Enabling Shape Memory Alloys as Actuators for Robotics

by

Austin Russell Gurley

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 16, 2017

Approved by

David Beale, Chair, Professor of Mechanical Engineering
Royall Broughton Jr., Professor Emeritus of Polymer and Fiber Engineering
David Bevly, Professor of Mechanical Engineering
John Hung, Professor of Electrical and Computer Engineering

Abstract

Does any existing technology provide a viable actuator for the high degree-of-freedom robots that will soon be ubiquitous in society? If a technology exists that is in any way superior to the electric motor for robotics – why has it not come into common use? What are the technology's benefits and flaws, and can these flaws be overcome? This thesis seeks to answer these questions. The author believes that, of all known devices which convert electrical energy to mechanical motion, shape memory alloy actuators provide the most convincing capabilities to fulfil the needs of complex robots over the next decades. The electro-magnetic motor has moved industry and automation for over a century, yet modern robotic machines are reaching the fundamental limits of its ability as mechanical systems increase in complexity while decreasing in size. Shape Memory Alloy (SMA) actuators have benefits of extraordinary high strength, high energy density, simplicity, and low cost. These benefits come along with obstacles of complex thermo-electro-mechanical behavior, difficult control, fatigue over time, and moderate speed – all of which can be overcome – as well as barriers of low energy efficiency and limited life-time which cannot be overcome. This thesis addresses all of the obstacles to enable more powerful and capable robots, answering the essential question:

How can we enable SMA technology so that it becomes useful for robotics?

Table of Contents

List of Figures

Nomenclature

Symbol	Name	Typical Value		Units
		Martensite	Austenite	
σ	Stress			MPa
T	Temperature			℃
ξ	Martensite phase fraction			
ϵ	Strain			
x	Position ($L(\epsilon - \epsilon_0)$)			m
F	Force (σA)			N
ϵ_L	Max. strain of SMA with 0 stress	.04		
Ω	Max. stress of SMA with 0 strain	3.0		GPa
θ	Coeff. of thermal expansion	0.55		MPa/℃
c_p	Specific heat	836		J/(kg°C)
ΔH	Latent heat of transformation	24200		J/kg
h	Heat transfer coefficient	121.2		W/(m^{2}C)
$\mathcal{M}$	Subscript for Martensite			
$\mathcal{A}$	Subscript for Austenite			
$x^d_{A\to M}$	Distance from M centerline			
$x^b_{A\to M}$	Distance from M_s line			
C_M , α	Slope of A, M lines	6.89 - 10.3		MPa /℃
k	Distribution factor	0.155 - 0.62		1/℃
ν_M	Distribution skew	0.0 - 2.5		
M, A	Peak transformation temp. at 0 Pa	67	93	℃
M_s, A_s	Transformation start temp. at 0 Pa	72	87	℃
M_f, A_f	Transformation finish temp. at 0 Pa	62	98	℃
ρ_0	Resistivity	75	86.5	$\mu\Omega$cm
α	Effect of Temperature on ρ	0.15	0.05	$\mu\Omega$cm/°C
C	Elastic compliance ($1/E$)	0.0357	0.013	1/GPa
E	Elastic Modulus	28	75	GPa
A	Wire cross-sectional area			m^2
d	Wire diameter			m
L	Wire length			m
A_s	SMA device surface area			m^2
m	SMA device mass			kg

Chapter 1
Introduction & Outline

Introduction

Does any existing technology provide a viable actuator for the high degree-of-freedom robots that will soon be ubiquitous in society? If a technology exists that is in any way superior to the electric motor for robotics – why has it not come into common use? What are the technology's benefits and flaws, and can these flaws be overcome? This thesis seeks to answer these questions. The author believes that, of all known devices which convert electrical energy to mechanical motion, shape memory alloy actuators provide the most convincing capabilities to fulfil the needs of complex robots over the next decades. The electro-magnetic motor has moved industry and automation for over a century, yet modern robotic machines are reaching the fundamental limits of its ability as mechanical systems increase in complexity while decreasing in size. Shape Memory Alloy (SMA) actuators have benefits of extraordinary high strength, high energy density, simplicity, and low cost (Table 1). These benefits come along with obstacles of complex thermo-electro-mechanical behavior, difficult control, fatigue over time, and moderate speed – all of which can be overcome – as well as barriers of low energy efficiency and limited life-time which cannot be overcome. This thesis addresses all the obstacles to enable more powerful and capable robots, answering the essential question:

How can we enable SMA technology so that it becomes useful for robotics?

Table 1 – The characteristics of SMA actuators. 'Ugly' characteristics have been overcome by this work

The Good	The Bad	The Ugly*
High specific strength	Energy efficiency (3-5%)	Low strain
High specific energy (work)	Life-time (100k-1M cycles)	Sensors bigger than wire
Silent operation		Nonlinear and hysteretic
Excellent DoF/$		Difficult training/processing
Excellent DoF/m^2		Bandwidth vs size

Actuators for robotics must work in a way that is complementary to both machine designers and to control engineers. Most SMA research today is focused on materials science rather than creating actuators. When used as an actuator, SMA provides an unconventional challenge because the material creates the motion: Models of the material and actuator are one-and-the-same. Most SMA actuators in modern studies are constrained to operate in very particular environmental and loading conditions. To realize the extraordinary specific energy[1] in SMA materials, the complete actuator and feedback system (feedback sensors and energy sources) must be considered. SMA wires are very compact actuators; they are so small that most feedback sensors are larger than the actuator, and therefore in feedback the effective SMA specific energy is often greatly reduced due to the added size and mass of the sensor. To enable SMA actuators for robotic actuators, the material behavior must be modeled under general loading and ambient conditions, feedback control must provide robust positioning accuracy in all scenarios, and mechanical designs must minimize the burden of sensors and energy sources in the feedback control system. Thus, we must understand electro-thermo-mechanical behavior, be able to model all of those components, and be able to design and control the actuator within their constraints. This thesis addresses all four primary areas of study that must be understood to create robust SMA actuators: mechanical, electrical, thermal, and feedback control design.

Mechanical Design & Modelling

SMA materials provide benefits of high strength, simplicity, and low cost, but provide those with obstacles of low physical strain (maximum of 8%), nonlinear behavior, and sensitivity to improper handling. Before diving into mathematical analysis, a sketch of the material behavior is provided (Chapter 2). Most work on SMA dives directly into multi-variable constitutive models for the material, hiding the useful properties in equations. Presenting a sketch of the behavior without a single equation allows students and beginners to expect certain reactions from smart materials which would not be seen in common engineering metals. After an intuitive sense of the material behavior is found, the models can be developed more clearly. 'Constitutive models' relate the actuator length and force to temperature and crystal phase transformation (Chapter 3).

[1] Specific energy generally refers to energy-per-unit-mass (J/kg), or energy-per-unit-volume (J/m^3). The latter is used throughout this work unless otherwise noted.

Crystal phase change affects the shape memory strain, elastic modulus, and electrical resistivity - and is the cause of the nonlinear behavior. A primary goal of this thesis is to create and control SMA actuators. With that objective, the modeling approach is often a compromise between clarity, simplicity, and adherence to underlying physics. The best engineering models are often not the best scientific models, but are the ones that describe the bulk behavior most accurately - even if that requires some empirical fitting and simplification of underlying physics. Most existing SMA models are not continuous in general conditions that an SMA actuator experiences – the author has developed a model that ensures continuity and simulation stability (Chapter 3).

Most material models assume the SMA material has been trained to provide highly functional properties. This is convenient for researchers who can purchases bespoke materials, but practicing engineers will find that specialized heat treating and training processes are required to get useful transformation properties from SMA material. These processes have been studied and schemes for proper training are developed in Chapter 4. Heat-treatment of the material sets the bulk properties, and cyclic training stabilizes their useful range of stress, strain, life, and actuation temperatures. The training – working of the material immediately after heat treating – stabilizes the material properties so that they do not fade or decay through the actuators life. The stabilized properties also ensure the material models will have consistent parameter values through the life of the actuator.

Engineers who might look to adopt SMA are not only disappointed by lack of clarity in material behavior, but also in converting the unusual motion of SMA into useful actuation in mechanical designs. The specific energy of an SMA actuator is very high – but this is due to high strength and low strain, where most engineering actuators provide high strain with low force. Decades of mechanical design research has overcome this obstacle in many ways; the most reliable designs and concepts, as well as novel designs created by the author, are presented in Chapter 5.

Electrical & Thermal Modelling

NiTi materials used in this work are actuated by heating and cooling. The heating is achieved by passing electric current through the material. Cooling occurs when the material dissipates heat into the surrounding air via convection. Controlling the electric power transferred into in the material heats it and thereby actuates it. The design of electrical circuits to drive electricity

through the material usually result in a compromise between component cost and resolution. In practical actuators, pulse-width-modulation (PWM) provides a reliable and efficient means of controlling the current in the material to heat and cool it. Circuits are designed so that they can simultaneously drive the heating current and measure it. These PWM systems are designed and compared to analog amplifiers which provide smooth control of input voltage, though at higher cost, in Chapter 6.

Heat transfer from the material to ambient air cools the material. Thermal behavior determines actuator efficiency and bandwidth, so it is critical to consider when matching SMA to a robotic application. The NiTi material has two interesting interactions to complicate this process. First, the material stores internal energy in both specific heat and latent heat of transformation: this causes the material to exhibit nonlinear mechanical behavior during heating and cooling. Methods and calculations to determine and design SMA actuator bandwidth are also presented in Chapter 6. Second, the electrical resistance of the material changes with crystal phase change; this can disrupt the expected current flow through the material and change the heating rate. If carefully measured, the change in resistivity can be used to help sense the material crystal state. The author's dual measurement self-sensing technique provides reliable strain sensing in all ambient and loading conditions, so that robust feedback control is achieved without external sensors (Chapter 7).

Feedback Control and State Estimation

These thermo-electro-mechanical models can be used to design a controller, or controllers can be created that are robust despite lacking a model. To test the models and controllers, a hardware-in-the-loop tensile testing machine was designed to simulate external loads on a physical SMA system under test (Chapter 8). In the actuator context, the objective is typically position control of linear or rotary actuators. The focus in this work is on control of actuator position (angle) and force (torque), not motion planning for the entire robotics system. A new modelling technique developed in this work presents the SMA models (of Chapter 3) with a 'Linear Parameter Varying' perspective. By considering an equivalent mechanical system instead of the bulk material model, a model can be framed as a state space with varying coefficients (Chapter 9). These coefficients vary as the actuator moves, but the primary system characteristics are maintained. State estimators are designed based on the LPV model. The

estimator is very useful to help understand the material behavior since the SMA temperature and crystal phase cannot be measured in practice. The LPV model is modified using the Exact Velocity Linearization (EVL) technique to make it more reliable for control (Chapter 10). The EVL and other linearizing controllers can greatly simplify the control design process. Over the life of an SMA actuator, most of these models will fail to remain accurate and so the controllers degrade over time. Sliding mode control provides a model-independent method to control the SMA material, immediately overcomes the primary nonlinearities of the system, and is robust to many disturbances (Chapter 11). Because the material is heated and cooled by rapidly switching input regardless (due to the PWM circuits), sliding mode provides an ideal solution to SMA control in most applications.

To validate the entire suite of modelling and control methods developed in this work, several demonstrations have been made which showcase the capability of SMA and the author's progress in creating truly practical actuators from SMA material (Chapter 12).

Contributions

Advances to enable SMA actuators made in this work fall into three categories; thermo-electro-mechanical models for wire actuators, the robust self-sensing discovery using the dual measurement technique, and linear-parameter varying perspective of SMA materials that lead to improved estimators and controllers. In previous research, SMA actuators have never been implemented in physical devices with reliable control using self-sensing techniques; these three advances are required to enable SMA actuators for robotics. In the course of this research, the author has:

- Developed a comprehensive one-dimensional, phenomenological thermo-electro-mechanical model for SMA wire. This model uses a new inner loop hysteresis condition that does not break down in controller simulation (as the common existing models do). By coupling the material to a heat transfer model for thin wire, it relates the material states (stress, strain, temperature), and constituent states (phase fraction, effective elastic modulus, electrical resistance), to driving input (electric power). (published 2016 [1], 2017 [2])
- Discovered and analyzed a self-sensing technique which provides a truly linear estimate of strain which surpasses all previous attempts to use electrical resistance as a strain sensor. The technique reduces RMS self-sensing error from 12% to below 1%; this 'dual-measurement' method is robust to ambient conditions, actuator life (fatigue), and most importantly it is accurate regardless of the force on the actuator. Self-sense strain measurement makes the SMA model fully observable with no additional sensor. (published 2016 [3], 2017 [4])
- Framed the one-dimensional SMA actuator model as a Linear Parameter Varying (LPV) system, creating linear estimators which can estimate both phase fraction and actuator stress. This has never been achieved before without relying on unrealistic simplifications about material behavior (which do not work well outside the laboratory) and is the first presentation of a technique which can accurately estimate SMA phase fraction and stress from simple measurement of position. (published 2017 [5])
- Modeled SMA behavior as a third order LPV system, with Martensite phase-fraction as the varying parameter. It was determined that the entire nonlinearity is contained in a varying scalar gain between temperature and phase fraction. Non-linear controllers based on gain-scheduling and sliding mode were created, providing control very near the ultimate limits due to thermal bandwidth. The control performance is validated in constrained testing equipment which simulates common loads for SMA actuators (published 2017 [6])
- Synthesized multiple SMA actuators which provide reliable transformation of SMA tension into joint torque and prevents premature fatigue of SMA wire due to strain of curvature. These mechanical joints are used to create a robotic arm which is accurate and powerful (for its weight), at a cost much lower than equivalent electric-motor based devices. The robotic arm is used as a platform for testing robust controllers (published 2017 [5]).

Chapter 2

Sketch of General Shape Memory Alloy Behavior

Shape memory behavior is often explained in a manner that involves Constitutive Models and concepts like Green Strain and Statistical Thermodynamics that make the topic inaccessible to new students of the science. Here we present an illustrated demonstration of the main features of SMA behavior that the designer will encounter, with sufficient detail for those with the intent of designing SMA-based systems. Additionally, a 'design guide' has been prepared to help designers choose whether SMA can be used in their system based on bandwidth, strength, size, and cost requirements.

What Causes Shape Memory Motion?

Shape memory materials achieve their remarkable properties due to crystal transformations that can be reversed without destroying the material. Most metals have multiple crystal phases that are different shapes – this is not unique to SMA. However, in most metals a phase change in the crystal structure does not change the macroscopic size of the material, and can only be changed by annealing in a furnace. SMA phases are unusual since the crystal phases are of very different size, and the transition between them is activated by temperatures that do not melt or soften the material. Heating and cooling the material causes a change in crystal phase and a corresponding change in the physical size of the material.

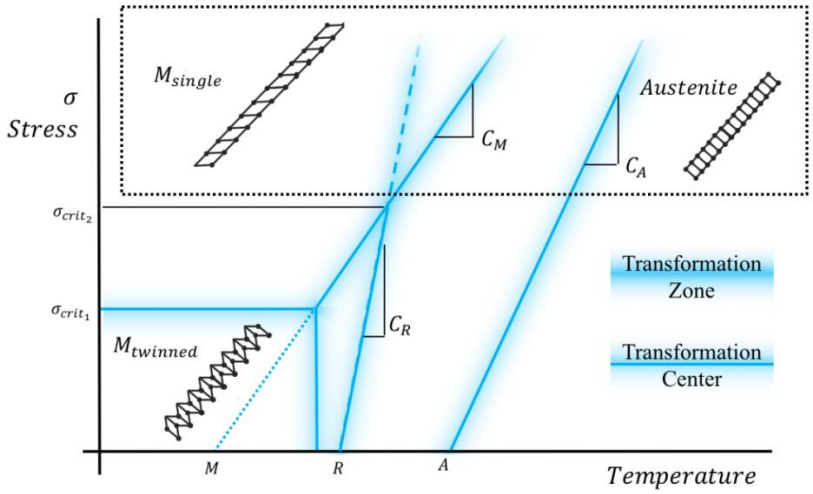


Figure 1 - The stress-temperature phase diagram shows where different NiTi crystal phases are guaranteed, as well as transition regions where the crystal phase percentage depends on the previous history of phase change

There are two predominant crystal phases in NiTi which can be drawn on the 'phase diagram' of stress vs temperature as shown in Figure 1. The high temperature, short and stiff phases is called 'Austenite' (Figure 1, right). The low temperature, high-stress, long and soft phase is called 'Martensite' (left). There is another special phase when both stress and temperature are low, called 'Twinned Martensite', where Martensite crystals are oriented such that the crystal is compact (lower left), but the material is soft and with some applied stress is easily stretched into the normal 'de-twinned' Martensite (upper left). For actuators that are repeatedly actuated, it is desirable to keep the stress high enough to ensure twinned Martensite never occurs – the safe region to prevent twinning is denoted with a dotted outline in Figure 1. The diagram also contains a band between Martensite and Austenite where the phase cannot be determined from stress and temperature alone: In this region, the SMA material will have some Martensite and some Austenite: the exact proportions depend on stress, temperature, strain, and the history of motion in the phase diagram. This middle region is very important – by careful application of temperature and stress, the material can be controlled to be partially Martensite and partially Austenite.[2] When both phases are present, the strain will be somewhere between the compact size of Austenite and the extended size of Martensite – this is how useful actuators are made from SMA. To keep track of how much of each phase is present, the 'Martensite fraction' is usually tracked, denoted by the Greek letter ξ. In many simple applications, designers want an actuator to move between two positions – in a wire this might be a certain short length corresponding to 100% Austenite and a long length corresponding to 100% Martensite. Examples where this is practical include activating a latch or switch, or opening and closing a door. A more general-purpose actuator or 'artificial muscle' needs to be controlled to intermediate lengths. Fortunately, by carefully controlling the stress and temperature, the entire body of material can have a fractional phase state.

There are two primary behaviors in SMA based on the way they are actuated: whether actuated by cycling stress or by cycling temperature. The first one is the 'super-elastic' (also

[2] This fractional phase does not mean that a single metal crystal could be, for instance, 'half Austenite and half Martensite'. It means that, in the bulk of the material, half the crystals are Martensite, and the other half Austenite.

called ‘pseudo-elastic’) effect (Figure 2, Top-Right). In common metals, application of stress leads to elastic strain and eventual plastic deformation which usually means the material has failed. However, in SMA the apparent plastic deformation is not permanent – even a large ‘plastic’ deformation disappears when the stress is released, and the material returns to the initial position. Real ‘plastic’ deformation is not recovered when stress is reduced, and real ‘elastic’ behavior does not have a nonlinear load path –the unique SMA stress-strain behavior is called ‘super-elastic’ motion. The second effect happens when stress is held constant, and the temperature is varied – because the temperature variation changes the crystal phase, the material exhibits a large strain (Figure 2, Bottom Left). The shape of the material at high temperature is determined by the way it is trained and annealed when manufactured. Since heating the SMA will return it to the ‘remembered’ shape, this is called the ‘shape memory’ effect. In an SMA wire actuator, the memorized shape is simply a short and straight wire. Both types of loading are considered in more detail below. For robotics applications, we are interested in using SMA materials as actuators and so most of the focus is on the shape memory effect.

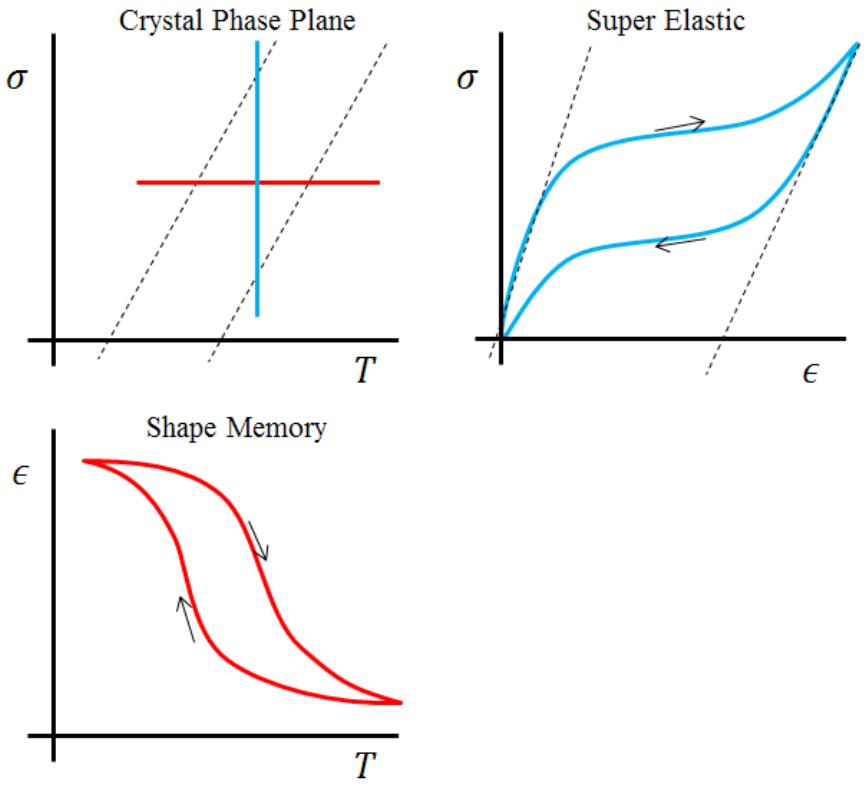


Figure 2- The two extraordinary characters of SMA materials: The super-elastic effect is stress-induced phase change. The shape-memory effect is temperature induced phase change.

Super-elastic and Shape Memory Effects

Now consider three scenarios where temperature is held constant (as in a thermal chamber) and stress is cycled (as if by a tensile testing machine). First, the shape memory effect which is common in magic tricks and simple two-state actuator devices (Figure 3, red). At low temperature, applied stress moves the material from twinned Martensite to de-twinned Martensite. When stress is released, the material does not return to the origin but rather it seems to have been truly plastically deformed. However, application of temperature can contract the material again by transitioning from de-twinned Martensite, to low stress Austenite, to twinned Martensite. This cycle is called the 'shape memory' effect, and is the most common effect seen in SMA magic tricks and gimmicks since it provides easily imagined 'memory recollection' behavior. Secondly, consider the 'super-elastic' effect (Figure 3, blue). At moderate temperatures, the material rests in Austenite when no stress is applied. Applied stress first elastically deforms the material, and then appears to plastically deform the material as the stress induces reorientation of the metal crystals from Austenite into Martensite. Eventually, the entire material is 100% Martensite, and it begins to 'stiffen' again with the elastic modulus of Martensite. The super-elastic effect becomes apparent as stress is relieved; the material 'un-plastically' deforms as stress is decreased and the crystals shift back into Austenite, leaving a flag-shaped stress-strain path. When stress is eventually reduced to zero, the material returns fully to the starting strain with no permanent deformations. This behavior is called the 'super-elastic' effect because it seems like the material should be ruined, but it has not actually been broken in any way. Finally, consider the case where temperature is very high (Figure 3, green). The material begins in the Austenite phase, if the temperature is sufficiently high, then applying stress will not cause the crystal phase to change to Martensite, perhaps even up to the material failing.

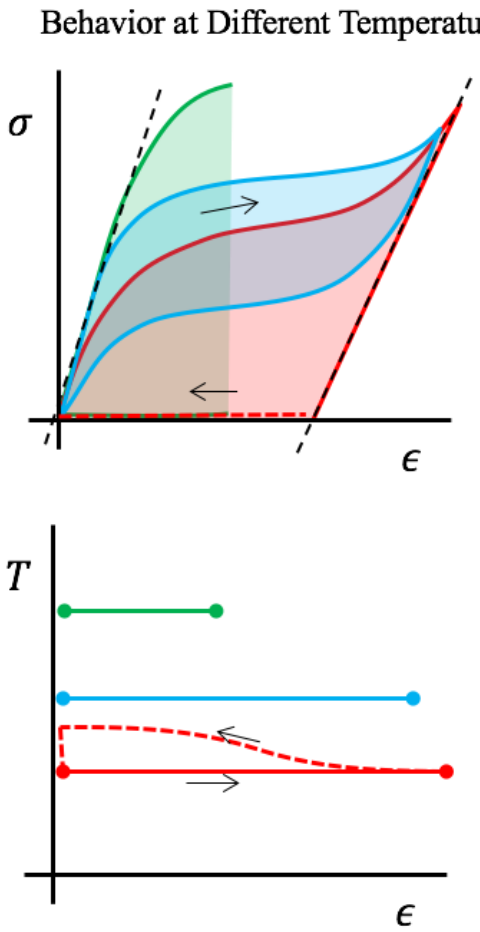


Figure 3 – SMA behavior at different temperatures. At high temperature, the material remains in Austenite and is mostly elastic. At moderate temperature, the material can be force into Martensite, but returns to Austenite when the stress is relieved. At low temperatures, the material exhibits the 'pseudo-plastic' effect; it deforms permanently until temperature is applied to contract it again.

Shape Memory Actuation & the Effect of Load and Loading Path

While motion on the phase plane can be very complex, in many practical applications the strain is easy to predict because the load applied to the SMA material is known. It is easiest at this stage to speak about SMA wires since the force and length are directly related to the stress and strain, respectively. Usually the force trying to extend the wire is called the 'antagonist' load – the SMA wire and this load together create an 'agonist-antagonist' pair, often called simply an 'antagonist' pair. The antagonist load can even be another SMA wire. For now, take the simple case where a constant force is applied to the wire and that material is heated and cooled in a temperature controlled chamber. This occurs, for instance, if a mass is suspended from the wire. Assume that the stress is high enough to de-twin any Martensite (if it is not then no motion would occur at all!). Looking to Figure 4, notice first the serpentine path of the temperature vs

strain; this path is hard to predict and will be discussed later, so for now we will ignore it and look only at the end conditions – where the material is 100% Austenite or Martensite. Note also the straight lines of stress – we chose a case with constant stress (hanging weight) so the path here is linear and easy to predict. Most importantly, notice the two lines which determine the boundaries of the motion – these lines define the 100% Austenite (left dashed line in Figure 4) and 100% Martensite (right dashed line) elastic modulus, or stiffness. The Martensite exhibits a lower elastic modulus and so has a lower slope. This provides a very interesting effect – the more stress applied to the material, the more strain is caused by the shape memory effect when the material is heated and cooled. Increasing the load increases the useful strain – increasing both stress and strain greatly increases the energy of the wire - for free. Thus, designers should always use SMA wires that are small enough to ensure they will be highly stressed during use – in fact stress should be kept high as possible without exceeding the limits for the design fatigue goals (discussed later).

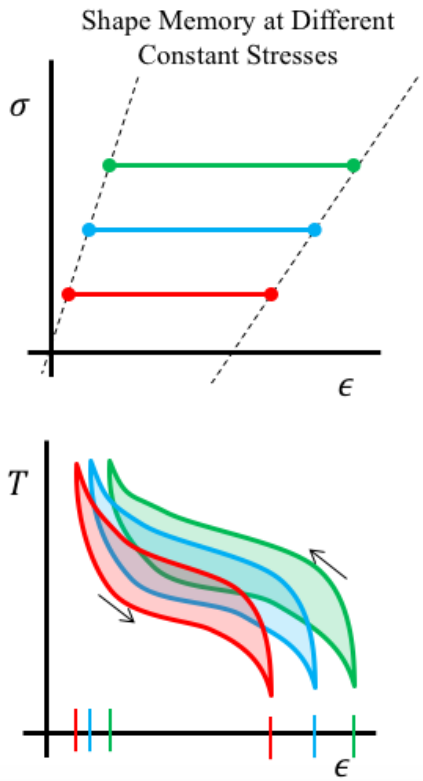


Figure 4 - The shape memory effect at several loads. Notice that higher loads produce larger shape memory strain. All cases are above the critical stress which prevents Martensite twinning.

Now consider cases that are slightly more complex than the constant load. These cases could produce a variety of different paths on the stress-strain diagram, but for now we consider only cases which result in a linear stress-strain path (Figure 5). First, the constant load is plotted for comparison (in blue). Now consider a linear spring which provides force proportional to its deflection with spring rate 'k'. Since the spring is arranged as an 'antagonist' load, the slope of the spring is opposite to the conventional direction, but it maintains the same slope 'k'. This path is shown in red on Figure 5. Because of the different modulus of Martensite and Austenite, the total strain between the hot and cold condition is *shorter than that of the constant load antagonist*. In other words, by actuating against a spring instead of a constant load, we decrease the total stroke of the SMA actuator! This is not good – typically springs are convenient antagonists yet are causing a loss of energy efficiency. Now consider a special spring – perhaps created with a mass and cam, spring and lever, or a nonlinear spring like a leaf spring – which provides a positive antagonist force but has a decreasing slope. This path (shown in green) is *longer than the actuating strain of common spring and the constant force*. Thus, by clever design of the antagonist load, we still create a linear path but get free work from the actuator. Many references cite the benefits of 'leaf-spring' antagonists without clarification on why it is beneficial; this provides a rational basis for that benefit.

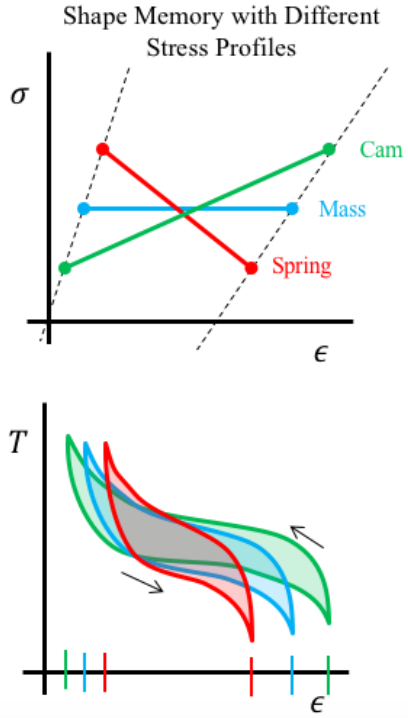


Figure 5 - The shape memory behavior with different antagonist loads - springs, masses, and special cases

Inner Loop Hysteresis

The forgoing discussion was possible without any explicit description of the hysteretic path of the material in temperature vs strain. Not only is the hysteresis complex to describe, but in cases with high speed or acceleration, the linear paths sketched in Figure 3 through Figure 5 are not reliable indicators of the SMA performance during actuation, even with simple loads. The inherent hysteresis in the material plays a large role in the material motion, and is critically important when creating high-performance feedback controllers. In the previous examples we have seen the serpentine path which the strain takes as temperature (or stress) varies. Now we look at the crystal phase alone, as it varies with temperature while maintaining constant stress (Figure 6). It is informative to see that the peak transformation rate ($\partial\xi/\partial T$) occurs on the transformation centerline in the phase diagram (by definition).

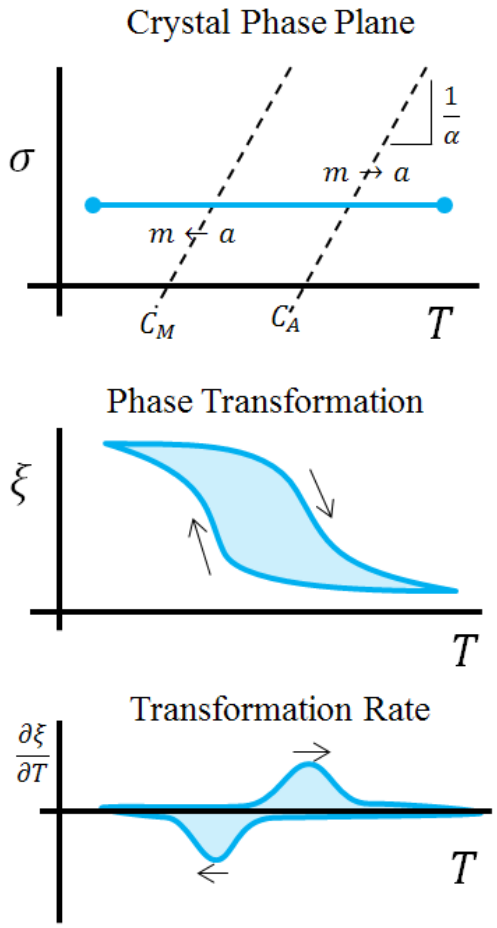


Figure 6 - The shape memory effect at constant load. The bands of transformation denote the peak of transformation and the band of stress and temperature at which the transformation will occur.

When modelling and controlling the material, the designer must be aware of complexity that result from switch directions in the middle of transformation. The key aspects can be illustrated in two examples. First (Figure 7, left), consider a material that begins in Martensite (1). The material is heated until it is 100% Austenite (2), and subsequently cooled until it is partially Martensite and partially Austenite (3). By heating again, we can reverse the transformation back to Austenite (path 4, red). But, this time only half the material needs transformation; the peak rate of transformation is only about half that of a full cycle as indicated by the smaller area under the $\partial\xi/\partial T$ curve.

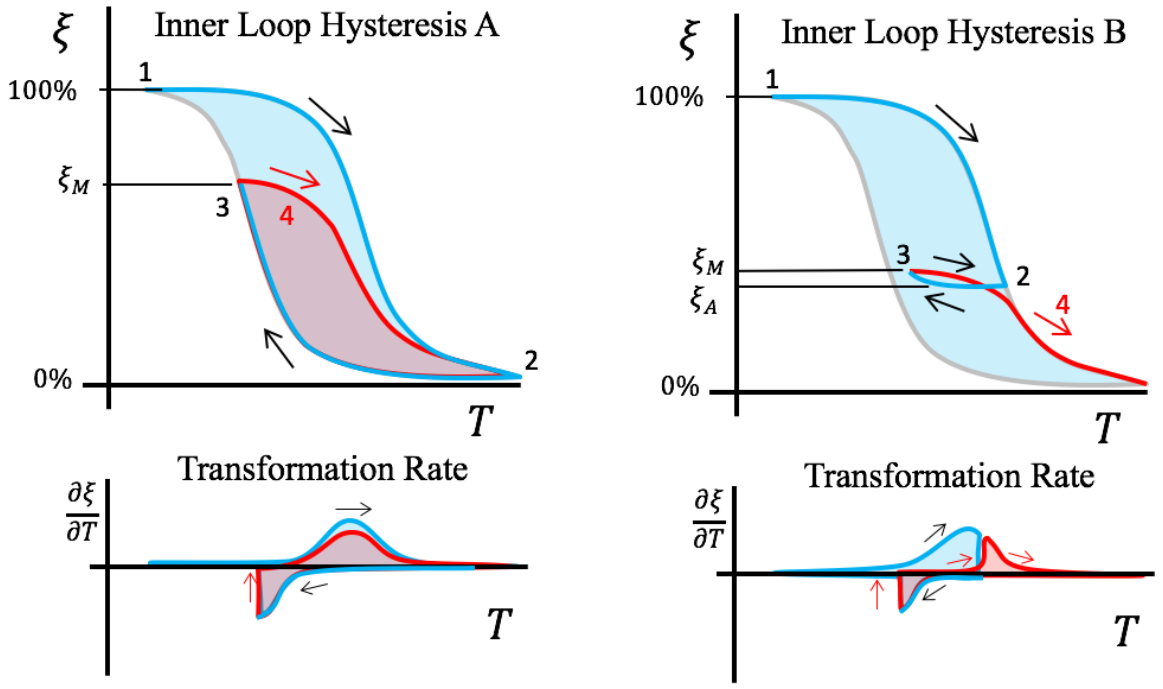


Figure 7 –(LEFT) The inner loop hysteresis for a change of direction in mid transformation. (RIGHT) The inner loop hysteresis for a short 'pause' in heating – many models have trouble describing this effect accurately

The second instance of hysteresis causes the 'inner loops' in the model (Figure 7, right). Consider a situation where the material begins in Martensite (1), is heated until it is partially Martensite and partially Austenite (2). It is then cooled slightly causing it to reverse transformation directions and start reorienting into Martensite again (3). However, this reversal is paused (3), and the material is heated (path 4, red) until it becomes 100% Martensite. In this case, the material will have only a small change in phase fraction during the delay, and then return to the original transformation path. During this last heating phase (path 4, red), notice the

way $\partial\xi/\partial T$ 'jumps up' to the original slope once the temperature exceeds the first original point (2). This behavior is unusual in the SMA material; most existing models cannot accommodate this behavior (this is remedied in Chapter 3). These hysteresis concepts were demonstrated with a driving temperature input, but can be applied analogously to the super-elastic case by replacing temperature in the foregoing discussion with stress.

Of course, in real actuators the stress and temperature effects can rarely be separated – the temperature and stress and crystal phase are all interrelated in even the simplest cases (see Chapter 6 for more details). One great way to see this is by stretching a super-elastic SMA spring quickly; it gets noticeably hot as it transforms. Not quite as intuitive is what happens when the spring is quickly released – it gets suddenly colder as it draws energy from the surroundings on transformation back to Austenite. These more complex coupled behaviors are unfortunately much harder to explain pictorially, and are easy to understand with a basis of the underlying physics, as described in depth later in this thesis.

SMA Actuator Design Equations

Before explaining SMA in detail, we provide a set of practical equations for SMA round wire actuators used as tensile actuators. The equations are mostly self-explanatory and are based on simple calculations and manufacturer specifications. Next, a design method for selecting appropriate SMA wire actuators is described. Finally, definitions of common terms used in SMA work are provided.

Table 2 - Units and parameters used in the design equations provided in

Diameter (mm)	D	mm
Length (mm)	L	mm
Voltage (V)	V	Volts
Austenite finish temp	A_f = 90	° C
Martensite finish temp	$M_f = 40$	° C
Shape memory strain	4	%
detwinning stress	100	MPa

Table 3 - Design calculations for SMA devices. These rules are conservative to provide long life (100k to 1M cycles)

Area	$A = \frac{\pi}{4} D^2$	mm^2	(1)
Max force (for 100k cycles)	$F_{max} = 170 * A$	N	(2)
Min. required preload (to ensure extension)	$F_{preload} = 100 * A$	N	(3)
Min. expected stroke	$\delta_{max} = 0.04 * L$	mm	(4)
Safe bending radius	$r_{min} = 50 * D$	mm	(5)
Cooling time (in still, room temp air 25 ° C)	$t_{cool} = D^2/0.0172$	seconds	(6)
Electrical resistance	$R = 0.001 * L/A = 0.0012 * L/D^2$	Ohm	(7)
Safe continuous current to prevent damage	$I_{max} = 20000 * A + 40$	mA	(8)
Safe continuous voltage to prevent damage	$V_{max} = I_{max} R$	Volts	(9)
Heating current	$I = V/R$	Amp	(10)
Heating time	$t_{heat} = 19000 * A/I$	Seconds	(11)
Cycle time	$t_{cycle} = t_{heat} + t_{cool}$	Seconds	(12)
Cost (MuscleWire from Dynalloy)	$L * 0.0045$	USD	(13)
SMA Energy (force*stroke)	$E_{sma} = \delta_{max}(F_{max} - F_{preload}) = 2.2 * L * D^2$	Nm	(14)
SMA steady-state power consumption	$P = 405 * D * L * (A_f - 20)$	W	(15)
Heat transfer coefficient	$65.5 e^{-\frac{d}{4}} (T - T_\infty)^{\frac{1}{6}}$	W/m^2/C	(16)
Actuator Efficiency	$\eta = \frac{0.025309}{\frac{0.000026}{d}\left(\frac{L}{V}\right)^2 + 1}$		(17)

Decide if SMA is practical based on energy consumptions

In general, it is easy to determine whether SMA actuators will work in a power limited application if we know the work done to fully actuate the system. The energy is typically the force required multiplied by the displacement (N*m) or, for rotary systems, the torque required multiplied by the angular displacement in radians (Nm*rad).

$$E_{req} = F * \delta \quad \text{or} \quad E_{req} = \mathrm{T} * \delta_{rad}$$

Very quickly, we can eliminate many applications by computing the power required to cycle the SMA wire continuously. Note the efficiency η is based on the ratio of mechanical work to thermal work (heating the SMA), but continual transfer of heat away from the wire into the air means that in practice a slow cycle time does not usually improve efficiency.

$$P_{min} = \frac{1}{\eta}\frac{E_{req}}{t_{cycle}} = \frac{E_{req}}{.02 * t_{cycle}}$$

If this power is not achievable (based on battery power or current limits, for instance) then SMA should not be considered. Remember, the value of SMA is energy in a small size, not energy efficiency.

