Linear Macro-Micro Positioning System Using a Shape Memory Alloy Actuator

by

Eric M.T. Ho

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2004

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The use of high-precision automated equipment is steadily increasing due in part to the progressively smaller sizes of electronic circuits. Currently, piezoelectric transducers (piezos) dominate as the actuation device for high precision machines, but shape memory alloys (SMA) may be a viable alternative to reduce monetary costs.

This work explores the implementation of a low-cost linear macro-micro positioning system. The system consists of a modified printer carriage to provide long range, macro scale linear motion (approximately 200 mm range and 200 μm precision) and a micro scale system (approximately 4 mm range and 5 μm target precision) that uses an SMA actuator. A detailed description of the design and implementation of the system is given in this research.

A model of the macro-stage is then generated by first identifying and inverting a simple friction model to linearize the system, thereby allowing for modified least squares (MLS) identification of a linear model. Various controllers are attempted for the macro-stage and compared with an experimentally tuned nonlinear PD controller that is implemented in the final design.

A model of the micro-stage is derived through analysis of the SMA actuator. The model for the actuator is separated into two portions, an electro-thermal model, and a hysteresis model. The hysteresis model is derived using the Preisach model, and the electro-thermal model through MLS identification. To control the micro-stage, a PI controller with anti-windup is developed experimentally.

The two stages are then executed together in closed loop and the resulting coupling between the two stages is briefly examined.

Experimental data used for the modelling and design is presented, along with results of the final macro-micro linear positioning system.

Acknowledgements

Thanks to my supervisor Robert B. Gorbet. Without his help and guidance this research would not have been possible. Thanks also to the readers, professors Jan Huissoon and David Wang for taking the time to review this thesis. Finally, I would like to thank my father, my mother, and brother for their continued support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

As manufacturing systems have become increasingly automated, active positioning systems have become an important step implemented in fabrication. As outlined in [1], the main applications for automatic positioning systems are to place machine tools for processing or probing parts, or to position parts with respect to another for assembly. However, many of the established positioning systems operate on macro-scale requirements. It is only fairly recently that the need for micron-scale systems has become an important issue, and unfortunately, macro-scale automation strategies are often not suitable.

In macro-scale fabrication, precision requirements rarely fall into the micron or sub-micron range, and the payloads are usually high, requiring large application forces. Common macro-scale actuators are electric motors and hydraulic/pneumatic pistons, both of which suit macro-scale requirements. However, as the scale of operation decreases, the automation requirements begin to change. Low force, high precision movements are desired and more importantly, mechanical limitations such as friction become increasingly significant as the scale of operation decreases. Electric motors are limited in precision by mechanical gearing and friction becomes increasingly difficult to overcome. Friction also greatly limits the performance of hydraulic/pneumatic pistons at small scales.

The decreasing size of electronic circuits has contributed greatly to the development of micron-scale manufacturing techniques. One development often employed in micro-automation is the coupling of a fine and coarse manipulator to provide micro-scale processing over a large workspace. This allows both for transport of a part or tool over large distances, and high precision processing via the fine manipulator.

The micro-scale processing often employs chemical, such as etching, and optical methods, such as exposing a mask. These methods are well suited to the micron scale as they essentially avoid many mechanical difficulties. Unfortunately, these methods cannot physically move objects, such as mask alignment, placing devices on a circuit board, or manipulating a tool at micron-scale precision. Shape memory alloy (SMA) based actuators offer a possible solution to this problem. SMAs can generate mechanical motion with theoretically infinite precision since continuous, as opposed to discrete motion, is provided.

1.1 Motivation

Piezoelectric transducers (piezos) dominate as the actuation device for high precision machines. Piezos provide high forces but microstrain [2]; thus the range of motion for piezo-based precision machines is very low unless a form of stroke amplification is used. As outlined in [3], amplification is mainly achieved through two general classes of flexure hinges, levers and frames. This increases the requirement for high precision parts and thus monetary cost.

SMAs may offer a viable alternative to piezos for some high precision positioning applications. SMAs provide larger strains, and possibly eliminate any need for stroke amplification. The main disadvantage of SMAs compared to piezos is reduced bandwidth. However, this may be offset by the potential cost reduction SMAs offer, making this configuration suitable for certain low-bandwidth applications.

1.2 Research Objective

The overall goal of this project is to design and build a low-cost (<$500) linear macro-micro positioning system using an SMA actuator for the micro-stage. The proof-of-concept prototype is shown in Figure 1.1.

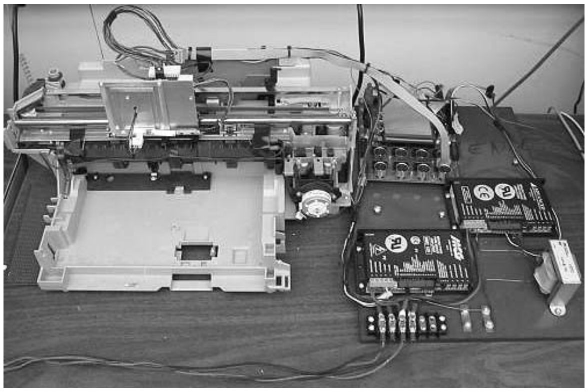


Figure 1.1: Prototype Positioning System

1.3 Scope

This thesis assumes that the reader has an engineering background and has an understanding of control systems and system identification techniques.

1.4 Outline

This work will describe the design and implementation of both the mechanical aspects and software control of the prototype positioning system. Both the macro- and micro-stages will be described from initial design to final implementation. Simulated and experimental responses will be shown as well.

Chapter 2 focuses on the SMA of the micro-stage actuator. An overview is given of SMA material properties and their origins. As well, the Preisach model is described to model the hysteresis SMAs exhibit. Also, the performance and cost of current micro-positioning technologies are reviewed.

Chapter 3 describes the experimental implementation of the prototype positioning system in detail. The interface of the macro system with a computer is described as well as the physical design and assembly of the micro-stage.

Chapter 4 identifies an experimentally tuned nonlinear model and develops a controller for the macro positioning stage. A model for the Coulomb friction is developed and then used to linearize the macro-stage for MLS identification. The resulting model is validated through comparison with experimental data. As well, various control techniques are attempted and compared to an experimentally tuned nonlinear PD controller that is used for the final implementation.

Chapter 5 attempts to develop a model for the SMA actuator within the micro-stage, and also derives a controller. The model is divided into two portions consisting of an electro-thermal model and a hysteresis model. The Preisach model is adopted to model the hysteresis, while MLS is used to identify the electro-thermal model. The resulting model is validated through comparison with experimental data. Then, an experimentally tuned PI controller with anti-windup is developed and implemented.

Chapter 6 gives experimental results from the full macro-micro positioning system. An analysis of its performance and cost is performed, and a brief overview of the issues arising from the coupling of the macro- and micro-stages is presented through experimental results,

Finally, Chapter 7 provides conclusions based on the results, and outlines future work that may be pursued as a result of this research.

Chapter 2

Background

Before the development and analysis of the macro-micro positioning system can begin, some important concepts must first be reviewed. First, a brief review of macro-micro manipulators is performed. Then, an overview of micro-positioning applications, and available micro-positioning technologies along with their costs is given. Also, the history and mechanics of operation of shape memory alloys is reviewed, as well as the Preisach model which is a well developed method of modelling hysteresis.

2.1 Macro-Micro Manipulators

A macro-micro manipulator helps to increase accuracy of the manipulator endpoint while maintaining a large workspace. For many applications that involve a large workspace and require high accuracy, the use of a single manipulator proves to be difficult. This is usually the case because a manipulator large enough to cover the workspace, often contains a degree of flexibility that produces significant error at the manipulator endpoint. Therefore, the flexibility of the manipulator contributes to reduced accuracy and lower bandwidth. A micro manipulator, added to the end of a macro manipulator, can provide fine adjustment to correct any error from the macro-stage, and thereby increase accuracy while maintaining a large workspace.

The concept of macro-micro manipulators is not new. A notable first example on the development of macro-micro manipulators was performed by Sharon and Hardt [4]. In [4], a hydraulically driven 5 degree of freedom micro manipulator was mounted on the end of a flexible beam macro manipulator.

Other examples of macro-micro manipulators include research by Chalhoub and Zhang [5], Narikiyo *et. al.* [6], and more recently, Lee and Cho [7], and Glisinn *et. al.* [8].

Chalhoub and Zhang [5] have developed a macro-micro manipulator by mounting a micro manipulator, a two-axis Cartesian robot driven by dc motors, on the end of a macro manipulator, a single flexible beam. The endpoint position of the macro-stage was measured by a laser-photo-detector system, and fed forward to actuate the micro manipulator. Narikiyo *et. al.* [6] have mounted a voice coil drive micro manipulator, on the end of an arc welding robot macro manipulator. This system was designed for use in the application of electrodischarge machining to complex parts. Lee and Cho [7] also used a voice coil driven micro-stage, but implemented a ball screw driven macro-stage. The micro-stage was used for precision force control during the surface mounting of components onto a printed circuit board. An example of an extremely small scale application, Gilsinn *et. al.* [8] developed a macro-micro motion system for use in a scanning tunneling microscope (STM). Here, a piezo driven flexure was used as the macro-stage (1-10 mm stroke and sub-micron resolution), and an STM tip was used for the micro-stage (10-500 nm stroke and sub-nanometre resolution).

In the research here, the micro-stage is to operate on the micron to sub-micron scale. Therefore, some micro-positioning technologies is reviewed next.

2.2 Micro-Positioning

Micro-positioning is becoming increasingly important for research and in automated manufacturing. As both electrical and mechanical devices become smaller, micro-positioning must be used to guide tools to manipulate and build them. For example, mask alignment in circuit manufacturing is increasingly difficult due to the decreasing feature sizes [1].

Also, research is continually examining and manipulating natural and manufactured objects at higher magnifications, therefore requiring micro-positioning tools to manipulate the objects. For example, as described in [9], atomic force microscopy involves the movement of an object under a probe at extremely high resolutions to scan objects. The

Table 2.1: Micro-Positioning Technology Comparison

System Type	Accuracy (nm)	Range	Speed	Load Capacity (N)	Approximate Cost ($ USD)
Miniature DC motor with lead screw and nut [10]	100 - 1000	<600 mm	0.1 μm/s - 100 mm/s	200	1500-5000
Ultrasonic piezo-drives [11] [12]	100 - 1000	unlimited	<130 mm/s	3	3000
Piezo stacks and Flexure guide [11] [12]	<1	<300 μm	at best 10KHz	100(pushing) 20(pulling)	4000
Piezo Inchworm Motor[13] [14]	20	<25 mm	1 nm/s - 1.5 mm/s	15	8000

scanning is performed by monitoring the atomic forces between the object and the probe, and then using the forces as feedback to move the object to follow its profile.

With the need for micro-positioning systems, many different technologies have been developed and implemented.

2.2.1 Current Technologies

There are many different technologies available for micro-positioning. A brief overview of existing micron-scale positioning systems is given in Table 2.1, and photographs of each system are shown in 2.1. As well, sub-micron positioning systems are also reviewed, as it is hoped that the result of this research will eventually lead to sub-micron positioning with SMAs. It should be noted that the given costs include a stage and feedback sensors, but not a controller.

Miniature DC or stepper motors coupled with a lead screw and nut system can provide both large travel lengths and relatively high resolutions. They are fast, and can provide extremely high forces.

Ultrasonic piezo-drives on the other hand, excite piezo actuators to produce elliptical oscillation. The elliptical motion of the piezo tip is used similar to a roller to push a stage along its track through friction. They are fast with a few millisecond settling times and strokes limited to the length of the track.

For resolutions in the sub-micron range, frictionless drives are better suited since they eliminate friction nonlinearities that hinder operation for extremely small displacements. Electromagnetic linear motors, voice coil drives and solid state piezo actuators are methods of frictionless actuation. Electromagnet linear motors and voice coil drives are, however, less commonly used in manufacturing. This is due to their high heat generation and induced magnetic field that can interfere with other processes (for example, electron beam steering [11]). Piezo actuators are highly popular due to their fast response time, high forces and lack of magnetic fields. They are, however, quite expensive due to additional stroke amplification and guides that are often required. There are also differences in the method of use of piezo actuators in the sub-micron range. Two popular designs are piezo stacks with flexure guides, and inchworm motors.

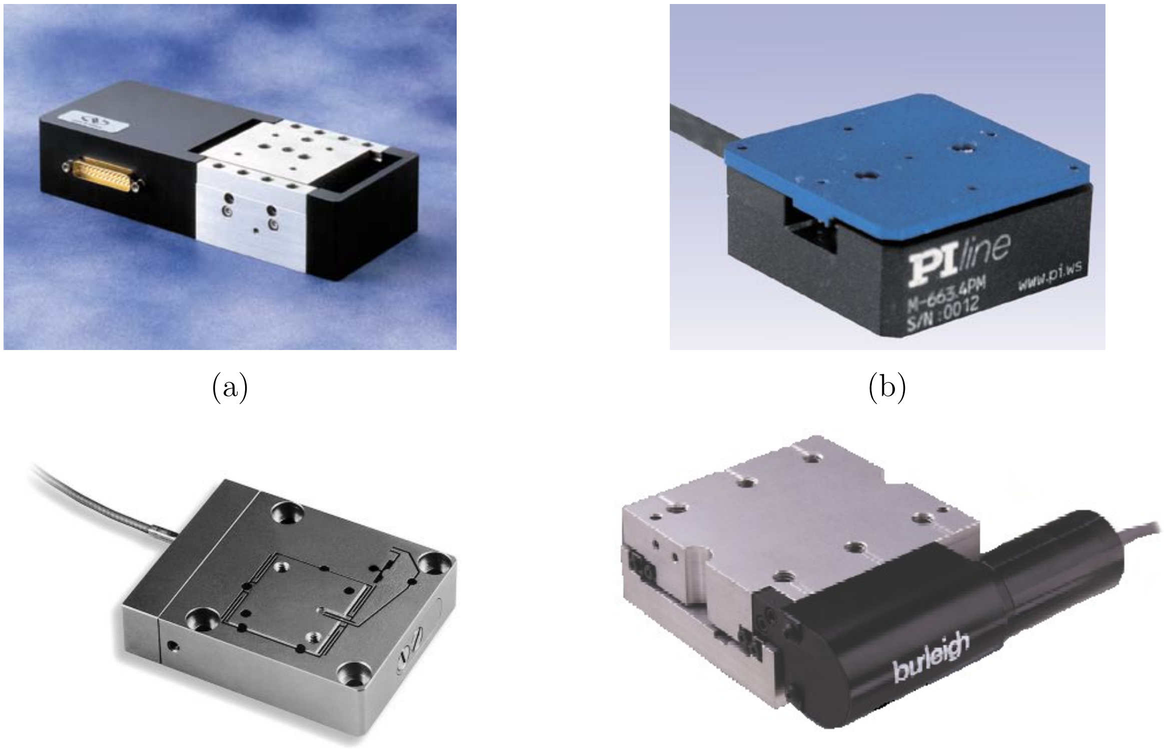
Piezo stacks are simply a group of piezos placed in series to increase the overall displacement. During actuation, the contraction or extension of each transducer is summed. Many nano-positioning devices couple the piezo stack with flexure guides to amplify the motion and provide frictionless motion guidance. However, these devices are nonlinear due to the hysteresis exhibited by piezos.

Inchworm motors on the other hand use three piezo actuators. One center stack parallel to the direction of motion, and two on each end mounted perpendicular to the shaft. The two end stacks act as clamps to hold the actuator in place, while the centre beam contracts and extends to move the assembly, much like a “worm”. A detailed description of its operation can be found in [13].

SMAs offer an alternative solution to the current available micro-positioning technologies.

2.3 Shape Memory Alloys

Shape memory alloys (SMAs) consist of materials that have the ability to “remember” shapes. This property allows them to undergo deformation at low temperatures, but subsequently return to their original state after heating. The most common shape memory alloys used in industry are copper zinc aluminum (Cu-Zn-Al) and nickel titanium (Ni-Ti). In fact, it was the discovery of the shape memory effect in Ni-Ti based alloys during the



(a) (b)

(c) (d)

Figure 2.1: Photograph of (a) Miniature Motor [10] (b) Ultrasonic Piezo-drive [11] (c) Piezo stack with Flexure Guide [11] and (d) Inchworm Motor [13] Positioning Stages

1960's at the Naval Ordinance Laboratory that first initiated great interest in the study of the shape memory effect [15].

The mechanism of the shape memory effect is very complex and difficult to understand. Thus, only a very basic overview of the mechanism for the shape memory effect, based on the information presented in [15], is given here.

2.3.1 Overview of Shape Memory Effect

The mechanism for the shape memory effect is rooted in phase changes within SMAs between austenite, also referred to as the parent phase, and martensite. To better understand the phase change in SMAs, the transformations that occur must be analyzed at a microscopic scale. Since the atomic changes are complex and require detailed knowledge of material properties, an effective two dimensional approximation is used to simplify the analysis. The unit micro-structure of the SMA material is approximated as quadrilaterals as shown in Figure 2.2; each node represents an atom and each connection an atomic bond. The parent phase, austenite, exists at high temperatures and the crystal structure it presents is highly symmetric. The unit micro-structure of austenite is approximated as a square, as shown in Figure 2.2a.

Figure 2.2: 2-D Schematic of Unit Micro-Structure of SMA in (a) Austenite and (b) Martensite Phases

To form an approximation of a piece of material in the austenite phase, a series of square unit micro-structures are combined to form the 2-D schematic illustration in Figure 2.3. It is important to note that the austenitic microstructure is unique and is an important property of the shape memory effect.

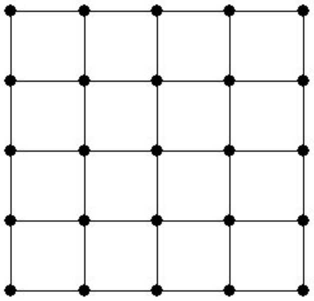


Figure 2.3: 2-D Schematic of Austenite Phase

The martensite phase exists at low temperatures and is formed upon cooling from austenite. The unit micro-structure of the martensite is different from that of austenite, and is approximated here as a rhomboid as shown in Figure 2.2b. However, the macro-structure must retain the same shape and size during the transformation from austenite to martensite. This is an accommodation step. To accommodate the change in micro-structure but maintain the same shape, the material microstructure must deform. During deformation, atoms move between each other to achieve the lowest state of energy and in accommodation, the dominant methods of atomic movement are slip and twinning. During slip, atomic bonds are broken allowing atoms to move and form new bonds. This produces irreversible changes to the micro-structure and permanent deformation. In twinning, atoms reshuffle without breaking any bonds by forming what are called twin boundaries; this change is reversible. An illustrated comparison of accommodation through slip and twinning is shown in Figures 2.4 and 2.5 respectively.

In SMAs, twinning is the dominant accommodation process. Therefore, microscopic changes during phase transformations are fully reversible.

An important property of twin boundaries is that they are of very low energy and thus very mobile. If stress is applied to deform the material, the twin boundaries will move before atomic bonds are broken to accommodate the stress. This process is referred to as detwinning and is illustrated in Figure 2.6. Detwinning results in macroscopic shape changes and also, significant strains (up to 8% in Ni-Ti) can be accommodated by detwin-

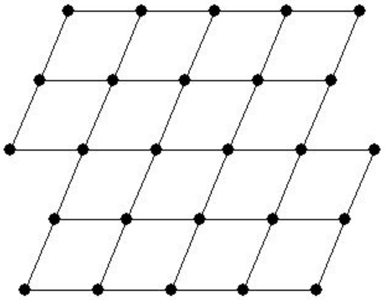


Figure 2.4: Accommodation through Slip

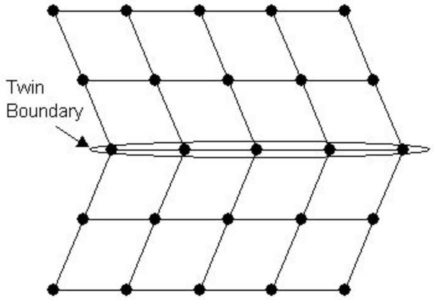


Figure 2.5: Accommodation through Twinning

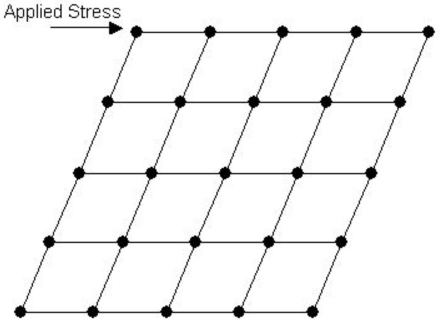


Figure 2.6: 2-D Schematic of Detwinning

ning.

If the deformed material is heated to the achieve the austenite phase, the microstructure reverts back to its original shape since it is unique, resulting in a memory effect. An overall schematic of the shape memory effect is illustrated microscopically and macroscopically in Figure 2.7.

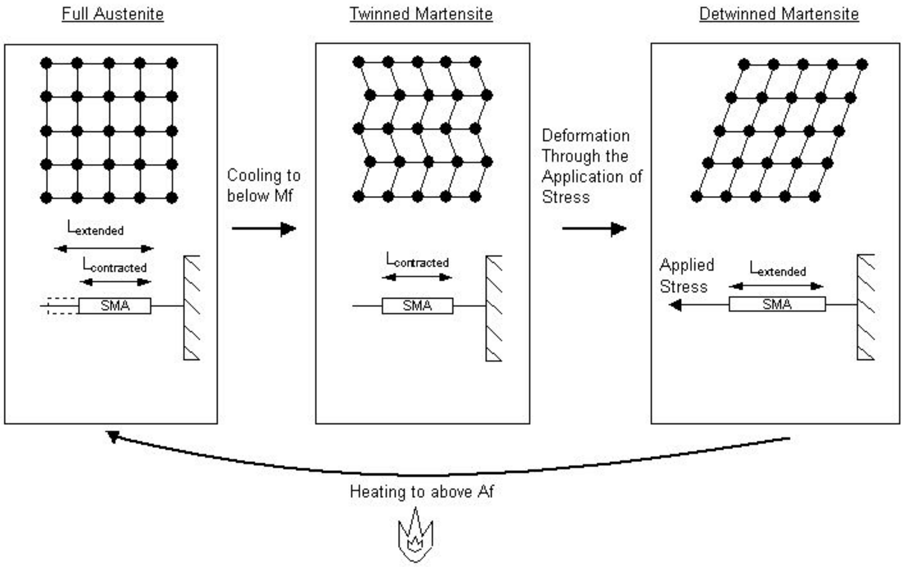
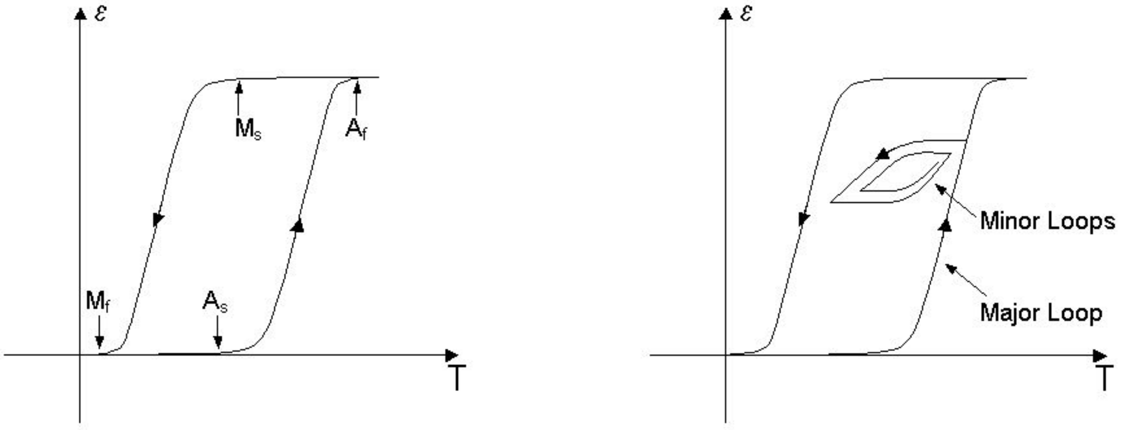


Figure 2.7: Microscopic and Macroscopic Illustration of Shape Memory Effect

The phase transformations do not occur instantaneously but over a temperature interval. During heating, the transition to austenite from martensite occurs over the range defined by the austenite start and finish temperatures, A_s and A_f respectively. Upon cooling, martensite formation occurs between the martensite start and finish temperatures, M_s and M_f respectively. The fact that the transformation temperatures are all different results in the tracing of a hysteresis curve as illustrated in Figure 2.8a. However, this hysteresis curve will only be followed if the temperature is monotonically increased or decreased during heating and cooling respectively. In addition to the “major loop” hysteresis curve of Figure 2.8a, there are “minor loops” that exist when the temperature is varied in-between the transition temperatures. These minor loops are illustrated in Figure 2.8b.

Hysteresis constitutes a major limitation for the use of SMAs and much research has been pursued to model and control this type of nonlinearity.

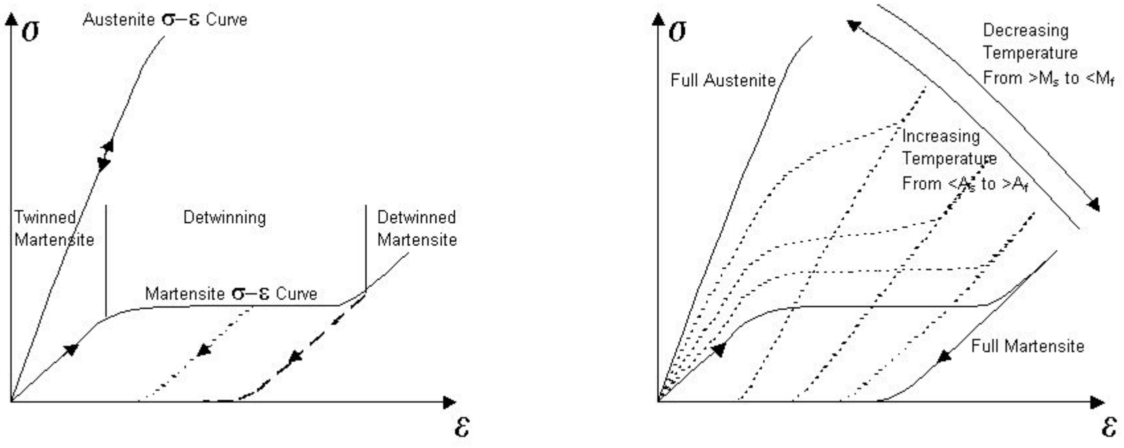


(a) (b)

Figure 2.8: Illustration of (a) Hysteresis and (b) Minor Loops in the Shape Memory Effect

Since austenite and martensite exhibit different micro-structures, their stress-strain curves also differ significantly. Figure 2.9a shows the stress-strain curve for austenite and martensite SMAs. As can be seen, in the austenite stage the stress-strain curve appears similar to most materials, but in the martensite stage, a plateau as well as different loading and unloading curves exist. This plateau is the area where detwinning occurs, shown by the area labelled "Twinned Martensite". Upon loading martensite, a "normal" elastic curve is initially followed until the plateau is reached. Then for a minor increase in stress, a large strain results due to detwinning. If the load is removed at any point on the plateau, a different unloading curve is formed shown in Figure 2.9a as the broken lines. The end of the plateau represents fully detwinned martensite. If the load is further increased, another elastic curve is seen.

During heating from A_s to A_f and cooling between M_s and M_f, the curves transition between each other. Figure 2.9b illustrates the transition from that of martensite to that of austenite and vice-versa. This transition plays an important role in allowing for the extraction of work from SMAs.



(a) (b)

Figure 2.9: Stress-Strain Curve of SMAs in (a) Full Martensite and (b) Full Martensite to Full Austenite

2.3.2 SMA Actuators

The mechanism for SMA actuators to perform work is based on the shape memory effect, and is best explained through a series of stress-strain curves during loading of SMAs.

The stress-strain curve of SMAs varies dependent on the amount of martensite and austenite that exists in the material, which can change according to temperature. An illustration of the stress-strain curves was shown in Figure 2.9.

An example of how the shape memory effect can be used to achieve work is shown in Figure 2.10. Starting at state A, an SMA wire in full martensite form is used to support two loads, L_a and L_b. It is assumed that at this state, the wire stretches until a state of fully detwinned martensite is reached.

If the wire is then heated, the stress-strain curve will shift such that for a constant stress, the strain decreases. As a result the wire length shrinks and lifts the two loads. Once the temperature reaches above A_f, the wire will be in full austenite and the system will reach state B of Figure 2.10.

Next, load L_a is removed from the wire. The system travels down the elastic curve of austenite to reach state C.

If the wire is then cooled, the stress-strain curve transitions back to that of martensite. Along this transition, the strain increases for a constant stress resulting in an elongation of the wire to state D.

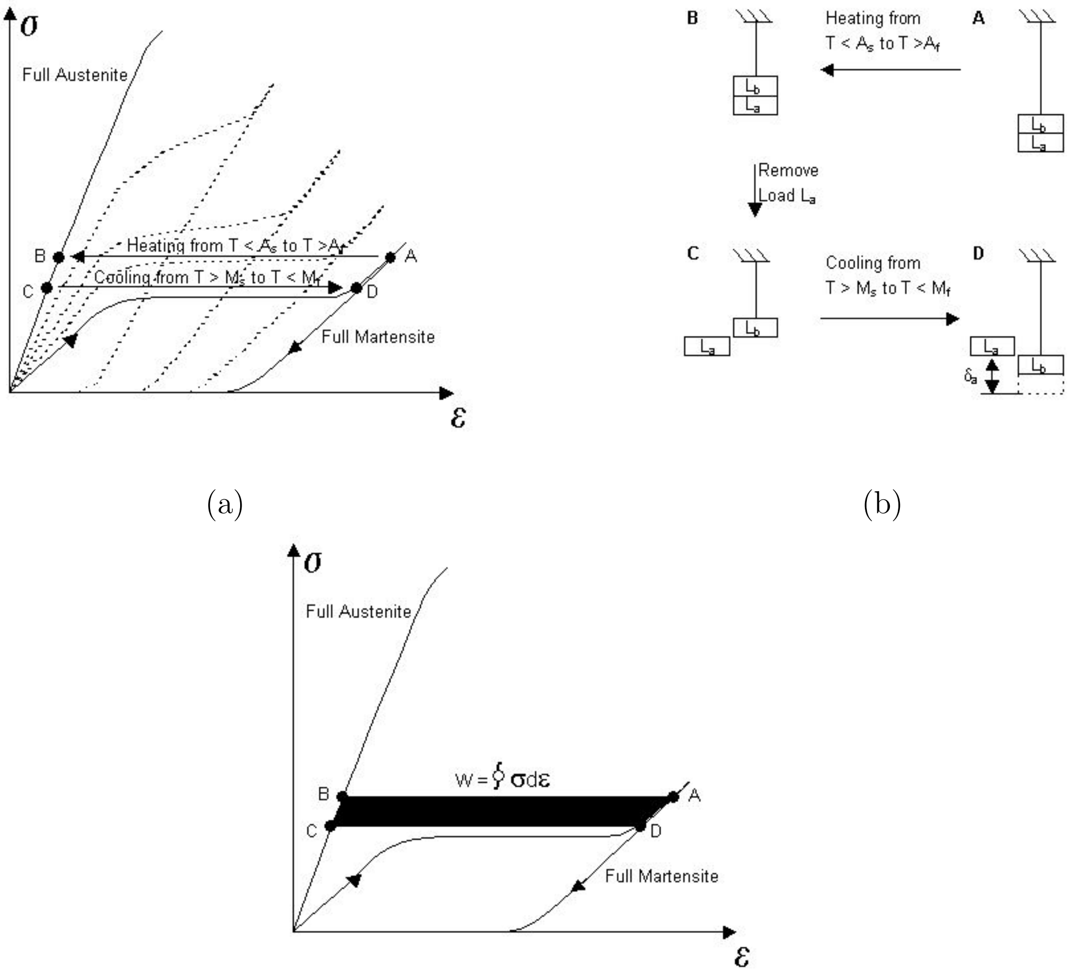
If another load of same magnitude as L_a is then added to the wire, the system will again be at state A. This results in a tracing of a hysteresis curve. The area within this curve, shown in Figure 2.10c, represents the work that is done to lift the load L_a a distance of δ_a.

It is important to note that to simplify the example, the bias load L_b was assumed to be constant. This, however, is not true for most actuators that use springs as a bias force, since they provide varying loads dependent on length. The only difference is that the stress does not remain constant, but increases or decreases during contraction and elongation dependent on the type of bias spring used. The overall analysis and achievement of work is however, similar.

2.3.3 Heating Methods

It has been shown that the shape memory effect of SMAs can be used to perform work by varying the temperature of the SMA between A_s, A_f, M_s and M_f. The rate at which the temperature can be changed is the main limitation of the bandwidth of SMA actuators. To best maximize the bandwidth, a fast and efficient method of varying the temperature is required. The most widely used method of heating the SMA is through Joule heating, while cooling is usually achieved through natural or forced convection.

Joule heating increases the temperature of a SMA wire as a result of current flow acting on the electrical resistance of the wire. A generally accepted heating relationship between the square of the current, i^2, and temperature, T, is given in (2.1) [16].



(a) (b)

(c)

Figure 2.10: (a) Strss-Strain Curve Illustration (b) Schematic Illustration and (c) Resultant Stress-Strain Curve Area of Work Using the Shape Memory Effect

$$\rho C_p V \frac{dT(t)}{dt} = Ri(t)^2 - hA(T(t) - T_\infty) \tag{2.1}$$

ρ = Density of Wire

V = Volume of Wire

A = Surface Area of Wire

R = Electrical Resistance of Wire

h = Convection Heat Transfer Coefficient

C_p = Specific Heat of Wire

T_∞ = Ambient Temperature

It is important to note that in this first order model the input current affects the temperature, and thus the length of an SMA actuating wire. However, for the SMA actuator used here, the current flow cannot be directly controlled. This is explained later in Section 3.2.1.

It has been shown that SMAs can be used to perform work through heating and transforming the material between full martensite and full austenite. However, for position control, the material is controlled to states that are between full martensite and full austenite. Between these two states, the operation is within the nonlinear region where hysteresis plays a major role. Therefore a model of this hysteresis would be beneficial.

