DESIGN AND DEVELOPMENT OF A HIGH FORCE ACTUATOR USING GIANT MAGNETOSTRICTIVE MATERIAL

By

Kam Fung

B. A. Sc. (Engineering Physics) University of British Columbia, 1992

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

ELECTRICAL ENGINEERING

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

1995

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

Department of Electrical Engineering

The University of British Columbia
Vancouver, Canada

Date Apr. 28/95

DE-6 (2/88)

Abstract

With the maturity of manufacturing process for the giant magnetostrictive material called Terfenol-D, high force and high precision actuators can now be realized at a reasonable cost. These actuators can be used in many applications, one of them being the control of machine chatter vibration. This violent phenomenon is one of the limiting factors on how fast a metal-cutting operation can proceed. In chapter 1 a discussion of chatter vibration is presented. Terfenol-D possesses some unique properties that affect actuator design; therefore an understanding of the magnetoelastic mechanism is essential. A basic discussion of magnetic material is first presented in appendix A; then properties unique to Terfenol-D are discussed in chapter 2. The design of the actuator at UBC is presented in chapter 3. It incorporates many features not found in existing designs. In particular, the push-pull configuration makes it suitable for applications requiring bi-directional movements. Preliminary test results of the device built are presented in chapter 4. Before Terfenol-D can be widely used, however, the problem of modelling and control has to be solved. Because of hysteresis and other nonlinearities, ordinary linear design techniques may not be applicable. Some promising nonlinear control strategies based on Preisach model and neural networks are discussed in the last chapter.

Table of Contents

Acknowledgement

This thesis could not have been completed if not for the kind assistance many people have provided.

Dave Fletcher, the machinist, have always been marvelous and patient when machining various parts from my ever-changing mechanical sketches. He, together with Leiff Kjolby, Donald Dawson, Chris Sheffield, Lloyd Welder, Tony Leugner and Alan Prince, reduced the complexity of this project to a minimum. My colleagues, Chia-tung Chen, Joseph Yan, Shyan Ku, Niall Parker, Ray Burge, Alison Taylor and Grey Grudic are not only fun to work with, but also provide much support and fuel for thought during these couple of long years. I would also like to thank Dr. Tim Salcudean and Dr. Peter Lawrence for their time; I learn a lot from them. I am indebted to my supervisor, Dr. Chris Ma. Not only did he guide me along during the course of this thesis, but also taught me how to become a better professional, and a better person overall.

The most gratitude I owe is to my parents, whom I don't think I can ever thank enough. Their support and love have always been boundless. Through all those long nights and times of anxiety, they stood by me and do whatever they can to help me focus on pursuing my dream. I dedicate this work to them.

Support of this thesis by an NSERC Graduate Scholarship is gratefully acknowledged.

Chapter 1

Introduction

1.1 Machine Chatter Vibrations

Chatter is a violent, relative vibration between the cutting tool and the workpiece. If not effectively dealt with, it will adversely affect work surface finish, machining accuracy and tool life. The most popular method to control machine chatter vibration is to lower production rate [1], by reducing either the feed-rate, or the depth of cut.

Fundamentally, chatter is caused by the feedback of dynamic cutting forces exciting the mechanical structure formed by the tool, tool frame, workpiece and drive. The most important characteristic of chatter is that it is self-induced and draws its energy from the cutting process itself. Chatter can occur on almost all metal cutting machines, including lathes and boring bars, milling and drilling machines, grinding machines, planing machines, surface broaching machines, and etc, since all machine operations are governed by the same general rules of self-excited vibration. Some machines, however, do exhibit individual characteristics within the general framework.

The main concerns of chatter include safety and quality of finish: the relative vibration can cause the tool to shatter, thus posting a threat to the operator, and it will also leave marks on the workpiece, although this latter problem might be tolerable during rough cutting. Consider the turning operation shown in figure 1.1. When the tool contacts a hard spot or some irregularity in the work surface (point O), it will bounce and vibrate relative to the workpiece, with frequency of vibration corresponding to one of the resonant

frequencies of the machine tool structure. After one revolution, the oscillation of the tool will result in a wavy work-surface. When the tool returns to this original cutting position and takes another cut (point A), it will contact the wave upon the work surface left by the preceding cut (if the tool traverse is less than the tool width, as is normal). The work surface wave will cause a dynamic variation of the steady cutting force. The cutting tool will again vibrate at the resonant frequency corresponding to the work surface waveform if the cutting speed has not been altered between the successive cuts. This will in turn oscillate the structure. When this regenerating dynamic force is just enough to sustain itself, the machine is said to be on the threshold of stability [1, page 51].

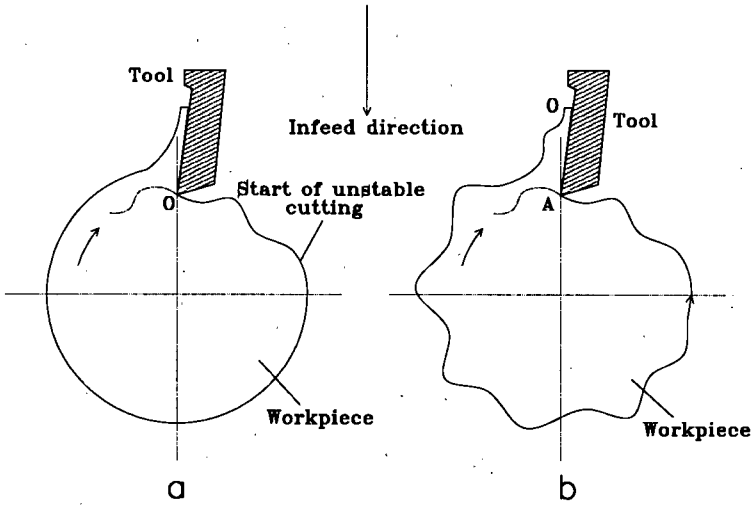


Figure 1.1: Unstable cutting on lathe or boring bar

To determine the stability of a particular machine, a Stability Chart can be constructed [2, page 35-47]. For a lathe or a boring bar, a typical chart is shown in figure 1.2.

The general shape of the chart is derived using a simplified model of the metal cutting process. The exact shape of the chart, however, depends on many factors, including the

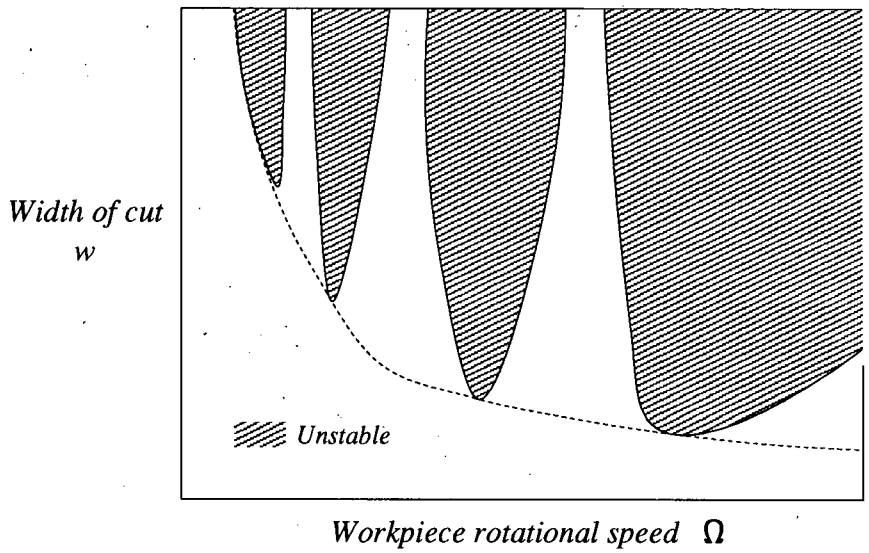


Figure 1.2: Stability chart corresponding to a particular lathe of boring bar

type of operation, hardness of the workpiece, overlap factor[1], and structural rigidity of the machine frame. From this chart, one can predict roughly when chatter will occur and could stay clear of those regions. However, this usually results in compromising of the production rate.

1.2 Controlling Chatter

Good machine design practice can reduce machine chatter vibrations. Using small widths of cut (ie. narrow tools), using large rake-angle tools and small depths of cut, cutting softer materials, and making the machine tool structure light, stiff and well damped will all help. But some of these will reduce the production rate, and others can only reduce, not eliminate, the chance of chatter from happening[2]. Therefore, it is desired to develop

[1]This factor, μ, is a measure of how much of the work surface has been worked on during the previous revolution. $\mu = 1$ represents completely regenerative cutting, whereas $\mu = 0$ means non-regenerative cutting.

[2]For a more detailed treatment of proper machine design for reduction of chatter, see [1]

a more reliable strategy to tackle this problem.

1.2.1 Indirect Approaches

In [3], Kasahara *et al.* investigate the behaviour of self-excited chatter in lathes from the space phase[3] point of view. They try to predict the amplitude of vibration and judging on that information, decide whether the resulting surface finish is acceptable. In addition, they suggest that varying the cutting speed might stabilize the operation. However, the approach taken is based mostly on experimental observations of the phase characteristics and behaviour of the vibration, without much consideration about the underlying dynamics of the structure. Therefore, it may not be easily applied to other types of operation.

In addition to changing the spindle speed, one can also control the feed velocity. Kanso [4] applies a model reference adaptive control strategy to an end milling machine, in an effort to minimize tool chatter and maintain stability of the whole system. First he derives a mathematical model to predict tool chatter, with parameters estimated on-line by an estimator. Then the adaptation algorithm generates the appropriate spindle speed and feed velocity that would deliver optimal machining conditions (represented by a reference model). Using on-line estimation, variations in the cutting process, which is very likely in practice, can be tolerated. However, the models chosen for the dynamics are linear ones; therefore it is doubtful that even the adaptation process can guarantee optimal performance given the various non-linearity of the cutting process (tool disengagement and feed-rate saturations for example). Furthermore, magnitude of the tool chatter after control actions have been applied is still in the order of fractions of a millimeter. The resulting surface finish would be quite rough and should not be acceptable for most applications. However, it should be pointed out that the exact machining conditions

[3]The space phase is define as the phase difference between outer and inner modulation.

are not given in the paper and no data regarding vibration amplitudes before and after application of control actions are given. Therefore, it is difficult to assess the overall effectiveness of the approach.

Suppressing chatter vibrations by controlling feed-rates can be viewed as indirect control. Since the vibration and the feed are orthogonal to each other, the control effort effects the muffling of chatter through cross-coupling in the machine structure (see figure 1.3): the feed-rate changes the amount of mechanical feedback induced in the

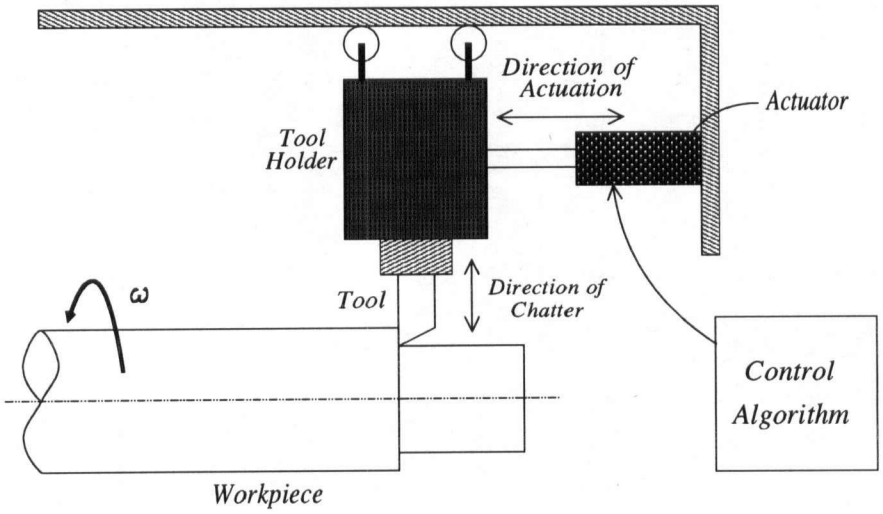


Figure 1.3: A typical control action: Not in the direction of vibration.

workpiece by varying the overlap factor μ, thus altering behaviour of the vibration. For this to work properly, the cross-coupling term has to remain relatively constant. However, temperature, properties of the lubricant, stiffness of the gear etc. can all affect this term. Therefore, it is reasonable to seek for a more direct approach.

1.2.2 Direct Approach

Towards this end, Ma and Altintas [5] develop an adaptive controller based on the sliding mode control theory for plunge turning (figure 1.4). They model the chatter dynamics

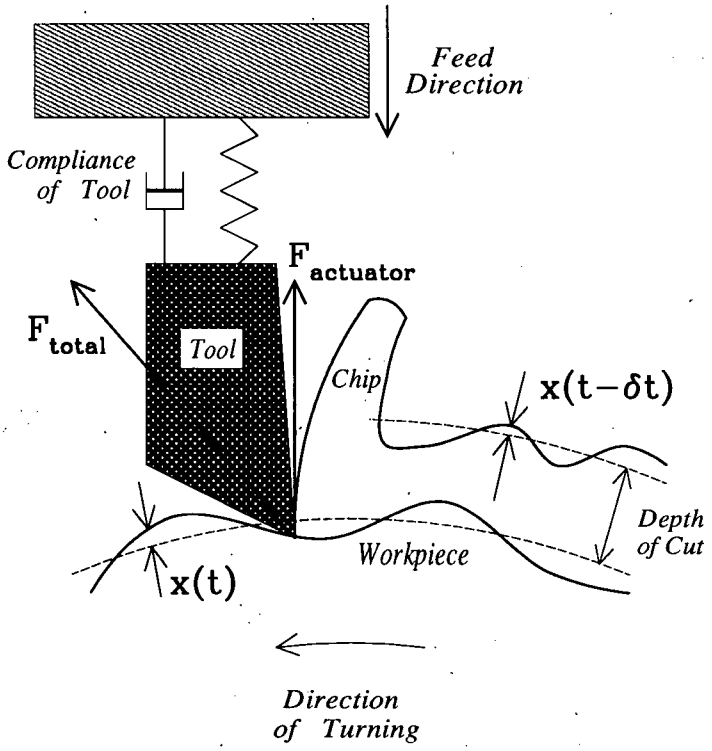


Figure 1.4: A plunge turning operation. The dash lines show the desired cutting surface, whereas the solid lines are those caused by tool chatter. Note that in this case, the feed direction is parallel to the possible direction of chatter vibration.

by a nonlinear process[4]. After simulating deflection of the tool under various cutting conditions, including abrupt changes in feed, they suggest that the proposed adaptive controller has promising potential for actual application in practice.

Verifying the results experimentally, however, necessitates an actuator with adequate performance; some requirements are listed below.

1. Able to generate high forces (in the order of a few thousand Newtons).
2. Able to response quickly. Since the frequency of chatter vibration can be quite high, it is desirable to have a mechanical bandwidth of at least several hundred hertz.

[4]Three main nonlinearities are considered: tool disengagement from the workpiece, feed motion being positive only, and saturation of the servo drive.

3. Good resolution. This figure is partially related to the achievable surface smoothness of the workpiece. No definite requirement, but sub-micron resolution is desired.

The actuator also has to be structurally rigid in order to survive the harsh environment encountered in metal cutting operations. Overall size is also a constraint, since space would be quite limited between the tool-carriage and the workpiece. The range of movement should be about the same as the expected amplitude of the chatter vibration, which is some fractions of a millimetre.

1.3 Actuator Technologies

Linear actuators come in many different flavours. Electromechanical is the more traditional type, while electrohydraulic actuators are becoming more popular in high power applications. Then there are piezoelectric materials and shape memory alloys. They can all generate movements in a linear fashion, and with reasonably high forces. The choice is dictated by the requirements of the application.

1.3.1 Electromechanical actuators

Electromechanical linear actuators can meet a wide variety of motion control needs. Using a DC or AC electric motor driving torque multiplying gear reduction, these units have leadscrew at the output to translate the circular motion into a linear one. Efficiency for the type with ball bearing leadscrew can be as high as 90 %. The maximum thrust depends on the gear reduction used, and can be up to a few thousand newtons for a typical 12-Vdc actuator with a 20:1 reduction ratios [6]. However, the response time and repeatability are not adequate for the application considered. Because the electrical motor has high inertia, the drive cannot response quickly enough to varying input command. In addition, end play between the nut and leadscrew, lead accuracy of the

leadscrew, backlash in drive mechanism, and end play of the leadscrew within its thrust system all prevent the actuators from achieving good accuracy and repeatability in both directions.

1.3.2 Electrohydraulic drive

Another type of actuator that can provide high force with increased precision is electrohydraulic drives. Using position feedback device, such as mechanically actuated linear variable differential transformers (LVDTs) for short stroke applications and magnetostrictive devices for medium and long stroke ones, these linear actuators can provide positioning accuracy to within 25.4 μm (0.001 in.) and speed up to 1.27 m/s (50 in./s) [7]. Furthermore, they offer advantages such as increased power density, good heat dissipation, simple mechanical design and self-lubrication. In fact, for high speed, long stroke applications, no other actuators can perform quite as well as electrohydraulic drives.

New electronically controlled proportional hydraulic valves[5], with addition of position sensing device, can have frequency responses to between 80 Hz and 90 Hz [9]. However, the actuator's natural frequency limits the control system's response to no more than the physical laws of motion would permit, and this frequency is dependent primarily on the mass being moved and the bore of the actuator. Other inherit limitations also make these actuators unsuitable for the controlling of chatter. First, because the valve will not react instantaneously to changes in signal from the controller, the output will exhibit time lag of at least a few milliseconds [9]. Second, a complete hydraulic drive system involves the installation of a pump, an accumulator, flow control devices, and all the connecting hoses. It would be expensive in terms of capital and space to require every metal cutting machine to have such a system installed. Therefore this option is not very attractive.

[5]A simpler type of flow control can be realized using two-position valves driven by pulse-width-modulation (PWM) signal. It provides reduced complexity and cost. See [8].

1.3.3 Shape memory alloys (SMA)

Shape memory effects are caused by a martensitic transformation between the austenitic and martensitic crystal structure [10]. When deforming a cold specimen some plate variants will be reoriented to cope with the external deformation. After unloading, the specimen will usually keep this new cold shape, and the original shape can be restored by heating the specimen. If the now heated sample is cooled again, thermal multivariant martensite is formed with no resultant shape change. This effect is called the one-way shape memory effect.

By several identical subsequent stress-free heating and mechanically-forced cooling, one can obtain a two-way shape memory effect. This procedure is called training of the material. Existing industrial or medical applications for SMA materials are now limited only to on/off type devices, like fire detector and clamp, although research efforts are directed towards some more continuously controllable actuators. Few practical designs are available.

Hirosi *et al.* [11] show that the efficiency of a SMA actuator is bounded by that of a Carnot cycle between the heating and cooling temperatures. Therefore it cannot compete with the classical drive types, but can excel in certain applications. When heated, SMA responds very quickly: they shrink, by returning to their original shape. Velocity of the contraction depends on the input power[6]. Maximum strain can reach about 3-4 % [10] and together with a stiffness in the order of 50 GN/m^2, a wire 1 mm in diameter can generate a force of about 1570 N. This is quite useful when the wires are put into a push-pull arrangement. On the other hand, however, the actuator is slow on cooling, especially when standing in free air. The rate of movement depends highly on the method of cooling, and is governed by thermal diffusion. In both heating and

[6]Hashimoto obtained, for example, for a TiNi wire of 1 mm diameter, length of 200 mm and no load, a velocity of 30 mm/s for an input power of about 200 W.

cooling, the thermal capacity of the material has to be considered, as well as the latent heat involved in the various phase transitions. Therefore even with forced liquid cooling, typical response time in the cooling phase of SMA actuators is still in the fractions of a second[7], making them too slow for chatter control applications.

1.3.4 Piezoelectric materials

Piezoelectric effect, in which certain crystals generate a charge when subjected to external forces and vise versa, was discovered in 1880 by Pierre Curie and his brother. In 1986, Uchino [13] presents a simplified model of piezoceramics that involves the notion of anions(-) and cations(+) connected by springs forming a crystal. These ions move in the applied E-field according to their polarity, thereby generating forces and cause the lattice to deform. Two kinds of deformations are observed. Piezostrictive materials have induced strain x roughly proportional to the applied electric field, E. Thus $x = dE$ (d is a material dependent tensor). Electrostrictive materials have induced strain proportional to the square of the applied field. Thus $x = mE^2$ (m is a material dependent tensor).

Piezoelectric materials have been used in many types of devices, including acoustic actuators, accelerometers, strain gages and force sensors. When used as an actuating material, piezoceramics require high drive voltages, and the range of movement is usually in the order of tens of parts per million (ppm) only. This situation has been alleviated somewhat with the development of layered lead zirconate titanate (PZT) devices. By sandwiching electrodes in between active materials, these devices can be operated with a much reduced voltage (less than 20 V is possible), and the maximum strain reported is up to about 150 ppm. Recently, there have been attempts to embed these actuators in smart structures for vibration control [14].

[7]Bergamasco *et al.* have developed an actuator using push-pull SMA coil springs with bandwidth of between 1 and 2 Hz. See [12].

1.3.5 Magnetostrictive materials

In 1840 James Joule discover that nickel contracts in the presence of magnetic field. Since then, many attempts have been made to adapt this and other magnetostrictive materials to commercial products. Micropositioners, low-frequency sonar transmitters and fast-acting switches and relays made with these materials are used quite extensively in the 1930s and 1940s [15]. However, they typically produce lower stains (λ) comparing to piezoceramics and electrostrictive materials. For example, nickel-based magnetostrictive alloys only produce strain in the order of 50ppm [16]. By the 1950s, piezoelectric ceramics has taken over magnetostrictive materials in most applications.

Then in the early 1970s, A. E. Clark and co-workers at the Naval Surface Weapons Centre discovered that when certain rare-earth metals, such as terbium (Tb), dysprosium (Dy), are combined with iron (Fe), the resulting alloys, $TbFe_2$[8] and $DyFe_2$ exhibit much higher magnetostrictive strains (specifically $TbFe_2$ at 3690 ppm and $DyFe_2$ at 1890 ppm, both at room temperature [17, Table 33]). Unlike Ni and Fe, $TbFe_2$ and $DyFe_2$ show positive magnetostrictive strains; that is, they expand in a magnetic field. Although both RFe_2 alloys exhibit the highest room temperature magnetostriction[9], the magnetic fields required to achieve such strains are also very high. Later, Clark succeeds in finding a ternary alloy $Tb_{0.3}Dy_{0.7}Fe_2$ with a comparable magnetostriction at room temperature, but with a much reduced magnetic field required for saturation. This new material, which he named Terfenol-D[10], causes the renew interest in applying magnetostrictive materials to many different areas.

One of the most promising applications of the highly magnetostrictive Terfenol-D alloy

[8] Or Terfenol, TERbium and FE alloy developed at NOL, Naval Ordnance Laboratory, where this alloy was discovered.

[9] Other rare-earth alloys, e.g.,RZn and RAl_2 compounds, also show outstanding magnetoelastic properties, but only at low temperature (T_c = 77 K for DyZn, with maximum magnetostrictive strain 5300 ppm @4.2 K) [17, page 305], and therefore are not useful for room temperature applications.

[10] Terfenol with Dysprosium added

is in micropositioning actuators. The high energy density of this material allows compact designs of direct drive positioning device for high precision, quick response and heavy load applications. Akuta makes a comparison with laminated PZT piezoelectric ceramics in [18] and the result (shown in figure 1.5) indicates that magnetostrictive materials have definite advantage when bulk expansion is considered. In addition, Terfenol-D requires

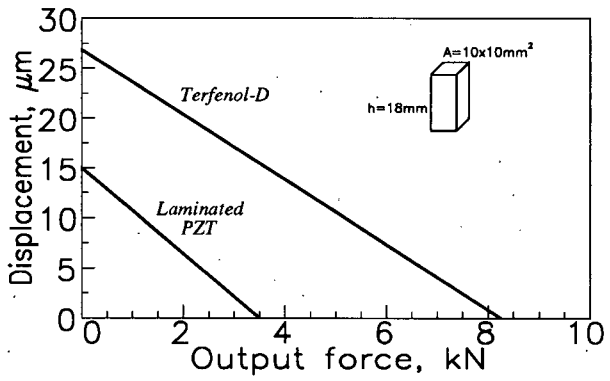


Figure 1.5: Displacement-Output force relation for two different actuators of the same dimensions a) Laminated PZT b) Terfenol-D, after Akuta [18].

only a low drive voltage to operate[11].

1.4 Choosing a Suitable Technology

By studying the requirements of the actuator for chatter vibration control (as stated in section 1.2.2), it is obvious that only two materials are suitable: PZT and Terfenol-D. However, the latter becomes more attractive after further considerations. Although piezoelectric materials might generate the necessary range of movement, its longer required length (due to the smaller absolute electroelastic strain) would violate the space specification. In addition, the reduced effective stiffness could affect the overall effectiveness of

[11]However, a higher drive voltage is desirable because it allows faster build-up of current.

the actuator. Therefore, the proposed device is based on Terfenol-D rods.

It should be noted that the price of Terfenol-D is relatively high. A rod 170 mm long and 21 mm in diameter cost about US$2,500. However, the price has been declining steadily over the past few years. Although probably will never become truly affordable, its unique properties still make this material very attractive for certain applications.

1.5 Controlling the Actuator

Although it is not the purpose of this thesis to find the best controller for the actuator developed, a few schemes are proposed. All magnetostrictive materials are inherently nonlinear and exhibit hysteresis. Therefore one should either keep the range of operation small and apply linear approximation, or seek for a better approach.

1.6 Thesis Overview

This thesis project work involves the design and fabrication of a high force actuator suitable for chatter vibration control, with Terfenol-D rods chosen as the active drive units. Some modelling techniques for this magnetostrictive material are described, and a few viable control frameworks are proposed.

The remaining of the thesis is organized as follows. Magnetostrictive properties of Terfenol-D are first discussed in chapter 2. Then the details of the actuator design are presented in chapter 3. Chapter 4 contains the proposed modelling and control strategies, and finally, conclusions and ideas for future works are presented in chapter 5.