Design Method A: Desired Energy → Desired Speed → Diameter → Length

Choose wire diameter based on required bandwidth

Because the wire is activated by temperature, we must be sure that the system can cool fast enough to actuate at the desired speed. An easy way to determine the worst-case scenario speed is to look at the cooling time in still, room temp air. Any small airflow will greatly decrease the time to cool. By rearranging equation (6), we can solve for a maximum diameter given a maximum cycle time:

$$D_{max} = \sqrt{0.0172 * t_{cool}}$$

Compute the length of wire needed

By considering the energy first instead of looking at a mechanical design (forces and torques) – we can calculate the minimum amount of wire we will need to achieve the desired energy, regardless of the mechanism design. Rearranging equation (14) and using the required mechanical energy computed above

$$L_{min} = \frac{E_{req}}{2.2D^2}$$

For rotary actuators and levers, determine a moment arm length

Because the wires only contract 4% of their length, in a rotary actuator the pulley or connection radius is often very small. Notice that the moment arm length can be adjusted to

directly trade stroke for torque – another benefit of the SMA's linear actuation. First, consider that we only have L_{arm} length between the fixed end of the wire, and the pulley. Then the total length of wire required does not change, but the system must have $n \geq L_{min}/L_{arm}$ wires in parallel. The moment arm or pulley radius is

$$r = 0.04 * \frac{L_{arm}}{n * \delta_{rad}} = 0.04 * \frac{L_{min}}{\delta_{rad}}$$

Based on the simple calculations above, it should now be clear whether SMA is a practical choice for your application. Remember that the benefits of SMA are in strength-to-weight, simplicity, and cost. They are not good for efficiency nor high speed. This lends them to be useful for miniature applications (high strength, small form) and applications with many unique degrees of freedom (lower cost, high strength in each DoF).

Design Method B: Force & Stroke → Length → Diameter & # Parallel Wires

In some cases, we know the desired linear stroke and cannot include a lever – this will require using a fixed length of SMA wire that will provide the stroke. Using equation (4), we can determine the wire length

$$L \geq \frac{\delta_{desired}}{0.04}$$

If the design must use a single wire, or a fixed number of wires in parallel (n), then the diameter can be directly computed

$$D \geq \sqrt{F_{des}/55n}$$

Alternately, the wire diameter can be determined using the method shown in the previous design method (based on a cycle speed), thus setting the minimum number of parallel wires required.

Electrical Considerations

The electrical resistance of an SMA wire increases with increased length and decreases with increased cross-sectional area (7). Now that the total length of wire has been determined, we determine the voltage needed to effectively heat the material. In the linear actuator case, this requires connecting the supply voltage the entire length of wire. However, in the rotary case or in any other case where multiple wires are mechanically in parallel, we have multiple options for electrical connection since the electrical path can zig-zag through the mechanically parallel

wires. To increase total resistance, the path should be as long as possible (perhaps the entire wire length) – this will increase the voltage required to effectively heat the system. Alternatively, by passing current through multiple wires at once the electrical resistance decreases – this will require more current to effectively heat the system. Taking the minimum power calculated above;

$$P_{min} \leq VI = \frac{V^2}{R} = I^2 R$$

With multiple wires, the resistance is based on the length from positive to negative electrical contact ($L_{electric}$) and the number of wires (n):

$$P_{min} \leq VI = \frac{V^2}{\left(\frac{0.0012 * L_{electric}}{nD^2}\right)} = I^2 \left(\frac{0.0012 * L_{electric}}{nD^2}\right)$$

This equation can be used to solve for voltage or current now that the power and resistance are known, and may help determine the ideal electrical path through parallel wires acting on the same load. If the power cannot be carefully controlled (as it could by using pulse-width modulation, for instance), be sure to also consider equations (8) and (9) which provide recommended limits to the electrical power so the wires are not overheated and destroyed.

Definitions

Because the study of shape memory alloys is not very well-known, researchers are still able to stick to their own terminology for different aspects of the field. These definitions clarify what the author means when using some of the common overlapping definitions.

1) "Shape Memory Alloy" or "SMA". A member of the general class of metals that exhibit the "shape memory effect" under some condition.
2) "NiTi". The chemical composition of a common shape memory alloy. It is prevalent due to fair cost and high performance.
3) "Nitinol". The name given to NiTi by researchers who discovered its shape memory properties at the Naval Ordinance Lab (NiTi-NOL).
4) "Super-elastic". A behavior of some shape memory alloys in which the material has nonlinear stress (it usually appears to deform plasticity) during straining, but which returns to zero stress without residual plastic or elastic strain when fully unloaded.

5) "Pseudo-elastic". Exactly equivalent to "super-elastic".
6) "Shape Memory Effect" or "SME". Material behavior in which a material shows apparently permanent deformation when strained, but the strain can be recovered by subsequently heating the material.
7) "One-way SME". A material which contracts (extends) when heated, and will stay contracted (extended) until stress is applied.
8) "Antagonist SME". A mechanism wherein a material which exhibits one-way SME (the agonist) is paired with a preload force (the antagonist – a weight, spring, or another SMA device) such that said material contracts when heated, and when cooled, it is extended by said preload force.
9) "Two-way SME". A material which contracts (extends) when heated, and extends (contracts) when cooled even when no stress is applied.
10) "Heat treatment" or "annealing". (Sometimes confusingly called "training"). The process of holding SMA material in a fixed shape, then raising the alloy temperature to a high temperature (for NiTi, above 300C) such that the material has diffusive stabilization of the metal crystals. Upon quenching, the material assumes said fixed shape whenever it is heated above the A_f temperature (for most NiTi, above 90C).
11) "Training". The process of heating and cooling SMA material repeatedly while under preload to cause the antagonist shape memory effect. This antagonist SME process behaves differently for each of the first few cycles (for NiTi, ~30 cycles) but stabilizes to a consistent and repeatable motion afterword.
12) "Functional fatigue". Fatigue caused by crystal rearrangement during SME, as opposed to conventional fatigue due to conventional elastic stressing and straining.

Chapter 3

Constitutive Model for Control Simulation

Decades of researchers have contributed to robust and well-validated models for shape-memory alloy thermomechanical behavior. The most common models contain the same core constitutive equations, but vary in their approach to transformation boundaries and phase transformation kinetics. These models have varying levels of complexity and accuracy – over time the models match better to experiment and become simpler to implement. Before constructing a model to use in this work, it is important to provide a clear explanation and categorization of these earlier developments. We review the development of 1D constitutive models of the past three decades, which are the most practical modelling technique when designing controllers for SMA actuators from fine wires, and expand on them to create a model with the best compromise of accuracy and simplicity for a servo-actuator model.

Many kinetics equations for SMA have been explored and validated in experiment. Unfortunately, for realistic choices of parameters and simulation conditions, these models have discontinuities that make control simulations fail. Here we present a solution to 'patch the gaps' in the kinetics model, such that it provides more accurate and more useful control simulation results. We introduce a weighted sum of two models: one of them an implicit solution to ensure continuity, and the other an outer-loop hysteresis model that 'attracts' the phase fraction toward the outer loop of its range of behavior.

Introduction

One-dimensional phenomenological models for SMA have been developing over the last 30 years, slowly incorporating newer experiments and realizations, seeking an optimal compromise of accuracy and complexity for controls design [7, 8]. Many contributions have been made, and several forgotten, over this time. Here we consider the common thread of 1-D SMA constitutive modelling initiated by Tanaka in 1986 [9]. Brinson and Huang reviewed the primary models developed before 1996 [7]. Paivo presented the most common SMA constitutive models in 2005, covering the main details of the models without experimental data [8]. Prahled and Chopra in 2001, and Sayyaadi et al in 2012, compared two common models to experimental data [10, 11].

Prahlad compared those common models focusing on experimental fitting of phase plane transition. For implementing real-time controls reference models, these 1-D tensile constitutive models show the best compromise of accuracy and simplicity. The 1-D restriction is not often limiting in practice since most SMA actuators are created from fine wires.

Fundamentals of Shape Memory Alloy Behavior

Shape memory alloys such as NiTi can be transformed between the two solid crystal phases, Austenite ($\mathcal{A}$) and Martensite ($\mathcal{M}$), due to changes in stress and/or temperature. NiTi has a special property in that the material has large strains (on order of 4%) in this transformation. The whole bulk of material does not instantly transform in its entirety from one state to another, and so fractions of the two phases can be present within the bulk of material at any given time. Conventionally, the instantaneous percentage of the two phases is tracked and denoted by ξ, where $0 \leq \xi \leq 1$ is the Martensite fraction of the material. The conditions under which the NiTi crystal phases exist can be shown on a 'phase diagram' (Figure 10). Lines M and A define the midpoint of transformation (50%) between Martensite and Austenite, along lines with slopes of C_M and C_A.

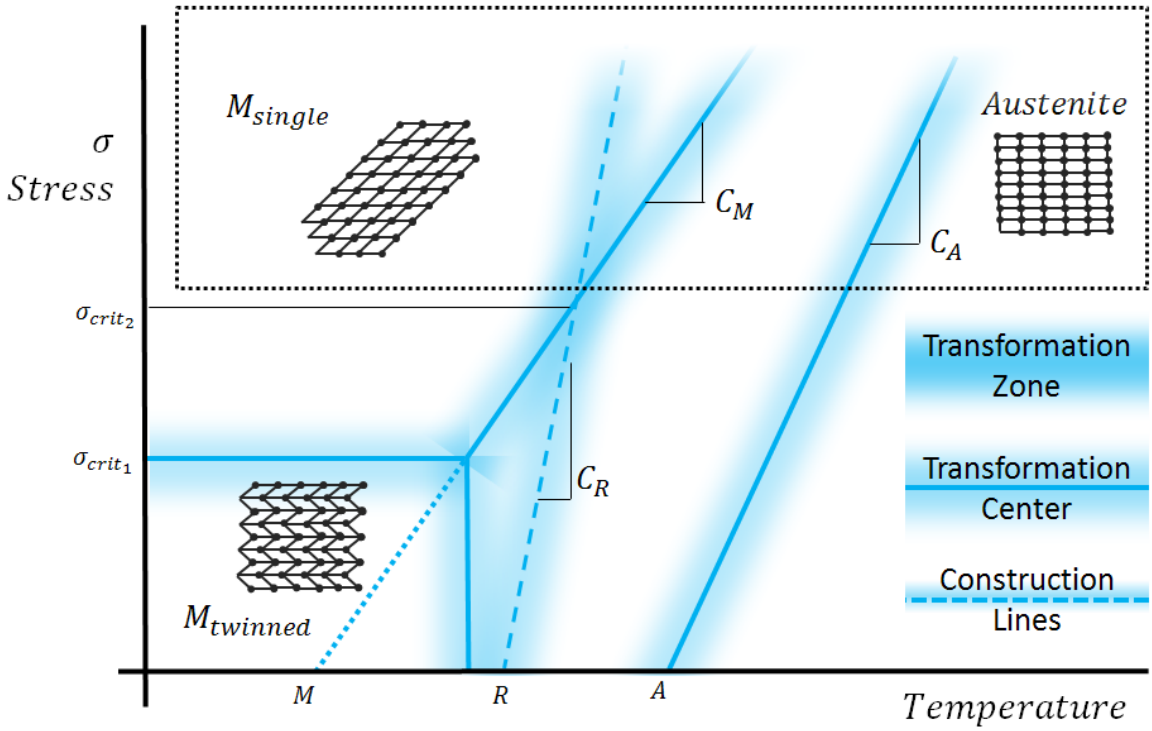


Figure 8 - Complete SMA Phase diagram including twinned and de-twinned Martensite states. SMA wire actuators are designed to operate in the highlighted (boxed) region.

When designing SMA actuators, the mechanical design of the device usually ensures that a preload stress is applied to the wire, ensuring it is extended when cooled. If this stress is not applied and temperatures are sufficiently low, the SMA material can enter a third crystal 'state', the twinned Martensite state. This has been considered by many researchers [12, 13, 14, 15], but is not important for actuator design and not described in detailed here. This minimum preloading stress has the added benefit of eliminating 'R-phase' transformation which complicates electrical characteristics of the wire [16]. This work therefore focuses on the highlighted region in Figure 8, which is shown (enlarged) in Figure 9. In some models, the boundaries of transformation are needed in addition to the midpoint [7]. These are noted in Figure 9– for instance the start of $\mathcal{M} \rightarrow \mathcal{A}$ transition is denoted A_s for 'Austenite Start' temperature. Similarly, $A_f, M_s,$ and M_f denote Austenite Finish, Martensite Start, and Martensite Finish temperatures, respectively. All parameters are measured at zero stress.

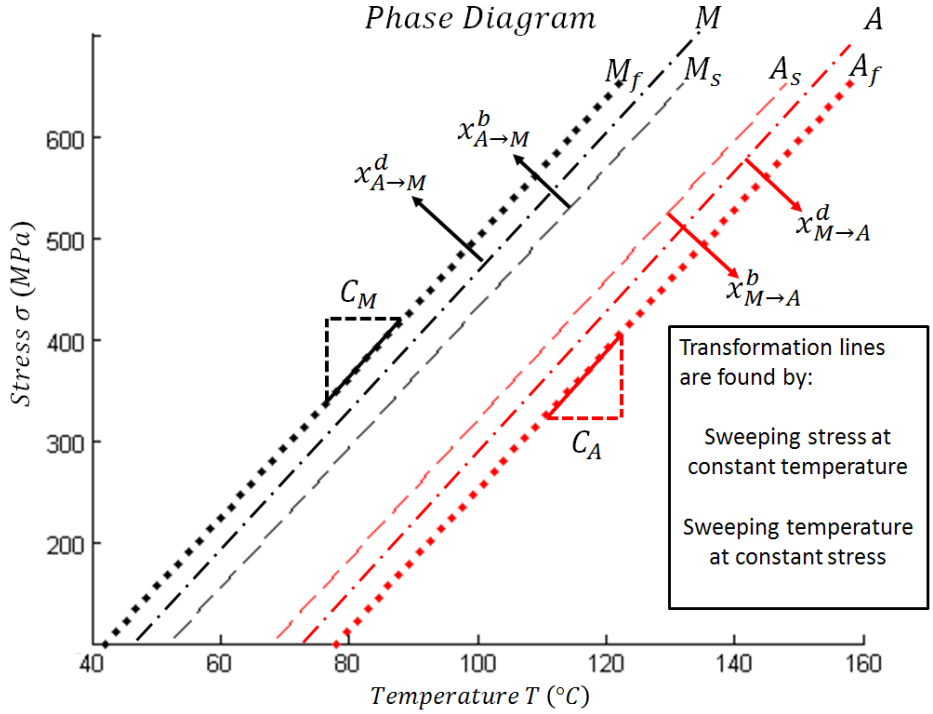


Figure 9 - Temperature-Stress Phase Diagram (at stress above 150 MPa) shows the variation of transformation regions with stress and temperature

Because the transformation regions for $\mathcal{A} \rightarrow \mathcal{M}$ and $\mathcal{M} \rightarrow \mathcal{A}$ do not overlap, the transition from one phase to another is said to have 'hysteresis', meaning the history of thermo-mechanical loading must be known to determine how the material will transform from the current state. In other words, when stress and temperature conditions place the SMA material between the M_f and

A_f boundaries, the crystal state of the wire cannot be known immediately without knowledge of the loading and thermal history up to that point. Models of this behavior will determine most of the material strain and are critical in defining the material model.

The energy required to transition from one phase to another can be observed experimentally, for instance in a Differential Scanning Calorimeter (DSC) experiment which measures heat flow vs temperature (Figure 10).

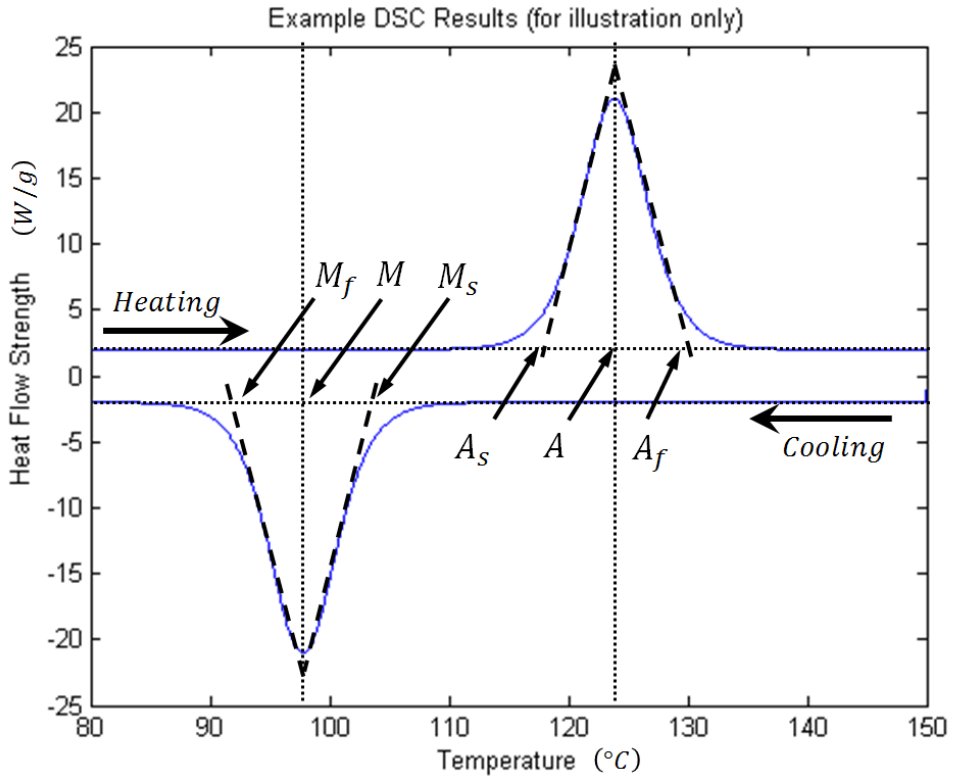


Figure 10 - Typical (simulated) DSC experiment shows temperature induced phase change from M to A, and back to M from A (with constant macroscopic stress)

As a verbal description of the behavior across the phases, two isolated causes of transformation have been named. Stress induced transformation is considered 'super-elasticity'; increased stress moves the material toward $\mathcal{M}$ from $\mathcal{A}$. Temperature induced transformation is the 'shape memory' effect; increased temperature drives the material toward $\mathcal{A}$ from $\mathcal{M}$. These loading paths are shown in Figure 11. Of course, in general the transformation does not follow either of these paths, but a combination of the two based on the thermo-mechanical loading involved.

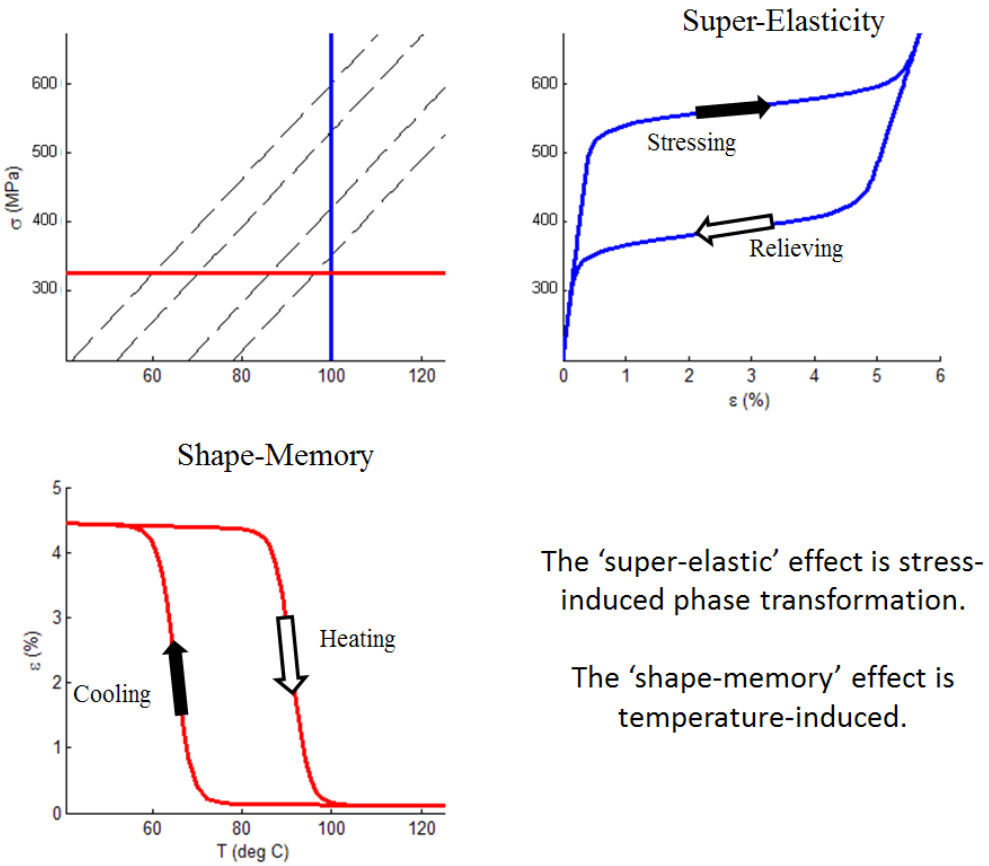


Figure 11 - Graphical explanation of the difference between the 'shape-memory' and 'super-elasticity' effects

Outline

To create a useful actuator model, the observed elastic and transformation behavior must be matched to mathematical models. In the context of actuators and designing controllers, it is more important that the models represent the 'phenomenon' of shape memory behavior rather than attempting to exactly match micro-mechanical behavior. Many SMA models are closely tied to the material physics and are not completely empirical. The key characteristics that will determine macroscopic material behavior through these transformations are comprised of four parts:

- Constitutive model - What are the relevant interacting variables?
- Mixing rule - Are the two phases acting in parallel, in series, or somewhere in between?
- Conditions for transformation - When does the phase transformation occur?
- Kinetics model of transformation - How does ξ change during transformation?

Each of these elements will be explored. All four stages have seen contributions from many researchers in many fields. In the conclusion, we provide a choice of components from each that provides the best compromise of accuracy and simplicity for SMA actuator design. This work

does not address 2D and 3D mechanics [17, 18, 19, 20], nor using thermal imaging and DIC strain field to detect transformation zones in physical samples [21, 22, 23, 24], nor finite element implementation [25, 26, 27], since these are not as immediately useful for SMA actuator control design. Our mantra in this model selection process is "minimum complexity with acceptable accuracy" – the focus is on ease of simulation and description of the material behavior for controls simulation; simple models that are sufficiently accurate are the most valuable ones.

Despite much progress, a complete model for actuator design and control still cannot be found in the literature. Most of the modelling research on SMA materials is validated using standard material tests. Unfortunately, these standard tests do not replicate some common scenarios that are encountered with physical SMA actuators. To correct this, this work develops a new model which is guaranteed to be continuous in all situations. The new model is shown to match both the old experiments and new experiments which would reveal the poor modelling of transformation direction change in old models.

Constitutive Modelling

The most common phenomenological model of SMA actuation was pioneered by Tanaka [9]. His contribution was a 'constitutive model' for super-elastic alloys derived from thermodynamics, expressed in rate form as

$$\dot{\sigma} = E(\xi)\dot{\epsilon} - \theta(\xi)\dot{T} + \Omega(\xi)\dot{\xi} \qquad (1)$$

There are four variables; stress (σ), strain (ϵ), temperature (T), and the Martensite fraction (ξ). The elastic modulus $E(\xi)$, coefficient of thermal expansion $\theta(\xi)$, and 'transformation tensor' $\Omega(\xi)$ are functions of the Martensite fraction. These quantities ($E, \theta,$ and Ω) could be more generally functions of all the constitutive variables, however they are usually approximately constant for the individual crystal phases, so their primary dependence on ξ describes the 'mixing' of properties during phase change.

Liang & Rogers [28] first noted that this model is only consistent for a certain form of $\Omega(\xi)$ based on the maximum shape memory strain with no stress ϵ_L

$$\Omega(\xi) = -\epsilon_L E(\xi) \qquad (2)$$

Brinson (with Huang in later publications) made many contributions to the model both mathematically and conceptually [7, 12, 29]. First, it was noted that the existing models cannot predict the shape memory effect at low stress and temperature because they do not distinguish between twinned and detwinned Martensite. This was corrected by separating the components of the Martensite into Twinned (ξ_T) and detwinned 'Single crystal' (ξ_S) components

$$\xi = \xi_S + \xi_T$$

There has been some controversy about the treatment of twinned and single crystal Martensite, but in general this addition must be accepted to fully describe the material [29]. In actuator design, the twinned Martensite is not typically encountered (see Figure 8); hence we will neglect the two Martensite variants unless specifically stated otherwise.

Mixing Rule

The arrangement of the two crystal phases in a 1D model is described using either parallel or series components. If the phases are modelled in parallel they have identical strain, if modeled in series they have identical stress. This modeling choice has a significant effect on the effective elastic modulus.

The original Tanaka and Liang models essentially used parallel modulus, despite being derived using thermodynamic principals. Ikuta derived essentially the same model form by assuming the material was comprised of two crystal phases in a parallel configuration [7] [30] leading to

$$\sigma = (1-\xi)\sigma_A + \xi\sigma_M = (E_A(1-\xi) + E_M\xi)\epsilon \quad (3)$$

(Curiously Ikuta used a *series* model for electrical resistance [30].) This form has been called the 'parallel' model, the 'Voigt' model, and the 'rule of mixtures' model. Ivshin and Pence were perhaps the first to use the series mode for strain [31]:

$$\epsilon = (1-\xi)\epsilon_A + \xi\epsilon_M = \left(\frac{(1-\xi)}{E_A} + \frac{\xi}{E_M}\right)\sigma \quad (4)$$

This form has been called the 'series' model and the 'Reuss' model. Brinson and Huang also recognized that the constitutive model can be derived from a micromechanics perspective

without thermodynamics [7]. They note that the constitutive model is most meaningfully rearranged as 'strain decomposition' which, again, is essentially a series model [29]:

$$\epsilon = \epsilon_{elastic} + \epsilon_{transformation} + \epsilon_{thermal} = \frac{\sigma}{E(\xi)} + \epsilon_L \xi + \frac{\theta(\xi)}{E(\xi)}(T - T_0) \quad (5)$$

This form has become the accepted constitutive model for over a decade. Brinson notes that the main controversy is no longer in the constitutive model, but in modeling the transformation kinetics [7]. For convenience, we use the model with a series mechanical Compliance (inverse of modulus) $C(\xi) = 1/E(\xi)$, such that the final complete constitutive equation is

$$\epsilon = C(\xi)\sigma + \epsilon_L \xi + C(\xi)\theta(T - T_0) \quad (6)$$

Where θ is now considered constant through the Martensite and Austenite phase. More complex mixing models such as derived by Boubaker [32] and Tanaka and Mori [33] are not considered here – they all provide a mixing rule that is somewhere between series and parallel behavior. Many authors significantly simplify the model and math involved by using the average modulus of the two phases instead of modeling the changing modulus. This introduces the error shown in the super-elastic simulation of Figure 12, where it is seen that varying modulus affects maximum strain by over 15%. In typical SMAs such as NiTi this simplification is not acceptable.

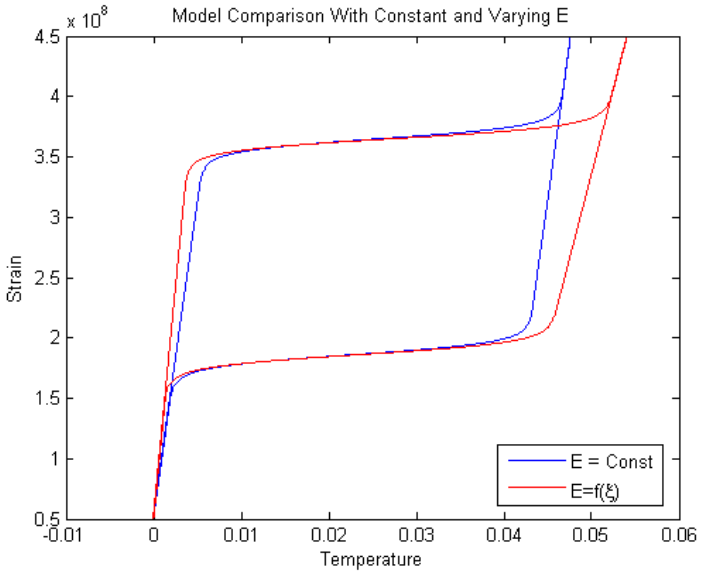


Figure 12 – Ignoring the phase-fraction dependent modulus causes large errors in the model (this simulation was performed after all the elements of this chapter have been incorporated)

Allowing the modulus to vary is important to accuracy (especially outside of the transformation region), though the choice to use series, parallel, or any intermediate mixing rule during transformation has only a small effect on the total system response (Figure 13). The effect on electrical and thermal properties is more pronounced.

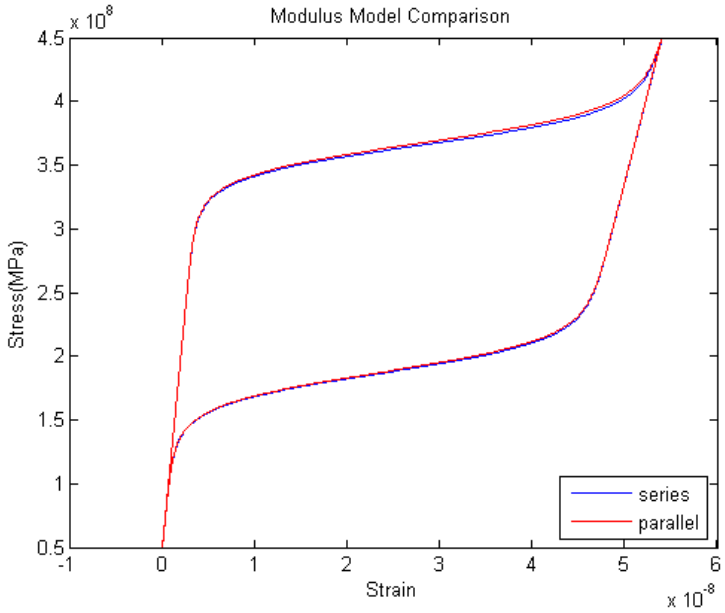


Figure 13 - Difference in response between series and parallel modulus models are not very significant.

Contribution of Linear Thermal Expansion

Because the effect of thermal expansion is small compared to the shape memory effect, the constitutive model is sometimes simplified [29]:

$$\epsilon = C(\xi)\sigma + \epsilon_L \xi \qquad (7)$$

The error from this assumption is shown in Figure 14. The error caused by this simplification is on the order of 3%, and so the simplification will not be employed for now, though it may be appropriate in some applications.

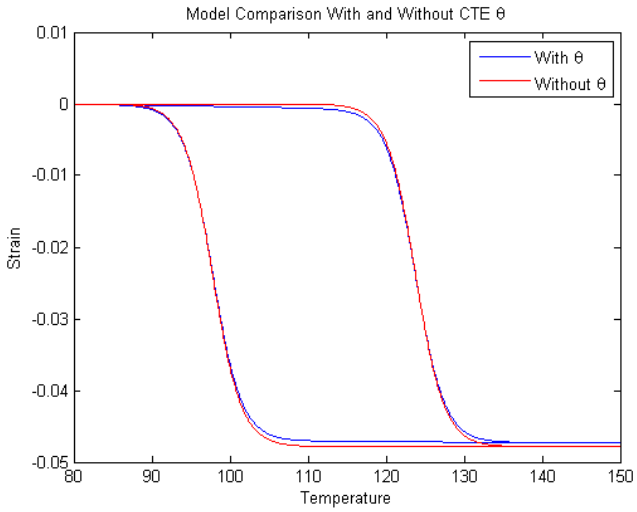


Figure 14- Ignoring thermal expansion in constitutive model causes error on order of 3% in a constant stress test

Transformation Criteria Development

Conditions for transformation determine when the phase change will occur. They are mostly based on the researcher's interpretation of the phase diagram (Figure 9). Several of these conditions sets were summarized by Elahinia [34, 35]. Tanaka first used the transformation conditions [9]

$$if \begin{Bmatrix} \dot{\sigma} < 0 \; and \; \sigma \leq -(T - A_s)C_A \\ \dot{\sigma} > 0 \; and \; \sigma \geq (T - M_s)C_M \end{Bmatrix} then \begin{Bmatrix} \xi_{M\to A} \\ \xi_{A\to M} \end{Bmatrix} \tag{8}$$

The symbol $\xi_{M\to A}$ is used to denote 'ξ is moving from Martensite to Austenite', and $\xi_{A\to M}$ expresses 'ξ is moving from Austenite to Martensite'.

Liang & Rogers [28] expressed the criteria in terms of temperature and bounded the transformation region to use his cosine kinetics model

$$if \begin{Bmatrix} \dot{T} > 0 \;\; and \; A_f + \frac{\sigma}{C_A} \geq T \geq A_s + \frac{\sigma}{C_A} \\ \dot{T} < 0 \; and \;\; M_s + \frac{\sigma}{C_M} \geq T \geq M_f + \frac{\sigma}{C_M} \end{Bmatrix}$$

$$then \begin{Bmatrix} \xi_{M\to A} \\ \xi_{A\to M} \end{Bmatrix} \tag{9}$$

Brinson [12] used these criteria but returned to a stress-based expression

$$if \begin{Bmatrix} \dot{\sigma} < 0 \; and \; (T - A_s) \geq \frac{\sigma}{C_A} \geq (T - A_f) \\ \dot{\sigma} > 0 \; and \; (T - M_f) \geq \frac{\sigma}{C_M} \geq (T - M_s) \end{Bmatrix}$$
$$then \begin{Bmatrix} \xi_{M \to A} \\ \xi_{A \to M} \end{Bmatrix}$$

(10)

The Brinson model requires additional criteria for low temperature detwinning transformation. Chung presents these criteria clearly in [15].

A more recent improvement was made by Elahinia & Ahmadian [34, 35]. Their major enhancement is a slight modification to the transformation criteria to account for isothermal stress change.

$$if \begin{Bmatrix} \dot{T} - \frac{\dot{\sigma}}{C_A} > 0 \; and \; A_s + \frac{\sigma}{C_A} < T < A_f + \frac{\sigma}{C_A} \\ \dot{T} - \frac{\dot{\sigma}}{C_M} < 0 \; and \; M_f + \frac{\sigma}{C_M} < T < A_s + \frac{\sigma}{C_M} \end{Bmatrix} then \begin{Bmatrix} \xi_{M \to A} \\ \xi_{A \to M} \end{Bmatrix}$$

(11)

Some models do not bound the transformation to lie within exact start and finish conditions [30, 31, 14]. While it is physically consistent with thermodynamics to have a discontinuous start and finish of transformation (by the Clausius-Clapyron relationship [9]), DSC experiments show that there is not a discontinuous start to transformation, and thus perhaps the start and finish conditions are merely the tails of a 'distribution' of transformation. Considering this concept, models can be built for which the transformation criteria are greatly simplified as they rely merely on the direction of temperature and stress perpendicular to transformation boundaries:

$$if \begin{Bmatrix} \dot{T} - \frac{\dot{\sigma}}{C_A} > 0 \\ \dot{T} - \frac{\dot{\sigma}}{C_M} < 0 \end{Bmatrix} then \begin{Bmatrix} \xi_{M \to A} \\ \xi_{A \to M} \end{Bmatrix}$$

(12)

For distinction, we will refer to models with a defined start and finish state as 'bounded' criteria, and those with a smooth distribution as 'distributed' models.

Kinetics Model Development

During transformation, the Martensite fraction is modeled by a function of the distance traversed across transformation boundaries (as seen in Figure 9). Several of these functions were reviewed by Zotov when developing the Richards function kinetics [36]. To clarify the following descriptions of the phase transformation 'kinetics', we define parameters that relates the stress/temperature state (within transition regions) to the distance to transformation boundaries (see Figure 9). For bounded models (note superscript $(\cdot)^b$), we look at perpendicular distance from the boundary of transformation:

$$x^b_{M\to A} \equiv -\frac{T-\frac{\sigma}{C_A}-A_f}{A_f-A_s}$$
$$x^b_{A\to M} \equiv -\frac{T-\frac{\sigma}{C_M}-M_s}{M_s-M_f}$$

(13)

For distributed models (note superscript $(\cdot)^d$), we look at perpendicular distance from the center band of transformation:

$$x^d_{M\to A} = T-\frac{\sigma}{C_A}-A$$
$$x^d_{A\to M} = T-\frac{\sigma}{C_M}-M$$

(14)

The simplest bounded model is a linear transformation, where the Martensite fraction is simply the distance across the boundary of (13):

$$\xi_{M\to A} = x^b_{M\to A}$$
$$\xi_{A\to M} = x^b_{A\to M}$$

(15)

This has been used in multiple studies for its simplicity [37]. On experimental and energy-balance grounds, Malukhin used square root kinetics. The square-root function is simply [38]:

$$\xi_{M\to A} = \sqrt{x^b_{M\to A}}$$
$$\xi_{A\to M} = \sqrt{x^b_{A\to M}}$$

(16)

Tanaka used an exponential model from his previous experience with phase transformation in other materials [9]:

$$\xi_{M\to A} = \exp[\ln(.01)\,(1 - x_{M\to A}^{b})]$$
$$\xi_{A\to M} = 1 - \exp[\ln(.01)\,x_{A\to M}^{b}]$$

(17)

Because the exponential does not have a defined finish point, the natural log determines how close to complete the transformation should be before rounding ξ to one. This model also does not account for conditions where the SMA element reverses transformation direction before completing the transformation from one phase to another. When transformation is not complete, the hysteresis is still present but creates inner loops of transformation. Because of the need to model incomplete transformation and because the discontinuous boundary in the Tanaka kinetics model is mathematically unpleasant, Liang and Rogers [28, 39] used the transformation

$$\xi_{M\to A} = \frac{\xi_{\mathrm{M}}}{2}\cos\left(\pi - \pi x_{M\to A}^{b}\right) + \frac{\xi_{\mathrm{M}}}{2}$$
$$\xi_{A\to M} = \frac{1-\xi_{A}}{2}\cos\left(\pi - \pi x_{A\to M}^{b}\right) + \frac{1+\xi_{A}}{2}$$

(18)

This cosine kinetics is widely used [12, 7, 34] because it has smooth boundaries at the start and finish of transformation. This model developed allows reversal of transformation when the phase fraction is not 0 or 100%. It does this by tracking the extent of transformation,

$$While \begin{Bmatrix}\xi_{M\to A}\\ \xi_{A\to M}\end{Bmatrix}, \quad \begin{Bmatrix}\xi_{A} = \xi\\ \xi_{M} = \xi\end{Bmatrix}$$

(19)

These ‘history variables’ are important for practical SMA actuator models that often reverse direction without complete transformation. The form of these history variables is the same regardless of other modelling choices.