2.4 Hysteresis Models

As noted previously, hysteresis constitutes a major drawback for the use of SMAs, and is the focus of much research. There are different methods of modelling hysteresis, two notable examples are the Preisach model outlined in [17] and the phaser model of [18], [19], and [20]. Here, the well established Preisach model provided by [17] is reviewed for use.

2.4.1 Preisach Model

The origins of this model can be traced back to a paper written by F. Preisach in 1935 to physically model hysteresis in magnetics. Since its first inception, it has undergone much modification and its application has expanded beyond the field of magnetism. It was not until the 1970s, when M. Krasnoselskii studied the model, that a purely mathematical form of the Preisach model was developed. This mathematical model allowed for a better understanding of hysteresis, as well as providing an effective modelling tool for hysteretic systems.

The Preisach model provides a mathematical model of hysteresis with nonlocal memory. Nonlocal memory is defined as follows: future values of output, $f(t)$ for all $t > t_0$, depend not only on the current value of output, $f(t_0)$, but also on past extremum values of the input. Whereas local memory is defined as follows: future values of output, $f(t)$ for all $t > t_0$, are uniquely predetermined by the output value at some instant of time, $f(t_0)$, and the input at all subsequent instants of time, $u(t)$ for all $t \geq t_0$.

In this model, hysteresis is essentially modelled as a weighted sum of counter-clockwise hysteresis operators, or relays, in parallel acting on a single input $u(t)$ as illustrated in Figure 2.11. Each relay, or operator $\hat{\gamma}_{\alpha\beta}$, is assigned different forward and reverse actuation thresholds (α and β respectively) with corresponding "on" and "off" values of 1 and -1 respectively. The relay outputs are then weighted according to a weighting function $\mu(\alpha, \beta)$ and summed. Thus the output can be mathematically summarized through equation (2.2).

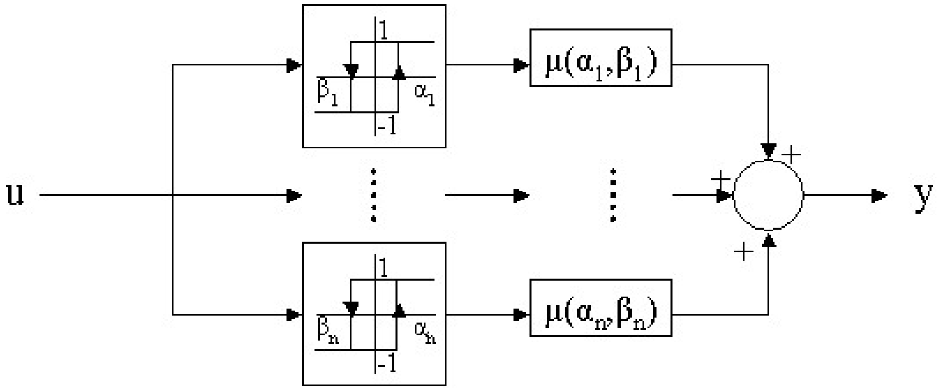


Figure 2.11: Relay Representation of Priesach Model

$$f(t) = \iint\limits_{\alpha \geq \beta} \mu(\alpha, \beta) \hat{\gamma}_{\alpha\beta} u(t) \, d\alpha \, d\beta \tag{2.2}$$

This structure gives rise to some interesting properties. The first of which is that though each single operator has local memory, the summation is nonlocal. Also, past input extrema history is stored within the operators, resulting in many possible branches of the hysteresis nonlinearity for the output to follow. The reasons behind these properties can be explained through the geometric representation of the Preisach model.

The geometric representation correlates the operators, $\hat{\gamma}_{\alpha\beta}$, to points on a half-plane defined by α and β where $\alpha \geq \beta$. This plane is referred to as the Preisach plane. The "on" and "off" threshold pair of each operator, (α, β) respectively, correspond to unique points on the plane. The half plane can be further constrained by a limiting triangle, T. T is a right triangle, whose hypotenuse is defined by $\alpha = \beta$, and sides defined as $\alpha = \alpha_0$ and $\beta = -\beta_0 = -\alpha_0$. α_0 and $-\beta_0$ define the maximum and minimum input values that affect the output, respectively. An illustration of this representation is shown in Figure 2.12.

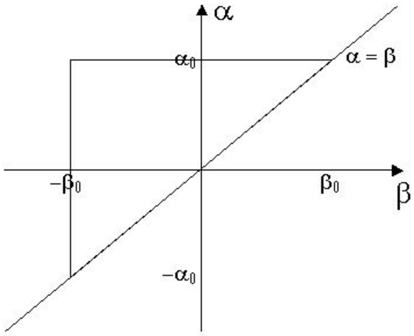


Figure 2.12: Preisach Plane

The operation of this system can be started by assuming the input is below $-\beta_0$ at some initial time t_0. At t_0, all the operators are "off" and output a value of -1. This is referred to as the state of "negative saturation", and the output value at this point is denoted as f^-. Conversely, if all the operators were "on", the state of "positive saturation" exists, and the output is denoted as f^+. It should be noted that $f^+ = -f^-$. Now assume a monotonically increasing input from u_0 to a value of u_1 at time t_1, where $-\beta_0 < u_1 < \alpha_0$. Every operator whose actuation threshold lies below u_1 switches "on" and outputs a value of +1. This leads to a subdivision of the triangle T into two sets, $S^+(t)$, the set of points whose operators are “on”, and $S^{-}(t)$, the set whose operators are “off”. The dividing line of the two sets, defined as *L(t)*, is $\alpha = u_1$. The state of the model at this point is shown in Figure 2.13.

If the input is now monotonically decreased to a value u_2 at time t_2, all the operators whose deactivation thresholds lie above u_2 are turned “off”. The division of the two sets, *L(t)*, now has a vertical and horizontal link as shown in Figure 2.14. The vertex of *L(t)* lies at $(\alpha, \beta) = (u_1, u_2)$.

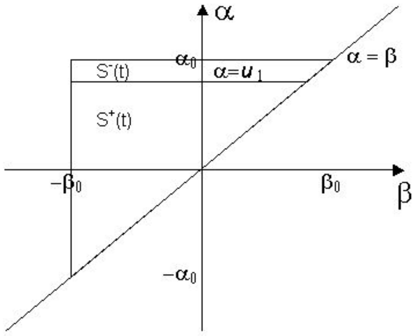


Figure 2.13: Preisach Plane with $u(t) = u_1$

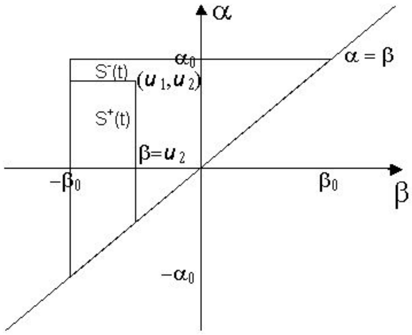


Figure 2.14: Preisach Plane with $u(t) = u_2$

If the input is then monotonically increased to a value u_3 at time t_3, where $u_3 < u_1$, a new horizontal link is generated and moves upwards on the plane until $\alpha = u_3$ is reached, as shown in Figure 2.15.

Finally, assuming that the input is monotonically decreased to u_4 at time t_4 where $u_4 > u_2$, a new vertical link is formed and moves left on the plane until $\beta = u_4$ is reached as shown in Figure 2.16. As a result a new vertex is formed at $(\alpha, \beta) = (u_3, u_4)$.

This process can be repeated any number of times to generate different vertices in *L(t)*. These vertices correspond to local input maxima and minima at past instants of time, resulting in a memory effect. Furthermore, the output of the system can be defined by the

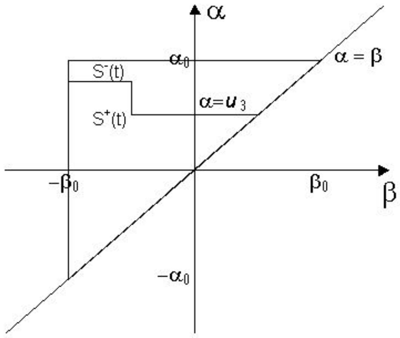


Figure 2.15: Preisach Plane with $u(t) = u_3$

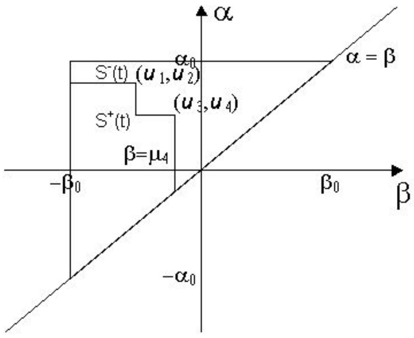


Figure 2.16: Preisach Plane with $u(t) = u_4$

summation of the two sets $S^+(t)$ and $S^-(t)$ given in equation (2.3).

$$f(t) = \iint\limits_{S^+(t)} \mu(\alpha,\beta)\hat{\gamma}_{\alpha\beta}u(t)d\alpha d\beta + \iint\limits_{S^-(t)} \mu(\alpha,\beta)\hat{\gamma}_{\alpha\beta}u(t)d\alpha d\beta \tag{2.3}$$

Since it is known that $\hat{\gamma}_{\alpha\beta} = +1$ for $(\alpha, \beta) \in S^+(t)$, and $\hat{\gamma}_{\alpha\beta} = -1$ for $(\alpha, \beta) \in S^-(t)$, (2.3) can be reduced to (2.4).

$$f(t) = \iint\limits_{S^+(t)} \mu(\alpha,\beta)d\alpha d\beta - \iint\limits_{S^-(t)} \mu(\alpha,\beta)d\alpha d\beta \tag{2.4}$$

The use of (2.4) for experimental modelling the response of SMA has been shown to be effective by Gorbet and Wang [21]. However, (2.4) is difficult to implement experimentally as it requires the evaluation of double integrals. Furthermore, the Preisach surface, $\mu(\alpha, \beta)$, is derived from differentiations of experimental data, resulting in possible noise error amplification [17]. Thus, a more robust experimental method has been derived.

First, (2.4) is slightly modified by recognizing $S^+(t)$ and $S^-(t)$ combined define the limiting triangle T, allowing for (2.5) to result. With this modification, it is noted that the first double integral is now over the entire triangle, which is the output at negative saturation, f^-, or inversely, $-f^+$.

$$f(t) = -\iint\limits_{T} \mu(\alpha,\beta)d\alpha d\beta + 2\iint\limits_{S^+(t)} \mu(\alpha,\beta)d\alpha d\beta \tag{2.5}$$

To evaluate the remaining double integral, the positive set S^+ is divided into a series of n trapezoids Q_k as shown in Figure 2.17.

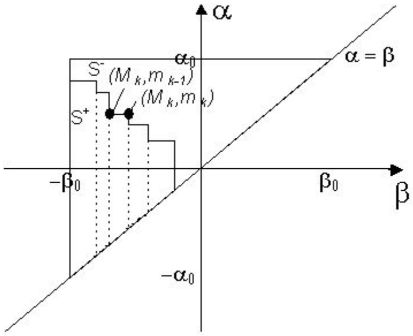


Figure 2.17: S^+ Preisach Plane Divided into Trapezoids

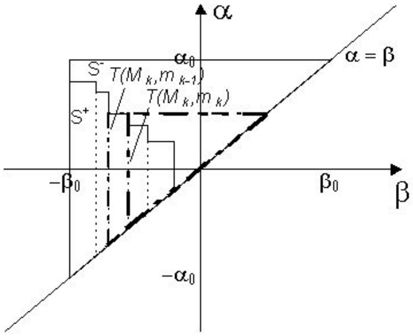


Figure 2.18: Illustration of Preisach Plane Triangles

The positive set S^+ can then be represented by the summation of the trapezoids through (2.6). Each corner is defined by its (α,β) pair of (M_k,m_k).

$$\iint\limits_{S^+(t)} \mu(\alpha,\beta)d\alpha d\beta = \sum_{k=1}^{n(t)} \iint\limits_{Q_k(t)} \mu(\alpha,\beta)d\alpha d\beta \tag{2.6}$$

Each single trapezoid, Q_k, can be represented by the difference of two triangles as highlighted in Figure 2.18. The triangles are defined by the hypotenuse $\alpha = \beta$ and corners (M_k,m_k) and (M_k,m_{k-1}). Thus, the integral over a single trapezoid can be represented through (2.7).

$$\iint\limits_{Q_k(t)} \mu(\alpha,\beta)d\alpha d\beta = \iint\limits_{T(M_k,m_{k-1})} \mu(\alpha,\beta)d\alpha d\beta - \iint\limits_{T(M_k,m_k)} \mu(\alpha,\beta)d\alpha d\beta \tag{2.7}$$

To derive the expression for the triangles, the first-order descending (FOD) curve is introduced. A first-order descending curve describes the output after a single reversal of the input. It is generated by starting the input below $-\beta_0$, producing the state of negative saturation. The input is then monotonically increased to some value $\alpha' < \alpha_0$ at t_1. During this increase the output follows the ascending branch of the major loop. The notation $f_{\alpha'}$ is used for the output along the major branch corresponding to an input of α' (see Figure 2.19). The FOD curve is then generated by monotonically decreasing the input between times t_1 and t_2. The notation $f_{\alpha'\beta'}$ is used to denote the output on the FOD curve attached to the major loop at $f_{\alpha'}$, and corresponding to an input value of $u = \beta'$.

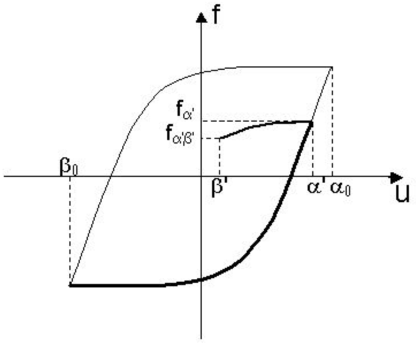


Figure 2.19: FOD Notation

Note that in the geometric representation, the triangle $T(\alpha', \beta')$ is traced during a first-order reversal, and relays in $T(\alpha', \beta')$ are moved from the set S^+ to S^-. $T(\alpha', \beta')$ can be thus derived as follows,

$$\begin{aligned}
\iint\limits_{T(\alpha',\beta')}\mu(\alpha,\beta)d\alpha d\beta &= \iint\limits_{S^+(t_1)}\mu(\alpha,\beta)\hat{\gamma}_{\alpha\beta}u(t)d\alpha d\beta - \iint\limits_{S^+(t_2)}\mu(\alpha,\beta)\hat{\gamma}_{\alpha\beta}u(t)d\alpha d\beta \\
&= \iint\limits_{S^-(t_1)}\mu(\alpha,\beta)\hat{\gamma}_{\alpha\beta}u(t)d\alpha d\beta - \iint\limits_{S^-(t_2)}\mu(\alpha,\beta)\hat{\gamma}_{\alpha\beta}u(t)d\alpha d\beta \\
\Rightarrow 2\iint\limits_{T(\alpha',\beta')}\mu(\alpha,\beta)d\alpha d\beta &= \iint\limits_{S^+(t_1)}\mu(\alpha,\beta)\hat{\gamma}_{\alpha\beta}u(t)d\alpha d\beta - \iint\limits_{S^+(t_2)}\mu(\alpha,\beta)\hat{\gamma}_{\alpha\beta}u(t)d\alpha d\beta \\
&\quad + \left[\iint\limits_{S^-(t_1)}\mu(\alpha,\beta)\hat{\gamma}_{\alpha\beta}u(t)d\alpha d\beta - \iint\limits_{S^-(t_2)}\mu(\alpha,\beta)\hat{\gamma}_{\alpha\beta}u(t)d\alpha d\beta\right]
\end{aligned}$$

Substituting (2.3),

$$\iint\limits_{T(\alpha',\beta')}\mu(\alpha,\beta)d\alpha d\beta = \frac{1}{2}(f_{\alpha'} - f_{\alpha',\beta'}) \tag{2.8}$$

Then relating $T(\alpha',\beta')$ to $T(M_k,m_k)$ results in equation (2.9).

$$\iint\limits_{T(M_k,m_k)}\mu(\alpha,\beta)d\alpha d\beta = \frac{1}{2}(f_{M_k} - f_{M_k,m_k}) \tag{2.9}$$

Combining each triangle (2.9) with (2.7), the equation of each trapezoid becomes (2.10).

$$\begin{aligned}
\iint\limits_{Q_k(t)}\mu(\alpha,\beta)d\alpha d\beta &= \frac{1}{2}(f_{M_k} - f_{M_k,m_k-1}) - \frac{1}{2}(f_{M_k} - f_{M_k,m_k}) \\
&= \frac{1}{2}(f_{M_k,m_k} - f_{M_k,m_{k-1}}) \qquad (2.10)
\end{aligned}$$

Back substituting (2.10) and (2.6) into (2.5), the final representation of the experimental implementation of the Preisach model is produced in (2.11), and is derived in [17].

$$f(t) = -f^+ + \sum_{k=1}^{n(t)-1}(f_{M_k,m_k} - f_{M_k,m_{k-1}}) + f_{M_n,m_n} - f_{M_n,m_{n-1}} \tag{2.11}$$

For cases where the final motion of the input is monotonically decreasing, $m_n = u(t)$. If the final direction of the input is monotonically increasing, then $M_n = m_n = u(t)$ and therefore $f_{M_n,m_n} = f_{u(t),u(t)} = f_{u(t)}$.

It has been shown in [17] that necessary and sufficient conditions for the existence of a Preisach model are that the system satisfy the Wiping Out and Congruent Minor Loop properties. It has been previously demonstrated in [22], at least in a frequency of interest, that SMAs do display this behaviour.

Chapter 3

Experimental System Description

The experimental system comprises a modified printer carriage to provide long range, macro-scale linear motion (approximately 200 mm range and 200 μm precision) and a micro-scale system (approximately 4 mm range and 5 μm target precision) that uses a NanoMuscle^{TM} SMA based actuator. However, the absolute resolution of the system is limited by the macro-stage precision of 200 μm. Therefore, for validation of the system, an Agilent laser measurement system is used to measure the absolute position of both the macro- and micro-stages to an accuracy of 2 μm. Both systems are monitored and computer controlled by Simulink^{TM} via a MultiQ-3 data acquisition board (MQ3) manufactured by Quanser Consulting.

3.1 Macro-Stage

The macro-stage consists of an ink-jet printer modified to interface with the MultiQ-3 D/A board and to accommodate the mounting of the micro-stage on the printer carriage. The actuating motor is controlled via two lead wires which are connected to a pulse width modulation (PWM) amplifier. The amplifier outputs a 20 V signal with the duty cycle tuned to 0 - 100 % for a 0 - 5 V analog output from the MQ3 board. A schematic block diagram of the electrical operation of the actuating motor is illustrated in Figure 3.1.

Position feedback is provided by an encoder that supplies approximately 200 μm resolution, already incorporated in the printer. It is connected to the encoder input of the

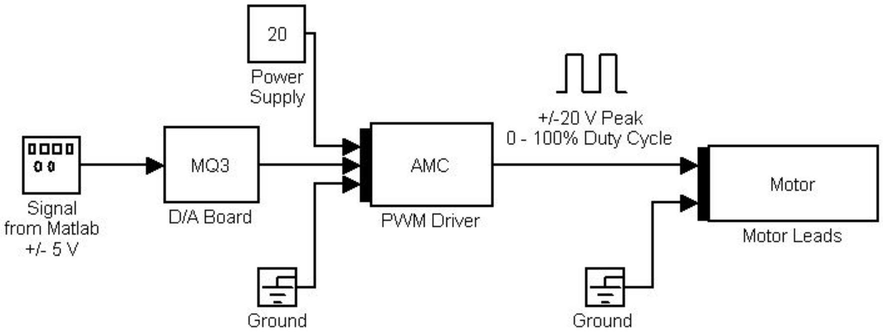


Figure 3.1: Block Diagram of Electrical Operation of Macro-Stage Motor

MQ3 board.

3.2 Micro-stage

The micro-stage consists of a NanoMuscle^{TM} SMA actuator in series with a linear flexure guide and a bias spring. Feedback is provided via a Hall effect sensor and magnets. A CAD illustration of the micro-stage is shown in Figure 3.2. It should be noted that the Hall effect sensor is not illustrated.

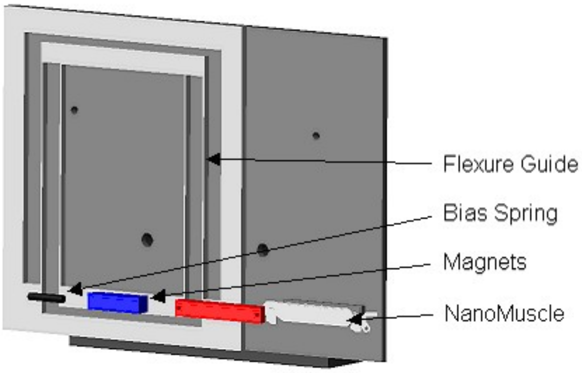
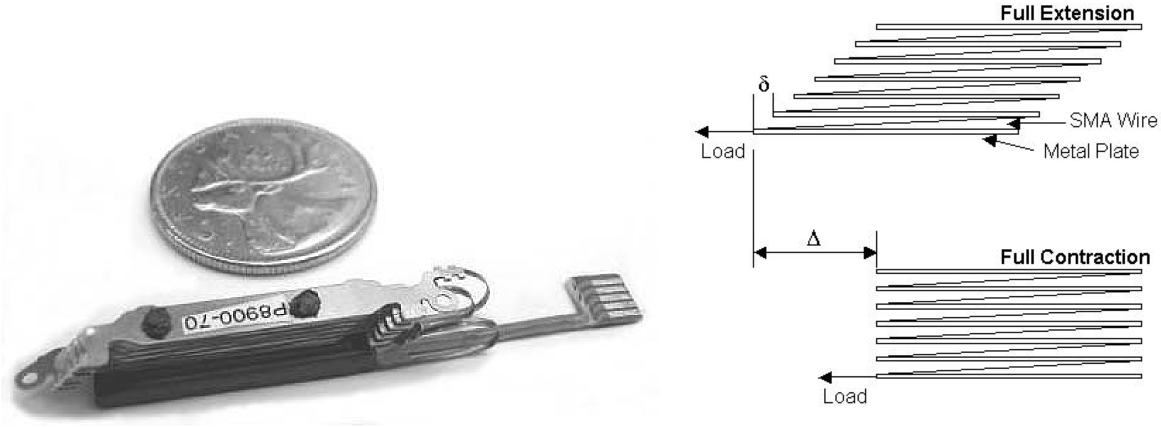


Figure 3.2: CAD Illustration of Micro-stage

3.2.1 NanoMuscle^{TM} SMA Actuator

The SMA actuator used, described in [23], is the NanoMuscle^{TM} NM70. It is a commercially available linear actuator that utilizes six NiTi SMA wires in series to generate the actuation force. Differential scanning calorimetry (DSC) curves for the SMA wires can be found in Appendix A. A photograph of the NanoMuscle^{TM} is shown in Figure 3.3a, and a schematic illustration of its operation is shown in Figure 3.3b. The NanoMuscle^{TM} consists of a series of alternating metal plates and SMA wires arranged in parallel configuration as seen in Figure 3.3b. The top metal plate is fixed while the bottom is attached to a load. When fully extended, the temperature of the SMA wires are below M_f and are in a state of detwinned martensite. Upon actuation, the wires are heated to above A_s causing the formation of austenite, and thus contraction of the wire length. Each wire pulls a plate by a distance δ relative to the plate above it. However, the total displacement experienced by the bottom plate, and hence the load, is the sum of all the δs, Δ. Δ is the stroke of the actuator which is 4 mm.



(a) (b)

Figure 3.3: (a) Photograph [23] and (b) Schematic Illustration of Operation of the NanoMuscle^{TM} Linear Actuator

The actuator's 4 mm stroke makes it a good physical match for the macro stage in this fine positioning application. In addition, the actuator is commercially available, and has a fully integrated digital interface, simplifying implementation. The digital interface provides access to information and control of the NanoMuscle^{TM}, through its internal circuitry. An electrical schematic of the NanoMuscle^{TM} is shown in Figure 3.4.

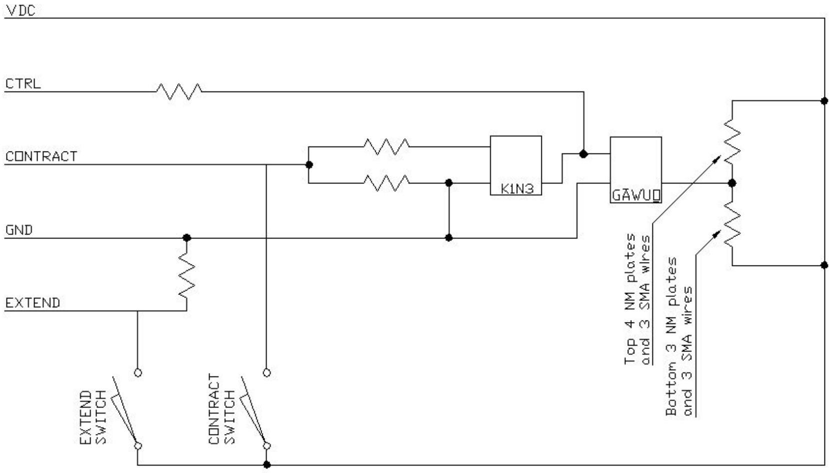


Figure 3.4: Electrical Schematic of NanoMuscle^{TM}

Information of whether the actuator is fully contracted or extended is provided by the signals *CONTRACT* and *EXTEND* respectively. It is important to note that Joule heating of the actuator is not directly controlled, but through a signal labelled as *CTRL*. A constant voltage is supplied to the wires through *VDC* and *GND*, and *CTRL* regulates the current flow through the SMA wires by way of an unknown component labelled as *GAWUO*. The fact that *CTRL* does not directly provide Joule heating of the actuator is important to note for actuator modelling, since a model of the internal actuator circuitry may be required.

Also, overheating is prevented by the *CONTRACT* signal. When this signal is pulled high to *VDC*, the component labelled as *K1N3* halts current flow through the SMA wires to prevent the actuator from overheating.

To operate the NanoMuscle^{TM}, a 4 V electrical power supply is connected to *VDC* and *GND*. *CTRL* is implemented through a 20 V PWM amplifier with the duty cycle tuned to 0 - 20 % for a 0 - 5 V analog output from the MQ3 board. The use of a PWM signal for the *CTRL* signal is recommended by NanoMuscle^{TM}. A schematic block of the electrical operation of the NanoMuscle^{TM} is illustrated in Figure 3.5.

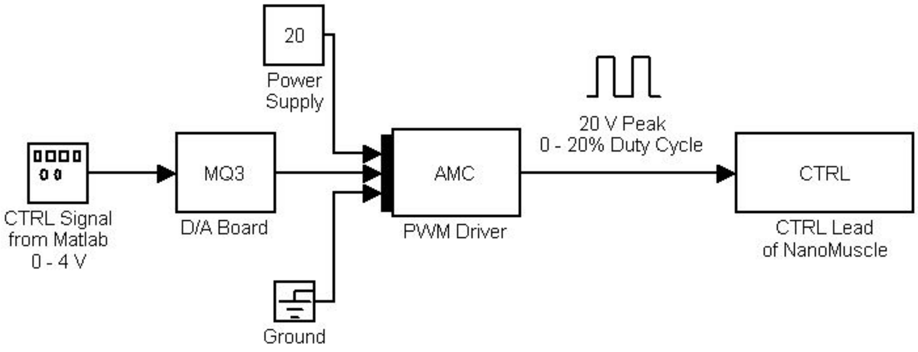


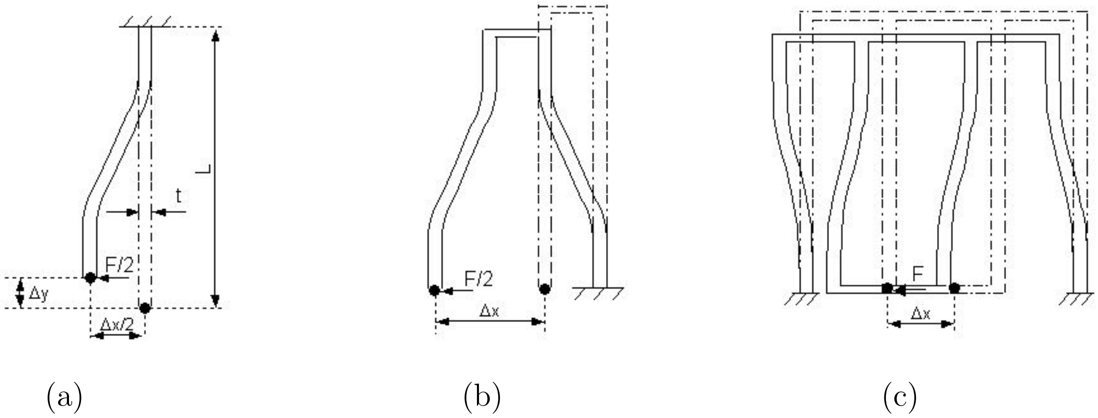
Figure 3.5: Block Diagram of Electrical Operation of NanoMuscle^{TM}

Although the actuator provides linear motion, a guiding system must be designed to support the sensors that will measure displacement, and to ensure linear motion is maintained during operation.

3.2.2 Linear Flexure Guide

Many possible guiding systems such as linear bearings or tracks can be used for the microstage, but they would introduce undesirable friction. Therefore, a linear flexure guide is implemented that eliminates friction completely from the guiding system. The flexure guide guarantees linear motion by following the principles of cantilever beams. The design adopted here follows the method that is outlined in [24].

As illustrated in Figure 3.6a, if the free end of a cantilever beam is subjected to horizontal force, the free end undergoes horizontal deflection, $\frac{\Delta x}{2}$. However, the beam must maintain a constant length, and thus the free end also deflects vertically by Δy. Note that the slope of the free end shown is constrained to zero, which is true for the final geometry of the flexure guide shown in Figure 3.6c. To compensate for the vertical deflection and maintain pure horizontal motion, two beams can be placed in parallel such that the vertical deflection of one beam is opposite to that of the second beam. This cancels any vertical motion of the free end. An illustration of this operation is illustrated in Figure 3.6b. Finally, four beams can then be combined to form a robust flexure guide that provides only linear horizontal motion as shown in Figure 3.6c.



(a) (b) (c)

Figure 3.6: Schematic of Flexure of (a) a Single Beam (b) Two Parallel Beams (c) Flexure Guide [24]

To design the flexure guide, only a single beam is first considered. Assuming a symmetric geometry and referring to Figure 3.6c, each pair of beams on either side of the guide will support half of the load, $\frac{F}{2}$, and accommodate the total displacement, Δx. Since each beam in a pair are connected end to end, the load is transmitted in full across, but the displacement is divided between each beam (see Figures 3.6a and 3.6b). Therefore each beam receives half the load, $\frac{F}{2}$, and undergoes half the displacement, $\frac{\Delta x}{2}$.

To assign dimensions to the beam, some design constraints must first be set. The stroke of the NanoMuscle is 4 mm and the maximum load is 70 gf, therefore the beam must deflect at least 2 mm for a 35 gf load. The material is chosen to be aluminum due to its low modulus of elasticity, 70 GPa, high yield strength, and high fatigue life. Due to its availability, the material thickness, w, is set at 2 mm. With these constraints only the beam length, L, and the thickness, t, need be set. The beam length can be constrained through limits on the size of the micro-stage. The mounting plate on the printer head on which the micro-stage is secured has a height of 80 mm, therefore the chosen beam length must be below this; a somewhat arbitrary length of 55 mm is chosen. The remaining beam thickness can be set through the analysis of the deflection under a load of 35 gf. The deflection of a beam with the ends constrained to zero slope follows the cantilever beam equation (3.1) [24].

$$\Delta x = \frac{FL^3}{12EI} \tag{3.1}$$

Where I, the moment of inertia for a rectangular beam is given by,

$$I = \frac{wt^3}{12}$$

Applying 3.1 with the stated constraints, a maximum t of 0.465 mm results; the chosen thickness is 0.45 mm. Since only the beam size is critical to the operation of the guide, the remaining dimensions required to form the flexure guide can be somewhat arbitrarily chosen to fit the mounting area on the printer head. The final design of the flexure guide is given in Appendix B.

With the guiding system and actuator of the micro-stage designed, only the feedback system still remains to be set.

3.2.3 Feedback System

Many possible sensing approaches to measure displacement of the micro-stage exist, however some requirements limit the available choices. The feedback system must be small enough to fit on the printer carriage mount, light enough for the guide to support, be of low cost ($< \$100$), and be non-contact to avoid introducing friction nonlinearities. The optimal feedback system that fits all the requirements was chosen to be a combination of a Hall effect sensor and magnets. The Hall effect sensor measures the magnetic flux from magnets mounted on the moving platform. Bipolar slide-by operation is used, where two magnets of opposite polarities are mounted beside each other forming a relatively linear flux profile similar to that of Figure 3.7. When the platform is moved linearly, the Hall effect sensor slides by the magnets detecting changes in flux and thus position. To implement this system, the magnets and Hall effect sensor must be selected.

Magnet Selection

The required range of motion is only 4 mm. Therefore, the magnet with the largest flux spread over 4 mm needs to be found. Honeywell/Microswitch readily supplies magnetic

Table 3.1: Magnet Comparison in Head-on Operation [25]

Distance (in)	Honeywell/Microswitch Magnet						
	103MG5	101MG2L1	101MG3	107MG1	103MG8	101MG7	103MG6
0	800	800	800	800	800	800	800
0.02	760	725					
0.04	510	550				800	
0.06	300	420					
0.08	200	320					
0.1	125	260	790	780	800	520	260
0.12	100	200					
0.2			425	410	375	210	
0.3			260		190	110	
Approximate Largest flux change over 1mm (0.04in)	460	305	144	146	167	122	213

flux profiles for their products, and thus, their magnets are used as a basis for study. To simplify the analysis, the largest change of flux over a range of 1 mm in head-on operation is used as the comparison variable since this information is easily obtained.

As can be seen in Table 3.1, the 103MG5 magnet has the largest magnetic flux spread over a range of 1 mm for head-on operation. For bipolar slide-by operation the flux profile for the 103MG5 is given in Figure 3.7.