Chapter 2

Magnetoelastic Properties of Terfenol-D

2.1 Magnetostriction

Magnetization induces strains, and magnetostriction is the change in shape and size of a body when its state of magnetization is changed. The state of magnetization can be changed by applying an external magnetic field, or by changing the temperature. While the magnetostriction can and does depend upon the orientation of the magnetization with respect to the crystal axes, symmetry forbids its depending upon the sign of the direction — a magnetization pointing north produces the same strain as one pointing south [17, prologue]. Magnetostriction is a general phenomenon which can be observed in any magnetic and deformable substance: it arises from the change of the internal energy and from the symmetry lowering experienced by a sample when it is magnetized [17, page 5]. More simply put, these dimensional changes result from the attraction or repulsion that individual dipoles of a magnetic persuasion has for each other in different configurations. In ferromagnetic materials, magnetostriction is a domain phenomenon in which magnetic field strength and mechanical strain are reciprocally related; in nonferrous materials, magnetic fields might produce strain, but reciprocity does not exist [19, page 81] (see appendix A for more detail).

2.2 The Magnetization of Soft Ferromagnets

Although Terfenol-D is a soft ferrimagnet, it behaves as a soft ferromagnet in terms of the stress dependence of the Joule magnetostriction (see section 2.3.2) and the field dependence of the elastic constants [17, page 48].

In ferromagnetic materials, magnetons[1] are associated in groups called domains. These are volumes of material supporting a net magnetomotive force. All the magnetons inside a domain are aligned in one direction, and a domain cannot exist by itself. Otherwise it would be required to support a magnetic field that extend beyond its own boundary, requiring energy stronger than what the exchange force can supply. If a number of domains form within a single crystal, however, they can mutually direct themselves in a manner that minimizes external fields, thus lowering the energy. Such an arrangement is temperature-sensitive because of the thermal effects on the relatively weak exchange forces between domains. At a temperature at which the thermal energy equals the exchange energy, the group action is destroyed, and the material becomes paramagnetic. For ferromagnetic materials, this happens at a temperature called the *Curie Temperature.*

2.2.1 Domain Walls

The boundaries separating domains are called Block Walls. These are not imaginary lines; rather, they are physical entities. The greater the angle between directions of magnetization for adjacent wall magnetons, the greater the force needed to maintain their wall against the self-aligning exchange forces. Actual sizes of the wall as well as the directions of the domains are a compromise, or balance, between the exchange forces, the forces that are required to maintain the angle between walls and the forces that minimize external fields.

[1] The term magneton is used to describe those atomic-sized, electrically charged particles that establish an associated magnetic field by spinning on their axes.

A ferromagnetic crystal is in a state of equilibrium in which all forces are in balance. Under these circumstances, there is no net lateral forces on the wall and a wall can be easily moved if it is exposed to an external force. For example, if an externally imposed force on a domain attempts to change its size, location, or direction, the walls will readily move to accommodate that change. After such movement, the wall-angle and exchange force balance is exactly the same as it had been before the movement [19].

2.2.2 Magnetic Anisotropy

There are forces within a crystal that attempt to align magneton axes with one of the crystal axes. In the absence of other forces, each domain field assumes one, out of some finite number, of the preferred crystalline directions. For instance, a cubic crystal has six preferred directions. This phenomenon is called *Magnetic Anisotropy.* These crystal-axis aligning forces can be countered by other forces and the magnetons consequently aligned in directions other than crystal-axis directions. *Magnetocrystalline Anisotropic Energy* is the force needed to rotate domain directions against crystal anisotropy [19, page 50].

An analysis of magnetization curves for ferromagnetic single crystals indicate that they exhibit a well-defined magnetic crystallographic anisotropy. This means that a ferromagnetic crystal has an axis of easy magnetization and an axis of hard magnetization. From figure 2.6 and 2.7 one can see that $<110>$ is the easy axis for Fe and $<111>$ is easy for Ni. The free energy of the magnetic crystallographic anisotropy for a ferromagnetic crystal is a function of the direction cosines of the angles between magnetization and the crystal axes. This energy is minimum when the magnetization is along the axes of easy magnetization[2]. Experiment indicates that work must always be performed by

[2]Absolute minimum of the free energy corresponds to the axis of easiest magnetization; other easy axes cause relative minima only.

outside forces in order to divert magnetization away from these easy axes. This interaction between external forces and internal magnetization is partly responsible for the magneto-mechanical interaction observed in magnetostrictive materials .

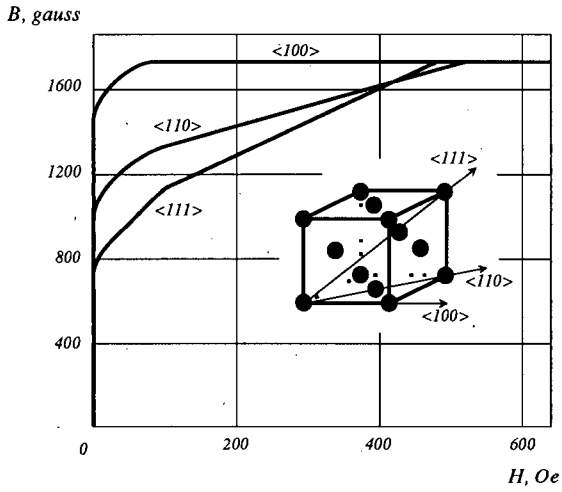


Figure 2.6: Magnetization curves of cubic Fe single crystals for three principal crystal axes $< 100 >$, $< 110 >$ and $< 111 >$ (after [20, page 941].

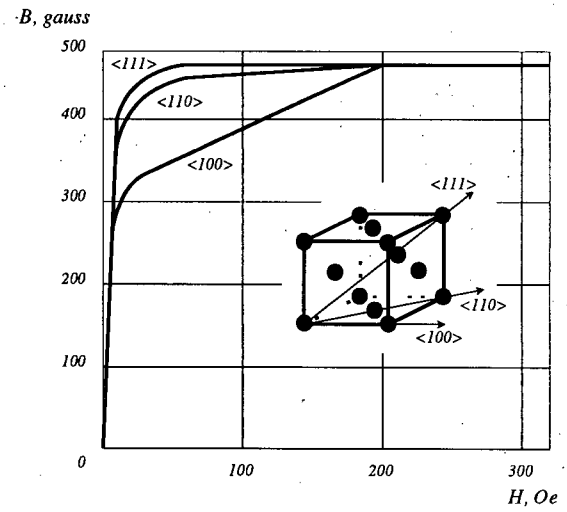


Figure 2.7: Magnetization curves of cubic Ni single crystals, for three principal crystal axes $< 100 >$, $< 110 >$ and $< 111 >$ (after [20, page 941].

2.2.3 Other Magnetic Energies

In addition to the anisotropy energy described above, there are a few other energies in play inside a crystal, and the magnetization processes of ferromagnetic substances is a result of the competition between these energies (see [17, page 38] for detail).

The Magnetostatic Energy When summed over the whole space, the magnetic energy is $W = -\frac{1}{2}\mu \int_V \mathbf{M}\, dV$ where $\mathbf{M}dV$ is an elementary magnetic moment submitted to local magnetic field, μ the permeability and V the volume. This energy would be noticeably reduced if the magnetic flux can be closed inside the material.

The Magnetocrystalline Anisotropy Energy Discussed in section 2.2.2.

The Heisenberg Exchange Energy This energy is of electrostatic and quantum origin, and decreases when the thickness of the domain walls increases. On the other hand, the magnetocrystalline anisotropy energy is minimized by thin walls. Thus the width of any wall is a result of the compromise between these two energies.

The Magnetoelastic Energy The last contribution to the total magnetic energy arises from the strain dependence of the three previous ones. Any magnet is not perfectly rigid, but is elastic, and it is this *magnetoelastic energy* that causes magnetoelastic strain. This energy may modify markedly the magnetization process when the material is submitted to external forces. The resulting effects, if harnessed, can be used to convert magnetic energies into mechanical ones and vice versa.

2.2.4 Hysteresis

The configuration of domains within a ferromagnetic crystal represents a minimum energy condition. When immersed in a weak magnetic field, domains that are oriented in the general direction of the imposed field become larger in volume and those that are not so oriented shrink in size. These changes are reversible however. Now when the imposed field is increased beyond some value, additional effects can take place, and one of them is caused by crystal imperfections. The magnetic fields associated with crystalline irregularities hold a wall in a fixed position until the external field is strong enough to break the wall loose from the imperfection. This is partially responsible for the hysteresis observed in the magnetization of ferromagnets.

The favoured domains take over those less favourably oriented until each crystal in a material becomes a single domain. Under these circumstances, the axes of the resulting domains lie along one of the finite number of possible crystallographic axes that is closest to the imposed field direction. When the field is further increased, the axis of

magnetization of each domain rotates from its preferred crystal axis until it is aligned with the external field. Figure 2.8 illustrates the different phase of magnetization in ferromagnets.

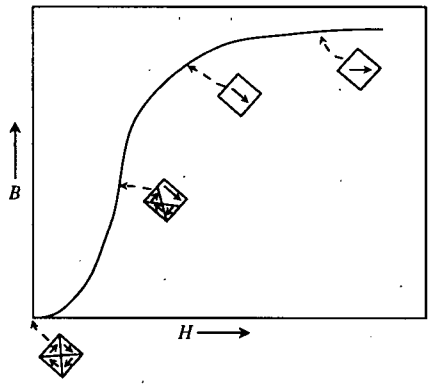


Figure 2.8: Schematic illustration of domain structure at various stages of the magnetization process.

2.3 Magnetostriction in General

Due to the dependence of magnetic interactions on the distance between the interacting particles, any elastic and magnetic body exhibits magnetoelastic phenomena: its dimensions and elastic properties depend on its magnetic state (*direct magnetoelastic effects*); and its magnetic properties are influenced by the applied and internal mechanical stresses (*inverse magnetoelastic effects*).

2.3.1 Volume Magnetostriction

When a nickel sphere is cooled down from, say, 800 K an anomalous thermal expansion is observed near the Curie temperature. This anomalous volume change, which is associated with the magnetism of any substance, is called its *volume magnetostriction.* This effect is quite general in magnets.

2.3.2 Joule Magnetostriction

If a magnetic field is now applied, an additional deformation discovered by W. P. Joule is observed: the sphere is transformed into an ellipsoid [17, page 46] (see figure 2.9). This anisotropic deformation depends on the field strength and reaches a maximum when the magnetic saturation is attended. Figure 2.10 shows this dependence of magnetostrictive strain, λ, on H for nickel.

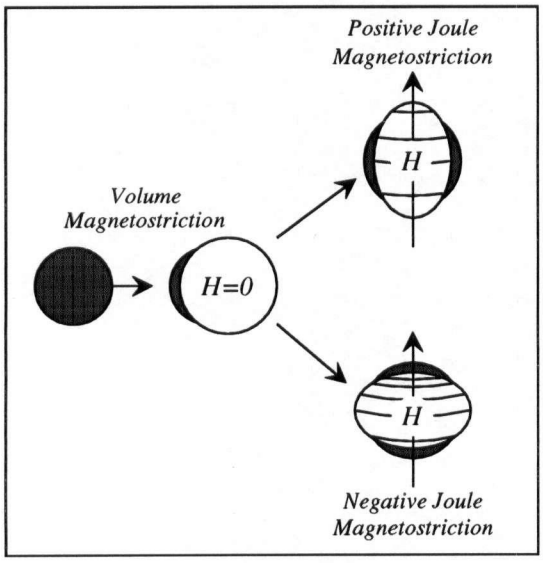


Figure 2.9: Volume and Joule magnetostriction of a spherical sample (after [17, page 46]).

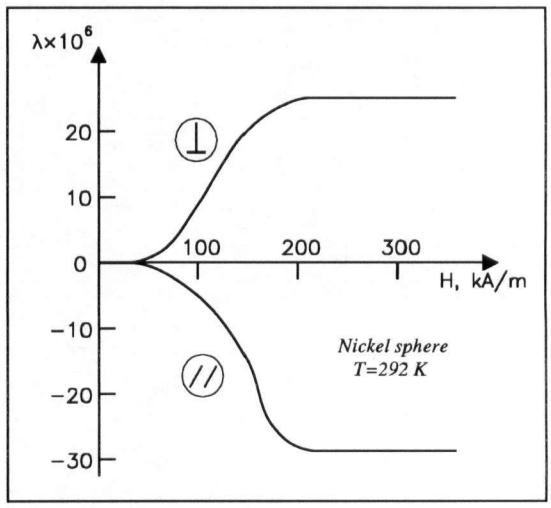


Figure 2.10: Field dependence of the Joule magnetostriction of a polycrystalline nickel (after [17, page46]).

If the material is initially isotropic, the magnetostriction ($\lambda_{//}$) along the direction of applied field should be twice that of the corresponding elongation ($\lambda_{\perp}$) in the direction perpendicular to the field (assuming negative magnetostriction). However, it is obvious from figure 2.10 that $-\lambda_{//}/2\lambda_{\perp}$ is smaller than unity. This indicates that the initial demagnetized state of the material is anisotropic. And it is for this reason that the saturated magnetostriction λ_s is now defined as $2/3(\lambda_{//} - \lambda_{\perp})$, instead of the old one

using $\lambda_{//}$ only. This new definition provides better consistency since it is independent of the demagnetized state.

The magnetization of ferromagnet is an odd function of H, whereas the magnetostriction is an *even* function, as can be seen in figure 2.11. This and the hysteresis shown are characteristics of ferromagnets. In addition, some ferromagnets (such as iron) show a sign change in λ at higher fields.

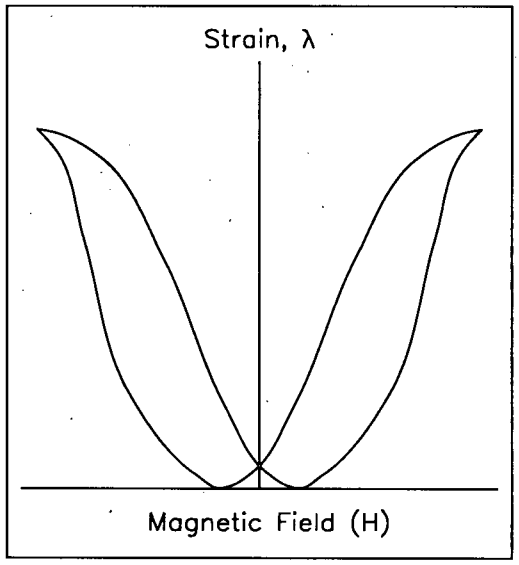
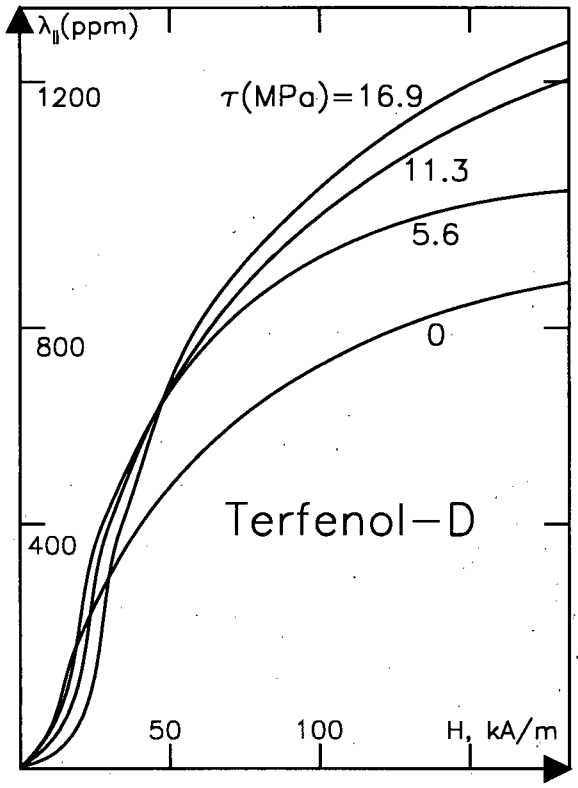


Figure 2.11: Magnetomechanical hysteresis loop (exaggerated). The loop is caused by the dependence of strain on the state of magnetization. [19]

In addition to external field, magnetostriction is also affected by external mechanical stress. This stress dependence presents itself mainly by the modification of the M(H) curves, and hence of the λ(H) curves. However, the value of the magnetostriction coefficient λ_s is usually stress-independent. Figure 2.12 shows the $\lambda_{//}$ curve for a $<11\bar{2}>$[3] polycrystalline rod of Terfenol-D under different prestress. The large increase of $\lambda_{//}$ with the compressive stress can be explained with the schematic in figure 2.13. The bottom

[3]The choice of this axis is explained in section 2.4.1.



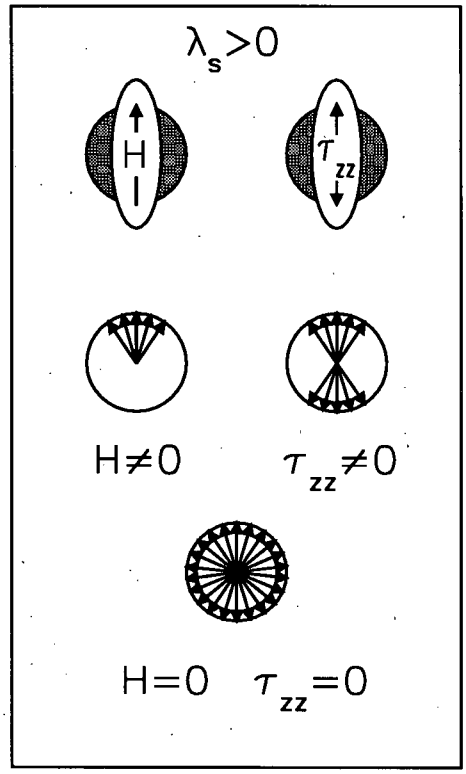


Figure 2.12: (Left) Typical magnetoelastic response for Terfenol-D at various compressive prestress (after [21]).

Figure 2.13: (Right) Effect of a magnetic field and of an uniaxial stress on the distribution of the magnetic moments in a ferromagnet. (after [17, page 48]).

sphere shows the isotropic spatial distribution of the magnetons under no external stress and magnetic field. When magnetized or stressed, the magnetons will orientate to the more favoured directions, making the sample anisotropic. The magnetostrictive strain is now associated with this distribution of magnetic moments. When a magnetic field is applied under these circumstances, the $\lambda_{//}$ value becomes a function of the applied prestress. If λ_s is positive, as in the case for Terfenol-D, the applied pre-stress will cause some magnetons to orientate perpendicularly to the stress direction (Oz). Then more moments are available to rotate under an applied magnetic field parallel to Oz. Hence

$\lambda_{//}$ is expected to be increased by the application of compressive pre-stress, as actually observed in figure 2.12.

2.4 Terfenol-D ($Tb_{0.3}Dy_{0.7}Fe_2$)

Many of the rare-earths have large magnetostrictions at low temperatures. Measurements of basal plane magnetoelastic strain of Tb and Dy at low temperatures show strains 100 to 10,000 times typical magnetostriction values [4] [23]. However, they show only negligible paramagnetic magnetostriction at room temperature which is above their Curie points. In 1971, a search for magnetostrictive materials with high magnetostriction at room temperature is started. Highly magnetostrictive rare-earths, Tb and Dy, are combined with the magnetic transition metal, Ni, Co and Fe. The highest ordering temperatures[5] are found for the RFe_2 compounds, such as $TbFe_2$ and $SmFe_2$. Because of the high concentration in rare-earths, these compounds have huge room temperature magnetostrictions. However, these alloys also possess the largest known magnetic anisotropies (see section 2.2.2) of cubic crystals at room temperature. Thus, very large fields are needed to harvest the full magnetoelastic strains (e.g., >10 kOe or 795 kA/m for $TbFe_2$).

Fortunately, it is possible to tailor compounds with optimum magnetostrictive properties and magnetocrystalline anisotropy simultaneously. Pseudobinary compounds can be constructed in such a way as to minimize the magnetic anisotropy while still maintaining a large positive (or negative) magnetostriction . Ternary compounds yielding large positive magnetostrictions at room temperature must contain Tb [23, page 561]. Hence acceptable anisotropy compensating systems are $Tb_{1-x}Dy_xFe_2$, $Tb_{1-x}Ho_xFe_2$ and

[4] Clark has recently reported a high power low temperature actuator using $Tb_{0.6}Dy_{0.4}$, capable of giving 6% strain at 77 K [22].

[5] Related to the Curie temperature of the material.

$Tb_{1-x}Pr_xFe_2$[6]. In particular, $Tb_{0.3}Dy_{0.7}Fe_2$ (Terfenol-D) displays very large magnetostriction at room temperature. This huge magnetostriction is due to the combination of the individual ones for the $TbFe_2$ and $DyFe_2$ alloys; on the other hand, the magnetocrystalline anisotropy vanishes near room temperature, as the positive contribution of $DyFe_2$ and the negative one of $TbFe_2$ cancel each other. Figure 2.14 shows the magnetostrictions at room temperature of polycrystalline samples of $Tb_{1-x}Dy_xFe_2$ for H = 10 kOe (795 kA/m) and H = 25 kOe (1988 kA/m). Note that near $x \simeq 0.7$, the magnetostrictions exhibit a peak for both stress levels, reflecting the near zero magnetic anisotropy at this concentration.

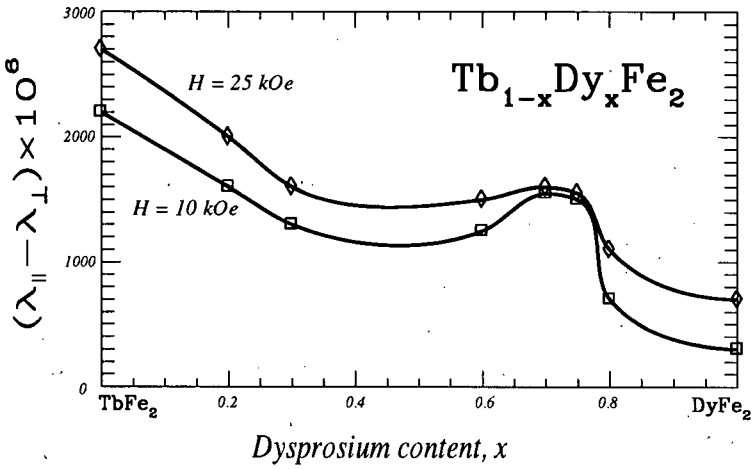


Figure 2.14: Magnetostriction of $Tb_{1-x}Dy_xFe_2$ at room temperature (taken from [23])

2.4.1 Magnetostriction in Grain Oriented Terfenol-D Rod

Commercially available Terfenol-D rods are manufactured by two methods. The free stand zone melt (FSZM) technique produces twined single crystal rods from 3 to 7 mm

[6]In fact, Williams and Koon have determined the composition of the pseudobinary alloy $(Tb_{0.3}Dy_{0.22}Fe_2)_x(Tb_{0.14}Ho_{0.86}Fe_2)_{1-x}$ for simultaneous minimization of the two lowest order anisotropy constants [23, page 564].

in diameters. Grain oriented rods from 8 to 50 mm are produced by the modified Bridgeman (MB) method. These latter rods are made up of many small crystals with aligned crystal axes, and therefore their performance is comparable to those with twined crystal. Transducer elements from both processes have the magnetostrictive growth axis along the rod axis. The ones used for the UBC design are manufactured by the modified Bridgeman method.

$< 112 >$ is the crystallographic axis that $Tb_xDy_{1-x}Fe_{2-y}$ most easily grow along. Unfortunately, this easy grown axis is not a principle (ie., easily magnetized) direction. At room temperature, the $<111>$ axes are magnetically easy and magnetization jumping can occur between one $<111>$ direction and a second $<111>$ easy direction which is closer to the rod axis. Huge changes in sample dimensions accompany with these jumps. However, the process is complex because of the slant axis. Clark [24] discusses the mechanism of jumping for rods with twined single crystal.

2.4.2 Surface Strain vs Bulk Strain

In many designs incorporating Terfenol-D (e.g., [25]), the magnetoelastic strains of the drive units are measured using strain gages. Recently there has been studies on the discrepancies between the surface strain (measured by strain gauges) and the bulk strain (actual extension of the rods). Schulze *et al.* [26] use an optical interferometer to measure spatial variations in the the magnetostrictive strains across the plane surface of a polished, grain oriented disc. DC and AC fields are applied to a direction perpendicular to the surface (i.e.,$<11\bar{2}>$ direction). At AC fields (4.5 kHz and 30 kHz), the circumference is found to be almost stationary, whereas the amplitude increases towards the centre. This variation is not in agreement with skin effect, and Schulze *et al.* suggest that it is attributed in part to demagnetizing field effects and surface anisotropy which hinders domain wall motion.

Therefore when designing sensor for the actuator, it would be more appropriate to measure the end to end movements instead of just the surface strains: strain gages will provide a qualitative, but not quantitative, indication of strain.

2.4.3 Eddy Current Loss

When an electrically conducting material such as Terfenol-D is magnetized dynamically, eddy current is induced inside the material. This flow of current generates its own magnetic field to counteract the external field. Therefore the effective magnetic field felt inside a magnetostrictive rod, such as Terfenol-D, is reduced, which makes the overall magnetization process more complicated. In [27] and [28], Kvarnsjö and Engdahl try to establish a model for the interaction between eddy currents and magnetoelasticity in Terfenol-D. They radially resolve the rod into sections, and calculate the individual behaviour . Then the interactions between the sections are taken into account in an effort to determine the overall behaviour of the rod.

2.4.4 Skin Effect

Eddy current also causes the skin effect. That is, when magnetized by an AC field, the centre of a Terfenol-D rod is subjected to less field than the surface. Not only will the centre expand less, it will also act as a load on the rest of the structure, further reducing the effectiveness of the unit. Since the skin depth is frequency dependent, it is important to design actuators accordingly based on the bandwidth expected. For example, a thickness of 11 mm is the maximum for a Terfenol-D rod expected to operate to 1 kHz [29].

Chapter 3

Designing the Actuator

3.1 Existing Applications and Designs

Since Terfenol-D rods became available commercially in the mid 1980s, there have been numerous designs utilizing this highly magnetostrictive material. Sonar is one of the first and most widely used applications, because Terfenol-D can deliver high power in the 0 to 5 kHz frequency range. More recently, there are increasing attempts to design high precision actuators using this material. Etrema Products Inc. has commercially made available Terfenol-D actuators for a few years now. Using a simple design [30], these units can produce a no-load displacement of 110 μm. Eda *et al.* [31] suggest that since the properties of Terfenol-D vary significantly with temperature, it is important to control the temperature inside the rod. They develop a water cooling system to keep the temperature to within 10 m°C, and achieve nanometre positioning accuracy without using position feedback. However, it is believe that with a good output sensing device, sensitivity to temperature changes can be minimized without resorting to complicated water-cooling system. Wang and Busch-Vishniac [32] connect two one-dimensional Terfenol-D actuators and develop a two degrees-of-freedom micro-positioner, with a 100 μm by 100 μm workspace.