Distributed models can use any function of temperature and stress which is bounded between 0 and 1 for all temperatures and stresses. Nearly any of the common population growth models [40] could be used for fitting; the function choice is a trade-off between number of fitting parameters and complexity. The most common model is the logistic function (the ‘3 parameter’ curve), commonly used in population growth modeling, which can be written [30, 14, 41]

$$\xi_{M\to A} = \frac{\xi_M}{1+\exp\left(kx^d_{M\to A}\right)}$$
$$\xi_{A\to M} = \frac{1-\xi_A}{1+\exp\left(kx^d_{A\to M}\right)} + \xi_A$$

(20)

Where k is a fitting parameter determined from DSC experiments that determines the distribution of the transformation band. Or equivalently [31] [7]:

$$\xi_{M\to A} = \frac{\xi_M}{2}\tanh\left(\frac{-kx^d_{M\to A}}{2}\right) + \frac{\xi_M}{2}$$
$$\xi_{A\to M} = \frac{1-\xi_A}{2}\tanh\left(\frac{-kx^d_{A\to M}}{2}\right) + \frac{1+\xi_A}{2}$$

Researchers have used constant k ranging from $5.44/(A_f - A_s)$ [14], $6.2/(A_f - A_s)$ [30] or higher based on how the data is extracted from a DSC experiment. Preferably, the value of k can be determined by the 'confidence' of A_f and A_s. If the A_f and A_s boundaries are defined as 1% transformation locations [42]:

$$k = \frac{\ln(1/.01^2)}{(A_f - A_s)} = \frac{9.21}{(A_f - A_s)}$$

This form of the kinetics has been widely adopted by controls engineers after Ikuta [30, 41, 43]. Of particular interest in simulation are the derivatives of this kinetics model. Jayender and Zhu provide analytical expressions for the transformation rate [43, 42]. The author prefers a different presentation based on the foundational logistic equation:

$$\left(\frac{\partial\xi}{\partial T}\right)_{M\to A} = -C_M\left(\frac{\partial\xi}{\partial\sigma}\right)_{M\to A} = \frac{-k\xi_m\exp(kx)}{(1+\exp(kx))^2} = k\left(\frac{\xi^2}{\xi_m} - \xi\right)$$
$$\left(\frac{\partial\xi}{\partial T}\right)_{A\to M} = -C_A\left(\frac{\partial\xi}{\partial\sigma}\right)_{A\to M} = \frac{-k\xi_a\exp(kx)}{(1+\exp(kx))^2} = k\left(\frac{(1-\xi)^2}{\xi_A} - (1-\xi)\right)$$

These functions have a maximum value at the center of transformation equal to $k/4$. This is an important consideration as it is the maximum 'gain' between temperature and phase fraction. The time rate of change can be computed similarly. Considering that Martensite fraction is a function of two variables, the time rate of change can also be presented, when convenient, as:

$$\dot{\xi} = \frac{\partial\xi}{\partial T}\left(\dot{T} - \frac{\dot{\sigma}}{C}\right) = \frac{\partial\xi}{\partial T}\dot{T} + \frac{\partial\xi}{\partial\sigma}\dot{\sigma}$$

More recently, Zotov et al studied the transformation using X-Ray powder Diffractograms (XRD) and saw that experimental data could be fit extremely well to the generalized logistic function (the 'Richards' function or '4-parameter' growth curve) [36, 40]. Zotov only presents the model for the Martensite to Austenite transformation, but both transformation directions can be modeled:

$$\xi_{M\to A} = \frac{\xi_M}{\left(1 + \nu_A \exp\left(k x_{M\to A}^d\right)\right)^{\frac{1}{\nu_A}}}$$

$$\xi_{A\to M} = \frac{1 - \xi_A}{\left(1 + \nu_M \exp\left(k x_{A\to M}^d\right)\right)^{\frac{1}{\nu_M}}} + \xi_A$$

(21)

To practically implement this from the common DSC parameters, find the inflection point of transformation by solving the following for ν:

$$\frac{A_f - A_s}{A - A_s} = (\nu_A + 1)^{\frac{1}{\nu_A}}$$

$$\frac{M_s - M_f}{M - M_f} = (\nu_M + 1)^{\frac{1}{\nu_M}}$$

(22)

And choose the distribution constant k to fit the data (though $k = \nu/2$ is a good choice if data is lacking). Of course, the Richards function reduces to the simple logistic function when $\nu = 1$. All these functions give a wide range of transformation characteristics. The responses are compared in Figure 15. In a super-elasticity experiment, these equations give rise to theoretical stress strain curves as shown in Figure 16.

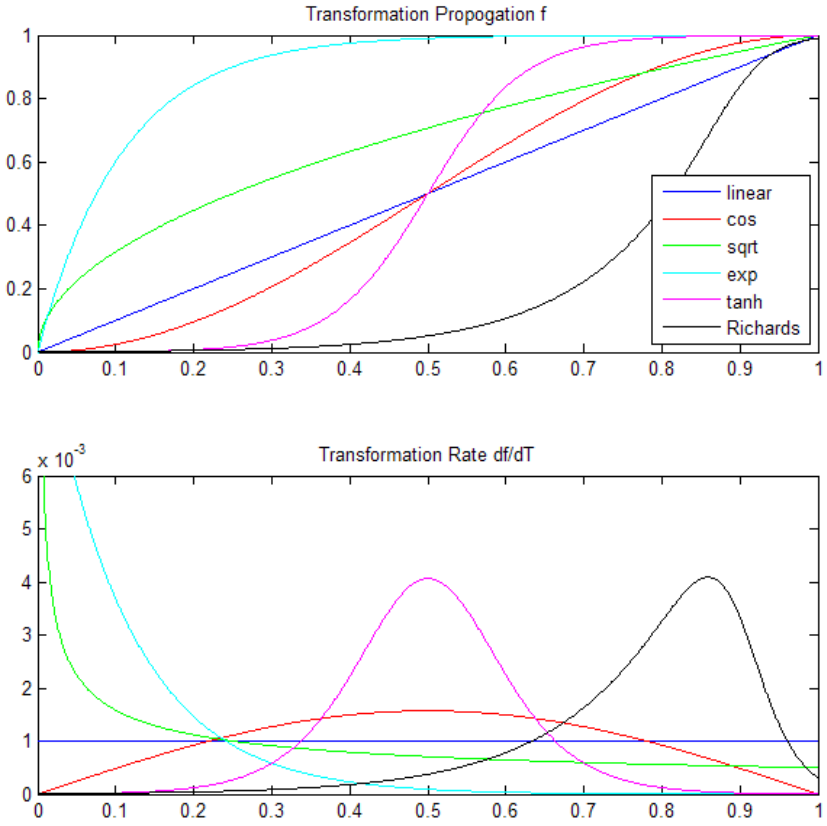


Figure 15- Comparing the transformation kinetics equations

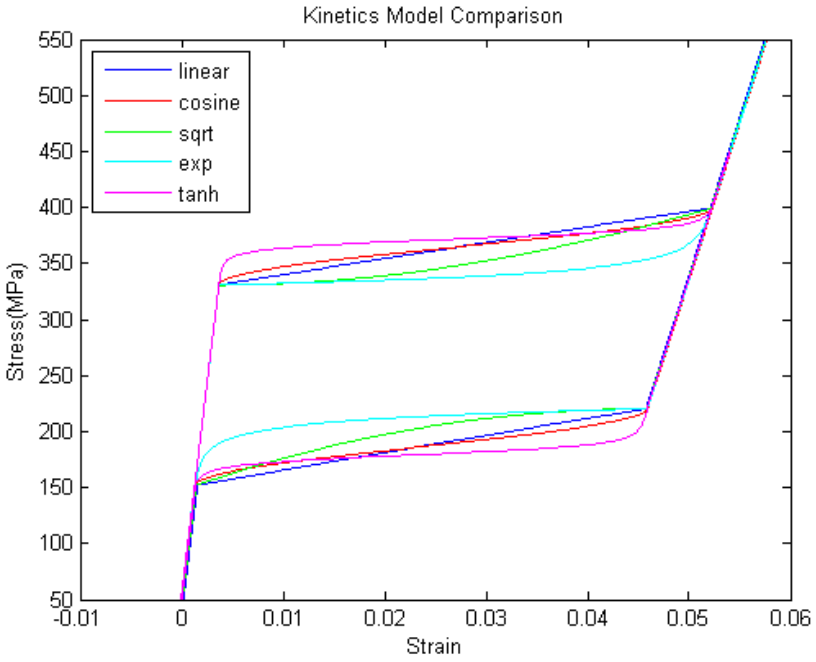


Figure 16 - Comparing the various kinetics models in a stress-strain plot showing super-elastic behavior at constant temperature

An additional, even more complex kinetics model was developed by Heintze and Seelecke based on the statistics of thermally activated processes for a single crystal of the material [44, 45, 46, 47, 48]. This model exceeds the complexity that can reasonably be used for controls modeling.

Summary of Mechanical Model

From the above development, we select a compromise of accuracy and implementation complexity from all model components. In the ends, equations (1), (4), (12), and (20) are employed in our model for constitutive model, transformation conditions, and transformation kinetics, respectively. These are here summarized for convenience:

$$\epsilon = C(\xi)\sigma + \epsilon_L \xi + C(\xi)\theta(T - T_0) \qquad (23)$$

$$if \begin{Bmatrix} \dot{T} - \frac{\dot{\sigma}}{C_A} > 0 \\ \dot{T} - \frac{\dot{\sigma}}{C_M} < 0 \end{Bmatrix} then \begin{Bmatrix} \xi_{M\to A} = \frac{\xi_M}{1 + \exp\left(k\left(T - \frac{\sigma}{\alpha} - A\right)\right)} \\ \xi_{A\to M} = \frac{1 - \xi_A}{1 + \exp\left(k\left(T - \frac{\sigma}{\alpha} - M\right)\right)} + \xi_A \end{Bmatrix}, \begin{Bmatrix} \xi_A = \xi \\ \xi_M = \xi \end{Bmatrix} \qquad (24)$$

In practice, the slope of the Martensite and Austenite transformation bands are identical, so a single term α is used in place of separate C_A and C_M parameters. The series compliance is a function of Martensite fraction:

$$C(\xi) = \left(\frac{(1 - \xi)}{E_A} + \frac{\xi}{E_M}\right)$$

These equations were implemented in a MATLAB simulation such that the model accepts arbitrary stress and temperature inputs as functions of time. Depending on the rate of strain, it has been found that equations such as these are numerically 'stiff'. Improved simulation times can be found using adaptive time-step solutions, however the author found that the strain equations did not require stiff solvers (c.f. [49, 50]). An example simulation is shown in Figure 17. These equations have been implemented in MATLAB and in SIMULINK. The way this model fits neatly as an actuator model in a controller simulation is shown in Figure 19.

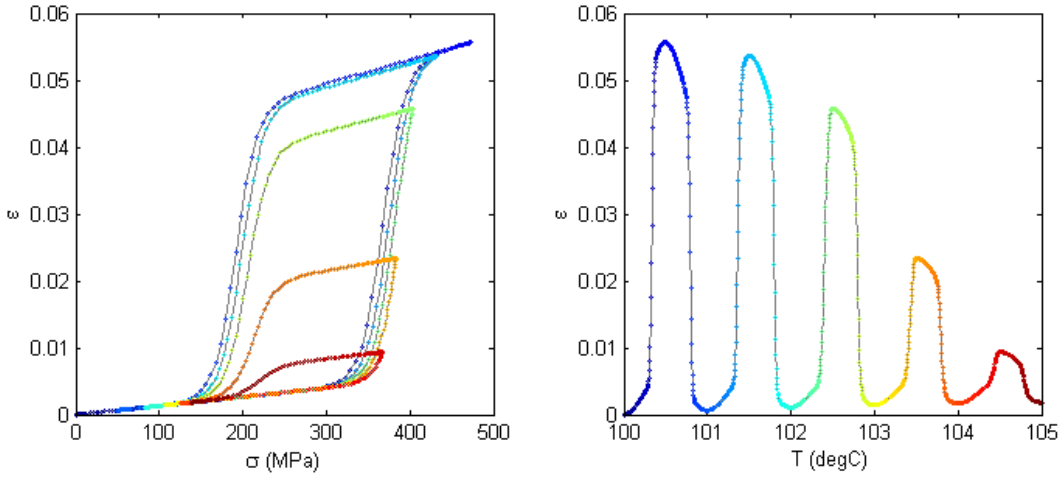


Figure 17 - Example strain (model output) response to arbitrary stress and temperature, where stress is cycled while temperature slowly increases. The sensitivity to temperature change is significant – a five degree change completely alters the behavior.

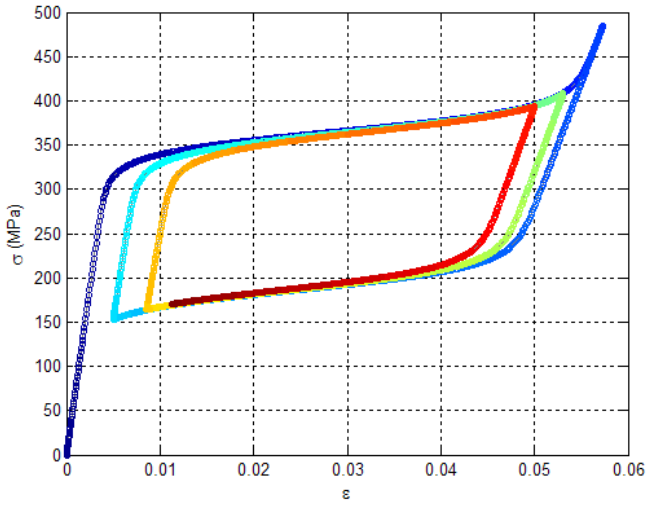


Figure 18 - Example stress (model output) response to sinusoidal strain of decreasing amplitude at constant temperature ($\boldsymbol{\epsilon = 0.03 * (1 - cos(t))/(1 + t/5)}$). Color denotes time (blue at time 0, to red at final time).

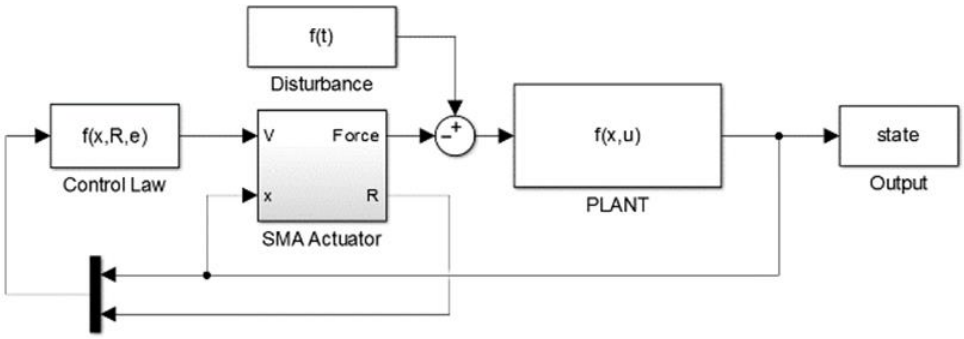


Figure 19 - SMA actuator model located in a feedback control block-diagram

Patching Holes in Existing Constitutive Models

Models for SMA materials have two main parts; the constitutive relationships between stress, strain, temperature, and crystal phase, and the kinetics equations which govern the way the crystal phase varies due to stress, strain, and loading history. In most one-dimensional models, the kinetics equations are chosen empirically to match experimental data as the phase transforms from one state to the other. In an experiment with a Differential Scanning Calorimeter, the transformation rate is assumed to be proportional to the heat flow – so good kinetics equations are those which match the transformation rates of the material. A few common kinetics equations are shown in Figure 15. These equations do a good job describing the transformation when a complete change is made from 100% Austenite to 100% Martensite, or vice-versa. One such equation which provides a good compromise of simplicity and accuracy is the logistic function, shown in equation (25). A complication arises when the model must be used when the loading path switches direction in the middle of the transformation. Typically, this is solved by tracking an intermediate history variable, denoted as ξ_A and ξ_M , which store the extent of transformation and therefor change the kinetics equation when the direction of transformation is reversed.

$$
\begin{gathered}
if \begin{Bmatrix} \dot{T} > 0 \\ \dot{T} < 0 \end{Bmatrix} \\
then \begin{Bmatrix} \xi_{M\to A}^{\text{inner loop}} = \dfrac{\xi_M}{1+\exp\big(k(T-A)\big)} \\ \xi_{A\to M}^{\text{inner loop}} = 1 - \dfrac{\xi_A}{1+\exp\big(-k(T-M)\big)} \end{Bmatrix} \\
if \begin{Bmatrix} \dot{T} > 0 \\ \dot{T} < 0 \end{Bmatrix} then \begin{Bmatrix} \xi_A = 1-\xi \\ \xi_M = \xi \end{Bmatrix}
\end{gathered}
\qquad (25)
$$

In most demonstrations of SMA models, the equations work well to describe the transformation even with inner loops. In this section, all examples show the transformation due to simple changes in temperature while stress is held constant. Consider the case where the SMA temperature starts at room temperature ($T = 20$), is raised to cause a 50% transformation ($T = 70$), dropped to near the nominal A_s temperature ($T = 50$), raised to a high temperature

($T = 100$) where Austenite is guaranteed, then reduced to room temperature ($T = 20$) (Figure 20). In this case, the transformation looks smooth and the path is realistic.

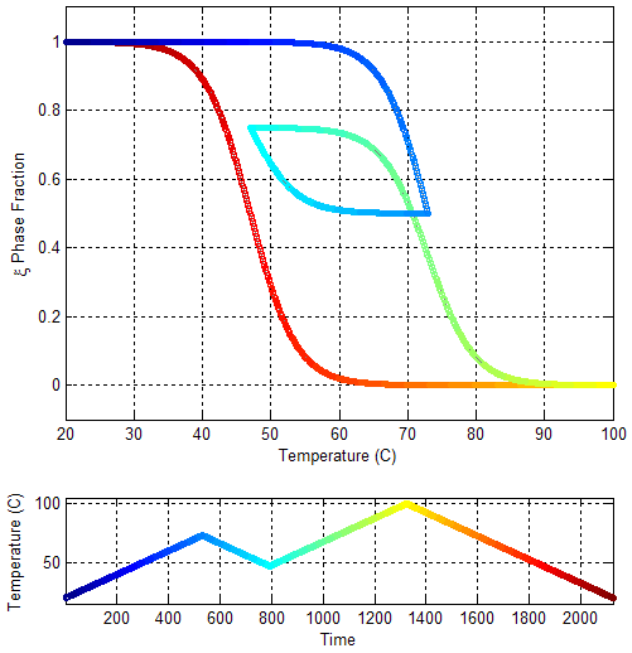


Figure 20 - Existing Kinetics models work well when the periods of transformation are wide. Color denotes the progress in time; blue is the start and red is the end of simulation

Now, consider the case where temperature is raised to cause a 50% transformation, dropped slightly, then raised to a high temperature where Austenite is guaranteed (Figure 20). In this case, the flaw in existing models is revealed – a large discontinuity appears in the response. This is caused by the poor modelling assumption that transformation is initialized at 50% (true) at a temperature that would ensure 0% Austenite (false!). Thus, the gap is due to forcing the phase fraction onto a path that doesn't correspond to its history.

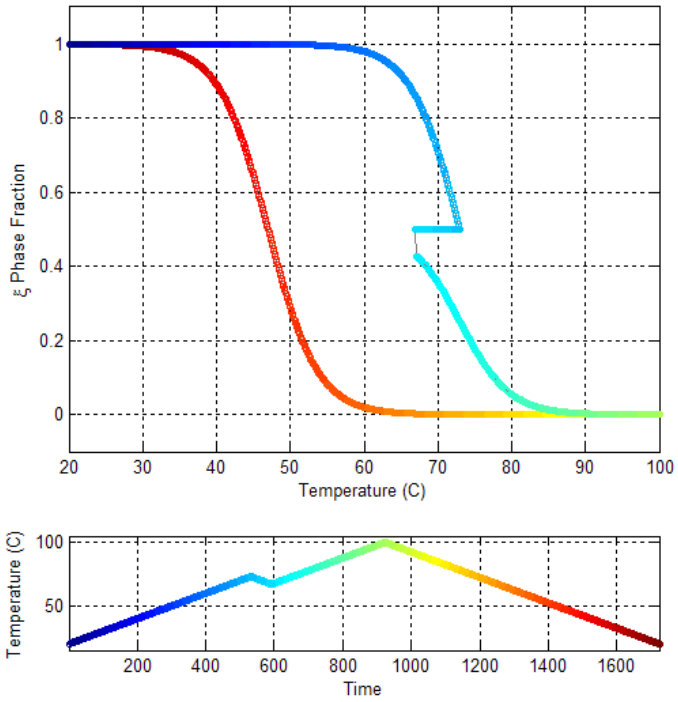


Figure 21 – Unfortunately, existing models do not work well when changing direction quickly within transformation bands – notice the large jump in transformation when direction is changed repeatedly

The first half of the proposed solution is to preserve the kinetics model as stated, but manipulate the history variables to ensure that the model is continuous. To do this, at each point where the transformation direction reverses, we compute the ξ_A or ξ_M that would have led, in a complete transformation, to the current phase fraction under the current stress and temperature conditions. Mathematically, the history variables are recomputed as follows

$$if \begin{Bmatrix} \dot{T} > 0 \\ \dot{T} < 0 \end{Bmatrix}$$

$$then \begin{Bmatrix} \xi_M = \xi\left(1 + \exp\left(k(T - A)\right)\right) \\ \xi_A = (1 - \xi)\left(1 + \exp\left(-k(T - M)\right)\right) \end{Bmatrix} \quad (26)$$

Again, this simply solves for the history variable that would have been necessary to reach this phase fraction at this temperature. The results clearly 'patch the hole' in the model as shown in Figure 22.

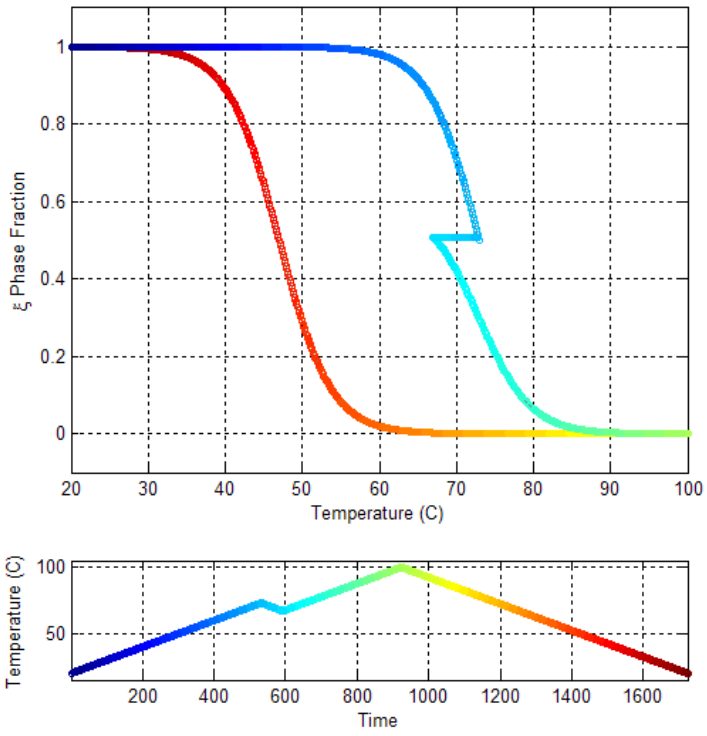


Figure 22 - One method to remove the discontinuity is to implicitly solve for the hysteresis variable such that the path is continuous

However, this model does not quite accurately represent the phase fraction motion seen in experiment. In experiment, the phase would not follow the long path inside the hysteresis boundaries – it would more quickly approach the outer band. It is well-known that the outer band (common) hysteresis model does not accurately describe SMA material behavior. However, in this case, it is not much less accurate than the implicit solution of equation (26). The equation that governs the outer-loop model is simple since we simply keep all phase change on the boundaries. The resulting path when the SMA material is subjected to the short change in direction is shown in Figure 23.

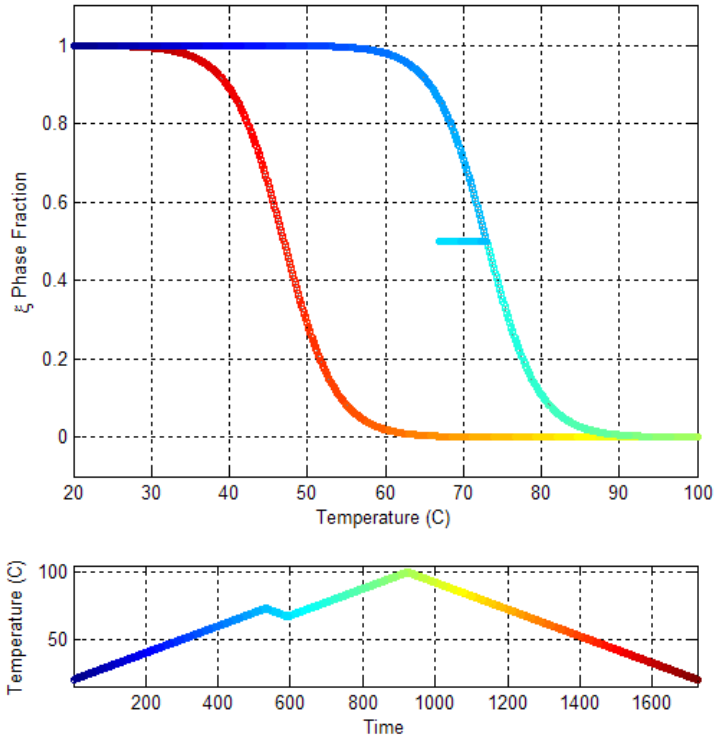


Figure 23 - Physically, the 'outer-loop' model is simpler and more accurate in controls context

The outer loop hysteresis can be modelled using:

$$if \left\{ \begin{matrix} \dot{T} > 0 \\ \dot{T} < 0 \end{matrix} \right\}$$

$$then \left\{ \begin{matrix} \xi_{M\to A}^{\text{outer loop}} = \min\left(\xi, \dfrac{1}{1+exp\big(k(T-A)\big)}\right) \\ \xi_{A\to M}^{\text{outer loop}} = \max\left(\xi, \dfrac{\xi_A}{1+exp\big(k(T-M)\big)}\right) \end{matrix} \right\}$$

Finally, we propose that a model which matches the experimental data best is not either of these, but rather a weighted sum of the two.

$$if \left\{ \begin{matrix} \dot{T} > 0 \\ \dot{T} < 0 \end{matrix} \right\}$$

$$then \left\{ \begin{matrix} \xi_{M\to A} = \alpha\, \xi_{M\to A}^{outer\ loop} + (1-\alpha)\xi_{M\to A}^{inner\ loop} \\ \xi_{A\to M} = \alpha\, \xi_{A\to M}^{outer\ loop} + (1-\alpha)\xi_{A\to M}^{inner\ loop} \end{matrix} \right\}$$

weighted with a proportional factor $0 < \alpha < 1$. The responses to small and large excursions of phase transformation path using this model are shown in Figure 24 and Figure 25, respectively. The weighting factor is 0.2, giving the outer loop model preference over the implicit inner model solution.

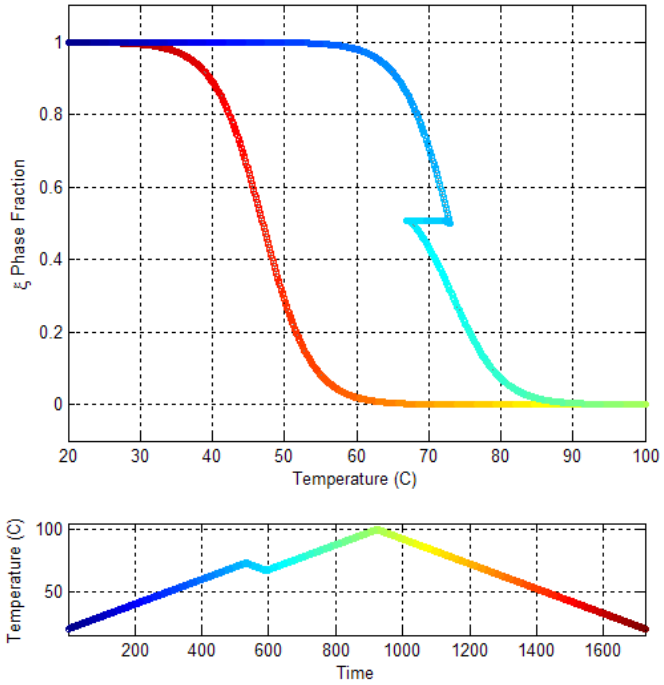


Figure 24 - The solution that matches physical data best is a weighted sum of the implicit and outer-loop methods. This is an example with a small drop in temperature during transformation

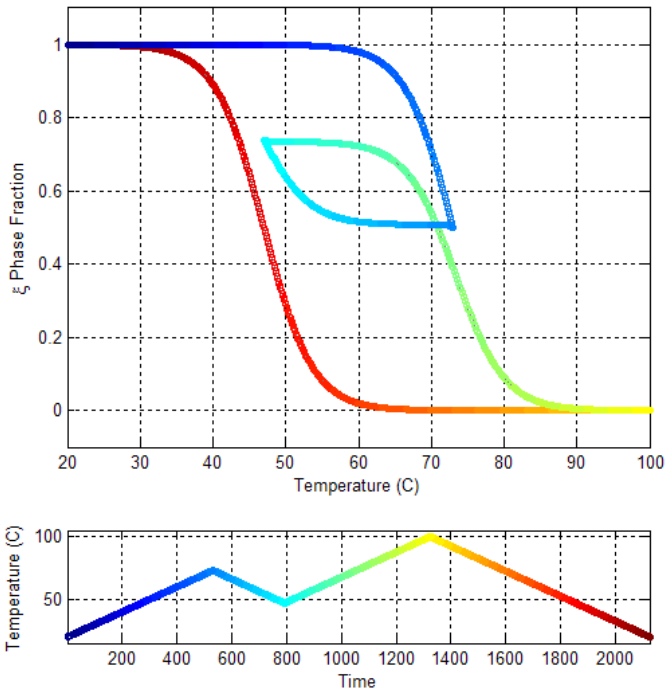


Figure 25 - This example shows the weighted model during a large drop in temperature during transformation

Experimental Results

To validate the 1D SMA model for actuators, a simple SMA wire was tested in super-elastic condition in the Hardware-in-the-loop testing machine developed as part of this thesis (Chapter 8). The super-elastic condition is easiest to test because varying temperature is difficult to control for fine wires, while varying stress is easy to apply. Two experiments were performed to show the outer loop characteristics and two different cases of inner loop hysteresis.

The first experiment demonstrates a large hysteresis loop where stress is applied, relieved to 50%, then reapplied to the maximum. This situation is usually handled accurately by the existing models because it shows such a large motion that any discontinuity when switching direction seems to vanish. In addition to matching the hysteresis behavior, the model accurately captures the Austenite and Martensite modulus (stress-strain slope) near the boundaries, as well as the total pseudo-elastic strain and hysteresis stress levels (Figure 26). The model does show an inaccuracy – it is symmetric when entering and exiting the pseudo-plastic region, whereas the physical material has different slopes entering and exiting the region. This discrepancy is expected: because the model we selected uses logistic function kinetics the nonlinearity is assumed to be symmetric. A higher order growth model (as used by Zotov) would slightly increase the accuracy, but with a large penalty on model complexity.

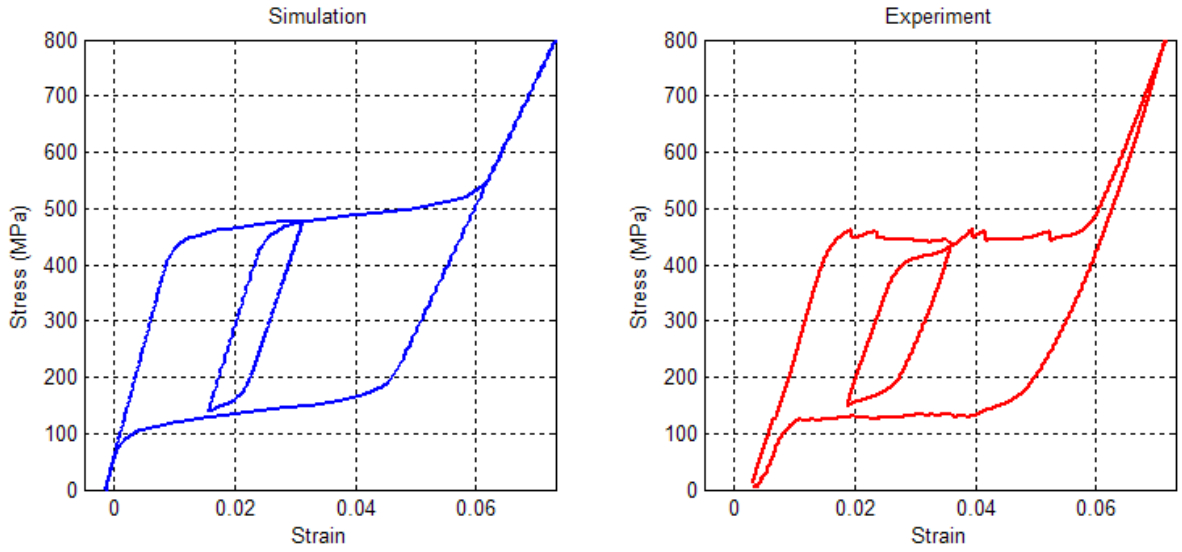


Figure 26 - Comparison between the model (left) and experiment (right) when a large stress oscillation causes an inner hysteresis loop. This situation is demonstrated in the literature often, and can be replicated by most models accurately.

The second experiment demonstrates the model response when a small (10%) drop in stress occurs during loading, and load subsequently increases to the maximum value. This deviation is small enough that most models would have 'skipped' when the load begins to increase again. The new model accurately returns quickly and smoothly to the outer loop boundary as predicted in Figure 24.

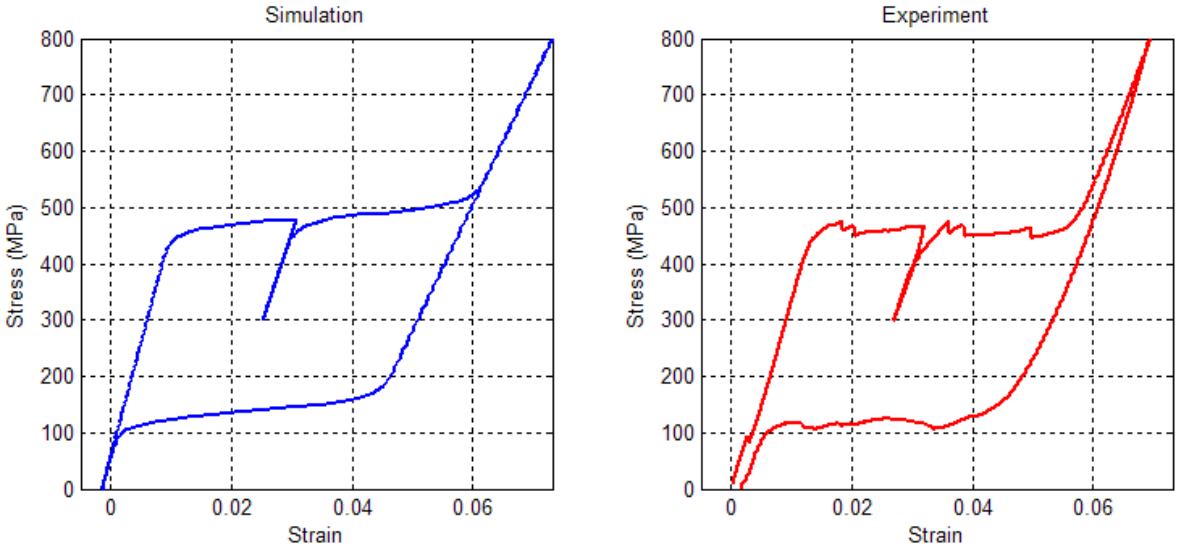


Figure 27 - Comparison between the model (left) and experiment (right) when a small stress deviation causes a hysteresis 'notch'. This situation causes most SMA models to fail with a large discontinuity. The model developed in this work is stable and accurately shows the deviation and return to the outer loop.

Conclusion

The intent of this model is to replicate the behavior of SMA materials used as actuators, in a model that is reliable in all conditions, and which can represent the phenomenon of shape memory behavior accurately. Most importantly, every development made by any researcher was evaluated and judged to see if it increased accuracy, and at what cost. The final model has been shown to match experimental behavior in all conditions in a simple form. It maintains the exact constitutive relationships required by the material science. It models crystal phase change based on a model that is common for crystal growth and propagation, which is appropriate for the crystal phase change seen in SMA materials. Some simplifying assumptions were seen to be acceptable for actuators (e.g. ignoring thermal expansion) and some were seen to have critical effects on the behavior (e.g. phase dependent elastic modulus). In experiment, the model accurately replicates situations which existing models reveal discontinuities and other flaws. This model can now act as a basis for simulations and design work for SMA actuators. Whenever

SMA material is simulated in other components of this work, the model described in this chapter is used as the reference. This model has been implemented in MATLAB. The coefficients used in these models are presented in the Nomenclature preceding Chapter 1. These coefficients were gathered experimentally when possible, and by using averages from values found in the literature if they could not be measured by the author.

Chapter 4

Heat & Surface Treatment, Training, and Fatigue

While straight Shape Memory Alloy (SMA) wire can be used directly in some applications, many of the benefits of shape memory require training the straight wire into a different memorized shape. Unfortunately, most treatment and training techniques for SMA actuators are proprietary. This chapter is a collection of published data about heat treatment, cycling to stabilize strain, two-way training techniques, and surface treatment procedures for NiTi shape memory actuators. Much of the existing data is contradictory and cannot be reconciled between each researcher's descriptions of the processes. However, we can conclude that most commercial SMA wire should be heat treatment near 500°C and quenched in water – this treatment provides a wire that can achieve high stress and strain, prevents R-phase transformations, and stabilizes within thirty complete actuation cycles. Two-way training in NiTi provides only weak effects, and is not recommended over antagonist wires or springs. Surface treatments typically do not improve the SMA behavior as actuators. However, some coatings and treatment methods can electrically insulate the wire surface.