The 103MG5 is the optimal choice for use in the feedback system. Unfortunately, it was not available in small quantities. After consulting with vendors, an equivalent magnet of similar size and material was chosen. The selected magnet supplied by Dexter Magnets, is a pressed samarium cobalt cylinder magnet (Part Number: PR28CO120B), with a diameter of 3 mm and length of 2.5 mm.

Hall Sensor Selection

To select the optimal Hall effect sensor, some basic requirements must first be laid out. The three requirements are:

- the operation voltage must be between ±5 Vdc, the same analog input range as the

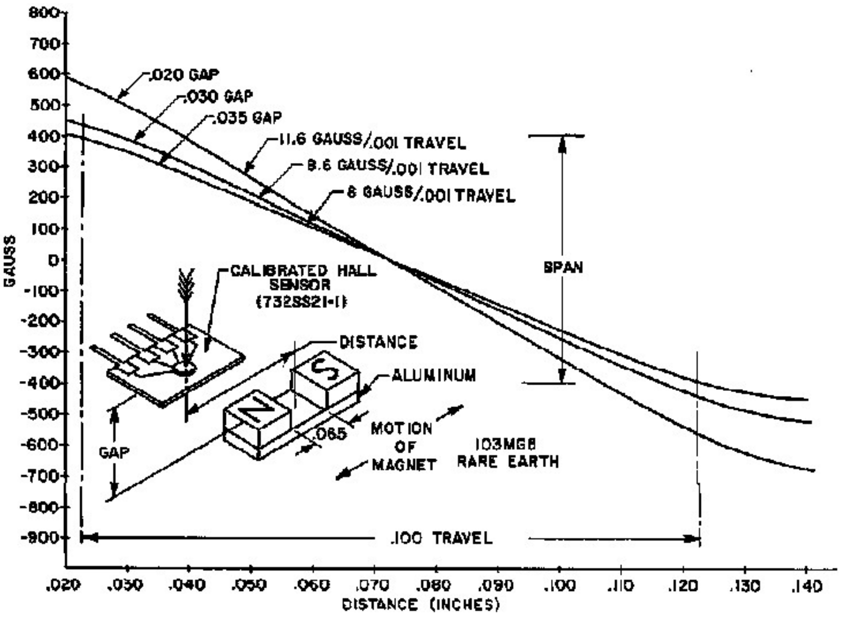


Figure C-21 (2) Rare earth .078" x .040" thick magnet.

Figure 3.7: Flux Profile of 103MG5 Magnet in Bipolar Slide-by Operation [25]

Table 3.2: Comparison of Hall Effect Sensors

	Honeywell/Microswitch Sensor [25]						Melexis [26]
	SS495A	SS495A1	SS495A2	SS496A	SS94A1B	SS94B1	90215
Range (Gauss)	±670	±670	±670	±840	±500	±670	n/a
Sensitivity (mV/Gauss)	3.125 ±0.125	3.125 ±0.094	3.125 ±0.156	2.5 ±0.100	1.875 ±0.100	3.125 ±0.063	0.5 - 14
Cost ($ USD)	∼10						∼100

MQ3.

- the maximum magnetic flux the sensor can measure must be greater than 650 Gauss, a slightly larger flux than the maximum provided by the 103MG5 magnet in Figure 3.7.

- the sensor sensitivity should be as large as possible. However, a lower limit on the sensitivity can be derived by dividing the average output range of Hall effect sensors, 0 - 4 V, by the span of the magnetic flux of the chosen magnet, ±650 Gauss. The resulting sensitivity should be at least 3 mV/Gauss.

- the monetary cost must be as low as possible.

Applying the general requirements gives 5 possible choices as seen in Table 3.2; they are the SS495A, SS495A1, SS495A2, SS94B1 and the Melexis 90215.

The Melexis sensor is programmable and meets all the requirements. It appears to be the most versatile with a variable sensitivity of up to 14 mV/Gauss and hence a variable magnetic range of operation. However, additional equipment to program the sensor incurs a high cost, and additional time is required to initially program the sensor. Also, sensitivity tolerances are not listed giving rise to questions about its accuracy.

Aside from the Melexis sensor, the rest of the sensors are identical in specification and cost except for the sensitivity tolerances. Since the SS94B1 sensor gives the highest precision, and is inexpensive, it is used here.

Sensor and Magnet Placement

To mount the sensor and magnets to the micro-stage, two magnets are press fit beside each other into a small teflon bar. They are 4 mm apart with opposite polarities facing the sensor. The magnetic assembly is then mounted onto the moving platform of the flexure guide. The sensor is glued with epoxy to a mount directly below the one of the magnets when the NanoMuscle^{TM} is fully extended. When the actuator is contracts, the magnets will move until the other magnet lies above the sensor at full contraction. Figure 3.8 shows the mounting and operation of the sensor system.

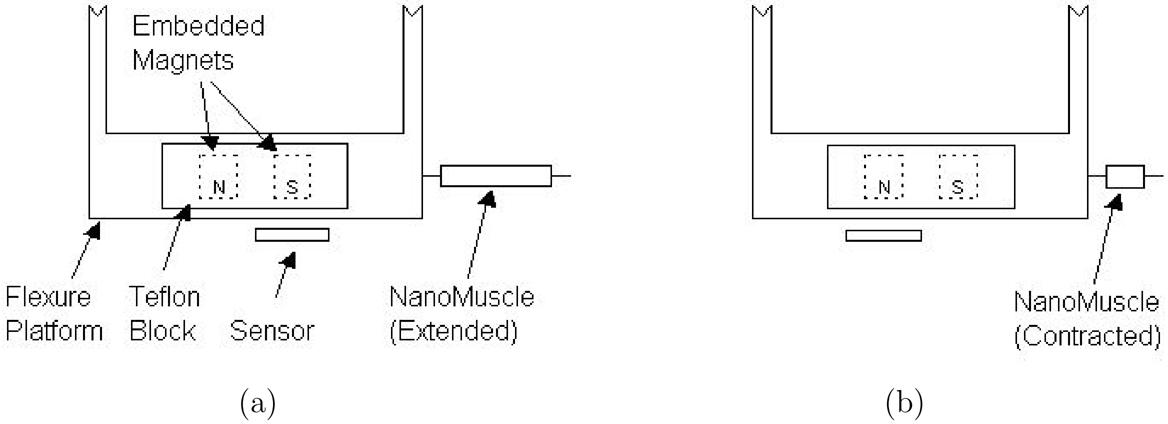


Figure 3.8: Illustration of the Micro-stage Sensor System at Full (a) Extenstion and (b) Contraction

The vertical spacing of the magnets from the sensor is important to produce the optimal flux profile for the sensor to measure. The spacing is set experimentally by positioning each magnet to touch and thus saturate the sensor to 0 V or 5 V. The sensor reads 2.5 V nominally. Each magnet is then raised to the point where the sensor is no longer saturated at each end of travel. However, the sensor must still be calibrated to correlate the flux reading to micro-stage position, due to the nonlinear relationship between field strength and relative position.

Target Accuracy

To estimate the accuracy that will be attained by the proposed system, the resolution of the A/D conversion is first found. The MQ3 board uses 12 bits plus a sign bit to represent input voltages between 0 - 5 V. This translates to a resolution of 1.22 mV/bit. Also, the chosen Hall effect sensor, the SS94B1, has a resolution of at most 3.125 mV/Gauss. If the two resolutions are then combined, the resulting resolution is 0.3904 Gauss/bit.

To convert magnetic flux to distance, the flux profile of the magnets is considered. Assuming a motion of 4 mm, the magnets will be used for a range of motion of 0.16 in, or between ±0.08 in. However, the worst case scenario for the magnetic flux occurs at the extremities of ±0.08 in. Therefore, the flux density at the extremities is used for analysis. Assuming that the magnetic profile matches that of Figure 3.7 for the 103MG5 magnet and that the flux curve is linear, the slope for the flux at the extremes of travel is approximately 5000 Gauss/in or 5 μm/Gauss.

Combining this slope of 5 μm/Gauss with the resolution of 0.3904 Gauss/bit found previously, a resolution of 1.952 μm/bit results. This is the theoretical resolution if no noise exists in the system. Since it is expected that electrical noise will exist, the expected accuracy is slightly decreased to an estimated accuracy of 5 μm. This is set as the target accuracy to attain.

Hall Effect Sensor Calibration

To calibrate the sensor, an Agilent laser measurement system is used to measure the displacement of the micro-stage to a precision of ±2 μm. A retroreflector, a three sided mirror, is mounted onto the platform of the flexure guide. It reflects a laser beam into a receiver where an interference pattern is produced when the reflected beam interacts with a reference beam. The receiver is then read by an A/D board to provide position data of the micro-stage. A detailed description of the operation of the laser can be found in [27].

The sensor is calibrated by correlating the Hall effect sensor reading to the actual position read by the laser, and averaging the results over runs in opposite directions. Figure 3.9 shows the experimental and the averaged profile of the position with respect to sensor reading. It must be noted that the experimental and averaged profiles lie atop one another and are therefore difficult to discern.

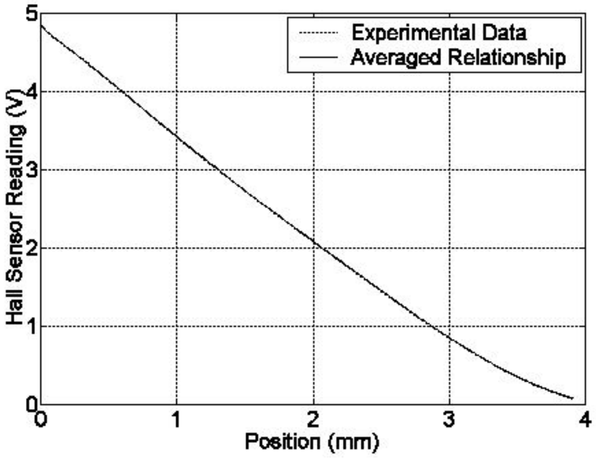


Figure 3.9: Position to Hall Effect Sensor Reading Relationship

From analyzing Figure 3.9, it can be seen that the sensor provides approximately 0.8 μm/mV resolution. Combining this with the MQ3 measurement resolution of 1.22 mV/bit results in an attainable resolution 0.976 μm/bit. This is better than the previously estimated resolution of 1.952 μm/bit for no noise. However, the expected accuracy during operation is still estimated to be 5 μm.

3.3 Electrical Interfacing

With the physical experimental system designed, the wiring and interface with a PC still remains to be completed. An electrical schematic diagram of the entire system can be found in Appendix C. To interface the macro-stage motor and encoder, and micro-stage SMA actuator and Hall effect sensor with Simulink for monitoring and control, the MultiQ-3 data acquisition board by Quanser Consulting is used. Information for the data acquisition board can be found in [28].

3.4 Complete Experimental System

With the entire system designed, its implementation is shown in Figure 3.10.

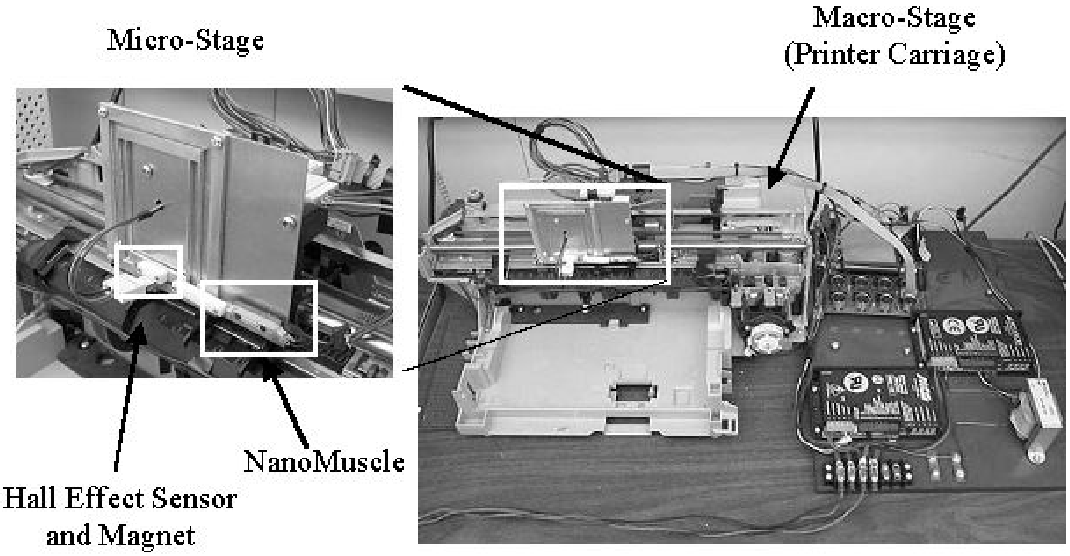


Figure 3.10: Final Setup of Macro-Micro Positioning System

Chapter 4

Macro-Stage Modelling and Control

The macro-stage, shown previously in Figure 3.10, provides long range motion of approximately 200 mm and precision of 200 μm through an optical encoder. The stage consists of an ink-jet printer modified to interface with the MultiQ-3 D/A board and to accommodate the mounting of the micro-stage on the printer carriage. To allow for offline controller design of the macro-stage, a mathematical open loop model of the stage is identified. The system is modeled as a sliding mass actuated by a DC motor as illustrated schematically in Figure 4.1. It should be noted that to prevent non-axial flexure of the drive belt, the sliding mass is guided by a bushing and rod within the printer. This ensures that only horizontal motion is possible. Also, to simplify the analysis, an assumption is made that the drive belt is highly rigid and renders nonlinearities due to belt flexure insignificant.

The system has significant friction nonlinearities, as well as control signal saturation imposed by the drive electronics. Assuming that the nonlinearities can be separated from the linear dynamics, friction and controller saturation can be modelled independently.

4.1 Modelling and Control Overview

Friction is a major nonlinearity affecting many systems, and many friction models exist. Notable examples are the Dahl and Bristle models described by Olsson *et al.* [29] and the LuGre model presented by Canudas de Wit *et al.* [30]. Most models attempt to develop a dynamic representation of friction to incorporate Coulomb, viscous and pre-

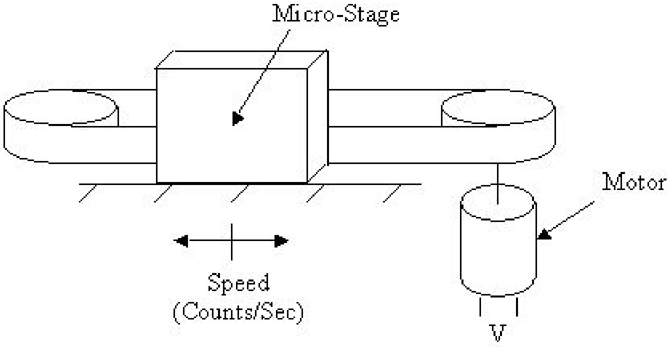


Figure 4.1: Schematic Illustration of the Macro-Stage

sliding frictional effects, resulting in models with very complex dynamics. However, in Kubo *et al.* [31], friction is modelled effectively as a dead zone. While this model is somewhat simplistic, it has the advantage of being easily invertible, a property which can be used in controller design. For a more thorough summary of friction models, refer to the survey of friction compensation methods published by Armstrong-Helouvry *et al.* [32].

4.1.1 Methodology

In open loop identification of linear systems, an input such as a step is applied and the output is measured and correlated with the input to derive a transfer function. One method of correlation used here is the modified least squares (MLS), as described by Ljung [33].

In this work, the simple dead zone model of Kubo *et al.* [31] is first applied to approximately linearize the system. After then applying MLS identification and evaluating the result, the dead zone model is extended to more accurately represent the system dynamics. The resulting identified models are subjected to different inputs in simulation and the results compared with experimental responses in order to evaluate and refine the model.

The MLS algorithm is implemented through the least squares algorithm with the addition of a loop. The additional loop filters bias noise from the applied experimental data. The least squares algorithm assumes an input to output relationship as follows:

$$Y = UB$$

Least squares attempts to find the optimal weighting sequence, B, for a given set of input and output data, Y and U respectively, through the following equation:

$$B = (U^T U)^{-1} U^T Y$$

To identify a model with the least squares algorithm, the format of the transfer function is assumed to be:

$$Y(z) = \frac{b_1 z + .. + b_m z^{-m}}{1 + a_1 z + .. + a_n z^{-n}} U(z)$$

The transfer function can be rearranged into a state space format as follows:

$$\begin{bmatrix} y_k \\ : \\ y_{k+r} \end{bmatrix} = \begin{bmatrix} y_{k-1} & .. & y_{k-n} & u_{k-1} & .. & u_{k-m} \\ : & : & : & : & : & : \\ : & : & : & : & : & : \end{bmatrix} \begin{bmatrix} -a_1 \\ : \\ -a_n \\ -b_1 \\ : \\ -b_m \end{bmatrix}$$

In this state space format, the least squares algorithm can be applied using experimental input and output data to find the optimal transfer function coefficients. The MLS algorithm is then implemented by "modifying" the least squares algorithm with the addition of a loop to filter bias noise from the experimental data. A flowchart of the implementation is shown in Figure 4.2.

After developing a model of the macro-stage, a closed loop controller is designed using the model and applied to the experimental system.

4.2 System Modelling

Since identification techniques try to match the simulated and actual output by fitting parameter values to a function given by the user, the mathematical form of the system is first postulated. Shown in Figure 4.3, the linear model of a motor exists within the the block diagram of the macro-stage. However, controller saturation that exists at the input

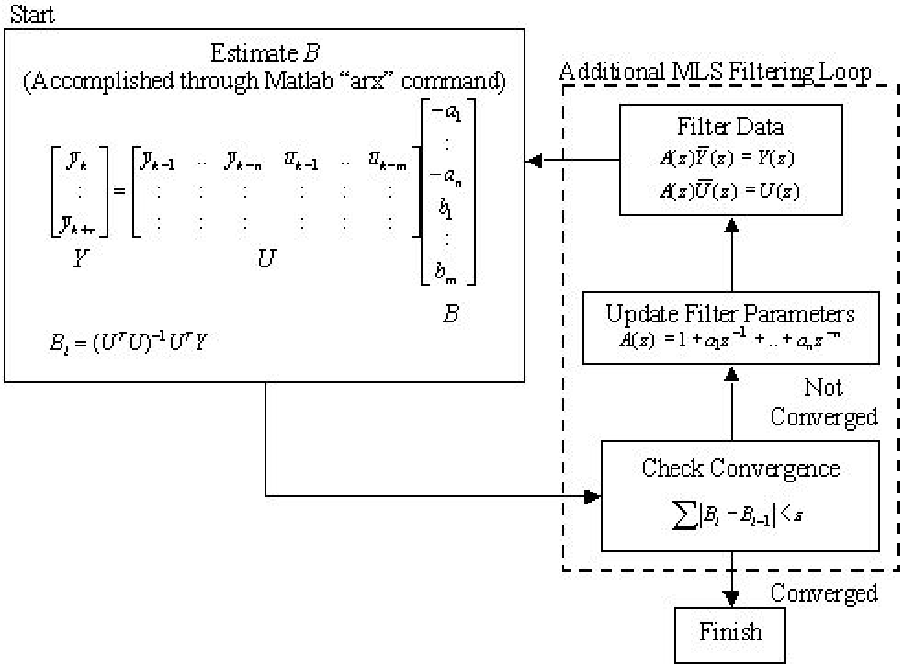


Figure 4.2: Flowchart of Modified Least Squares Implementation

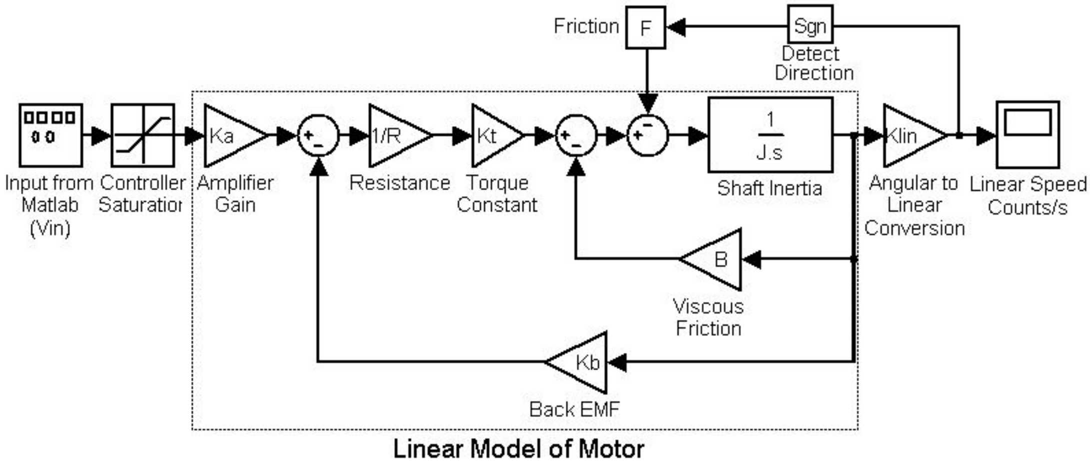


Figure 4.3: Block Diagram of Macro-Stage Motor

to the linear model of the motor, and the sliding friction that acts as a torque disturbance to the motor shaft, introduces nonlinearities. These nonlinearities must be removed to discover the linear representation of the motor.

4.2.1 System Nonlinearities

The system displays both friction and drive saturation nonlinearities. In modelling, these are accounted for as follows:

1. Saturation

Saturation in the drive electronics is assumed to be separable from the mechanical drive system dynamics. System identification inputs are chosen so as to limit the drive signal to less than 5 V, in order to avoid saturation during identification.

2. Friction Effects

Viscous friction effects are assumed to be captured in the linear motor dynamics, and is represented in Figure 4.3 by the damping coefficient, B, multiplied by the motor shaft speed. The Coulomb friction is assumed to be constant. The adopted model of Coulomb friction is shown in Figure 4.4.

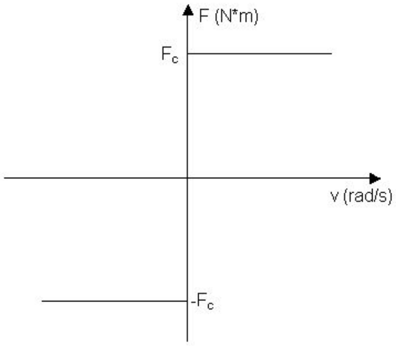


Figure 4.4: Input-Output Response of Coulomb Friction Model

This friction model represents F in Figure 4.3 and introduces a nonlinearity in the system. However, only the input to output behaviour is important. Therefore, F can be shifted elsewhere in the block diagram without affecting the input-output characteristics. As a result, if F is moved to the input, the resulting block diagram is shown in Figure 4.5.

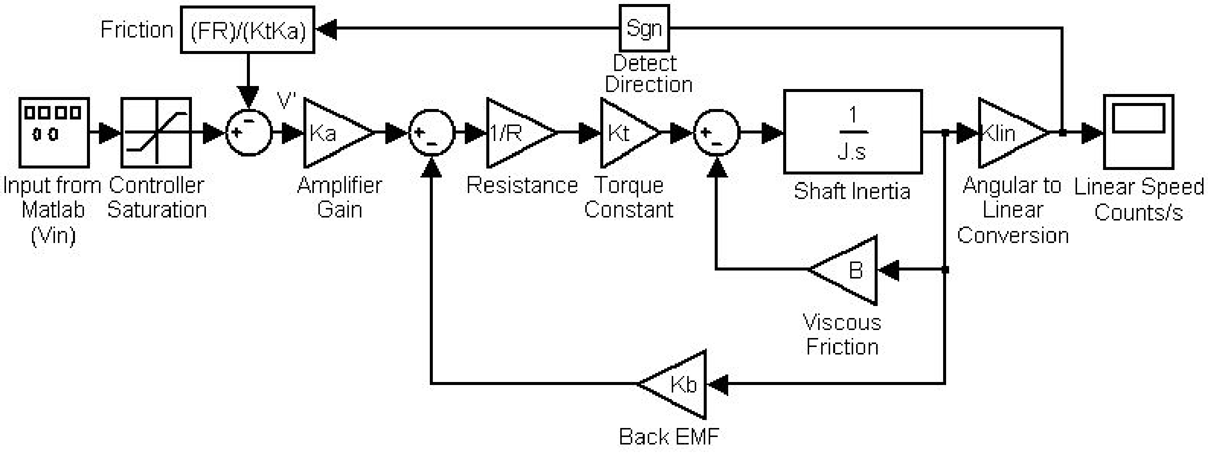


Figure 4.5: Block Diagram of Motor with Coulomb Friction Moved to Input

As a result, the remaining system that lies ahead of the friction is that of a linear motor.

4.2.2 Linear Dynamics

The macro-stage actuator is a DC motor for which a common state-space model is represented in (4.1).

$$
\begin{aligned}
\begin{bmatrix} \dot{i} \\ \dot{\Theta} \end{bmatrix} &= \begin{bmatrix} -\frac{R}{L} & \frac{K_b}{L} \\ \frac{K_t}{J} & -\frac{B}{J} \end{bmatrix} \begin{bmatrix} i \\ \Theta \end{bmatrix} + \begin{bmatrix} \frac{K_a}{L} \\ 0 \end{bmatrix} V \\
\dot{y} &= \begin{bmatrix} 0 & K_{lin} \end{bmatrix} \begin{bmatrix} i \\ \dot{\Theta} \end{bmatrix}
\end{aligned}
\tag{4.1}
$$

K_a = Amplifier Gain
R = Armature Resistance
L = Motor Inductance
Kb = Back EMF Gain
Kt = Torque Constant
B = Viscous Damping
J = Motor Shaft Inertia
i = Armature Current
V = Applied Voltage
Θ = Angular Position
K_{lin} = Angular to linear speed Conversion
y = Linear Position

It must be noted that the common assumption that the motor inductance L is insignificant compared to armature resistance R is made here. This is often true since the inductance is generally two magnitudes smaller than the armature resistance, allowing for first order behaviour of the motor [34]. If this assumption is false, the open loop behaviour would better resemble a second order response. Applying this assumption to the state-space model and converting to a transfer function, the result is given in (4.2).

$$
\frac{\dot{Y}}{V} = \frac{\frac{K_a K_{lin} K_t}{JR}}{s + \frac{BR + K_t K_b}{JR}} \tag{4.2}
$$

For MLS identification, the transfer function is converted into discrete time and replaced with the first order model of (4.3).

$$
\frac{\dot{Y}}{V} = \frac{b}{z + a} \tag{4.3}
$$

Where,

$$
\begin{aligned}
a &= e^{-[\frac{BR+K_tK_b}{JR}]T} \\
b &= -\frac{K_aK_{lin}K_t}{BR+K_tK_b}\left[e^{-[\frac{BR+K_tK_b}{JR}]T}-1\right]
\end{aligned}
$$

The remaining system after the friction in Figure 4.5 can now be exchanged with (4.3) to give the block diagram of Figure 4.6.

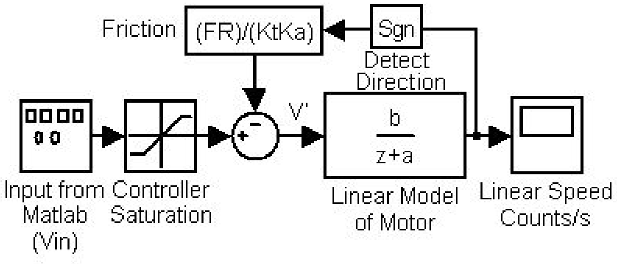


Figure 4.6: Block Diagram of First Order Model of Motor with Coulomb Friction

The parameters a and b can be identified by MLS, however the system is nonlinear due to saturation and friction.

4.2.3 Linearization for Identification

To approximately linearize the system, the friction can be identified and linearized. In the system diagram of Figure 4.6, the signal V' will be non-zero only if $|V_{in}| > \frac{F_sR}{K_tK_a}$. This is equivalent to a dead zone to the input signal. Therefore pre-sliding friction, or stiction, is initially modeled as a dead zone in the control signal to produce a simplified model of the macro-stage shown in the block diagram of Figure 4.7.

To characterize the dead zone, various input voltages are applied to the motor, by way of the MQ3 D/A board, until uninterrupted motion is observed. The results are shown in Figure 4.8.

As can be seen in Figure 4.8, the threshold voltage for motion varies depending on the position, as indicated by the highly nonlinear motion at lower drive voltages. The time periods where little motion is seen indicates areas where larger magnitudes of friction, and

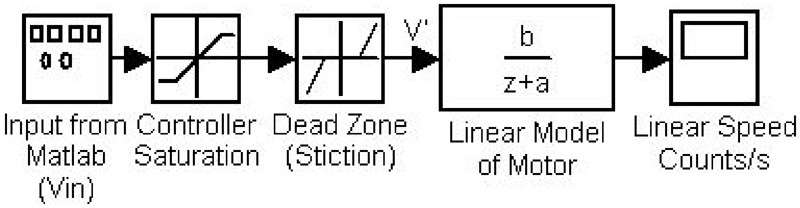


Figure 4.7: Block Diagram of First Order Model of Motor with Dead Zone as Coulomb Friction

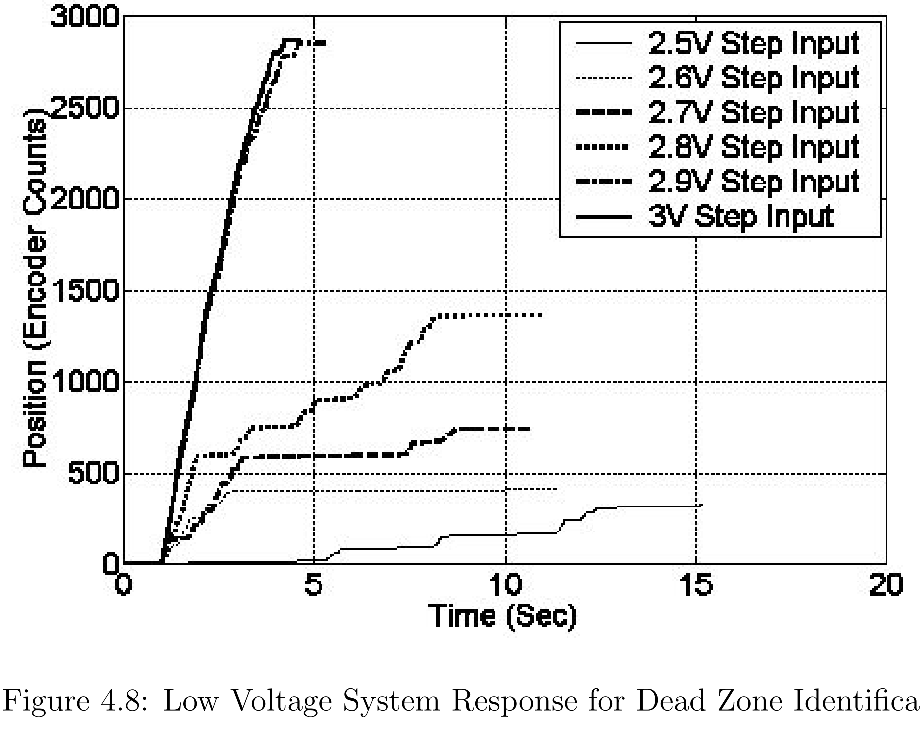


Figure 4.8: Low Voltage System Response for Dead Zone Identification

thus threshold voltage, exists. The maximum threshold voltage appears to be approximately 3 V, since this produces relatively smooth motion throughout the entire range of travel of the system. Thus, the dead zone width is set at 3 V. Open loop compensation for this model of friction can be easily achieved, since the dead zone is invertible. This is similar to the compensation scheme pursued by Kubo *et al.* [31]. The resulting model and its linearization is illustrated in Figure 4.9.

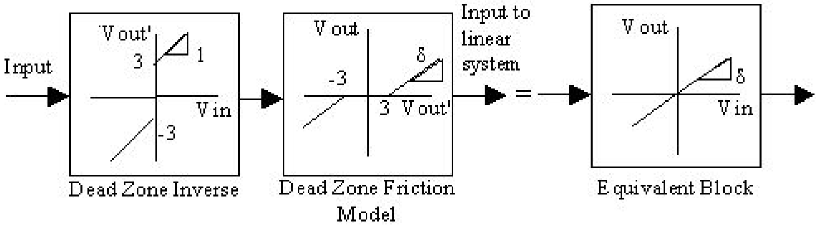


Figure 4.9: Initial Friction Model and Linearization

It should be noted that with this linearization scheme, the inputs are offset by ±3 V through the dead zone inverse. Since the output of the dead zone inverse is applied to the D/A board which has a ±5 V saturation limit, to avoid controller saturation the inputs must be limited to ±2 V.

4.2.4 Open Loop System Model

Combining the linear and nonlinear components of the system, the resulting open loop model is shown in Figure 4.10. The system model shown in Figure 4.10 is identified using MLS. The Actual Plant Model represents the system hardware; the Dead Zone Inverse is implemented in software. Assuming the validity of the dead zone friction model, and that test inputs are limited to avoid controller saturation, the system to be identified appears linear from input to output. This identification results in a model for the "Linear Dynamics block", which is combined with the dead zone and saturation nonlinearities to give the "initial nonlinear model". After evaluation, the dead zone friction model is extended and tuned, resulting in the "tuned nonlinear model".