Vibration generation and isolation are also areas where magnetostrictive materials are found to be useful. Vingsbo and Schön [33] use a Terfenol-D actuator to generate vibrating action for fretting testing at low amplitude (less than 2 μm). For isolation,

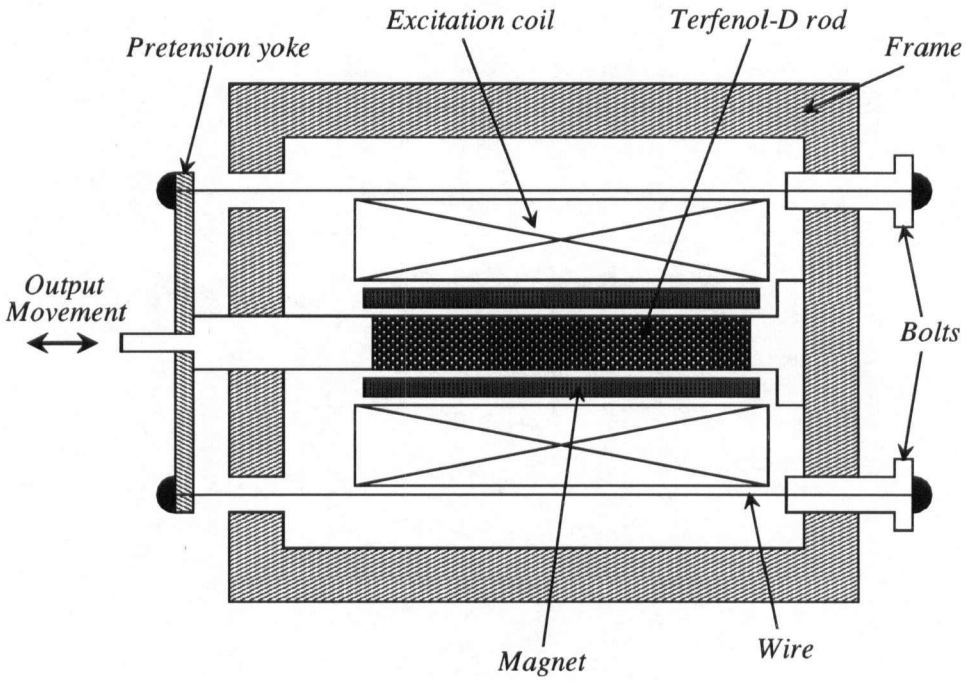


Figure 3.15: A typical design of Terfenol-D actuator.

Bryant *et al.* [34] investigate the performance of a motion stabilizing table with three Terfenol-D actuators as table legs; Geng and Haynes [35] propose a multiple degrees-of-freedom vibration isolating system using Terfenol-D actuators as trusses for a Stewart platform-type mechanism.

3.2 Limitations of Existing Designs

Most of the above-mentioned actuators have the same basic structure: a Terfenol-D rod at the centre, a coil surrounding it to generate magnetic field, magnets at the outside or top and bottom for magnetic bias, and bolts or wires for prestress (see figure 3.15). Although simple and compact, this general design has a few deficiencies, making it not suitable for chatter vibration control. These shortcomings are discussed in the following,

together with some Terfenol-D properties that affect actuator design.

1. A prestress is needed to increase the operating magnetoelastic strain. However, as discuss in the previous chapter, this external stress also affects other properties of the drive unit (e.g., shape of the magnetization curve), and therefore should be kept constant in order to get a consistent response from the actuator. It is obvious that a simple spring type arrangement is not adequate because as the rod expands, the prestress increases according to Hook's law and vise versa.

2. Since the expansion of Terfenol-D is independent of the polarity of the external field (see figure 2.11), a dc magnetic bias is required if the drive unit has to deliver bi-directional movement . This could be accomplished by using a dc field coil (with control signal superimposed on top); however, the heat loss would be quite high. In most cases, permanent magnets are employed. If a soft iron magnetic circuit is used at the same time to help concentrate the field, care should be taken not to increase the overall hysteresis of the actuator by too much. Beadle [36] even suggests that it is debatable whether to have a soft iron magnetic return path at all. Furthermore, the permanent magnet should not form part of the ac magnetic circuit, since this would give rise to a large hysteresis. Therefore, it is desired to have a separate static and dynamic magnetic path: the DC path can use soft iron and powerful magnets to create an adequate and even magnetic bias, whereas the AC field should encounter as little hysteretic material as possible.

3. Although Terfenol-D can withstand great compressive stress (many MPa), it is quite weak in the tensile direction[1]. This means, it can only exert forces in one direction: push, but not pull. Therefore in all of the above mentioned designs, the Terfenol-D rod is responsible only for the pushing movement, and the pulling is realized by the

[1]In fact, the author *did* manage to break a rod 21 mm in diameter, with a price tag of almost $3,000, by just bending it slightly.

prestress mechanism. This may be adequate at low frequency; but at high frequency the spring-type system might not be able to generate forces as high as the rod can supply. This would result in unbalanced actions.

4. The amount of prestress is crucial to the performance of a Terfenol-D actuator. However, all designs found have no way of measuring this stress directly. Most of them depend on some kind of threaded adjustments. Theoretically, one can calculate, for example, the increase in prestress per turn of a thumb nut by knowing the spring constant of a wire, bolt or spring, but this does not take into account effects like friction and compliance of the frame. Therefore, direct measurement of the prestress is desired.

5. Skin effect limits the depth of penetration of the magnetic flux. Therefore one should take this into account when deciding on the dimensions of the Terfenol-D rod. If the radius required (because of, for example, output force requirement) is bigger than the skin depth at the desired bandwidth, then the centre portion not only would not expand, but would even act as a load on the expanding portion, severely reducing the output motion and force. The rod should be cut and re-laminated so that the minimum distance to free space inside the rod at any point is shorter than the skin depth at any operating frequency.

The actuator design described in this thesis tries to address some of the above stated requirements.

3.3 Essential Components of an Actuator Employing Terfenol-D

Below is a recapitulation of the essential components for a Terfenol-D actuator:

- A prestress mechanism. It should provide a constant and adjustable prestress to the Terfenol-D rods. The force capacity required is roughly (cross-sectional area of the

rod) $\times$ (tens of MPa).

- A magnetic bias, if bi-directional movement is desired. Either a magnet or dc current in coil can be used, with or without a soft iron return path.
- A coil to generate dynamic magnetization. This field should not pass through the magnet; otherwise hysteresis would increase.
- A position sensing device, if one wants feedback operation. It should measure the bulk strain instead of the surface strain, have resolution better than one micron, and be operable under strong, varying magnetic field.

Other useful, but not essential, components include a sensor for B-field inside the rod, temperature sensor, cooling for the coil and etc.

3.4 The Design Here at UBC

After studying the various existing Terfenol-D actuators, it is decided that a better design is required for the control of machine chatter vibration. The desired performance of this actuator is specified by studying the requirements posted by the control problem.

- It should have a linear motion range of more than 100 μm.
- It should be able to produce forces up to 2000 N over the full reachable space.
- It should have a motion bandwidth of about 1 kHz.

The following sections describe in detail the main features of the approach taken.

3.5 The Push-Pull Design

A push-pull type structure is chosen (see figure 3.16), primarily because Terfenol-D is not very good at pulling. It is felt that relying on the prestress to generate pulling forces would increase the nonlinearity of the system and reduce the effectiveness of the overall

design. In this design, the rods are driven by signals 180 degrees out-of-phase, an idea similar to a class $\mathcal{A}$ push-pull electronic amplifier. Although the output does trace a circular path rather than a straight line, it should be tolerable if the range of movement (about 0.15 mm) is small comparing to the arm length (a few cm). Furthermore, external structures can be used to correct for this.

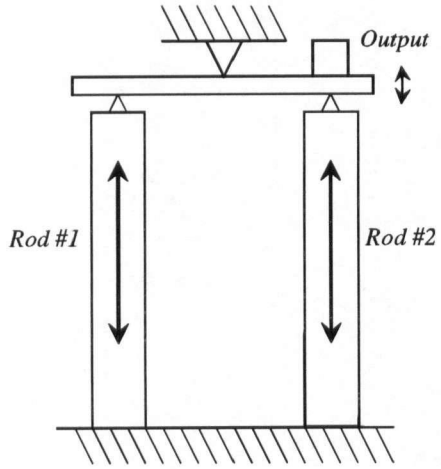


Figure 3.16: The basic push-pull design.

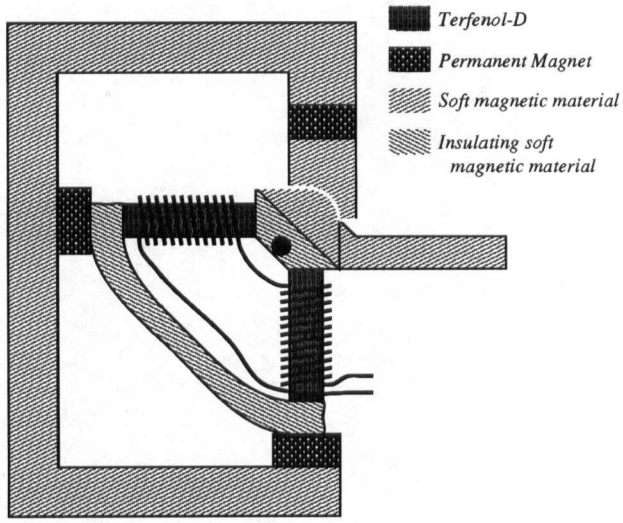


Figure 3.17: Actuator proposed by Cedell *et al.* [37] (after [17, page 355]).

3.5.1 A Survey

Two designs that also employ two active rods instead of one are found referenced in [17, page 355]. The one proposed by Cedell *et al.* (figure 3.17) has a separate DC and AC magnetic path, and it generates rotating output motions. However, no prestress mechanism is shown, and because the biasing magnets are in series with the rods, the stiffness felt at the output would be compromised. The other one mentioned is developed by Kvarnsjö and Engdahl [38]. This design uses four Terfenol-D rods and four coils to achieve a push-pull motion (figure 3.18). It allows the drive units to operate under

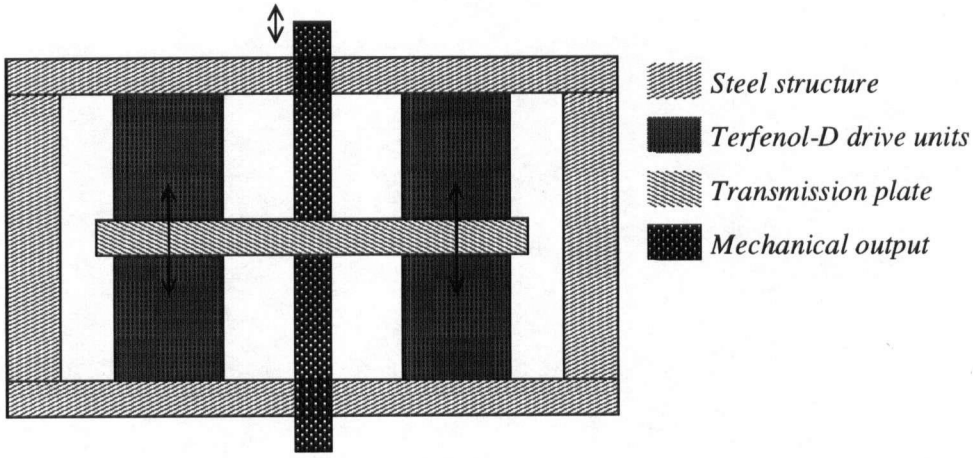


Figure 3.18: The push-pull design by Kvarnsjö and Engdahl [38].

constant prestress, since the upper units contract when the lower ones expand and vice versa. A hydraulic transmission mechanism is incorporated into the transmission plate, making impedance matching of mechanical load possible. This mechanism is also critical in the balancing of any discrepancy in response between drive units. If the transmission plate only provided a rigid connection, the output rod would be rubbing against the steel structure if the two sets of drive units do not move exactly in synchronization. However, it is felt that the hydraulic circuit inflicts undue complexity on the device. Also, no provision is made for a separate magnetic paths for DC and AC field.

3.5.2 Is Push-Pull a Waste

At the first sight, using one rod for pushing and one for pulling seems wasteful since Terfenol-D is so expensive. However, further analysis shows that the two rods are effectively acting in parallel: that is, the stiffness felt by the external load is twice that of a single rod with the same dimensions. Consider figure 3.19. Here, the rods are represented by two springs. When there is no external load, the forces exerted by the rods are both $\frac{F_o}{2} = kx_o$, where F_o is the force from prestress, k the spring constant for each rod, and

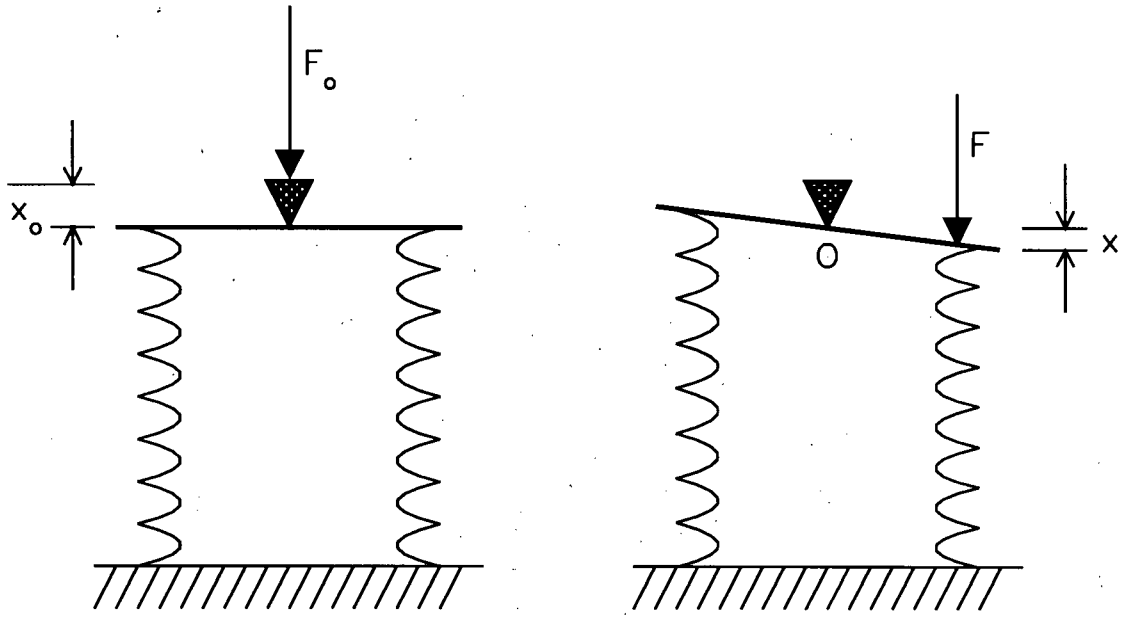


Figure 3.19: (left) Rods with prestress applied but no external load. (right) With external load applied.

x_o the initial change in length. If an external force F now causes both springs to change length further by x, then one can relate x to F by summing moment around point O in the clockwise direction.

$$\sum \overset{\curvearrowright}{M_O} \quad \text{gives} \quad k(x_o - x)l - k(x_o + x)l + Fl = 0 \tag{3.1}$$

$$-2kxl + Fl = 0 \tag{3.2}$$

$$\therefore \quad F = 2kx \tag{3.3}$$

where l is half the length of the lever. From equation 3.3, it is obvious that the external load is feeling the effect of a single spring with stiffness $2k$, ie., twice the stiffness. Therefore the combined cross-sectional area of the rods determines the load capacity of the actuator. This results in roughly the same cost in terms of raw material for Terfenol-D,

when compared to a single-rod design[2].

3.5.3 Symmetry of Actuator Response

Another benefit of using a push-pull structure is that the output response of actuator is made symmetric. In figure 3.20, the response for a single Terfenol-D rod is shown. A positive input on top of the bias will give a different response from a negative one. This lack of symmetry complicates the controller design. With a push-pull configuration, however, when one rod is in the +ve region, the other will be in the -ve region: any dissimilarity will be averaged out. The overall response is quite similar to a normal hysteresis loop for the magnetization of a ferromagnet. This reduction in nonlinearity should reduce the complication of the controller.

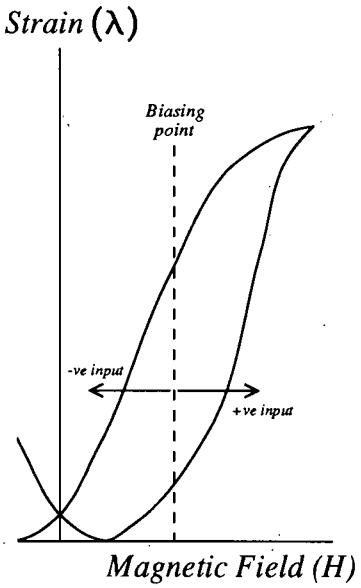


Figure 3.20: Output response of a Terfenol-D rod.

[2]Although it does cost more to acquire two rods with half the cross-sectional area (due to more labour involved), the thinner rods require fewer cuts and re-laminations to satisfy the skin-depth requirement.

3.5.4 Dimensions for the Drive Units

The output forces of the actuator are generated by straining the Terfenol-D rods, and is given by

$$F = EA\Delta l/l$$

where F is output force, E the Young's modulus, A the combined cross-sectional area of the rods, l the length of the rods without load, and Δl the change in length caused by F. The thicker the rods, the less will they be compressed under the same load, and thus more strain is left for output motion. However, this will also make the actuator more expensive, bigger and bulkier. Obviously, there is a compromise between the diameter and the length of the drive units. In this particular design, more consideration is given to the length, because the actuator's outside dimensions, which are mostly determined by this length, should be kept to a minimum.

Some rough estimations indicate that an active rod length of 120 mm is suitable. Assuming a maximum magnetoelastic strain of about 1000 ppm[3], a motion range of 0.12 mm (120 μm) can be achieved. If a maximum load of 2000 N applied at maximum displacement causes 10% (100 ppm) of compressions, then

$$E = \frac{F/A}{\Delta l/l} \tag{3.4}$$

$$\begin{aligned} \Rightarrow \quad A &= \frac{F}{E} \cdot \frac{1}{\Delta l/l} \\ &= \frac{2000N}{25 \times 10^9 N/m^2} \cdot \frac{1}{100\,ppm} \\ &= 8 \times 10^{-4}\, m^2 = 800\, mm^2 \end{aligned} \tag{3.5}$$

where E=25 GPa is the Young's modulus of Terfenol-D (a worst case figure given by the

[3]Although the absolute maximum strain for Terfenol-D can go up to 1500 ppm or even 2000 ppm, the response becomes increasingly nonlinear as the strain goes beyond about 1000 ppm.

supplier, Etrema Product Inc.). Therefore area for each rod should be around 400 mm^2. The diameters would then be

$$D = \sqrt{\frac{4A}{\pi}} = \sqrt{\frac{4 \times 400mm^2}{\pi}} = 22.6\, mm$$

This figure is bigger than the maximum thickness 11 mm recommended by Etrema for operating frequencies to 1000 Hz [29] (due to skin effect). Therefore, the rods need to be sliced once along their axes and re-laminated. To allow for some deviations, two rods with diameter 21 mm and length 170 mm[4] are ordered. The resulting maximum effective thickness in the horizontal direction is 10.5 mm.

3.6 Frictionless Joints

Although the push-pull design is found to be superior to its single-rod counterpart in many respect, one major problem needs to be addressed: How to provide frictionless connections between the Terfenol-D rods and the supporting structure? This is an important issue because of the large forces involved. Both the prestress and the external force are in the order of thousands of newtons. Therefore using simple pin-pivoted joints would result in significant stick-slip nonlinearities, thus hampering the performance of the actuator.

Many different approaches have been studied, and the most promising solution turns out to be very simple and elegant, and it is called flexure pivots or flexure supports[5]. Consider figure 3.21. If the range of lateral movement is kept small, the rod is effectively pivoted around a line somewhere along the flexure plate, and only a small force is required to rotate the rod around this line, no matter how big F is.

[4]The active length remains at 120 mm, because the top and bottom portions are used for magnetic interface.

[5]See [39] and [40] for more information on flexure pivot design.

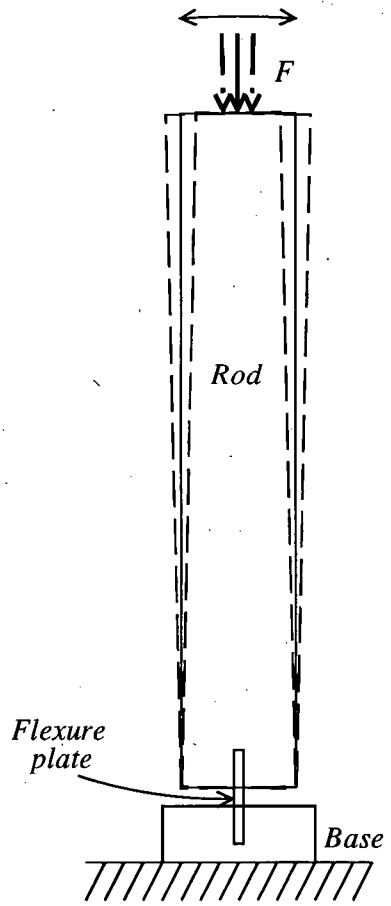


Figure 3.21: A flexible plate (flexure) can provide a frictionless support between the rod and the base.

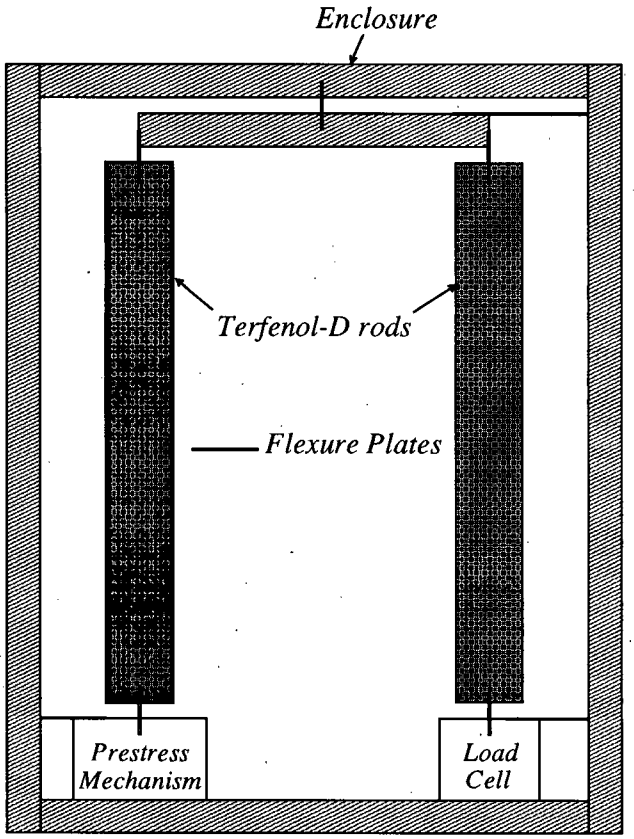


Figure 3.22: Arrangement of the flexure plates in the final design.

A total of eight flexure plates are used in the final design (figure 3.22): five for vertical support, three for lateral. The two rods, the vertical flexure plates connected to their ends, and the lever at top form a quadrangular structure. When the rods contract and expand as driven by the dynamic excitation, the flexure plates allow the lever to rotate around the pivot at top of the structure. Output can then be taken from end of the lever.

There are several requirements for the material used as flexure plates. High compressive strength is needed, since the cross-sectional areas of the plates are relatively small. In addition, it should have low modulus of elasticity in order to make bending easy. Furthermore, because of the very slight hysteretic losses in all materials, alloys with small internal hysteresis, such as phosphor bronze and beryllium copper, should be used if high precision is needed [40]. For this particular application, only the compressive strength and the ease of bending are important, and stainless steel sheets 1 mm thick are used. The free-standing length of the vertical flexure plates is chosen to be 3 mm.

The maximum force experienced by the vertical plates is the sum of the prestress and the external load. Area of each rod $\pi(10.5\,mm)^2 = 346\,mm^2$, together with a prestress of 7 MPa (suggested by Etrema), make the force of prestress $(346\times10^{-6}m^2)\cdot(7\times10^6 N/m^2) = 2422\,N$. Adding the maximum external load felt by each rod (2000 N shared by two rods), the vertical flexure plates should be able to withstand at least 3422 N. This force results in a compressive stress of $(3422\,N)/(1\,mm\times16\,mm) = 214\,MPa$ inside the plates (width of flexure plates is 16 mm).

On top of this stress, however, there is an additional stress induced by bending. Consider figure 3.23. If the flexure plate assumes a circular cross-section, then $l = 3\,mm = R\theta$, and the length at the outside of the flexure is

$$l + \Delta l = (R + \frac{1}{2}d)\,\theta \tag{3.6}$$

Therefore the induced strain is

$$\varepsilon = \frac{\Delta l}{l} = \frac{(R+\frac{1}{2}d)\theta - R\,\theta}{l} = \frac{\frac{1}{2}d\,\theta}{l} = \frac{1}{2}\cdot\frac{0.5\,mm}{3\,mm}\cdot\theta = 0.0833\,\theta \tag{3.7}$$

The angle of movement can be estimated by dividing the range of output movement by half the length of the lever, ie., $\theta = \frac{0.15\,mm}{42.5\,mm} = 0.0035\,rad = 0.2°$. This bending would result in a strain of $0.0833 \times 0.0035 = 2.9 \times 10^{-4}$, and translate into a stress of

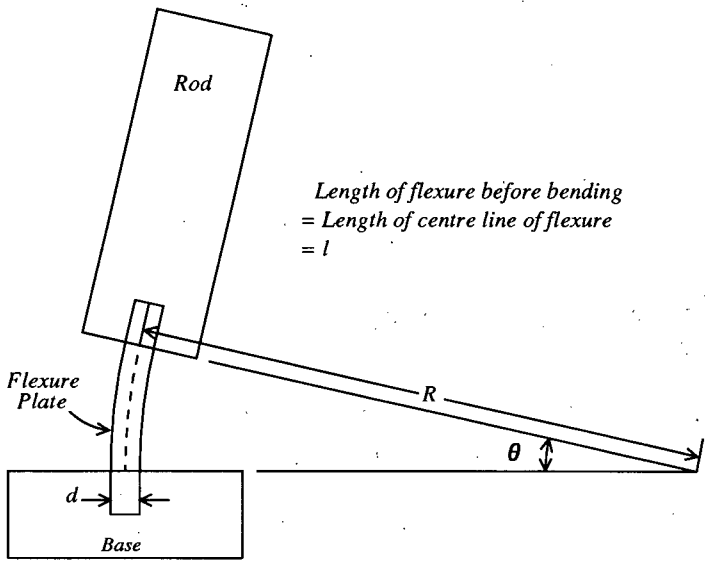


Figure 3.23: Bending of a flexure plate (exaggerated view). A circular profile for the bending is assumed.