Introduction

NiTi SMA is often purchased by researchers in the 'as-drawn' condition – wire in this condition is cheap and available in bulk. This wire usually has a manufacturer designed transformation temperature, but usually also has significant residual stress from the drawing process which makes it stiff and not useable for actuation. To utilize the shape-memory capability of the SMA, the wire must be annealed to remove the residual stress and the transition temperature must be changed to something above room temperature so that the material is Martensite at room temperature, and heating causes a transition to Austenite. The transition temperature can be set be annealing the material at controlled temperatures. Conveniently, the same annealing step used to set transition temperature also sets the Austenite 'memorized' shape. The final shape, the transition temperature, and the maximum recovery/actuation stress achievable are all set simultaneously. Wires that are purchased that already exhibit the shape memory effect (such as Dynalloy Flexinol [51]) can be retrained in the same manner – despite

the manufacturer's warning that this will decrease the performance we have not found that to be the case when proper treatment and training procedures are followed.

Alloy Composition

Most research labs purchase off-the-shelf SMA wire or bar and do not have the facilities required to create new alloys. However, the alloying determines most of the wire's ability to demonstrate shape memory effects. The most common NiTi shape memory alloys are nearly 50% Ti; the following suggestions relate to materials of that composition. Transformation temperature can be changed in alloying by changing the ratio of nickel to titanium in the metal (Figure 28), though this is out of reach to the end-user and is typically set by the manufacture (commonly at 70C or 90 C). Dynalloy, one of the largest shape memory NiTi manufacturers in the USA, claims that in their Flexinol wire "both nickel and titanium atoms are present in the alloy in almost exactly a 50%/50% ratio." Experimental testing of transformation temperature vs atomic fraction of Nickel was performed by Duerig [52], showing that a large range of transformation temperatures can be achieved within a range of only 3 percent near 50%/50% alloy (Figure 28). Adding other metals to the alloy can have large effects on performance, but almost always with a loss of SMA performance or an increase in cost that makes it less useful for most users.

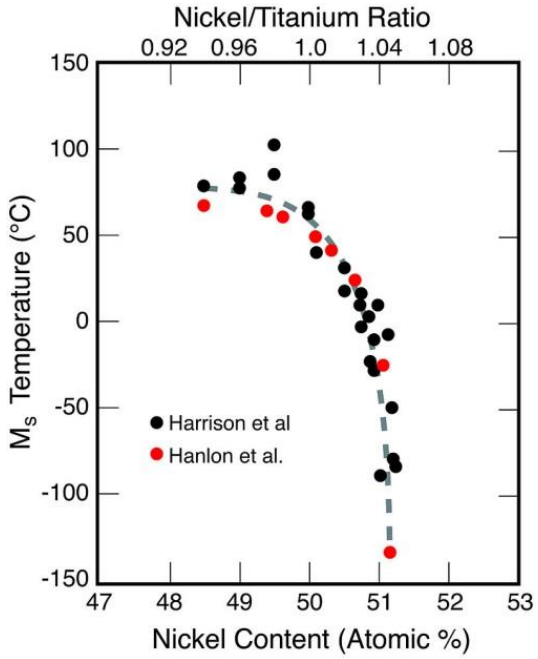


Figure 28 – The effect of alloy composition on Martensite Starting temperature (Tom Duerig [52])

Heat Treatment Temperatures

The most obvious processing parameter in NiTi training is by changing the heat treatment (i.e. annealing) temperature, disclosed in a US patent [53]. It has since been studied widely with much agreement in the general effects but some disagreement about the exact resulting SMA properties. Sadiq et al concluded that the transformation temperature is generally increased by raising annealing temperature, though in a nonlinear manner (Figure 29) [54]. By annealing at high temperatures, the R-phase transformation can be eliminated. Sadiq et al and Huang found R-phase was eliminated above 500 C [54, 55], Uchil et al found 410C was adequate to remove it [56, 57, 58] . Elimination of R-phase is very useful for creating actuators that will use electrical resistance for strain estimation. They also concluded that annealing above 600 C significantly reduces recovery stress – the stress required to de-twin the Martensite when the wire is cooled. The recommended maximum training temperature in that study was 450 C for general actuator creation [54].

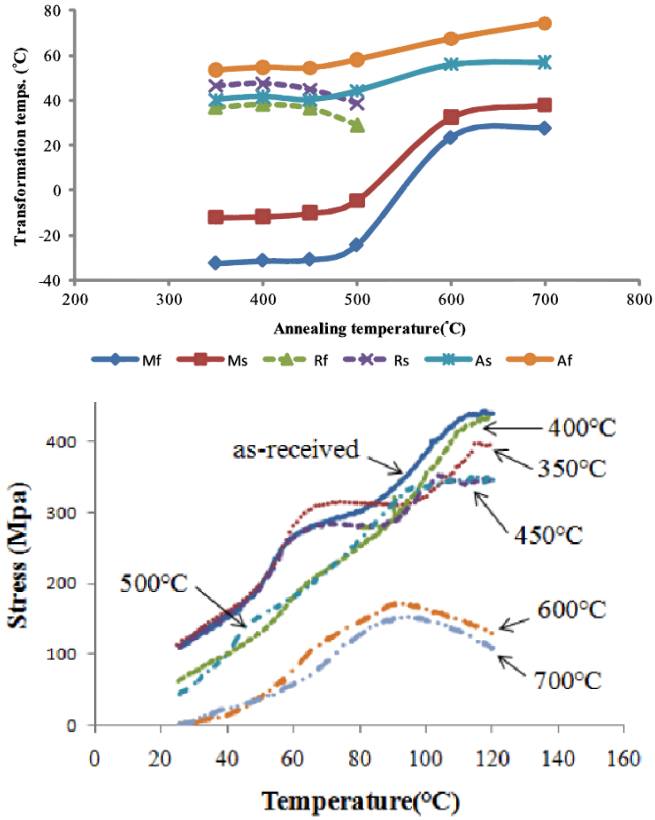


Figure 29 - Effect of annealing temperature on transformation temperature and on recovery stress at 4% strain, from Sadiq et al 2010

The annealing temperature also affects the permanent deformation in the first cycles after the heat treatment; permanent deformation is significantly increased for annealing temperatures above 400 C [59]. The heat treatment time has a large effect on the transformation stress and ultimate tensile strength as well. Models by Morgan show that higher training temperature decreases transformation temperature, decrease stress required to transform from Austenite to Martensite, and decrease ultimate tensile strength [60]. Longer heat treatment times have the same effect, controlling super-elastic stress more than other parameters (Figure 30).

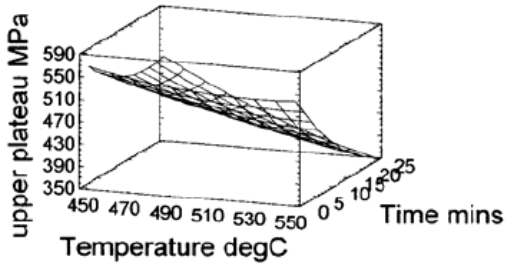


Figure 7: Surface response of upper plateau stress.

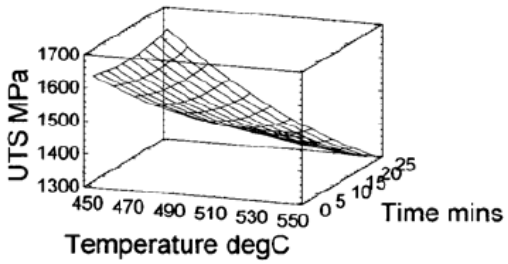


Figure 8: Surface response for ultimate tensile stress.

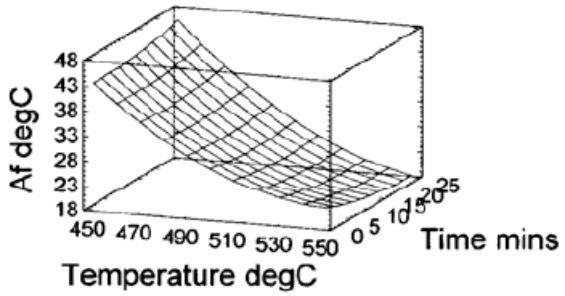


Figure 9: Surface response for A_f temperature.

Figure 30 - Effect of heat treatment temperature and time on super-elastic behavior [60]. 'Upper plateau stress' is the stress required to force Austenite into Martensite.

Curiously, the material after annealing is so exactly crystalline that the shape memory behavior is extremely high for the first few cycles after training. Thus it has proven necessary to cycle the material for as many as 30-150 cycles before shape memory deformation characteristics are reliable [57, 61, 62]. The number of cycles required is marginally reduced for low annealing temperatures (below 500 C) [57]. The resistance stabilizes in a similar number of cycles [63]. In some alloys, forcing the actuator to a large strain while at high temp (Austenite) will immediate stabilize the transformation strain behavior [64]. There is an additional consideration beyond the training heat cycle – Aging. By aging near 400 C, the transformation temperatures can be increased; more time in the forge leads to higher transformation temperatures [65].

From these suggestions, we recommend and employ the following training process for one-way shape memory:

1. Constrain into desired hot (memory) shape
2. Anneal sample at a temperature from 490℃ to 510℃ for 5 min (while constrained)
3. Quench in cold water
4. Remove constraints
5. Cycle 30 times to stabilize range of motion

Two-way Memory Effect

By a multi-stage training process, most shape-memory alloys can be trained to exhibit some two-way shape memory effect [66]. By cold-working into the desire cold shape, and recalling the memorized hot shape, the material gains internal flaws which guide the cooling material into a preferred orientation when cold. Unfortunately, this training into two-way memory decreases the total recovery strain, and decreases the transition temperatures of the material [66]. Luo & Abel found that only one of the published two-way training techniques was effective in non-axial recovery stress. The training process is outlined as follows:

1. Heat sample well above A_f, but below annealing temperature
2. Load it (while heated) to a desired cold shape
3. Hold in position and cool to below M_f
4. Unload completely
5. Repeat 1-4 for at least 30 cycles

In some cases, it is possible to introduce two-way effect by compressive load cycling - this leads to a maximum two-way recovery stress of nearly 2.2% [67].

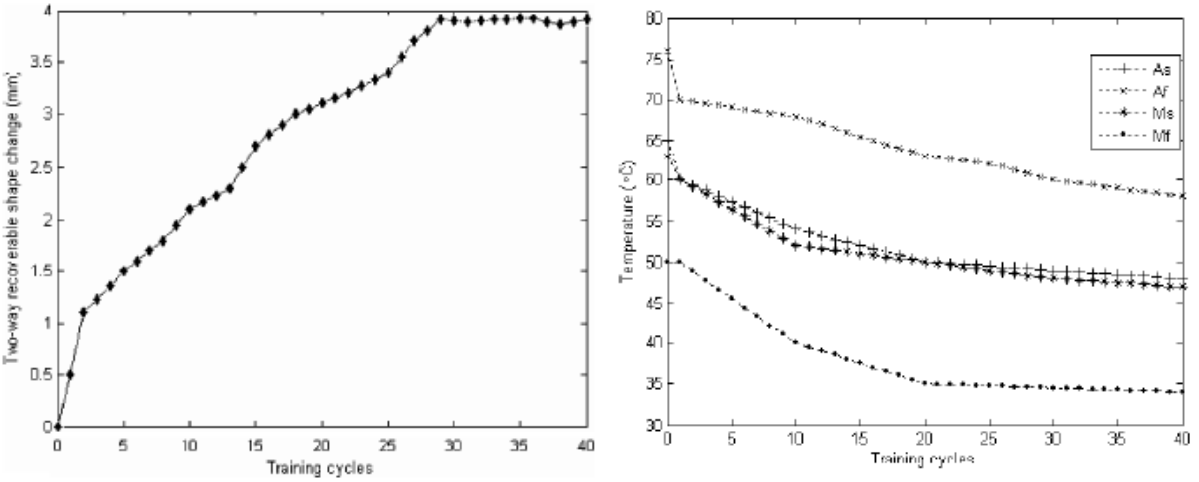


Figure 31 - Two-way training is completed in approximately 30 cycles, the transition temperature is significantly decreased by the cold working [66]

Fatigue

Fatigue life of super-elastic SMA has been studied in detail previously [68]. The fatigue during thermal cycling, termed 'functional fatigue' in the SMA community, has also been studied to a lesser extent [68, 69]. It has already been stated that SMAs do not achieve repeatable behavior until after actuating at least 30 cycles. Unfortunately, as with all metals SMAs also fatigue after repeated use. New research shows some alloys (e.g. $Ti_{54}Ni_{34}Cu_{12}$) are significantly more durable than current commercial alloys [70], but for now the engineer must accept the fatigue limit of common NiTi SMAs and design accordingly. As with all fatigue processes, the stress, strain, and temperature variation amplitudes have a large effect on the number of cycles before failure. A thorough review of fatigue mechanisms was made by Robertson et al using both their own data and a collection of past researchers data [71]. The focus was not on actuators with electrical heating and stress cycling, but on loading of medical devices within the body. They showed that survival to 10e7 cycles was only possible with strains below approximately 0.5%. Tobushi showed that 'fatigue training' NiTi SMA actuators greatly improved their fatigue properties [72]. 'Fatigue training' refers to simply cycling several times at near maximum strain (~6%) before operation at low strains (~1%-2%) [73]. That is, initial training at high strains greatly increases the number of cycles at subsequent low strains that can be sustained before failure. It has been shown there is large benefit to using only partial transformations in addition to decreased cycle strain [74, 75]. Because surface imperfections initiate failure in most cases,

protection during heat treatment can greatly affect the fatigue life – vacuum heat treatment is generally much better than treatment in air [76]. Also, polished surface treatment of samples can create a ten-fold difference in lifetime [77]. Many of the fatigue tests are performed on dogbane samples, and results may not be comparable to fine wire results due to surface effects – a commercial NiTi wire producer suggests that finer wires have superior lifetimes compared to thick wires and strips [78].

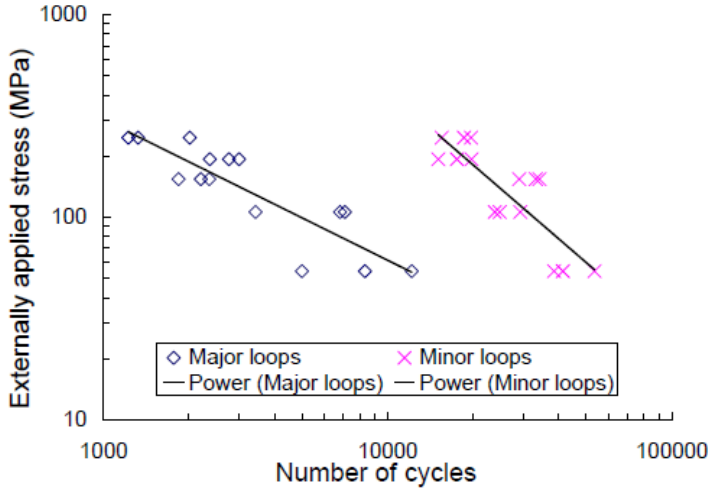


Figure 32 - Lagoudas shows the fatigue life varies with applied strain and transformation range [74]

Bending and Curving

There are many applications where it is necessary to warp or curve a NiTi wire around a curved surface. Even if the bending radius does not change, this introduces an additional fatigue consideration since the strain in the outer edge of the wire experiences increased stress which oscillates in magnitude every actuation cycle. Cyclic bending fatigue tests show that strain of greater than 1% in bending leads to premature failure of NiTi wire in bending [79, 80]. To prevent this mode of failure, the bending radius should be greater than 50 times the wire diameter (which keeps the bending strain below 1%) [81].

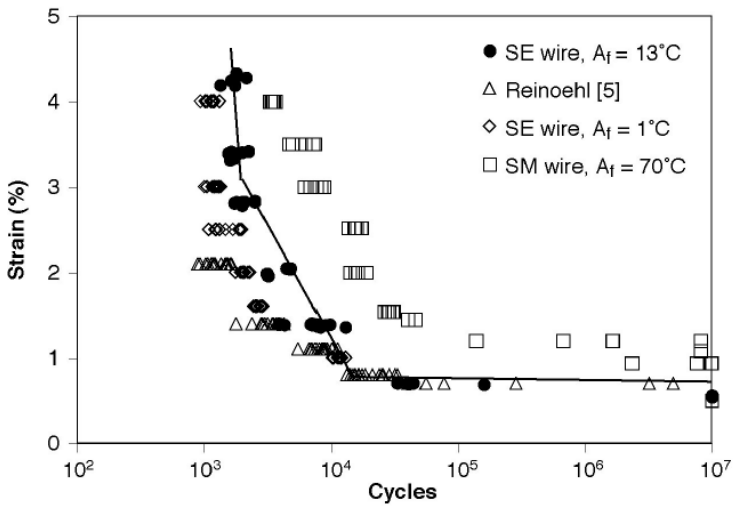


Figure 33 - Bending strain greater than 1% greatly reduces super-elastics NiTi lifetime [79]

Surface Treatment

It is essential to NiTi actuator success that the wire is electrically conductive. However, it would be desirable in many cases to have an insulating coating on the wire in actuators to prevent accidental electrical shorting. An ideal insulating coating would be very thin, flexible enough to match the strain of the NiTi wire (~8%), and would not have a high thermal resistance. Anodizing of common titanium alloys develops an insulating coating with a thickness proportional to the voltage applied during anodizing. Subsequently, the anodized surface is insulated against all lower voltages (unless surface defects like cracks form). Unfortunately, the Nickel in NiTi prevents direct oxidation from producing an insulating layer (from the author's experiments and [82]). High voltage anodizing in Sodium Acetate buffer was effective in growing oxides on the NiTi surface, but did not remove the Nickel and simply formed Nickel oxides alongside the Titanium oxides [83]. High temperatures can be used to form controlled oxides - however they are generally brittle, and raising a formed NiTi actuator above 300C would destroy the shape training so this method is ineffective in practice [84]. In the medical community, it is a common concern that toxic Nickel in the NiTi will leach into medical implants after insertion, as well as to prevent corrosion. This has led them to seek to isolate Titanium and its oxides (and to strip the Nickel) from the NiTi surface [82]. This same process can be an effective tool for the electrical insulation - isolate Titanium in the surface, and anodize to create an insulating barrier. A simple low-temperature solution is to chemically etch and

simply boil the wire in water, providing a 4-fold corrosion resistance [85]. However, a more effective method is chemical etching and boiling in Hydrogen Peroxide - the hydrogen peroxide etching releases Ni from the NiTi at the surface, and simultaneously oxidized the remain titanium

$$\mathrm{Ti(s)} + \mathrm{O_2(g)} \rightarrow \mathrm{TiO_2(s)}$$

$$\mathrm{Ni(s)} + \frac{1}{2}\mathrm{O_2(g)} \rightarrow \mathrm{NiO(s)}$$

This is particularly successful since the resulting oxidation forms a continuous gradient from surface inward, rather than a discrete layer between oxidized scale and interior NiTi [86]. Several sources suggest that long-term aging at high temperature (near 400℃) can lead to predominantly titanium oxides in the surface increasing both corrosion [65] and electrical [87] resistance. SMA wire can also be coated in plastics such as PEEK [88] and PTFE.

Conclusion

A brief review of training and treatment techniques for SMA materials was presented. The heat treating process is accessible to most engineers and is able to greatly affect the SMA material properties. In general, straight wire should be purchased in the desired condition to ensure uniformity. If a shape other than the original straight wire is needed, heat treatment near 490℃ will not change the shape memory performance, but will set the new shape precisely. Many studies on fatigue of SMA actuators show that strains in the range of 4% at stress near 100 MPa is sustainable for 1e5 cycles or more. Lower stress and strain leads directly to longer life, even for functional fatigue caused by shape memory actuation. Surface treatments in practice do not have a large effect on the wire, though coatings of various polymers can insulate the material from short-circuiting.

Chapter 5

Review of SMA Actuators in Robotic Applications

Shape Memory Alloys (SMA) wires that move around when heated and cooled make good party tricks with little effort. Creating truly useful actuators is much more difficult. Typically, electric actuators are self-contained devices that connect *onto* robotic mechanisms. To fully realize the benefits of SMA strength, cost, and simplicity, the SMA actuator must be designed *into* the robotic mechanism. The designer must match the strength, speed, range of motion, and cost to their application. This review covers the mechanical design of robotic actuators with built-in SMA actuation. It is shown how these techniques are applied to general robotics, and specifically focuses on implementation for anthropomorphic robotic hands. Finally, a review of control methods from previous SMA research shows the state of the art in speed and accuracy of SMA actuators.

Introduction

What is it about robots that makes them so robotic? Concepts like dexterity and bio-fidelity are easy to see, hard to define, and even harder to achieve. Their absence in modern robots is apparent at a glance. A clear obstacle to creating robot dexterity is in limits to their versatile poses – the degrees of freedom in their motion. While computer science is quickly finding ways to control high DoF robots (deep learning, for instance), **the hardware is not keeping pace**. Computer graphics in film can reconstruct 244 degrees of freedom in the human body and convince millions of movie-goers that they have seen a living being [89]. The most pressing obstacle to dexterity in robots is physical degrees of freedom.

Today, **actuator design is the limiting factor** – motors, pneumatics, and hydraulics are simply too heavy and too big to package hundreds of DoF in a single body. This is particularly apparent at small scales – common ants have over 20 useful degrees of freedom, but there is no mechanical motor today that can match an ant's leg in size and strength. For commercial applications to become viable, the actuators must be extremely low cost because many more will be needed in each robot than traditional systems. The ideal actuator must provide extremely high

strength-per-weight, while allowing many to fit in a small space (DoF/m^2), at low cost ($/DOF).

Shape Memory Alloys (SMA) are special alloyed metals that undergo a reversible change in crystal geometry when heated and cooled, causing the alloy to expand and contract with large force. This motion can be controlled by heating using electric current and cooling in ambient air. SMA actuators – in many cases simple fine wires - can provide huge benefits in cost reduction, miniaturization, and other aspects of robotic devices. Especially in high degree-of-freedom and miniature robotic systems, SMA exceeds all alternate choices.

This review will study the application of shape memory alloy actuators as artificial muscles for humanoid robots. Previous reviews have focused on other applications of SMA such as automotive [90], aerospace [91], medical [92], or comparison of SMA against other 'smart' materials [93]. However, no review has focused specifically on the application to humanoid robots. Humanoid applications require high dexterity and strength; the combination of refined resolution and superior strength could potentially be an excellent place to apply SMA artificial muscles. This review will determine their feasibility. We study in particular the use of SMAs for anthropomorphic hands, which provide an excellent example where many degrees of freedom are needed in a tight space, with high strength.

Types of Smart Materials

A smart material in the present context is a *material* which provides coupling between different realms of physics: This stands apart from engines and generators which are *machines* that do the same. 'Smart' implies intelligence, though in this context it means something closer to 'unexpectedly/highly reactive' more than anything 'cognitive' or 'understanding'. Typically, these materials strongly couple mechanical motion with one of electricity, heat, light, or magnetism. It is curious that we say only materials which have a mechanical motion are 'smart'; for instance, simple iron wire couple electricity and magnetism but that coupling is not considered 'smart'. Tzou et al group smart materials that they consider to be common into a dozen groups [94]. A comprehensive review of SMA and comparison between different actuators (both smart and conventional) was made in a review by Jani et al. in 2014 [95]. Chopra also reviewed smart actuators of this nature and reported on their various strengths and weaknesses [96].

Piezo- materials are typically crystalline, exhibit very low actuation strains, though with very high bandwidths. They create large forces during this very low strain. For the current work, the low strain of piezo-electric materials is not acceptable, though their speed and linear response character would be desirable [96].

Shape-memory alloys provide a much higher strain capability (near 10% strain), they create high forces though they are relatively slow and exhibit large hysteresis [96]. In general, these materials react to a change in temperature by changing their crystal structure. The geometric difference between the low temperature and high temperature structure creates a meaningful strain (or stress if constrained) that is large compared to ordinary thermal expansion [51] [97]. The two crystal states are Martensite (when cold) and Austenite (when hot). Upon cooling, the material becomes twinned Martensite, and requires applied stress to return to a fully extended length. The most prevalent shape-memory alloy for the past half-century has been Nitinol, a Nickel Titanium alloy developed by William Buehler in the Naval Ordnance Lab and patented in 1965 (US3174851) [98].

Shape Memory Polymers compare to SMAs with characteristics of increased strain capabilities but much lower stress [95]. Shape Memory Polymers are far less expensive (by mass) than SMAs, their activation is typically based on glass transition and they are much more efficient than the shape memory alloys [95]. Electroactive polymers are grouped into two major classifications – ionic (based on diffusion of ions in the material) and electronic (driven by Maxwell forces such as in a capacitor) [99]. The ionic muscles are not very high stress and require an electrolyte medium for actuation. The electronic types are very efficient, very fast, and exhibit high strain – but require very high voltages to be activated. They also have a benefit of holding steady position under DC current.

Inexpensive coil-based actuators from fishing line were developed by the NanoTech Institute, University of Texas at Dallas [100]. By putting large amounts of twist into plastic lines such as Nylon 6, an inherent coiling winds the line into a spring with major diameter to minor diameter ratios near 1.1. By heating the (aligned) polymers in the cable, the contraction of the line internally leads to untwisting motion – which in turn creates a reaction force in the coil causing a linear force. These actuators were found to be comparable to nitinol – they have faster bandwidth, very small hysteresis, and are far cheaper than shape memory alloys. Their volume specific work output is smaller, but the mass-specific output is on the same order as nitinol. The

overall efficiency is not very impressive (like Nitinol). To actuate these muscles using electricity they must be clad with a resistive and conductive layer (source used silver) or otherwise exposed to direct heating or cooling.

Previous reviews have covered the fundamentals of actuator selection without explicit details of any actuation method [101] [93] [102]. Other reviews have focused on applications of SMA [91] [103] [104] [105] [106], specific areas of application such as automotive [90], aerospace [107], medical [92], small scale [108], mechanism design [109, 110], comparison between SMAs and Shape Memory Polymers (SMPs) [111], and other 'smart' materials [93]. The 'holy grail' of shape memory actuation is to create artificial muscle which is equivalent or superior to its biological namesake. The concept of biologically inspired actuation is this; if an artificial muscle existed which had the weight, strength, stroke, speed, accuracy, and efficiency of a biological muscle, then robots could be made based on natural creatures in which we observe accurate and fluid motion with abilities no modern robot currently can achieve. In the quest for such motion, materials are sought which create linear motion from energy input (heat, light, etc.). Each material has some beneficial features and some obstacles. Shape memory alloys, biological muscle, and some other technologies are compared in Table 4. 'Specific work' provides a useful metric for the energy of the system because mechanical advantage can exchange strength for stroke and vice versa. Specific work is best normalized against weight, volume, or cost as appropriate for the application at hand. Shape memory alloys have the highest power density of all known actuators [112].

Table 4 – The characteristics of SMA and Biological Muscle. Most smart materials have some beneficial feature – else they would not be studied. The best technology for each characteristic is also listed.

Characteristic	**Units**	**NiTi**	**Muscle** [102]	**Best Technology**
Weight (Density)	kg/m^3	6450	1060	Polymeric
Strength	MPa	200	0.35	SMA
Stroke	%	10	20	SMA, Polymeric
Mass-specific work	kJ/kg	1.178	0.066	SMA
Volume-specific work	kJ/m^3	7600	70	SMA
Speed (Bandwidth)	Hz	4	20+	Piezo
Mass-specific power	W/kg	4713	1320	Piezo
Volume-specific power	kW/m^3	30400	1400	Piezo
Efficiency (Eout/Ein)	%	3	30	Electromagnetic
Cost	$/g	$3.40	~$0.25	Polymeric

SMA Design Considerations

There are several trade-offs in SMA mechanism design – the primary trade-off will be between complexity, speed, and strength (Figure 34). The strength is determined by the total cross-sectional area being stressed. Speed is determined by the total surface area. Cost is driven by the mechanism complexity (the pulleys and routing components required to implement the actuator), and by the total length of wire required. Single SMA wires provide a direct trade-off between speed and strength. By arranging multiple wires in parallel, we can get both strength and speed, but at higher cost. Training the SMA into shapes besides wires can provide other benefits. Solid rods and torsion tubes are very strong, but expensive and slow. Forming the SMA wire into coil springs or zig-zag 'snakes' is inexpensive and increases the effective strain (which is useful), but usually decreases total strength.

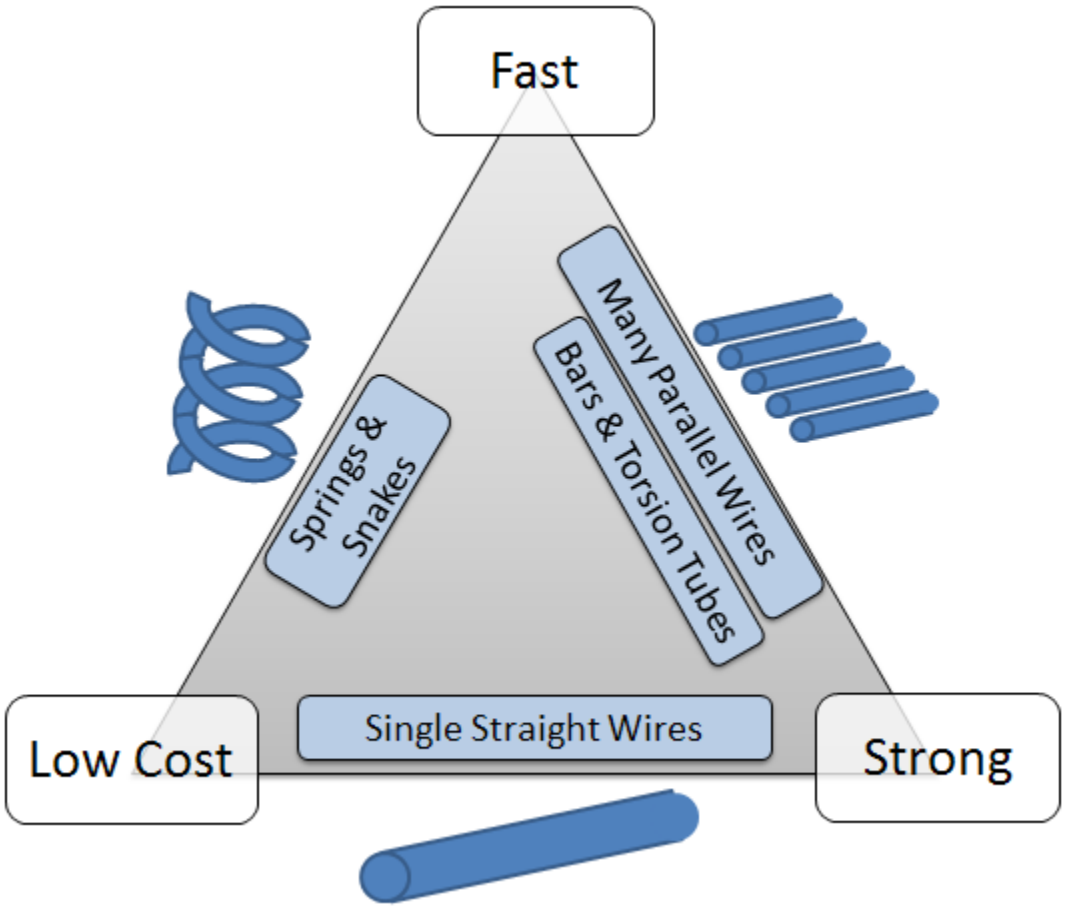


Figure 34 - The primary trade-off in SMA mechanism design is between speed, strength, and complexity.

Mechanisms

The trade-offs between speed, strength, and cost for the wire alone are clear since they merely require knowledge of the size, surface area, and length of wire used. However, none of this addresses the primary reason for designing an SMA mechanism: increasing strain. SMA wires have a typical actuating strain of 4% - for reference, human muscles provide strains on the order of 40% to 100%. The approaches to solve this problem fall into two groups based on the needs of linear or rotary actuators. A previous review identified several stand-alone actuator mechanisms [113] – this review focuses instead on implanting SMA into the mechanical design.

Linear Actuators

Linear actuators practically are constrained to changing the form of the wire into something that amplifies the small strain into linear motion. Common examples are springs [59, 114] and snakes [115]. Clever arrangements of springs can even provide powered extension and contraction (typical SMA actuators only contract strongly) [116]. More complex shapes can be 3D printed [117] or cut using lasers or wire EDM [115]. Some clever mechanisms can amplify the strain by creating tensegrity shapes [118] [119]. Redundant pulley systems allow longer wire length to be fit into a smallspace – the free end provides more stroke than a single length [120, 121]. A special antagonist system called a Bowden or ‘push-pull’ cable allows the SMA wire to be laid down out of the way of the mechanism [5, 122].

Figure 35 - Coil-spring greatly increases strain [59] [114] (device #1)

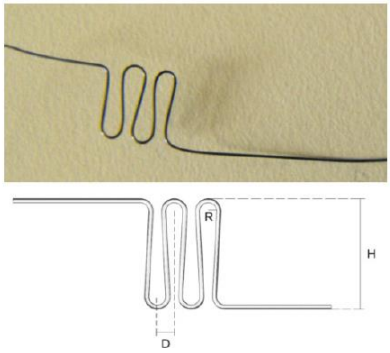


Figure 36 - Snake-like shape in wire or laser-cut sheet [115](device #2)

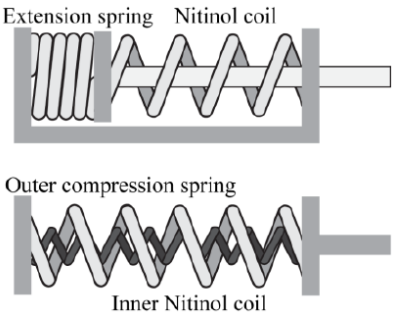


Figure 37 - Parallel, passive antagonist spring [116](device #3)

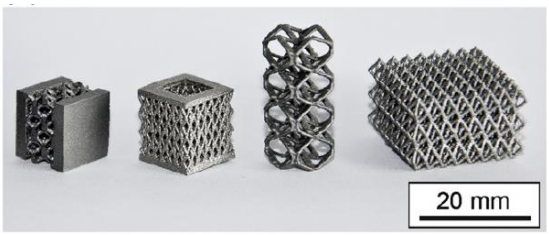


Figure 38 - Fabrication using Selective Laser Melting [117](device #4)

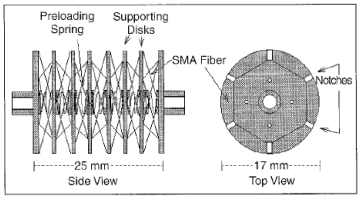


Fig. 1. Shape memory alloy actuator.

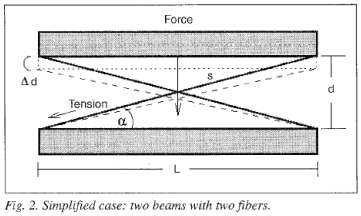


Fig. 2. Simplified case: two beams with two fibers.

Figure 39 - Tensegrity Strain-amplification [118] [119](device #5)

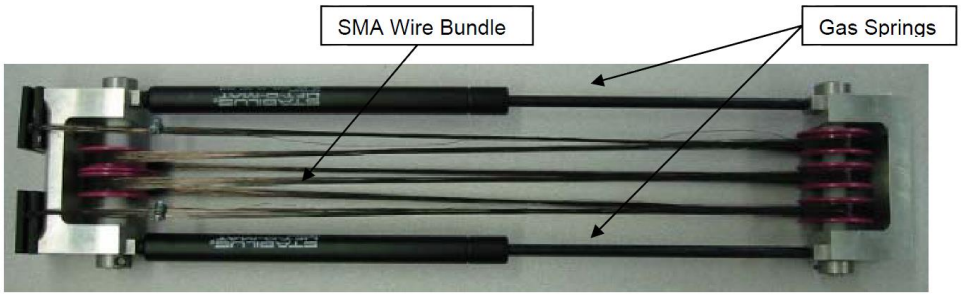


Figure 40 - Redundant pulley system – longer wire length simply provives more stroke [120] (device #6)

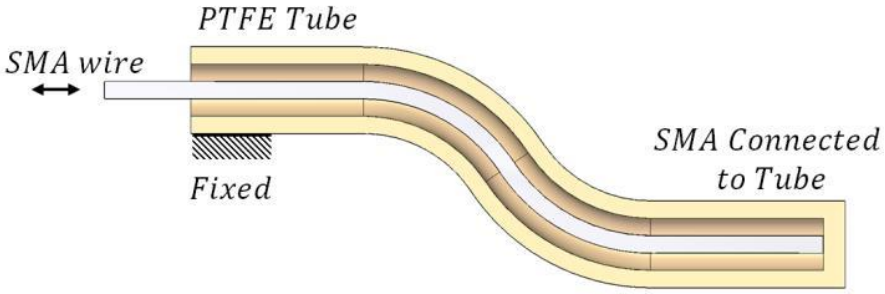


Figure 41 - Bowden tube developed by the author [5]. Asimilar technique was used in [122]. (device #7)

Rotary Actuators

As shown above, there are limited techniques for increasing the linear strain of SMA mechanisms. Fortunately, most robotic actuators rotate *hinged* joints - so actual linear straining is not the end goal (think of fingers or elbows). Rotary actuators have many more options for strain amplification. By connecting straight SMA wires close to the hinge, the angle of rotation can be very large even though the total strain of the wire is small [123]. This can also be done with springs for an even larger stroke [124, 125, 126]. Alternately, a long length of wire can be would around a spool or pulleys to increase the total length being heated and cooled [127].