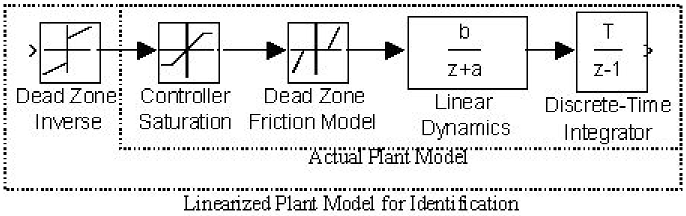


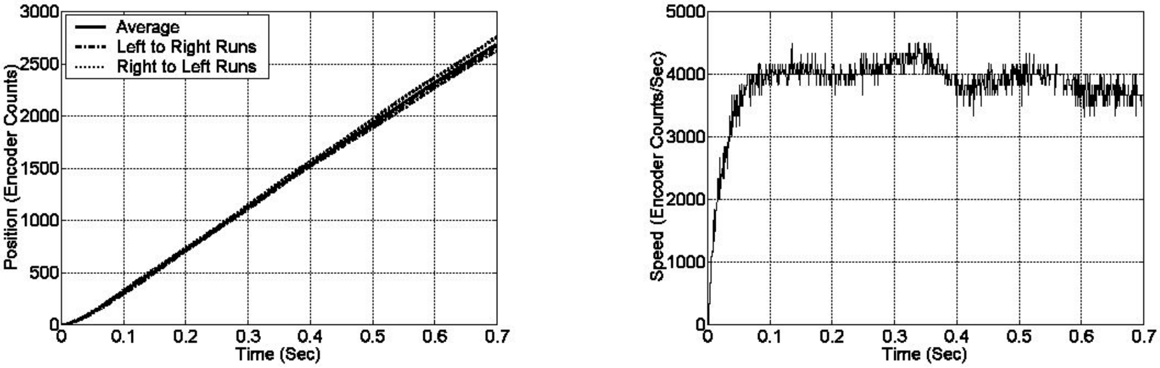
Figure 4.10: Block Diagram of the Open Loop System with Initial Friction Model

4.2.5 Open Loop System Identification

The open loop system is identified using a series of step input functions that are less than two volts (less than five volts, after dead zone compensation to avoid saturation). The resulting output position is then differentiated to calculate the speed, to which MLS is applied to determine parameters a and b from (4.3). The resulting transfer function is multiplied by an integrator to give the transfer function from input voltage V to output position Y.

Identification Data

The data used for identification is summarized in Figure 4.11. A series of runs in opposite directions are averaged together and then differentiated to calculate the resulting speed.



(a) (b)

Figure 4.11: Open Loop (a) Position and (b) Speed Response for 1 V Step Input

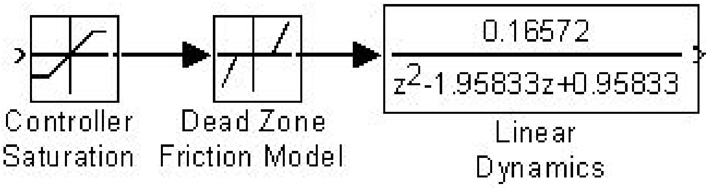


Figure 4.12: Macro Stage Initial Nonlinear Simulation Model

The “noise” which appears in the speed signal is due to differentiation of the quantized position data. This output noise, if assumed to be similar measurement noise, is easily handled by the MLS identification technique [33].

Linear Model

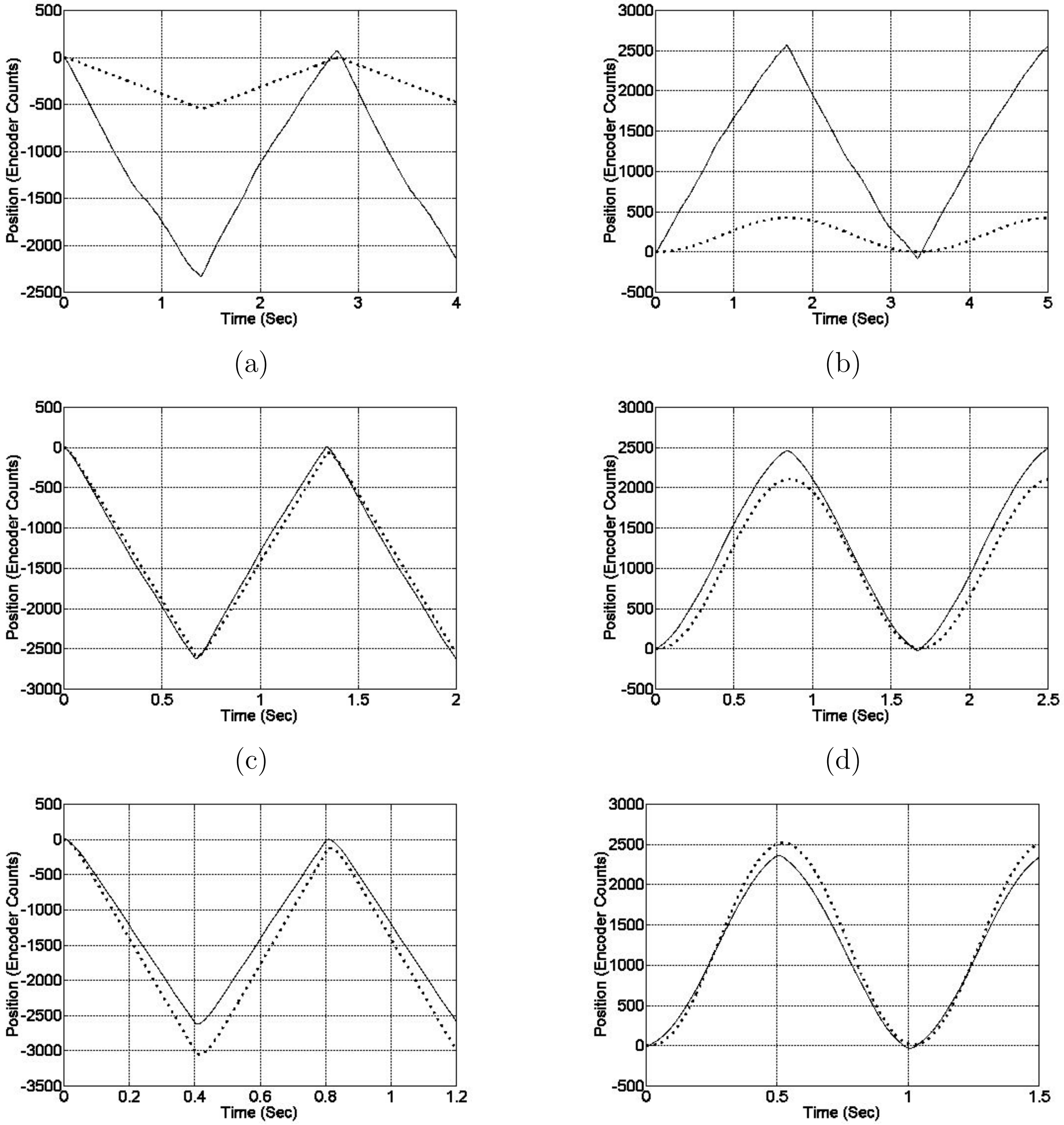
Applying MLS to the identification data gives $a = -0.95833$ and $b = 165.72$. Multiplying (4.3) by a discrete integrator sampling at 1000 Hz, the resulting linear model is shown in (4.4).

$$\frac{Y(z)}{V(z)} = \frac{0.16572}{z^2 - 1.95833z + 0.95833} \tag{4.4}$$

Initial Nonlinear Model

The model of (4.4) is combined with the dead zone and saturation nonlinearities. This forms the initial nonlinear model for simulation. To validate this model, different input signals are applied to the actual system and to the simulation model for comparison. The simulation model is shown in Figure 4.12, and the results are shown in Figure 4.13. Note that tests are done for varying amplitudes of control voltage. The position response has different frequencies, and as a result, the time scales are different.

The simulated open loop system closely matches for the 1 V amplitude square wave case. This is expected since the model was based on data from a 1 V step input. However, as the input signal amplitude is increased or decreased, or changes in shape, the model estimate diverges from the actual results.



(a) (b)

(c) (d)

(e) (f)

Figure 4.13: Experimental and Initial Nonlinear Model Estimate Responses to (a) 0.1 V, (c) 1 V and (e) 2 V Square Wave Inputs, and (b) 0.1 V, (d) 1 V and (f) 2 V Sine Wave Inputs. Solid lines present experimental and dotted lines estimated results

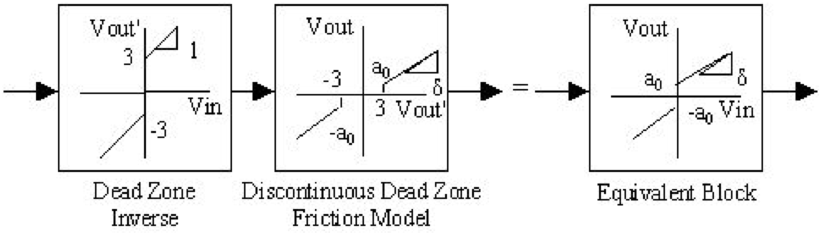


Figure 4.14: Tuned Friction Model and Linearization

Tuned Nonlinear Model

Two possible reasons for the discrepancy noted in Figure 4.13 are, that the assumption of friction as a dead zone is too simplistic and does not accurately reflect the system, and the open loop compensation of friction may not be effective in removing the nonlinearity. However, to maintain as simple a model as possible, the dead zone friction model with open loop compensation is kept, but slightly modified. The approximation of friction as a dead zone with a threshold of 3 V is not accurate. In fact, the simple dead zone model of friction would suggest that the actual voltage seen by the system should be $\delta(V_{in} - 3)$ for $V_{in} > 3$ (c.f. Figure 4.9). In the limit as V_{in} approaches 3 V from above, this would suggest no motion. However, it is known from Figure 4.8 that 3 V does generate good motion in the system. In some cases, motion can be seen at inputs below 3 V. Including a discontinuity in the dead zone at $V_{in} = 3$ V produces a more appropriate, though still relatively simple, friction model. Figure 4.14 shows this modified dead zone model along with the previous dead zone linearization block. The combined behaviour is represented by (4.5).

$$V_{out} = (a_0 + V_{in}\delta)sgn(V_{in}) \tag{4.5}$$

It should be noted that the discontinuous dead zone model does not accurately model the friction that exists in the system. In reality, motion can still exist at inputs below the 3 V dead zone threshold, but the motion is highly nonlinear. By combining the dead zone inverse with the discontinuous dead zone model, these nonlinearities are effectively avoided.

The identification of the threshold output voltage a_0 and slope δ must then be performed. They can be determined by examining the steady state speed at different inputs. Since the responses seen in Figure 4.13 appear symmetric, only the positive direction of motion is considered. Therefore, (4.5) becomes a linear function, and can be combined with (4.3) for use with the final value theorem. This results in (4.6) for $V_{ss} > 0$.

$$\dot{Y}_{ss} = V_{ss} \lim_{z \to 1} \frac{b}{z+a} \tag{4.6}$$

Where

$$V_{ss} = (a_0 + V_{inss}\delta)$$

Parameters a and b identified from MLS and two steady state results can then be applied to (4.6). The two steady state responses used are for the lowest and midpoint input voltages of 0.1 V and 1 V. Although the 2 V case was not used, it gives similar results. From Figure 4.13a, an approximate average of 2300 encoder counts for 1.35 seconds results in $\dot{Y}_{ss} \approx 1700$ for 0.1 V. From Figure 4.13c, an approximate average of 2585 encoder counts for 0.65 seconds results in $\dot{Y}_{ss} \approx 3977$ for 1 V. Applying the steady-state velocities to 4.6 gives:

@ Input = 0.1 V

$$\lim_{z \to 1} = \frac{(a_0 + 0.1\delta)b}{z+a} \approx 1700 \tag{4.7}$$

@ Input = 1 V

$$\lim_{z \to 1} = \frac{(a_0 + 1\delta)b}{z+a} \approx 3977 \tag{4.8}$$

Applying the previously derived values of a and b and solving results in $a_0 = 0.364$ and $\delta = 0.636$. The dead zone block of Figure 4.10 can now be replaced with the new discontinuous dead zone friction model as shown in Figure 4.15, resulting in the tuned nonlinear model of the system.

Tuned Nonlinear Model Validation

Applying test inputs to the experimental system and to the tuned simulation model of Figure 4.16, the results shown in Figure 4.17 are produced. As can be seen the results are much improved and more closely match actual experimental results. However, in the future, a more accurate friction model, as well as a more effective friction compensation

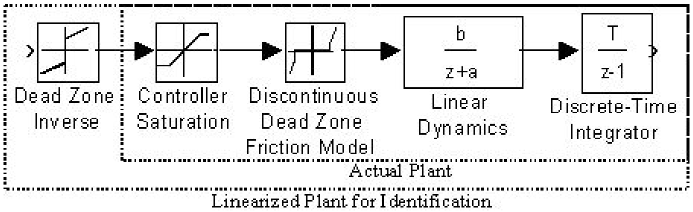


Figure 4.15: Block Diagram of the Open Loop System with Tuned Friction Model

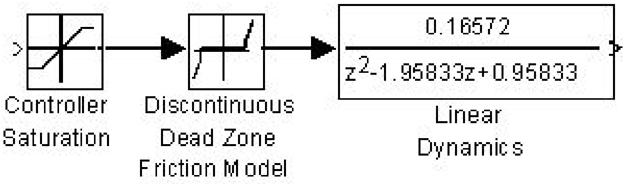


Figure 4.16: Tuned Nonlinear Simulation Model

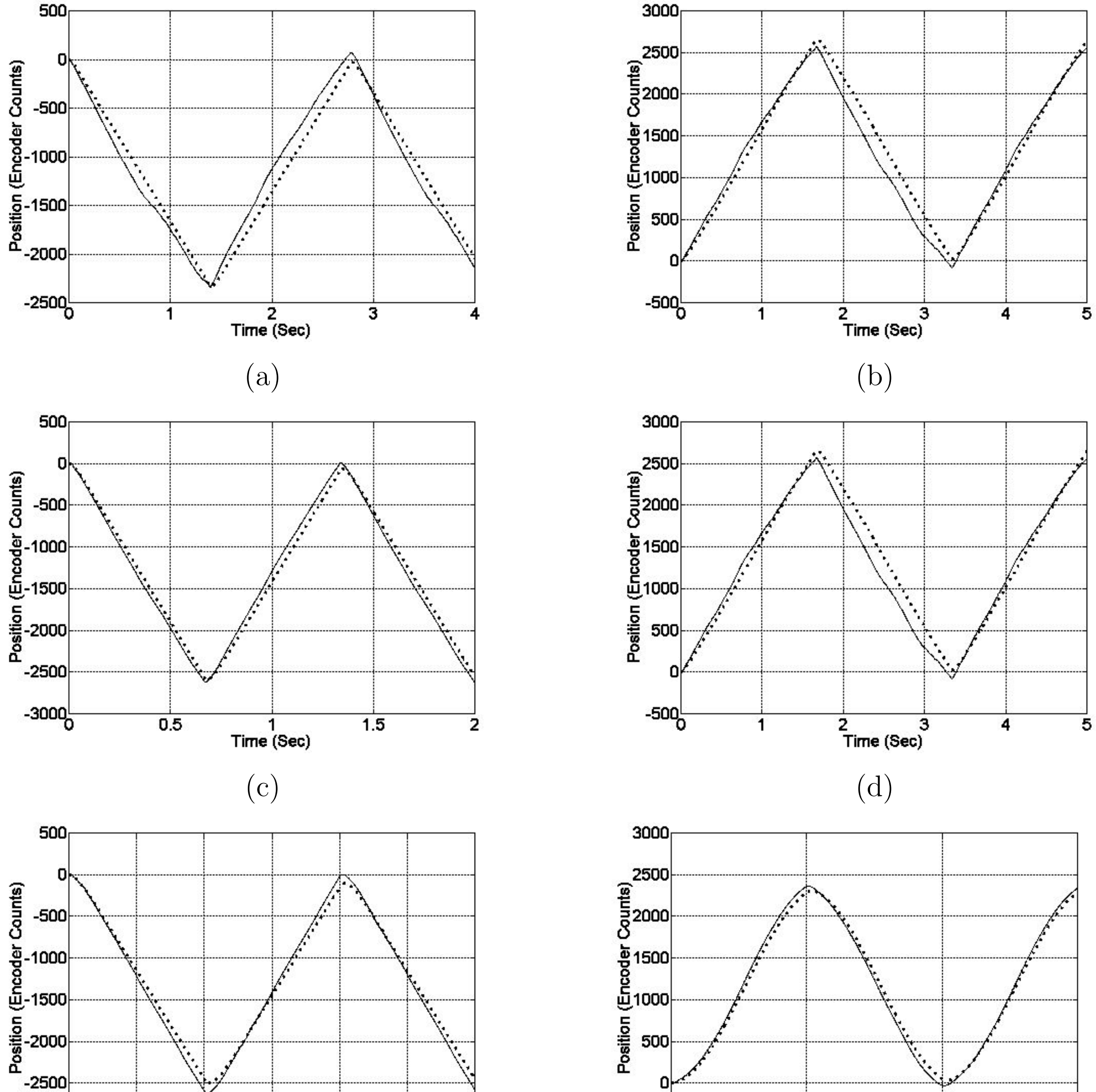
method, may be required to remove the remaining discrepancies between the model and experimental system.

4.3 Closed Loop Control

In this section, the value of having a good simulation model, which incorporates even a simple friction model, is illustrated. Three controllers are designed and compared: a "benchmark PID" controller tuned experimentally and not accounting for the nonlinearities present, an "initial nonlinear controller" designed using only the linear system model (4.4), and a "tuned nonlinear controller" designed based on (4.4) but tuned in simulation on the tuned nonlinear model of Figure 4.15.

4.3.1 Benchmark PID Controller Design

Using performance objectives of 700 millisecond risetime, zero steady state error, and zero overshoot yields a controller with proportional, integral, and derivative gains of 0.02, 0.007, and 0.0025 respectively. The result of this controller is shown in Figure 4.18 for a 2800



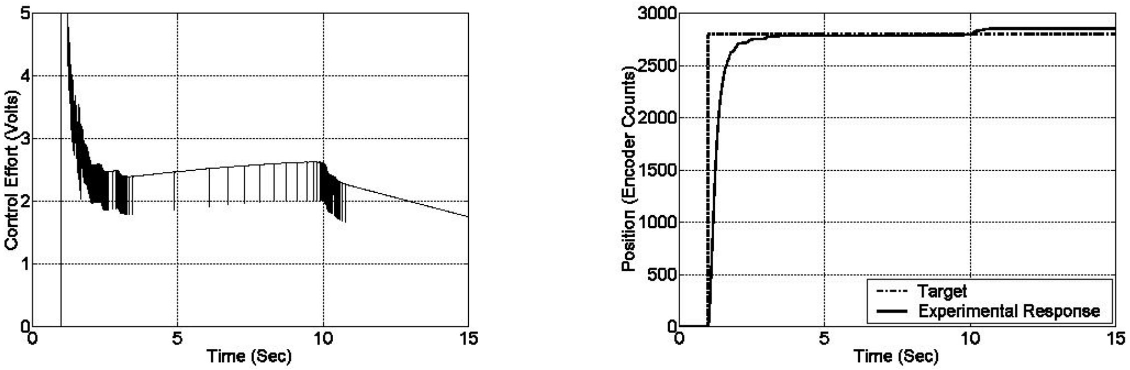
(a) (b)

(c) (d)

(e) (f)

Figure 4.17: Experimental and Tuned Nonlinear Model Estimate Responses to (a) 0.1 V, (c) 1 V and (e) 2 V Square Wave Inputs, and (b) 0.1 V, (d) 1 V and (f) 2 V Sine Wave Inputs. Solid lines present experimental and dotted lines estimated results

encoder count step input signal, corresponding approximately to full macro-stage travel.



(a) (b)

Figure 4.18: Macro-Stage (a) Controller Output and (b) System Response using Benchmark PID Controller

As seen in Figure 4.18, the rise time is quick. However, the main difficulty lies in steady state error reduction. When the system nears the target, the control signal can fall below the dead zone threshold, preventing further motion. This is evident by the extremely slow response from 4 to 10 seconds where the controller output is ramping up below 3 V. To overcome this with a linear controller, an integrator is used. However, if a minor amount of overshoot occurs, the controller must then ramp the control signal from positive to negative three volts before motion in the opposite direction occurs, resulting in a very slow response. This is evident in Figure 4.18 where, after 10 seconds, overshoot occurs and the effort ramps down. Yet the system remains motionless. The only method of speeding the integrator is to increase the integrator gain. However, this also results in greater overshoot and undesirable integrator windup. This repeated cycling results in a limit cycle characteristic of systems with a dead zone nonlinearity, and a controller behaviour known as “hunting”. Thus, incorporating friction nonlinearities in the controller design is critical.

4.3.2 Initial Nonlinear Controller Design

Using the identified linear system of (4.4), a simple proportional-derivative controller is designed. The controller is obtained by assigning a desired damping ratio of 1 to eliminate

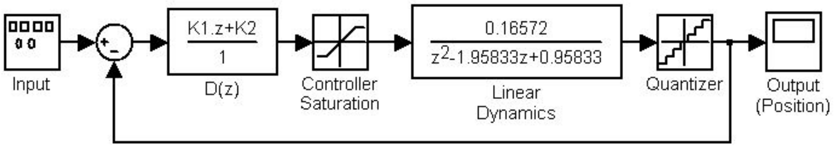
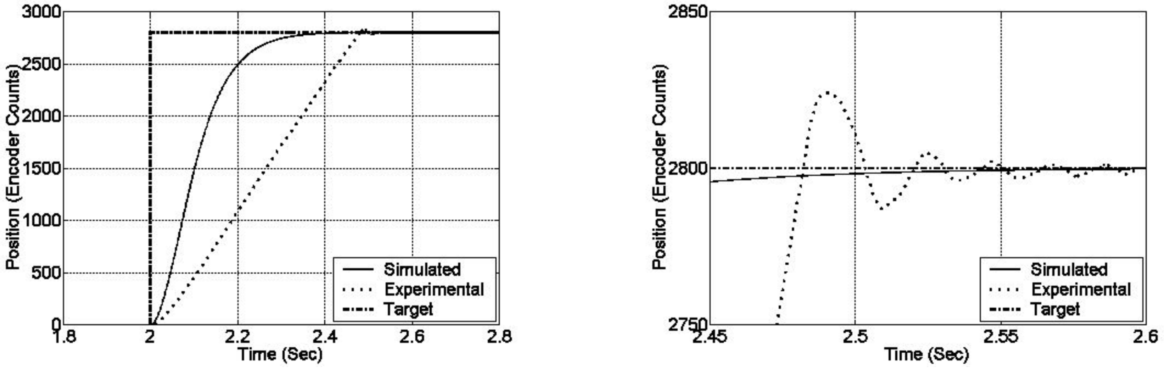


Figure 4.19: Macro-Stage Linear Closed Loop Simulation Model

overshoot, and thus determine a desired phase margin for PD design for a linear system [34]. The resulting linear controller is given in (4.9). With the addition of the dead zone inverse block, the initial nonlinear controller is formed. Coupling with the initial nonlinear simulation model of Figure 4.12 results in the linearized simulation model of Figure 4.19. Note that the dead zone friction is completely linearized by its inverse, and therefore is not shown. Testing this controller in simulation and on the experimental system yields the response shown in Figure 4.20.

$$D(z) = \frac{0.00245865z + 0.0000998447}{z} \tag{4.9}$$



(a) (b)

Figure 4.20: (a) Full and (b) Zoomed View of Simulated and Experimental Response using Initial Nonlinear Controller and Initial Nonlinear Model

As can be seen, the initial nonlinear controller does not perform very well for the experimental system and requires tuning. The inconsistencies between the simulated and

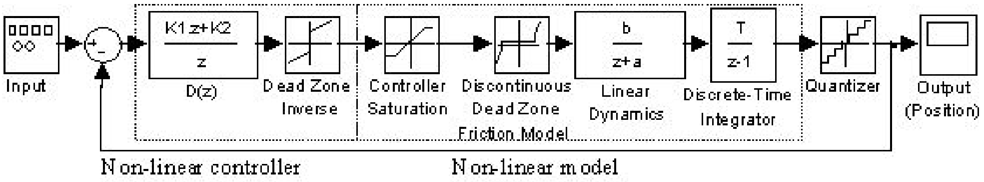


Figure 4.21: Macro-Stage Tuned Nonlinear Closed Loop Model

actual responses are due to nonlinearities that exist. The difference is due to an incorrect friction model previously discussed in open loop identification.

4.3.3 Tuned Nonlinear Controller Design

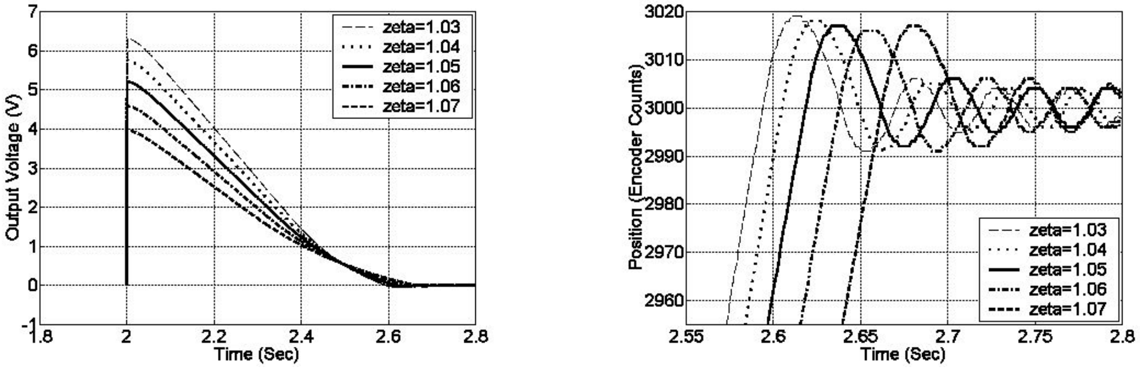
Revisiting the controller design stage, a nonlinear closed looped model is first formed using the tuned nonlinear model previously developed. The resulting model is shown in Figure 4.21.

The linear model (4.3) is still used as the basis for initial controller design, but tuning is then performed in simulation on the full nonlinear model of Figure 4.21.

Tuning to Avoid Controller Saturation

Controller saturation is avoided by tuning the damping ratio to slow the response until the controller output of the simulated system is below 5 V for a step input of 3000 counts, the maximum travel of the system. The effect of changing the damping ratio and choosing the lowest phase margin frequency that satisfies PD design criterion [34] is shown in Figure 4.22. The damping ratio is chosen to be 1.05, producing a maximum controller output near the saturation level of 5 V.

From Figure 4.22, the nonlinear response is quite good with a relatively small amount of overshoot. However, overshoot is not desired. The dead zone introduced by friction is linearized by applying its inverse that was introduced in Figure 4.9; this block ensures output signals have a magnitude greater than 3 V. Thus, if overshoot occurs, the controller output changes sign resulting in a discontinuous change in output signal. This discontinuous change results in motor chatter and undesirable oscillation.



(a) (b)

Figure 4.22: Plot of Nonlinear PD Closed Loop (a) Controller Output and (b) Zoom of Response for 3000 Step Input

Tuning for Overshoot Reduction

To further tune the controller, the phase margin frequency is varied to reduce overshoot. As noted in [34], generally lower phase margin frequencies result in a lower bandwidth and thus rejection of higher frequency oscillation. The frequency is tuned by reducing it until the response of the nonlinear model displays little or no overshoot. This is shown in Figure 4.23. The phase margin frequency is chosen to be 4.91 rad/sec.

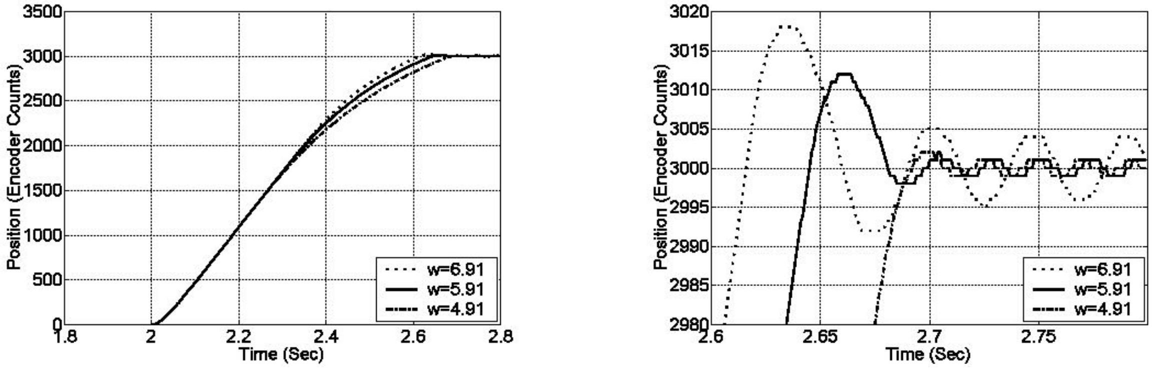
Controller for Tuned Nonlinear Model

The resulting linear controller with the desired damping ratio and phase margin frequency is given in (4.10).

$$D(z) = \frac{0.0124593z + 0.0112178}{z} \tag{4.10}$$

4.3.4 Comparison of Closed Loop Responses

For the final test of the controller and model, the linear controller (4.10) is coupled with the dead zone inverse. The resulting tuned non-linear controller is tested in simulation on



(a) (b)

Figure 4.23: (a) Full and (b) Zoomed Plot of Nonlinear PD Closed Loop Response With Varying Phase Margin Frequency

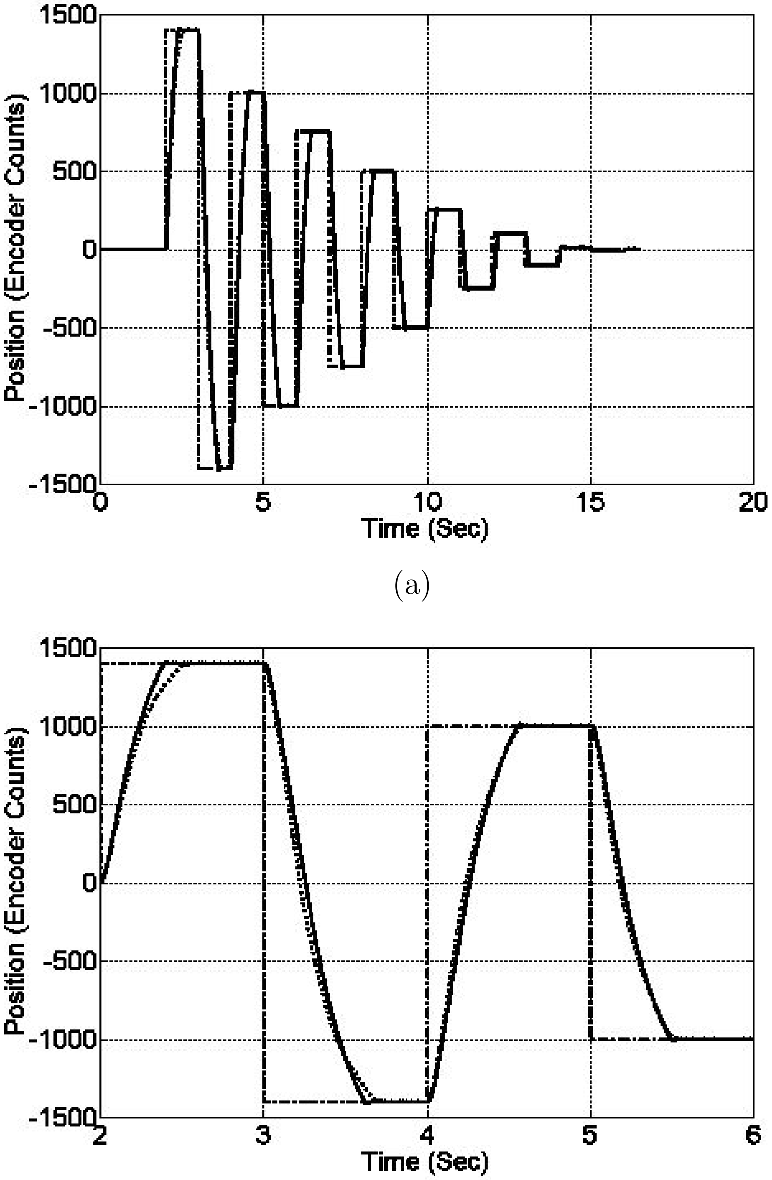
the tuned nonlinear model, and experimentally. The results for a decaying square wave input are compared in Figure 4.24.

Aside from some slight mismatch at high displacements, the experimental and modelled responses match extremely well, indicating an accurate model. Therefore, the developed tuned nonlinear model provides an excellent basis to design and tune a controller offline.

Figure 4.25 compares the performance of each controller on the experimental system using a decaying square wave input. It can be seen that the benchmark PID controller performs poorly with much difficulty in tracking the target. The initial nonlinear controller tracks quite well. However, when overshoot occurs significant chatter results, as illustrated in the zoomed view of Figure 4.25. The tuned nonlinear controller performs the best with good tracking performance and no oscillations.

