position residual set by NLC design specifications. Artificially stiffening a mechanical support using classic feedback techniques and a piezo-electric actuator proved disappointing.

The fundamental resonant reaction was eliminated but overall vibration attenuation was only a factor of 4 improvement over the uncontrolled reaction to 12.1 nm RMS.

The extension of classic PID with resonance notching served to marginally increase gains which served to reduce RMS vibrations to 9.3 nm.

More advanced control techniques would hopefully yield better performance with regard to feedback. However, once vibrations have been limited such that the principle remaining vibration is caused by narrow frequency spikes, vibration modeling with accumulating resonators is capable of eliminating on the order of 50% of remaining RMS levels. In our case such modeling reduced our best RMS result to 4.1nm.

Appendix A NLC Vibration Control Editorial

Let us look at the problem posed by the NLC requirements:

- The final focus magnets contribute the largest component of beam position to the particle beam prior to collision. The final focus magnets are subject to movement due to environmental motion. The motion is a function of the magnet support design combined with the magnitude and frequency of the man-made and natural vibrations. The aim of the particle beam needs to be accurate to approximately 1 nanometer RMS at the point of beam collision.

Most obviously, the man-made vibrations can be reduced. F. Le Pimpec et al. [2] examines the vibrational stability of NLC Linac Accelerating Structure. F. Asiri [4] et al is a study of near-field vibration sources for the NLC linac components. Conclusions are twofold: one, the final magnets should probably not be fluid cooled and are therefore required to be permanent magnets [2]; two, near-field vibrations can and should be reduced through careful consideration of vibration during facility engineering.

However, vibrations should be considered to still be on the order of an urban university lab to provide a safety design margin[3].

An exacerbating point regarding the expected vibration levels is that the ground motion is only the disturbance to any support structure. Actual vibration experienced by the focussing magnets will be much larger than the raw ground motion since it will include the resonance response of the support structure. The "pessimistic" model of ground motion may not be pessimistic enough to encompass the actual vibration encountered by the immediate support of the final quadrupole magnets.

My recommendation to the actual overall problem is as follows:

1. Eliminate vibration at its source to the limit of facility design.

2. Design the support structure as stiffly as possible to an upper limit of perhaps 15 Hz main resonance. Incorporate maximum passive damping in this design. Do not exceed 15Hz for the principle resonance; this will insure that rotating machinery vibrations will not resonate with the support structure (generally, rotating machinery will resonate at 30 Hz or more).

3. Measure vibration inertially of the support structure near the magnetic supports.

4. Use the inertial vibration measurement of the support structure in a feedforward manner to the piezo supports for the quadrupole magnet.

5. Measure vibration remaining with an accelerometer mounted on the magnet itself. Use a feedback signal into the same piezo supports.

6. Using an optimal Kalman estimator of any residual vibrations to feedforward into electromagnetic dipole steering magnets.

7. Inertial measurements may be made directly if a magnetic-field resistant device can be purchased or manufactured. Alternately, an inertial standard can be passively supported on the ground and referenced using optical-anchor technology.

An Argument for Inertial "Hard" Support Vibration Control:

When first given this problem, ones first instinct is to apply isolation to the system. Then try to use suspension systems and ultra-low frequency support schemes: fundamentally passive systems. Indeed, SLAC has undertaken a project designed to produce just such a system for this problem.

Using a suspension system draws upon the one reference frame easily accessible technologically to both magnets: the inertial reference frame. Furthermore, by controlling the magnets to the inertial frame, no high order modes of vibration are induced in the magnets since the purpose of suspension is to eliminate all accelerations.

However, in my view, It is the inertial reference frame that provides the functional key to the solution, not the soft suspension. One misconception that can easily be made is assuming that a system using an inertial reference as a position reference also should include a soft passive support system. It is a natural assumption since the soft support system inherently produces an inertially quiet system. However, when dealing with large masses, possible vacuum forces and large magnetic forces, a soft support system becomes difficult to implement because it cannot deal with large static forces easily. My comment on this matter is that the obvious actual engineering solution to the magnetic support system is to establish an inertial setpoint but control using hard supports such as piezo elements which can withstand very large static forces.

Appendix B Deadbeat Gain

Deadbeat gain is the name given to the gain which results in a minimum time to correct all state values to zero from an arbitrary starting condition. It has a unique non-iterative solution.

Starting with a common discrete system definition given below:

$$x_{n+1} = \mathbf{A}x_n + \mathbf{B}u_n \qquad \text{Eqn(57)}$$

But with perfect knowledge of the system states:

$$y_{n+1} = x_{n+1} \qquad \text{Eqn(58)}$$

x	State Vector
$\mathbf{A}$	System Propagation Matrix
$\mathbf{B}$	System Response to Output Operator
u	System Output Vector
y	Measurement vector

And where the system output is defined with a constant gain Operator:

$$u_n = \mathbf{K}x_n \qquad \text{Eqn(59)}$$

$\mathbf{K}$	Gain Row Vector

Combining gives:

$$x_{n+1} = (\mathbf{A} + \mathbf{BK})x_n \qquad \text{Eqn(60)}$$

Define the Propagation Operator **P** as:

$$\mathbf{P} = \mathbf{A} + \mathbf{BK} \qquad \text{Eqn(61)}$$

The deadbeat gain operator $\mathbf{K}_d$ is defined as the gain vector **K** which produces a propagation operator **P** with ALL EIGENVALUES EQUAL TO ZERO. This will produce an identically zero system state from any initial state within m iterations where m is the order of the system.