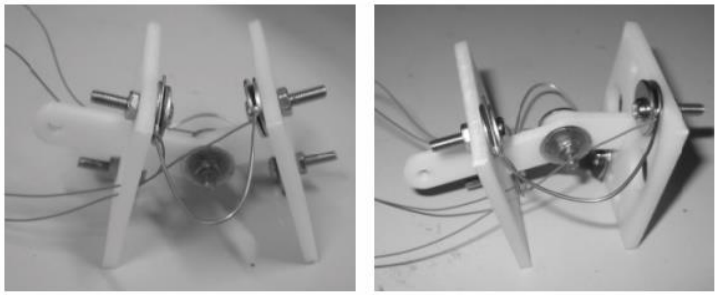


Figure 42 - Convert wire bending and straightening to rotation [128](device #8)

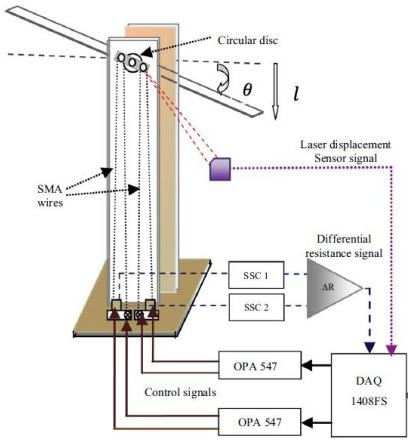


Figure 43 - Example of the very common three-bar arrangement [123], see also [129](device #9)

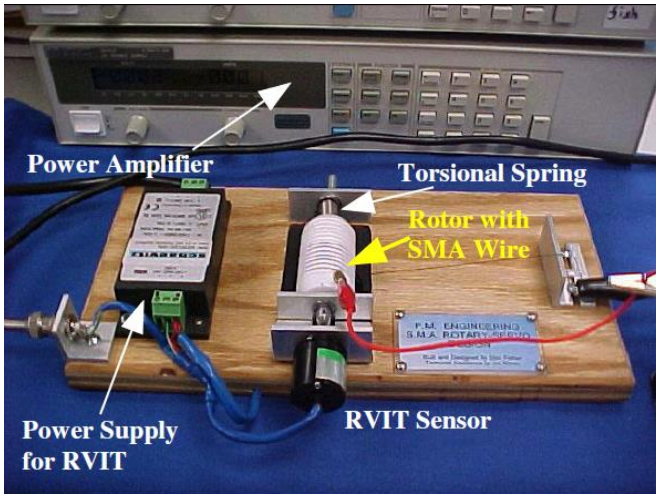


Figure 44 - Wrapping a long length of SMA wire around a single pulley [127](device #10)

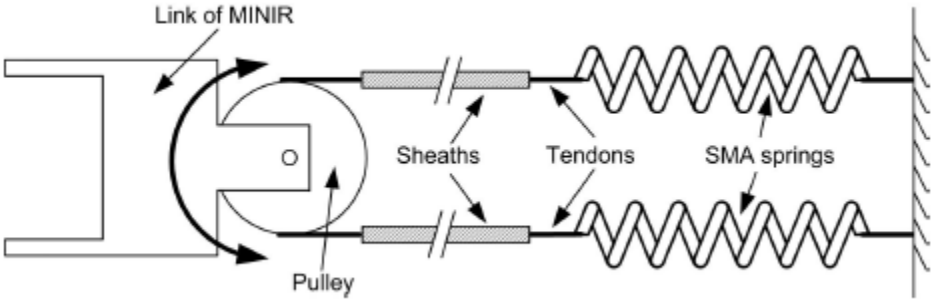


Figure 45 - Springs acting as antagonists on a single joint [125](device #11)

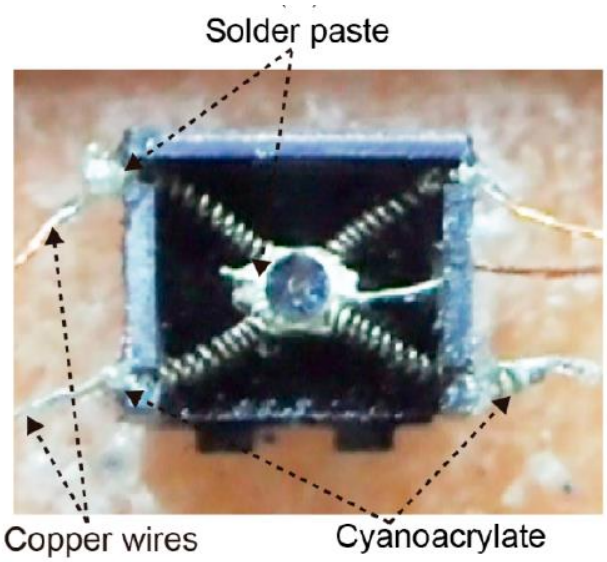


Figure 46 - Four springs arranged around a central rotating spindle [124](device #12)

Flexing and Bending

Some actuators are designed to bend. These are not usually useful for robotic machines except in soft robotics where they can completely replace the need for small legs, fins, or wings [130, 131]. Using extremely fine wires can allow speeds up to 35 Hz [132].

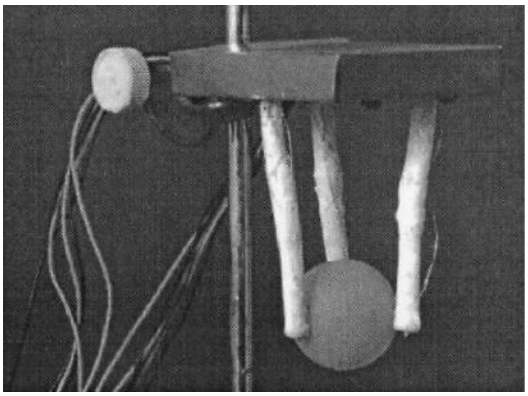


Figure 47 - Simple flexures used as fingers [130](device #13)

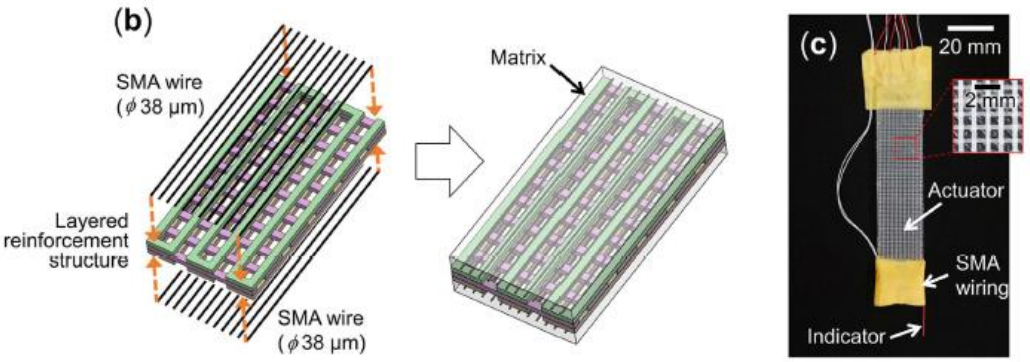


Figure 48 - 35 Hz bending actuator with 38 um (0.0015 inch) wires in a thin sheet [132](device #14)

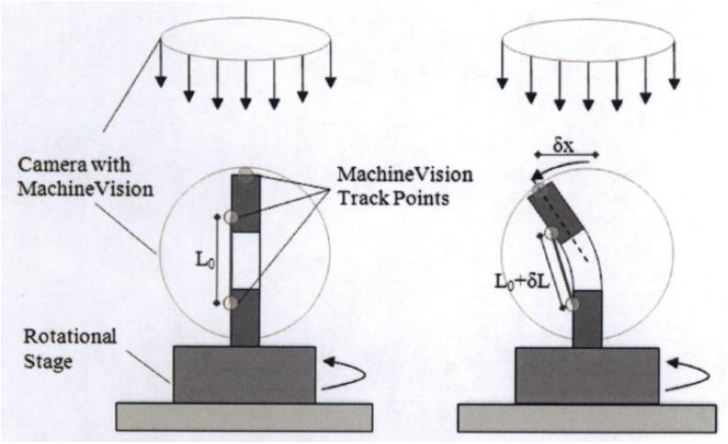


Figure 49 - Flexure with axial SMA wires, used to control a nozzle [131]. Similar to [133]. (device #15)

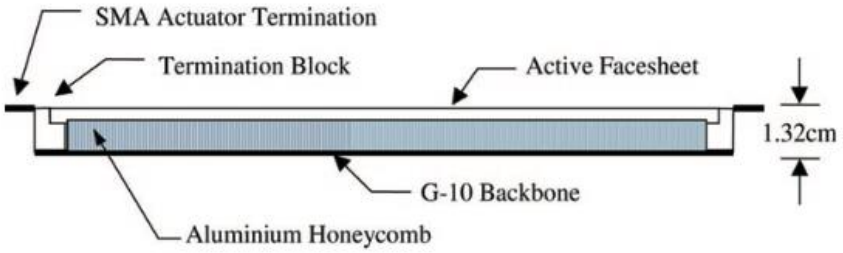


Figure 50 - Using SMA as the skin of a honeycomb composite structure [134](device #16)

Comparison

As a final comparison, the techniques of each reference are plotted in Figure 51 on the original trade-offs diagram to show how they fit the constraints of complexity, strength, and speed. Each device listed above was considered for the published speed, cost, stroke, and strength reported by the creator. The compromise of these four characteristics is compared in Figure 51. Each device is placed to show the relative speed, cost, and strength, and the marker for each is colored by the device' ability to get the high strain needed for practical actuators.

Devices near the center provide a fair trade-off of all characteristics. One of the biggest advantages of SMA over other technologies is the simplicity. Based on these performance trade-offs presented here, the best results for robotics will be to use straight SMA wires (cheaper and simpler than coil springs), and find a way to route the required length of wire through the robot so that sufficient strain is achieved. The following section describes ways that these different methods were applied in actual design situations.

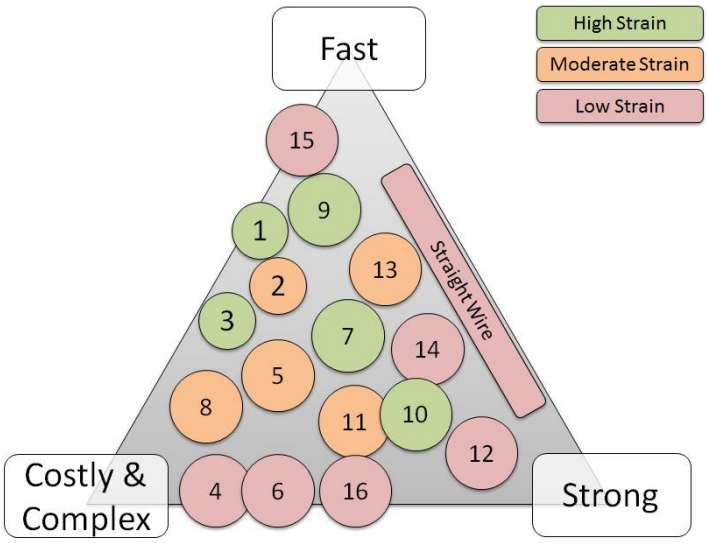


Figure 51 - The references in this report are compared for strength, speed, complexity, and ability to amplify strain

Complete Robotic Humanoid Hands

The techniques reviewed above have been implemented in many complete robotic systems. Though none of them have become commercial products with large success, they all show the great potential for SMA, and continue to solve the problems of implementing SMA tightly into the robot design. Prosthetic or robotic humanoid hands provide the ultimate test of actuators. A human hand has 24 discrete degrees of freedom all fit in a very tight space. Creating a robotic hand with electric motors usually requires significant gear reductions to get sufficient strength – this leads to slow and loud actuation.

Many researchers focus on design of single finger – this is already a complex system with four distinct degrees-of-freedom [135]. In most cases the finger is 3D printed, and straight wires are routed like tendons along each side. Generally these designs fail to solve the critical issue – where will the length of wire be beyond the finger?

Another hand was made with five fingers, this one was controlled by selectively heating the SMA wires in a grid that allowed precise control of the stroke [136]. The complete length of wire needed to curl the finger around into a fist is present – the wire rides in a cradle in the wrist area. This robotic hand provides a huge range of control (24 degrees of freedom in total), and could create all the gestures of American sign-language. Unfortunately, the research seems to have stalled – this hand is not practical because the big grid of SMA wires is not compact and integrated into the robot's wrist.

One solution was found by using SMA springs instead of straight wires [137]. By using SMA springs instead of straight wires, the actuators can be fit entirely in the palm – a good choice for prosthetics where only a hand needs to be replaced. Unfortunately, the springs are too large, and only eight total actuators could be fit into the palm.

Recently, a research group created a simple three-finger prosthetic hand using straight SMA wires [138]. In each joint of each finger, six wires are placed in parallel to increase the stress. The wires extend down through the wrist and into a base, providing a long length of wire and more strain. Each finger can be actuated at 1 Hz – slow compared to humans but fast enough to make a practical prosthetic. This hand was successful in showing the strength and speed of SMA is adequate. However, the design does not provide means for accurate control (no sensors) so some work remains to be done.

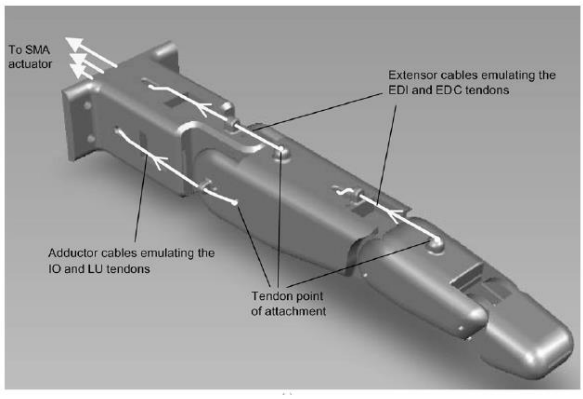


Figure 52 - A single finger with several degrees of freedom [135, 139]

Figure 53 - Prosthetic hand with clever open-loop control system [136] [140]

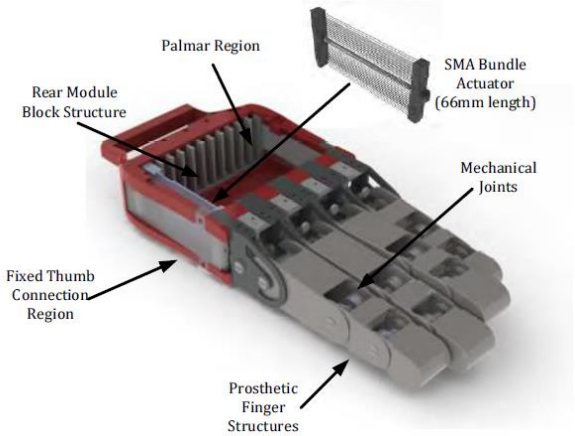


Figure 54 - Using many SMA springs in parallel to get high stroke and high strength [137]

Figure 55 - Prosthetic hand [138]

Control Methods

There are always many ways to solve control problems, especially to nominally difficult problems such as SMA control, and researchers test many complex techniques making small gains in performance. The ideal control schema for SMA is self-sensing, robust, and simple enough that it can be repeatedly implemented. Control methods that rely on external sensing are included here, as well as literature focusing on sensing ability alone and of course controls utilizing self-sensing. The schemes are compared for tracking ability, ability to be implemented using self-sensing, and to focus only on necessary complications. Elahinia reviewed SMA controls in 2010 [141]. He concluded several key concepts:

1. Bandwidth is limited by passive cooling
2. Rapid heating to near transformation temperature should always be included
3. Adaptive control laws help maintain stability to disturbances and un-modeled parameters
4. A biasing force is required for controlled, two-way motion
5. Combining position, force, and temperature control helps reduce sensitivity to disturbances and un-modeled parameters

Another review by Sreekumar et al covered nonlinear control methods with no strong conclusions about which models were most capable, but suggesting that resistance feedback would soon be perfected for micro-actuator applications (though this is known to be

unobservable for strain control, see Chapter 7) [142]. Fuzzy control, adaptive control, and techniques that account for actuator hysteresis were found to be most effective.

It was found experimentally that the open-loop phase response of NiTi wires (force output / power input) is shifted proportional to the wire diameter and independent of wire stress and strain (similar to first order low-pass filter if cutoff frequency is proportional to diameter) [143]. The magnitude response is affected by stress and strain equally at all frequencies. The conclusion is made that decreasing wire diameter farther will continue to increase bandwidth, in agreement with others, as the speed energy is transferred away from the wire is the limiting factor [143] [144]. Some of the many control techniques applied to SMA are compared in Table 5, along with their accuracy and speed if reported:

Table 5 - Comparison of control schemes for SMA

Category	**Method**	**Mechanism**	**Sensing**	**Demonstrated Positioning (Force) Accuracy**	**Demonstrated Bandwidth**
Feedback Control					
	Modified Prandtl-Ishlinskii FF + PI [145]	NiTi Springs, rotary joint	Angular Potentiometer	19.87% PI, 8.65% FF+PI, F rejection	0.5 Hz P tracking
	anti-slack, rapid-heating, anti-overload, differential PID [146] [147]	antagonist pair	Angular encoder, load cells	(0.017%), F rejection	2 Hz F tracking, <1Hz P tracking
	Generalized Prandtl-Ishlinskii FF [148]	passive antagonist beam	Linear potentiometer	not specified	.02 Hz
	Fermi-Dirac Statistics Model, Robust H_infinity loop shaping [149]			2.78%, F rejection	N/A
	Nonlinear PID Temperature-feedback [150], heat transfer FF	antagonist pair	temperature	0.14 degrees C	.2 Hz
	'Relay' control. Force feedback [151]	antagonist pair	load cell	2%, but has limit cycles	2 Hz
	VSC sliding mode [152]	antagonist spring	rotary pot		
	Temperature feedback [153]				

	Neural Network Feedforward [154]	antagonist spring	LVDT	15%	0.1 Hz
	Sliding Mode [127]	antagonist spring	Rotary potentiometer	0.2%	1/60 Hz
	Feedback Linearization polynomial model [155]	beam bending spring	Strain gage(?)		1/4 Hz
	Neural Network Feedforward [156]	beam bending	strain gage		
	Nonlinear stress-model based control VSC [157]	antagonist spring and mass	potentiometer		~1/2 Hz
	Self-tuning fuzzy PID [158]	antagonist mass	potentiometer		0.033 Hz
	FF Neural Network and Sliding Mode [159]	bias spring	LVDT	0.8%	1/30 Hz
	Sliding Mode [160]	bias spring	laser displacement	1.3% RMS	1/15 Hz
	PID-P^3 [161]	beam spring			
Sensing					
	Resistance modelling [30] [162] [163] [164] [165] [166] [167] [168] [169] [61]	variable load	resistance	N/A	N/A
	Inductance modelling and circuit for measuring [170]	static strain	inductance	N/A	N/A
Self-Sensing Control					
	Polynomial strain-resistance model, fuzzy-tuned PID [171]	Preloaded Wire	Resistance	3% RMS, load independent, F independent	0.2 Hz P tracking
	Polynomial strain-resistance model, FF hyst. comp., PID [172]	Preloaded antagonist pair	Resistance	1.093%, 1.628% PID only, F rejection	0.15 Hz P tracking
	Fuzzy tuned PID [173]	Antagonist pair	Differential Resistance	0.4%	1.0 Hz Tracking
	assumed linear R [130]	beam flexure	Resistance	N/A	N/A
	SMC and PID [123]	antagonist pair	Differential Resistance		
	Neural Network resistance feedback [174]	antagonist spring	resistance	7% Steady state	1/30 Hz

Concepts					
	Resistance priming used to quickly heat from ambient temp [175]				
	Rapid Heating technique using circuits and control laws [147] [176]				
	Nonlinear state estimator (not actually a Kalman filter as they claim) [177] [178]				
	L2 Stability Analysis, PI control [41]				
	Dissipativity Stability study – model is probably too simple for this to be meaningful [179]				
	Preisach hysteresis model system identification [180]				
	Relating Power consumption to airspeed estimation [61]				
	Strain-rate dependent properties can affect control at very high speed (1000/s strain) [181]				
	Uncertainty quantification for better modelling and estimation [182]				

Conclusion

While highly complex humanoid robotic systems are still out of reach, technology such as shape memory alloys are bringing the world closer to realizing lifelike actuation. This report reviewed the success over the last 30 years of work to bring SMA actuators closer to this goal. Electromagnetic motors have become such an entrenched method for robotic motion, it is hard to push designers and dreamers away from the standard tool. We must realize that electric motors are hitting fundamental limits of capability – they can never be as strong (for their size) and accurate as human muscles, and are so complex that their cost cannot be decreased much further. Shape memory alloys may not be the final solution, but with benefits of high strength, simplicity, and low cost, these 'artificial muscles' have a chance to redefine the future of robotics.

Chapter 6

Energy Balance and Electrical Heating

While in practice it is nice to think of SMA actuators as electro-mechanical devices, the shape memory effect is entirely driven by temperature change. Electric current is used to heat the wire, and ambient air or conductive heat sinks are used to cool it, but the crystal phase change is caused directly by change in temperature. Passing electric current through the material heats it by Ohmic or 'Joule' heating which converts directly to thermal energy in the material and mechanical work. Circuits and techniques for driving the current are discussed. The temperature of the material is decreased only by dissipating to the environment (primarily by convection). The heat energy is unfortunately all dissipated. Most SMA actuators are made from fine wires and are cooled in open air; this heat transfer scenario is considered in detail here. Calculations and experiments confirm that heat transfer from round wires is a limiting factor to SMA actuation rate, and is also an important consideration in control and mechanical design.

Energy Equations

Consider a nitinol wire as a single thermal mass. Energy can enter, be stored in, and leave the system a few ways as shown in Figure 56.

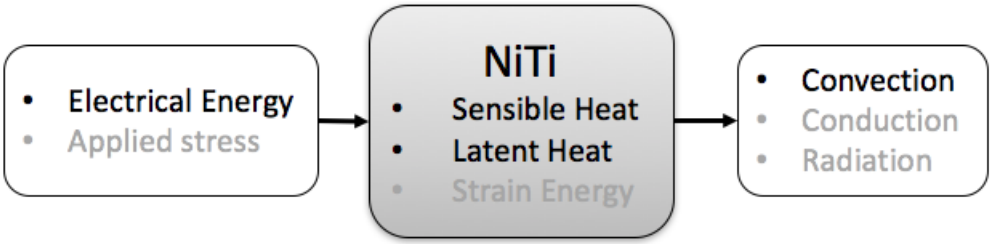


Figure 56 - Energy balance for a 'lumped mass' of shape memory material. The material receives energy from electrical current passed through it, stores most energy internally as sensible and latent heat, and dissipates energy mostly via convection

The primary source of energy is from the electrical power, while most energy is stored as sensible and latent heat, and dissipated by convection [42]. Other sources and sinks are relatively small, even the strain energy which we are using as the actuator – the efficiency from electrical heating to mechanical work is on the order of 4% [183]. The balance of power in the system results in a change in temperature, summarized as follows:

$$mc_p\dot{T} - m\Delta H\dot{\xi} = P + AL\sigma\epsilon_L\dot{\xi} - hA_s(T - T_\infty)$$

(27)

Which is equivalent to:

$$\left(mc_p - m\Delta H\frac{\partial\xi}{\partial T}\right)\dot{T} = P + AL\sigma\epsilon_L\dot{\xi} + m\Delta H\frac{\partial\xi}{\partial\sigma}\dot{\sigma} - hA_s(T - T_\infty)$$

(28)

Where P is electrical power, m is the material's mass, c_p is the specific heat, ΔH is the latent heat of transformation, h is the coefficient of convective heat transfer, A_s is the surface area, and T_∞ is the ambient air temperature (and other terms have been previously defined). Consider the terms in this equation individually. First, sensible heat is stored in the metal as with all materials:

$$mc_p\dot{T}$$

Second, the material can also store energy in the crystal as latent heat:

$$-m\Delta H\dot{\xi} = -m\Delta H\left(\frac{\partial\xi}{\partial T}\dot{T} + \frac{\partial\xi}{\partial\sigma}\dot{\sigma}\right)$$

This is similar to ice melting and freezing – there is energy stored in phase transformation. The negative sign is required because transformation from Austenite is exothermic (it releases the energy increasing the temperature of the material or environment), while the Martensite to Austenite transformation is endothermic (it draws heat from the environment or decreases the material temperature). The input power P is typically input by passing electric current through the material. This is discussed at length in a following section. Generally, this can only be a positive input except with a special device such as a thermo-electric generator. Heat is dissipated to the environment mostly by convection:

$$-hA_s(T - T_\infty)$$

This is most true for SMA wires in air, although in some cases conduction or radiation would be the dominant heat transfer mechanism. Finally, the pseudo-elastic (seemingly 'plastic') deformation during the super-elastic motion is also coupled in the energy balance:

$$AL\sigma\epsilon_L\dot{\xi}$$

This term is simplified by assuming the only plastic-like deformation occurs due to phase change. This effect is very similar to the 'pressure ($A\sigma$) times volume change ($L\epsilon_L\dot{\xi}$) studied in elementary thermodynamic analysis.

The effects of the latent heat can be clearly seen when heating or cooling a normal mass (with sensible heat alone) compared to a mass with both sensible and latent heat during a phase change (Figure 57). In this simulation, a common material is heated with a constant power source until reaching steady state (blue). The response is compared to a material which changes phase during the heating period (red). The shape memory effect – diffusionless crystal phase change with small change in temperature – propagates through the material and causes contraction or expansion at the speed of sounds once the transformation is initiated. The practical limit to actuation speed is only restricted by the speed at which the temperature can be changed. The topics of this chapter are: (1) determining how to generate the input power P from an electrical power source, (2) computing the total energy stored in the material and needed to raise it to a certain temperature, and (3) determining the rate of heat loss to convection – which is the driving factor in SMA actuator bandwidth limitations.

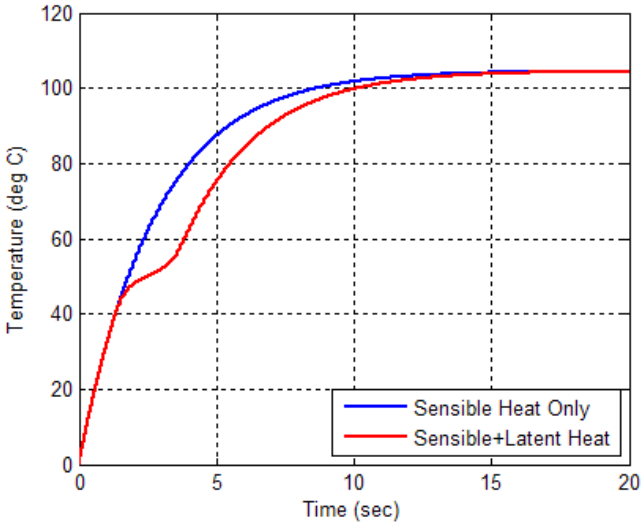


Figure 57 - The energy stored as latent heat is clearly seen when comparing the step response of a standard mass (blue) to a mass with phase transformation (red). In this simulation, phase change occurs from 40 to 60 ℃.

When the material is in the super-elastic state, the interaction between stress, strain and temperature provides some exciting effects. Consider a simulation of perfectly insulated super-elastic SMA wire which is being stretched and released (Figure 58). The wire temperature at first increases as the exothermic A to M transformation begins – the temperature rises as energy is stored as sensible heat in the material since it can no longer be stored as latent heat. The transformation also heats the material due to the plastic deformation. When stress is relieved, the energy from latent heat is perfectly recovered (dropping the material temperature), but the energy

of the plastic deformation cannot be restored. The final temperature of the material is raised exactly proportional to the energy dissipated by the area inside the hysteretic stress-strain plot:

$$\int AL\sigma d\epsilon = \int AL\sigma\epsilon_L d\xi = mc_p(T_{final} - T_\infty)$$

(29)

During the round-trip, the total elastic energy is conserved so that component of the integral is not considered.

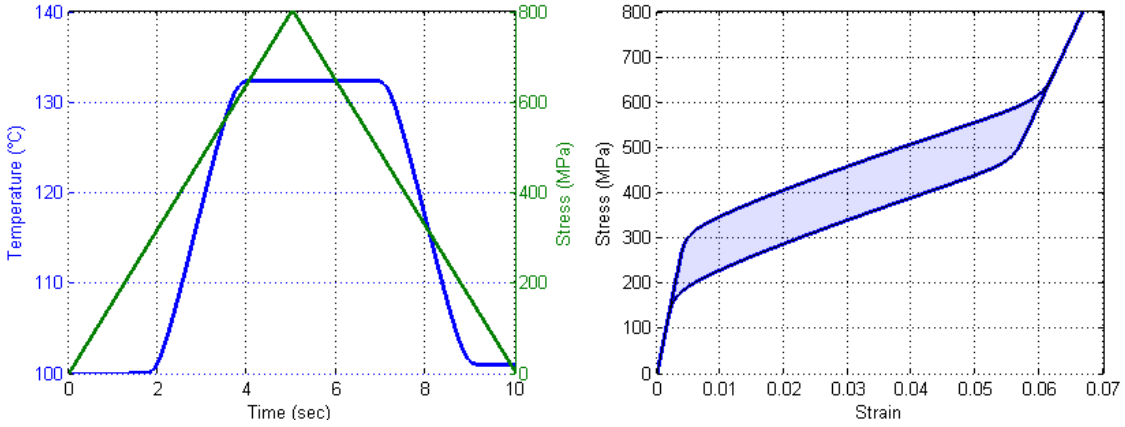


Figure 58 - A perfectly insulated, super-elastic SMA wire is stretched and released. After stress is relieved, the material temperature has risen. The material acts like a damper: energy is dissipated by the process.

Consider the same SMA wire but with no insulation, allowing convection to the surrounding air (Figure 59). In this case, the system is able to constantly release or absorb energy from the air with first order dynamics. When the stress is increased, the wire temperature begins to increase – but this energy is slowly released to the environment. When the stress is relieved, the endothermic transformation from Martensite to Austenite draws energy from the temperature of the wire and eventually stabilizes with the environment. This causes the curious effect of the wire cooling to below room temperature when the stress is relieved. The heating and cooling greatly affect the stress-strain behavior of the material (compare Figure 58, right and Figure 59, right). The balance between latent heat oscillation and dissipation of heat to the ambient air was studied in detail previously [184, 185].

This effect can be easily reproduced in the lab with a super-elastic SMA spring. Have someone quickly pull the spring and touch it while holding it extended – it will be hot for a few seconds until it cools to room temperature. After it cools, quickly release the stress and feel the spring now – it will be cold, even to the point of creating condensation.

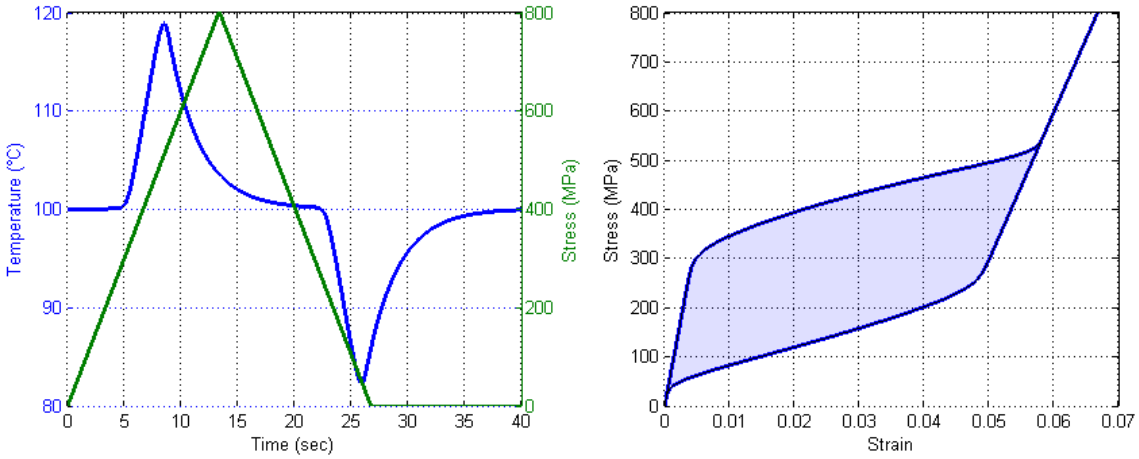


Figure 59 - An un-insulated SMA wire is stretched and released in an ambient 100°C environment. Due to the endothermic M to A transformation, the temperature of the wire can decrease below ambient

Electrical Heating

As with most conductive materials, when electric current is passed through the material an SMA wire converts some electrical energy into heat. In most systems, this would be considered inefficiency, but when creating NiTi actuators we will use it to our advantage. Fortunately, it is usually possible to apply sufficient electrical current to heat SMA wires very quickly. As the most common shape memory material, NiTi will be used in examples in this chapter. The most common form for NiTi devices is still round wires, so all effects will be quantified in terms of the diameter and length of those wires.

Common NiTi has an electrical resistivity ρ of 76e-8 Ohm*m when in the Martensite phase, and 84e-8 Ohm*m when in the Austenite phase [51]. This variation can have a large effect on the behavior of the wires when heated using constant voltage or constant current sources [61]. It has also been used as a strain estimation tool, though the author has shown that to be unreliable (c.f. Chapter 7). For some simple comparisons, take the average of 80e-8 Ohm*m as the typical value. The resistivity of Copper is 1.7e-8 Ohm*m and the resistivity of Nichrome is 100e-8; 47 times lower and 1.2 times higher than NiTi average, respectively. The relationship between the metal's intrinsic resistivity ρ and a wire's total electrical resistance R is

$$R = \frac{\rho L}{A} = \frac{4\rho L}{\pi D^2} \qquad (30)$$

where L is the wire length from power to ground, A the cross-sectional area, and D the diameter.

For now, consider only the heating and storage without dissipation. Assuming any power source desired could be obtained, it is easy to say that the power input P to the wire can be driven by any current I or voltage V source:

$$P = I^2R = \frac{V^2}{R} = VI$$

In practice, most supplies provide a fixed voltage and can supply current up to a maximum limit. Once a supply voltage and length of wire (chosen, perhaps, to obtain a given strain) is selected, the choice of wire to be heated does not affect the heating rate. This can be shown with the following computation. Since the energy stored in the material is proportional to its mass, consider the balance of power assuming a perfectly insulated wire:

$$P = \frac{V^2}{R} = V^2\frac{\pi D^2}{4\rho L} = \delta * \frac{\pi D^2 L}{4} C\dot{T}$$

(31)

Where δ is the density ($\mathrm{kg/m^3}$) and C is the energy stored per degree of temperature increase (average of specific and latent heats). Then the heating rate $\dot{T}$ can be computed from (31) as:

$$\dot{T} = \frac{1}{\rho\delta C}\frac{V^2}{L^2}$$

(32)

and is completely independent of the total mass of wire. Thus, to improve the heating time of any SMA device, we must increase the allowable voltage, or decrease the length of wire between power and ground.

A complimentary concept is the effect of current on the heating rate. Some circuits can be designed (as detailed later in this chapter) to provide a regulated current through the SMA material, regardless of the voltage required to do so. If such a supply is provided:

$$P = I^2R = I^2\frac{4\rho L}{\pi D^2} = \delta * \frac{\pi D^2 L}{4} C\dot{T}$$

Rearranging;

$$\dot{T} = \frac{16\rho}{\pi^2\delta C}\frac{I^2}{D^4}$$

(33)

The heating rate is a function of current and the diameter – but is independent of the length. Datasheets (e.g. [51]) commonly provide suggested current for a given diameter – because it is

more likely for a designer to know their wire length and cut a ‘useful’ length from it, than to know an exact length beforehand.

To connect SMA wires to a computer or microcontroller, the power must be sourced in a way to heat the wires quickly while ensure no damage is done to the controller. The most effective and efficient way to heat SMA wire is to use a field-effect transistor as a voltage controlled, solid-state, ‘valve’ which allows the low power controller to regulate the flow of current from a high power (possibly high voltage) source. The circuit to connect an N-channel, enhancement mode MOSFET (a very common variety) as such a valve is show in Figure 60. In this case, microcontroller provides a voltage to switch the FET on, and the supply voltage provides power to the SMA – no high current passed through the microcontroller itself. A FET can be used as a variable valve or as a switch. Both modes of operation have benefits and problems.

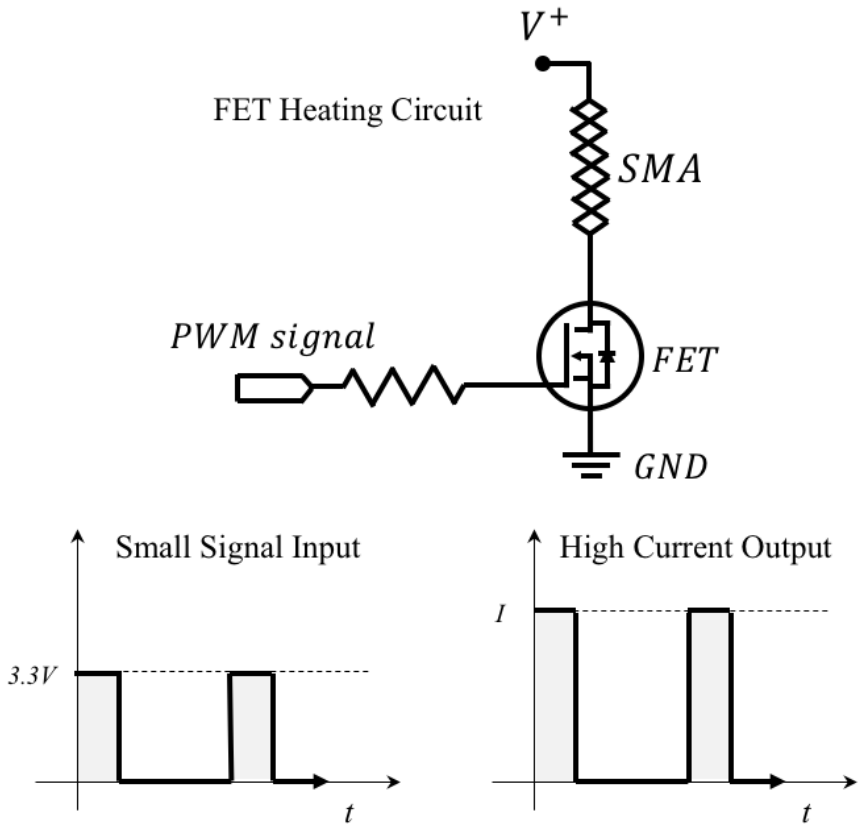


Figure 60 - Circuit for electrically heating SMA actuators using a PWM signal from a microcontroller, and using an n-channel MOSFET as a solid-state switch. The diamond element is an SMA wire.

FET as a Switch for PWM

In the first case, the design uses a pulse-width modulated (PWM) signal as shown in Figure 60 to switch the current on and off. This switch happens hundreds or thousands of times per second – so much faster than the heating and cooling speed of the wire that the wire only feels the average effect. PWM is a clever substitute for adjustable voltage; by keeping the switch closed (connected) for a long time and open (not connected) for a short time, the wire receives significant power. By keeping the switch closed for only a little time, and open for a long time, the wire receives little power. This percent of time the switch is turned on (closed) is called the 'duty cycle' $Duty$. The duty cycle proportionately affects the power in the circuit, not the voltage, following the equation:

$$P = \frac{V_s^2}{R} * Duty$$

While it is usually only complicating to consider it this way, the equivalent average voltage can also be written:

$$V_{average} = \sqrt{Duty} * V_s$$

Note that this effect is very convenient for SMA control – since we wish to control temperature, and temperature (in steady-state) is proportional to power, there is nominally a linear relationship between $Duty$ and T while ambient conditions are stable. However, this is not exactly right, since the electrical resistance of the SMA wire changes with temperature and phase fraction. To calculate the power in any case, both the voltage across the wire and the current through it must be measured (this also allows calculation of the SMA resistance). Typical microcontrollers measure voltage directly, but not current. To measure the current, a very small (0.1 Ohm or less) shunt resistor is place in series with the SMA wire. By measuring the voltage across the known-value shunt resistor, the current can be directly calculated. A possible location for the shunt resistor is shown in Figure 61 – note that this circuit would not be a good design if the supply voltage was higher than the micro-controller logic voltage since V_{SMA} is connected to the high power supply.