4.4 Review of Design

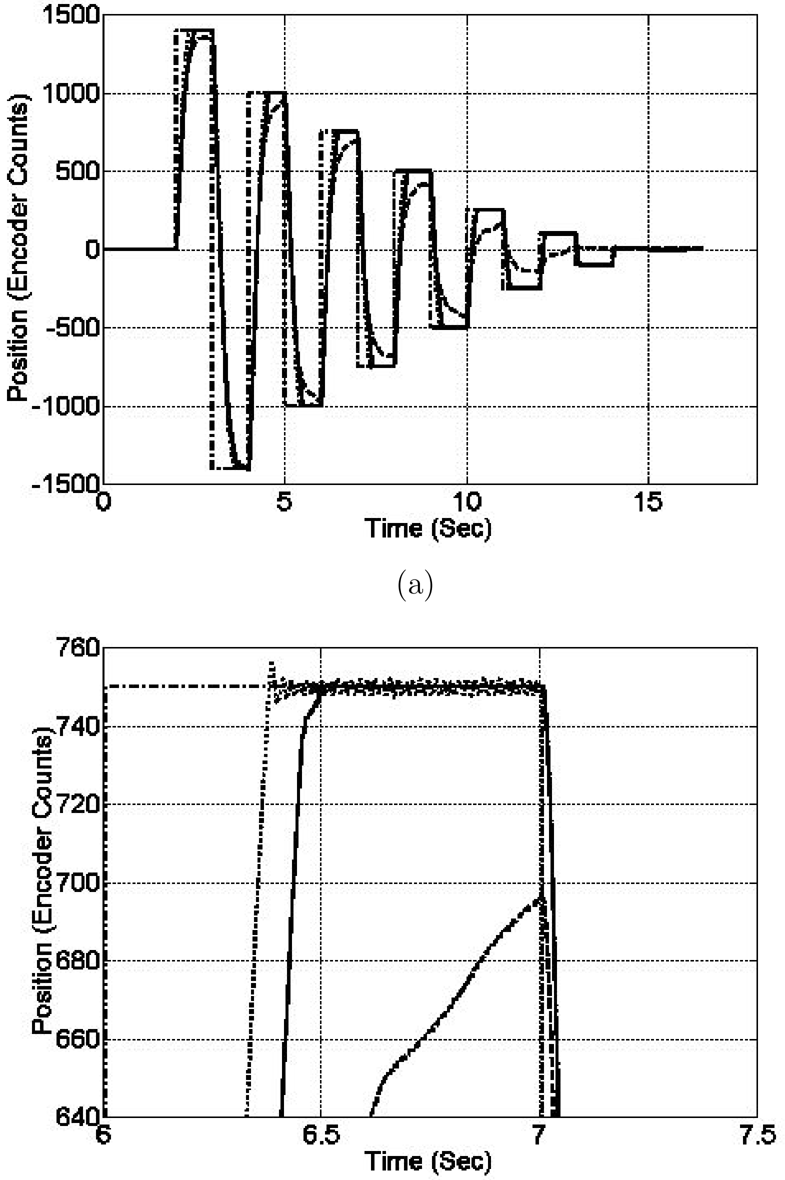
The theoretical equations of a DC motor were used to form a basis for open loop MLS identification of a macro-positioning system. The system is highly nonlinear due to friction, which was approximated by introducing a dead zone applied to the control signal. Linearization of this dead zone was performed using a dead zone inverse to modify the control



(a)

(b)

Figure 4.24: (a) Full and (b) Zoomed Comparison of Experimental and Simulated Closed Loop Responses with Tuned Nonlinear Controller. Legend: Target (Dash-dot), Simulated (Solid), Experimental (Dotted) Response



(a)

(b)

Figure 4.25: (a) Full and (b) Zoomed Comparison of Experimental Closed Loop Responses with Each Controller. Legend: Target (Dash-dot), Tuned Nonlinear Controller (Solid), Initial Nonlinear Controller (Dotted), Benchmark PID Controller (Dashed)

signal to the motor. This resulted in a system, assumed linear, to which MLS identification could be applied. MLS identification yielded a linear model that was found to match poorly the experimental system. However, by slightly modifying the friction model, introducing a discontinuity in the dead zone, the resulting model produced excellent matching with experimental results.

To illustrate the importance of developing the nonlinear model and controller, an experimentally tuned linear PID controller, as well as a PD controller that was developed based on the linear model were applied. The resulting closed looped systems performed satisfactorily. To avoid tuning the controller online, the nonlinear model developed was used to tune the damping ratio and phase margin frequency of the closed loop system. Applying the resulting tuned PD controller along with a simple nonlinear friction compensation scheme, the simulated and actual results matched extremely well and the controller provided good performance. This demonstrates that the developed tuned nonlinear model is quite accurate and provides a solid basis for the macro stage of the macro-micro positioning system. The micro-stage of the system is analyzed next.

Chapter 5

Micro-Stage Modelling and Control

The micro-stage consists of a NanoMuscle^{TM} SMA actuator in series with a linear flexure guide and a bias spring. Feedback is provided via a Hall effect sensor and magnets. In this chapter the micro-stage is mathematically modelled and a control strategy is designed.

5.1 Modelling and Control Overview

The dynamics of the SMA actuator includes linear and nonlinear properties. As described previously, the current to the SMA wires in the NanoMuscle^{TM} is not directly controlled. An input signal, $CTRL$, controls the current through unknown driver circuitry. A model of the operation can be represented through the block diagram of Figure 5.1.

In an attempt to more easily model the actuator, an initial assumption is made that hysteresis is the sole nonlinearity in the actuator. Since the dynamics of the driver circuitry are unknown, as an initial approach they are assumed linear and combined with the thermal dynamics as shown in Figure 5.2 to form the "Electro-Thermal Model".

5.1.1 Methodology

The electro-thermal model is formed by applying MLS to experimental data, in order to correlate the input $CTRL$ voltage to the output temperature of the actuator. However, the output temperature cannot be measured directly during operation. Therefore, a method

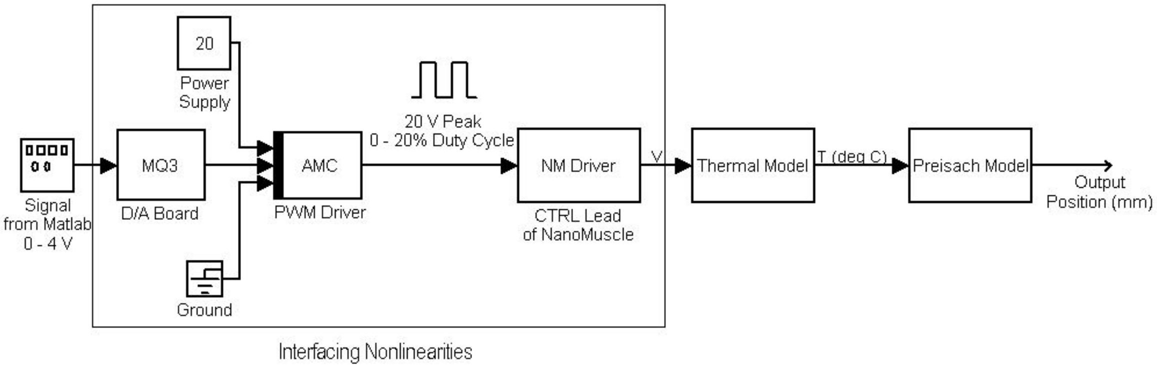


Figure 5.1: Block Diagram of Actuator Model

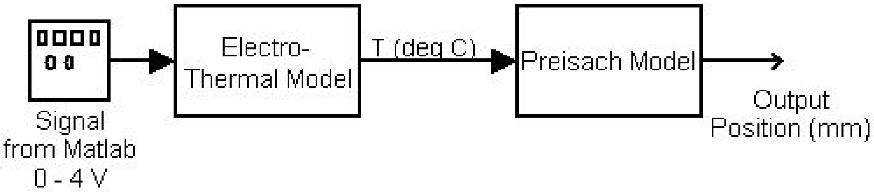


Figure 5.2: Block Diagram of Simplified Actuator Model

of inferring the actuator temperature from output contraction length is used. The adopted temperature inference method uses experimental data obtained by heating the actuator in a kiln, thereby controlling temperature directly, and measuring the output contraction length. If it is assumed that minor loops are not travelled for step $CTRL$ inputs to the actuator when initially fully extended, then the experimental data collected can be used to calculate the actuator temperature from contraction length.

The hysteresis is modelled using the experimental implementation of the Preisach model. Using the same experimental data used for inferring the actuator temperature, the resulting model provides a hysteretic conversion of input temperature to output contraction length.

Finally, an experimentally tuned PD controller with anti-windup is designed and implemented to give experimental closed loop results.

5.2 System Modelling

The first analysis performed is to generate a model representing the inherent hysteresis of the actuator. The Preisach model is adopted and applied using data gathered through open loop heating of the actuator. The model is validated through comparisons of simulated and actual results from experimental tests.

An electro-thermal model for the actuator is then derived through the use of MLS on the same experimental data used in the Preisach model.

5.2.1 Hysteresis Model

One major aspect of SMAs that makes their use in position control difficult is hysteresis. This nonlinearity plays an important role in the simulation and analysis of its behaviour for controller design. It is ideal to have a mathematical model of this nonlinear behaviour. Particularly in position control, it is important that the model reproduce minor loop behaviour. The Preisach hysteresis model described in Section 2.4.1 is one such model, which has been used in the past by Gorbet *et al.* [21], and Hughes and Wen [35] for SMA hysteresis and is adopted here. The implementation of the Preisach model is restated as follows:

$$f(t) = -f^{+} + \sum_{k=1}^{n(t)-1} (f_{M_k,m_k} - f_{M_k,m_{k-1}}) + f_{M_n,m_n} - f_{M_n,m_{n-1}} \tag{5.1}$$

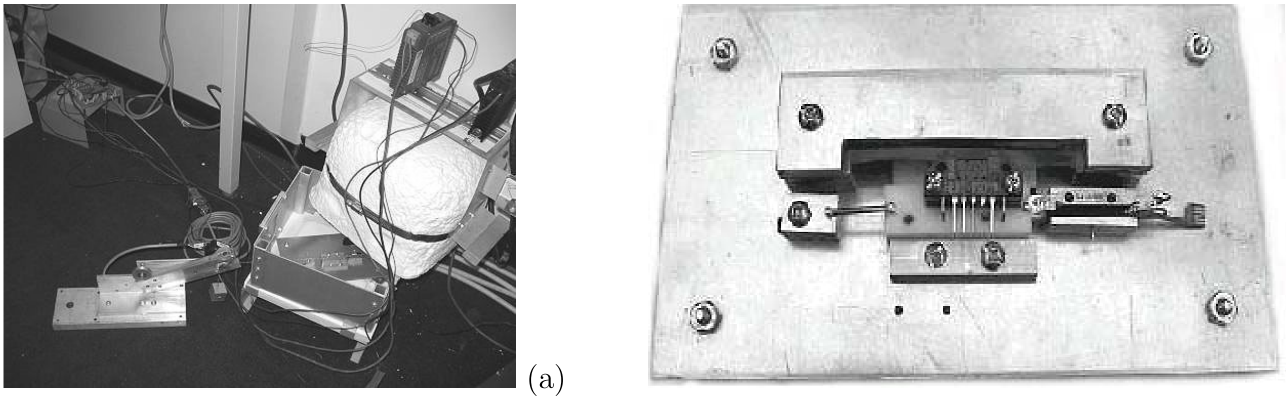
In order to implement (5.1) the values of f_{M_k,m_k} must be found through identification. This process involves the measurement of data on "first-order descending" (FOD) curves which are descending branches of the hysteresis curve. They are generated by starting the actuator temperature below the point of negative saturation, heating to a temperature M_k below positive saturation, and then cooling the actuator while measuring the output at discrete points m_k over the temperature range. The measured outputs correspond to specific values of f_{M_k,m_k}.

NanoMuscle^{TM} Preisach Model Identification

To identify the data required for the NanoMuscle^{TM} Preisach model, a series of FOD curves were generated to develop a look up table of the output at specific maximum and minimum input values. The curves were generated using a programmable kiln to heat the actuator, thus providing direct control of the temperature. The setup of the kiln is shown in Figure 5.3. Current controlled by a solid-state relay generates heating through a resistive heating element in the kiln. The temperature of the NanoMuscle is measured using a thermocouple fixed to the top plate of the actuator. The actuator is mass-biased with a 30 gram weight, and displacement is measured using an optical encoder with a linear resolution of approximately 14 μm. The temperature is monitored, and the relay contactor of the kiln controlled, by a micro-controller via a LabView^{TM} interface on a PC.

Using the kiln test bed, the NanoMuscle was repeatedly heated to different maximum temperatures M_k, then cooled back down to ambient temperature, generating FOD curves. With each heating cycle, M_k was decreased. As the actuator cooled, measurements were taken at decreasing values of temperature, m_k, and assigned to FOD data point f_{M_k,m_k}.

Temperature Profile An example temperature profile that was measured from the actuator is shown in Figure 5.5a. This profile was obtained by the thermocouple attached to the actuator. Figure 5.5a also illustrates a problem that existed in the setup during



(a) (b)

Figure 5.3: Experimental Setup using a Kiln (a) Entire Setup (b) NanoMuscle^{TM} Fixture

experimental testing. It can be seen that seemingly erratic transient temperature responses are measured at specific temperatures. The source of this transient behaviour is electrical interference from the thermocouple. However, the experimental setup was not rectified until after the data presented here was gathered. To smooth the transient spikes, a function was developed to search for areas where the slope is larger than expected. A window is placed around these points and the transient is replaced with a linear curve. A flowchart of the smoothing algorithm is given in Figure 5.4.

If the algorithm is applied to the example temperature profile of Figure 5.5a, the result is shown in Figure 5.5b.

Averaging of Data With a smoothed temperature profile, the measured hysteresis loop of the example data can be plotted as shown as the solid line in Figure 5.6. The data contains noise, therefore an averaging algorithm is used to remove the noise and to also reduce the data file size. The algorithm consists of a moving window that calculates the mean of all the points within the window. If this is applied to the solid lined hysteresis curve of Figure 5.6, the resultant averaged data points are illustrated as diamonds.

Resulting Data The smoothing and averaging functions can be applied to all the experimental data gathered to form the FOD curves. A plot of the resultant FOD curves

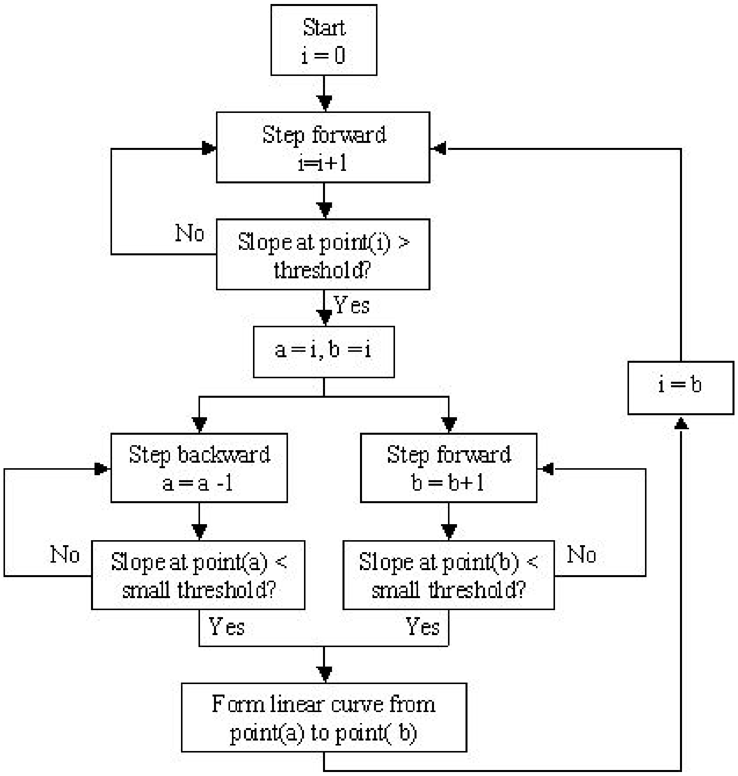
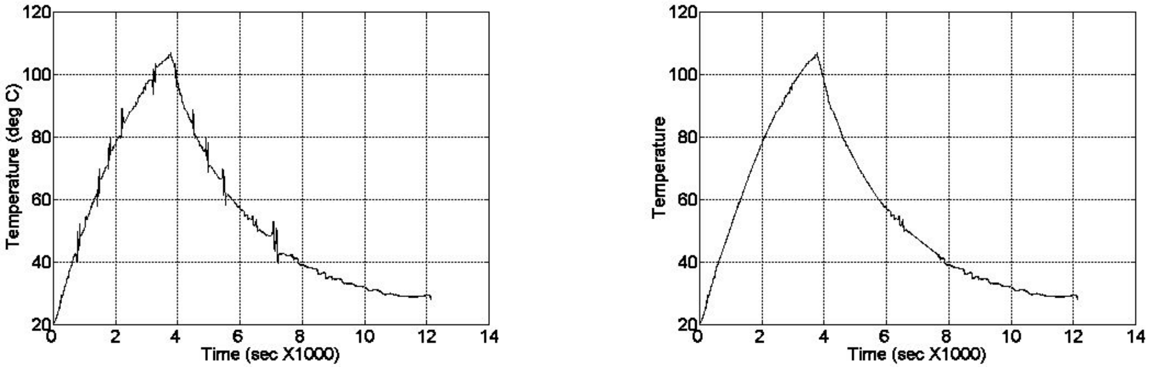


Figure 5.4: Flowchart of Smoothing Function



(a) (b)

Figure 5.5: Example (a) Experimental and (b) Smoothed Temperature Profile

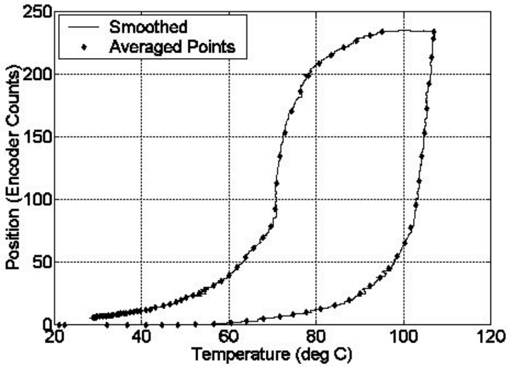


Figure 5.6: Hysteresis of Smoothed Temperature Example Data

is given in Figure 5.7. M_k, m_k, and contraction length can be used to form the axes of a FOD lookup table as illustrated in Figure 5.8.

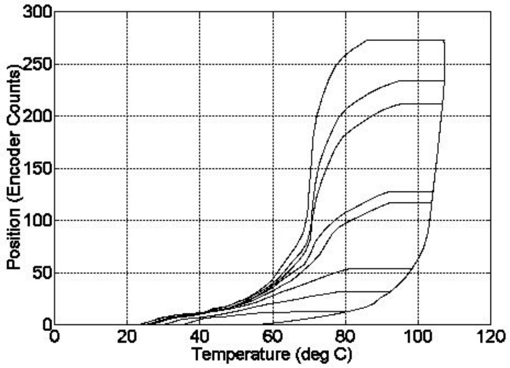


Figure 5.7: FOD Data used for Preisach Model

Simulation Results

To test the validity of the model, (5.1) was implemented in Matlab along with the FOD data of Figure 5.8 to create a simulation of the system. A new series of temperature profiles from data obtained from the kiln are entered to the simulator and compared with actual results. The comparison of the resulting curves with the actual data is shown in Figure 5.9. The match is quite good.

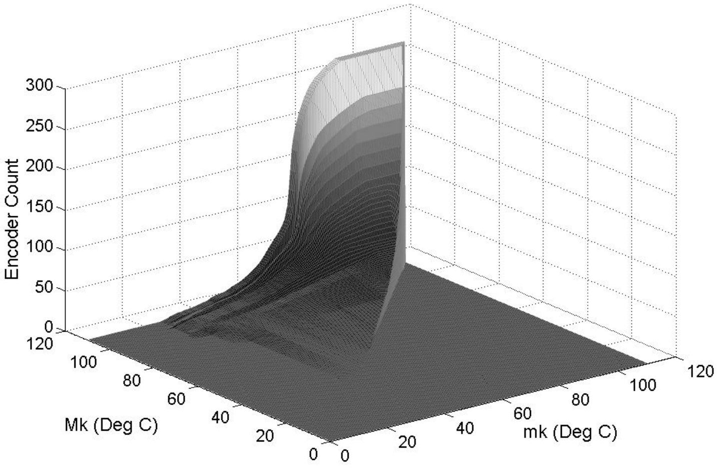


Figure 5.8: 3D Plot of FOD Data used for Preisach Model

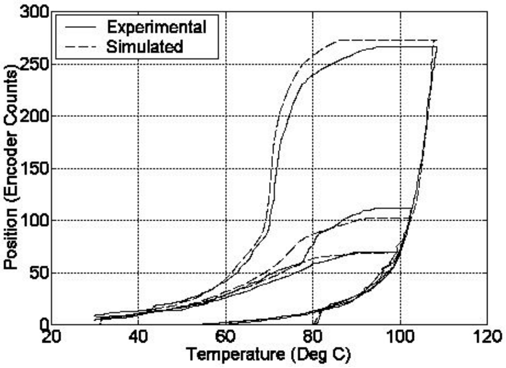


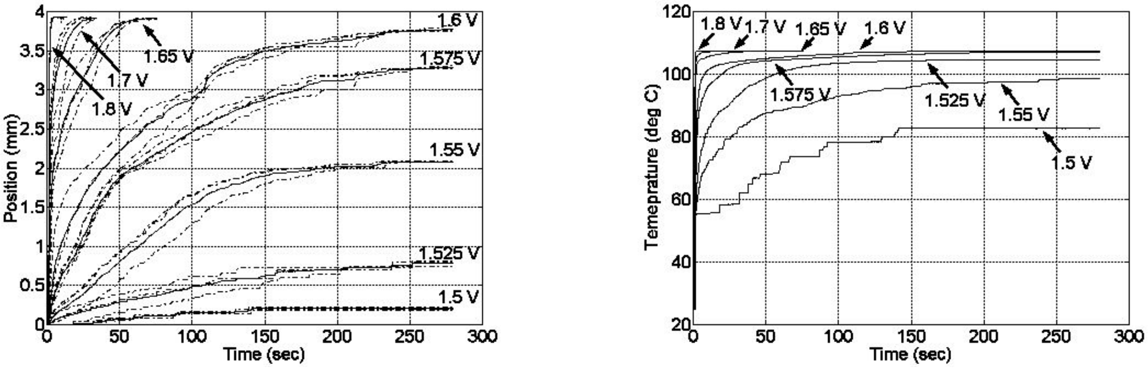
Figure 5.9: Comparison of Preisach Model Simulated and Actual Results

5.2.2 Electro-Thermal Model

For the identification of the Preisach model, the actuator was directly heated in the kiln. However, the full model of the actuator will have an input of voltage and an output of contraction length. Thus, an electro-thermal model is required to convert input voltage to output temperature, so that the Preisach model can be applied to convert temperature to contraction length. Here, two separate electro-thermal models for heating and cooling are generated.

Electro-Thermal Heating Model

To find a heating model, the micro-stage described in Section 3.2 is operated in open loop. Various step input voltages are applied to the actuator and the output position is tracked using the laser based measurement device described in Section 3.2.3. Figure 5.10a illustrates the output for various input signals. The broken lines for each input signal represent the output for three different trials, and the solid line the mean of the three trials. It must be noted that the actuator has a mechanical stop at a contraction of approximately 3.98 mm, which is evident when the input signal exceeds 1.65 V in Figure 5.10a. Therefore, only responses that do not reach saturation can be analyzed by MLS to find a linear model.



(a) (b)

Figure 5.10: Micro-Stage (a) Position and (b) Temperature Response for Varying *CTRL* Step Input Voltages

At this stage, it is assumed that the actuator temperature is below M_f when the actuator is fully extended and the input is 0 V. With this assumption, no minor loops in the hysteresis curve are followed when the actuator begins and ends at full extension, and the input applied to the $CTRL$ lead of the NanoMuscle^{TM} steps from 0 V to >0 V and back down to 0 V. In this case, the actuator temperature will follow an FOD curve during operation. Thus, the experimental FOD data of Figure 5.8, originally gathered for the Preisach model, can also be used to infer the actuator temperature from contraction length. Note that this inferencing can only be done for cases of monotonically increasing and decreasing input signals where the actuator is initially fully extended. Using this inferencing technique on the experimental data of Figure 5.10a results in the temperature responses shown in Figure 5.10b.

If the steady state temperatures are examined by plotting them with respect to input $CTRL$ as shown in Figure 5.11, it can be seen that the electro-thermal model is not linear. The steady-state relationship in fact appears to be exponential. Therefore, the initial assumption of a purely linear electro-thermal model is incorrect. The source for the exponential relationship likely lies within the unknown driver circuitry. In the future, this circuitry may need to be identified for better modelling of the actuator. (It should be noted that the ambient temperature is assumed to be 25^0C and therefore the values are shifted by -25^0C to set the initial conditions to zero).

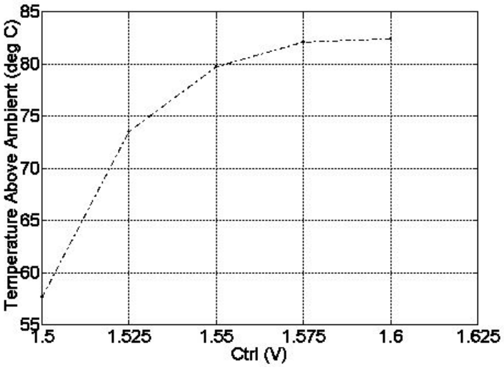


Figure 5.11: Steady-State Relationship of $CTRL$ Input Voltage to Output Temperature

Nonlinear Electro-Thermal Heating Model

The relationship between input $CTRL$ to output temperature appears to be nonlinear. This can be expected since the electrical circuitry within the NanoMuscle^{TM} is unknown, and was only assumed to be linear. A nonlinear model appears to be required but it is uncertain what type of nonlinearity the actuator displays. The nonlinearities could be dynamic, the transfer function changing due to changes in any number of different system parameters, or they could be as simple as a nonlinear gain coupled with linear dynamics. To simplify the analysis, it is still assumed that the dynamics of the NanoMuscle^{TM} driver circuitry are linear and can therefore be combined with the thermal dynamics to form an "electro-thermal model" that can be identified with MLS. However, a nonlinear gain is assumed to produce the observed nonlinear steady-state behaviour. The proposed model now contains a nonlinear gain function, ϕ, along with a linear transfer function, Θ, representing the dynamics of both the driver circuitry and thermal dynamics. A block diagram is shown in Figure 5.12. The output of nonlinear gain is T_d, the desired steady-state temperature. The nonlinear gain can be characterized through the steady-state relationship between input $CTRL$ and steady-state actuator temperature.

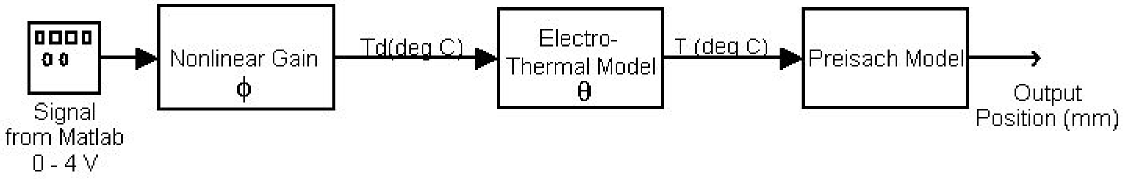


Figure 5.12: Block Diagram of Simplified Actuator Model with Nonlinear Gain

Revisiting Figure 5.11, the general form of the curve appears to match the exponential curve of the form given in (5.2) where x_0 and y_0 represent the x and y offsets from the origin. If the input $CTRL$ signal is attributed to x and the output temperature to y, by analyzing Figure 5.11 x_0 and y_0 can be assigned the values of 1.5 V and 57.66^0C respectively.

$$y = B(1 - e^{-A(x-x_0)}) + y_0 \tag{5.2}$$

The optimal method of discovering the parameters would be to use a fitting algorithm such as least squares to fit experimental data to the exponential function. However, this is difficult due to the exponential form of the proposed function. A nonlinear least squares search algorithm would be required since the function is nonlinear, and the search algorithm may not provide a result that is the global minimum. Therefore, a less formal method using the experimental data is adopted to assign parameter values to the nonlinear gain.

Parameter B can be found by analyzing the curve as x tends to infinity as in (5.3). From Figure 5.11 an asymptote appears to exists for an output temperature of approximately 83^0C. Substituting $y_\infty = 83$ into (5.3) results in an approximate value of $B = 25.5$.

$$y_\infty = \lim_{x \to \infty} B(1 - e^{-A(x-x_0)}) + y_0 = B + y_0 \tag{5.3}$$

At this point the model can be linearized for MLS identification of A by rearraging (5.2) as follows:

$$\begin{aligned} A(x - x_0) &= ln\left[1 - \frac{y - y_0}{B}\right] \\ Ax' &= y' \end{aligned} \tag{5.4}$$

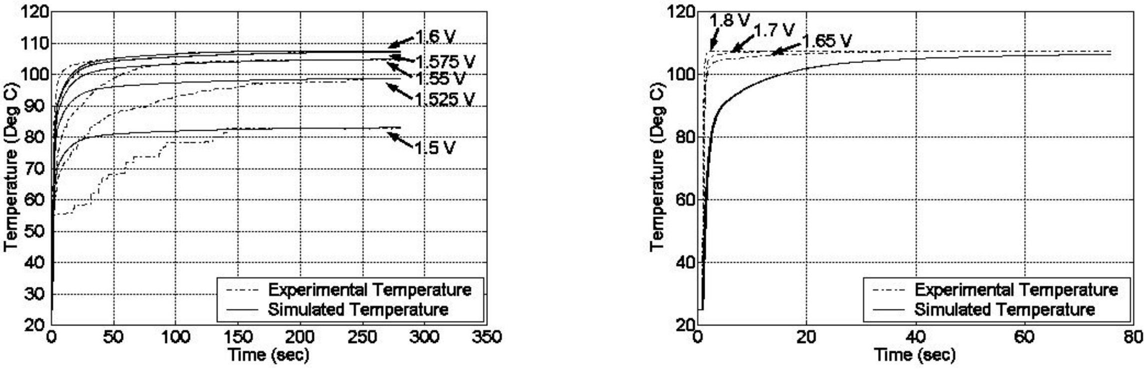
The transformed input and output values, x' and y', can be calculated with the previously found parameter values for x_0, y_0, and B, and experimental input and output values x and y. By applying MLS to the experimental data it is found that $A = 37.86$. The resulting steady state relationship is thus best represented by (5.5). The relationship is treated as a nonlinear gain on the input signal, converting input $CTRL$ voltage to desired steady-state temperature, T_d.

$$T_d = \phi(CTRL) = 25.5(1 - e^{-37.86(CTRL-1.5)}) + 57.66 \tag{5.5}$$

What remains to be analyzed is a model for the electro-thermal heating dynamics. To generate a dynamic model, MLS is applied to the median step response for which saturation does not occur. The output for an input of $CTRL = 1.575$ V is chosen. Also, it is found that the assumption of a 3rd order model gives the best result. The resulting transfer function is given in (5.6).

$$\frac{T}{T_d} = \Theta = \frac{1.001s^2 + 0.1483s + 0.001552}{s^3 + 1.491s^2 + 0.1604s + 0.001537} \tag{5.6}$$

This model is validated by comparing its simulated responses with the various other step responses of Figure 5.10b. The results are shown in Figure 5.13.



(a) (b)

Figure 5.13: Heating Model Validation Temperature Response for (a) 1.5-1.6 V and (b) 1.65-1.8 V Step Input

It can be seen that the heating model unfortunately does not match the responses for input signals of $CTRL > 1.6$ V. In fact if only the responses where saturation occurs are considered, the response time appears to exponentially decrease, which is evident in Figure 5.13b.

Unfortunately, the linear steady state temperature output cannot be discovered for the cases in Figure 5.13b because of saturation. Therefore, modification of the heating model can only be done through experimental tuning. Although a varying dynamic model would appear to be useful since the response characteristics change with different input values, a varying dynamic model is difficult to implement in practice. Therefore, the derived dynamic model is retained without modification, and only the nonlinear gain is modified. It is noted that if the steady state temperatures for input values $CTRL > 1.6$ V are increased, the resulting time constant would appear to decrease. That is, the time required to reach the saturation temperature of 107.5^0C shortens. It is found that the addition of another exponential curve to the nonlinear gain previously developed produces satisfactory results.

To join the two curves, a weighting function is employed to provide a smooth transition between the two. The weighting function, γ, used by Gorbet [21], is an exponential of the form given in (5.7). The final nonlinear gain is limited to a minimum of 0 and is given in (5.9) and plotted in Figure 5.14.

$$\gamma = \frac{1}{1 + e^{-40(-u+1.7)}} \tag{5.7}$$

$$\begin{aligned} T_d = \phi(CTRL) &= \gamma\ max([0\ 25.5(1 - e^{-37.8631(CTRL-1.5)}) + 57.66]) \\ &\quad +(1-\gamma)(10e^{10(CTRL-1.625)} + 82.581682) \end{aligned} \tag{5.8}$$

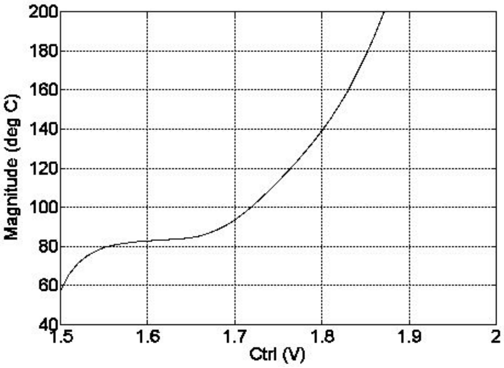
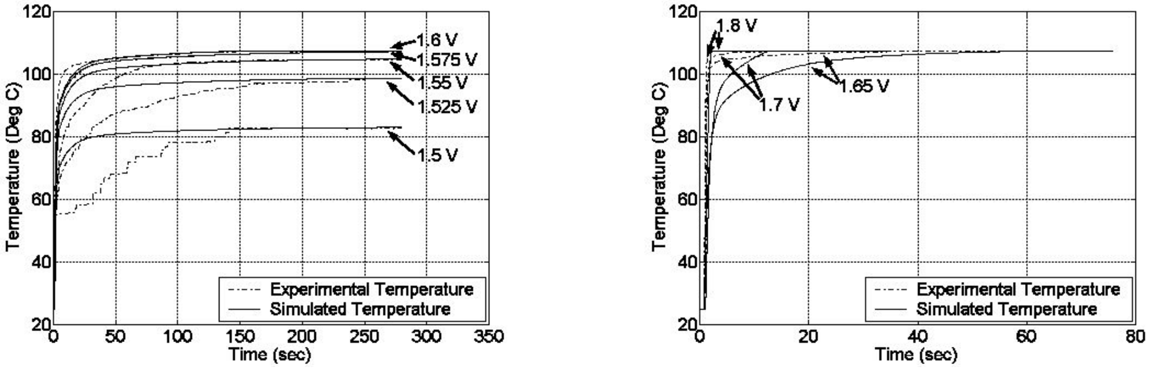


Figure 5.14: Weighted Steady-State Relationship of $CTRL$ Input Voltage to Output Temperature

The gain function ϕ of Figure 5.12 is replaced with the weighted nonlinear gain, (5.9), to be combined with the dynamic model of (5.6). This gives the final heating model to be combined with the Preisach model. The output of the final model is compared with the experimental data in Figure 5.15. It should be noted that the simulated results match the experimental at steady state, but only provides qualitative matching dynamically.



(a) (b)

Figure 5.15: Weighted Heating Model Validation Temperature Response for (a) 1.5-1.6 V and (b) 1.65-1.8 V Step Input

Cooling Model

Aside from a heating model, a cooling model is also formed. A separate model for cooling is used, as a first order model better matches the cooling response. This is evident in Figure 5.16. It must be noted that the cooling response can change over time with repeated cycling of the actuator as was noted in [36]. This paper can be found in Appendix D. The time constant increases as the base temperature of actuator rises from room temperature, but an approximation is made that a linear model is acceptable if the assumed time constant is some median between the slowest and fastest cooling times.