The following procedure solves for the deadbeat gains assuming a controllable system when the output, u, is scalar.

Generate a series of states starting from the zero state but excited with a unit output:

$$x_0 = \mathbf{0} \qquad \text{Eqn(62)}$$

$$x_{n+1} = \mathbf{A}x_n + \mathbf{B} \qquad \text{Eqn(63)}$$

Continue the process until there are m states generated where m is the order of the system.

Construct a matrix with columns consisting of these constructed states:

$$\mathbf{T} = \left(x_1 \mid x_2 \mid x_3 \mid \ldots \mid x_n\right) \qquad \text{Eqn(64)}$$

Then the deadbeat gain vector can be explicitly calculated as:

$$\mathbf{K}_d = bottom - row - of(\mathbf{T}^{-1}\mathbf{A}^m) \qquad \text{Eqn(65)}$$

$\mathbf{T}$	Excitation Series Matrix
$\mathbf{K_d}$	Deadbeat Gain Vector

When stability of a system is evaluated by the largest eigenvalue of the propagation matrix, P, then the deadbeat solution is the most stable system possible.

However, the deadbeat solution is "similar" to an LQR solution in the limit as the cost-of-output, (usually known as R), goes to zero. The deadbeat solution is thus just as unreasonable for actual application in most cases as a low cost-of-output LQR gain.

"Similar" gain in this case can be taken to mean gains whose elements are of the same close order of magnitude and same sign.

Appendix C Notch Filter Algorithm

We implemented our notch filter with a 'resonator' which is a numerical technique using a state vector designed to resonate at a single frequency excited by an input signal. This resonator vector is a numerical representation of a sine and cosine component (or real and imaginary component) of a single frequency amplitude.

$$x = \begin{bmatrix} X_{real} \\ X_{imag} \end{bmatrix} \qquad \text{Eqn(66)}$$

The vector x is updated every iteration of a numerical system through a rotation operation and an innovation as follows:

$$x_{k+1} = \mathbf{R}x_k + \mathbf{G}\Delta_k \qquad \text{Eqn(67)}$$

Where G and Δ are defined shortly and where the matrix **R** is a standard rotation matrix:

$$\mathbf{R} = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \qquad \text{Eqn(68)}$$

The angle alpha is defined by the algorithm iteration time to be the angle of phase transited during a single algorithm loop time at the resonant frequency:

$$\alpha = 2\pi fT \qquad \text{Eqn(69)}$$

Where f is the frequency of the resonator in hertz and T is the loop time for the algorithm in seconds. Both **R** and α are constant and don't have to be calculated every loop iteration.

The matrix **G** is the gain matrix which allows the measurement to affect the filter amplitude and is given by:

$$\mathbf{G} = \begin{pmatrix} w \\ 0 \end{pmatrix} \qquad \text{Eqn(70)}$$

Where w defines the width of the notch filter as follows.

$$w = 2\pi f_w T \qquad \text{Eqn(71)}$$

Where f_w is the width of the resonator response (or notch filter) in hertz.

The residual is defined as the measurement minus an estimate of the resonator amplitude as follows:

$$\Delta_k = m_k - (0 \quad 1)\mathbf{R}x_k \qquad \text{Eqn(72)}$$

Where m_k is the measurement.

The notched measurement stream is simply the first element of the residual Δ_k. The measurement minus the predicted amplitude of the resonator is the notched result. In short, the residual **is** the notched measurement stream.

To obtain mutiple resonant notches, every resonator prediction must be subtracted from the measurement to obtain an overall residual:

$$\Delta_k = m_k - \sum_{j=1}^{\text{MaxResonators}} (0 \quad 1)\mathbf{R}_j x_{jk} \qquad \text{Eqn(73)}$$

The real component of x_j can be used as a real-time measurement of a single spectral component of the signal. An overall amplitude of the signal can be obtained by taking the norm of the vector x_j.

Thus the entire algorithm takes about five multiplication operations and about the same number of additions per iteration to provide either a specral line amplitude monitor or an effective digital notch filter.