$\varepsilon E = (2.9 \times 10^{-4})(193 \times 10^{9}) = 56\,MPa$. Adding in the loading, the maximum stress felt by a vertical flexure plate is $214 + 56 = 270\ MPa$. This figure is within the yield stress 345 MPa for stainless steel. However, it should be noted that any bending larger than, say, 0.5° or so would result in permanent deformation and work hardening of the plates, and therefore should be avoided.

Even though the flexure plates would not yield due to compressive stress, they could still fail because of buckling[6]. Buckling occurs some time before the yield stress in compression is reached, and it could happen because the plate is not perfectly straight initially, the load is not applied exactly along the centre, and/or there is some lack of uniformity in the material properties throughout the plates. By looking at the differential equation for an axially loaded beam, one can see that the solutions are linear combinations of a set of eigenfunctions. When the load is smaller than a certain value, P_e, the only possible solution is the trivial solution, and no buckling would occur. This limiting load

[6]For a more detailed study on buckling, see any book on mechanical design, eg. [41].

is given by the Euler theory,

$$P_e = \frac{4\pi^2 EI}{L^2} \tag{3.8}$$

for plates with fixed ends (I is the moment of inertia and L the length of the beam). Now

$$E = 200\,GPa$$

for stainless steel,

$$I = \frac{1}{12}bh^3 = \frac{1}{12}(16\,mm)(1\,mm)^3 = 1.33 \times 10^{-12}\,m^{-4}$$

for a rectangular cross-section, and

$$L = 3\,mm$$

$$\therefore \quad P_e = \frac{4\pi^2(200 \times 10^9\,N/m^2)(1.33 \times 10^{-12}\,m^{-4})}{(0.003\,m)^2} = 1170\,kN \tag{3.9}$$

This load would result in a compressive load of $(1170\,kN)/(16\,mm^2) = 73\,GPa$, which is much bigger than the ultimate strength. Therefore, the plates would fail compressively well before buckling would occur.

3.7 Adjusting the Prestress

Another advantage of the push-pull design is that it allows better adjustment and maintenance of the prestress. As stated before, common prestress mechanisms rely on springs, and therefore is prone to vary as the rod changes in length. The exact amount of static prestress is also difficult to adjust because of friction. In the UBC design, the actual prestress is monitored directly by a load cell, eliminating any uncertainty caused by friction and other factors. In addition, the among of prestress is constant throughout the motion range, provided that the case is made rigid enough. This advantage is illustrated in the following.

Figure 3.22 shows the general layout of the prestress mechanism and the load cell. Prestress is applied by two sliding wedges, adjusted by a thumb screw outside the case (the detail view is shown in figure 3.24). As the rod on the left is being pushed upward by

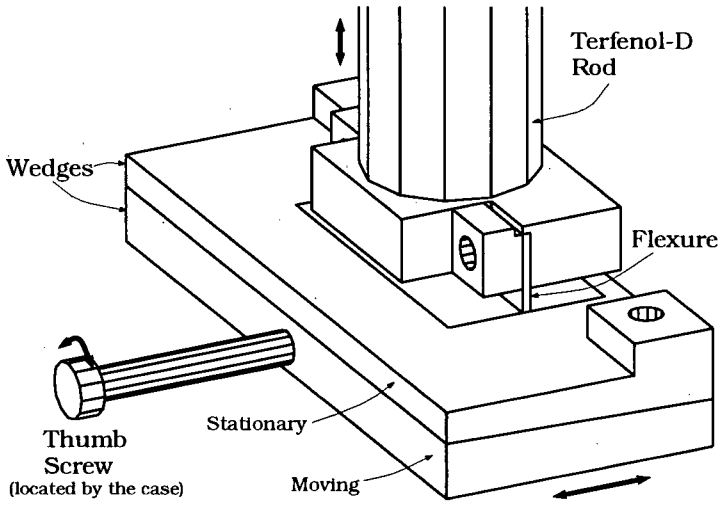


Figure 3.24: The prestress mechanism. The operator turns the thumb screw, which moves the bottom wedge in or out, and the load on the rod is changed accordingly.

this mechanism, the stress is transmitted through the lever at the top and onto the right rod. If the small forces for bending the flexure plates are neglected, the prestress in both rods would be equal and can be directly measured by the load cell at the bottom right. Any initial loading on the loadcell, including weights of the rods and other components, can be compensated for before any prestress is applied. In addition, any externally applied forces can also be measured.

Now consider figure 3.19. If the height of point O relative to the base of the loadcell stays constant, then, at least for small movements, the prestress stays the same at kx_o no matter how big F is. This allows more accurate investigations into how the shape of the response curve for Terfenol-D changes with different prestress values.

Because the height of the actuator should be kept to a minimum, a space-efficient

clamp for the flexure plates is designed (figure 3.25). The clamp can be held in place

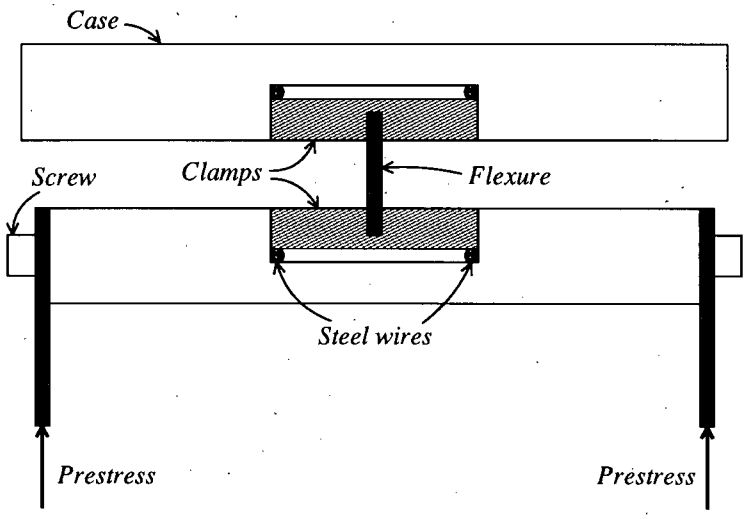


Figure 3.25: A space saving clamp for flexure plates (not to scale).

simply by the compressive force, or with the help of some screws. As the prestress is applied, the flexure would push into the clamp, and the action of leverage would tighten the grip on the plate. Therefore, the harder it pushes, the better it clamps.

3.8 The Magnetic Bias and Excitation Coils

Since magnetostrictive materials expand under magnetic field independent of the direction of excitation, bi-directional movements are only possible with the help of a magnetic bias. The idea is similar to a simple transistor amplifier operating in class A. The operating point is chosen to be at the middle of the output range. Positive inputs to the excitation coil, adding to the bias, cause the rod to expand further, whereas negative inputs subtract from the bias, causing the rod to expand less, which can be seen as a contraction from the point of bias.

This magnetic bias can be provided by either a permanent magnet or a dc current flowing in the excitation coil, with the latter method resulting in a simpler magnetic

circuit. However considerable amount of heat would be generated even when the actuator is not moving. Therefore the former route is chosen. Using magnets, however, posts some unique challenges. First, in order to reduce hysteresis it is highly desirable to have a separate dc and ac magnetic circuit. That means the dynamic field would not excite the magnet. Second, because of Terfenol-D has relatively low permeability ($\mu_r = 5$–10), magnets with high coercive forces are needed to deliver the required biasing magnetic field into the rods.

3.8.1 The Magnetic Circuit

The solution proposed here meets both requirements, and is shown schematically in figure 3.26. When there is no current flowing in the coils, the magnet induces downward-

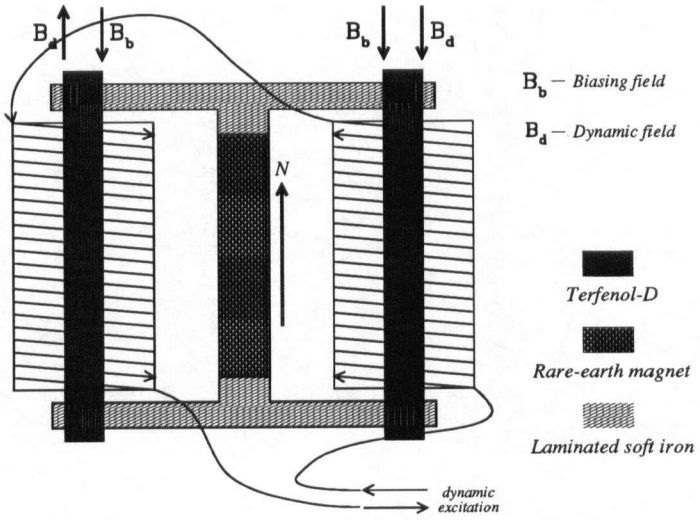


Figure 3.26: Components of the magnetic circuit. Note that the ac and dc paths for the flux are separate.

pointing B-fields with equal strength in both rods, and this will provide the magnetic bias. Now, since the coils are connected in series, a positive current would cause the field to increase in one rod, and decrease in the other. This dynamic field will only flow in

the outside loop, however, if the magnetic circuit is built symmetrically. This is because the opposing magnetomotive forces from the two coils would cancel each other out inside the magnet. Therefore the dynamic field will not pass through the highly hysteretic rare-earth magnet.

From data supplied by Etrema Inc., a bias point of H=45 kA/m is chosen. With a mechanical prestress of about 7 MPa, having H varies from 0 to 90 kA/m should result in an output range (with no load) of $1000\ ppm \times 120\ mm = .12\ mm$. The magnets used are premium grade neodymium-iron-boron with $BH_{max} = 42\,MGOe$. The required dimensions are calculated by considering the space occupied by the Terfenol-D rods as air gaps [42, page82]. Figure 3.27 shows the demagnetization curve for the magnet chosen, and figure 3.28 shows the load line assumed by a permanent magnet.

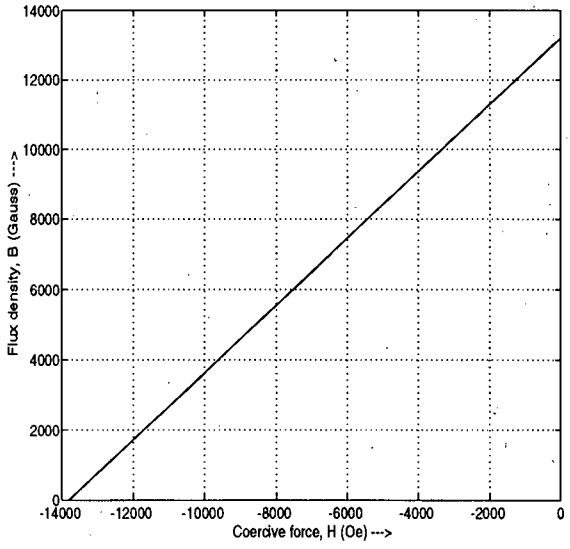


Figure 3.27: Demagnetization curve for 42H NdFeB premium grade magnet (from [43, page 5-12]).

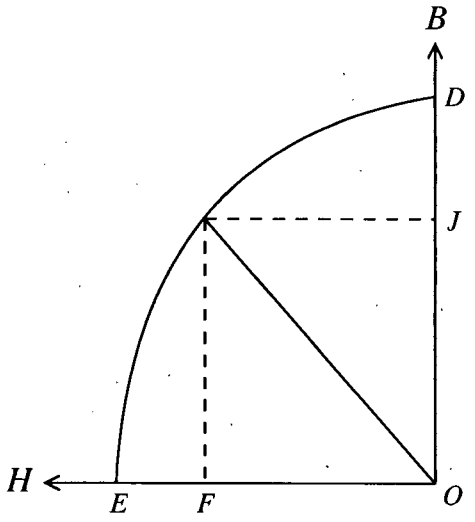


Figure 3.28: Load line of a permanent magnet.

Let l_g be the length of the air gap, which is the space occupied by the Terfenol-D

rods and equals to the full length 170 mm; and a_g be the gap's cross-sectional area. Due to the small relative permeability of Terfenol-D ($\mu_r \simeq 4.5$ to 10), leakage from the soft iron path will affect quite significantly the requirements of the magnet. It is estimated that the total exposed area of soft iron is about 10 times of the combined cross-sectional area of the rods. Therefore, the effective area of the air gap is chosen to be (worst case)

$$a_g = (1 + \frac{10}{4.5}) \times a_{Terfenol-D} \simeq 3 \times 2 \times \frac{\pi}{4}(21\ mm)^2 = 2078\ mm^2$$

Also let l_m be the length of the magnet, and a_m be its cross-sectional area. Now, the required H-field inside Terfenol-D rods is 45 kA/m. So the total magnetic flux inside the "air gap" is

$$B_{Terfenol-D} = B_{gap} = \mu H_{gap} = \mu_r \times 45\ kA/m \times 4\pi \times 10^{-7}$$

$$= \begin{cases} 0.254\ Tesla & \text{if } \mu_r \text{ is 4.5,} \\ 0.565\ Tesla & \text{if } \mu_r \text{ is 10.} \end{cases}$$

Since the inherent mmf of a magnet is the coercive force multiplied by the length of magnet, and the magnetomotive force (mmf) required for a path with length l, permeability μ and flux B is Bl/μ, therefore

$$OF \times l_m = \frac{B_{gap} l_g}{\mu} \tag{3.10}$$

$$OJ \times a_m = B_{gap} a_g \tag{3.11}$$

where OF and OJ define the operating point of the magnet (see figure 3.28). Although any load line can be chosen, the volume of the magnet will be minimized if OF $\times$ OJ is at the maximum. In this case, the corresponding OF and OJ are about 6500 Oe = 517 kA/m and 6500 Gauss=0.65 T respectively. Equation 3.10 gives

$$l_m = \frac{B_{gap} l_g}{OF \cdot \mu} = \frac{\mu H_{gap} l_g}{OF \cdot \mu} = \frac{H_{gap} l_g}{OF} = \frac{(45\ kA/m)(0.17\ m)}{517\ kA/m} \tag{3.12}$$

$$= 14.8\ mm \tag{3.13}$$

This would give, from equation 3.11,

$$a_m = \frac{B_{gap} a_g}{OJ} = \frac{\mu(45\ kA/m)(2078\ mm^2)}{0.65\ T} = \mu_r \cdot 181\ mm^2 \tag{3.14}$$

$$= \begin{cases} 815\ mm^2 & \text{if } \mu_r = 4.5, \\ 1810\ mm^2 & \text{if } \mu_r = 10. \end{cases} \tag{3.15}$$

Space limitation makes this choice not feasible, and another operating point is sought. A smaller a_g means longer magnets, and in order to simplify mechanical design, a length of 90 mm is chosen. Equation 3.12 now gives

$$OF = \frac{(45\ kA/m)(0,17\ m)}{90\ mm} = 85\ kA/m = 1069\ Oe$$

and from figure 3.27, the corresponding B_{gap} is about 1.22 Tesla. Equation 3.11 gives

$$a_m = \frac{\mu(45\ kA/m)(2078\ mm^2)}{1.22\ T} = \mu_r \cdot 96.3\ mm^2$$

$$= \begin{cases} 433\ mm^2 & \text{if } \mu_r = 4.5, \\ 963\ mm^2 & \text{if } \mu_r = 10. \end{cases}$$

A medium value, $a_m = 800\ mm^2$ is chosen.

3.8.2 Confirming the Results

In order to verify the calculations, a finite element analysis (FEA) is performed. Due to the lack of a suitable 3-D package here at UBC, a 2-D program[7] is used. Since the problem now has to be adapted to a reduced dimension, the accuracy is expected to suffer. However, it should still give an indication of how well the circuit will perform.

The whole magnetic circuit is transformed into a continuous 2-D model with the appropriate changes in dimensions. Results of the analysis are shown below. Figure 3.29

[7]The program is a shareware for DOS called ELCUT, and it can be obtained from the archive site ftp://OAK.Oakland.Edu:/pub/msdos/plot.

shows a mesh plot of the H-field inside one rod. The y-axis is in a direction along the length of the rod, while the x-axis represents the lateral direction. Except at the two corners, the field inside the rod is relatively constant. This is important because any variation will not only reduce the expanding portion, but the material expanding less will also act as a load, further reducing the overall expansion. A μ_r of 7 is used because it represents the average of the possible relative permeability for Terfenol-D. In figure 3.30, a y-z projection of the same plot is shown.

The results of this analysis agree quite well with the calculations in the previous section.

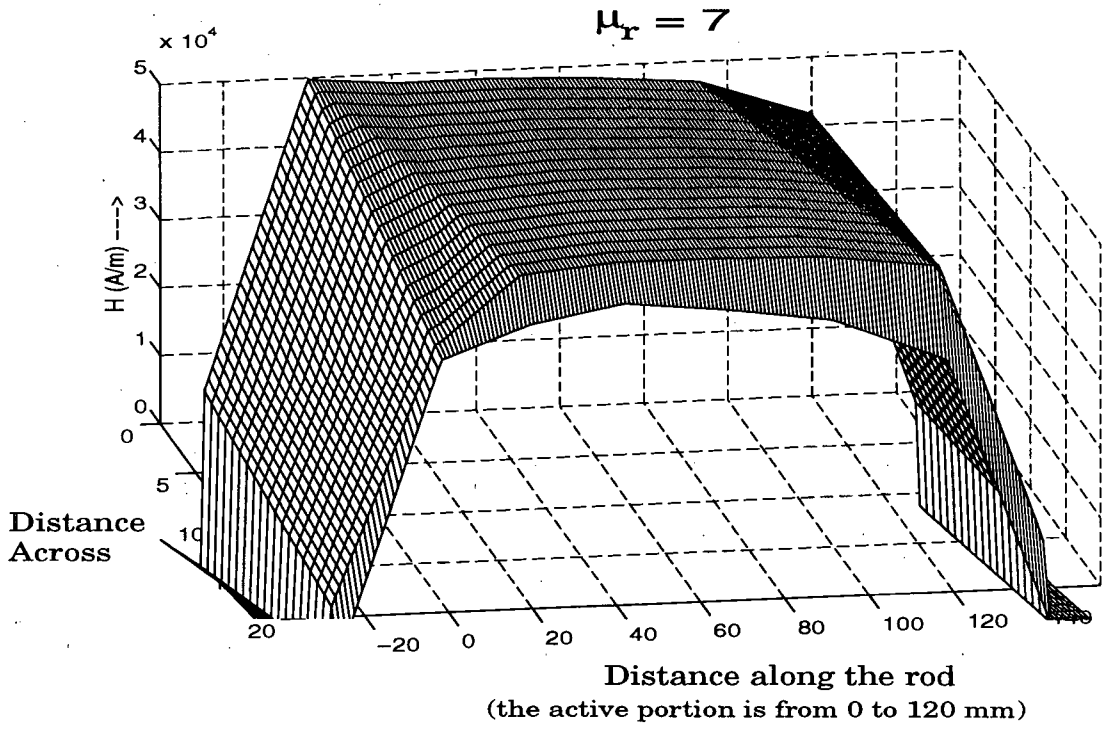


Figure 3.29: Distribution of magnetomotive force (mmf) inside the Terfenol-D rod (obtained by FEA).

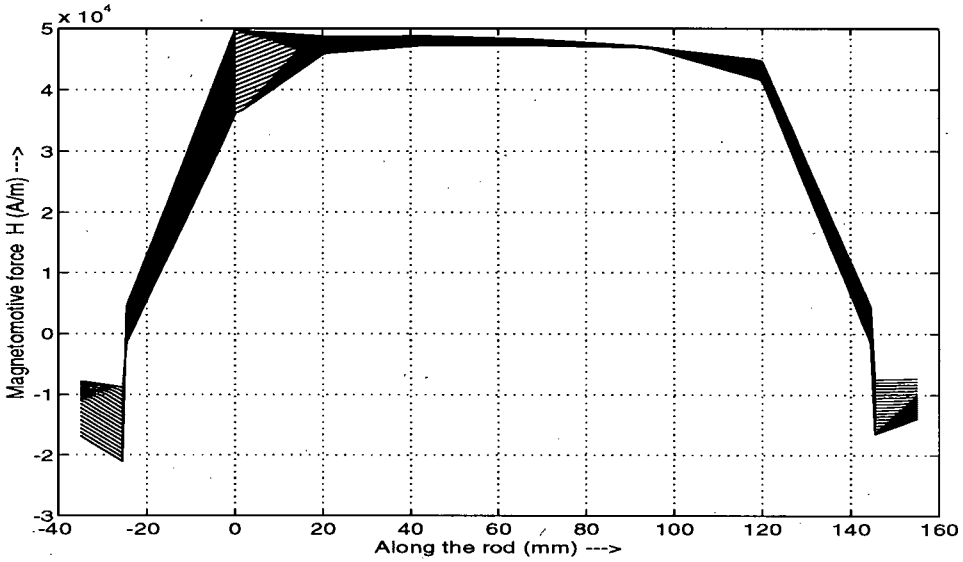


Figure 3.30: Side view of the magnetomotive force (mmf) inside the Terfenol-D rod (obtained by FEA).

3.8.3 Coils

Outside diameter of the coils is chosen to be 75 mm. Using a simple spreadsheet program supplied by Etrema, the following data for coils wound with different wire sizes are obtained.

AWG	10	12	14	16	18
R (Ω)	0.22	0.57	1.43	3.63	9.17
L (mH)	0.007	0.018	0.045	0.114	0.287
Field (A/m @ 1A)	3170	5040	8014	12742	20261

The heat generated by each of these coils for the same H-field is roughly the same. Therefore, the choice of wire size relies solely on which one allows faster build-up of current. This is important because the faster the fields get established, the faster the rods response. For a power amplifier with a maximum output voltage V_o driving an

initially relaxed coil, the quickest way to build-up current is to output V_o from t = 0 until $H_{desired}$ is reached. Under these circumstances, $i = \frac{V_o}{R}(1 - e^{-Rt/L})$, where i is current flowing through coil. It is found that bigger wires (eg., AWG 10) are better in this regard. For example, a coil with AWG 10 wire takes 0.221 msec to establish a current for H = 10 kA/m when driven by a 100 V source, while one with AWG 15 wire requires 0.721 msec.

In order to increase packing density and thermal conductivity, ribbon wire with rectangular cross-section is chosen. The resulting coils have 462 turns each, and would produce 3850 A/m of H-field for each ampere of current. Full range output requires about 11 to 12 A of current input. Although only a relatively low voltage is required to sustain this current inside the coil, a high voltage servo amplifier is still desired, because it would allow quick build up of current.

3.9 Position sensing

The actuator is capable of delivering sub-micron positioning accuracy and repeatability if the output position is measured with enough precision. A sensor based on a position sensing device (PSD) is developed by Dr. Chris Ma here at UBC for this application. It uses a lens to optically increase the sensitivity of the original PSD, and a novel high-gain amplification circuitry is also developed.

Basic structure of the sensor is shown in figure 3.31. The displacement to be measured is input by moving the lens vertically. This movement in turn causes the focused light spot to move on the PSD sensing surface. This movement is then translated into an electrical signal. The main reason for using a lens instead of, say, just a pin-hole is the increase in intensity of light available to the PSD. Therefore higher signal-to-noise ratio is realized. In addition, the alignment of the lens is not critical, since misalignment would

not affect the point of focus.

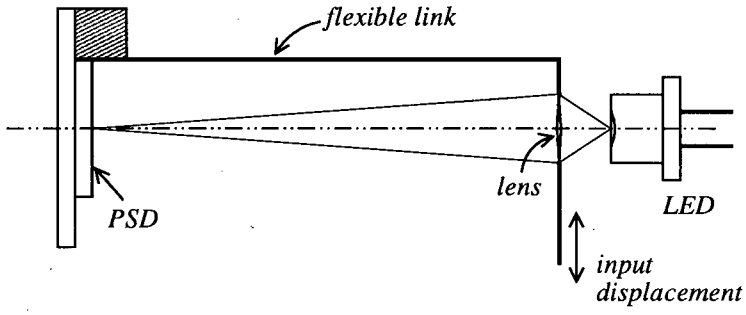


Figure 3.31: A high precision sensor based on PSD.

Chapter 4

Experimental Results

The proposed actuator has been build in-house here at the University of British Columbia. In this chapter, some of the results obtained from preliminary testings are presented.

4.1 Equipment Setup

General setup of the equipment for the experiments discussed in the following sections are shown in figure 4.32. A 486-based PC is used as the host computer, with all the external interface and computation of the control signal handled by a TMS320C30-based DSP board; the main CPU is responsible only for user interface and disk I/O. The reason for this arrangement is to allow proper operation under OS/2, a multitasking and non-realtime operating system.

The PWM servo amplifier is put into current mode, for which the output current is proportional to the command signal. This simplifies requirements of the controller, since the amplifier automatically provides maximum voltage to drive the current into the actuator when the difference between commanded and actual current is big. It does, however, make the overall response susceptible to internal dynamics of the amplifier.

Current flowing into the actuator (I_d) is measured by a LEM current sensor. This sensor generates an output voltage (I_a) proportional to the current going through it. Displacement of the actuator caused by this current is measured by the position sensor discussed in the previous chapter. This sensor is calibrated using a precision dial gauge.