The experimental data used is for an actuator that has been cycled a few times to allow the actuator to warm-up. Experimental data from a step input of 1.575 V to 0 V is used as the basis of the model. From Figure 5.16, for a 63% risetime, the time constant is estimated to be approximately 1 second, thus forming the model given in (5.9).

$$\frac{1}{z+1} \tag{5.9}$$

With an electro-thermal model of the system, a full model for the SMA actuator can be formed.

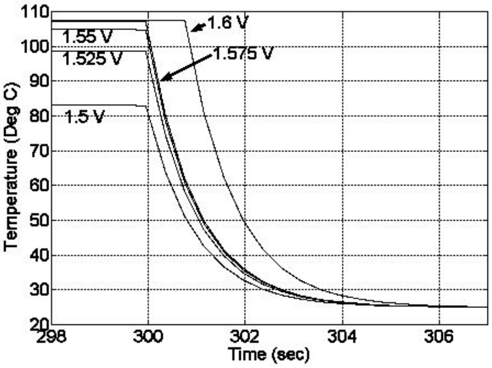
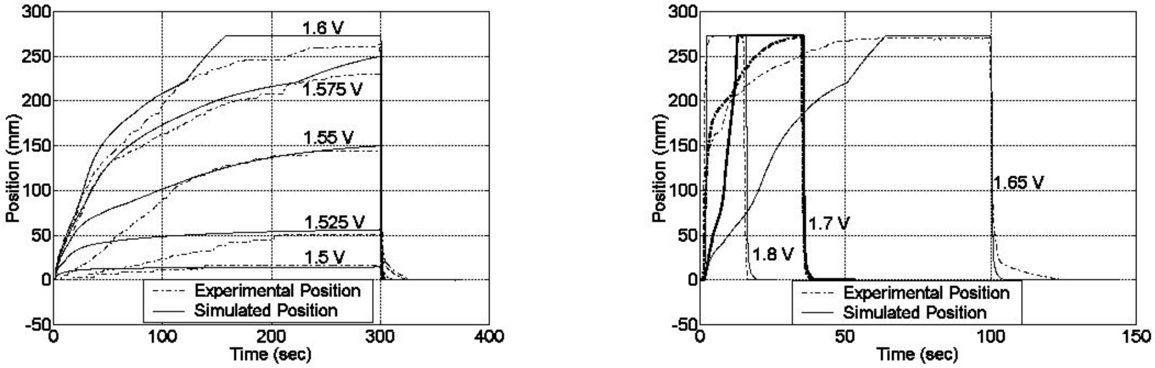


Figure 5.16: Micro-Stage Cooling Temperature Response for Varying *CTRL* Step Input Voltages

5.2.3 Full Model of Actuator

The electro-thermal model is combined with the Preisach model to form a full model for the SMA actuator, as was shown in the block diagram of Figure 5.12. The resulting model is validated with experimental data as given in Figure 5.17.



(a) (b)

Figure 5.17: Complete SMA Actuator Model Validation for (a) 1.5-1.6 V and (b) 1.65-1.8 V Step Input

As can be seen in Figure 5.17, the resulting model gives good qualitative results, especially in steady state, but poor matching at large input signals. It should be noted that the large 1.8 V step simulated response appears to match the experimental, but are in fact quite dissimilar when the time scale is taken into consideration.

Two difficulties that contribute to the poor matching are the mechanical stop and the sensitivity of the actuator at high temperatures. The mechanical stop introduces a saturation nonlinearity that prevents the identification of actuator dynamics at high input voltages. Also, at high input voltages the sensitivity of the actuator is quite large. This is attributed to the exponential characteristics of the developed Preisach model at high temperatures. At high temperatures the sensitivity of the position to changes in temperature dramatically increases. Thus, a minor temperature error results in a large position error. This high sensitivity generates a large amount of error in simulation for large contractions, and also introduces imprecision in the critical step of inverting experimental position data to infer actuator temperature.

While the model provides good qualitative matching to allow for simulation estimates of actuator response for various inputs, it is not accurate enough to use as a basis for controller design. The poor matching at large input voltages is problematic for controller design. The actuator must operate well in this area for large contractions. However, since the model does not accurately reflect the experimental system in this area, a controller cannot be accurately designed offline using the model. Therefore, the model is abandoned and not used for the design and validation of a closed looped controller of the micro-stage.

5.3 Closed Loop Control

Since the full model of the actuator provides only qualitative agreement with the SMA actuator within the micro-stage, it is not used to design the controller, $D(z)$, of the closed looped system given in the block diagram of Figure 5.18. The controller is instead designed experimentally through ad-hoc design methods. There are a few formal methods of controller design for SMAs such as dissipativity-based control described by Gorbet *et al.* [37], or an adaptive proportional controller implemented by Dickinson and Wen [38]. However, they require knowledge of the SMA properties which are unknown here, and the driver circuitry that exists in the NanoMuscle^{TM} introduces nonlinearities that may not be handled by the control strategies. As a first attempt, a PI controller is tuned experimentally on the micro-stage. However, integrator windup becomes a significant problem due to the hysteresis of the SMA actuator. Therefore, an anti-windup scheme is implemented and refined experimentally.

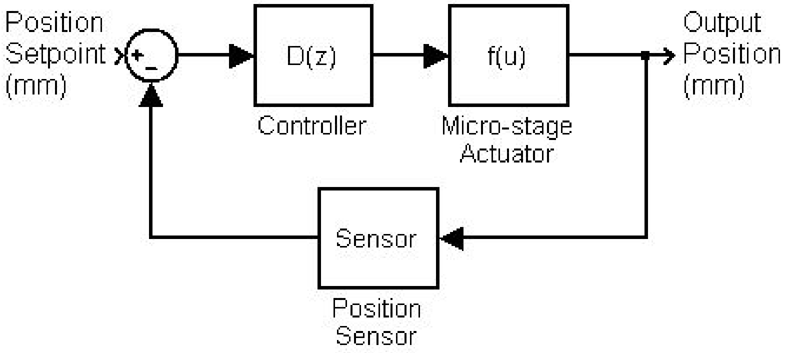


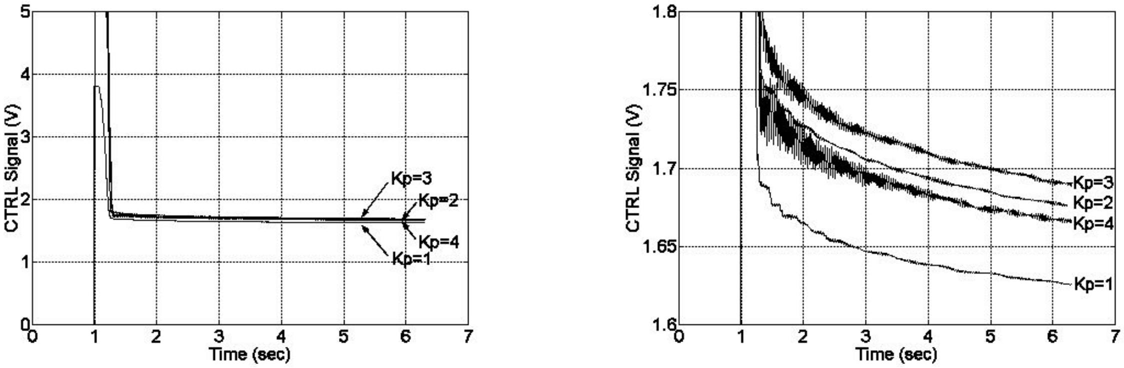
Figure 5.18: Closed Loop Block Diagram of Micro-Stage

5.3.1 PI Controller

To develop a controller experimentally, the proportional gain, K_p, is first set. This is done by analyzing the step response of the system for a target of 3.8 mm, which is near the maximum stroke of the system, for varying controller gains. Applying proportional control to the micro-stage with various gains, the resulting $CTRL$ signal to the actuator, the controller output, is shown in Figure 5.19, and the output position is shown in Figure 5.20.

Generally, a larger K_p provides a faster response. However, if the gain is increased too high, oscillations can result. This is most apparent in Figure 5.19, where it can be seen that, with increasing gain, the amount of resulting oscillation increases. Therefore, the largest gain that provides minimal oscillation is chosen experimentally. The selected gain is $K_p = 3$.

What is also apparent from Figure 5.20 is that steady state error for the micro-stage cannot be eliminated solely with proportional control. This is in agreement with the results noted in [22] and [38] that proportional feedback alone for SMA actuators cannot remove steady state error. Therefore, integral control is implemented.



(a) (b)

Figure 5.19: (a) Full View and (b) Zoomed View of Controller Output for Micro-Stage with Proportional Control

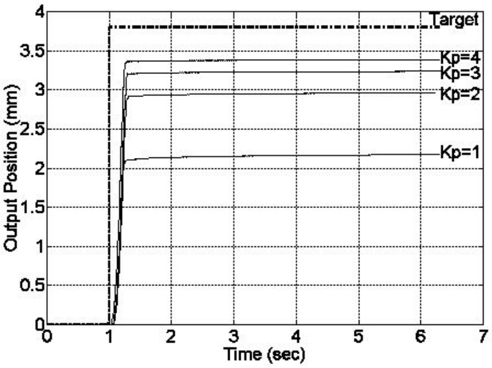


Figure 5.20: Position Response for Micro-Stage with Proportional Control

The block diagram of the discrete time PI controller to be designed is shown in Figure 5.21

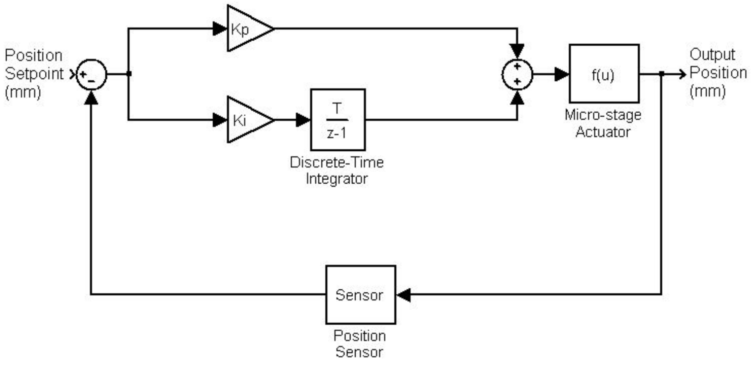
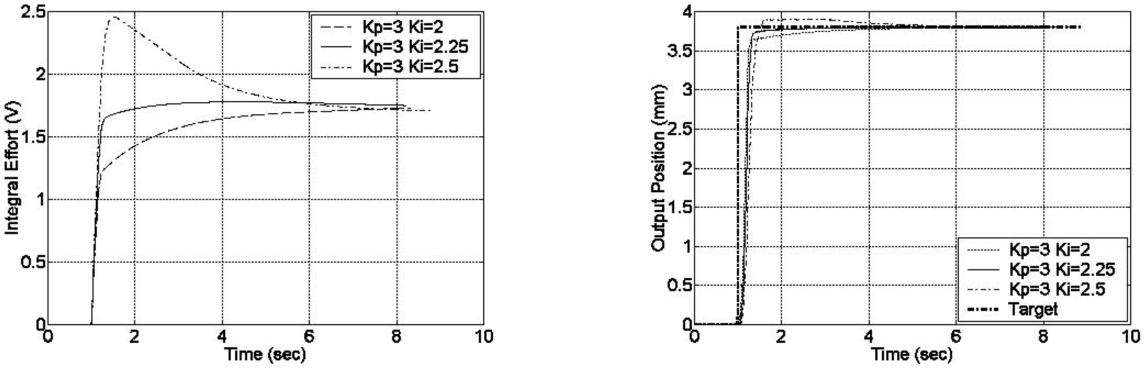


Figure 5.21: Block Diagram of Micro-Stage Using PI Control

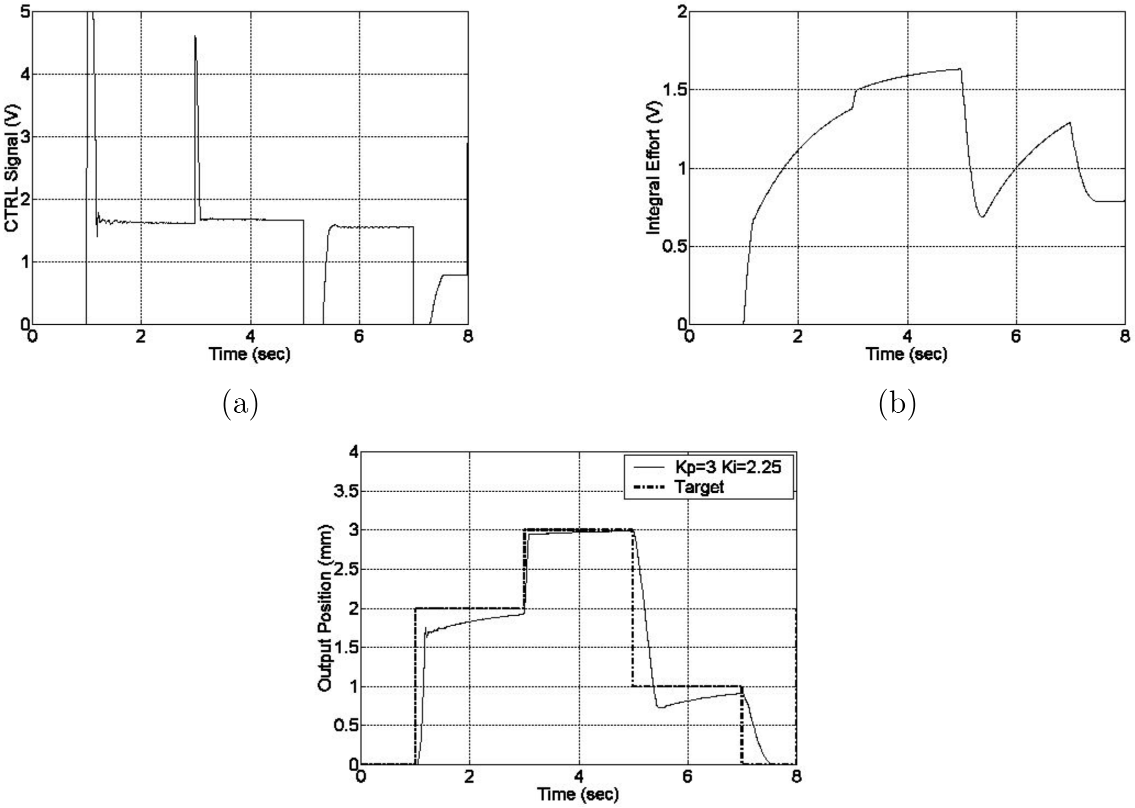
To design the integrator gain K_i, the closed loop response of the system with various integrator gains is examined. Figures 5.22a and 5.22b illustrate the integrator effort and the position output for various integral gains and $K_p = 3$ respectively.



(a) (b)

Figure 5.22: Resultant(a) Integral Effort and (b) Position Response of Micro-Stage with PI Control

As can be seen in Figures 5.22a and 5.22b, a small K_i can result in a slow ramp up of the integrator effort to reduce the steady state error, resulting in a slow response. However if the gain is slightly increased, overshoot can occur and a slow ramp down of the integrator effort can occur, again resulting in a slow response. A value of $K_i = 2.25$ is chosen since the resulting integrator effort reaches steady state the quickest without overshoot. If a sequential step input is applied to the system to test the resultant PI controller of $K_p = 3$, and $K_i = 2.25$, the resultant output is shown in Figure 5.23.



(a) (b)

(c)

Figure 5.23: Sequential Step Input (a) Controller Output, (b) Integral Effort, and (b) Position Response of Micro-Stage for PI Control

It can be seen from Figure 5.23c that the resultant response does not match the target input very well. The reason for the discrepancy can be uncovered from Figure 5.23b. The integrator effort is slow to respond to reduce steady state error for the two increasing steps at 1 and 3 seconds. This would suggest a larger K_i is required. However, for the decreasing step that occurs at 5 seconds, the integrator effort overshoots, suggesting a lower K_i is required. The overshoot is caused by windup of the integrator. During actuator extension, the controller output saturates at 0 V to provide maximum cooling, this is evident in Figure 5.23a at the 5 and 7 second marks when a decreasing step input occurs. However, during this cooling period the integrator continues to windup, resulting in overshoot.

This error can be remedied by the implementation of an anti-windup scheme. An anti-windup scheme will also allow for a larger integral gain, since it minimizes overshoot of the integrator. Integrator overshoot, and therefore windup, is the main drawback of increasing K_i too aggressively.

5.3.2 PI with Anti-Windup Controller

Integrator windup is a common problem in systems with saturation. If a control error exists during saturation in a system with integral control, the integrator will continue to "windup" the control signal in an attempt to eliminate steady-state error. This windup results in degraded performance in closed loop.

The classic anti-reset windup technique outlined in [39] is adopted for use here. A block diagram of the anti-windup system is shown in Figure 5.24. The static nonlinearity that causes the windup to occur is taken as the saturation block of Figure 5.24. In addition to the proportional and integral gains or K_p and K_i respectively, an additional anti-windup gain value, K_{aw} is required. To simplify the analysis, the condition is set that the integrator and anti-windup gains must be equal. This condition essentially requires that the time constant of the integrator and the anti-windup control are equal.

First, the limits of the saturation block of Figure 5.24 must be found. The first most evident form of saturation that exists in the micro-stage occurs during actuator extension, when the actuator is cooling the controller saturates at 0 V. In fact, at input $CTRL$ signals below 1.3V the actuator does not respond. Therefore, the lower saturation limit of the integral controller is set at 1.3 V. The upper saturation limit is not as obvious. It was also shown from Figure 5.13b that, in general, as the input signal is increased, the response time decreases. This is true for input signals between 1.3 V and 1.8 V. For input signals greater than 1.8 V, the change in response time is minimal. A slightly larger value of 1.9 V is chosen as the upper limit for performance, and thus can be considered to be the upper saturation limit on the integrator.

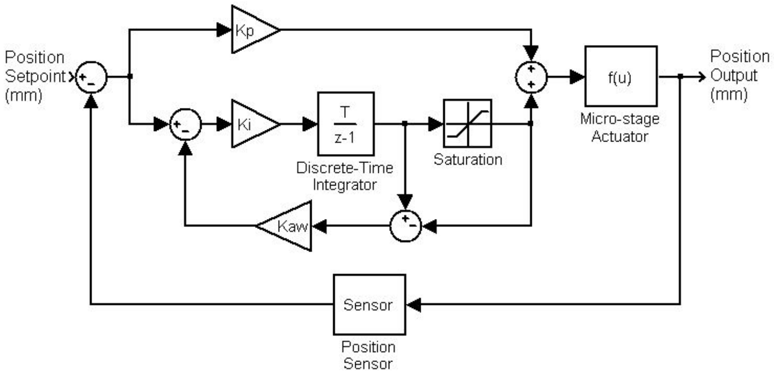


Figure 5.24: Block Diagram of Micro-Stage Using PI Control with Anti-Windup

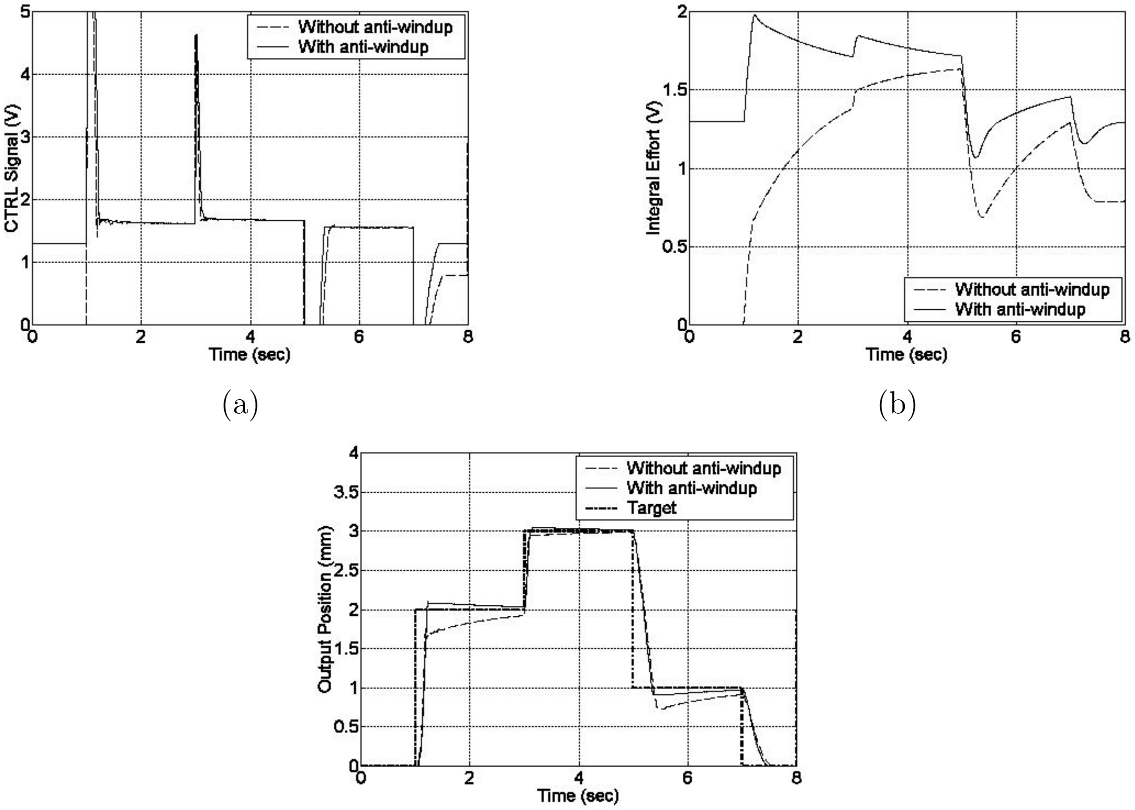
The addition of the anti-windup loop to the PI controller results in more effective operation of the controller, as shown in the responses of Figure 5.25.

To improve performance by more quickly reducing the steady state error, K_i can be increased. However, if K_i and thus K_{aw} are increased too aggressively, instability can result. Therefore the optimal K_i and K_{aw} values are found by searching for the largest values that do not induce oscillation. The 3.8 mm step response is revisited and the integrator effort and position response of the micro-stage for various values of K_i is shown in Figures 5.26a and 5.26b respectively.

From Figure 5.26a, $K_i = 30$ appears to be the largest gain value before oscillation begins to appear, and is therefore adopted. The final PI controller with anti-windup control is implemented and tested with a sequential step input. The result is shown in Figure 5.27. As can be seen, the controller performs extremely well, producing little overshoot and a settling time of approximately 0.5 seconds in contraction and 1 second for extension. As can be seen in Figure 5.27b, the 5 μm target accuracy for the micro-stage has been met.

5.4 Review of Design

The micro-stage consists of a NanoMuscle^{TM} SMA actuator in series with a linear flexure guide and a bias spring. Feedback is provided via a Hall effect sensor and magnets.

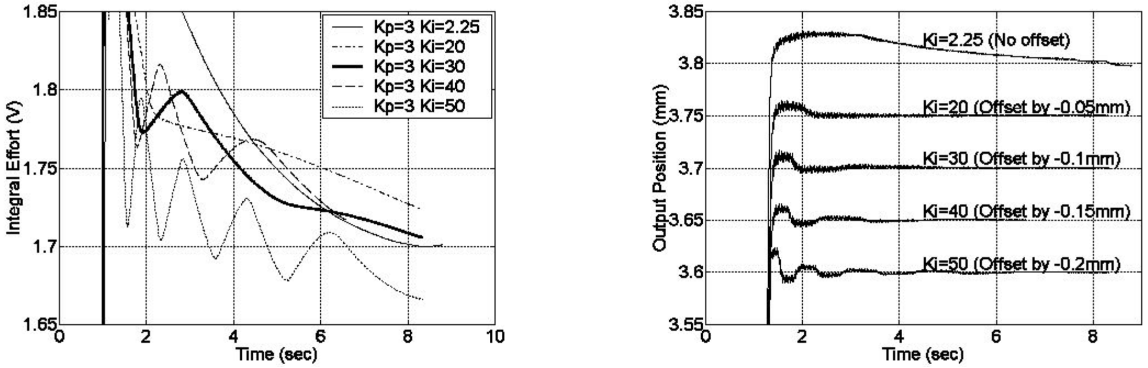


(a)

(b)

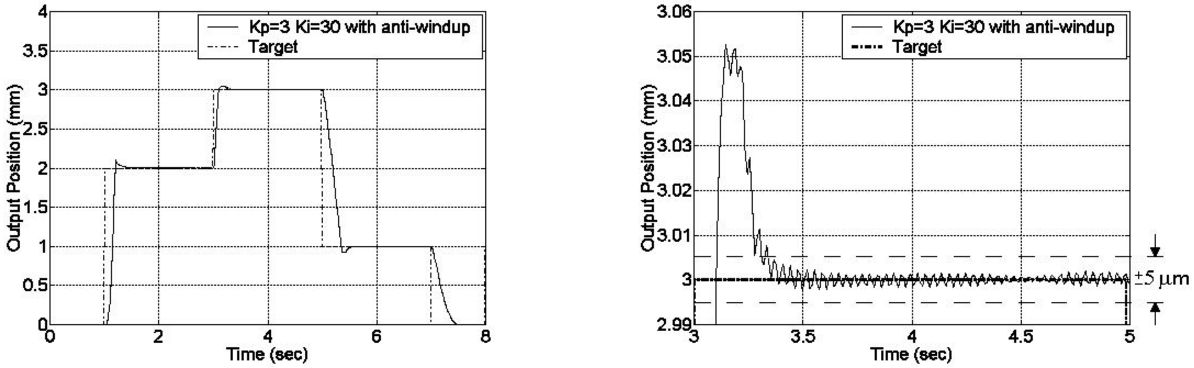
(c)

Figure 5.25: Sequential Step Input (a) Controller Output, (b) Integral Effort, and (b) Position Response of Micro-Stage for PI Control with Anti-Windup



(a) (b)

Figure 5.26: Step Input (a) Zoomed view of Integrator Effort and (b) Position Response of Micro-Stage for PI Control with Anti-Windup and Various Values of K_i



(a) (b)

Figure 5.27: Sequential Step Input (a) Full and (b) Zoomed Response of Micro-Stage with PI Control with Anti-Windup. ($K_p = 3$ and $K_i = K_{aw} = 30$)

The dynamics of the SMA actuator were divided into linear and nonlinear properties. The linear portion was assumed to consist of the thermal and electrical dynamics, which proved to be incorrect, while the SMA hysteresis formed the nonlinear portion of the model. Also, it was assumed that the linear and nonlinear dynamics were separable and could thus be considered independently.

An electro-thermal model was formed by applying MLS to correlate the input voltage to the output temperature of the actuator. Due to the fact that the output temperature cannot be measured directly during operation, a method of inferring the actuator temperature from output contration length was used. This method of inferring the temperature required experimental FOD data of the actuator.

The Preisach model of the actuator hysteresis was identified using experimental data. The data was obtained by heating the actuator in a kiln to control temperature directly.

Upon analysis of experimental data, the electro-thermal behaviour was found to be nonlinear. However, by dividing the model into a nonlinear gain function and a linear dynamic system, a satisfactory model was developed. The nonlinear gain was generated from the steady state relationship between input voltage and output temperature, and the linear dynamics were formed by applying MLS to experimental data.

The resulting electro-thermal model was combined with the Preisach model to form a full model of the SMA actuator and thus the micro-stage. While the full model showed good qualitative agreement with system behaviour, the match was deemed unsatisfactory for controller development. Therefore, the model of the micro-stage was abandoned and not used for controller design.

Closed loop control was then implemented experimentally using a tuned PI controller. However, integrator windup proved to be a problem and degraded performance. Thus, the classical anti-windup technique was implemented and tuned experimentally to allow for large integrator gains to increase performance without windup. The resulting controller provided excellent performance, meeting the 5 μm target accuracy set initially.

With closed looped control implemented for both the macro- and micro-stages, both systems can now be used to form the full macro-micro positioning system.

Chapter 6

Results

With both the macro- and micro-stages designed and implemented in closed looped control, the integrated macro-micro positioning system can be tested. The control strategy for the macro-micro system is described, and experimental results are presented for a test input of a decreasing square wave for the macro stage and a sequential step for the micro stage. Vibration issues from the macro-stage are briefly discussed, as well as the changes in properties of the micro-stage due to cycling. Finally, the performance and cost of the system is reviewed.

6.1 Integrated Control Strategy

The final goal is to have the macro-stage reach a setpoint and then allow the micro-stage to operate with precision within a small area. Therefore, the controllers of the two stages are operated independent of each other. The macro-stage is assigned a setpoint and allowed to reach steady-state before the micro-stage is invoked. A block diagram of the control strategy is shown in Figure 6.1.

Though the feedback is provided by the Hall effect sensor, the laser measurement system is used to measure the displacement to verify correct operation of the system. It is noted that the absolute resolution of the system is limited by the macro-stage precision of 200 μm. However, the goal of the system is to use the macro-stage to position the micro-stage, at coarse accuracy, with respect to a feature. Thus allowing the micro-stage to work with

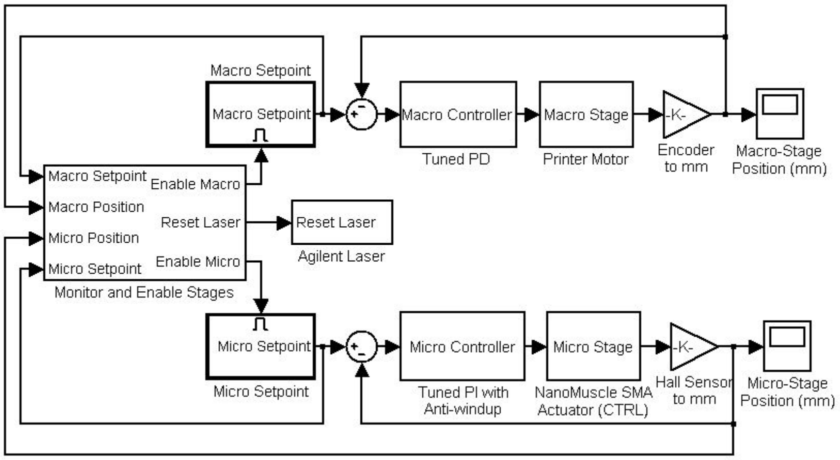


Figure 6.1: Block Diagram of Complete Macro-Micro Control Strategy

precision within a small area. Therefore, the laser is reset after the macro-stage reaches its setpoint. This initializes the laser measurement system to the current position of the micro-stage.

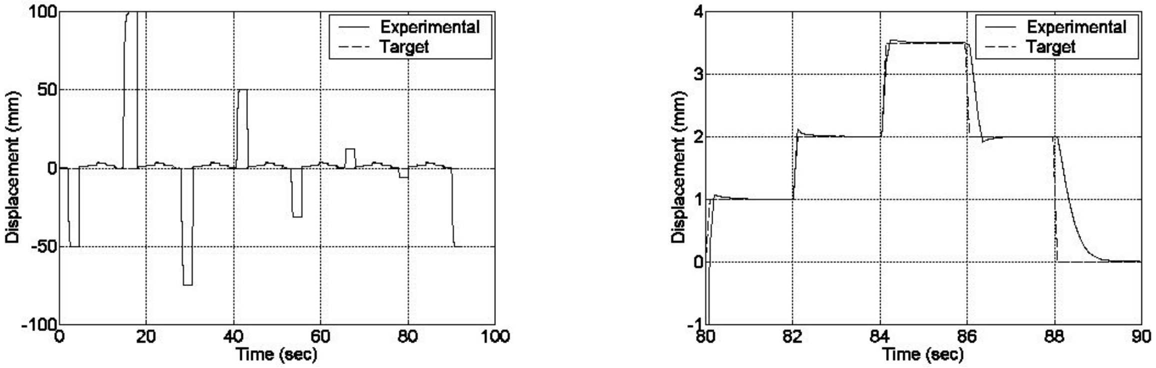
What must also be noted is that a two second delay is introduced between the time at which the macro-stage reaches steady state and when the micro-stage is enabled and the laser reset. The reason for this is to allow oscillations to decay. These oscillations are described in the next section.

A test input consisting of a decaying square wave for the macro-stage, and a sequential step for the micro-stage after each macro movement, is applied to the system to gauge its performance. The resultant response of the full system with the test input is shown in Figure 6.2.

6.2 Macro-Stage Vibration

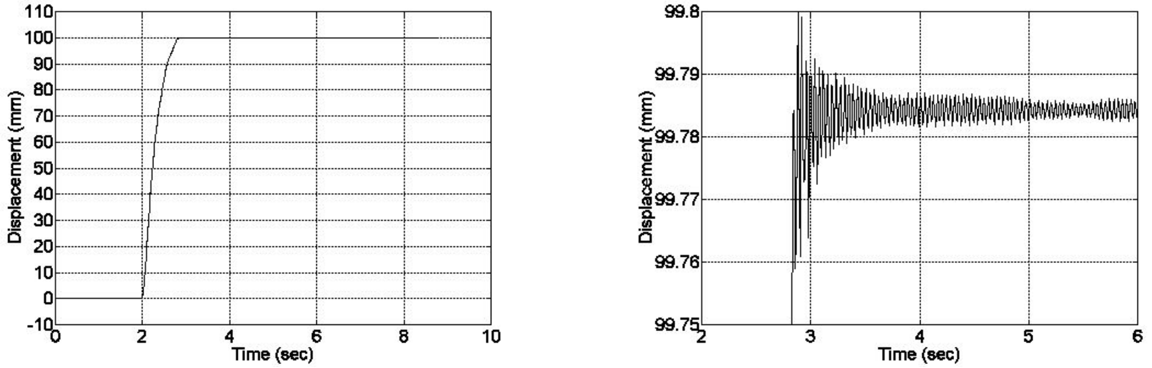
The macro-stage induces a small amount of vibration in the system when it reaches its setpoint. The main cause of the oscillation, shown in Figure 6.3, is likely the flexibility of the drive belt.