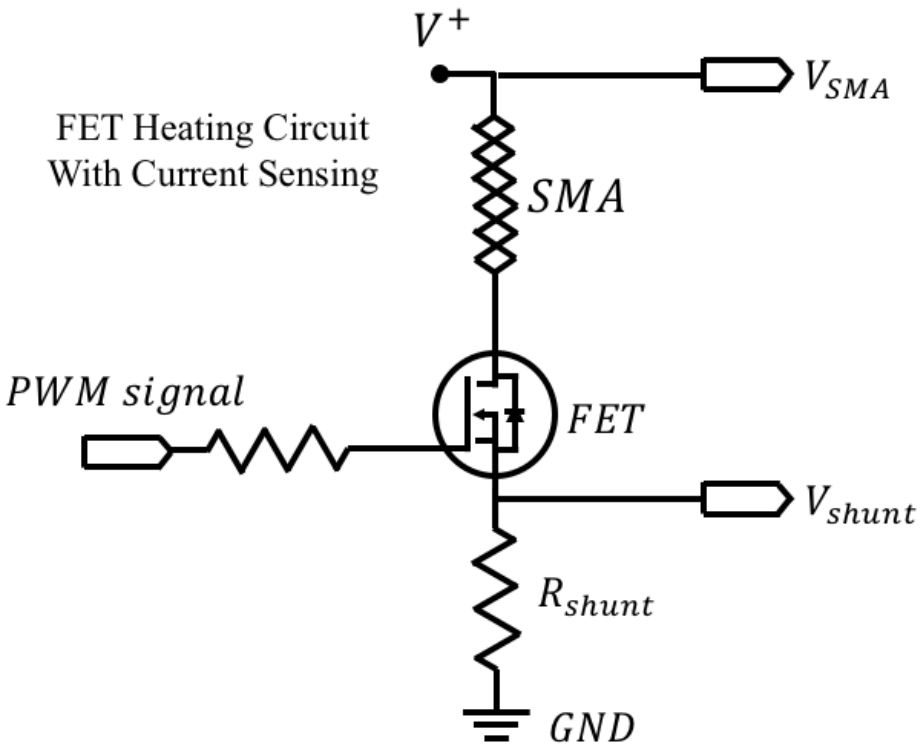


Figure 61 - An alternate circuit for PWM heating, with additional components to measure voltage and calculate current in the SMA device

It is also important when using this circuit to synchronize the pulsing and measuring times, since the voltage measurements must be taken when FET switch is closed and current is flowing through the SMA material. A slight modification of the circuit allows a single ADC input to measure both the supply and shunt voltage (Figure 62). When the FET is conducting, a measurement is taken which can be used to determine the current in the SMA material. When the FET is open, a measurement is taken which is equal to the supply voltage, and can be used to compute the resistance of the wire, the total power, and to ensure the supply is stable. This technique requires a very stable supply voltage that does not dip or spike when the FET is switched.

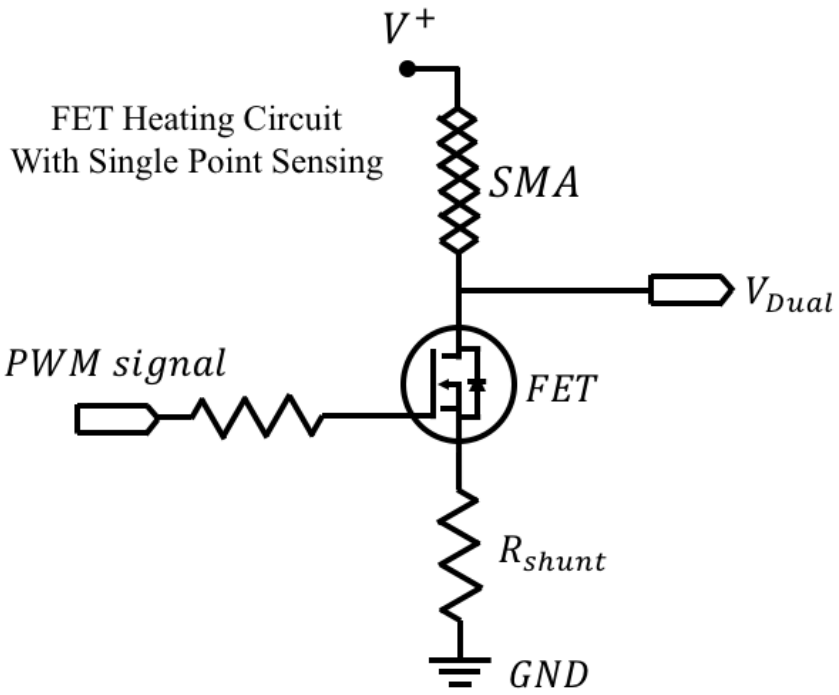


Figure 62 - Another variation of the PWM heating circuit which uses only a single ADC channel. When the FET is open, the ADC measures the supply. When the FET is closed, it measures the shunt voltage. These two measurements can be used to calculate wire voltage drop, current, resistance, and power.

This is the circuit that is implemented on most devices designed during this research since it is very inexpensive and powerful. When cost is a critical factor, the FET switch was replaced with Bi-junction Transistors. When accuracy was most important, the R_{DS} of the FET must be very low – DMN1019 N-FETs were an effective choice though only available as surface mounted devices.

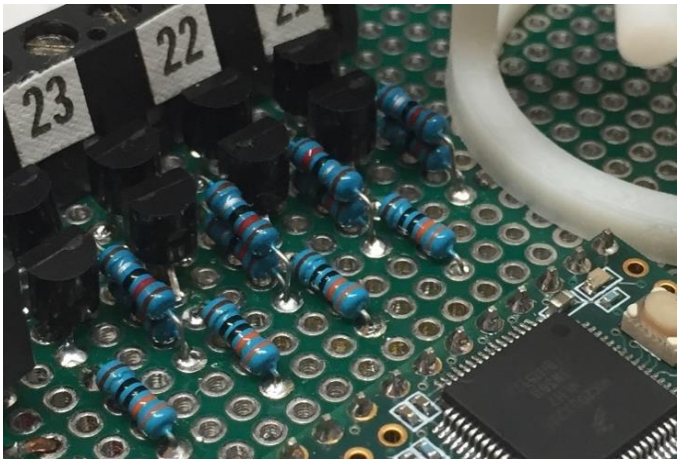


Figure 63 - Four switching/sensing circuits shown in a prototype SMA driver. The switch here is a BJT for low cost, although FETs provide higher performance in most cases

FET as an Amplifier

For some control and characterization experiments, the PWM signal interferes with the ability to take measurements of the material since the voltage pulse must be synchronized with the measurements. For this and other situations, it may be valuable to control the power in the SMA continuously without any effect from the PWM. This capability comes with increased cost in electrical components, but the primary burden is efficiency: a PWM signal reduces the total power draw based on the duty cycle, while most 'amplifier' circuits will dissipate significant energy in heating transistors and resistors in the circuit. For the precise Hardware-In-The-Loop equipment (detailed in Chapter 8), the need for smooth and stable control was more important than component cost, so an amplifier was built. When testing general wires of known diameter but generally unknown length (for instance, in a NiTi coil-spring), it is valuable to drive a known current through the wire instead of a known voltage – this ensures the heating rate is independent of the wire length and prevents melting and short circuits from destroying testing equipment. Another desirable characteristic is to perform all regulation on the high voltage end of the SMA device being testing – this ensures it remains grounded when unheated and is not a threat to short circuit equipment nearby. A high-side driven analog voltage-to-current amplifier was designed and built to accomplish this, developed based on a similar concept by Texas Instruments [186]. The schematic of this device is shown in Figure 64. The circuit has three stages. First, a 1^{st} order low pass RC filter converts the microcontroller PWM into a smooth voltage signal with bandwidth of 40 Hz. A second stage uses an op-amp feedback circuit which controls the voltage across a shunt resistor using a FET – this voltage produces a known current since the shunt resistance is known. A second op-amp feedback loop forms a 100x current amplifier between the shunt current and the SMA wire. This design was validated in LTspice software and built; the prototype device is shown in Figure 65. An INA219 IC is used to precisely measure the resulting voltage and current through the SMA device which is under test.

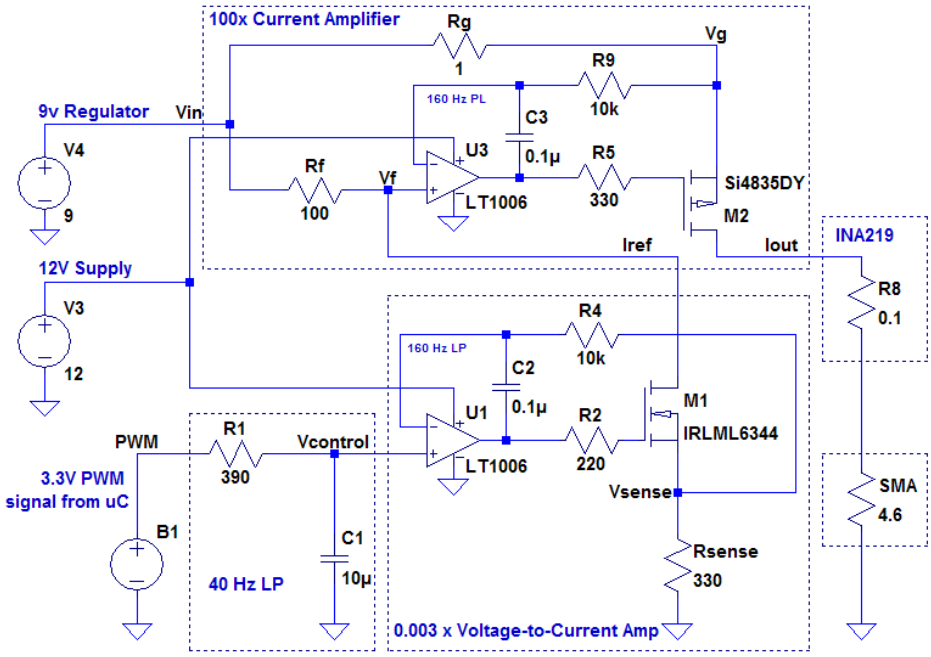


Figure 64 – PWM-to-current converter and amplifier

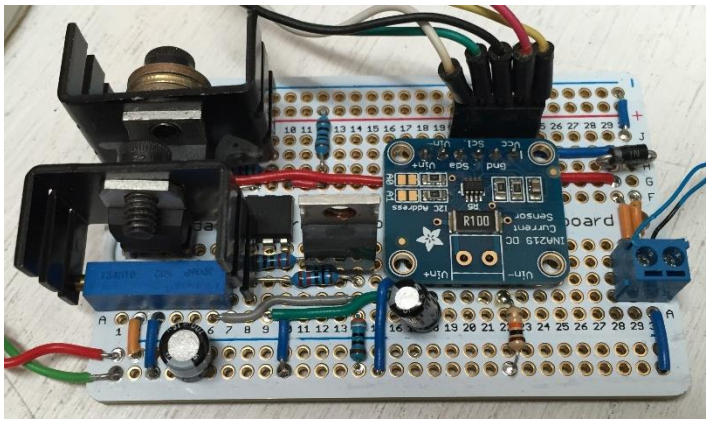


Figure 65 – PWM-to-current converter and amplifier prototype

Binary control of SMA Heating

One other method of controlling the heating of SMA is quite different than these two. By providing many connections to the SMA at spaced intervals, small segments of the wire can be heated without heating the entire device. Instead of trying to delicately balance the device temperature so that the inner structure is partially transformed from Martensite to Austenite, fixed regions are selected for heating and are heated as to guarantee 100% Austenite. The best resolution of control with the fewest number of electrical connections is made by spacing the taps in a binary spacing – doubled length with every added tap (Figure 66). This method was

abandoned when the author could control the position of SMA wires faster and more accurately using PWM across the entire device. It remains a useful tool for situations where it is exceedingly difficult to maintain a balanced temperature in the device, or where the position (or stiffness) control needed is not variable, but remains constant for long periods. Another benefit is that this scheme does not require a sensor for position – it can be 'open-loop' so long as we are ensured that the heated segments receive enough power to achieve complete Austenite transformation. Because of the stepwise motion and open-loop nature, these devices are to a normal SMA wire as a stepper motor is to a servo. A prototype is shown in Figure 67.

Figure 66 - Binary segments allow control resolution based on the number of divisions

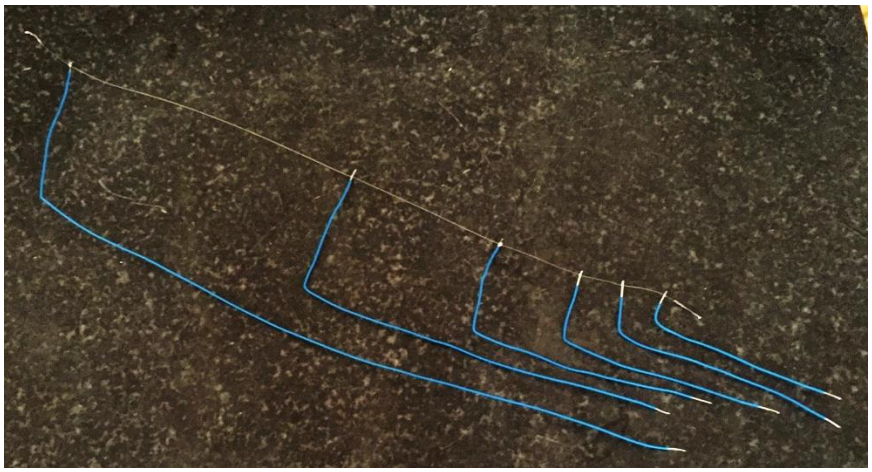


Figure 67 - A 5 'bit' segmented SMA device. For this simple wire it is not very useful, but for slow and stable motion on larger objects, the segmented heating is predictable and reliable.

Heat Dissipation to the Environment

In practice, the most limiting factor in SMA adoption is the slow cooling speed which directly leads to slow actuation speeds. By understanding the mechanism behind cooling, appropriate SMA actuators can be selected when bandwidth is a limiting factor, and techniques can be developed that increase the speed of heat transfer. The science of heat transfer uses equations that are notoriously complex and nonlinear and empirical – when considering SMA wire actuators we reduce the complexity by treating the wire as a lumped mass with a single 'heat

transfer coefficient' h which contains the complexity. Choosing a constant h is adequate for most applications, but if the most accurate models of the wire temperature many additional considerations can be made, as studied by the author's lab group [1]. The temperature of an SMA wire typically follows the heat transfer law presented in equation (27). Consider the homogeneous case without applied power, and ignore for now the latent heat of transformation. The mass is easily calculated using the density and volume, resulting in a first order differential equation:

$$\frac{\rho \pi L c_p}{4} \dot{T} = P - hD^2 LT$$

$$\dot{T} = \left(\frac{4h}{\rho D c_p} \right) T = \lambda T$$

The constant λ is the eigenvalue of the equation, which determines the time it takes for the actuator to cool from a given temperature to room temperature. It is noteworthy, and logical, that the wire length does not play a role in determining the time constant, and that the time constant is inversely proportional to diameter. This equation is only straightforward because it hides significant complexity in determining h. The heat transfer coefficient for small SMA wires was measured experimentally by considering the first order heat transfer equation in steady state:

$$P_{ss} = h\pi D^2 LT$$

A coil-spring formed from 0.25mm SMA wire was heated using a known current source. The temperature was measured using an infrared camera, and h was computed to be 121.8 $\mathrm{W/m^2K}$. The experimental setup is shown in Figure 68.

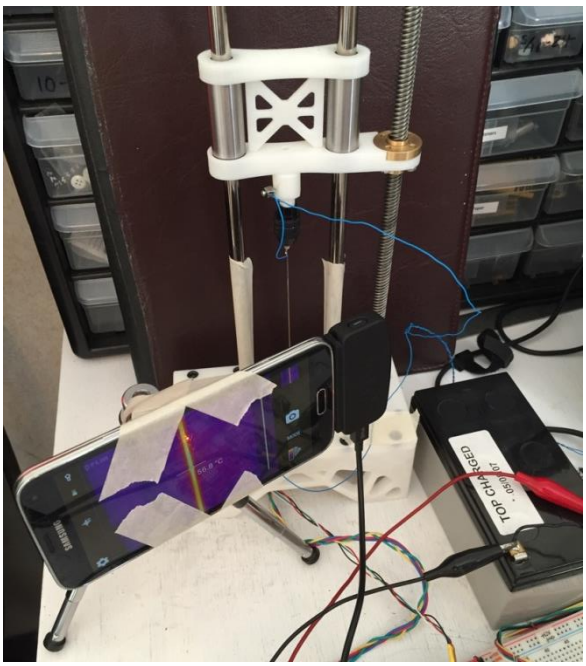


Figure 68 - Experiment to determine heat transfer coefficient using the steady state temperature measured by a FLIR camera

Curiously, more detailed modeling by the author's lab reveals that the diameter, wire orientation, and ambient temperature influence h though these effects are relatively small [1]:

$$h(T, T_\infty, D) \approx 65.5 e^{-\frac{D}{4}} (T - T_\infty)^{1/6} \quad (34)$$

The computed value for small NiTi wires was in the range 100-130 $\mathrm{W/m^2K}$. After validating equation (34) and comparing to experimental values of h [1], the eigenvalue for SMA wire near room temperature in still air is seen to be a function of D^2:

$$\lambda \approx \frac{0.0086}{D^2} \ (Hz) \quad (35)$$

The time it takes to cool to room temperature is approximately $3/\lambda$. This provides a practical lower limit to the bandwidth of the actuator. Changes such as blowing air across the device (forced convection) and immersion in water can greatly increase the speed. In practice, the actuator can be fully cycled at speeds that exceed this bandwidth because this eigenvalue determines the bandwidth of cooling completely to room temperature air – but the nonlinear SMA behavior occurs in a temperature range M_f to A_f. Different SMA alloys and heat treating

process can lead to higher transformation temperatures – the higher these are above the ambient, the faster the actuator operates and the more power it consumes. All these must be considered when designing an SMA actuator since they lead to a compromise of simplicity, strength, and speed (Figure 69).

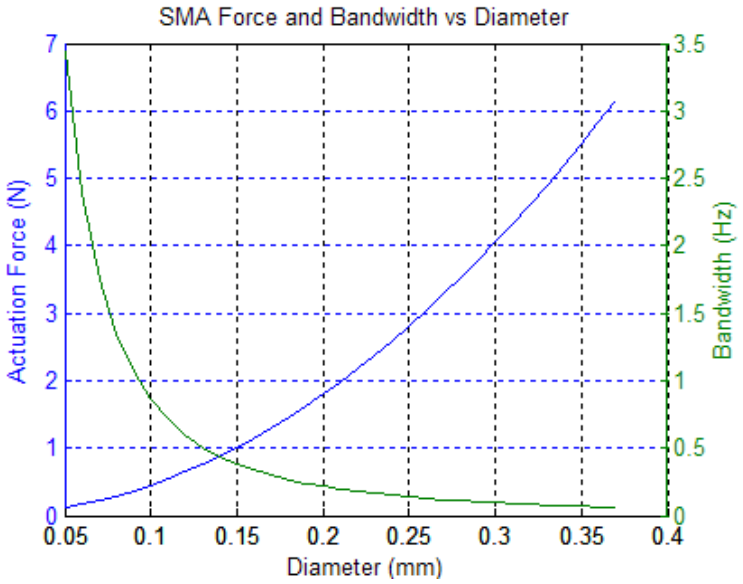


Figure 69 - the compromise between actuation force and bandwidth as wire diameter changes

Conclusion

It is critical to understand the heat transfer of SMA wires when designing controllers and mechanisms. Shape memory effect is completely driven by temperature; temperature change stands as the intermediate block between electrical input and shape memory motion. Increasing heat transfer can improve actuation speeds but also introduces disturbances. Consider a control situation in which the heating bandwidth is faster than the closed-loop control bandwidth (Figure 70). In this case, the controller performs well when the air around the wire is perfectly still (in a testing chamber), but is inadequate in ‘still’ room temperature air. As a rule, mechanical designers should increase bandwidth up to the needs of the control situation, but stop there: any increased bandwidth usually results in higher cost (finer wires), lower efficiency, and less natural stability to temperature disturbance.

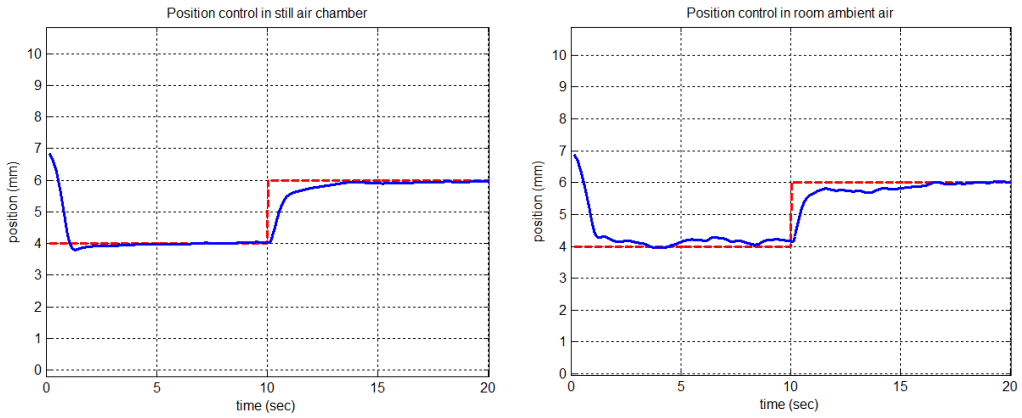


Figure 70 (Left) Control performance in a thermal chamber shows smooth regulation. (Right) Normal 'calm' ambient air has fast changes in heat transfer that greatly affects this slow controller

Using the relationships developed in this chapter, a mechanical design guide can be formed which balances the required actuator strength and bandwidth as in Figure 69. The relationships also suggest design improvements: it is known that NiTi films (sheets) can actuate faster than wires due to their high surface area, although such sheets are currently prohibitively expensive. Any method to increase the surface area is beneficial – the author has formed SMA material into braids and cords which have much higher surface area and cool faster than single wires (Figure 71). Another experimental device – a 'Bowden tube' - made by the author can increase the bandwidth by placing the SMA wire in an insulating tube which increases greatly the heat transfer surface area, as well as providing other mechanical benefits [5].

Figure 71 - Braided NiTi rope offers improved flexibility, much lower cost, and much faster heat transfer than a solid rod of equal weight

Chapter 7

Dual Measurement Self-Sensing Technique

Using a Shape Memory Alloy (SMA) actuator as both an actuator and a sensor provides huge benefits in cost reduction and miniaturization of robotic devices. Despite much effort, reliable and robust self-sensing (using the actuator as a position sensor) had not been achieved for general temperature, loading, hysteresis path, and fatigue conditions. Prior research has sought to model the intricacies of the electrical resistivity changes within the NiTi material. However, for the models to be solvable, nearly every previous technique only models the actuator with very specific boundary conditions. Here, we measure both the voltage across the entire NiTi wire and of a fixed-length segment of it; these dual measurements allow direct calculation of the actuator length without a material model. We review previous self-sensing literature, illustrate the mechanism design that makes the new technique possible, and use the dual measurement technique to determine the length of a single straight wire actuator under controlled conditions which show the method is reliable despite loading and temperature changes. This robust measurement can be used for feedback control of wire length in unknown ambient conditions and unknown end loads.

Introduction and Motivation

A clear obstacle to creating robot dexterity is in limits to their versatile poses – the degrees of freedom in their motion. While computer science is quickly finding ways to control high degree-of-freedom (DoF) robots (deep learning, for instance) [187], the hardware is not keeping pace. For example, computer graphics in film can reconstruct 244 degrees of freedom in the human body and convince millions of movie-goers that they have seen a living being [188]. The most pressing obstacle to dexterity and dexterous manipulation in modern robotics applications is physical degrees of freedom [189].

Today, actuator design is the limiting factor – motors, pneumatics, and hydraulics are simply too heavy and too big to package hundreds of DoF in a single system. This is particularly apparent at small scales – common ants have over 20 useful degrees of freedom that have been mimicked in prior research [190], but there is no mechanical motor today that can match an ant's

leg in size and strength. For commercial applications of highly dexterous robotics to become viable, the actuators that give the robots their dexterity must be extremely low cost because many more actuators will be needed than in more traditional systems. The ideal actuator must provide extremely high strength-per-weight, while allowing many to fit in a small space (DoF/m^2), at low cost ($DoF/\$$). This research outlines how Shape Memory Alloys (SMA) can be utilized to create actuators that fit this mold. A primary discovery of this work is a self-sensing technique that allows accurate control of simple SMA wire actuators.

Because both the models and controls are very complex, SMA actuators have not become prevalent in robot or machine design except in places where binary (two-position) motion is all that is needed. Binary control allows the model to be as simple as two known positions; the state of the SMA when heated and when cold. A more useful actuator will be easily commanded to obtain a desired position (or force) anywhere within its range of capability, not just at the limits. In order to accomplish this, a position (force) sensor is generally used in a feedback loop with the actuator. While sensor feedback is feasible in many cases, the great benefits of SMA actuators (high payload-to-weight ratios, simplicity, and small size) cannot usually be realized in this configuration because the added sensor adds complexity, weight and volume. Ideally the changing electrical properties of the SMA material itself could be used as a position (load) sensor. Because the behavior of the SMA motion and electrical properties are nonlinear and exhibit hysteresis, it has proven difficult to create a robust self-sensing control scheme [191, 192, 173]. No previous research has successfully modeled the electrical properties accurately for general loading and ambient conditions – this is crucial for a robotic actuator since the 'artificial muscle' must control motion under varying loads and reject disturbances during use.

We have developed an SMA actuator "dual measurement" method that allows self-sensing without complex models. This method is simple, only requiring one additional voltage measurement at a stationary point along the length of the actuator, and can easily be incorporated in practical robotic devices. By making the secondary measurement, the length of the actuator can be calculated directly and accurately with no assumptions about ambient conditions or loads except that they are constant along the entire actuator length. The method is not affected by hysteresis, twinning, or R-phase transformations which cause other methods to fail.

In this work, we begin by reviewing existing SMA self-sensing techniques, in particular resistivity modeling and attempts to utilize measurements of wire resistance to determine

actuator length. While these models do seem to model the material correctly, this work finds that no previous self-sensing method is robust to both unknown ambient conditions and unknown loads. Our robust dual measurement technique is described, and relevant equations are derived. An experiment is performed that shows the robust length sensing ability of the actuator. The dual measurement technique is compared to the use of resistance alone, and shown to be superior. A control example with several relevant scenarios demonstrates the ability of the dual-measurement condition to remain accurate in conditions that cause the resistance-based methods to fail. Finally, the accuracy improvements of this method, and potential sources of inaccuracy if the method is incorrectly applied, are discussed.

Literature Review

Resistance Modeling

It has been generally accepted in the SMA community for many years that the resistance of SMA materials can be used for self-sensing of strain. However, when trying to apply the concept, it is found that strain is only one component of the resistance variation; the actual response is affected by strain, stress, temperature, and Martensite phase fraction. In many cases, the easy solution is to assume the wire position is proportional to heating power applied [192], or to operate the actuator in only two positions at the end of the travel [130, 138]. Because of the large effect of transformation on strain (compared to other terms in the constitutive model) it has been proposed in several works that the relationship between strain and resistance is sufficiently linear to use for control [168, 193]. Wu has shown that the resistance is a very linear function of strain for super-elastic wires at constant temperature [63].

However, a more general sensing technique should allow control to intermediate positions. Most researchers approach self-sensing by modeling the resistivity of the material and the change with temperature, load and crystal phase fraction. A model employed by several researchers is the linear phase-fraction model with temperature dependent coefficients. Ikuta (1991) modeled the electrical resistivity (ρ) of an SMA wire as a function of Martensite phase fraction (ξ) consistent with the 'Voigt model' of the SMA [162, 30]:

$$\rho = (1 - \xi)\rho_A + \xi\rho_M \qquad (36)$$

ρ_A and ρ_M denote resistivities of the material in Austenite and Martensite phases, respectively. The resistivity of each phase can be computed as a function of the wire temperature:

$$\rho_A = \rho_{A0} + \alpha_A(T - T_0)$$
$$\rho_M = \rho_{M0} + \alpha_M(T - T_0)$$

(37)

T is the temperature of the wire actuator, T_0 is the initial measured temperature of the wire, α_M and α_A are the temperature dependence of the resistivity, and ρ_{M0} and ρ_{A0} are known quantities of the resistivities at the initial measured temperature. This model is simplified by neglecting the additional terms that account for the R-phase; R-phase transformation confuses many experimental results as it seems to behave macroscopically like Martensite, but has very different electrical resistivity than either Martensite or Austenite [165]. It has been demonstrated in many cases that transformation to the R-phase can be precluded by maintaining actuator stress above a critical value, often given as approximately 150 MPa [168]. The R-phase can also be eliminated by shape-setting at high temperatures, though this may have other adverse effects [57]. To be certain the change in electrical resistivity is due to crystal transformation and not another mechanism, Antonucci confirmed the effect of the crystal phases on resistivity using differential scanning calorimetry [165].

However, attempts at self-sensing with varying temperatures and stresses further complicate the model. Novak uses a similar model to Ikuta and adds an additional correction term for the resistivity dependence on wire stress (β_A, β_M) [163]:

$$\rho_A = \rho_{A0} + \alpha_A(T - T_0) + \beta_A\sigma$$
$$\rho_M = \rho_{M0} + \alpha_M(T - T_0) + \beta_M\sigma$$

(38)

These equations again neglect the R-phase components of the model in this equation, since it can be eliminated by preloading. This does indeed fit better to experimental data than the Ikuta model, especially for constant stress [163, 164, 194, 166].

There is, however, a realistic and more practical interpretation of the impact of applied stress than the addition of a stress term to the formulation of the resistivity. Because the stress causes strain (change in length) and contraction of area (due to Poisson contraction), the additional resistance can be found to be caused by this elastic strain and modeled by the definition of

resistance [167, 61, 45]. This can be modelled by noting that the wire resistance is defined as linearly proportional to the resistivity:

$$R = \frac{\rho L}{A} \tag{39}$$

R denotes the total resistance of the wire, L denotes the length of the SMA actuator and A denotes the cross-sectional area of the SMA wire. Both properties can be found in terms of initial wire dimensions and the wire stress and strain [16]:

$$L = \ell_0(1 + \epsilon) \tag{40}$$

$$A = \pi r_0^2 \left(1 - \frac{\nu\sigma}{E}\right)^2 \tag{41}$$

ϵ denotes the elastic component of strain, r_0 the initial wire radius, ℓ_0 the initial wire length, and ν is the wire's Poisson ratio.

It is clear that models to predict resistivity and resistance grow more complex, but are not accurate enough to use for practical controllers. These techniques have not yet explained:

- The plastic deformation the actuator incurs over its life
- Determination of constants for general SMA materials
- Temperature measurement techniques for very fine wires
- Temperature prediction techniques where measurement is not possible.

Empirical Techniques

All of the aforementioned resistance models are difficult to employ for self-sensing since they do not include any model of the hysteresis, and require knowledge of parameters that are difficult or impossible to measure (such as the phase fraction and temperature). These concerns are magnified for very thin wire actuators. Attempts have been made to reduce the error of neglecting hysteresis by employing polynomial models to fit the relationship between strain and resistance separately on the heating and cooling paths. However, these models rely on explicit knowledge of the wire stress [192, 191, 195, 196].

One clever solution for elimination of some nonlinear error is to arrange two SMA actuators in antagonistic manner and measure the difference in resistance between them [173, 123]. However, the material behavior is different for antagonist pairs than with single wires, and

requires additional consideration during implementation [197]. This method also does not sufficiently eliminate the effects of hysteresis, and cannot account for the effect of differing stress in the two antagonist wires.

Empirical techniques are good for cases where load is known, but not for general actuators. In particular, these techniques still cannot account for:

- The plastic deformation throughout the actuator's lifetime
- In situ fitting to data requires a temporary position sensor
- Arbitrary changing loads during actuation, or simultaneously changing the loading and temperature conditions of the wire [34].

Desired Self-Sensing Characteristics

A complete model of shape memory resistivity must model many aspects of the behavior to provide robust self-sensing. The following are some broad characteristics that a robust self-sensing scheme should include:

- Ability to obtain accurate measurements independent of the wire loading and stresses.
- Ability to obtain accurate measurements for all ambient air temperatures and currents
- Elimination of any errors associated with the hysteresis associated with both the shape memory and super-elastic behaviors of SMA actuators
- Ability to account for wire strains due to fatigue
- Ability to obtain accurate measurements independent of the specific alloy used.
- Ability to obtain accurate measurements without an exact knowledge of the SMA temperature

None of the self-sensing models considered to-date can account for all of these important factors. Most existing self-sensing controller schemes employ some form of a simplifying assumption or constraint on the physical model, such as:

- Electrical resistance is linearly proportional to Martensite phase fraction [198, 199, 130]
- Hysteresis between heating and cooling curves is adequately small [168]
- SMA temperature is a known property [165]
- SMA temperature is proportional to input power [61, 45, 192]
- SMA temperature is a constant property [63]
- SMA stress versus actuator length is known or held constant [194, 166, 195]
- For two SMA wires in an antagonist orientation, the wires have the same stress [123, 173]
- Hysteresis is not affected by stress [191]

The Dual-Measurement Technique

The measurement system is comprised of two main components. The first is a high-side measurement of voltage across the entire SMA wire (V_3) across and current (I) through the entire SMA wire. The high-side voltage lead is directly connected to the end of the SMA wire, such that it moves as the wire changes length. The second component is a set of two voltage probes (V_1 and V_2), set a fixed distance (l_0) from the stationary end of the wire and a fixed distance from one another (l_s) (Figure 72) .

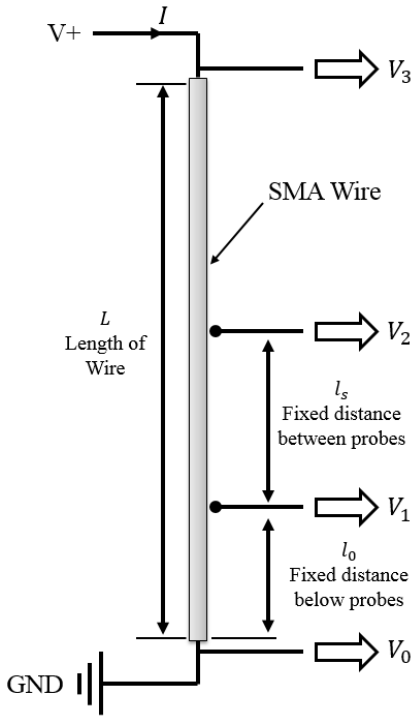


Figure 72. Dual Measurement Technique Schematic

The useful quantities that can be computed using the dual measurement are now presented. It is easily shown that, if l_s and l_0 are constant, then

$$L = l_0 + l_s\left(\frac{V_3 - V_1}{V_2 - V_1}\right) \quad (42)$$

Thus we can directly solve for actuator length with no knowledge of the actuator properties besides initial length and slider position, nor a model of the material. This is only accurate if the material properties are equal along the entire length of the wire, thus it provides a better estimate

for larger l_s. Because sensor noise is unavoidable, it is best to maximize the denominator in Equation 9 by having l_s as large as is practical, and l_0 as small as is practical. Thus, if possible, V_1 should be measured at ground ($l_0 = 0$), and l_s should be as close to the contracted length of the wire as possible. This also reduces inaccuracies arising due to varying conditions along the length of the wire. If these recommendations are made, then the model reduces to

$$L = l_s\left(\frac{V_3}{V_2}\right) \tag{43}$$

Notice that this provides the total length without ever measuring the total length. This alone could be beneficial in manufacturing where the lengths may not be consistent. The electrical resistivity is easily calculated using:

$$\rho = \left(\frac{A}{l_s}\right)\frac{(V_2 - V_1)}{I} \tag{44}$$

Experimental Setup

The device being tested is a simple, straight SMA wire. The physical characteristics of the wire are listed in Table 6, where A_f is the Austenite finish temperature at which the wire is nominally fully contracted, according to the manufacturer [200].

Table 6. Example SMA actuator characteristics

Characteristic	Value
Initial (Hot) Length	160 mm
Cold Length	168.1 mm
Diameter	0.125 mm
Mass	12.6 mg
Expected A_f	90 °C
Treatment	Kroll's Reagent
Manufacturer	Dynalloy Flexinol HT

The experiment is performed on a custom-built dynamic tensile testing machine that implements the features of Figure 72 (Figure 73). This machine controls the electrical input to the SMA actuator, and simulates a system for the SMA actuator to control. In this experiment, the simulated system is a simple suspended mass subjected to a constant force.

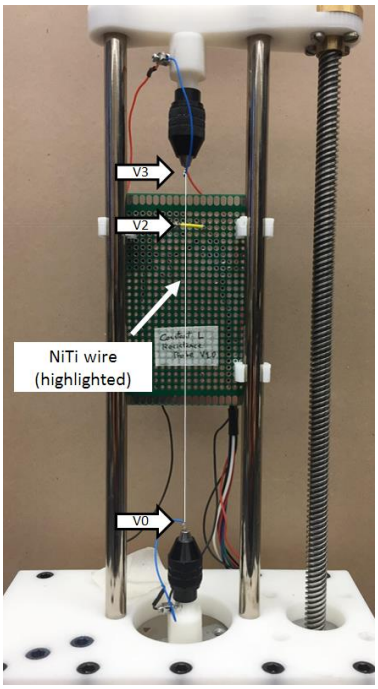


Figure 73. Tensile Testing Machine

The V_3 voltage lead is soldered to the end of the SMA wire, such that it moves as the wire changes length. Two voltage probes (V_1 and V_2) make sliding contact with the SMA wire on a 1.5mm Chrome-plated steel cylinder (Figure 74).

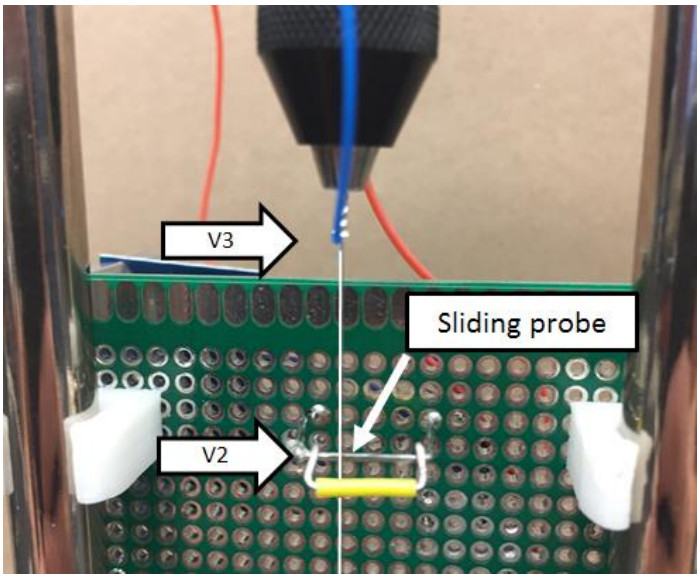


Figure 74. Sliding Probe for second measurement. Note that the yellow bar prevents wire tangling while slack

These probes are mounted on a plate in the testing machine such that they can be moved up and down to change l_0. Electric current is controlled through the SMA wire in a slow saw-tooth

manner, and the load is regulated. Three simulated masses are used; 100 gm, 200 gm, and 300 gm. By repeating the test with these different loads, we will see that the model can estimate actuator length regardless of the stress applied – a valuable result for robust sensing. The ambient temperature is approximately 25°C throughout the duration of the test.

Characterization Experiment

The experimental results validate the dual measurement technique. The experiment is performed in 3 parts. The driving input is SMA current, driven from 40 to 250 mA in a saw-tooth manner, and the load is regulated to 300, 200 and 100 gf during three current sweeps (Figure 75).