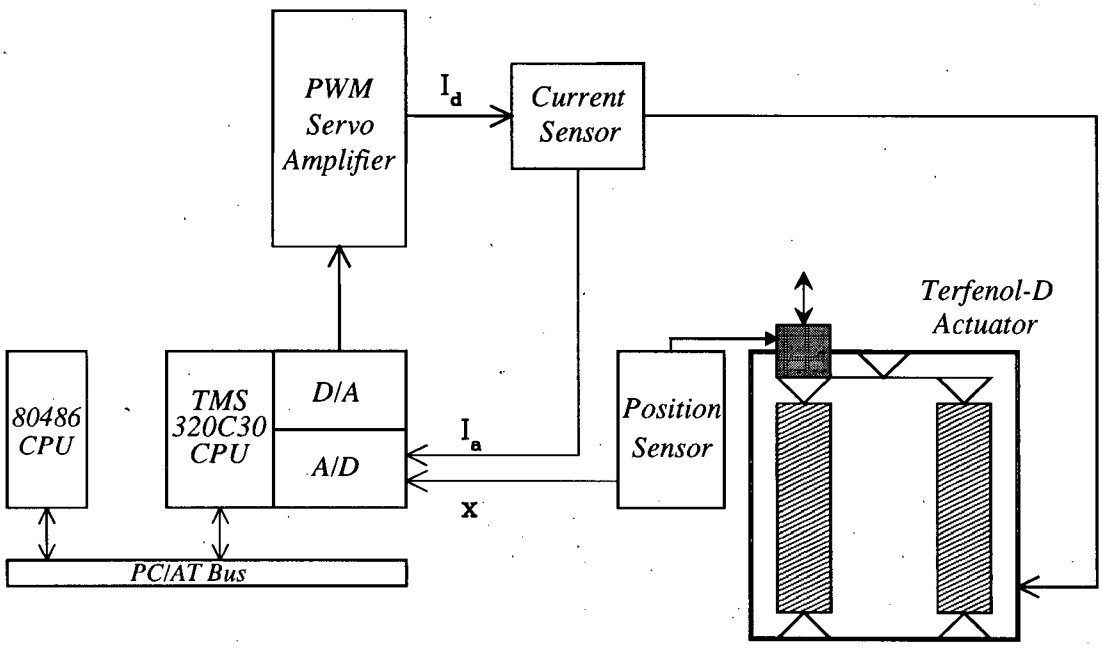


Figure 4.32: Equipment setup for the experiments discussed in this chapter.

Although this mechanical calibrating device does provide a crude reference, a more accurate method is much desired (eg. an interferometer-based device). Gain of the sensor has been set to 0.169 $V/\mu m$.

To estimate the resolution of the position sensor, one can look at the noise level of its output. Figure 4.33 and 4.34 show the raw data from the ADC when the power amplifier is turned off, which means the output of the sensor should be zero. It is apparent that the ADC itself is being affected by some noise source other than the sensor, and the reason for this is still under investigation. By comparing the two plots, it is difficult to estimate any effects, if any, that the sensor has on the ADC readings. Therefore, the effectiveness of the sensor is now limited by the analog-to-digital conversion and can obviously be improved. Eliminating the source of noise is obviously the best solution, but filtering in the digital and/or analog domain should also help immensely.

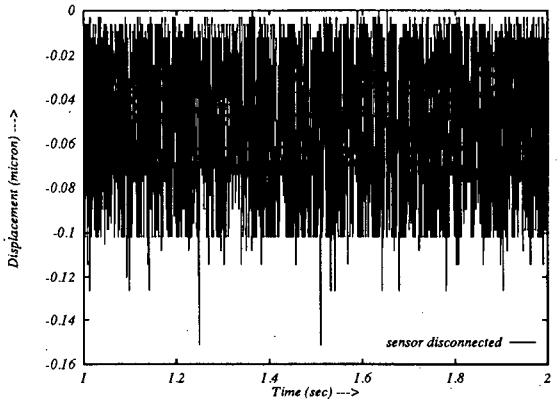


Figure 4.33: Readings of the ADC with sensor disconnected.

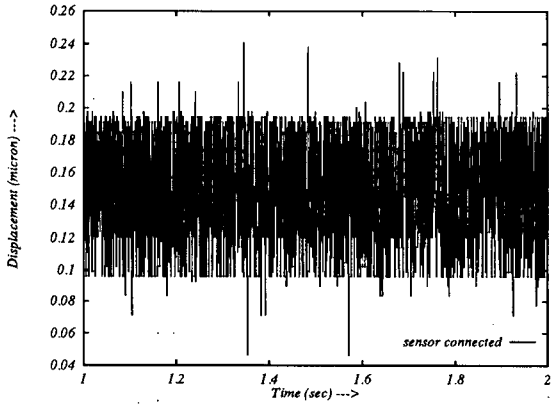


Figure 4.34: Readings of the ADC with sensor connected.

4.2 DC Response

Since Terfenol-D possesses significant amount of hysteresis, it is important to determine the severity of this property. By applying a slowly varying excitation, the output displacements are recorded and plotted in figure 4.35. The amount of hysteresis observed is

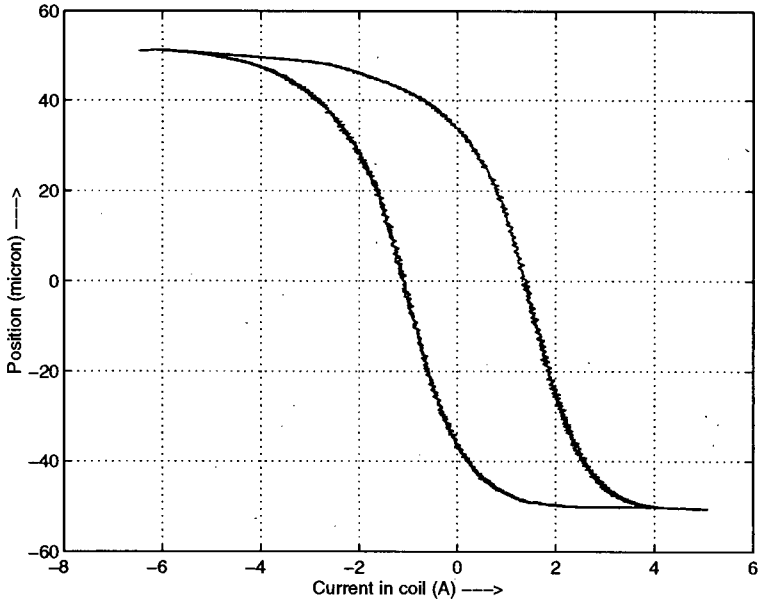


Figure 4.35: Response of the push-pull actuator (note the hysteresis loop).

roughly the same as that for a single rod (test result supplied by Etrema Inc. is shown in figure 4.36). However, the response is more symmetric, and this should make precise position control of the actuator easier than a single rod design.

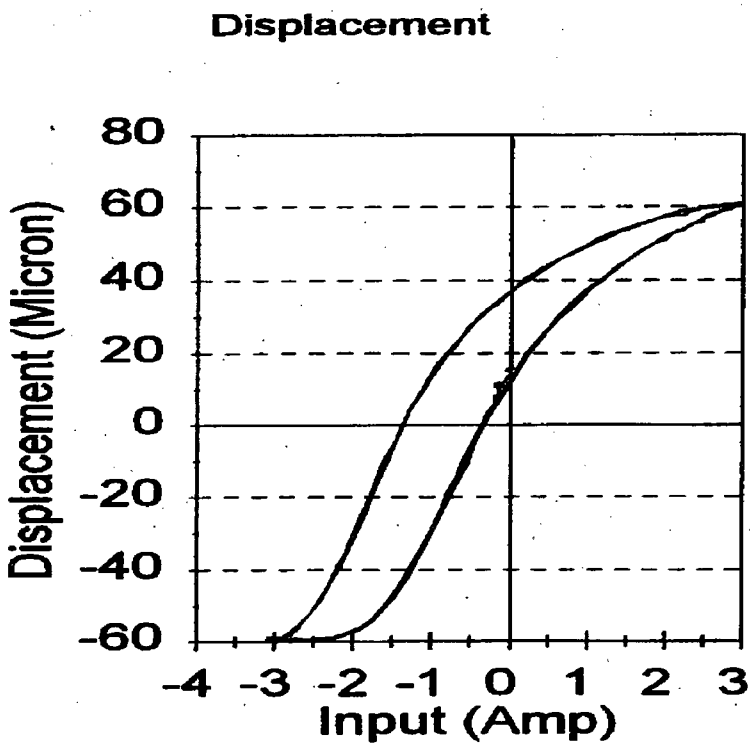


Figure 4.36: Test result for one of the Terfenol-D rod (supplied by Etrema Inc.).

Range of the actuator, as can be seen in figure 4.35, is about 100 μm. While this meets the design goal, it is felt that improvement is possible. One area that needs further investigation is the magnetic circuit, because insufficient magnetic bias would hamper the performance of the actuator. Since there is no convenient way to measure magnetic field strength inside the Terfenol-D rods, more testings have to be done to determine whether the magnetic circuit is performing as designed. Furthermore, the load cell used inside the actuator has a full scale deflection of 0.001" to 0.003" (25 μm to 75 μm); therefore,

it is possible to recover some lost range by improving the stress-sensing mechanism.

4.3 Bode Plots

The actuator is nonlinear because of hysteresis and saturation; however, Bode plots obtained by small excitations should still provide valuable information about characteristics of its frequency response. Figure 4.37 shows the results of four trials in measuring the frequency response. The excitations are kept small (about 20% of full range) to prevent the actuator from saturating.

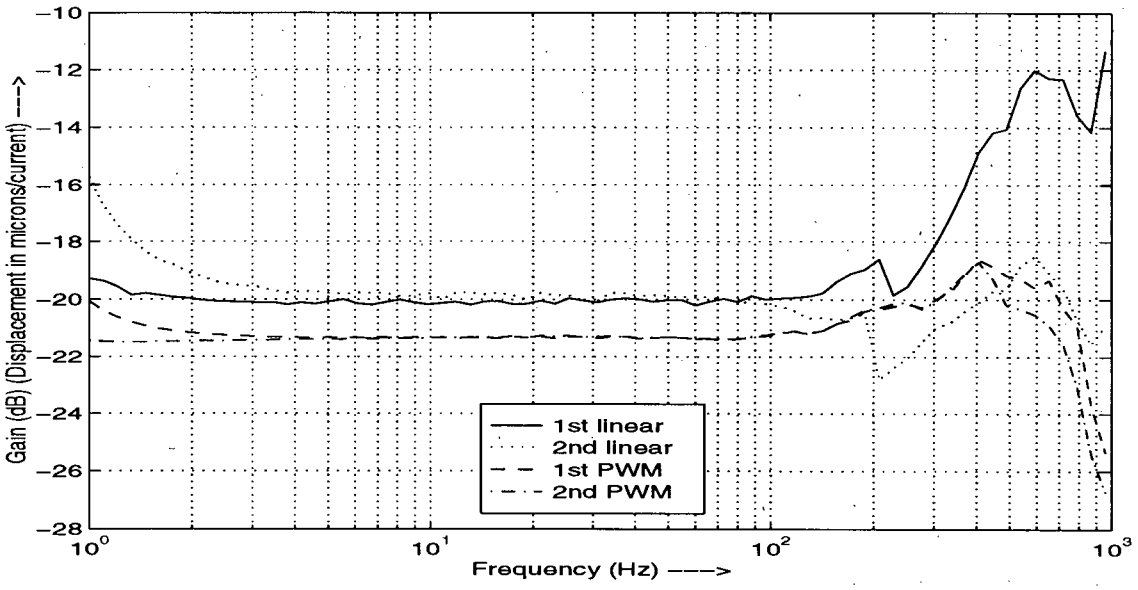


Figure 4.37: Frequency response plot of the actuator. Two runs for each of the analog and PWM amplifier are performed.

Two of the lines are obtained by using an analog audio amplifier; the other two by a PWM servo amplifier. Overall shapes of the Bode plots are rather consistent: relatively flat from dc up to about 100 Hz; a peak at 400 Hz to 600 Hz; roll-off at higher frequencies. However, there are discrepancies between individual plots. One possible reason for these variations are nonlinearities: a change in the initial magnetic state of the Terfenol-D rods

would change the small signal response of the actuator. Therefore when designing the controller, it is important to track or estimate the internal magnetic states of the device.

4.4 Time Responses

A simple PID controller is implemented. The control effort (current), u, at time k, is obtained by

$$u(k) = u(k-1) + (K_P + K_I + K_D)e(k) + (-K_P - 2K_D)e(k-1) + K_D e(k-2)$$

where $K_P = 0.0017$, $K_I = 0.0010$, and $K_D = 0.0004$ are the coefficients for the P, I, and D term respectively, and $e(k)$ is the position error at time k. The coefficients are obtained by tuning the controller for minimum rise time with no overshoot.

4.4.1 No-load response

Time response to step input with no-load attached is shown in figure 4.38. Rise time is quite satisfactory, but there is slight "rounding off" when the set point is far away from zero. This may have been due to saturation of the actuator.

4.4.2 Response under external disturbance

Since external disturbance is expected to be present when the actuator is used in actual applications, the above experiment is repeated with a varying load (figure 4.39). Due to the lack of a better source, a human hand is used to effect the changing load. Range of the disturbance is estimated to be from 0 to about 100 N. Although this force is much smaller than that expected in actual metal cutting processes, the steady-state output is already being affected by up to 1 μm. Theoretically, a force of 100 N should cause a deflection of

$$\Delta x = \frac{(length)(force)}{(Young's\ modulus,\ E)(Area)}$$

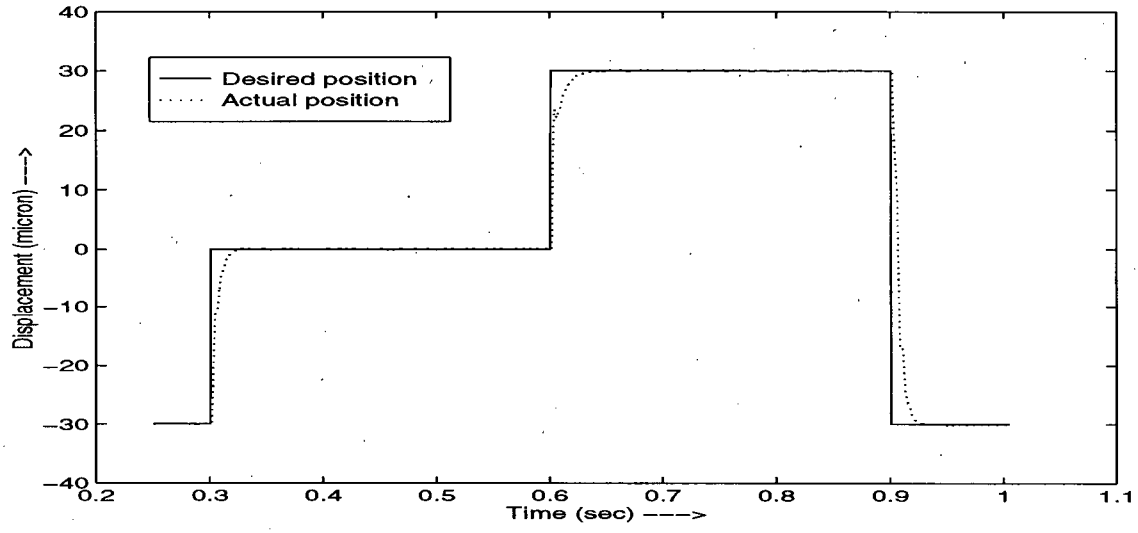


Figure 4.38: Time response of the actuator under a PID controller (with no disturbance).

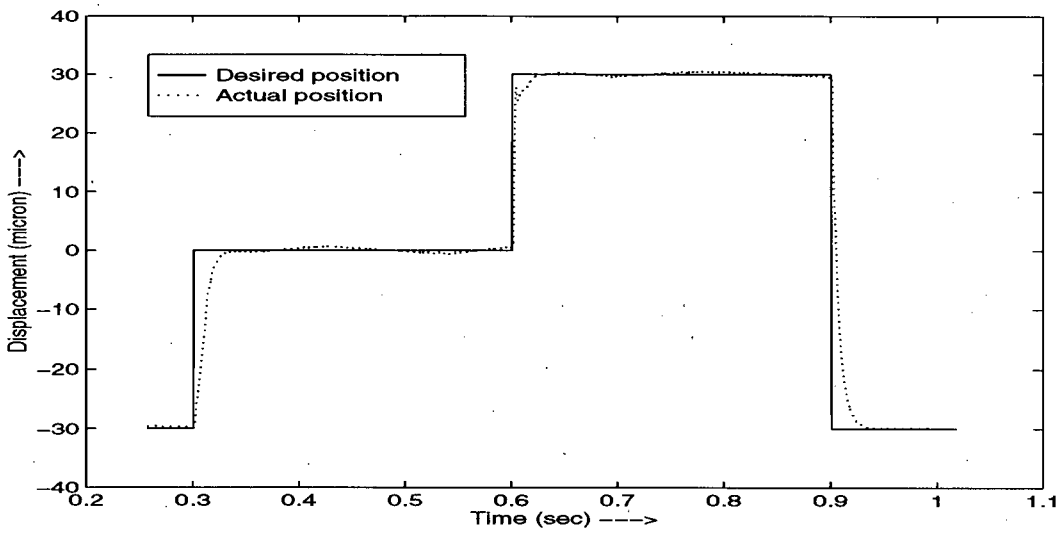


Figure 4.39: Time response of the actuator under a PID controller (with disturbance).

With $length = 0.17\ m$, $Area = 2 \times \pi(0.0105\ m)^2 = 692 \times 10^{-6}\ m^2$ and $E = 2.5 \times 10^{10}\ N/m^2$,

$$\Delta x = 0.98 \mu m,$$

which is comparable to what's being noticed. Therefore the PID controller is not very effective in rejecting, at least this kind of, disturbance.

4.4.3 Response under load

In order to get an idea of how well the actuator would perform under heavy load, an experiment is performed with a mass sitting on top of the actuator. If the full force of 2000 N as described in the specification is to be tested, a mass of about 200 kg (or 440 lb) is needed. Not only is this difficult to handle, it is also quite dangerous[1]. Therefore, a lesser mass is used. With a 70 kg weight resting on top of the actuator, the step response is shown in figure 4.40. Other than some vibration around 0 μm, this response is quite similar to that for no-load.

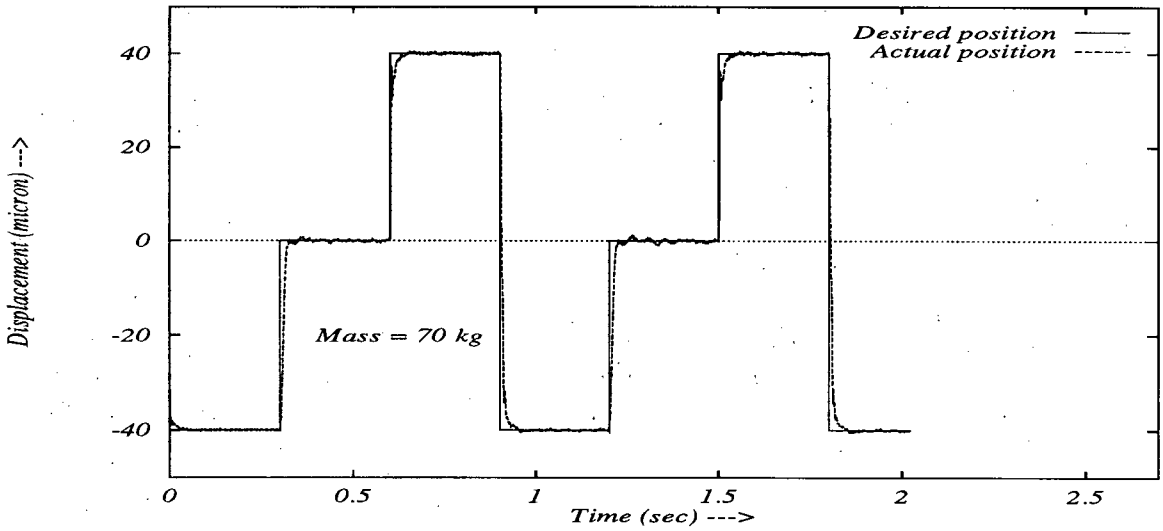


Figure 4.40: Step response of the actuator with a 70 kg mass attached.

[1] And heavy too.

4.5 The Need for a Better Controller

From the above observations, it is obvious that a PID controller alone is not adequate to ensure high-precision position tracking that is essential in the control of chatter vibration and non-circular turning. In the following chapter, a few alternate approaches are proposed.

Chapter 5

Modelling and Control

Due to time limitation, full development and implementation of a suitable control strategy for the newly developed Terfenol-D actuator is not possible. However, a few promising candidates have been studied and are discussed below.

5.1 Modelling

The simplest model for Terfenol-D is to view it as linear and proportional. That is, $\varepsilon = kH$, where ε is the strain, H the applied magnetizing field, and k a proportional constant. This approach is used in quite a number of studies (eg., [32], [35] and [31]), and when used with output feedback, can work quite effectively. However, some intrinsic properties of Terfenol-D render this inadequate when the output strain is pushed to the limit. Figure 2.11 shows a typical, but somewhat exaggerated, characteristic curve of a Terfenol-D rod. The most noticeable features of this "butterfly" curve are the hysteresis and saturation; they are both related to the process of magnetization. The exact shape of this curve is also affected by the external stress applied on the rod. Therefore, the output strain at any given time is determined by at least three factors: the external H-field, the external stress σ and the past history of output (due to hysteresis); a proportional model would be too simplistic.

In the following, some attempts to model Terfenol-D more accurately are described. They include a sixth-order linear model, a simple but elegant mathematical model, and

a promising approach using neural networks[1].

5.2 A linear model

A sixth-order linear model for Terfenol-D actuators in audio range has been developed by Bryant and Wang [45]. It accommodates physical effects such as inverse magnetostrictive effect and back emf by including feedback loops in the model. It also claims that the interaction between the first and second vibrational modes gives rise to a non-minimum phase zero, thus making active control over an extended bandwidth difficult.

Figure 5.41 shows the block diagram of the model. The solenoid block $\frac{1}{Ls+R}$ converts the input voltage V_{in} into current in the coil, i. This current then gets multiplied by K_h to generate magnetization H. The block $\frac{K_e}{s+\omega_e}$ is used to represent the effects of magnetic energy storage and loss associated with flux inside the Terfenol-D rod. The resulting flux causes domain rotations and magnetostrictive expansion (K_f). This mechanical expansion generates a force (F), and together with that applied externally (F_d), eventually excite the fundamental mode (top block) and the second harmonics (bottom block) of the rod; the output is taken as the sum of the two responses. The output x in turn affects the magnetic flux and input voltage through the feedback loops, representing the inverse magnetostrictive effect and back emf.

After simplification, the transfer function between inputs and output is given by

$$x = \frac{B}{sBAK_fK_{xv} + (1 - BK_fK_{x\phi})} F_d + \frac{BAK_f}{sBAK_fK_{xv} + (1 - BK_fKx\phi)} v_{in} \tag{5.16}$$

$$= C\,F_d + D\,v_{in} \tag{5.17}$$

[1] Xu and Refsum have made some brief comparisons between five models for systems with hysteresis. See [44].

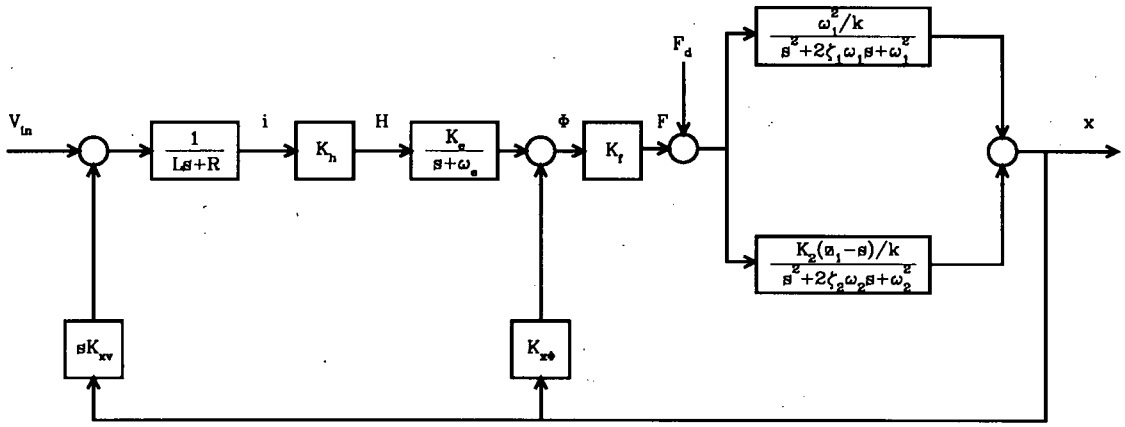


Figure 5.41: Block diagram of the linear model proposed by Bryant and Wang (after [45]).

where

$$A = \frac{K_H K_e / L}{s^2 + s(R/L + \omega_e) + R\omega_e / L} \tag{5.18}$$

and

$$B = \frac{-K_2 s^3 + s^2(\omega_1^2 + K_2 z_1 - 2\xi_1\omega_1 K_2) + s(2\xi_2\omega_2{\omega_1}^2 + 2\xi_1\omega_1 K_2 z_1 - \omega_1^2 K_2) + (\omega_1^2\omega_2^2 + \omega_1^2 K_2 z_1)}{k(s^2 + 2\xi_1\omega_1 s + \omega_1^2)(s^2 + 2\xi_2\omega_2 s + \omega_2^2)} \tag{5.19}$$

The numerous coefficients are determined experimentally either by direct measurements (eg., R and L), or from plots of frequency response between x and the various inputs (i, v_{in} and ϕ). Basically, the model is adjusted to fit the measured bode plots. For the rod used in the paper, the frequency response plot using the proposed model is shown in figure 5.42[2]. The model agrees reasonably well with the measured data. However, it should be noted that the essence of this model is still the K_f block, which assumes linear relationship between magnetostrictive expansion and flux. Maybe adequate for

[2]There is a minor numerical error in the original paper: the simplified transfer function is twice as big as what it should be.