The induced vibration has a maximum amplitude of approximately 20 μm, and a decay



(a) (b)

Figure 6.2: (a) Full and (b) Zoomed View of Integrated System Response



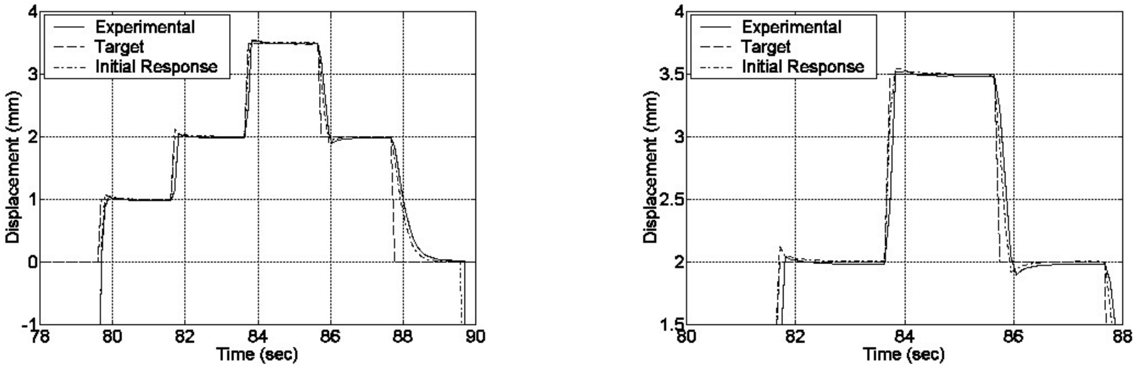
(a) (b)

Figure 6.3: (a) Full and (b) Zoomed View of Macro-Stage Vibration

time of approximately two seconds. As this oscillation is quite large relative to the target precision of 5 μm precision, damping techniques may be required in the future.

6.3 Repeated Cycling

As indicated in [36], the characteristics of the NanoMuscle^{TM} actuator can change over time over repeated cycling . This is due to the gradual warming of the actuator to higher base temperatures. If the macro-micro system is continually cycled with the test input, Figure 6.4 displays the output of the micro-stage after ten cycles of the macro-micro system, or approximately 15 minutes or operation.



(a) (b)

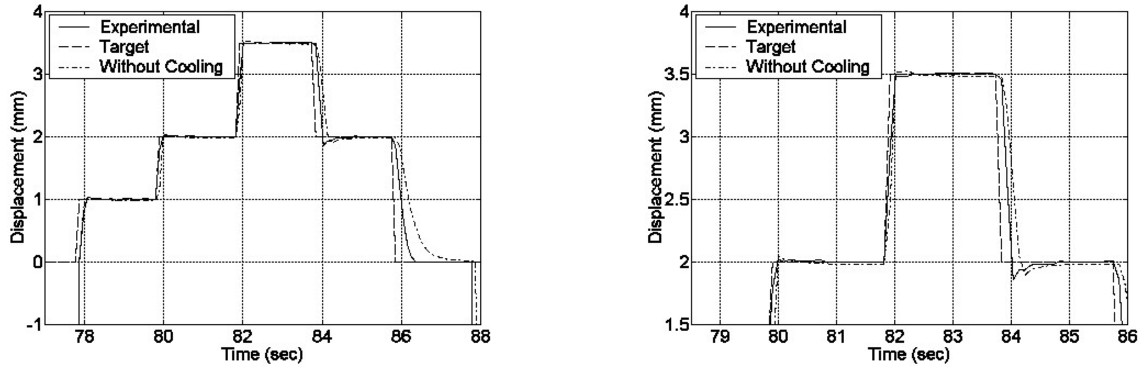
Figure 6.4: (a) Full and (b) Zoomed View of Micro-Stage Response After 10 Cycles of Macro-Micro Operation

As can be seen, the performance of the micro-stage is degraded after ten cycles. The extension, and thus cooling time, increases as expected. More importantly, the stage appears to no longer meet the setpoint displacement at steady state. This is only partially true. As the actuator warms, it becomes increasingly difficult for it to cool and fully extend. Since the laser measurement system is reset after every cycle, the laser can be reset before the actuator cools to the fully extended position. This results in the measurement bias seen here.

These changes in actuator performance errors may be remedied by keeping the actuator cool, or by limiting bandwidth and overall duty cycles.

6.3.1 Forced Cooling

To test the hypothesis that a larger cooling rate would improve cycling performance, a fan is placed directly in front of the experimental system to provide forced convection. As a result, the cycling characteristics are improved. Figure 6.5 gives the micro-stage response after ten cycles of the macro-micro system with forced cooling.



(a)

(b)

Figure 6.5: (a) Full and (b) Zoomed View of Micro-Stage Response After 10 Cycles of Macro-Micro Operation With Cooling

With the forced convection, the extension time remains fairly fast at one second. Also, the steady state problem that was seen in Figure 6.4 no longer exists.

6.4 Performance

With the limitations of the system reviewed, a brief analysis of performance can be made. Since there are as yet no specific applications for the system, the highest achievable performance is researched. The two performance characteristics that are of interest in most positioning applications are accuracy and bandwidth.

6.4.1 Accuracy

As reviewed in Section 3.2.3, the target accuracy of the system using the Hall effect sensor and magnets was 5 μm and was achieved, as shown in Chapter 5. If the laser system is used instead, the resolution improves only slightly because the accuracy of the system is limited by noise. If the system is placed at rest and laser measurement is used, the result can be seen in Figure 6.6.

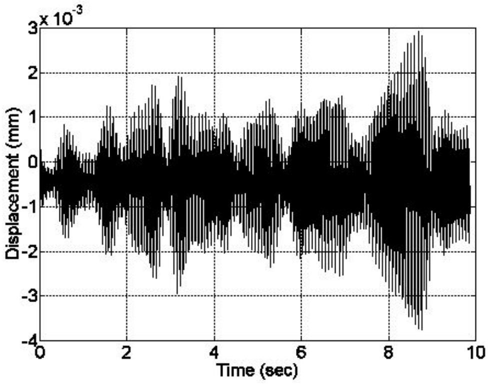


Figure 6.6: Apparent Noise in System Using Laser Measurement

From Figure 6.6, the noise produced has an amplitude of 2 μm, thus the maximum attainable accuracy from the laser is 2 μm. The accuracy of the laser is highly dependent on environmental conditions such as air pressure, temperature, and humidity. Minute changes in the environment, along with mechanical vibration are likely sources for the noise. To increase the accuracy in the future, this noise must be reduced.

6.4.2 Bandwidth

The bandwidth of the system is limited by the speed of the printer and the cooling of the SMA actuator.

It takes the macro-stage approximately 0.75 seconds for it to move from one extremity to the other in closed loop control. This translates into an approximate macro-stage bandwidth of 0.67 Hz in closed loop control.

Table 6.1: Cost Breakdown of Micro-Stage

Component	Approximate Cost ($ USD)
Hall Effect Sensor	10
Magnets	5
Flexure Guide	100
NanoMuscle with Spring	100
Magnet Mount	20
NanoMuscle to Flexure Guide Linkage	20
Total Cost	255

Table 6.2: Cost Comparison of Micro-Positioning Systems

Technology	Accuracy	Approximate Cost ($ USD)
SMA Micro-Stage	5 μm	255
Miniature DC motor with lead screw and nut	0.1 - 1 μm	1500
Ultrasonic piezo-drives	0.1 - 1 μm	4000
Piezo stacks and Flexure guide	<1 nm	4000
iezo Inchworm Motor	20 nm	8000

As for the micro-stage with the conditions analyzed here with a cooling fan, a cycle time of approximately 2 seconds is achieved resulting in a bandwidth of 0.5 Hz. However, with further cooling of the actuator it may be possible to reduce the cycle time.

6.4.3 Cost

To estimate a cost for the micro-stage, the cost of material and manufacture for each individual component is summed. A summary of the cost is shown in Table 6.1.

The total cost for manufacture of the prototype micro-stage is found to be approximately $255 USD. This is lower than the cost of other micro-positioning technologies outlined previously in Table 2.1, though the accuracy is lower. A comparison of the costs are shown in Table 6.2.

Chapter 7

Conclusions

The developed experimental system comprises a modified printer carriage to provide long range, macro scale linear motion (approximately 200 mm range and 200 μm precision) and a micro scale system (approximately 4 mm range and 5 μm precision) that uses a NanoMuscle^{TM} SMA based actuator. Both systems are monitored and computer controlled by Simulink^{TM} via a MultiQ-3 data acquisition board (MQ3) manufactured by Quanser Consulting.

A model of the macro-stage was identified with MLS identification, but first the Coulomb friction was modelled as a dead zone for linearization through its inverse. As a result an experimentally tuned nonlinear model was developed for the macro-stage, and validated through comparison with experimental data. The model was then used to develop a tuned nonlinear controller offline and shown to produce excellent results.

The micro-stage consists of a NanoMuscle^{TM} SMA actuator in series with a linear flexure guide and a bias spring. Feedback is provided via a Hall effect sensor and magnets. The current to the SMA wires in the NanoMuscle^{TM} is not directly controlled but through an unknown driver circuit by an input signal, $CTRL$.

The dynamics of the SMA actuator were divided into two halves. The thermal and electrical dynamics formed one half, and the SMA hysteresis formed the second half. It was assumed that the two halves were separable and therefore were considered independently.

The electro-thermal model was formed by applying MLS to correlate the input $CTRL$ voltage to the output actuator temperature. The actuator temperature was inferred from the output contraction length using experimental FOD data of the actuator. The SMA hysteresis was modelled with the experimental implementation of the Preisach model, using the FOD data obtained by heating the actuator in a kiln.

The electro-thermal model was further divided into a nonlinear gain function and a linear dynamic system. The nonlinear gain was discovered from the steady state relationship between input voltage and output temperature, and the linear dynamics were formed by applying MLS to experimental data.

The resulting electro-thermal model was combined with the Preisach model to form a full model of the SMA actuator and thus the micro-stage. While the full model showed good qualitative agreement with system behaviour, the match was deemed unsatisfactory for controller development. Closed loop control for the micro-stage was implemented using an experimentally-tuned PI controller with anti-windup control. The resulting controller provided excellent performance.

Both the macro and micro stages were then operated together in closed loop to form the integrated macro-micro system and a laser measurement system was used to verify the operation of the micro-stage. The following results were found:

- It was found that the macro-stage induced vibration whenever it reached its steady state position. This vibration has an initial magnitude of 20 μm and an approximate 2 second decay time.

- If the system is continually cycled, the SMA actuator warms and cooling becomes difficult resulting in a degradation of performance since actuator extension becomes slower. The warming of the actuator can also lead to the micro-stage not being able to fully extend.

- If forced convection is provided with a fan to cool the micro-stage, the performance over repeated cycling improves.

- The target accuracy of 5 μm for the micro-stage was met.

- Using the laser measurement as feedback for the micro-stage, it was found that the maximum attainable precision is 2 μm. This precision is limited by the accuracy of the laser measurement system in its current setup. Under operation, it was found that an oscillation of 2 μm exists in the measured position. The source of this oscillation must be found and compensated for in the future.

- The bandwidth of the macro-stage was found to be approximately 0.67 Hz while that of the micro-stage is approximately 0.5 Hz.
- The cost of the micro-stage was lower than other micro-positioning technologies. The estimated cost of the micro-stage is $255 USD.

7.1 Future Work

With a proof of concept macro-micro positioning system designed and implemented in closed loop, some issues concerning the system have arisen.

The macro-stage induces minute vibration in closed loop operation. This vibration results in micron-range vibration of the micro-stage. Therefore, the coupling between the macro- and micro-stages must be studied to discover a method of reducing this vibration. The oscillation may be reduced with different control strategies, possibly requiring a model of the flexibility within the macro-stage drive belt.

Also, a more sophisticated friction model for the macro-stage can be developed to resolve discrepancies between the simulated and experimental system. With a better friction model, a more effective friction compensation method can also be developed.

For the micro-stage, it was shown that the developed model was poor. The observed nonlinear behaviour in the experimental system was due to the unknown driver circuitry of the actuator, and the nonlinear properties of SMA. Future research into the driver circuitry would be beneficial to develop a better model of the micro-stage. A better model would also allow for the formal design of a controller for the micro-stage.

Also, the precision of the micro-stage is so far limited by the accuracy of the feedback system. Currently the maximum precision is 2 μm using a laser measurement system. However, the laser is in theory capable of measuring lengths in the nanometre range (~5 nm). Unfortunately, measurement noise limits the laser system from measuring in the sub-micron range. To increase precision of the micro-stage, this noise must be analyzed and remedied.

Finally, further into the future, the bias spring, flexure guide, and SMA actuator of the micro-stage may be replaced with a single monolithic SMA device. A monolithic SMA actuator, similar to the actuator developed by Toews [40] for use in a miniature robotic device, would reduce the complexity, and possibly cost, of the micro-stage. Also, a monolithic SMA actuator would likely remove the uncertainty that exists from the NanoMuscle^{TM} driver circuitry.

Bibliography

[1] I.J. Busch-Vishniac. Micro-Automating Semiconductor Fabrication. *IEEE Circuits and Devices*, 7:32–37, 1991.

[2] Sensor Technology Limited. Piezoelectric Stack Actuators. http://www.sensortech.ca (current Aug 1, 2004).

[3] T. King, M. Pozzi and A. Manara. Piezoactuators for 'Real-World' Applications. *Power Engineering Journal*, 14(3)(3):105–110, 2000.

[4] A. Sharon and D. Hardt. Enhancement of Robot Accuracy using Endpoint Feedback and a Macro-Mmicro Manipulator System. *Proceedings of the American Control Conference*, pages 1836–1842, 1984.

[5] N.G. Chalhoub and X. Zhang. Reduction of the End-Effector Sensitivity to the Structural Defections of a Single Flexible Link: Theoretical and Experimental Results. *Transactions of American Society of Mechanical Engineering*, JDSMC 115:658–666, 1993.

[6] T. Narikiyo, H. Nakane, T. Akuta, N. Mohri and N. Saito. Control System Design for Macro/Micro Manipulator with Application to Electrodischarge Machining. *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*, 2:1454–1460, 1994.

[7] Duk-Young Lee and Hyungsuck Cho. Precision Force Control via Macro/Micro Actuator for Surface Mounting System. *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*, 3:2227–2232, 2002.

[8] J.D. Glisinn, B.N. Damazo, R. Silver and H. Zhou. A Macro-Micro Motion System for a Scanning Tunneling Microscope. *Proceedings of the 5th Biannual World Automation Congress*, 14:280–289, 2002.

[9] David Baselt. Atomic Force Microscopy. http://stm2.nrl.navy.mil/how-afm/how-afm.html (current Aug 1, 2004).

[10] Newport. *The Newport Resource.* Newport, California, 2003.

[11] Physik Instrumente. Products and Specs. http://www.physikintrumente.de (current Aug 1, 2004).

[12] Polytec PI Inc. *Polytec PI Inc. Price List.* Physik Instrumente, Germany, 2000.

[13] EXFO Burleigh. EXFO Products-Precise Positioning. http://www.exfo.com/en/products/ProductsFamily.asp?Family=216 (current Aug 1, 2004).

[14] Burleigh Instruments Inc. *Nanopositioning Instrumentation Price List.* Burleigh Instruments, Inc., New York, 2000.

[15] T.W. Duerig, K.N. Melton, D. Stöckel and C.M. Wayman. *Engineering Aspects of Shape Memory Alloys.* Butterworth-Heinemann Ltd., London, England, 1990.

[16] Daniel R. Madill and David Wang. The Modelling and L_2-Stability of a Shape Memory Alloy Position Control System. *1994 IEEE International Conference on Robotics and Automation*, 1:293–299, 1994.

[17] I.D. Mayergoyz. *Mathematical Model of Hysteresis.* Springer-Verlag, New York, 1991.

[18] J.M. Cruz-Hernandez and V. Hayward. On the Linear Compensation of Hysteresis, TR-CIM-97-08. Technical report, McGill University, Montreal, Canada, 1997.

[19] J.M. Cruz-Hernandez and V. Hayward. An Approach to Reduction of Hysteresis in Smart Materials. *Proceedings of the 1998 IEEE International Conference on Robotics and Automation*, 2:1510–1515, 1998.

[20] J.M. Cruz-Hernandez and V. Hayward. Phase Control Approach to Hysteresis Reduction. *IEEE Transactions on Control Systems Technology*, 9(1):17–26, 2001.

[21] R.B. Gorbet, D.W.L. Wang and K.A. Morris. Preisach Model Identification of a Two-Wire SMA Actuator. *Proceedings of the 1998 IEEE International Conference on Robotics and Automation*, 3:2161–2167, 1998.

[22] C.A. Dickinson, D.C. Hughes and J.T. Wen. Hysteresis in Shape Memory Alloy Actuators: The Control Issues. *Smart Structures and Materials*, 2715:494–506, 1996.

[23] NanoMuscle Inc. 70 Gram HS/HE Linear Actuator Manual. http://www.nanomuscle.com (current Aug 1, 2004).

[24] S.T. Smith and D.G. Chetwynd. *Foundations of Ultraprecision Mechanism Design.* Gordon and Breach Science, Pennsylvania, 1992.

[25] Honeywell International Inc. Magnetic Position Sensors. http://content.honeywell.com/sensing/prodinfo/solidstate.html (current Aug 1, 2004).

[26] Melexis. Hall Effect Sensors. http://www.melexis.com/prod_hall.asp (current Aug 1, 2004).

[27] Agilent Technologies. *Laser and Optics User's Manual.* Agilent Technologies, Santa Clara, California, 2002.

[28] Quanser Consulting. *MultiQ-3 Programming Manual.* Markham, Ontario. http://www.quanser.com (current Aug 1, 2004).

[29] H. Olsson, K.J. Astrom, C. Canudas de Wit, M. Gafwert and P. Lischinsky. Friction Models and Friction Compensation. *European Journal of Control*, 4(3):176–195, 1998.

[30] C. Canudas de Wit, H. Olsson, K.J. Astrom and P. Lischinsky. A New Model for Control of Systems with Friction. *IEEE Transactions on Automatic Control*, 40(3):419–425, 1995.

[31] T. Kubo, G. Anwar and M. Tomizuka. Application of Nonlinear Friction Compensation To Robot Arm Control. *Proceedings of the 1986 International Conference on Robotics and Automation*, 3:722–727, 1986.

[32] B. Armstrong-Helouvry, P. Dupont and C. Canudas de Wit. Survey of Models, Analysis Tools, and Compensation Methods for Contol of Machines with Friction. *Automatica*, 30(7):1083–1138, 1994.

[33] L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, New Jersey, 1987.

[34] C.L. Phillips and H.T. Nagle. *Digital Control System Analysis and Design*. Prentice-Hall, New Jersey, 1987.

[35] D. Hughes and J.T. Wen. Preisach Modeling of Piezoceramic and Shape Memory Alloy Hysteresis. *IEEE International Conference on Control Applications*, pages 1086–1091, 1995.

[36] E. Ho and R.B. Gorbet. Potential of the NanoMuscle SMA Actuator for Fine Position Control. *3^{rd} Annual CanSmart International Workshop on Smart Materials and Structures*, 2003. Montreal, QC.

[37] R.B. Gorbet, K.A. Morris and D.W.L. Wang. Passivity-Based Stability and Control of Hysteresis in Smart Actuators. *IEEE Transactions on Control Systems Technology*, 9(1):5–16, 2001.

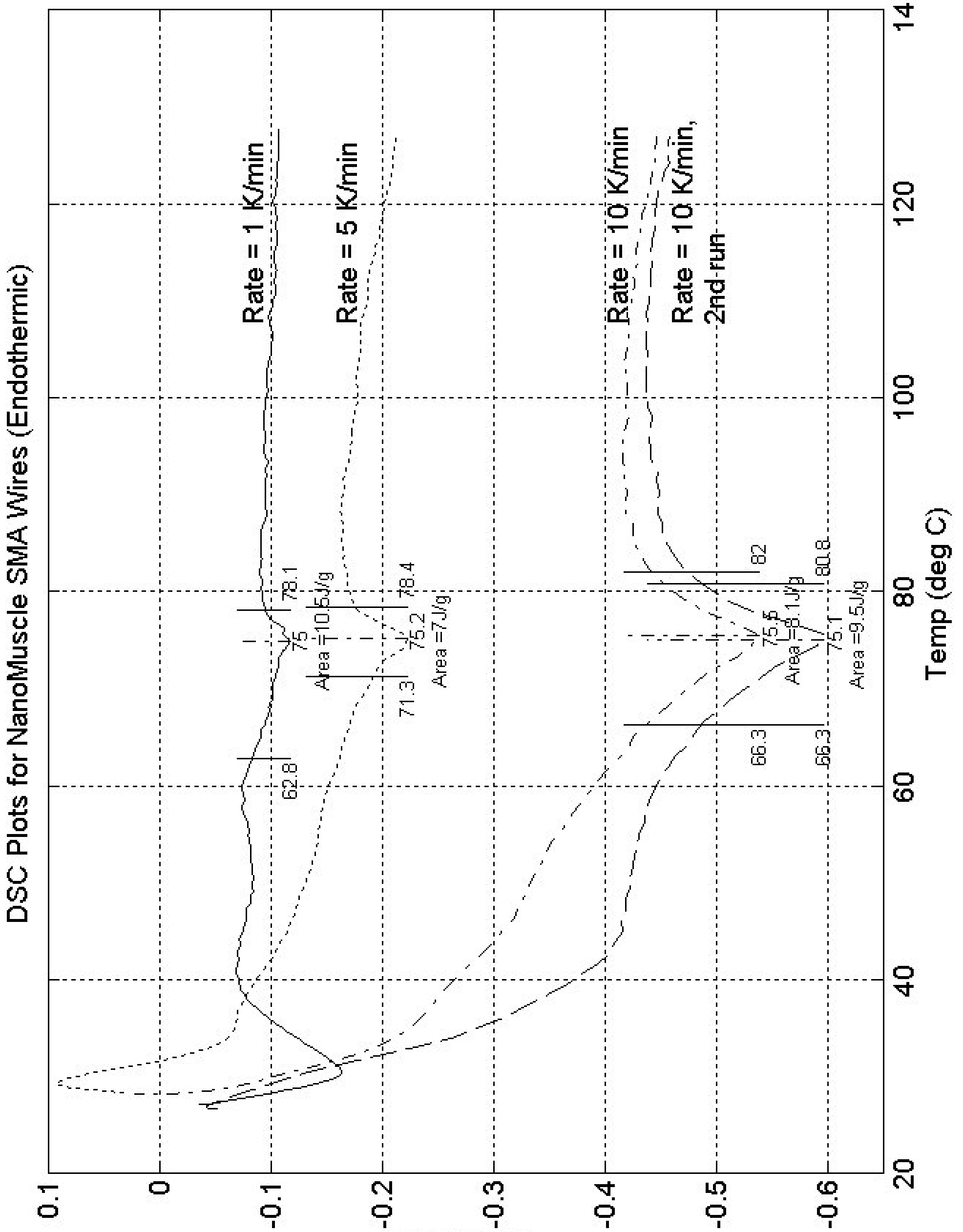
[38] C.A. Dickinson and J.T. Wen. Feedback Control Strategies for Shape Memory Alloy Actuators. *Journal of Intelligent Material Systems and Structures*, 9(4):242–250, 1998.

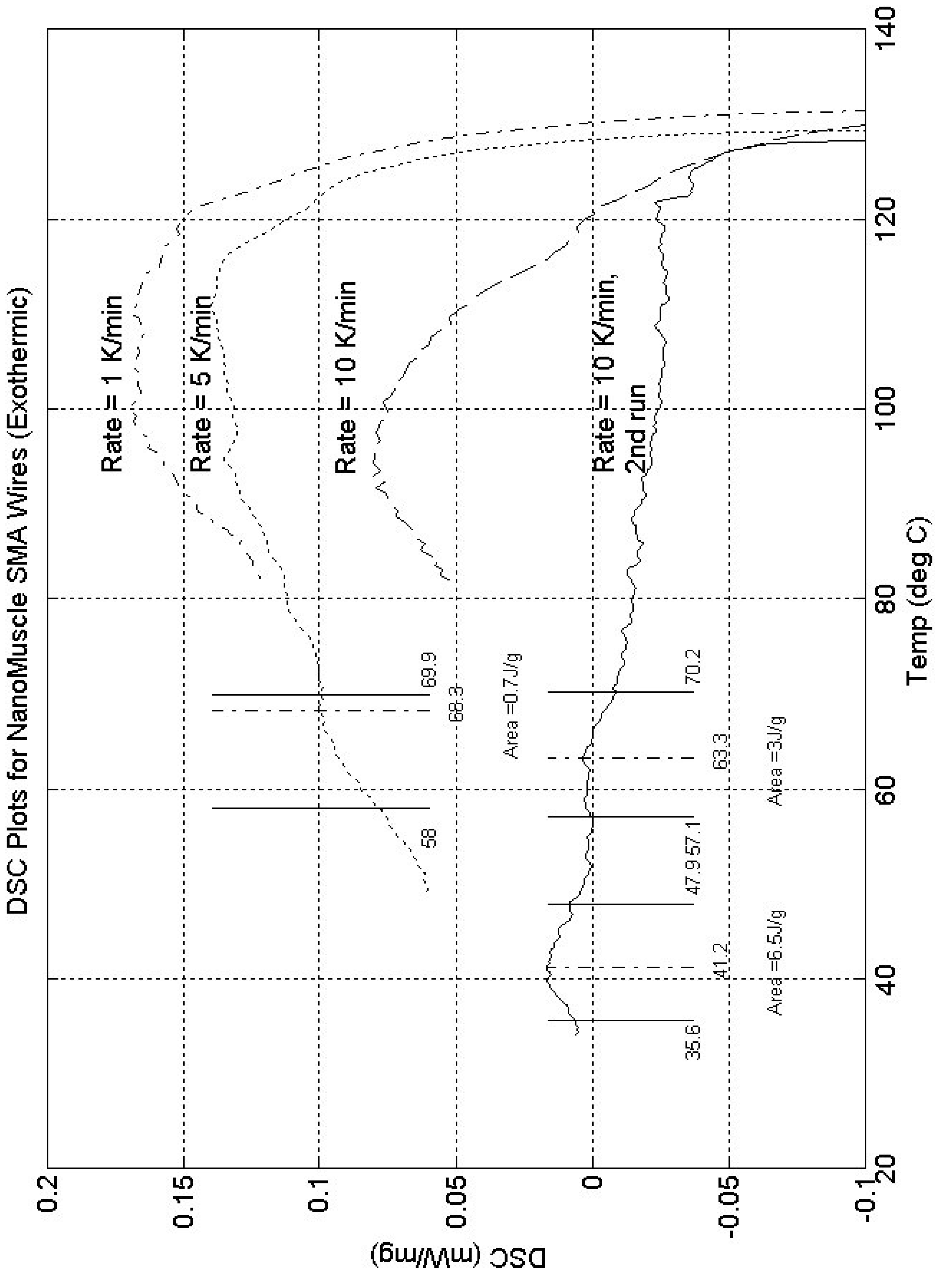
[39] G.C. Goodwin, S.F. Graebe and M.E. Salgado. *Control System Design*. Prentice-Hall, New Jersey, 2001.

[40] L.M. Toews. The Development of a Monolithic Shape Memory Alloy Actuator. Master's thesis, University of Waterloo, Waterloo, Ontario, 2004.

Appendix A

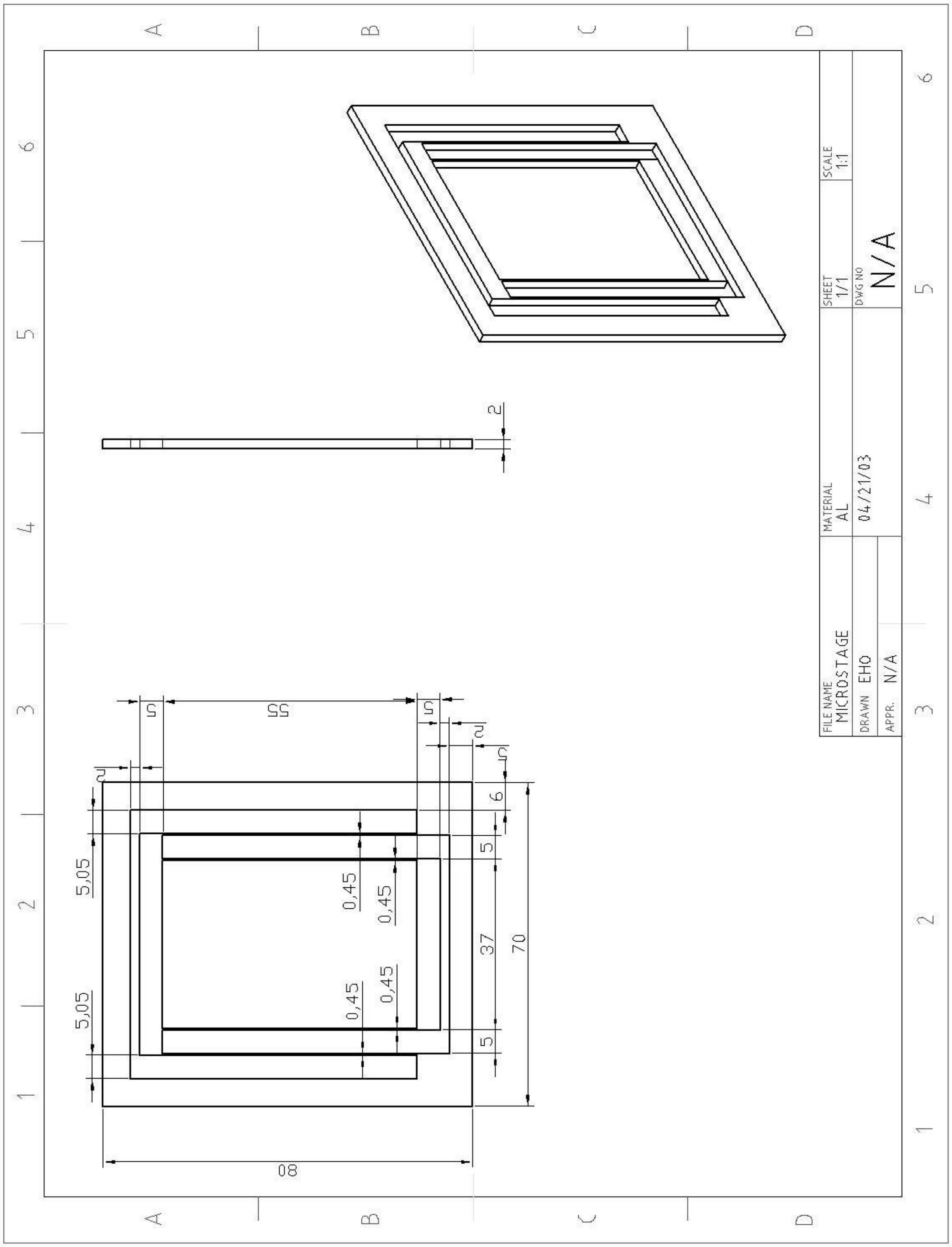
DSC Plots for NanoMuscle^{TM} SMA Wires


DSC Plots for NanoMuscle SMA Wires (Endothermic)
DSC (mW/mg)
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
Temp (deg C)
20
40
60
80
100
120
140
Rate = 1 K/min
Rate = 5 K/min
Rate = 10 K/min
Rate = 10 K/min, 2nd-run
62.8
75
78.1
Area =10.5J/g
71.3
75.2
78.4
Area =7J/g
66.3
75.5
82
Area =8.1J/g
66.3
75.1
80.8
Area =9.5J/g


DSC Plots for NanoMuscle SMA Wires (Exothermic)
DSC (mW/mg)
Temp (deg C)
0.2
0.15
0.1
0.05
0
-0.05
-0.1
20
40
60
80
100
120
140
Rate = 1 K/min
Rate = 5 K/min
Rate = 10 K/min
Rate = 10 K/min, 2nd run
69.9
68.3
58
Area =0.7J/g
70.2
63.3
47.9
57.1
41.2
35.6
Area =3J/g
Area =6.5J/g