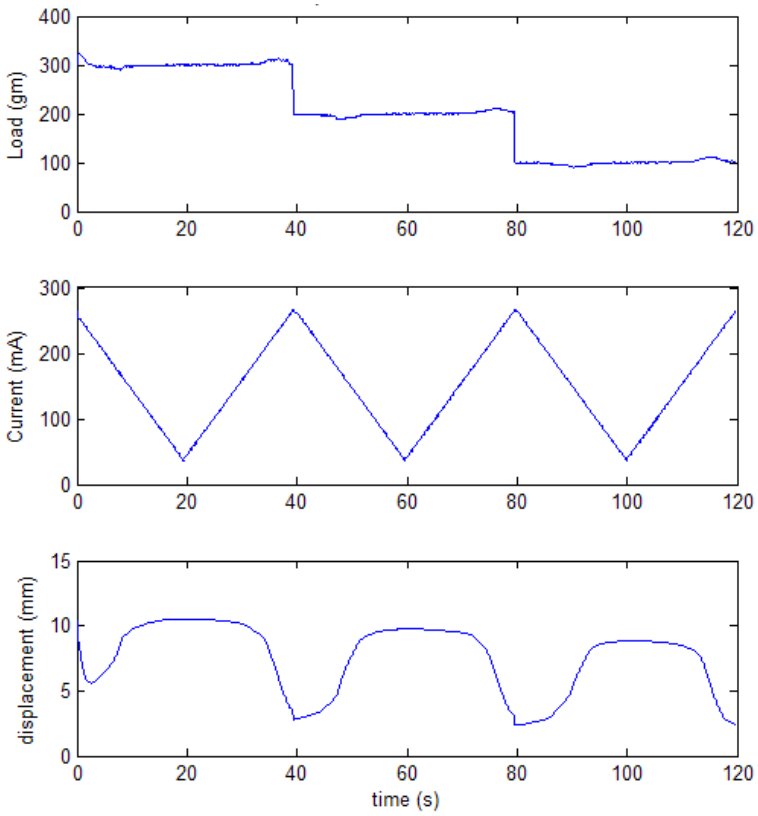


Figure 75. Driving inputs for the experiment are three loads held constant (top) while current is swept in a saw-tooth manner (middle) which extends and contracts the actuator (bottom)

The stresses due to these loads are approximately 80, 160, and 240 MPa, respectively. The ambient temperature is approximately 25 °C throughout the duration of the test, though no enclosure is provided to limit air currents since that is not typically realistic for general actuators. Before each new load, the actuator is allowed to stabilize at the new equilibrium position –

measurements during this transition are not recorded (causing data discontinuity at 40 and 80 seconds). The initial position, with the wire completely heated with no load is listed in Table 7.

Table 7. SMA actuator parameters required for length calculation

Distance	Value
L	160 mm
l_s	55 mm
l_0	91 mm

Characterization Test Results

The test shows the inadequacy of assuming the resistance directly relates to actuator strain; electrical resistance is plotted against true actuator displacement in Figure 76. Also, a line is fit through the data using least squares, and percent error is calculated from the difference between resistance and this line.

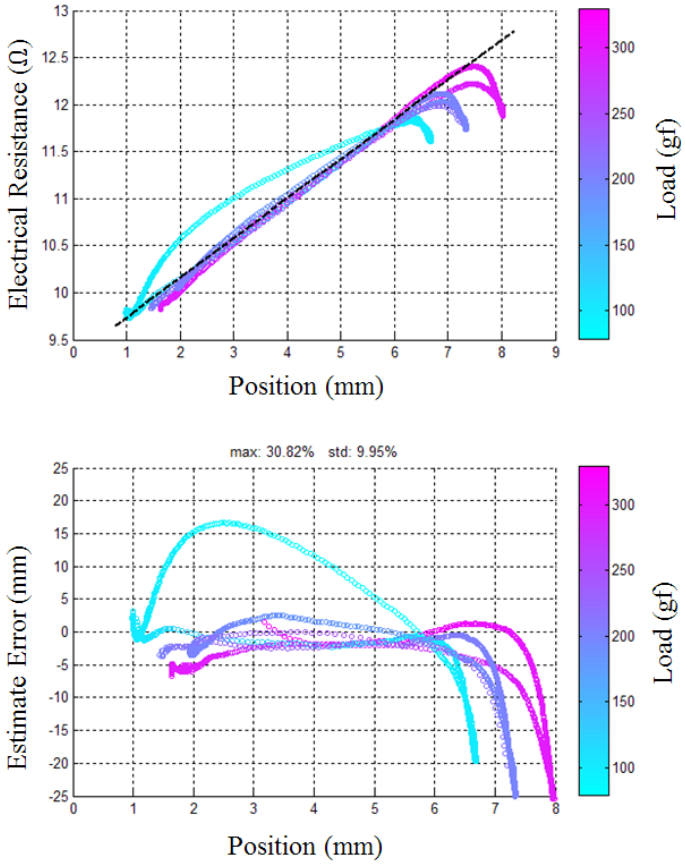


Figure 76. Self-Sensing Displacement Measurements and Errors using SMA Resistance versus actual Displacement (colored by loading condition). Resistance of the SMA wire, by itself, makes a poor displacement sensor due to combined effects of elastic strain, temperature, and shape memory hysteresis.

Several observations can be made about this plot:

- The main trend of resistance versus actuator displacement is linear away from the transformation boundaries
- There is large hysteresis when near 100% Martensite (longest length)
- There is a very wide hysteresis band at low load (100 gf)
- The relationship between resistance and displacement reverses near the ends due to temperature effects.

By computing a length estimate using Equation 9, we can directly calculate the actuator length as seen in Figure 77. The estimate is very linear and matches well to the measured position. The estimate becomes accurate soon after the wire touches the probe – this can be seen in the middle of the plot as the wire initially heats and contacts the sliding probe (near 3 mm).

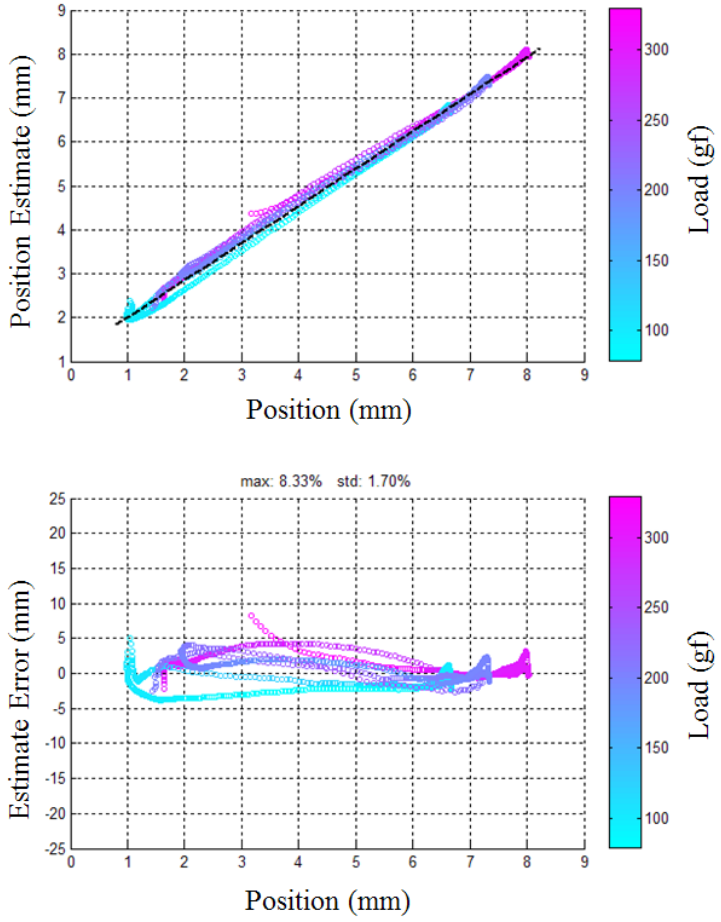


Figure 77. Self-Sensing Displacement Measurement and Errors using Dual Voltage Measurement Technique (colored by loading condition)

Feedback Control Experiment

An experiment was performed to determine the ability of the sliding-sensing method, compared to feedback with a position encoder, and feedback using the assumption that length is linearly related to wire length. The linear fits used are shown in Figure 76 and Figure 77 as dashed black lines.

The dynamic tensile testing machine was used to provide a constant antagonist force while the SMA wire was controlled using a PI feedback controller. The closed loop system is Type 1, and as such has zero steady state error with an unbiased measurement of position. The bandwidth of the machine is much higher than the bandwidth of the SMA actuator, which is characteristically very low [1, 201, 202, 203], and so its dynamics do not directly affect the SMA controller. The control experiment maintains a constant load on the SMA wire while a position set-point is changed. The load is regulated to 100, 200 and 300 gf during each experiment, respectively, to determine whether the two self-sensing techniques are reliable with changing loads. The experiment using the encoder serves as a reference to ensure the controller is capable of tracking the reference, and so any error is due to flaws in the self-sensing method.

Controller Test Results

The three trials showed nearly consistent results, with the encoder providing the most accuracy. The dual measurement self-sensing technique provides nearly as accurate a measurement as the encoder, while the resistance based position measurement lacks the accuracy of the encoder or the dual measurement technique's measurements. The control reference was set to two different points, ensuring the sensing method was accurate across the range of SMA wire motion. To understand the control motion and the meaning of positioning accuracy, the performance of the control using encoder feedback is provided (Figure 78. Wire position feedback with 200 gf load.

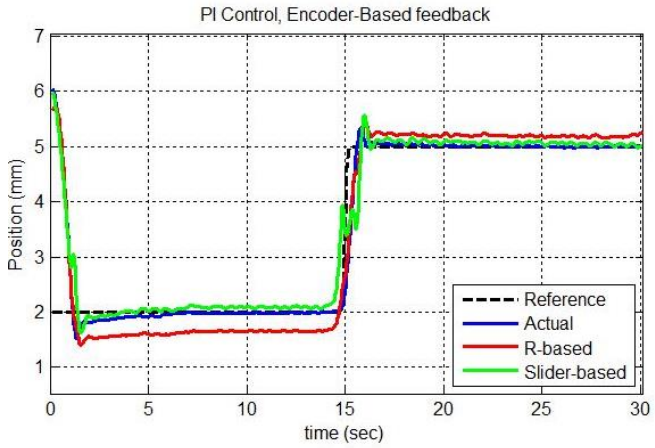


Figure 78. Wire position feedback with 200 gf load

The controller quickly drives the system to the desired set-points, settling in less than five seconds at each point. By observing the resistance and slide-sense position estimates during this test, we get a first understanding of the inaccuracy they will induce. The resistance-based method has large deviations from the true position both dynamically and in steady-state. The slider-based self-sensing is accurate and nearly matches the encoder position measurement.

A better presentation of the relative accuracies of these three methods is given by using the respective position estimates for controller feedback. By comparing the tracking error between different methods (encoder, resistance, and slide-sense feedback), and loadings (100, 200, and 300 gf loads), the range full range of ability can be seen in Figure 79.

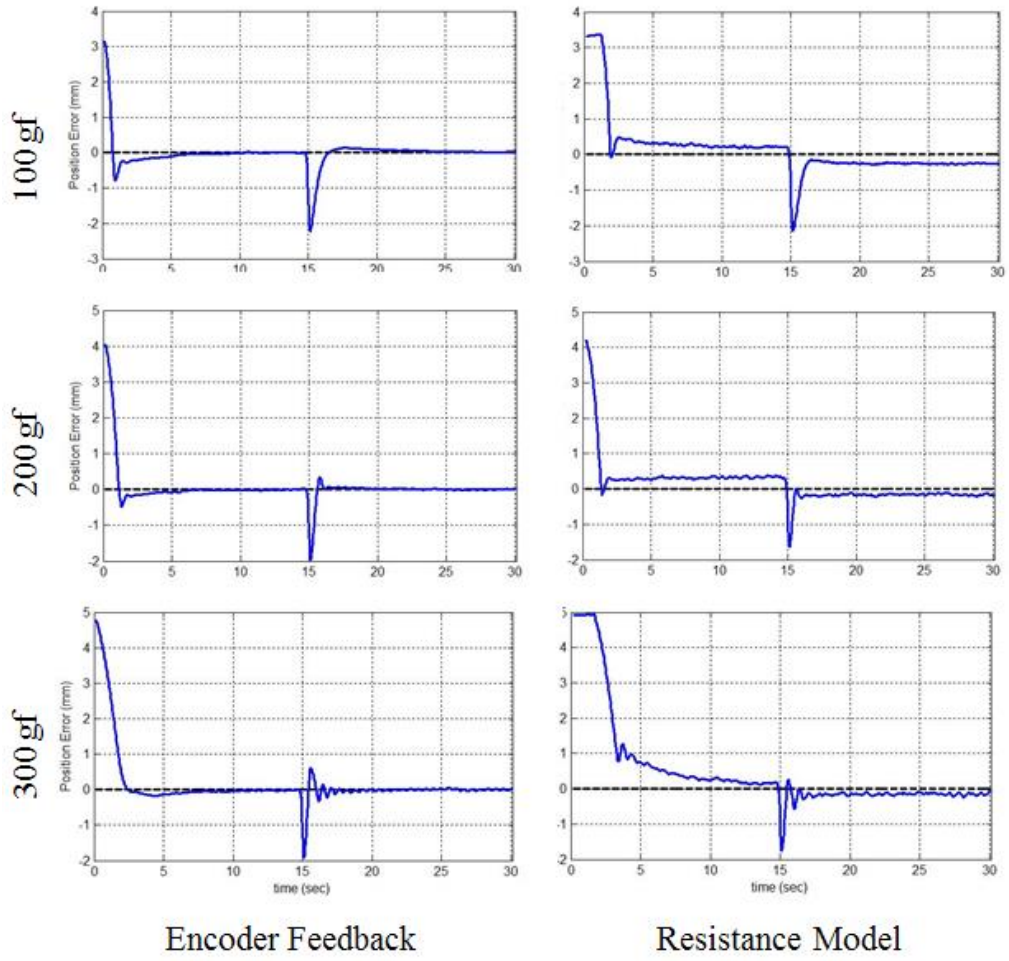


Figure 79. Position error using encoder feedback and the resistance model of self-sensing at three antagonist loads

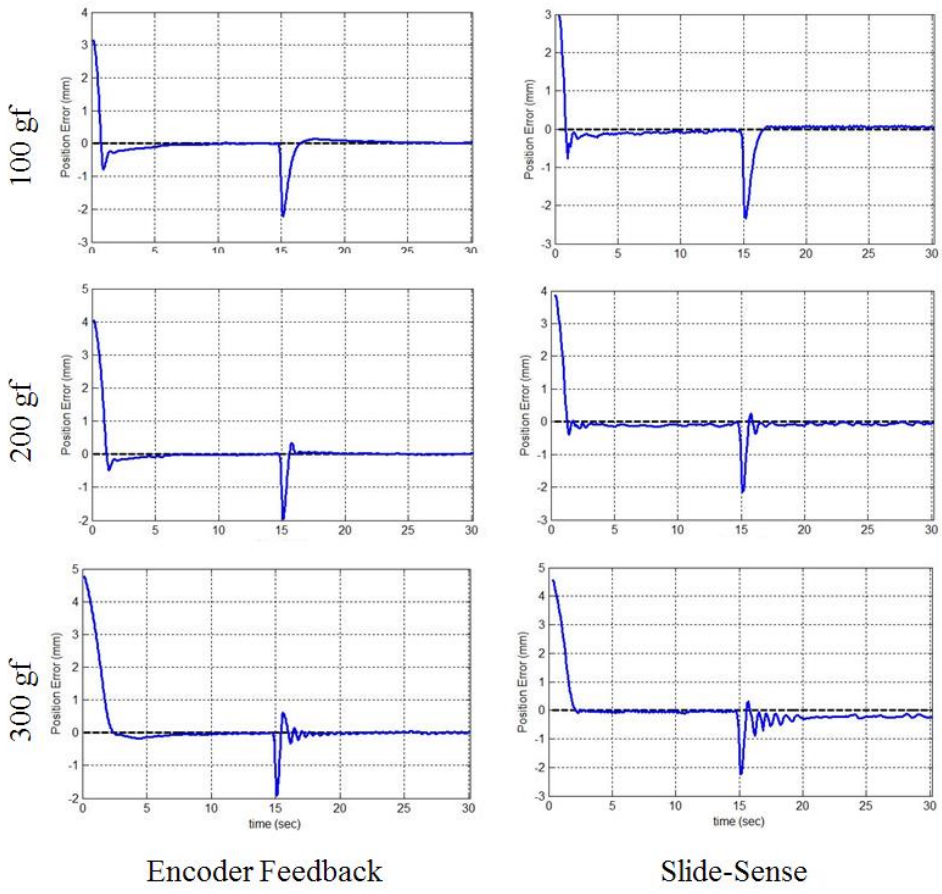


Figure 80. Position error using encoder feedback and the dual measurement technique at three antagonist loads

First looking to the encoder feedback, we see the performance of the PI controller. In all cases, the controller is able to regulate the position with no steady-state error. The controller stabilizes the system within five seconds in all cases. At high loads, the controller has some difficulty stabilizing at the second set-point when the wire is soft (cold, long) after fifteen seconds in each test. This is due to the decrease in the plant natural frequency (lower spring rate, increased load), which has shifted the problem such that the controller gain is too high for the plant. This is not a problem in this experiment, but it shows the value of reviewing encoder feedback to know this is an effect of the controller, not the sensing method.

The resistance-based feedback shows two primary flaws. First, the initial accuracy in the measurement is very poor, as this is when the wire is completely cold and in either a twinned Martensite or R-phase state, so the hysteresis error is exaggerated. Because of the simple PI control law, this causes a ‘windup’ in the integrator, and impacts the stability of the system (slowing down tracking and causing nonlinear overshoot). At high loads, the windup of the

integrator is particularly poor. While this can partially be avoided by incorporating anti-integrator windup feedback, a more accurate measurement technique could allow for an avoidance of this complication altogether. The second primary concern of this measurement technique is that the hysteresis in the heating and cooling response for the electrical resistance causes steady-state error in the regulator – the system is never driven to the correct position because the resistance measurement has error that is never resolved.

The dual-measurement feedback quickly provides an accurate estimate of the position. The controller has no steady state error because the method provides a linear, unbiased estimate, except in the case of a very high load. At such a high load, the controller scheme does have some steady state error because the physical sliding probes deflected slightly under the pressure of the taut wire and this physical movement altered the measurement. For a more robust conductive probe, this concern is eliminated. The dual measurement technique does fall short of the encoder with respect to the signal noise, as the sliding does generate some electrical noise, and the resolution of the voltage measurements is a limiting factor.

Discussion & Conclusion

In more than thirty repeated tests, the measurement error is typically banded within 0.5 mm of the true length. This is true even as the wire shows permanent deformation in the form of fatigue with repeated cycling. Error is calculated by subtracting the estimated length from the true length; the error in all experiments shows standard deviation of error of 0.154 mm. As an actuator, we often care about the error as a percentage of the total travel, not length. With an effective stroke of 8.2 mm in the 160 mm wires used, this implies 1.7% standard deviation. For comparison, the linear resistance assumption provides a 9.95% standard deviation of error. Noise in the estimate appears to be solely related to noise in the voltage measurements made. Hardware or software filtering can effectively improve the noise.

In a control scenario using a generic, stable PI control law, the slide-sense method provides an unbiased, accurate estimate of position. The method is consistent until the test setup was physically bent due to high load on the probe. For comparison, the resistance feedback method shows integrator windup errors and steady-state error, both due to real hysteresis in the wire.

The dual measurement technique has allowed us to estimate the length of a general SMA wire without prior knowledge of:

- Macroscopic electro-mechanical material properties of the SMA actuator
- Applied loads and temperature conditions of the wire
- Ambient conditions of the cooling fluid
- The electric current being passed through the SMA actuator
- The initial wire geometry or configuration.

Thus we have created a robust mechanism for self-sensing in SMA actuators for use in robotics applications. Without a sensor for feedback, accurate robots cannot be built; the dual measurement technique has the potential to make SMA actuators more cost effective by eliminating the need for position sensors. It also allows the actuators to be smaller and to fit more degrees of freedom in a smaller space. Dual measurement provides a reliable and simple technique to implement robust self-sensing in SMA actuators.

The dual measurement method eliminates the major sources of error in existing methods, but introduces some new concerns:

- The wire surface must be conductive, meaning to use this method a surface treatment will be required.
- Corrosion of probe contact surface may affect the quality of the measurement over time
- Sliding contact may decrease the life of the SMA actuator
- Electrical noise in the form of Nyquist-Johnson noise or inaccuracies due to discretization in the ADC.
- End crimps or solder affects measurement accuracy for short wires.
- The assumption of constant resistance across the wire may not hold for short wires due to the effects of boundary conditions (this concern is alleviated for wires over 14.8 cm in length) [204].

Future work will study these potential issues. Superior surface finish of the wire could prove to eliminate corrosion and maintain electrical conductivity. Improved mechanical design of the sliding contact can eliminate wear – techniques from electric motors and slip rings may be adopted for this purpose. Electrical noise is directly related to the quality of the ADC and the level of filtering used. Empirical evidence suggestions that placement of the sliding probe affects the amount of electrical noise. In situations where the sample rate is fast compared to the required sensing bandwidth, software or hardware filtering can significantly eliminate the noise.

Chapter 8

Hardware In-the-Loop Testing Machine

To facilitate Shape Memory Alloy (SMA) robotics research, an experiment station is needed to test physical actuators as Hardware In-the-Loop (HIL). Finding no commercially available alternative, it was necessary to design and construct a tensile testing machine appropriate for SMA actuator research. No commercial tensile testing machine within this project's budget had the required features:

- High speed measurement and controls feedback
- Ability to simulate dynamic loads
- Programmable actuation inputs in coordination with measurements
- High strain rates and accelerations (requires low cross-head inertia)
- A regulated environment control chamber

A new machine was designed for this research that contains all those features. The SMA actuators studied in this thesis have a bandwidth below 2 Hz – the control system has a mechanical bandwidth near 25 Hz and sample rates of 50 Hz to ensure testing device dynamics do not interfere with the experimental data. Several common experiments are presented in this chapter that demonstrate the ability of the machine, more detailed experiments are shown in other chapters where this system was used to gather characteristic data about SMA wires and other devices, to validate material models, and to test control laws and heating methods.

Design Requirements

The design requirements for this machine are derived directly from the expected hardware to be tested; primarily, the devices studied in this research are formed from NiTi wires ranging in diameter from 0.075mm to 0.25mm. These are often used as straight wires (with high strength and low strain), but can be formed into springs, snakes, and braids as well (see Chapter 5). The devices range in length from 2 cm to 30cm – practical lengths for actuators in robots from insects to humanoid hand and arm muscles. A straight SMA wire even 30cm long would only have a strain of a centimeter or less, but a coil-spring or braid can exhibit strains beyond 400% of their contracted length. NiTi wire of 0.075mm diameter with an A_f temperature near 90 °C in room

temperature air has a bandwidth near 1.5-2.0 Hz. The machine therefore should have a bandwidth well above 2 Hz to ensure its tracking dynamics do not intermingle their effect with the measured effects of the NiTi. A 20Hz bandwidth is selected as the target for tracking control – this is 10x the device-under-test bandwidth (and compatible with the 50Hz sampling limits of the electronics). While high strength devices can be formed from many SMA wires in parallel, this machine will be used to measure simpler mechanisms. Straight, super-elastic SMA wires can be completely deformed into Martensite under stress of 400 MPa. For 0.25mm diameter wires, the largest expected in this research, this corresponds to an axial stress of 25 N (2.55 kgf). The machine should accept loads of 50 N (5.1 kgf) as a safe overload. Finally, the maximum actuation speed is determined based on the 2Hz bandwidth and maximum practical stroke amplitude of 2 cm: under sinusoidal inputs this requires a speed of 25 cm/s. These requirements are summarized in Table 8.

Table 8 - Design requirements for the HIL testing machine

	Max Value
Travel	30 cm
Load	50 N (5.1 kgf)
Speed	25 cm/s
Bandwidth	20 Hz

Design Details

The testing machine has five main components; crosshead positioning, load measurement, SMA power control, SMA electrical monitoring, and thermal chamber. All the components were purchased to provide acceptable range and resolution, and to keep the entire project budget below $200.00 (2016). The entire system is controlled using an ARM CORTEX M4 microcontroller; a Teensy 3.1 development board. Programming of the embedded electronics was completed in C and compiled for the ARM microprocessor. The user interface for the machine was built in MATLAB for recording and plotting experimental results. A computer model of the final design is shown in Figure 81. The complete machine with all mechanical and electrical components labeled is shown in Figure 82. The final machine has the specifications and performance capabilities shown in Table 9.

Table 9 - Final components and performance specifications of the HIL testing machine

	Mechanism	**Detail**	**Range**	**Precision**	**Accuracy**
Linear Positioning	Stepper Motor	NEMA 17, 8mm lead screw	0- 250mm	0.04mm/step 0.625um/μstep	≈ precision
Load Measurement	Load Cell	shear beam cell & HX711 24bit ADC	0-1kg 0-5kg	0.425 μg 2.127 μg	0.15g 0.75g
SMA Power Input (A)	PWM Voltage	16 Bit PWM 738 Hz	0-12 V	0.18 mV	≈ precision
SMA Power Input (B)	Direct current drive	16 Bit input, analog filter and amp	0-2 A	30 μA	≈ precision
SMA Measurement	Voltage/power monitor circuit	INA219 sensor	0-16V 0-3.2A	6 mV 0.8 mA	≈ precision
Environment Control	Thermo-electric cooler/heater	40 Watt TEG	0-95 °C	0.25 °C	1.0 °C

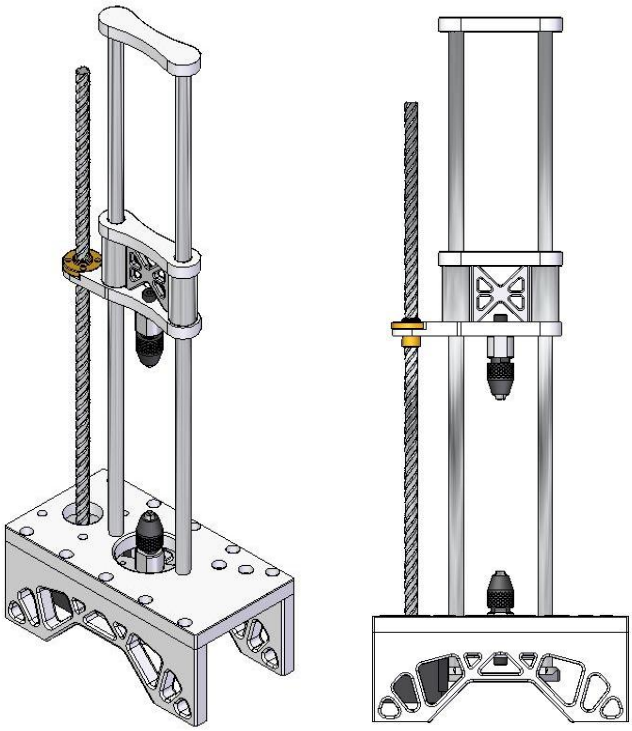


Figure 81 - CAD model of the HIL Testing Machine

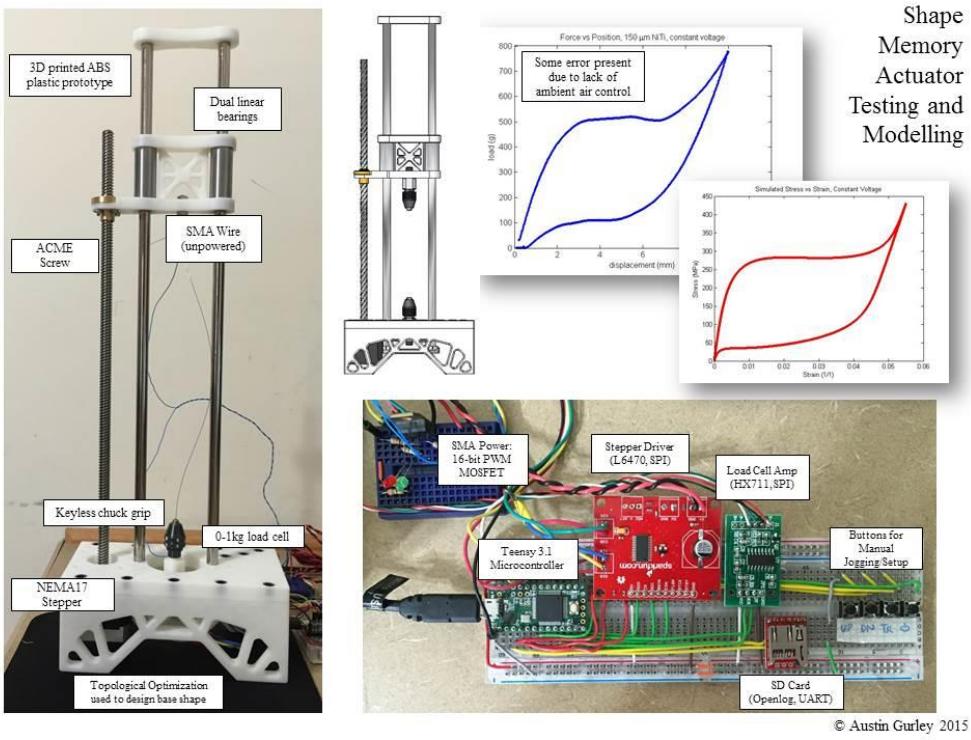


Figure 82 - Layout of the mechanical and electrical components that comprise the testing machine.

Position Control

Position is controlled using a NEMA17 size stepper motor. A dedication integrated circuit, the ST Microsystems L6470, is used to offload the computing of stepper motion from the main computer. The controller sends desired position, speed, and acceleration via SPI serial, and the L6470 drives the stepper to follow the commanded profile autonomously. Positioning is calculated using the number of steps taken by the motor in open-loop. A stepper motor of sufficient torque was selected such that it will not skip steps until the load on the crosshead exceeds 20kg – well beyond the range of use. The maximum speed is approximately 400 mm/s, which is fast enough for most SMA control simulations with SMA actuator bandwidths near 2 Hz, and strokes of less than 2 cm. A lead screw drives the crosshead along linear bearing rails. The total stroke length is 250 mm – in most experiments the range used is much less than an inch – but the long range allows many forms of actuators to fit in the one machine. Positioning resolution is directly determined by the number of steps per revolution of the stepper motor (400 steps/rev) and the lead screw pitch (16 mm/rev). Without micro-stepping, the resolution is 0.04mm per step. Using micro-stepping, the resolution is 0.625 micrometers. For reference, the

typical stroke of an SMA actuator is 4mm – so the resolution is as low as 0.01% of the actuator stroke. Because of the plastic machine frame, there is some potential for compliance between the screw motion and SMA movement. Much of this is eliminated by gravity keeping the components preloaded. Experiments in the useful range from 0-5kg load showed the compliance to be near 0.01mm, near the resolution of the stepper. The machine compliance was precisely calibrated by measuring deflection of the machine when pulling on a rigid rod, and is removed from the test data in the embedded software.

Load Control

A shear beam load cell is used for measuring force in the SMA actuator. These are special arrangement of strain gages that mechanically compensates for temperature variation and is insensitive to misaligned loading. The strain gages in this transducer are read using a bespoke integrated circuit (HX711) with a 24-bit differential ADC. Due to the choice of IC, this component is limited to 80 Hz sampling – though much faster than the bandwidth of the SMA actuator this IC limits the sample rate used for the entire machine logging and control system. However, with a desired sample rate of 50 Hz, this is not a concern. Two load cells are used – one with a maximum load of 1 kg, and the other with a maximum load of 5kg. The more sensitive cell is used for testing high strain devices (springs and braids) while the larger cell is used most often for straight wire devices, and straight wire characterization experiments.

SMA Electrical Driver and Measurement

The SMA power is controlled by applying a pulse-width-modulated (PWM) signal to a filter & current amplifier. The circuit for this device is presented in Figure 83 (A detailed explanation of this system was provided in Chapter 6). The PWM signal is easily generated by the microcontroller with a duty of 16-bit resolution, 11 kHz frequency. This signal is filtered using a first order low-pass RC filter with a cutoff frequency of 40 Hz. The filtered signal is input to a voltage-to-current conversion circuit, which uses op-amp feedback to maintain the current through the circuit proportional to the input voltage, regardless of load. This current is then used as the input to final circuit which acts as a 100x current amplifier – the high current side is connected to the SMA load. A ‘high-side monitor’ IC, the INA219, is used to measure the current and voltage across the SMA device. These measurements are not used for feedback – the circuit independently maintains current proportional to duty, and the INA219 is used to determine the voltage, resistance and power through the SMA device. This system is provided a

feedback controller within the embedded software so that it can quickly track any desired voltage, current, or power reference. For example, a constant power input can hold an SMA wire at constant temperature –but with constantly changing electrical resistance, the actual current must be adjusted (in feedback) to maintain constant electrical power.

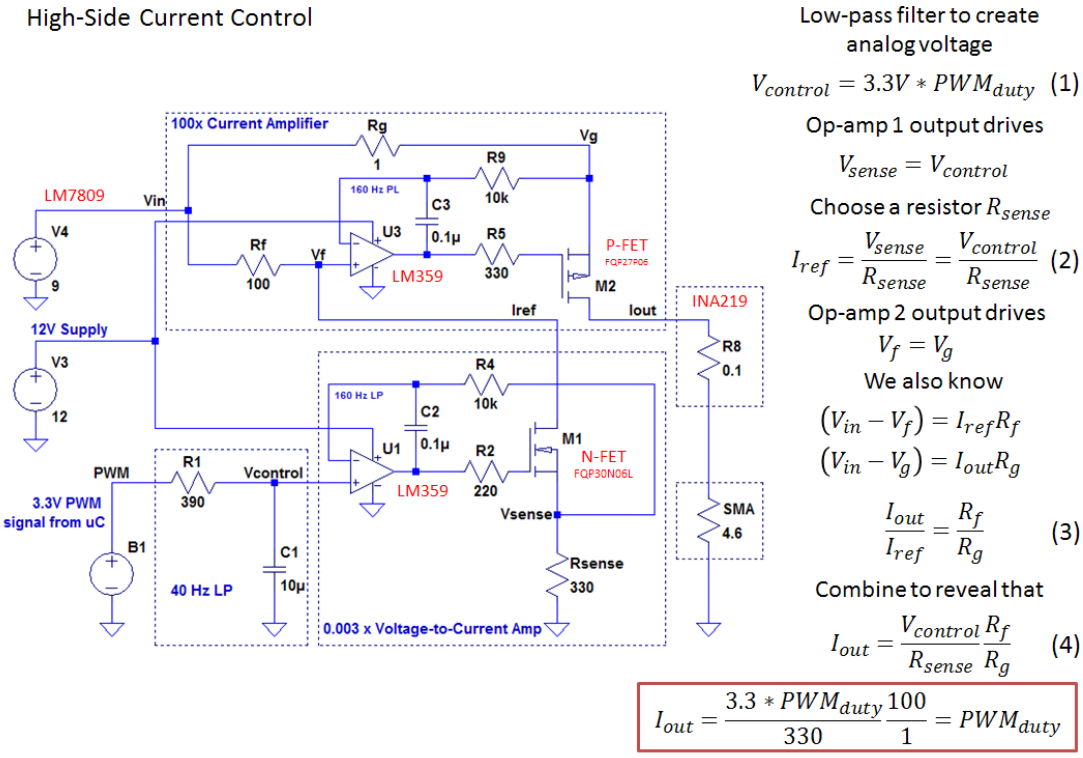


Figure 83 - Analog Circuit for converting PWM signal at 3.3 V into controlled current. The current is maintained using feedback (op-amps) regardless of the load on the circuit

Mechanical Frame

These components are all mounted in a 3D-printed ABS frame. The base houses the stepper motor and load cell. The crosshead (with no load-cell nor motor) is very light and therefore can be moved more quickly. The base dimensions are 6 inches wide by 3.5 inches deep. The entire machine is 18 inches tall. Most features of the device have proven satisfactory for this research. An improved design would provide for two identical steppers, one on each side of the crosshead, which would eliminate nearly all compliance. The load cell could be monitored using a faster IC, such that the machine sample rate is not limited to 80 Hz. The materials and electronics choices in other components are satisfactory.

Thermal Chamber

An enclosed chamber has been built around the tensile testing machine to regulate the ambient air currents, which can have a large effect on the motion of fine SMA wires. This chamber was later equipped with two, 40W Thermo-electric generators (TEGs) which can actively cool and heat the chamber from 0C to 95C. These are monitored using a K-type thermocouple (converted using a MAX31850 IC). A dedicated PID controller regulates the temperature to the user's desired set-point.

Possible Improvements

Although this HIL testing does not simulate every disturbance and load that will be seen in practice, it covers the most common functional situations where NiTi actuators are used. It characterizes the nonlinear actuators in terms of some common linear control concepts (such as open-loop and closed-loop bandwidth from sine-sweep inputs) and is an excellent way to test controllers against each-other with repeatable loads. The one situation which has been most difficult to simulate with this device is high-speed Sliding Mode Control (SMC). The rapid switching interferes with measurements of the wire properties during simulation, and can produce oscillations outside of the controllable bandwidth of the machine.

Another potential improvement would be to determine a robust method for measuring SMA wire temperature. Unfortunately, thermocouples are very large compared to the SMA wires and their mass skews the temperature near the probe. They are also generally conductive unless isolated from the wire – adding further bulk and mass. Potential solutions might include narrow-beam IR sensors or thermal cameras.

Simulation Suite

A suite of standard test procedures was implemented in the embedded software. The tests cover material characterization, open-loop device characterization, and closed loop control testing. Some of the procedures are for material characterization: direct strain (load) control simply oscillates the cross-head at constant speeds between a lower and upper set-point of position (force). This can be used to plot the shape memory or super-elastic effects. Other procedures compute the open-loop performance of the device: by sweeping the power input between set limits while holding force (position) constant, the effective bandwidth of the device with no control law is determined. These tests can be run with step, saw-tooth, or sinusoidal inputs of voltage, current, or power into the SMA device. Finally, closed loop control allows

control laws to be tested back-to-back in identical conditions. The simulation provides the cross-head a set ‘mass’, and a Proportional-Integral controller tracks the position to maintain a constant ‘gravitational’ load – thus simulating a hanging weight (the proportional term provides some ‘damping’ stability). During these tests, step disturbances in load can be introduced at any time. Simple position regulation, and tracking saw-tooth or sinusoidal waves, allows controllers to be tested and quickly tuned in software. Tracking sine waves with increasing frequency (sine-sweep) reveals the closed loop bandwidth of the feedback control device. These testing and simulation methods are built into a total of seven testing routines:

- **Direct Strain Control** – oscillates at constant speed between specified strains, while the device is heated with a fixed voltage/current/power
- **Direct Load Control**– oscillates at constant speed between specified loads, while the device is heated with a fixed voltage/current/power
- **Constant Force Sine Sweep** – load is held constant regardless of strain while the input voltage/current/power is swept in a step/saw-tooth/sinusoid of increasing frequency
- C**onstant Strain Sine Sweep** – position is held constant regardless of load while the input voltage/current/power is swept in a step/saw-tooth/sinusoid of increasing frequency
- **Simulated Plant Position Regulation** – a series of reference positions are toggled every few seconds for a custom feedback controlled actuator to try to match
- **Simulated Plant Position Tracking** – a smoothly varying reference position is provided for a custom feedback controlled actuator to try to match
- **Simulated Plant Tracking Sine Sweep** – a smoothly varying reference position of increasing frequency is provided for a custom feedback controlled actuator to match

Software Interface

Embedded electronics are powerful for collecting and recording data, but they are not adequate for analyzing it. To remedy this, a graphic-user interface program was built in MATLAB GUIDE which can connect to arbitrary text data streams (such as the one generated by the HIL machine) and plot and parse the data at high speed. The tool was designed to

1) Easily connect to a generic USB serial stream

2) parse the stream regardless of number of data channels being reported

3) quickly plot the data for quick inspection and

4) saves the data to the MATLAB workspace for post-recording analysis

The final interface is shown in Figure 84, where a device is attempting to track a sinusoid reference position that is beyond its bandwidth. This tool currently only allows reading and

recording the data. A future addition will add output control so that control laws can be written in MATLAB environment and directly implemented in the physical system. Currently the sample rate is sufficient (up 500 Hz*channels), but the latency (round-trip delay in communication) is too high for this to be practical.