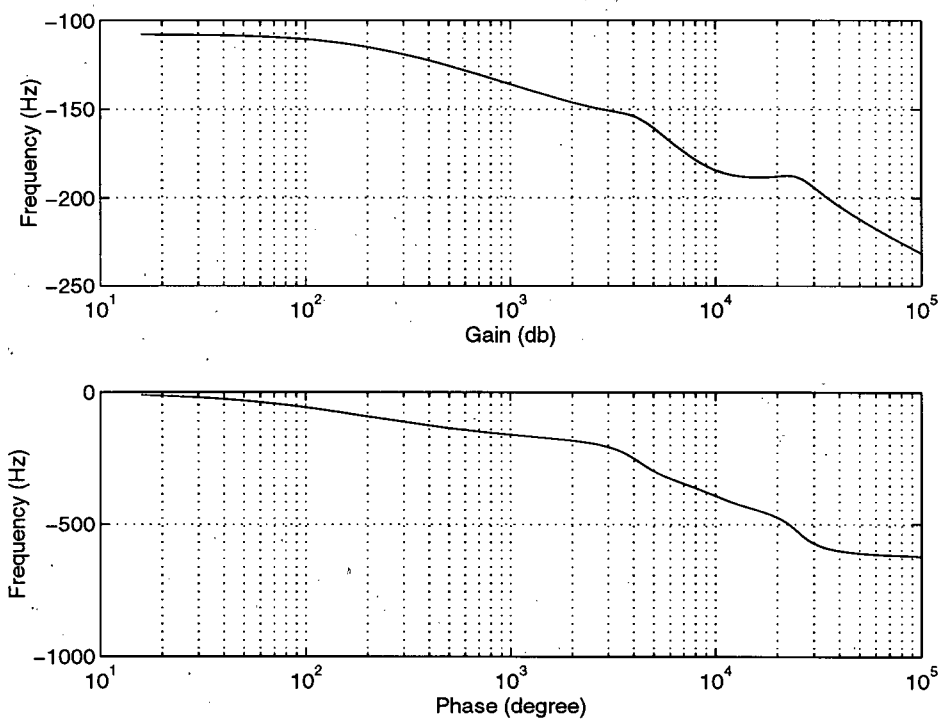


Figure 5.42: Bode plot of the transfer function between x and v (linear model by Bryant and Wang).

small displacement, it is obviously not accurate enough if the full motion range is to be realized.

Some modifications are required to adapt this model to the UBC design. The two rods and the lever in between can be treated as the simplified system shown in figure 5.43. The rods generate forces F_{d_1} and F_{d_2}, which excite the lever M_o through two flexure plates represented by spring-dashpot systems with spring constants k and damping coefficient c. An external force F also acts on M_o. Using Newton's 2nd law,

$$F_{d_2} - F_{d_1} - F = M_o \ddot{x}_o$$

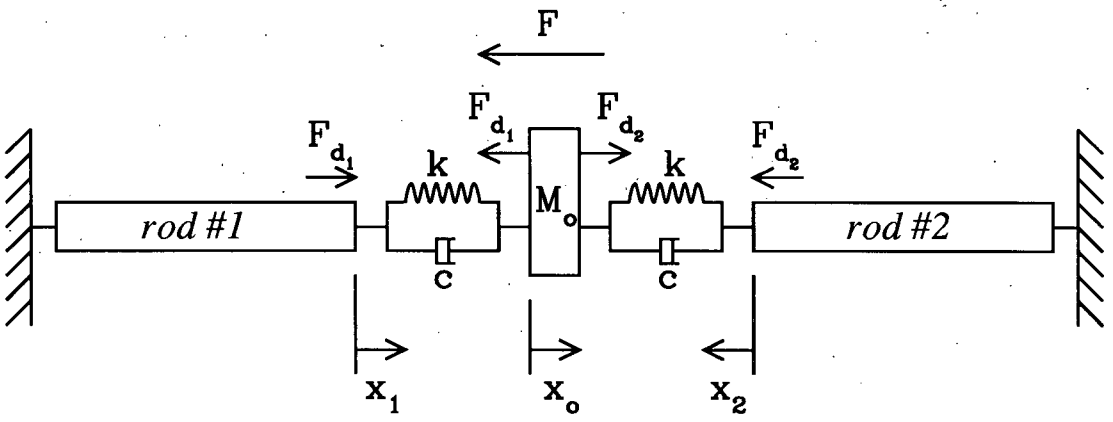


Figure 5.43: Simplified layout for simulation of the UBC design.

Now

$$F_{d_1} = k(x_o - x_1) + c(\dot{x}_o - \dot{x}_1) \Rightarrow F_{d_1} = (sc + k)x_o - (sc + k)x_1$$

$$F_{d_2} = k(-x_o - x_2) + c(-\dot{x}_o - \dot{x}_2) \Rightarrow F_{d_2} = -(sc + k)x_o - (sc + k)x_2$$

$$\therefore \quad k(x_o - x_1) + c(\dot{x}_o - \dot{x}_1) - k(-x_o - x_2) - c(-\dot{x}_o - \dot{x}_2) = -M_o\ddot{x}_o - F$$

or

$$M_o\ddot{x}_o + 2c\dot{x}_o + 2kx_o = k(x_1 - x_2) + c(\dot{x}_1 - \dot{x}_2) - F$$

$$\Rightarrow (s^2 + \frac{2c}{M_o}s + \frac{2k}{M_o})x_o = \frac{1}{M_o}(cs + k)x_1 - \frac{1}{M_o}(cs + k)x_2 - F \cdot \frac{1}{M_o}$$

$$\Rightarrow \boxed{X_o = \frac{(cs + k)\frac{1}{M_o}}{s^2 + \frac{2c}{M_o} + \frac{2k}{M_o}}(x_1 - x_2) - \frac{\frac{1}{M_o}}{s^2 + \frac{2c}{M_o} + \frac{2k}{M_o}}F} \tag{5.20}$$

Since data are not yet available for the rods actually used, numbers from the paper are used to perform simulations. Matlab[3] function linmod is used to find the overall state-space representation from a Simulink model (shown in figure B.63). Figure 5.44 is an example with $k = 200 \times 10^6$ and $c = 0.2\sqrt{2k}$, and $M_o = 1$ kg.

[3]Matlab and Simulink are computer programs from The Mathworks Company. Version used is 4.2.

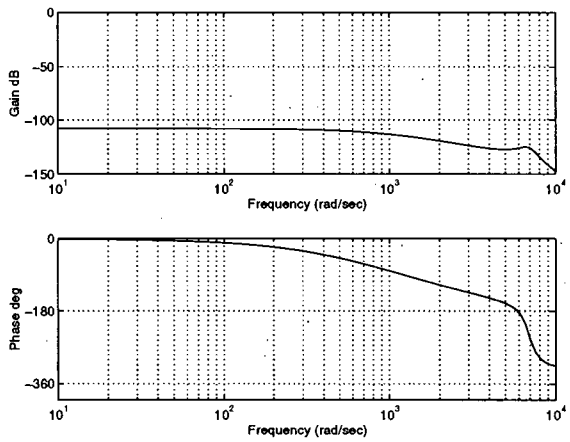


Figure 5.44: Bode plot for overall system with $M_o = 1$ kg.

5.3 The Preisach Model

Since Terfenol-D exhibits a fair amount of hysteresis nonlinearity in its magnetoelastic strain, a mathematical model originally proposed by F. Preisach in 1935 [46] can be very useful in describing it. The classical Preisach model is based on some hypotheses concerning the physical mechanisms of magnetization. Although originally It contains some intrinsic limitations, generalizations have been attempted to remove or relax some of them[4]. Basically, the approaches taken are to describe and generalize experimental facts using purely mathematical models; therefore there isn't much theoretical support from the physics point of view. Nevertheless, they can be powerful tools for designing magnetic devices. Due to the complexity involved, only the classical Preisach model of hysteresis is described here.

The hysteresis nonlinearity in Terfenol-D can be classified as hysteresis with nonlocal memories [17]. That is, future values of the output $f(t)$, $(t \geq t_o)$ depend not only on the current value of output $f(t_o)$ and future input values $u(t)$, but also on past extremum values of input as well. Note that only the extrema are of importance, not the whole

[4]See chapter II of [47] for a discussion of the generalization.

history of input. These extrema would determine the appropriate *branch* for the output to follow, and branching can happen at any point with a change in direction of movement.

Consider an infinite set of the simplest hysteresis operators $\hat{\gamma}_{\alpha\beta}$. Each of these operators can be represented by a rectangular loop on the input-output diagram (figure 5.45). The numbers α and β correspond to the "up" and "down" switching values respectively.

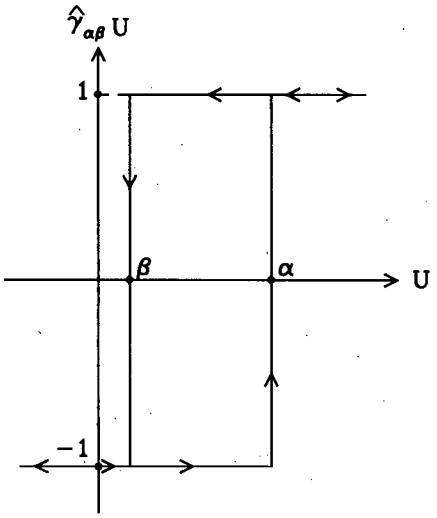


Figure 5.45: The simplest hysteresis operator $\hat{\gamma}_{\alpha\beta}$.

Output of this operator can assume two values only: 1 and -1. Now consider an arbitrary weight function $\mu(\alpha, \beta)$, which is often referred to as the Preisach function, and the Preisach model can be written as

$$f(t) = \hat{\Gamma}u(t) = \iint\limits_{\alpha \geq \beta} \mu(\alpha, \beta)\hat{\gamma}_{\alpha\beta}u(t)\, d\alpha\, d\beta \tag{5.21}$$

Here $\hat{\Gamma}$ denotes the Preisach hysteresis operator on $u(t)$. This model can be viewed as a set of two-position relays connected in parallel. The same input is connected to all of them, and each has different switching thresholds. The individual output, multiplied by $\mu(\alpha, \beta)$, are summed to generate the output of the model. In a sense this model can be

interpreted as a decomposition of a complicated hysteresis operator $\hat{\Gamma}$ into many simpler operators $\hat{\gamma}_{\alpha\beta}$.

There is a one-to-one correspondence between operators $\hat{\gamma}_{\alpha\beta}$ and points (α, β) of the half plane $\alpha \geq \beta$ (on a plot of β versus α). In other words, each point of the half-plane $\alpha \geq \beta$ can be identified with only one particular $\hat{\gamma}$-operator whose "up" and "down" switching values are respectively equal to α and β coordinates of the point.

Now consider a right-angle triangle T (figure 5.46). Its hypotenuse is a part of the line $\alpha = \beta$, while the vertex of its right angle has the coordinates α_o and β_o with $\beta_o = -\alpha_o$. This triangle will be called the limiting triangle and it is assumed that the function $\mu(\alpha, \beta)$ is equal to zero outside this area.

The output of a system with hysteresis can now be traced by looking at the difference between numbers of "up" and "down" operators. First assume at some point of time t_o, the input is at a value smaller than β_o; therefore all $\hat{\gamma}_{\alpha\beta}$ operators are negatively saturated and have output of -1. Output of the model becomes $(-1)\times$(total number of $\hat{\gamma}_{\alpha\beta}$ operators). Now assume the input is monotonically increased until it reaches some value u_1. All $\hat{\gamma}_{\alpha\beta}$ with α smaller than u_1 will now have switched to $+1$. Geometrically, it leads to the division of T into two parts: S^+ consists of points (α, β) for which the corresponding operators $\hat{\gamma}_{\alpha\beta}$ are in the "up" position, and S^- for those corresponding to "down" position. This subdivision is made by the line $\alpha = u(t)$ (figure 5.47), which moves upwards as the input is being increased. The output will now be less negative than before.

Next, the input is monotonically decreased from u_1 to some minimum value u_2. Movement of the interface is now in the horizontal direction. All $\hat{\gamma}_{\alpha\beta}$ operators with "down" switching value β larger than u_2 will be turned back to the "down" position. The new $S^+(t)$ and $S^-(t)$ areas are shown in figure 5.48.

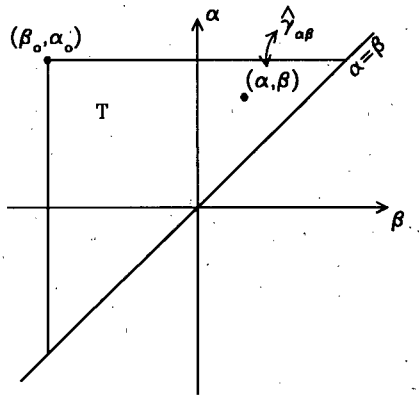


Figure 5.46: Negatively saturated.

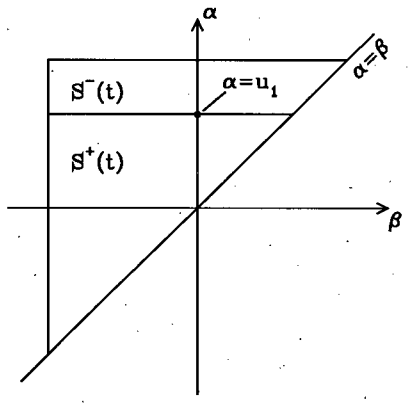


Figure 5.47: Input increased to u_1

Any further changes in the input are dealt with in a similar fashion. Rising inputs create an upward moving horizontal portion of boundary between S^- and S^+, while falling inputs cause the formation of a new vertical link which moves from right to left. Therefore at any instant of time, the triangle T is subdivided into two sets: $S^+(t)$ and $S^-(t)$. The interface $L(t)$ in between is a staircase line whose vertices have α and β coordinates coinciding respectively with local maxima and minima of input at previous instances of time. An example interface is shown in figure 5.49, with the corresponding input shown in figure 5.50. Note that this interface line is always a monotonically decreasing function

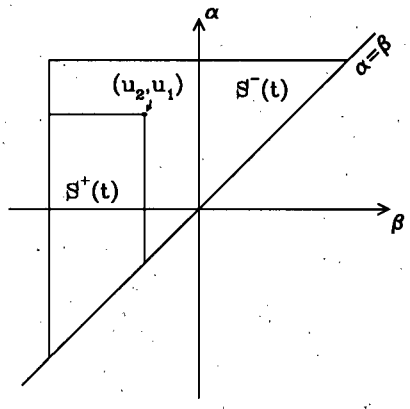


Figure 5.48: Decreased to u_2.

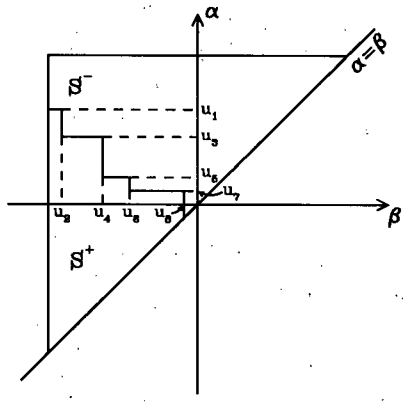


Figure 5.49: A typical boundary.

from left to right and is single-valued.

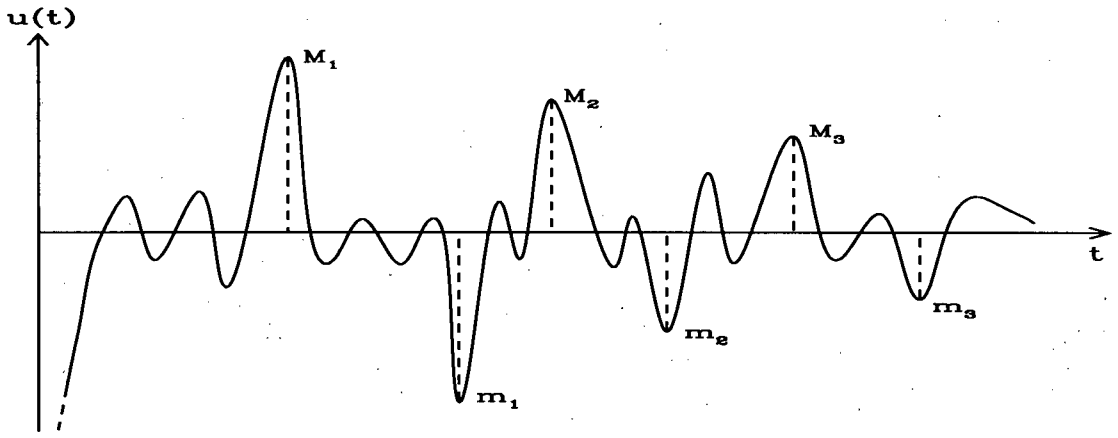


Figure 5.50: Input history for the interface line shown in figure 5.49.

The integral in equation 5.21, according to the above mentioned, can be separated into two parts.

$$f(t) = \hat{\Gamma}u(t) = \iint\limits_{S^+(t)} \mu(\alpha,\beta)\hat{\gamma}_{\alpha\beta}\, d\alpha\, d\beta + \iint\limits_{S^-(t)} \mu(\alpha,\beta)\hat{\gamma}_{\alpha\beta}\, d\alpha\, d\beta \tag{5.22}$$

Since

$$\hat{\gamma}_{\alpha\beta}u(t) = +1 \qquad \text{if}(\alpha,\beta) \in S^+(t) \tag{5.23}$$

and

$$\hat{\gamma}_{\alpha\beta}u(t) = -1 \qquad \text{if}(\alpha,\beta) \in S^-(t), \tag{5.24}$$

therefore

$$f(t) = \iint\limits_{S^+(t)} \mu(\alpha,\beta)\, d\alpha\, d\beta - \iint\limits_{S^-(t)} \mu(\alpha,\beta)\, d\alpha\, d\beta \tag{5.25}$$

That means an instantaneous value of output depends on a particular subdivision of the limiting triangle T into positive and negative sets $S^+(t)$ and $S^-(t)$. This subdivision is

determined by a particular shape of interface $L(t)$, which in turn depends on the past extremum values of input. This is how the model *memorizes* the history of input, and it is formed as a result of two different rules for the modification of the interface $L(t)$.

Two major features of the classical Preisach model limit its usefulness as a good candidate for modelling of hysteresis. First, the output stays constant when saturated. This happens when S^+ or S^- expands to the whole triangle T. Any changes in $u(t)$ will not affect the output at all if $u(t) < \beta_o$ or $u(t) > \alpha_o$. This contradicts with observations for, say, magnetization of ferromagnets. Usually, the output would still vary slightly inside the region of saturation (eg., due to the permeability of free space). Second, all minor loops generated by this model corresponding to back-and-forth variations of inputs between the same two consecutive extremum values are congruent. That is, they have the same shape. Again, this is not always observed in the real world.

Dispute the above mentioned weaknesses, the Preisach model still provides a very useful framework within which a good model for a particular system with hysteresis can be constructed. In the case of magnetostrictive materials, hysteretic behaviour is observed in the variations of two variables: magnetic field and applied stress. Extension in this regard has been made by several groups of scholars, including Adly, Mayergoyz and Bergqvist [48]; and Kvarnsjö, Bergqvist and Engdahl [49]. In addition, generalization to include dynamic effect has been proposed by Bertotti [50].

5.4 Modelling Using Artificial Neural Networks

Recently artificial neural networks have been considered for the intelligent control of nonlinear systems, mainly because of their learning and nonlinear modelling capabilities. A neural network is loosely modelled on the complex operation of the brain and is a lattice of simple processors called neurons. These neurons can be interconnected in

many different ways. One of the most common architectures is the multi-layer perceptron (MLP) [51, chapter 4]. This type of network consists of an input layer, a number of hidden layers (typically only one or two hidden layers are used) and an output layer (figure 5.51). The hidden and output layers are made up of a number of neurons (figure 5.52), whereas

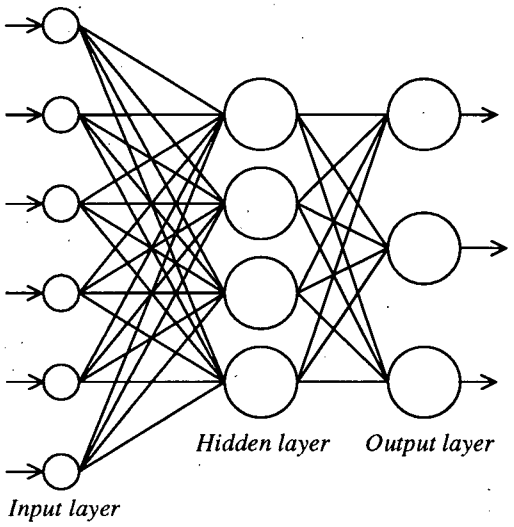


Figure 5.51: Multi-layer perceptron with one hidden layer.

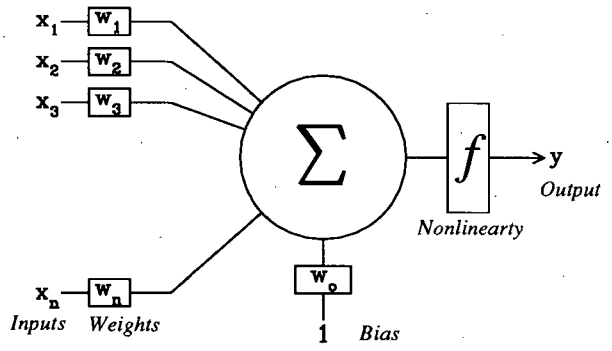


Figure 5.52: Model of a neuron.

the input units are merely "fan-out" units; no processing takes place in these units. The output of each node in a layer is connected to the inputs of all the nodes in the subsequent layer. Each node, except those in the input layer, sums up all the outputs with individual weight from the previous layer. After adding in a local bias, this sum is past through a non-linear function (can be a hard limiter, tanh or sigmoid function) to generate output for the node. Data flows through the network in one direction only, from input to output; hence, this type of network is called a feedforward network.

When the network is used for modelling, the optimum set of weights is obtained by training the network with known input-output data pairs. The most popular algorithm for this purpose is called back-propagation learning rule [52]. It involves two phases.

During the first phase the input is presented and propagated forward through the network to compute the output values for each output unit. The second phase involves a backward pass through the network during which the error signal is passed to each unit in the network and appropriate weight changes are calculated. The directions of adjustment for the weights are obtained by the method of steepest descent: change the weights such that the sum of error square for the outputs decrease most rapidly. The idea is to make a change in weight proportional to the negative of the derivative of the error as measured on the current data pair with respect to each weight. After the weights are trained to satisfactory conditions, a set of new data is used to check whether the resulting model is a good generalization of the data. The basic steps are illustrated in the following example.

Consider figure 5.53. This MLP has one hidden layer with two neurons and the output layer has only one neuron. The nonlinear functions used are $\mathcal{F}$, which can be any monotonic and differentiable function. w_j denotes the weight for each connection. All the biases are assume to be zero. The initial weights could be generated randomly, although this would lead to slow training in sigmoid layers (better methods are available). Define

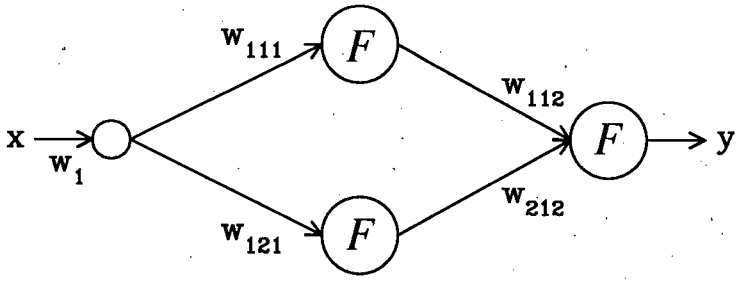


Figure 5.53: A simple example for backpropagation.

the error term

$$E = \frac{1}{2}[y - y_d]^2 \tag{5.26}$$

where y_d is the output in the input-output data pair (x, y_d) supplied, and y is the actual

output from the network. When the network is activated by an input value of x, the output becomes

$$y = \mathcal{F}\big(w_{112} \cdot \mathcal{F}(w_{111} \cdot w_1 x) + w_{212} \cdot \mathcal{F}(w_{121} \cdot w_1 x)\big) = \mathcal{F}(u) \tag{5.27}$$

The error can then be differentiated with respect to individual weights.

$$\frac{\partial E}{\partial w_i} = (y - y_d) \cdot \frac{\partial y}{\partial w_i} \tag{5.28}$$

For example,

$$\frac{\partial y}{\partial w_{111}} = \frac{\partial \mathcal{F}}{\partial u} \cdot \frac{\partial u}{\partial w_{111}} = \frac{\partial \mathcal{F}}{\partial u} \cdot \big(w_{112} \cdot \frac{\partial \mathcal{F}}{\partial (w_{111} w_1 x)} \cdot w_1 x\big) \tag{5.29}$$

The changes in weight are then set to

$$\Delta w_i = -\gamma \frac{\partial E}{\partial w_i} \tag{5.30}$$

where $\gamma > 0$ determines the learning rate. This results in a gradient descent on the error surface, and the weights would move towards a minimum on this surface (for a simpler recursive algorithm for the calculation of Δw_j, see [51, page 34–35]).

Despite the apparent success of the back-propagation learning algorithm, some aspects of this method make the algorithm not guaranteed to be universally useful; the most troublesome being the long training time and the inability to escape local minimum. True gradient descent requires infinitely small learning rate and in practise, the rate is chosen to be as large as possible without leading to oscillation. However, this step size will usually make learning time quite long. Oscillation at large step size can be reduced by including a *momentum* term: make the change in weight dependent of the past weight change, as shown in equation 5.31 [51, page 37] [53].

$$\Delta w_i = -\gamma \frac{\partial E}{\partial w_i} - \beta (w_i^t - w_i^{t-1}) \tag{5.31}$$

where $\beta > 0$ is a factor which is used to influence the degree of this dependence and $(w_i^t - w_i^{t-1})$ is the previous change in weight. Adding this term would also alleviate the problem of getting trapped in a local minimum. With pure gradient descent and small step size, the learning process will stop when the output errors have reached a local minimum: the gradient is zero, and therefore no more learning is needed. The momentum term allows a network to respond not only to the local gradient, but also to recent trends in the error surface. Acting like a low pass filter, momentum allows the network to ignore small features in the error surface. However, this may also prevent the weights from reaching the global minimum, because the momentum could make the weights to move back and forth in the valley of global minimum. Therefore, there is a compromise between the learning rate, chance of oscillation and chance of being trapped in a local minimum. Many scholars have proposed improvements of and extensions to the basic back-propagation algorithm. A survey can be found in [51, page 39].

An increasingly popular optimisation technique known as "simulated-annealing" could be useful for training large networks. It is basically formulated within the "random walk" framework. First initialize the weights with small random values. Present an input and calculate the output, together with the error term E. Then generate a Gaussian distributed random vector and add it to the weights. If the new error function becomes smaller, keep the new weights; otherwise, retain the old values. This process is repeated until the error function converges. The advantages of this algorithm include the elimination of clumsy back-error propagation, possibly faster convergence, potential of avoiding local minima and ease of computation.