References

1) J. Frisch et al., *Active Vibration Suppression R&D for the NLC* in Proceedings of the 8th International Conference on Accelerator and Large Experimental Physics Control Systems, 2001, San Jose, California (published electronically)
Retrieved June 2, 2004 from http://epaper.kek.jp/ica01/proceedings.html

2) F. Le Pimpec et al., *Vibrational Stability of NLC Linac Acceleration Structure* in proceedings of EPAC 2002,Paris (published electronically)
Retrieved June 2, 2004 from
http://accelconf.web.cern.ch/AccelConf/e02/KEYWORDS/K109.htm

3) S. Allison et al., *Active Vibration Suppression R+D for the Next Linear Collider* in *Chicago 2001, Particle accelerator* 1261-1263

4) F. Asiri, F. Le Pimpec, A. Seryi, *Study of Near-Field Vibration Sources for the NLC Linac Components* in Proceedings of PAC 2003 Portland (Or), 2003, SLAC-PUB-9815

5) Dutton,Thompson,Barraclough, *The Art of Control Engineering* (Addison Wesley Longman 1997). ISBN 0-201-17545-2

6) Malcolm B. Gray et al., A simple high-sensitivity interferometric position sensor for test mass control on an advanced LIGO interferometer Optical and Quantum Electronics 31: 571, 1999.

7) T. Mattison, R. Greenall, T. Downs, *Vibration Control Feedback R&D at University of British Columbia* in Proceedings of the 26th Advanced ICFA Beam Dynamics Workshop On Nanometer Size Colliding Beams (Nanobeam 2002), Sep 2002, Lausanne, Switzerland edited by Zimmermann, Assmann

8) Grewal, Andrews. *Kalman Filtering Theory and Practice* (Prentice Hall, Englewood Cliffs, New Jersey 1993). ISBN 0-13-211335-X

9) Zhou, Doyle, Glover. *Robust and Optimal Control*
(Prentice-Hall, Upper Saddle River, New Jersey, 1996)
ISBN 0-13-456567-3

10) Pedrotti, Pedrotti. *Introduction to Optics Second Edition*
(Prentice Hall, Upper Saddle River, New Jersey, 1993)
ISBN 0-13-501545-6

11) Franklin, Powell, Workman. *Digital Control of Dynamic Systems* (Third Edition, Addison-Wesley USA, 1998) ISBN 0-201-82054-4

12) *Modern Control Engineering* (Third Edition, Ogata, Prentice Hall, Upper Saddle River, New Jersey, 1997) ISBN 0-13-227307-1

13) A Mohazab (private communication April 2003)

Corporate References

14) Jager Computergeu MesstechnikGmbH, Rheinstr. 4,64653
Lorsch,Germany
Manufacturer of Adwin-Gold realtime computer system

15) WaveMetrics, Inc., P.O. Box 2088, Lake Oswego, OR 97035, USA
Manufacturer of "IGOR" data aquistion/ manipulation software

16) RTS Consulting Ltd, 1 Noble Court, Port Moody, B.C. V3H-3Z5 Canada.
Manufacturer of ADWIN-IGOR connection driver software.

17) Waterloo Maple Inc.
450 Philip Street, Waterloo Ontario
Canada N21 5J2

18) The Mathworks Inc. (Publishers of MATLAB)
3 Apple Hill Drive
Natick, MA 01760-2098
UNITED STATES

19) Physik Instrumente (PI) GmbH & Co. KG
Auf der Roemerstrasse
D-76228 Karlsruhe/Palmbach
Germany

Index

accuracy, 57
actuator, 3
Adwin, 31
Algorithm Tests, 57
Apparatus, 29
Beam Based Systems, 1
beam deflection, 1
beam diameter, 1
Coherent Disturbance Correction, 74
coherent ground disturbance, ii
control response, 16
control term, 16
control transducer, 34
Convergence, 50
critical damping, 8
damping, 4, 6
Data Acquisition, 32
deadbeat gain, 21
degrees of freedom, 45
Direct Parameterization, 37, 39
discrete parameterization, 16
effective control, 5
explicit model, 39
Explicit S Iteration, 49
feedback, 5
feedforward, 5
first order iterant, 48
flexure mode, 23
geometric parameterization, 45
Geometric Parameterization, 43
Geometric Parameterization Validity, 44, 47
Inertial Based Systems, 2
Interferometer, 37
Interferometer Based Systems, 1
Interferometer Setup, 37
interferometric instrument, ii
lissajous, 43, 44, 50, 51, 53, 57
lock-in amplifier, 32, 33
luminosity, 1
Mechanical tuning, 57
modes, 2
Mohazab, 45
monlithic mass, 23
natural system evolution, 16
Next Linear Collider, 1
Non-Correlated Noise Test, 59
non-linear systems, 43
Notation convention, 16
optical anchor, ii, 1, 2
Optimal tuning, 57
Parameter Iterations, 47
Parameters of the geometric model, 44
Particle accelerators, 1
phase-locked, 32
PID, 8
piezo, 33, 34, 58
propagation matrix, 17
quadrature, 40
second order response, 5
Simple Continuous Model, 3
spectral system response, 32
stability, 4
stability contour, 21
stalled trajectories, 53, 54
symmetric matrix, 50
system evolution, 4
System Model, 5, 9, 13
system response, 4
systematic error, 57
test mirror, 58