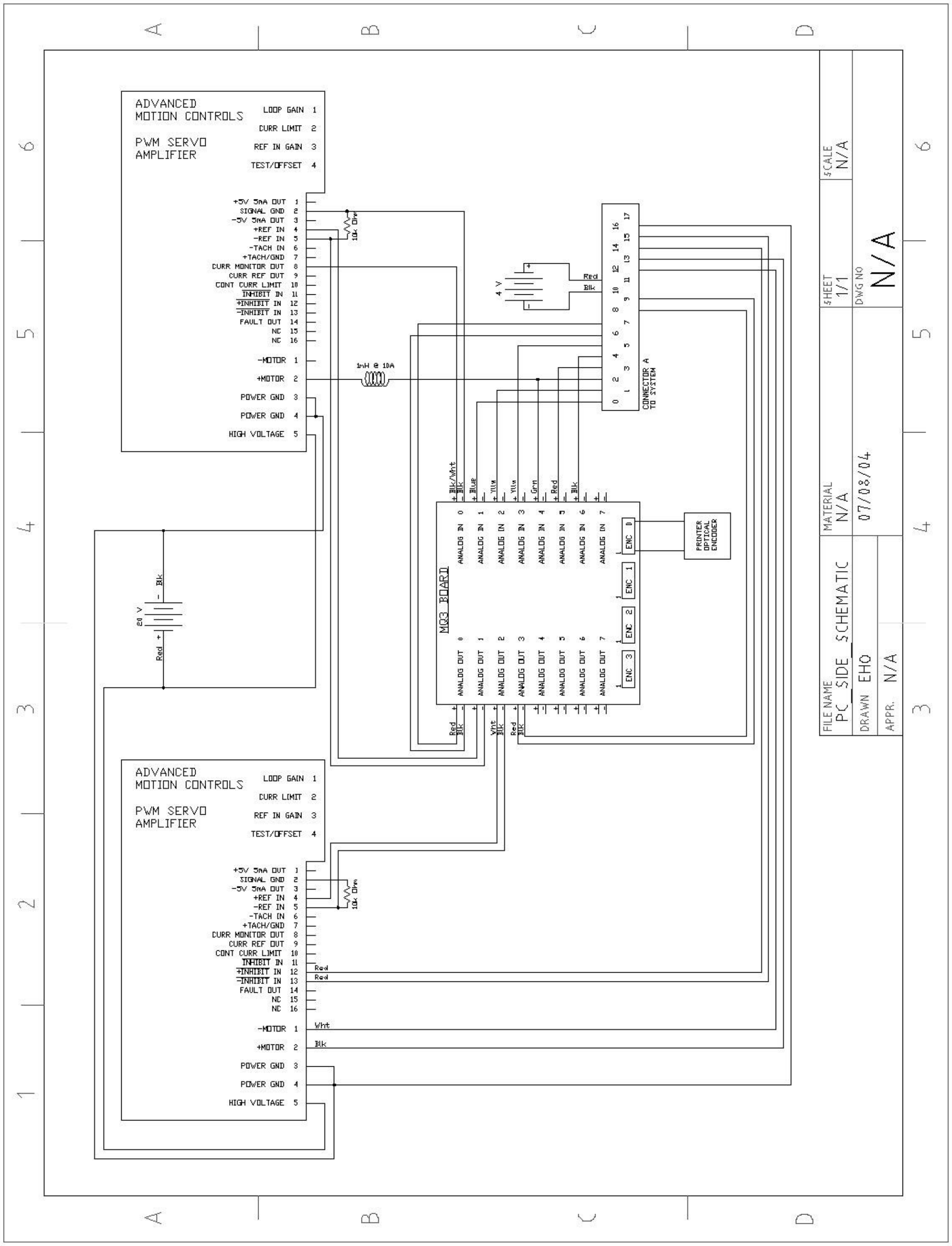
Appendix B

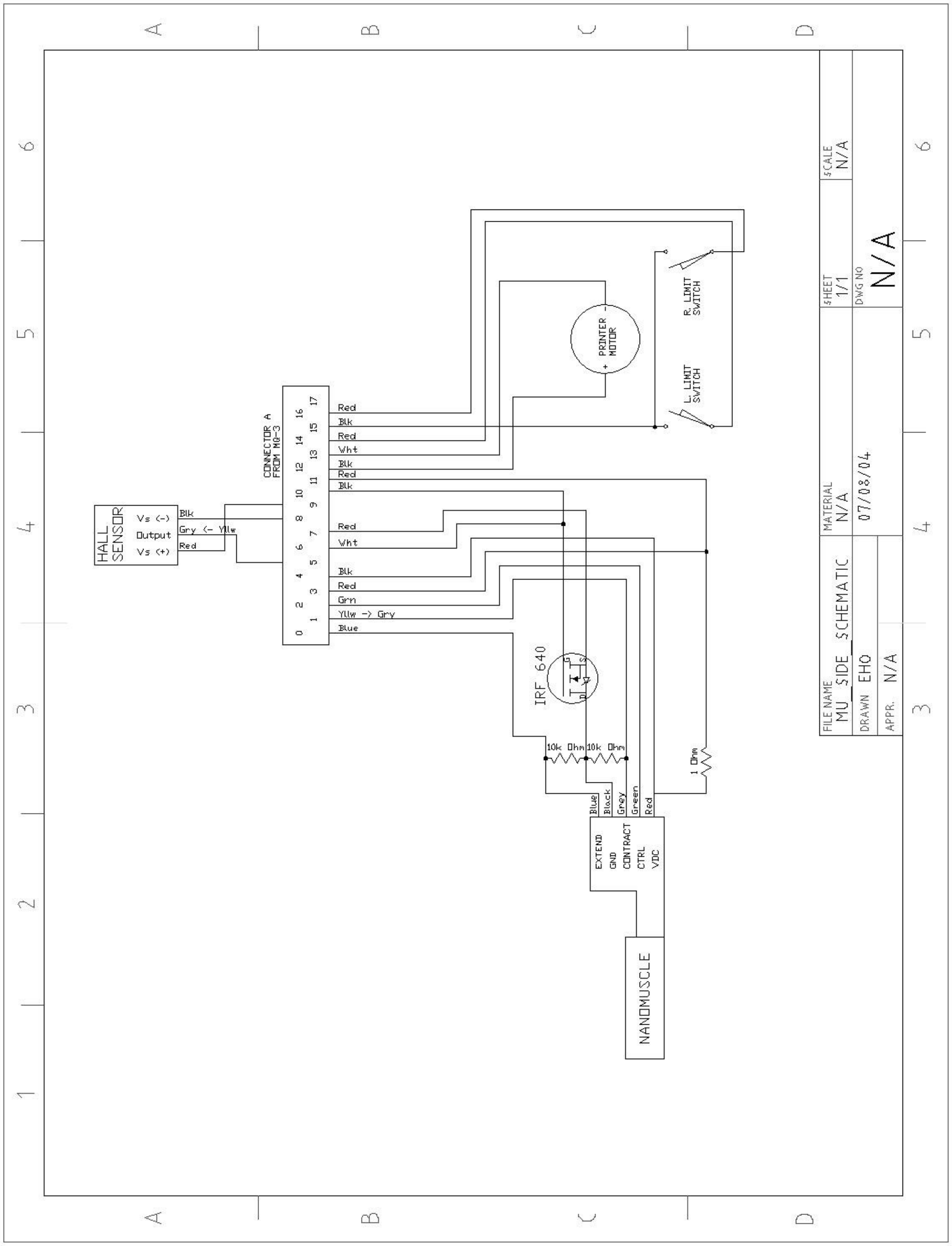
CAD Drawing of Flexure Guide


FILE NAME
MICROSTAGE
DRAWN EHO
APPR. N/A
MATERIAL
AL
04/21/03
SHEET
1/1
DWG NO
N/A
SCALE
1:1
80
70
55
37
5
5
5
5
2
2
2
6
5,05
5,05
0,45
0,45
0,45
0,45
A
B
C
D
1
2
3
4
5
6

Appendix C

Electrical Schematic of Macro-Micro System


ADVANCED MOTION CONTROLS
PWM SERVO AMPLIFIER
LOOP GAIN 1
CURR LIMIT 2
REF IN GAIN 3
TEST/OFFSET 4
+5V 5mA OUT 1
SIGNAL GND 2
-5V 5mA OUT 3
+REF IN 4
-REF IN 5
-TACH IN 6
+TACH/GND 7
CURR MONITOR OUT 8
CURR REF OUT 9
CONT CURR LIMIT 10
INHIBIT IN 11
+INHIBIT IN 12
-INHIBIT IN 13
FAULT OUT 14
NC 15
NC 16
-MOTOR 1
+MOTOR 2
POWER GND 3
POWER GND 4
HIGH VOLTAGE 5
10k Ohm
1mH @ 10A
4 V
Red
Blk
CONNECTOR A TO SYSTEM
28 V
MQ3 BOARD
ANALOG IN 0
ANALOG IN 1
ANALOG IN 2
ANALOG IN 3
ANALOG IN 4
ANALOG IN 5
ANALOG IN 6
ANALOG IN 7
ANALOG OUT 0
ANALOG OUT 1
ANALOG OUT 2
ANALOG OUT 3
ANALOG OUT 4
ANALOG OUT 5
ANALOG OUT 6
ANALOG OUT 7
ENC 0
ENC 1
ENC 2
ENC 3
PRINTER OPTICAL ENCODER
Blk/Wht
Blue
Yllw
Grn
Wht
SCALE N/A
SHEET 1/1
DWG NO N/A
MATERIAL N/A
07/08/04
FILE NAME PC_SIDE_SCHEMATIC
DRAWN EHO
APPR. N/A
A
B
C
D
1
2
3
4
5
6


A
B
C
D
1
2
3
4
5
6
HALL SENSOR
Vs (-)
Output
Vs (+)
Blk
Gry <- Yllw
Red
CONNECTOR A FROM MQ-3
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
Red
Blk
Red
Wht
Blk
Red
Blk
Red
Wht
Blk
Red
Grn
Yllw -> Gry
Blue
+ PRINTER MOTOR -
L. LIMIT SWITCH
R. LIMIT SWITCH
IRF 640
G
D
S
10k Ohm
10k Ohm
1 Ohm
Blue
Black
Grey
Green
Red
EXTEND
GND
CONTRACT
CTRL
VDC
NANOMUSCLE
FILE NAME
MU_SIDE SCHEMATIC
DRAWN EHO
APPR. N/A
MATERIAL
N/A
07/08/04
SHEET
1/1
DWG NO
N/A
SCALE
N/A

Appendix D

CanSmart Conference Paper

Potential of the NanoMuscle SMA Actuator for Fine Position Control

Eric Ho
emtho@engmail.uwaterloo.ca

Robert B. Gorbet
rbgorbet@uwaterloo.ca

Department of Electrical & Computer Engineering
University of Waterloo
Waterloo, ON N2L 3G1

ABSTRACT

The use of high-precision automated equipment is steadily increasing due in part to the progressively smaller sizes of electronic circuits. Currently, piezoelectric transducers (PZTs) dominate as the actuation device for high precision machines, but shape memory alloys (SMA) may be a viable alternative to reduce monetary costs.

The research we are involved in explores the implementation of a low-cost linear macro-micro positioning system. The system consists of a modified printer carriage to provide long range, macro scale linear motion (approx. 200mm range and 200µm precision) and a micro scale system (approx. 4mm range and 5µm target precision) that uses an SMA actuator. In this talk, a commercial SMA actuator, the Nanomuscle, is studied for practical use. The presentation will include actuator characterization results such as cycle times, degradation of performance over repeated cycling, hysteresis modeling, and a preliminary Preisach model. Also presented are the test bed used to obtain the results as well as a brief overview of the design for the micro positioning system.

INTRODUCTION

Motivation

Due to the decreasing sizes of electronic circuits, the demand for high-precision automated equipment has been steadily increasing. Piezoelectric transducers (PZTs) currently dominate as the actuation device for such high precision machines. PZTs provide high forces but very low stroke (in the range of microns for stack actuators[e.g., 4]); thus the range of motion for PZT-based precision machines is very low unless a form of stroke amplification is used. This increases the requirement for high precision parts and thus monetary cost. Shape memory alloys (SMA) may offer a viable alternative to PZTs for some high precision positioning applications. SMAs provide larger strokes possibly eliminating any need for stroke amplification. The main disadvantage of SMAs compared to PZTs is reduced bandwidth. However, this may be offset by the potential cost reduction SMAs offer, making this configuration suitable for certain applications.

Goal

The overall goal of this project is to design and build a low-cost linear macro-micro positioning system. The proof-of-concept prototype is shown in Fig. 1.

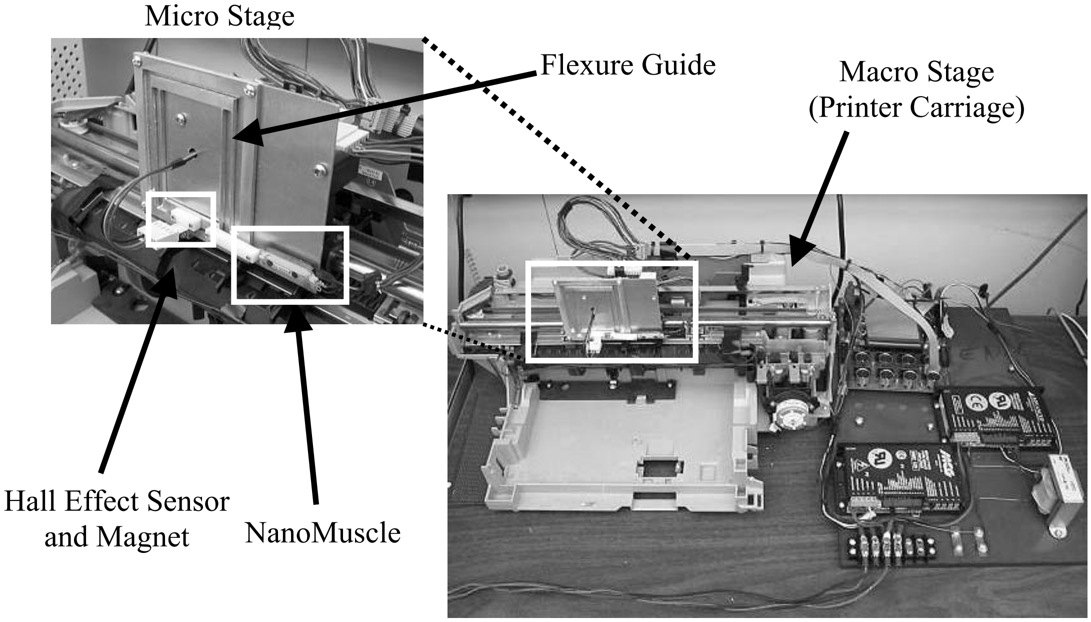


Fig. 1 – Final Setup of Macro-Micro Positioning System

The system utilizes a modified printer carriage for macro scale motion. The carriage provides 200mm of linear motion, and an optical encoder and encoder strip with a resolution of 200μm provides positional feedback.

The micro scale system consists of a linear flexure guide to create linear motion as well as a bias force, and a NanoMuscle NM70 linear actuator as the driving force. Position feedback is provided through two magnets (Micro Switch 103MG5) and a Hall effect sensor (Honeywell SS94B1).

The entire system is monitored and controlled through Simulink via a MultiQ-3 data acquisition board (Quanser Consulting MQ3). The A/D resolution of the MQ3 and the Hall effect sensor and magnet properties give a micro stage linear measurement resolution of approximately 5μm. A block diagram of the system is illustrated in Fig. 2.

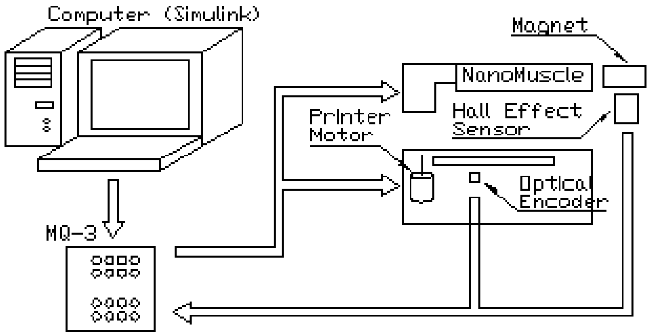


Fig. 2. –Block Diagram of Overall System

Since the SMA actuator plays a significant role in the system, it must be carefully studied for use. Thus, the goal of this paper is to evaluate the NanoMuscle SMA actuator for use in the proposed system.

NanoMuscle Description

The SMA actuator to be used is the NanoMuscle NM70, a commercially available linear actuator that utilizes a series of six NiTi SMA wires in series as the physical drivers. A detailed illustration of the NanoMuscle is shown in Fig. 3. The actuator's 4mm stroke makes it a good physical match for the macro stage in this fine positioning application. In addition, the actuator is commercially available, and has a fully integrated digital interface, simplifying implementation. It is electrically powered by a five volt supply. Current to the SMA wires is linearly regulated through internal circuitry by a zero-five volt control signal. [3]

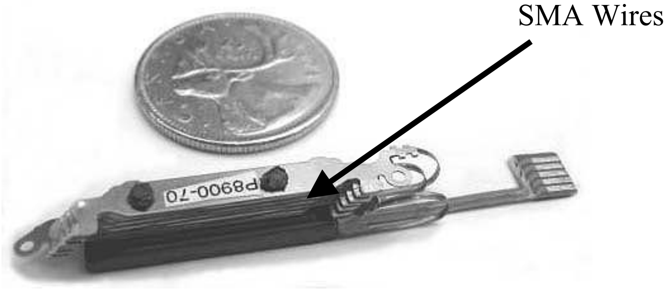


Fig. 3 – NanoMuscle Linear Actuator

Since the actuator uses SMA as the driving mechanism, non-linear behaviour is inherent. The following sections look at the identification of a Preisach model for the actuator hysteresis, as well as closed loop testing of extension and contraction times over repeated cycling and steady state characteristics in closed loop control.

HYSTERESIS MODELLING

The first analysis performed is to generate a model representing the inherent hysteresis of the actuator. The Preisach model is adopted and numerically applied using experimental data gathered through open loop heating of the actuator. The model is validated through comparisons of simulated and actual results from experimental tests.

Preisach Model

One major aspect of SMAs that make their position control difficult is hysteresis. This non-linearity plays an important role in the simulation and analysis of its behaviour for controller design, thus it is ideal to have a mathematical model of this non-linear behaviour. Particularly in position control, it is important that the model reproduce minor loop behaviour. The Preisach hysteresis model[1] is one such model, which has been used in the past for SMA hysteresis.

In this model, hysteresis is modeled as a weighted sum of relays in parallel acting on a single input $u(t)$. Each relay, $\gamma_{\alpha\beta}$, is assigned different forward and reverse actuation thresholds (α and β respectively) with corresponding 'on' and 'off' output values of 1 and –1 respectively. The relays are then weighted according to a weighting function $\mu(\alpha, \beta)$ and summed. Thus the output can be mathematically summarized through equation (1).

$$y(t) = \iint \mu(\alpha, \beta) \gamma_{\alpha\beta} u(t) \, d\alpha \, d\beta \tag{1}$$

Mayergoyz[1] has shown that necessary and sufficient conditions for the existence of a Preisach model are that the system satisfy the *Wiping Out* and *Congruent Minor Loop* properties. It has been previously demonstrated, at least in a frequency range of interest, that SMA does display this behaviour [e.g., 5].

Experimental Implementation of Preisach Model

One method of simulating the Preisach model is to evaluate the Preisach weighting function $\mu(\alpha, \beta)$ by curve-fitting to measured identification data and numerically integrate equation (1) [2]. However, this method is inherently imprecise and lacks robustness. If the only goal is simulation, and not analysis, another implementation exists, which does not require the analytical form of $\mu(\alpha, \beta)$. Using the so-called *numerical Preisach model*[1], the output is computed using only identified data and a finite summation (2):

$$f(t) = -f^{+} + \sum_{k=1}^{n-1}\left(f_{M_k m_k} - f_{M_k m_{k-1}}\right) + f_{M_n u(t)} - f_{M_n m_{n-1}} \tag{2}$$

In (2), f is the model output, f^{+} is the positive saturation value of output, and f_{Mkmk} is the measured output at input maximum M_k and minimum m_k.

In order to implement equation (2), the values of f_{Mkmk} must be found through identification. This process involves the measurement of data on “first-order descending” (FOD) curves which are descending branches of the hysteresis curve. They are generated by starting the actuator temperature below the point of negative saturation, heating to a temperature M_k below positive saturation, and then cooling the actuator while measuring the output at discrete points m_k over the temperature range. The measured outputs correspond to specific values of f_{Mkmk}.

For further details on the Preisach model, its identification, and its numerical implementation, the reader is referred to [1].

NanoMuscle Preisach Model Identification

To identify the data required for the NanoMuscle Preisach model, a series of FOD curves were generated to develop a look up table of the output at specific maximum and minimum input values. The curves were generated using a programmable kiln to heat the actuator, thus providing direct control of the temperature; the setup of the kiln is shown in Fig. 4. Current controlled by a solid-state relay generates heating through a resistive heating element in the kiln. The temperature of the NanoMuscle is measured using a thermocouple fixed the top plate of the actuator. The actuator is mass-biased, and displacement is measured using an optical encoder, with a linear resolution of approximately 14μm. The temperature is monitored and the kiln contactor controlled by a microcontroller, via a LabView interface on a PC.

Using the kiln testbed, the NanoMuscle was repeatedly heated to different maximum temperatures M_k, then cooled back down to ambient temperature, generating FOD curves. With each heating cycle, M_k was decreased. As the actuator cooled, measurements were taken at decreasing values of temperature, m_k, and assigned to FOD data point f_{Mkmk}. Examples of the FOD curves are shown in Fig. 5. Fig. 6 shows a plot of the FOD data.

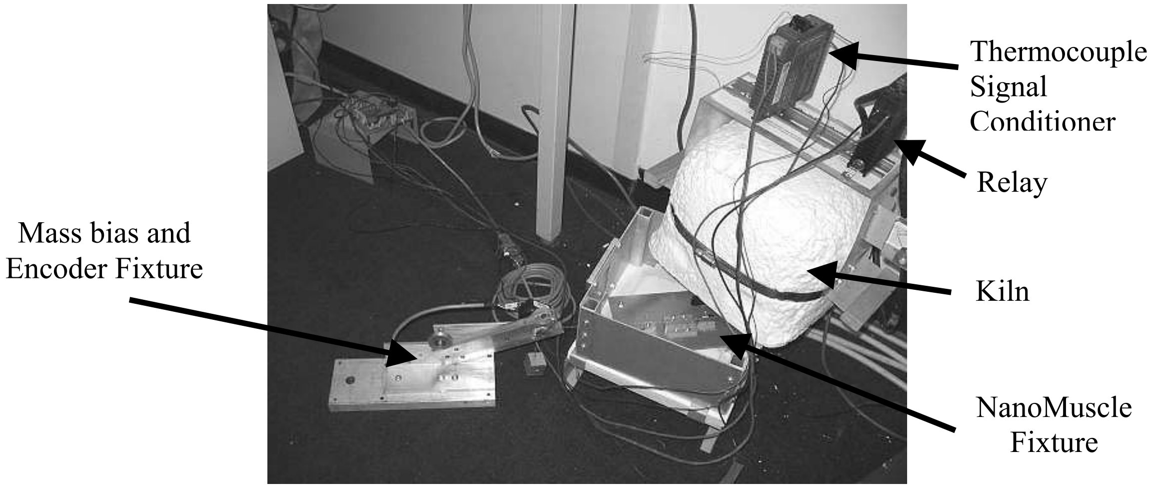


Fig. 4 – Experimental Setup using a Kiln

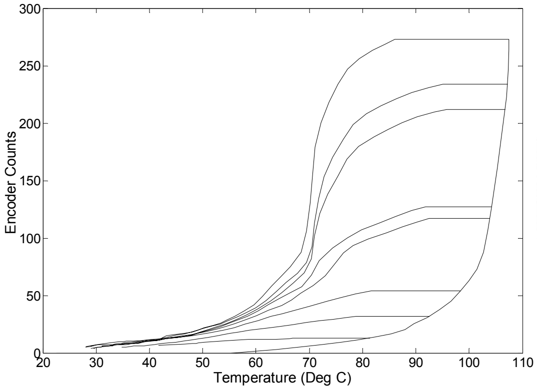


Fig. 5 – FOD Curves

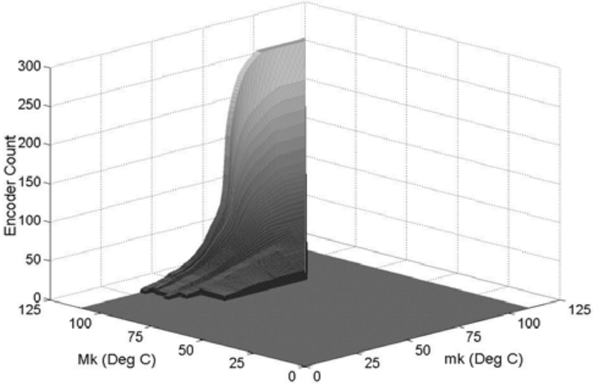


Fig. 6 – Plot of FOD Data

Simulation Results

To test the validity of the model, equation (2) was implemented in Matlab along with the FOD data of Fig. 6, to create a simulation of the system. A new series of temperature profiles from data obtained from the kiln are entered to the simulator and compared with actual results. The comparison of the resulting curves with the actual data is shown in Fig. 7; the match is quite good.

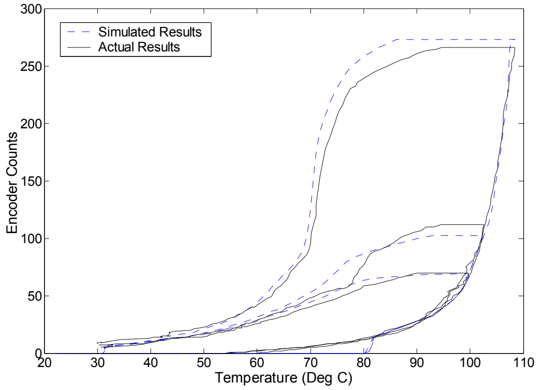


Fig. 7 – Comparison of Simulated and Actual Results

ACTUATOR CYCLE TESTING

After achieving a model of the actuator using direct heating in open loop control, the system was tested under electrical operation and closed loop control, to provide some general characterizations of the actuator.

Experimental setup for Cycle Testing

To test actuator characteristics, an experimental setup was built consisting of a NanoMuscle in series with an optical encoder and a bias spring; a picture of the system is shown in Fig. 8. The optical encoder and strip provide a resolution of approximately 17μm. The system is controlled and monitored through Simulink via the MQ3.

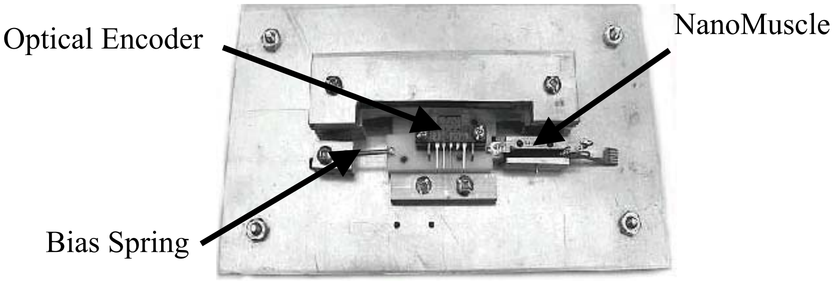


Fig. 8 – Experimental Setup Used for Cycle Testing

Cycle Testing

The first experiments performed were to study the changes in actuator characteristics over repeated cycling under closed loop control. Using an optical encoder as a feedback sensor, a PID controller where K_P=0.5, K_D=0.1 and K_I=0.5 was implemented. A periodic pulse train reference signal with a peak value of 150 encoder steps (approximately 2.55 mm), a 4s period and a 50% duty cycle was applied to the system. The change in actuator response over each cycle was analyzed with a sample-time of 0.01s. What should be noted is that the output is dependent on the input period. Shorter periods provide less cooling time, thus giving rise to shorter contraction and longer extension times. A period of 4s was found to allow repeated full extension and contraction cycles.

The first characteristics studied are the rise (contraction) and fall (extension) times in closed loop control; plots of these times over repeated cycling are shown in Fig. 9 - 12. Each figure consists of 11 runs shown as light lines, and the mean shown as the bold solid line. Each run was performed with an approximate half hour rest between each trial and an ambient temperature of approximately 23°. Two definitions of rise-time are examined; the first, τ_1, is the time elapsed from when the step input signal begins to when the encoder counts reaches 149, shown in Fig. 9. The second, τ_2, is defined as the time for the encoder counts to rise from 0 to 149 shown in Fig. 10. The difference τ_1-τ_2 is a measure of the delay in actuator response. That this difference is non-zero confirms that the actuator is cooling fully to the martensite phase in each cycle. The rise-time of the actuator decreases over cycling and stabilizes at about 0.23s and 0.18s for τ_1 and τ_2, respectively. This can be attributed to the base temperature of the actuator increasing and stabilizing higher than at cold start (yet still below the phase transition temperature).

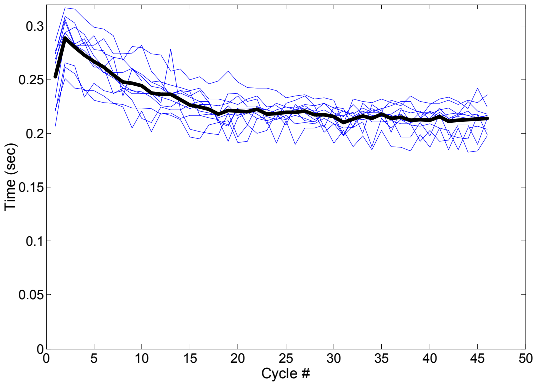


Fig. 9 – Plot of τ_1 vs. cycle

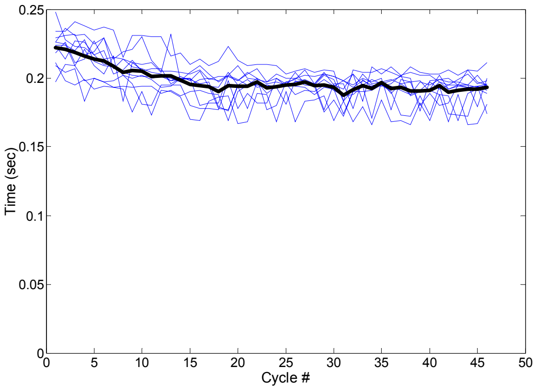


Fig. 10 – Plot of τ_2 vs. cycle

The fall-time, shown in Fig. 11, is calculated as the time required for the encoder counts to drop from 150 to 0. The fall-time increases and appears to flatten out at approximately 1.2s; this also is attributed to warmer actuator temperatures and thus longer cooling times. The total time for the average contraction and extension times is shown in Fig. 12. This is a measure of the actuator bandwidth at the target displacement.

It appears from Fig. 9-12, that there may be a trend from one run to the next (shown in light lines) in addition to the cycle-to-cycle behaviour described. However, data from the different runs cross and overlap; there is no discernable trend from run to run.

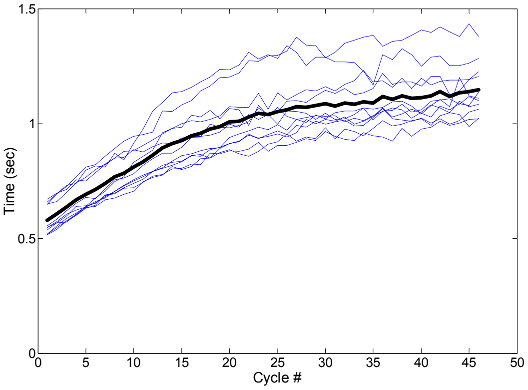


Fig. 11 –Plot of fall time vs. cycle

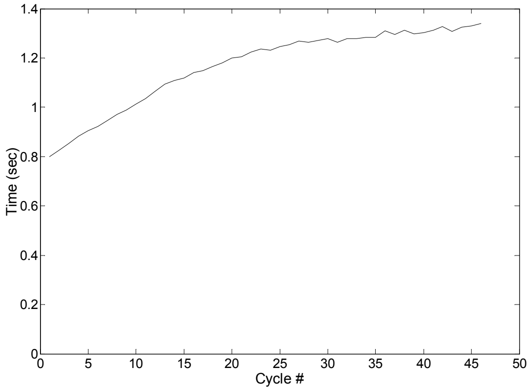


Fig. 12 – Plot of total time vs. cycle

Another important characteristic that is studied is the efficacy of a simple controller on the actuator over repeated cycling. One characteristic analyzed is the ability of the controller to maintain a steady state position. Shown in Fig. 13 and 14 are the percentage of 1000 sample points that read 149, 150, 151 and 152 encoder steps. The figures show ratios for the first 1000 points after the output first reaches the setpoint (Fig. 13), and for the last 1000 points before the output drops in response to the setpoint change (Fig. 14). Each curve is an average of 11 runs.

As can be seen, the number of points that read 150 encoder counts increases over time in both cases but is slightly higher in the second half of the peak. This gives an indication that the settling time is reduced, and the ability of the control system to maintain the setpoint is increased as the actuator warms up.

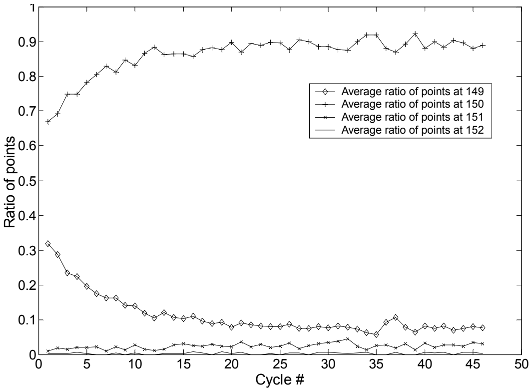


Fig. 13 – Ratio for first 1000 points

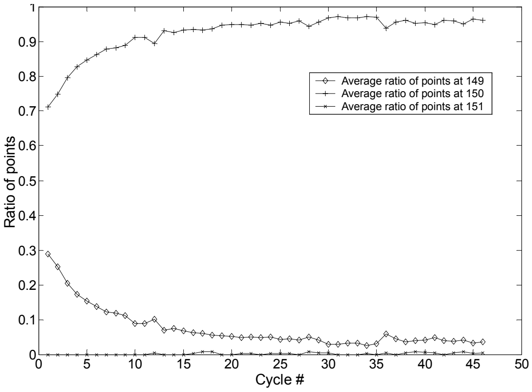


Fig. 14 – Ratio for last 1000 points

CONCLUSIONS

The overall goal of the project is to build a low-cost linear macro-micro system for fine positioning. The use of a commercial SMA actuator, the NanoMuscle, is proposed for the micro positioning stage and is studied here. Using the numerical Preisach model, a model of the actuator was developed using a series of first order descending curves experimentally generated by directly heating and cooling the actuator with a kiln. The model was then verified with a series of curves, different from those used for modelling. The model was found to provide a good reflection of the actuator behaviour.

Also performed were a series of closed loop tests on the actuator to assess its characteristics under normal operation. As expected, the actuator characteristics change over repeated cycling due to increases in temperature; the contraction time decreases and extension time increases as the actuator warms up. Also, the ability of a simple PID controller to maintain a steady state position increases after repeated cycling.

Future development will involve the use of the generated Preisach model for controller design for the SMA actuator, and its integration into the final system. In addition, the appropriate actuation duty cycles and periods for stable, repeatable behaviour will be assessed.

REFERENCES

[1] I.D. Mayergoyz, “Mathematical Models of Hysteresis”, Springer-Verlag, 1991.
[2] R.B. Gorbet, D.W.L. Wang, K.A. Morris, “Preisach Model Identification of a Two-Wire SMA Actuator”, IEEE International Conference on Robotics & Automation, pages 2161-2167, May 1998.
[3] NanoMuscle Inc., “70 Gram HS/HE Linear Actuator Manual”, NanoMuscle.
[4] Sensor Technology Limited, http://www.sensortech.ca (current Aug 21, 2003)
[5] D. Hughes, J.T. Wen, “Preisach modeling of piezoceramic and shape memory alloy hysteresis”, IEEE Conference on Control Applications, Albany, New York, 1995.