In addition, it is possible to use a Newton-like algorithm which includes second derivative information [54]. More computational overheads will be involved, but it is likely to be offset by the improved speed of convergence.

It has been shown [55–58] that multi-layer feedforward networks with only one hidden layer can approximate any function with finitely many discontinuities to arbitrary precision, provided the activation functions of the hidden units are nonlinear. Therefore they can be used as output predictors for nonlinear processes: only a set of examples is needed to generate a model for the underlying structure. If the network is able to generalize correctly, input values which are not presented as learning patterns will still result in correct or reasonable output values. Possible applications include inverse kinematic transformation for a robot arm [59].

As powerful as it is as mapping devices, feedforward networks are static. A given input will generate the same output independent of the time and past input history. This limitation makes them unsuitable for modelling of dynamic systems, in which the outputs depend on both the instantaneous and past input values.

5.4.1 Recurrent Networks as Dynamic Models

Feedback, also known as recurrent, neural networks, introduced by Hopfield [60], provide an alternative neural network structure useful for dynamic systems. A variation of this model consists of two layers: an input layer and a Hopfield layer (figure 5.54). The nodes in the latter layer are neuron models previously described (figure 5.52) with either hard limiting or sigmoidal activation functions. The major difference from a feed-forward network is that the outputs of the Hopfield nodes are weighted and fed back to the inputs of all of the other nodes.

An adaptation of recurrent networks suitable for SISO dynamic modelling is shown in figure 5.55. Its structure is similar to a feedforward network, except that all but one of the inputs are time delayed version of the output. In addition to the ability to extract features in a dynamic system, this network can also be trained by a relatively simple back-propagation algorithm.

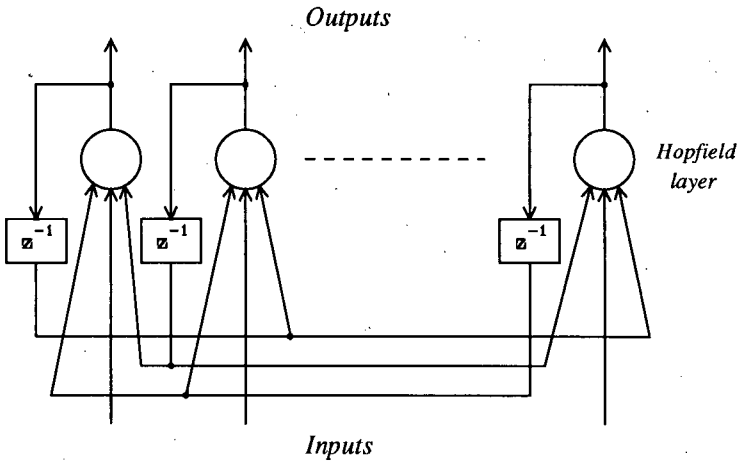


Figure 5.54: Hopfield network. z^{-1} is the unit delay operator.

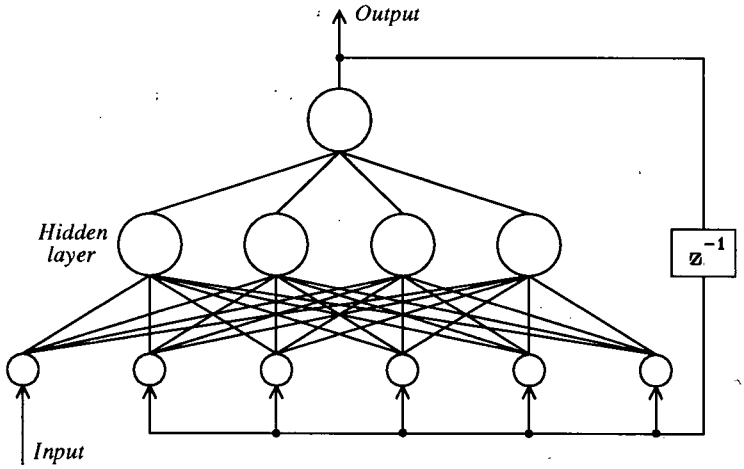


Figure 5.55: A recurrent network suitable for SISO dynamic modelling.

For modelling of Terfenol-D, which possesses hysteretic properties in both the magnetic field and external stress, further enhancement can be made to the above model. From the study of classical Preisach model, it is apparent that the output of such a system is affected by the past extrema. Therefore, including these values as input to the network should allow more accurate modelling without increasing the complexity of the network (figure 5.56).

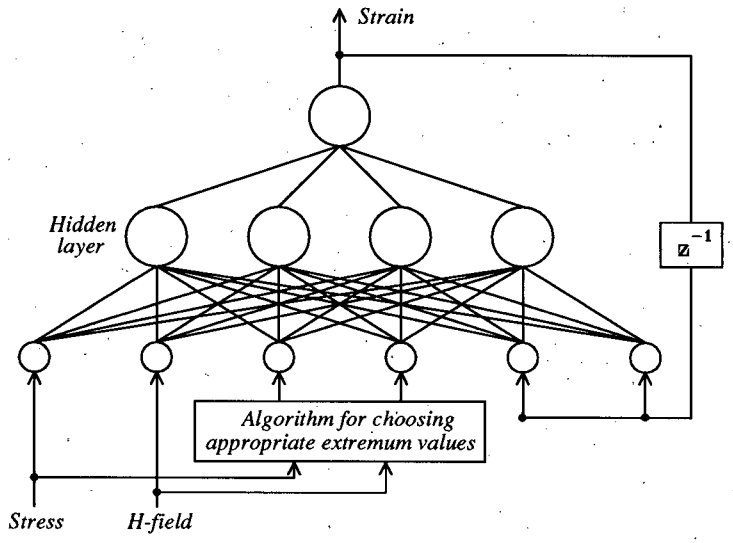


Figure 5.56: A structure of recurrent network suitable for modelling of Terfenol-D.

5.5 Control Strategies

Since Terfenol-D exhibits significant amount of nonlinearity, control algorithms based on linearization would not perform adequately if high performance is desired. It is for this reason that all the approaches proposed below are nonlinear ones based on neural networks.

Behaviour of Terfenol-D drive units depends very much on the past extrema of inputs. This means two similar actuators would behave differently when subjected to the same series of inputs, even from power-off states; the two could have been in different magnetic state initially, and this initial difference does not decay away exponentially as would happen in non-memorizing systems. Therefore either the controller has to compensate for any variance in initial conditions, or some mechanism of guaranteeing consistency between runs has to be built in. One possible solution is to magnetically saturate (positively or negatively) the drive units at the very beginning of each run, thus creating a consistent starting point.

Hughes has recently proposed a control scheme for hysteretic shape memory alloy using the passivity of the the inverse Preisach model in his research proposal for doctoral candidacy examination. Interested reader should consult [61].

Only one attempt to control Terfenol-D actuators using neural network has been found in the literature. Bryant *et al.* [34] use a fairly simple feedforward network to control a vibration isolating system using Terfenol-D actuators. The network is used only as a nonlinear mapping function, and the performance enhancement, comparing to a well-tuned PID controller, is about 2.7 dB (-6.9 versus -4.2 dB).

5.5.1 A One-step Ahead Control Problem

Yip and Pao [62] has proposed a technique based on a recurrent neural network (figure 5.57) to provide control actions for nonlinear dynamic systems. It is formulated as a one-step ahead control problem, and can be stated as: "It is required that control action be found such that the square of error, defined by equation 5.32, is minimal."

$$E = \sum_{i=1}^{q} [y_{d,i}(k+1) - \hat{y}_i(k+1)]^2 \tag{5.32}$$

where

q is the number of output variables,

$y_{d,i}(k+1)$ is the desired response of the ith output,

$\hat{y}_i(k+1)$ is the response of the ith output of neural net emulator.

The optimization method chosen is gradient-descent incorporated in an emulator net. Vary $u(k)$ so that $\hat{y}_i(k+1,t)$, which is defined by equation 5.33, tends to $y_{d,i}(k+1)$

$$\hat{y}_i(k+1,t) = G_i(\underline{u}(k,t)) \tag{5.33}$$

where the index t indicates the time sequence of the optimization procedure. The index k denotes the time index for digital control, while t serves as a label for the successive

computational steps. At every value of k, $\underline{u}(k,t)$ is evolved to yield $G_i(\underline{u}(k,t)) = \hat{y}_i(k+1,t) \rightarrow y_{d,i}(k+1)$.

If, during each k, the control $\underline{u}(k,t)$ is modified by

$$\frac{\partial u_i(k,t)}{\partial t} = \eta \sum_{j=1}^{q} \frac{\partial \hat{y}_j(k+1,t)}{\partial u_i(k;t)} (y_{d,j}(k+1) - \hat{y}_j(k+1,t)) \tag{5.34}$$

where η is a positive proportionality constant, then the error will be a non-increasing function[5]. This modification is carried out digitally using iterations. The control action $u_i(k,t)$ will converge to the correct value, and this can be confirmed by ensuring the final $u_i(k)$ will generate $y_d(k+1)$ as output. According to Yip and Pao, when using the previous control action $u(k-1)$ as the initial seed for $u(k)$, convergence is achieved within 10 iterations.

Below is a summary of the steps involved.

1. Train the neural network $G(u)$ with input-output data from the plant.
2. Read the next desired output $y_d(k+1)$.
3. Set the initial value for the next control action to $u(k-1)$.

[5]The prove is given in [62].

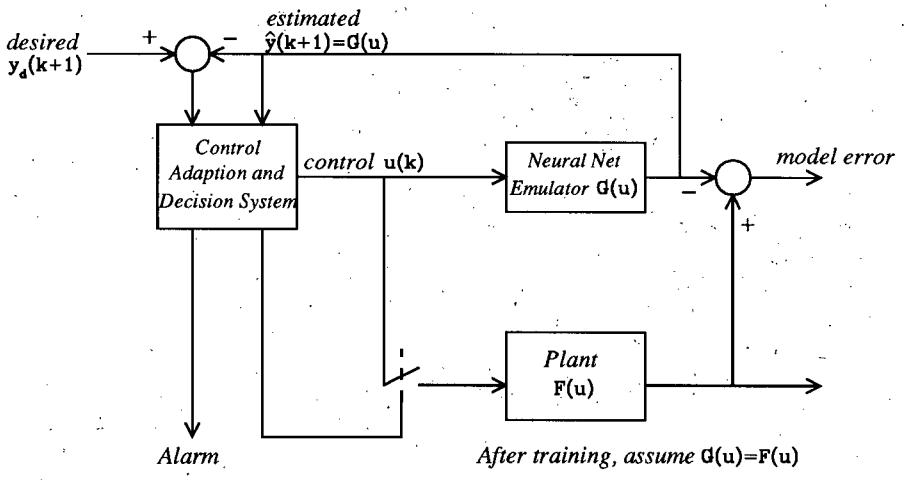


Figure 5.57: Neural network control system proposed by Yip and Pao [62].

4. Making use of the trained neural network $G(u)$, modify $u(k,t)$ according to equation 5.34 until the error E converges.
5. Apply new $u(k)$ to plant.
6. Repeat.

In essence, this is equivalent to solving for the inverse of the plant. With the help of a good model (ie., the neural network), this task is made easier.

In the case of Terfenol-D actuators, y_d will be the desired strain, and u will be the current to the coil. However, the neural network should have an additional input: the external stress. Initially, the network is trained with varying current and stress. After training is completed, this two-input model is used inside the one-step ahead framework to evaluate the input current required, under the amount of stress at that moment, to bring the output to the desired strain.

A potential problem arises when the neural network used is dynamic instead of static. Any change in input to the network during step 4 above by the control adaptation and decision system will leave the actuator and the network in different states. This would greatly reduce the accuracy and effectiveness of the model the next time around. However, this problem can be alleviated if the perturbations are kept small, or the original states of the network are restored after every iteration,

The algorithm can also be used with a Preisach model, in which case it will replace the neural network. If it is determined that even a recurrent network is incapable of capturing enough details of the dynamics for a Terfenol-D drive units, then modelling techniques based on the classical Preisach model can be used.

5.5.2 Model Predictive Control

If it is possible to model the system dynamics adequately with a neural network, the trained network can be used to provide predictions of the plant response over a prediction horizon t_h. These predictions can then be used by an optimization routine to produce the optimum control output for the current time step. The idea is similar to the generalized predictive control proposed by Clarke [63], and both of them try to minimize the cost function

$$J = \sum_{k=N_1}^{N_2} [y_{ref}(i+k) - y_{model}(i+k)]^2 + \sum_{k=1}^{N_u} \lambda_k \Delta u(i+k) \tag{5.35}$$

where y_{ref} represents the output of the reference model (ie., the desired output), y_{model} the output of the plant model, and Δu the change in control input. Below is a summary of the steps involved [64].

1. Predict the system output over the range of future times using the plant model.
2. Assume that the future desired outputs are known.
3. Choose a set of future controls, $\hat{u}$, which minimizes the future errors between the predicted future output and the future desired output.
4. Use the first element of $\hat{u}$ as the current control effort and repeat the whole process at the next instant.

Figure 5.58 shows a predictive control scheme proposed by Murray-Smith *et al.* for non-linear plants using a neural network as the model [65]. Both networks have two inputs: the control signal and the previous output, which would in effect make them recurrent networks. The *prediction model* is used to provide the prediction data to the optimization algorithm, while the *model* is used to track the dynamics of plant.

To minimize the functional 5.35 a simple gradient descent or other more efficient and complex algorithms can be used. In the case of gradient descent the derivatives of the

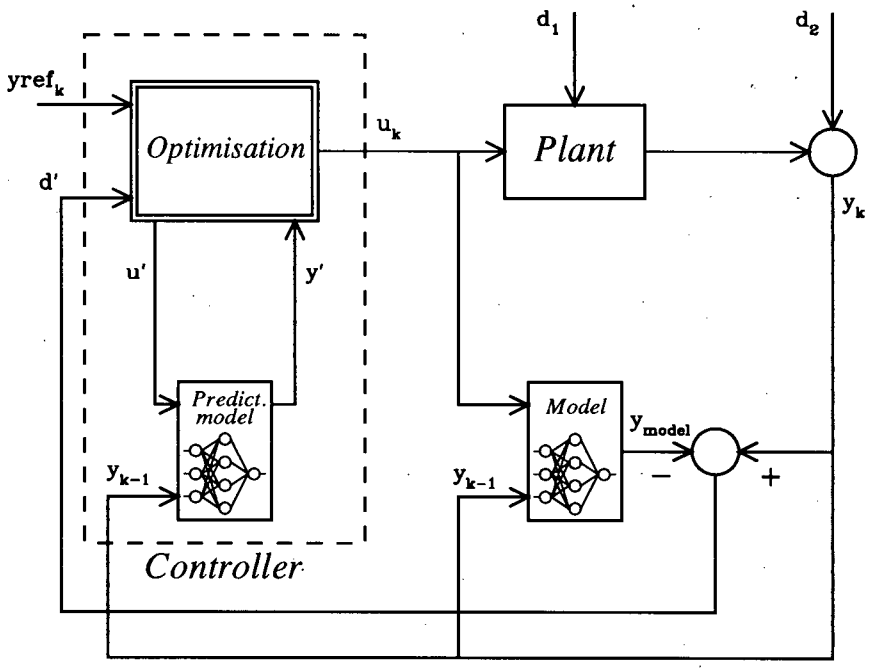


Figure 5.58: Nonlinear predictive control using a neural network for the model.

output against an input variable are estimated from the model and used to calculate the gradient in each iteration. Detail of the procedures can be found in [64].

5.5.3 Inverse time method

A very interesting approach has been suggested by Andersen of the University of Queensland [66,67] (figure 5.59). The basic idea is this. If one wants to take a plant from state

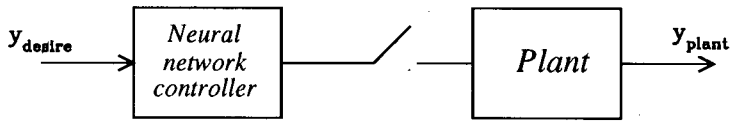


Figure 5.59: Using a neural network as the controller in the Inverse Time method. The switch is closed when gathering information from the plant; open when training the network.

$\mathcal{A}$ to state $\mathcal{B}$, this request is fed into the neural network which then generates a control signal C. This has the effect of taking the plant to state $\mathcal{B}'$. So it is now known that signal C takes the plant from state $\mathcal{A}$ to state $\mathcal{B}'$. Now disconnect the controller and feed the request for taking the plant from state $\mathcal{A}$ to state $\mathcal{B}'$ into the network. A dummy control signal C' will be generated (C is the desired value). The difference, $(C - C')$ is used as the error to backpropagate through the network. After training is complete (providing convergence is achieved), the neural network will act as an inverse of the plant. A controller, knowing the present state of the plant, can now issue command in the form of "go to state $\mathcal{B}$", and the network will generate the appropriate control effort to realize the action.

5.5.4 Using a Reference Model

Another possible scheme is shown in figure 5.60. Here the neural network is trained as the sole controller, with the reference model (eg., a first order low pass filter) determining the desired dynamics of the overall system. Discrepancies between the actual plant outputs and the model outputs are used to train the network, using either backpropagation (may not be possible) or simulated annealing (see section 5.4). After training is done, the controller should generate control actions driving the plant to follow output of the reference model. The overall response can be adjusted by simply modifying this model.

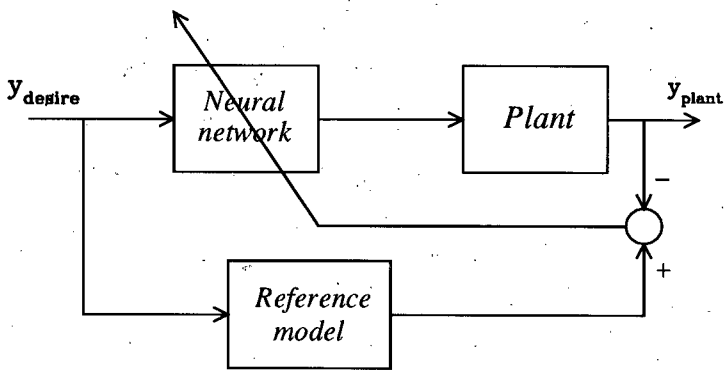


Figure 5.60: Neural network control using a reference model.

Chapter 6

Conclusions

6.1 Contributions

A high force actuator suitable for controlling machine tool vibrations has been designed and fabricated using a giant magnetostrictive material called Terfenol-D. This device has the following advantages over existing designs.

- A push-pull configuration
  - It allows active actuation in both directions, instead of relying on the prestress mechanism to generate pulling forces.
  - It has been established that this configuration is not wasteful: the two rods act in parallel towards external load.
  - It makes the response symmetric by cancelling a portion of the nonlinearity exhibited in the drive units. This makes controlling of the device significantly easier.
  - It allows the prestress inside drive units to stay constant throughout the motion range.
- It incorporates a novel mechanism for mounting flexure plates in tight spaces.
- By using a magnetic circuit with separate static and dynamic field-paths, hysteresis loss due to the biasing mechanism is minimized.

A discussion of the physical properties for Terfenol-D has been presented. Special considerations to features and limitations of this material have been made during the

design of the actuator.

The device has been fabricated and preliminary testings have been performed. Results obtained are encouraging.

Based on the properties of drive units, several promising modelling techniques and control strategies have been proposed.

6.2 Future Works

- Operation of the position sensor has to be confirmed. It shows great performance when tested alone, but might be affected by the large magnetic field around the actuator.
- The magnetic circuit seems to be saturating when full scale drive signal is applied. A new circuit has been designed and fabricated. More testing needs to be done in order to pin-point the exact cause.
- Apply the various proposed modelling and control strategies to the actuator and determine which scheme is the most suitable and effective.
- After a stable control framework is set up, the actuator can be used to implement various algorithms for control of chatter vibrations.
- The actuator can also be used in non-circular turning. Elliptical or even polygonal workpiece can be machined directly on a circular lathe. Potential applications include elliptical cylinders heads for automobile engines.

Bibliography

[1] G. Sweeney, *Vibration of Machine Tools*, The Machinery Publishing Co. Ltd., London, 1971.

[2] D. B. Welbourn and J. D. Smith, *Machine-tool Dynamics: An Introduction*, Cambridge University Press, London, 1970.

[3] N. Kasahara, H. Sato, and Y. Tano, "Phase characteristics of self-excited chatter in cutting", *Journal of Engineering for Industry*, vol. 114, Nov. 1992.

[4] Wajih Kanso, "Model reference estimation and control of tool chatter", in *Proceedings of the 33rd Midwest Symposium on Circuits and systems*, New York, NY, 1991, vol. 2, IEEE.

[5] C. C. H. Ma and Y. Altintas, "Adaptive sliding mode control of chatter vibrations in cutting", Internal report CICSR-TR92-003, CICSR, University of British Columbia, 1992.

[6] D. Alfano, "Selecting electromechanical linear actuators", *Machine Design*, vol. 63, July 11 1991.

[7] D. Caputo, "Electrohydraulic drive offer high power, precise positioning", *Chilton's I & CS*, vol. 64, Aug. 1991.

[8] Gavin D. Jenney, "Simple fly-by-wire actuator", in *National aerospace and electronics, 1991 conference(NAECON)*. 1991, IEEE.

[9] Burton K. Au, "Trends in hydraulic positioning systems", in *Rubber and plastic industries, 1993 conference (RAPCON)*. 1993, IEEE.

[10] Dominiek Reynaerts and Hendrik Van Brussel, "Development of a sma high performance robotic actuator", in *Advanced Robotics, 1992 International Conference*. 1991, IEEE.

[11] S. Hirose, K. Ikuta, and Umetani Y., "A new design method of servo-actuators based on the shape memory effect, theory and practice of robots and manipulators", in *Proceedings of RoManSy '84 - 5th CISM IFToMM Symposium*. 1985, Hermes Publishing.

[12] M. Bergamasco, F. Salsedo, and P. Dario, "A linear sma motor as direct-drive robotic actuator", in *Robotics and automation, 1989 IEEE international conference*, 1989.

[13] K. Uchino, "Electrostrictive actuators: materials and applications", *Ceramic bulletin*, vol. 65, pp. 647–652, 1986.

[14] Edward F. Crawley and Javier de Luis, "Use of piezoelectric actuators as elements of intelligent structure", *AIAA Journal*, p. 1373, Oct. 1987.

[15] Mel Goodfriend, "High force, high strain, wide band width linear actuator using the magnetostrictive material, Terfenol-D", in *Active and adaptive optical components.* 1991, vol. 1543, SPIE.

[16] Mel Goodfriend, "Material breakthrough spurs actuator design", *Machine design*, pp. 147–150, Mar. 21 1991.

[17] Etienne du Trémolet de Lacheisserie, *Magnetostriction: theory and applications of magnetoelasticity*, CRC Press, Boca Raton, Florida, 1993.

[18] T. Akuta, "An application of giant magnetostrictive material to high power actuators", in *Proc. 10th intern. workshop on rare-earth magnets and their applications, Kyoto, Japan, 16-19 May 1989*, The society of non-traditional technology, Ed., Tokyo, Japan, May 16-19 1989, p. 359.

[19] Harry E. Burke, *Handbook of magnetic phenomena*, Van Nostrand Reinhold Company, New York, 1986.

[20] S. V. Vonsovskii, *Magnetism*, Halsted Press, New York, 1974, Translated from Russian by Ron Hardin.

[21] Reacton Terfenol Data Sheet Number 3, "Characteristics of REacton Terfenol", Cheshire.

[22] A. E. Clark, "High power magnetostrictive transducer materials", in *Proc. 3d Intern. Conf. on new actuators, Berlin.* VDI-VDE-Technologiezentrum Informationstechnik, 1992, p. 127.

[23] A. E. Clark, "Magnetostrictive rare earth-Fe_2 compounds", in *Ferromagnetic materials*, E. P. Wohlfarth, Ed., vol. 1, chapter 7. North-Holland publishing company, Amsterdam, 1980.

[24] A. E. Clark, "High power rare earth magnetostrictive materials", *Journal of Intelligent Material Systems and Structures*, vol. 4, pp. 70–75, Jan. 1993.

[25] Mark B. Moffett et al., "Characterization of Terfenol-D for magnetostrictive transducer", *Journal of Acoustic Society of America*, vol. 89, no. 3, pp. 1448, mar 1991.

[26] M. P. Schulze et al., "Variations in strain amplitude and phase in a cylindrical specimen of $Tb_{0.3}Dy_{0.7}Fe_{1.95}$ in an ac magnetic field", *IEEE Transactions on Magnetics*, vol. 28, no. 5, Sept. 1992.

[27] L. Kvarnsjö and G. Engdahl, "Examination of eddy current influence on the behavior of a giant magnetostrictive functional unit", *Journal of Applied Physics*, vol. 67, no. 9, May 1 1990.

[28] L. Kvarnsjö and G. Engdahl, "Examination of the interaction between eddy currents and magnetoelasticity in Terfenol-D", *Journal of Applied Physics*, vol. 69, no. 8, Apr. 15 1991.

[29] Mel J. Goodfriend, "Eddy current critical frequency dependence on rod diameter", Fax reply from Etrema Products, Inc.

[30] Etrema Products Inc., "Etrema Terfenol-D magnetostrictive actuators", Product Catalog, Ames, Iowa.

[31] Hiroshi Eda, Takashi Onoma, Masashi Sahasi, and Tadahiko Kobayasi, "Study on magnetostrictive actuator, development of device with high power and ultra precise positioning", *Journal of the Japan Society of Precision Engineering*, vol. 57, pp. 532–7, Mar. 1991.

[32] Wanjun Wang and Ilene Busch-Vishniac, "A high precision micropositioner based on magnetostriction principle", *Review of Scientific Instruments*, vol. 63, pp. 249–54, Jan. 1992.

[33] Olof Vingsbo and Joakim Schön, "Giant-magnetostrictive vibrator system for fretting testing at low amplitudes", *Wear*, vol. 162-164, pp. 1129–1138, 1993.

[34] Michael D. Bryant, Benito Fernández, Ning Wang, Venkatesh V. Murty, Viswanath Vadlamani, and Timothy Scott West, "Active vibration control in structures using magnetostrictive Terfenol with feedback and/or neural network controllers", *Journal of Intelligent Material Systems and Structures*, vol. 4, Oct. 1993.

[35] Z. Jason Geng and Leonard S. Haynes, "Six degree-of-freedom active vibration control using the stewart platforms", *IEEE Transactions on Control Systems Technology*, vol. 2, no. 1, Mar. 1994.

[36] R. M. Beadle, C. A. McMahon, and R. D. Adams, "Identification and measurement of essential performance parameters for actuators and other engineering devices incorporating giant magnetostrictive materials", in *Actuator '90. Proceedings of the 2nd International Technology-Transfer Congress*, 1990, pp. 219–221.

[37] T. Cedell, L. Sandlung, and M. Fahlander, "New magnetostrictive alloy for rapid conversion of electric energy to mechanical motion", in *Proceedings of 2nd International Technology-Transfer Congress.* VDI-VDE-Technologiezentrum Informationstechnik BmbH, Berlin, 1990, p. 156.

[38] L. Kvarnsjö and G. Engdahl, "A new general purpose actuator based on Terfenol-d", in *Proceedings of 3rd International Conference on New Actuators.* VDI-VDE-Technologiezentrum Informationstechnik BmbH, Berlin, 1992, p. 142.

[39] Arthur G. Thorp, II, "Flexure pivots—design", *Product Engineering*, pp. 192–200, 1953.

[40] Fred S. Eastman, "The design of flexure pivots", *Journal of the Aeronautical Sciences*, pp. 16–21, 1937.

[41] E. J. Hearn, *Mechanics of Materials—An Introduction to the Mechanics of Elastic and Plastic Deformation of Solids and Structural Components*, Pergamon Press, Oxford, 2nd edition, 1985.

[42] Edward Hughes, *Electrical Technology*, Longman Group Limited, London, 1977.

[43] Magnet Sales & Manufacturing Company, Culver City, California, *High Performance Permanent Magnets*, 1993.

[44] Q. F. Xu and A. Refsum, "Analysis of some numerical models of hysteresis loop", in *IEE 2nd International Conference on Advances in Power System Control, Operation and Management*, Dec. 1993, p. 734.

[45] Michael D. Bryant and Ning Wang, "Audio range dynamic models and controllability of linear motion Terfenol actuators", in *Recent Advances in Adaptive and Sensory Materials and Their Applications Conference on Recent Advances in Adaptive and Sensory Materials and Their Applications.* Virginia Polytechnic Institute and State University, 1992, pp. 398–408.

[46] F. Z. Preisach, ", *Physics*, vol. 94, 1935.

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

[48] A. A. Adly, I. D. Mayergoyz, and A. Bergvist, "Preisach modeling of magnetostrictive hysteresis", *Journal of Applied Physics*, vol. 69, no. 8, Apr. 15 1991.

[49] L. Kvarnsj o, A. Bergqvist, and G. Engdahl, "Application of a stress-dependent magnetic preisach hysteresis model on a simulation model for Terfenol-D", *IEEE Transactions on Magnetics*, vol. 28, no. 5, Sept. 1992.

[50] Giorgio Bertotti, "Dynamic generalization of the scalar preisach model of hysteresis", *IEEE Transactions on Magnetics*, vol. 28, no. 5, Sept. 1992.

[51] Ben J. A. Kr ose and P. Patrick van der Smagt, *An Introduction to Neural Networks*, University of Amsterdam, Amsterdam, fifth edition, 1993.

[52] P. J. Werbos, "Backpropagation through time: What it does and how to do it", *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, Oct. 1990.

[53] Howard Demuth and Mark Beale, *Neural Network Toolbox User's Guide*, The MathWorks, Inc., Natick, Massachusetts, Apr. 1993.

[54] M. J. Willis, C. D. Massimo, G. A. Montague, M. T. Tham, and A. J. Morris, "Artificial neural networks in process engineering", *IEE Proceedings-D*, vol. 138, no. 3, May 1991.

[55] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators", *Neural Network*, vol. 2, pp. 359–366, 1989.

[56] K. I. Funahashi, "On the approximate realization of continuous mappings by neural networks", *Neural Network*, vol. 2, pp. 192–193, 1989.

[57] G. Cybenko, "Approximation by superpositions of a sigmoidal function", *Mathematics of Control, Signals, and Systems*, vol. 2, pp. 303–314, 1989.

[58] E. J. Hartman, J. D. Keeler, and J. M. Kowalski, "Layered neural networks with gaussian hidden units as universal approximations", *Neural Computation*, vol. 2, pp. 210–215, Summer 1990.

[59] G. Josin, "Neural-space generalization of a topological transformation", *Biological Cybernetics*, vol. 59, pp. 283–290, 1988.

[60] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities", in *Proceedings of the National Academy of Sciences*, 1982, vol. 79, pp. 2554–2558.

[61] Declan C. Hughes, "Piezoceramic and SMA hysteresis modeling and passivity analysis", Research Proposal for Doctoral Candidacy Examination, July 1994.

[62] Percy P. C. Yip and Yoh-Han Pao, "A recurrent neural net approach to one-step ahead control problems", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 24, no. 4, pp. 678, Apr. 1994.

[63] D. W. Clarke, C. Mohtadi, and P. S. Tuffs, "Generalized predictive control", *Automatica*, vol. 23, 1987.

[64] D. Sbarbaro and K. J. Hunt, "A nonlinear receding horizon controller based on connectionist models", in *Proceedings of the 30th Conference on Decision and Control. Brighton, England.* Dec. 1991, pp. 172–173, IEEE.

[65] Roderick Murray-Smith, Dietmar Neumerkel, and Daniel Sbarbaro-Hofer, "Neural networks for modelling and control of a non-linear dynamic system", in *International Symposium of Intelligent Control*, 1992, p. 404.

[66] Hans Andersen, "Usenet news posting", Oct. 19 1994.

[67] H. C. Andersen and A. C. Tsoi, "Single net indirect learning architecture", to be published in IEEE Transactions on Neural Networks, 1994.

[68] E. A. Nesbitt and J. H. Wernick, *Rare earth permanent magnets*, Academic press, New York, 1973.

Appendix A

Some Elementary Aspects of Magnetism

Note: For a more thorough treatment, see [19, page 41-48] and [68].

A.1 Atomic Magnetism and Magnetic Alignment

The magnetic moment of an atom is due to the spin and orbital motions of unpaired electrons. For the ferromagnetic metals Fe, Co and Ni, in which the unpaired electrons are the outermost ones (3d electrons), the spin contribution to the magnetic moment is the most important. For the rare-earths, however, the orbital motion of electrons in incomplete inner shells contributes significantly to the magnetic moment of the atom.

Exchange energy or exchange forces, quantum mechanical in origin, are responsible for producing magnetic ordering in a crystal to give a net magnetic moment. Although the exchange energy is electrostatic in nature, it is sometimes convenient to think of it as a huge internal magnetic field. The elementary atomic moments should be thought of as magnetic dipoles. Exchange forces are necessary in aligning these magnetic dipoles to produce spontaneous collective atomic behaviour which results in ferromagnetic, ferrimagnetic, and antiferromagnetic behaviour. In Fe, Co, and Ni, the exchange interaction appears to take place via direct overlap of 3d-electron wave functions. In the case of oxides (e.g.,ferrites), their interaction takes place via the non-magnetic intervening oxygen ions and is known as superexchange. In the rare-earth metals, the exchange is via polarization of conduction electrons.

The four types of magnetic behaviour are shown in figure A.61. The arrows are used

to indicate preferred directions of the atomic moment (also referred to as a dipole). The lengths of the arrows indicate the magnitude of the moments.

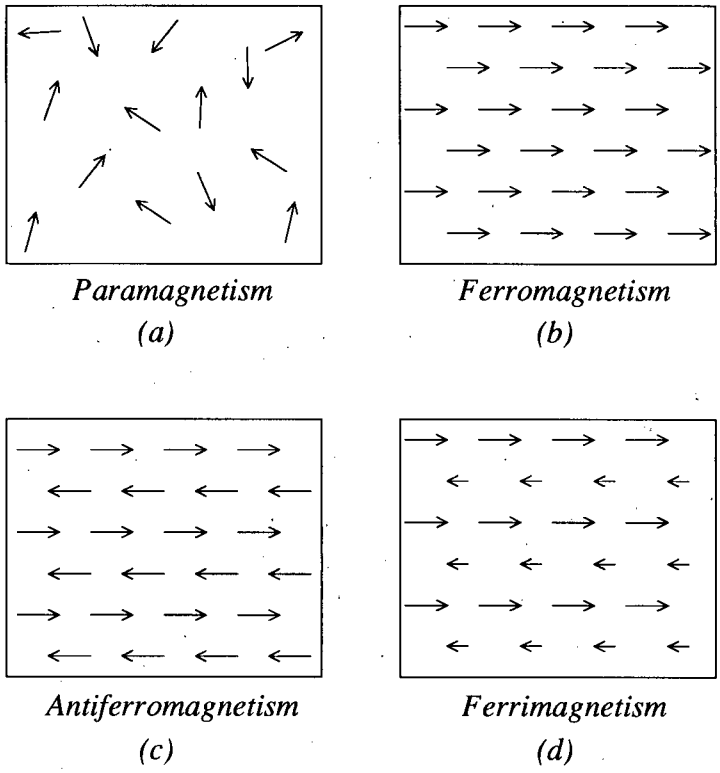


Figure A.61: Types of spin arrangement: (a) paramagnetism, no long range order of magnetic dipoles; (b) ferromagnetism, positive exchange interaction between magnetic dipoles; (c) antiferromagnetism, negative interaction between magnetic dipoles; (d) ferrimagnetism, negative interaction between unequal moments.

In the paramagnetic case, the exchange interactions between the atomic dipoles are so weak that the magnetic moments are essentially free and independent of each other. It is usually impossible to magnetize materials completely at room temperature (i.e.,induce magnetic order) because the thermal agitation of the atoms is so great that one cannot supply sufficient magnetic field energy to overcome it.

When ferromagnetism exists, there are strong, positive exchange interactions between the atomic dipoles. The simple picture shown in figure A.61 indicates a parallel array

of atomic moments on one magnetic sub-lattice. These materials are relatively easy to magnetize, and they have a high magnetic permeability.

Antiferromagnetism occurs when the exchange interaction is negative and the opposing moments are equal. In figure A.61, equal magnetic moments on two sub-lattices point in opposite directions. The net atomic moments of each sub-lattice cancel and result in a zero net moment. These materials are difficult to magnetize and have low magnetic permeability.

Ferrimagnetism occurs in solids containing more than one magnetic sub-lattice. Crystals of a pure substance have well-defined lattice structures extending in three-dimensional space. If two materials are combined by dissolving one in the other, the result can be a new crystal in which each constituent type of molecule assumes a unique position in a more complicated lattice pattern. The resulting structure can be considered to be two crystals of two different substances that simultaneously occupy the same space. Although these "two" crystals are really only one, each substructure can assume characteristics that differ from those of the other substructure in certain circumstances. Ferrimagnetic as well as ferromagnetic materials form domains, and the phenomenon of the double lattice leads to a wide variety of possible magnetic properties.

A.2 Magnetic Domain

Although spontaneous alignment or order of the atomic moments exists in ferro- and ferrimagnets, these materials often do not show a net moment in the absence of an applied field. This is due to the existence of magnetic domains in the material. Domains are small regions within which the elementary magnetic dipoles or atomic moments are oriented by the exchange energy so that they are parallel. Figure A.62 shows a schematic view of such domains in a section of a demagnetized bar. In this state, the net moment

of the material as a whole is zero. The reason for this is that the domains are oriented in such a way that their moments cancel so as to minimize the total magnetic energy.

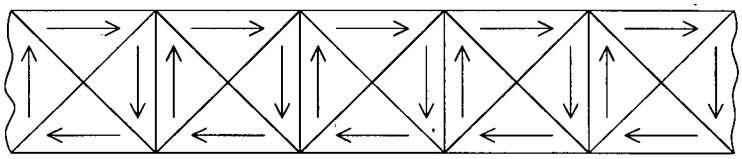


Figure A.62: Schematic illustration of domain in a demagnetized iron bar. Each triangular area represents a magnetic domain composed of many dipoles oriented in parallel.

The transition region between domains is called a domain wall. Domain widths frequently range between 10^{-2} and 10^{-5} cm. Domain wall thicknesses are of the order of 1000 atoms in an iron crystal, and approximately 10^{-6} cm in rare-earth alloys.

The domain structure and the behaviour of domains and domain walls in an applied magnetic field determine many of the magnetic properties in ferro- and ferrimagnetic materials. In materials that are easy to magnetize and demagnetize (known as soft magnetic materials) such as iron-silicon alloys, the magnetic domains are usually large (visible under a ordinary microscope). For materials that are hard to magnetize and demagnetize (hard magnetic materials), on the other hand, the domains are very small.

Appendix B

Simulink model

Below is the simulink model used for simulation of the linear model for the overall system.

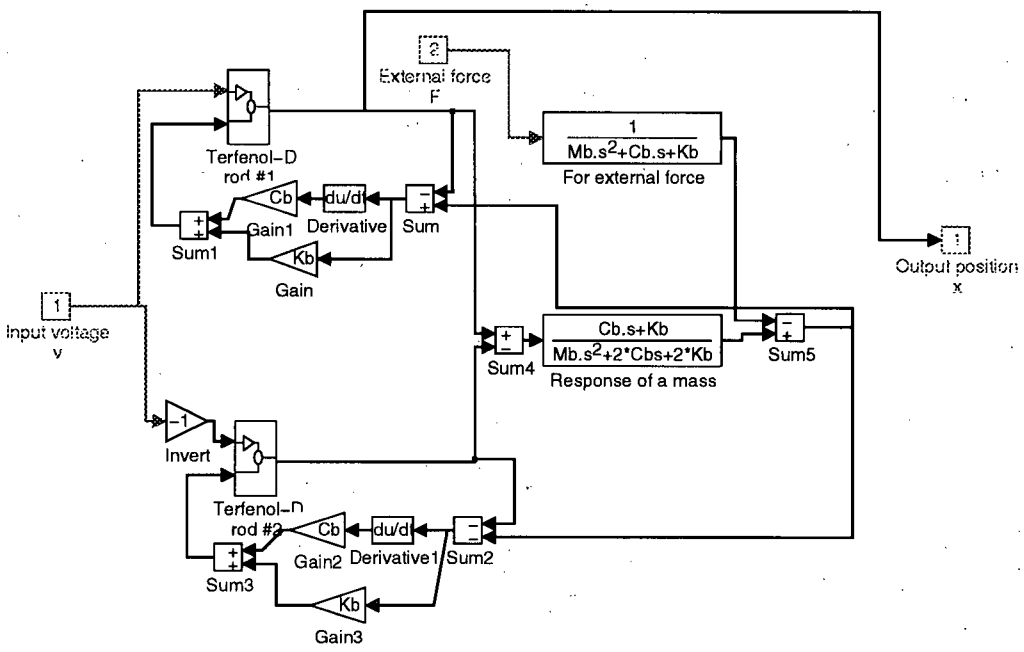


Figure B.63:

Appendix C

Mechanical Drawings

All the mechanical parts are designed using the Advanced Model Extension (AME) of AutoCAD version 12. In fact, the whole actuator *exists* as a 3-D model inside the program before any part is machined. This allows the elimination of many possible errors.

Below are some composite views of the various parts.

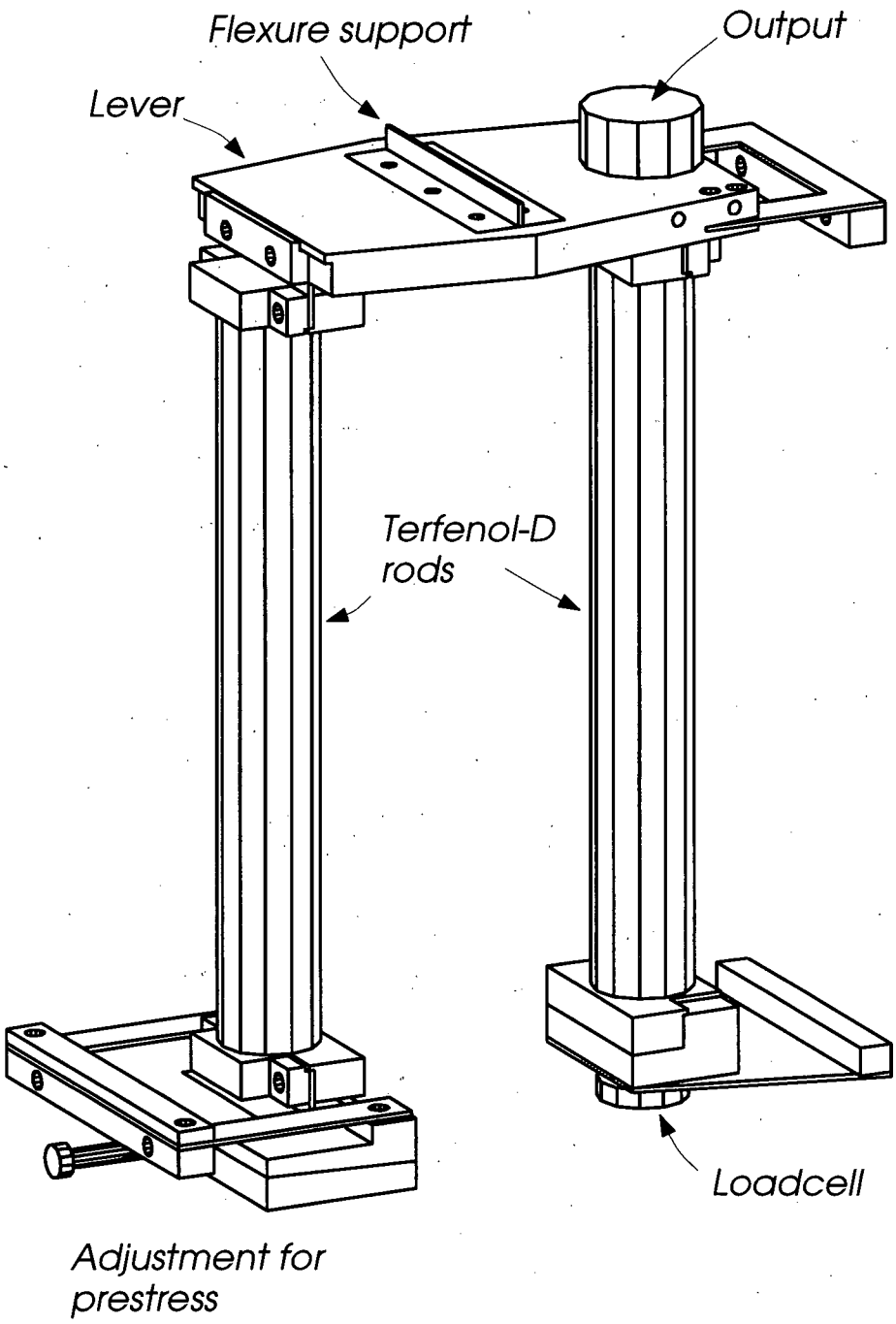


Figure C.64: The Terfenol-D rods are shown here with the loadcell, the prestress mechanism and the lever at top.

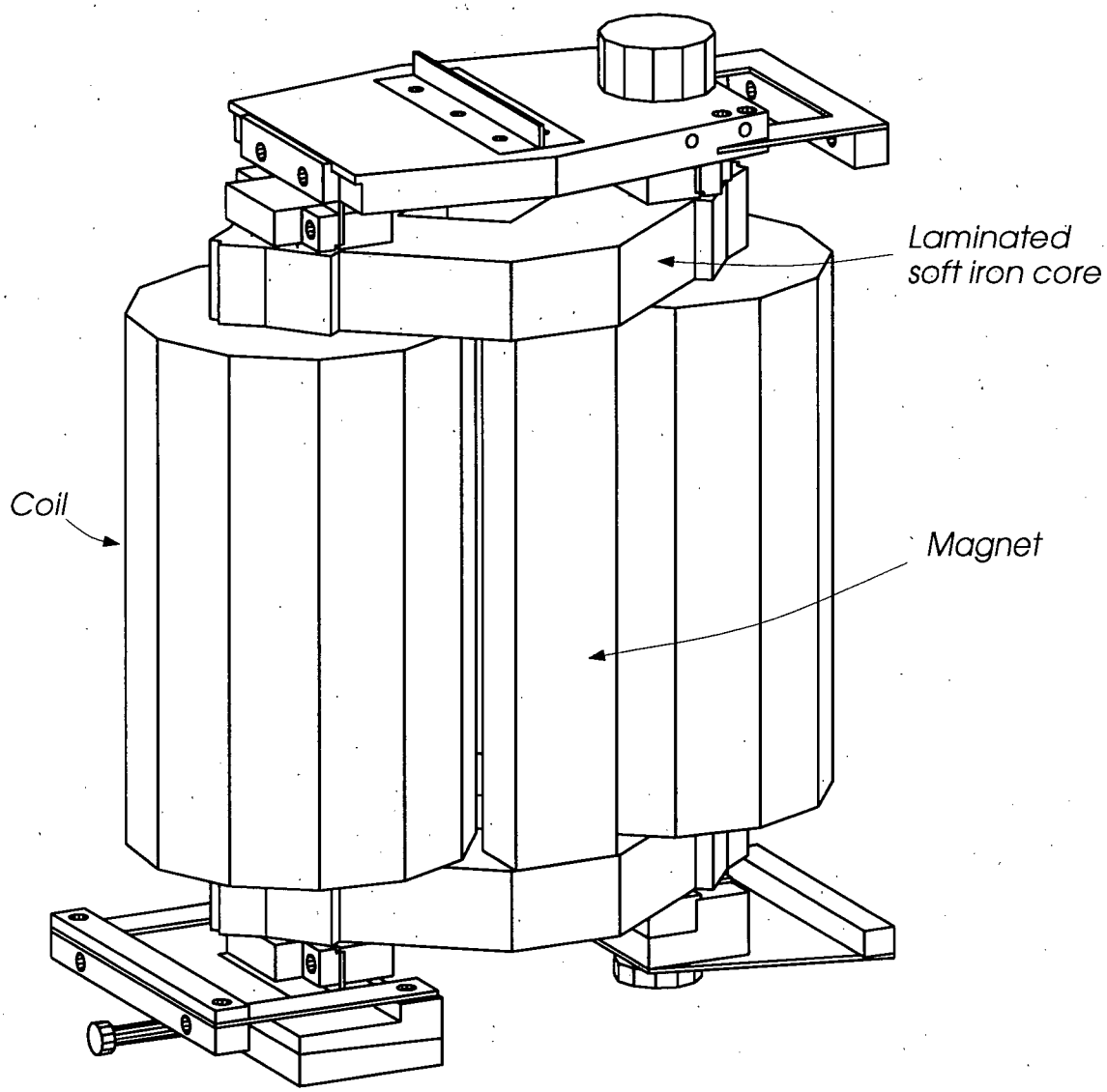


Figure C.65: With the magnetic circuit included.

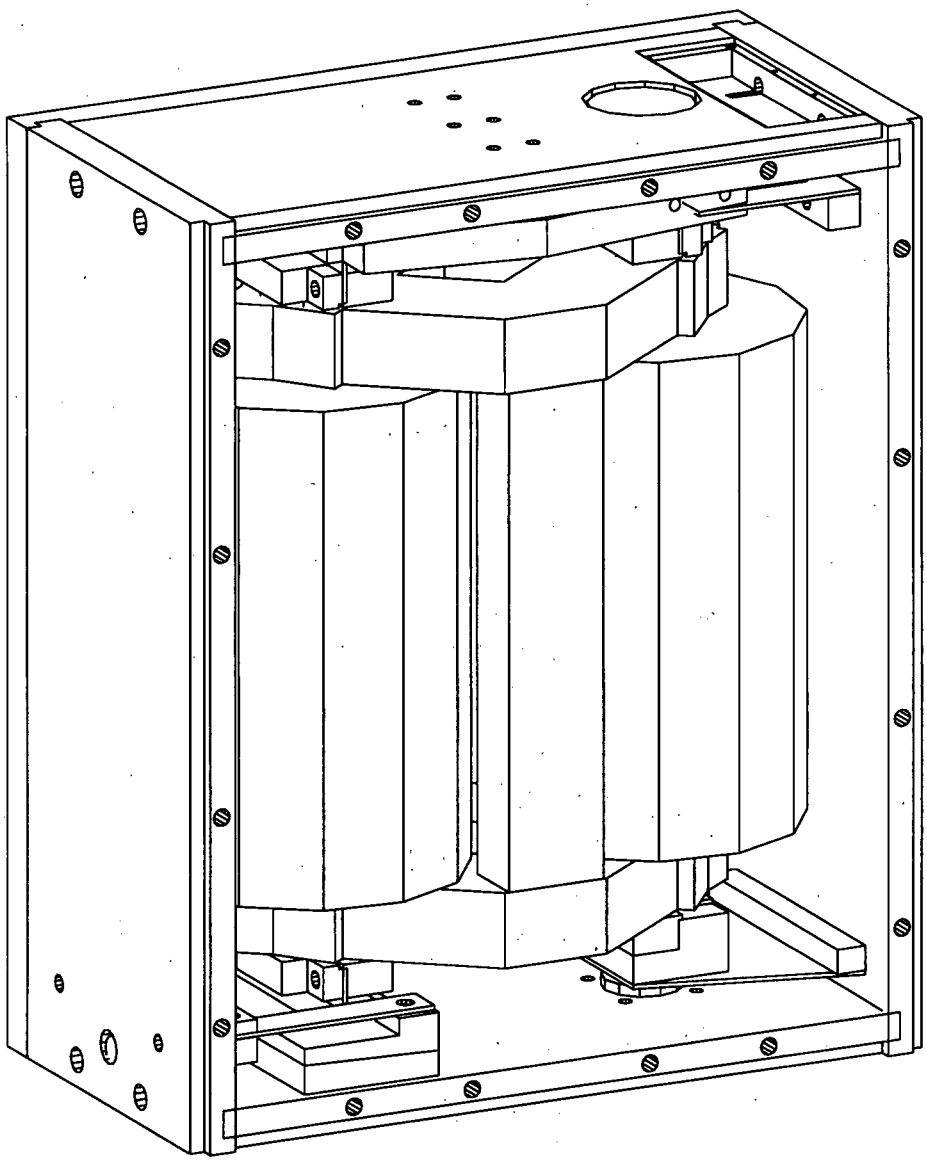


Figure C.66: With the case.