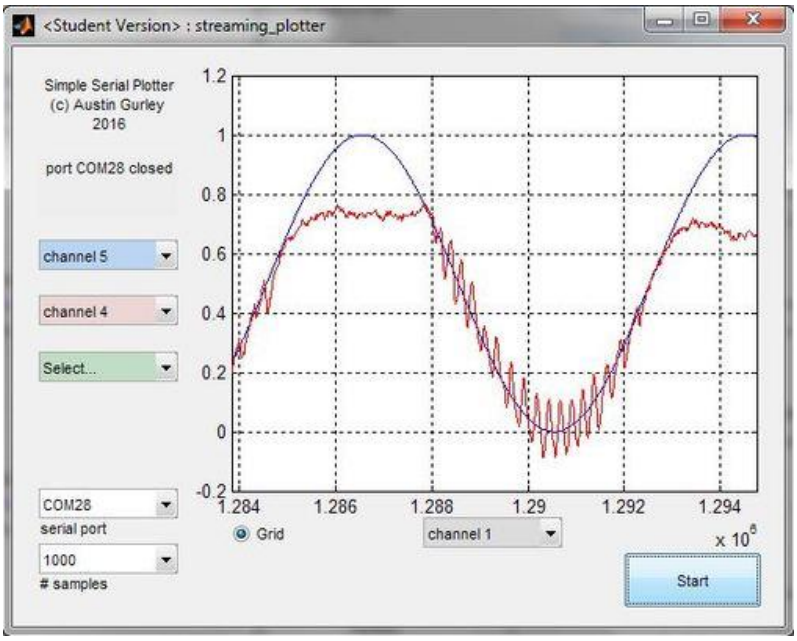


Figure 84 - Streaming Serial Plotter - connects to an arbitrary USB/UART serial stream and plots the incoming data

Example Experiment 1: Material Characterization

The results of a characterization experiment performed with this machine are shown in Figure 85. This test was performed to determine the stress-strain behavior of an SMA wire as temperature is varied. To maintain a constant temperature in the wire, the power (voltage*current) input to the SMA wire was regulated. In each of seven tests, the electrical heating power was held constant and the machine controlled the strain in the wire in a saw-tooth oscillating manner between 0 and 12 mm. The electrical power was increased in each experiment until the material would return to Austenite upon returning to the initial position (instead of slacking into twinned Martensite) – thus demonstrating the ‘super-elastic’ response of the SMA wires.

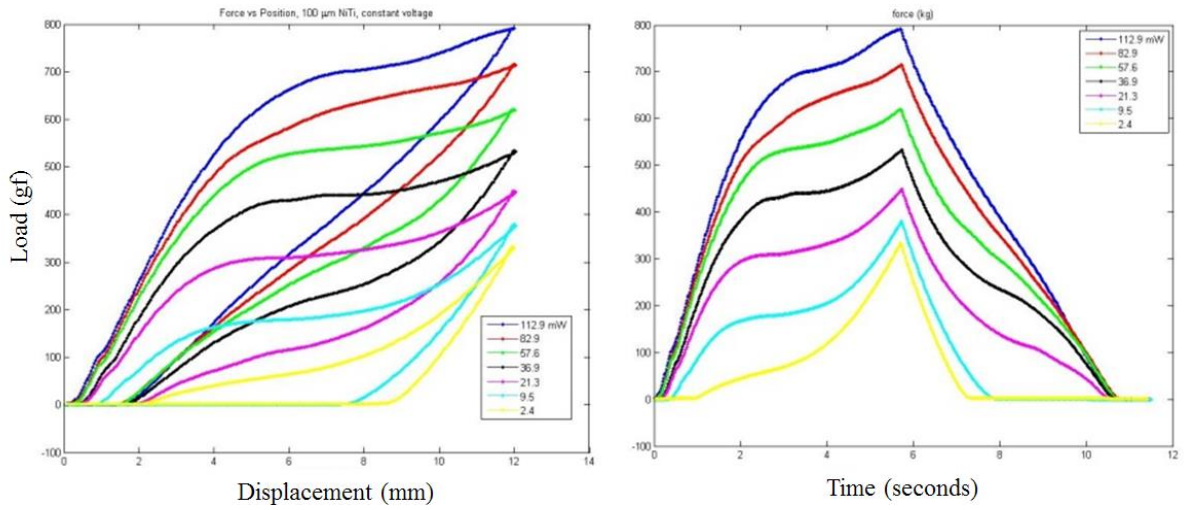


Figure 85 - A 100 micron wire was tested with several different constant-power heating inputs. On the left, load vs displacement curves show the behavior ranging from shape memory at low power (yellow) to super-elastic at high power (blue). On the right, load vs time is shown for each of the tests.

Example Experiment 2: PID vs Sliding-Mode Control

The most productive application of this HIL machine is in comparing control laws while maintaining identical ambient and plant conditions, for fair comparison. This particular data was generated from a test using a NiTi coil-spring (Figure 86). Coil-spring actuators have large strain (on the order of 200%), but provide low force compared to straight wires (20gf vs 200gf). Because the device is formed from a single wire, it can be heated electrically just like a straight wire, and cools approximately as quickly. This experiment compares a well-tuned PID controller and first-order sliding-mode controller when controlling a (simulated) 20 gf mass hanging in earth gravity with no damping. The results are seen in Figure 88 and Figure 88: The linear PID control does regulate the position with zero steady-state error, but is poorly damped in closed loop. This oscillatory performance is consistent with expectations of such a controller as presented later in Chapter 10. The sliding mode controller here was chosen with a slow sliding surface time constant. The nearly exponential decay to steady state is the expected response, which the switch controller tracks nicely. However, the SMC also reveals a flaw in this simulated plant: Because the switching interferes with measurements, the SMC switch is reduced to the controller's 50 Hz bandwidth which is not fast enough to prevent it from exciting motion in the machines internal feedback and tracking system. This simulation relic results in the 'twitching' in Figure 88, right. This problem is eliminated with faster sample rates in the physical system.

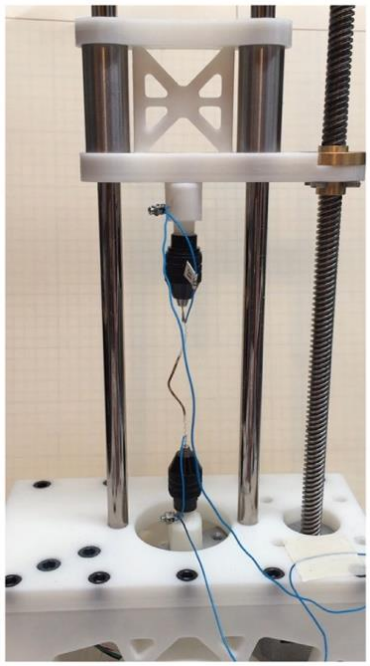


Figure 86 - A NiTi coil-spring actuator with a contracted length near 50mm and extended length of 125mm. The maximum load this small spring can create is only 30 gf.

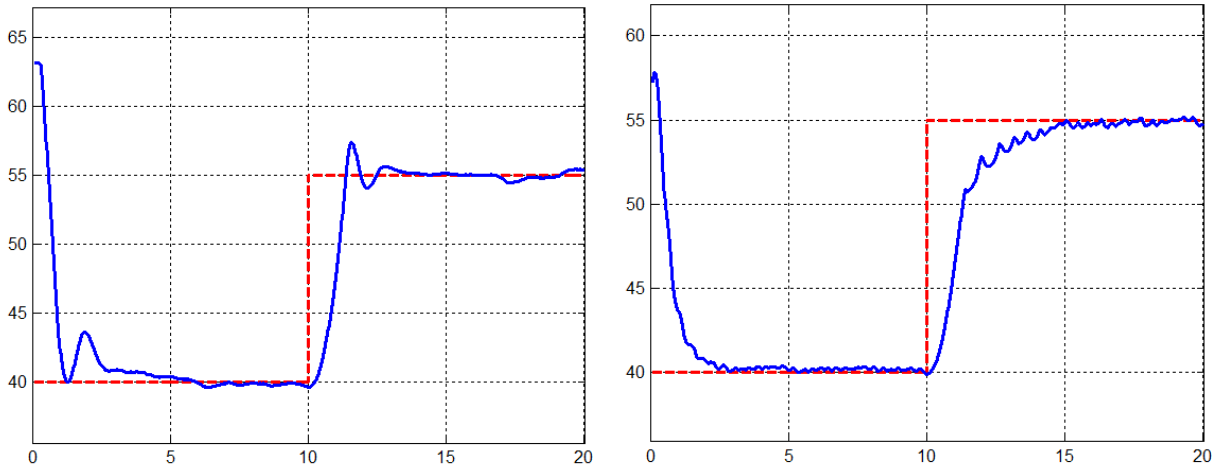


Figure 87 - Comparing the tracking control of a PID controller (left) with a sliding mode controller (right) during a step change in the position reference

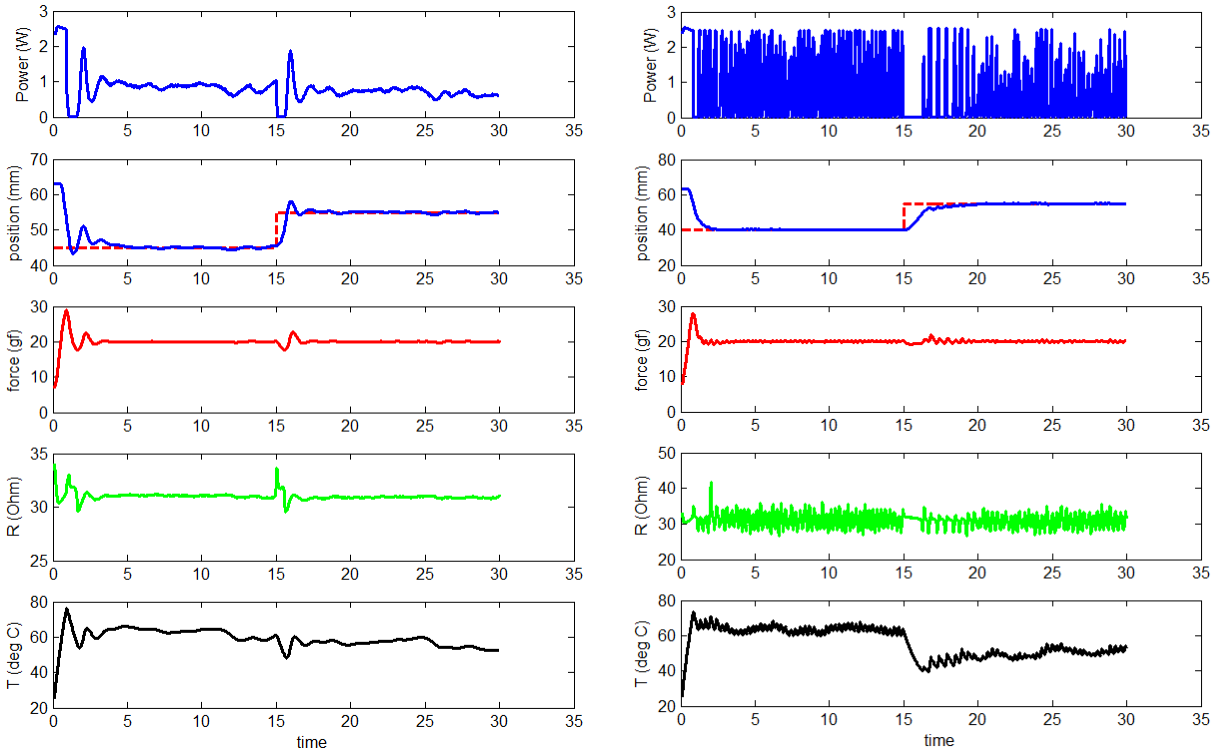


Figure 88 - Comparing the tracking control of a PID controller (left) with a sliding mode controller (right). All quantities measured by the testing machine are shown, plotted against time.

Conclusion

A Hardware In-the-Loop testing machine was designed for experimenting with SMA actuators. In the context of this research, the hardware is an SMA device, and the loop is a virtual mass-spring-damper system or an explicit load/position profile. The design requirements for the machine were derived from the range of expected properties in NiTi wire devices. The machine can perform accurate characterization and control tests, and allows rapid fitting of SMA models and validation of control schema. In the other parts of this thesis, all experimental data for modelling or control was collected using this HIL system, unless specified otherwise.

Chapter 9

Linear Parameter Varying Modelling & Estimation

Shape memory alloys such as NiTi exhibit a nonlinear and hysteretic relationship between their input heating power, and their motion. The hysteresis and nonlinearity is caused solely by the crystal phase transformation. Because the instantaneous crystal phase fraction determines the nonlinearity, the SMA constitutive and kinetic models can be written in Linear Parameter Varying (LPV) form, with phase fraction as the varying parameter. LPV for general hysteretic systems has proven useful for improved controller design as it allows some linear controls design to be applied to the nonlinear system under specific conditions. No existing literature demonstrates the creation of a complete LPV model for SMA state estimation. The estimator was implemented with a physical system that represents many common robotic systems such as robot arms. Using only position measurement (such as the dual-measurement of Chapter 7) the estimator can predict states such as temperature and phase fraction which cannot be measured – the performance agrees well with simulation. Estimators designed using the LPV method will enable controllers (Chapter 10) that take full advantage of SMA actuator performance.

Introduction

Analysis tools for linear systems provide a wide range of robust solutions to control and estimation problems [205, 206, 207]. Many nonlinear control and estimation techniques attempt to either frame the nonlinear problem in a format where the linear tools apply or to invert nonlinearities with the control such that the dynamics appear linear [208]. This chapter focuses on the former method. Because the SMA material model has quickly varying nonlinearities and hysteresis, the most common methods like Taylor-series linearization are not relevant; the method must constantly adapt to the varying nonlinearities. The power of Taylor-series linearization is that it reveals the eigenvalues of the system, knowing the eigenvalues leads to selection of control laws and easy tuning of their transient and frequency response characteristics. Here, we frame the SMA control problem as a Linear Parameter Varying (LPV) state-space [209]. For many nonlinear systems this is not possible, but the form of the SMA 'constitutive' equations works well in this format. The method has proven useful for general

hysteretic systems [210]. In LPV system analysis, the system is written, with no simplifying assumptions, into a state-space format where the coefficients in the state matrices can vary but the state contains the system states directly. The system is said to be 'parameter varying' because the state matrices vary as the state variables themselves change (as opposed to 'time varying' where the evolution in time is independent of the system motion). In practice, the range of eigenvalues which the state space assumes as the parameters vary will be used to design controllers. This does not guarantee stability for all systems, but in practice it is very effective. The average or typical eigenvalues can be used to design a decent linear feedback controller or estimator, or gain-scheduling techniques can be used to adapt the control gains as the state matrices vary. For the SMA system, the ability to use this model for estimation is greatly beneficial since temperature is the major driver of SMA motion and is difficult to measure, and phase fraction cannot be measured in a live device in any practical manner. SMA actuators has never been framed as an LPV system with the model presented in Chapters 3 & 6, nor used for estimation of phase fraction (c.f. [211]), though gain-scheduling controllers have been designed with very simplified models [149].

This chapter reviews the SMA model developed in earlier chapters and removes some components that do not have a large influence on the control behavior, for simplicity. In particular, super-elastic effects where stress changes the phase fraction are ignored. Two equivalent LPV models are presented and compared – one uses stress as a state variable while the other uses crystal phase fraction. The range of instantaneous eigenvalues this model exhibits as the parameters vary is shown. Essentially, the SMA actuator controlling position of a Spring-Mass-Damper (SMD) plant provides a stable, fourth order system with one zero. State estimators are designed using the average and varying model and shown to track simulated states very well. The fourth order LPV model is not useful for controller design because it is not fully controllable – phase fraction and temperature have an instant or 'kinematic' relationship (this fact produces the open-loop system zero). Techniques in Chapter 10 consider a reduced-order system for controller design.

Model Review & Simplification

The following is a review of the final models developed in Chapters 3 and 6, with a spring-mass-damper plant under control by SMA wire. Previous sketch of the SMA behavior arises from

material behavior that can be characterized by three equations. Primarily, the strain has constitutive components due to temperature, stress, and phase fraction.

$$\epsilon = \frac{\sigma}{E} + \epsilon_L \xi + \Theta T$$

Where E is the elastic modulus, ϵ_L is the maximum shape memory strain, and Θ is the coefficient of thermal expansion. Because the thermal expansion is small compared to the other strain components, it is usually ignored. For controller design, the elastic modulus is considered to be constant. It is valuable now to rearrange such that stress rate is the dependent variable, the first important constitutive equation:

$$\dot{\sigma} = E\dot{\epsilon} + E\epsilon_L \dot{\xi}$$

(45)

The second equation governs the hysteresis of the crystal phase fraction. It is useful to characterize the transformation with a population growth equation to represent the propagation of the crystal transformation through the material. We chose the piecewise logistic function which accurately models the hysteresis, with enhancements that ensure the model is continuous. During simulation the Martensite phase fraction hysteresis behavior can be represented using conditions for transformation derived from the phase diagram:

$$\boldsymbol{if}\ \left(\dot{\mathbf{T}} - \frac{\dot{\boldsymbol{\sigma}}}{\boldsymbol{\alpha}} > \mathbf{0}\right) \qquad \boldsymbol{else\ if}\ \left(\dot{\mathbf{T}} - \frac{\dot{\boldsymbol{\sigma}}}{\boldsymbol{\alpha}} < \mathbf{0}\right)$$

$$\xi = \frac{\xi_M}{1 + \exp\left(k\left(T - \frac{\sigma}{\alpha} - A\right)\right)} \qquad \xi = \frac{1 - \xi_A}{1 + \exp\left(-k\left(T - \frac{\sigma}{\alpha} - M\right)\right)} + \xi_A$$

$$\xi_A = (1 - \xi)(1 + \exp\left(-k\left(T - \frac{\sigma}{\alpha} - M\right)\right) \qquad \xi_M = \xi(1 + \exp\left(k\left(T - \frac{\sigma}{\alpha} - A\right)\right)$$

$$\frac{\partial \xi}{\partial T} = k\left(\frac{\xi^2}{\xi_M} - \xi\right) \qquad \frac{\partial \xi}{\partial T} = k\left(\frac{(1 - \xi)^2}{\xi_M} - (1 - \xi)\right)$$

$$\frac{\partial \xi}{\partial \sigma} = -\frac{1}{\alpha}\frac{\partial \xi}{\partial T} \qquad \frac{\partial \xi}{\partial \sigma} = -\frac{1}{\alpha}\frac{\partial \xi}{\partial T}$$

(46)

And the time rate of change of ξ is computed using the partial derivatives:

$$\dot{\xi} = \frac{\partial \xi}{\partial T}\dot{T} + \frac{\partial \xi}{\partial \sigma}\dot{\sigma}$$

(47)

Where A (M) is the temperature for the center of Austenite (Martensite) transformation with no stress, $1/\alpha$ is the slope of the transformation centerline in the phase diagram, k determines the width (or distribution) of the transformation region, and ξ_M denotes the extent of the previous transformation (it drives the hysteresis). In addition to the material model, it necessary to model the energy balance between electrical heating power ($P = VI$), wire temperature, and ambient air temperature:

$$mc_p\dot{T} = P + m\Delta H\dot{\xi} + AL\sigma\epsilon_L\dot{\xi} - hA_s(T - T_\infty)$$

(48)

Where m is the SMA mass, c_p the specific heat, h is the heat transfer coefficient, ΔH is the latent heat of transformation, and A_s is the wire surface area. Many SMA materials have a response that is nearly independent of stress. Also, in many applications the stress of preloading is large compared to dynamic load variations. In these cases, the material model is simplified, when designing controllers, to assume stress cannot affect the phase fraction (there is no super-elastic behavior), resulting in the following simplification to equation (47):

$$\dot{\xi} = \frac{\partial \xi}{\partial T}\dot{T}$$

(49)

This reduces the complexity of (48) greatly:

$$\left(mc_p - m\Delta H\frac{\partial \xi}{\partial T}\right)\dot{T} = P - hA_s(T - T_\infty)$$

(50)

Notice that the effect of transformation effectively changes the thermal mass in an otherwise linear ODE. Equations (45),(46), (49), and (50) provide a complete thermo-mechanical model for controls design and analysis.

Why is it acceptable to ignore the effect of stress on phase fraction? When SMA wire is used in the super-elastic condition, it is an important factor. However, when an SMA actuator is used to control a reasonable payload, the temperature is the primary driver of phase change. This can

be seen in a simple demonstration where a constant power is applied to heat an actuator that is driving a spring-mass-damper (Figure 89). It was shown in (47) that phase change rate is comprised of stress and temperature components. In a simulation, the phase change rate is tracked as well as these two components. It is clear that the temperature component (red) has a much larger effect than stress (blue); the total phase change (black) is almost equal to the temperature component alone.

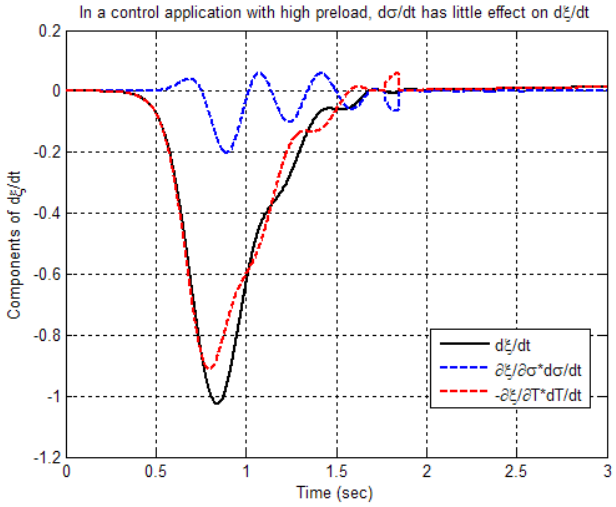


Figure 89 – The phase change rate (black) is comprised of stress (blue) and temperature (red) components. Temperature is the primary driver of phase change for actuators, not stress. This data is from the same simulation shown in Figure 97 later in this chapter.

Actuating a Mechanical System

Consider an SMA wire controlling the position of a linear spring-mass-damper system with external preload or disturbance force. This general model can represent many physical systems such as robot arm joints, linear positioning systems, etc. Based on the simplified equations just reviewed, the convention and physical view of the SMA actuator control system is shown in Figure 90. This model has an SMA wire connected between a fixed wall and the spring-mass-damper plant. The actuation arises from change in the internal state of the material. However, the change of internal material states is not intuitive. For clarity, the system can be reformed into an equivalent mechanical system as in Figure 91. This model breaks the single SMA material with internal changes into two parts; the inelastic ‘shape memory strain’, and an elastic spring. This view of the system makes it clear where the nonlinearity occurs (between temperature change and crystal phase fraction), and eliminates the need to keep stress as a state variable.

The spring-mass-damper has equation of motion:

$$M\ddot{x} + B\dot{x} + Kx = -F_{SMA} + F_{preload} + F_{ext}$$

(51)

Where x is the position, M is the plant mass, B its damping, and K its spring rate. The force of the SMA actuator (F_{SMA}) has a negative contribution, as the coordinates in the figure are labelled, since tensile stress is positive. To prevent Martensite twinning (see Chapters 2&3), a minimal external preload ($F_{preload}$) on the SMA wire is required. Other external forces might arise from external payloads or disturbances (F_{ext}). When considering the material behavior, stress and strain are more convenient variables than force and position. They can be easily transformed using

$$F = A\sigma$$

$$x = L\epsilon$$

The SMA material has an elastic modulus E which results in an equivalent spring rate K_{SMA} of the entire wire:

$$K_{SMA} = AE/L$$

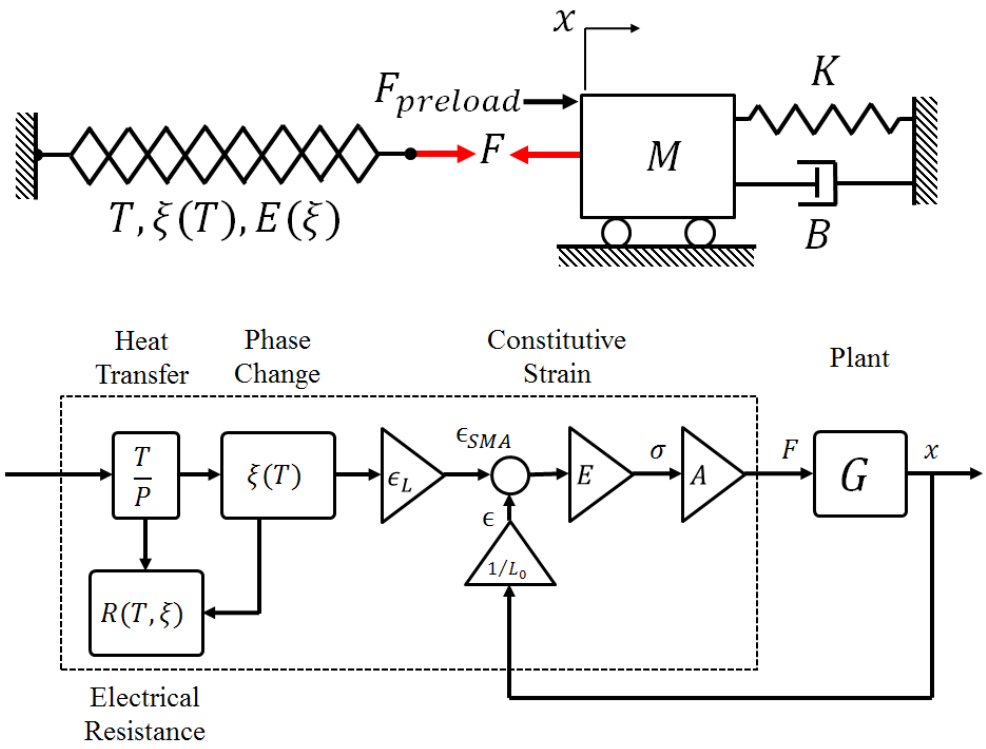


Figure 90 - A representation of the SMA material controlling an SMD system which isolates the SMA device entirely from the plant.

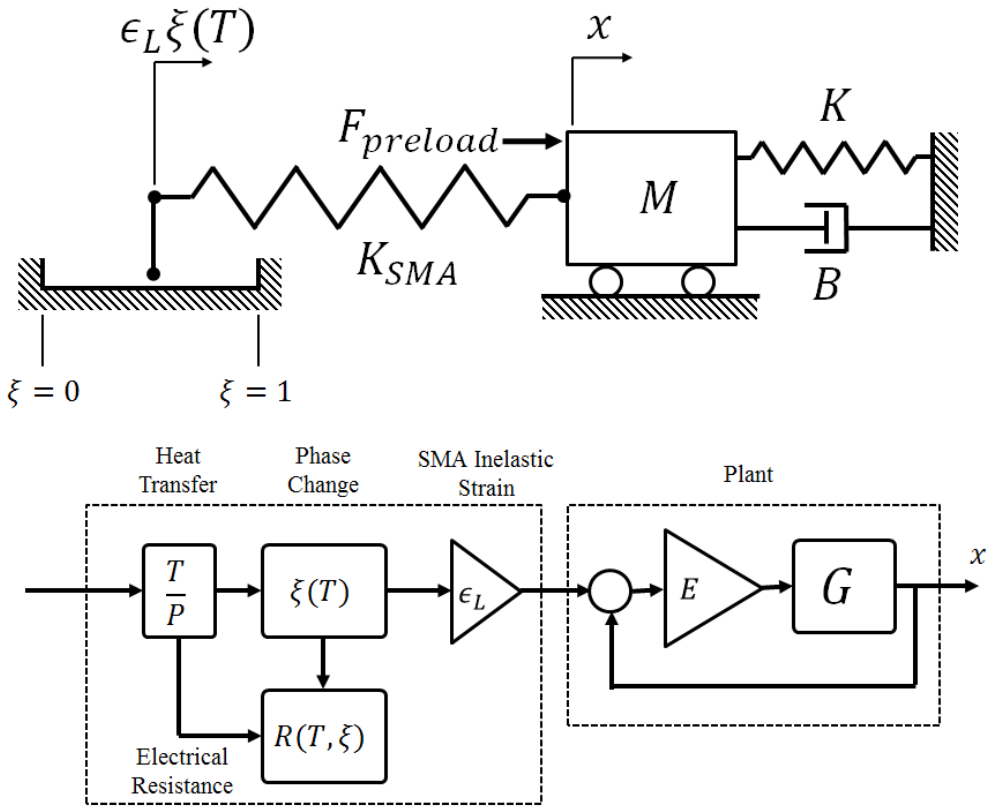


Figure 91 - An equivalent SMA system illustrates the behavior in a useful manner for controller design

Linear Parameter Varying Model

The model presented above can framed in a LPV state space in two different ways; either with stress or phase fraction as a state variable. As represented in Figure 90, the material perspective considers the stress as a result of material behavior. This completely isolates the SMA actuator equations from the plant creating an actuating force ($F = A\sigma$) that pushes on the plant to be controlled.

$$\begin{Bmatrix} \dot{x} \\ \ddot{x} \\ \dot{\sigma} \\ \dot{T} \end{Bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -K/M & -B/M & -\mathrm{A/M} & 0 \\ 0 & E/L & 0 & E\epsilon_L hA_s/Q \dfrac{\partial \xi}{\partial T} \\ 0 & 0 & 0 & -hA_s/\mathrm{Q} \end{bmatrix} \begin{Bmatrix} x \\ \dot{x} \\ \sigma \\ T - T_\infty \end{Bmatrix} + \begin{bmatrix} 0 \\ 0 \\ -\dfrac{\partial \xi}{\partial T}/Q \\ \mathrm{W} \end{bmatrix} P$$

$$Q \equiv mc_p - m\Delta H \frac{\partial \xi}{\partial T}$$

(52)

The new variable Q combines all the thermal energy storage mechanisms (sensible heat, latent heat, and plastic deformation) into a single parameter that acts like a varying 'thermal mass'. Note that this thermal mass can only be larger than the common thermal mass (mc_p) since $\partial\xi/\partial T$ is always negative. An alternate view uses the phase fraction as the state as shown in Figure 91. This equivalent model eliminates stress as a state variable by considering the SMA material as an elastic spring (K_{SMA}) being pushed by the shape-memory strain ($e_L\xi$). This reframing is compatible with (45). Consider the state space representation of Figure 91:

$$\begin{Bmatrix} \dot{x} \\ \ddot{x} \\ \dot{\xi} \\ \dot{T} \end{Bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -(K+K_{SMA})/M & -B/M & \mathrm{K_{SMA}}L\epsilon_{\mathrm{L}}/\mathrm{M} & 0 \\ 0 & 0 & 0 & -hA_s\dfrac{\partial\xi}{\partial T}/Q \\ 0 & 0 & 0 & -hA_s/\mathrm{Q} \end{bmatrix} \begin{Bmatrix} x \\ \dot{x} \\ \xi \\ T-T_\infty \end{Bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \dfrac{\partial\xi}{\partial T}/Q \\ 1/\mathrm{Q} \end{bmatrix} P \qquad (53)$$

The model uses the same 'varying thermal mass' Q as in (52). Usually this model is preferable since it is more directly clear how temperature drives phase change, and phase change subsequently acts to move the plant. The stress can be recovered as a 'measurement' by using a new output matrix, if desired, for state estimation or other purposes.

$$y = \begin{Bmatrix} \epsilon \\ \dot{\epsilon} \\ \xi \\ T-T_\infty \\ \sigma \end{Bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -E & 0 & E\epsilon_L & 0 \end{bmatrix} \begin{Bmatrix} \epsilon \\ \dot{\epsilon} \\ \xi \\ T-T_\infty \end{Bmatrix} \qquad (54)$$

As equivalent systems, these two models have identical eigenvalues; the following discussion focuses on the model with phase fraction ξ as a state. The only varying parameter is the partial derivative of phase fraction with respect to temperature. This parameter takes on a value in the range

$$-\frac{k}{4} < \frac{\partial\xi}{\partial T} < 0$$

As seen in Chapter 3, this parameter has a minimum in the center of phase change when a small change in temperature greatly affects the phase fraction. The parameter is near zero when the temperature is far from the transformation range, and when the hysteretic phase change direction has recently reversed. The variation changes both the gain and the first order eigenvalue of the response from power to temperature and from power to phase fraction (Figure 92). The resulting

response has two different effects when near and far from the peak transformation region.

When $\partial\xi/\partial T$ is small (far from transformation region, or immediately after phase reversal) the gain from power to temperature is very high, but the gain from power to phase change is negligible. A small change in power will affect the temperature greatly since only mc_p stores thermal energy and not latent heat. A small change in power will not affect the phase fraction because of hysteresis or because temperature is outside the transformation region. Because the thermal mass Q is lowest when far from the transformation region, the homogeneous decay of temperature is faster when far from the transformation region. However, phase is most 'stable' when far from the transformation region, so the pole is smallest in this condition.

When $\partial\xi/\partial T$ is large (in the middle region of transformation) the gain from power to temperature is low because the effective thermal mass Q is at a maximum. However, the gain from power to phase fraction is very high because a small change in temperature has a relatively large effect on the phase fraction. For the same reason, the homogeneous decay of temperature is slow when the effective thermal mass is highest in the center of transformation. However, a small change in temperature affects the phase change greatly in this condition, so the homogeneous response of phase fraction is highest here.

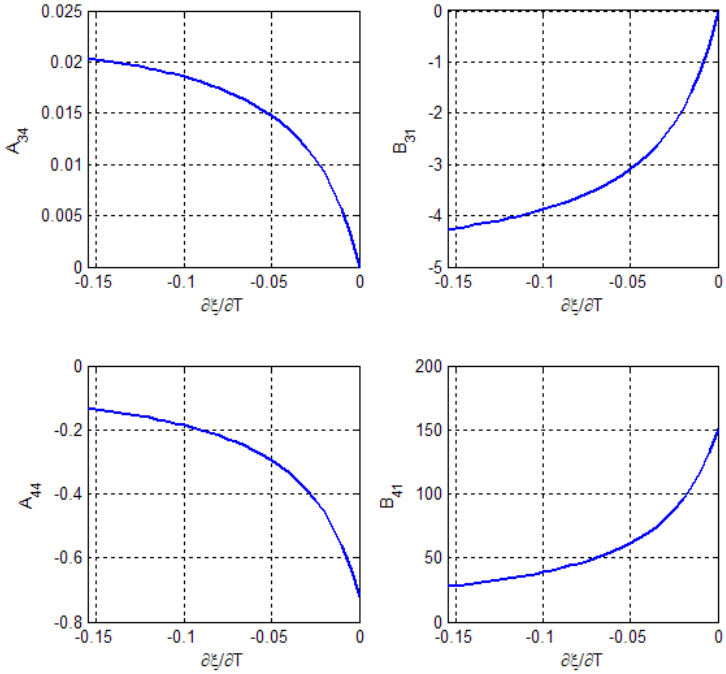


Figure 92 - Varying Real pole and Gain in the model with phase fraction (ξ) as a state variable. The variables are labelled according to their location in the state-space model (53)

Consider a 'snap-shot' of these models where $\partial\xi/\partial T$ is frozen (though not acceptable in general, this is fair since the nonlinearity is first order with negative definite eigenvalue [208]). Both models are observable, but not controllable. The lack of controllability can be seen in two ways. Consider (54) where ξ is a state: Mathematically, the third and fourth rows are clearly not linearly independent. This leads to a controllability matrix with rank 3 – proving the system is not controllable for any value of $\partial\xi/\partial T$. Practically, it is easy to see that we cannot drive the system to have a unique temperature and unique phase fraction simultaneously because temperature and phase fraction are constitutively coupled.

State Estimator with Gain-Scheduling

The LPV model presented was used to create a state estimator[3]. This estimator has gains designed based on pole placement, with the estimator time constants selected to be more than ten times faster than the closed loop system behavior, so that the estimator does not interfere with the control transient response. The estimator is implemented using:

$$\dot{\hat{x}} = \bar{A}\hat{x} + \bar{B}u + L(y - \hat{y}), \quad \hat{y} = \bar{C}\hat{x}$$

$$\dot{\hat{x}} = (\bar{A} - L\bar{C})\hat{x} + \bar{B}u + Ly$$

If the assumed model $(\bar{A}, \bar{B}, \bar{C})$ and the exact system (A, B, C) coincide, then the dynamics of the error are determined by:

$$e \equiv x - \hat{x}$$
$$\dot{e} = \dot{x} - \dot{\hat{x}}$$

$$\dot{e} = (A - LC)e$$

Consider an example SMA wire actuator of 0.125mm diameter. The natural pole of the temperature response is $\lambda_T = -0.7228$ (1/s). The estimator is designed using poles that are faster than ten times this: $Real(\lambda_{est}) < -7.228$. In the simulations presented below, estimator closed loop poles were chosen to be negative real, all faster than this minimum speed.

In practice, the estimator gain is too complex to recompute (on a microcontroller) after each sample while the controller is running. However, the state transition matrix A and input gain B

[3] A review of linear system concepts, including state estimators, is provided in the Appendix

can be easily updated if $\partial\xi/\partial T$ is known or is estimated. There are three ways to address this: by using a constant matrix that represents the average properties, by switching between a few controllers based on the region of operation, or by continually adjusting the state model as the system varies. The first method is only reliable for the states that are not derivatives of nonlinear terms – so it can track the position, speed, and phase fraction accurately, but not temperature. However, using the gain scheduled estimator, the temperature and all other states are accurately recovered after initial offsets are removed by the estimator. A simple (poorly tuned) PI controller is used to demonstrate both estimators in simulation. The complete non-linear thermo-mechanical models of Chapters 3 & 6 were used as the physical system, and the LPV model developed in this chapter was used for the estimator design. Figure 93 demonstrates the estimator using average properties, and Figure 94 demonstrates the gain-scheduled estimator. It is clear that both estimators accurately track the position and phase fraction. Both struggle to track temperature when far from the transformation region (where ξ is changing and $\partial\xi/\partial T$ is near zero). The average constant-coefficient model cannot accurately determine temperature. The gain scheduled controller can accurately track temperature once initial errors are eliminated, as expected.

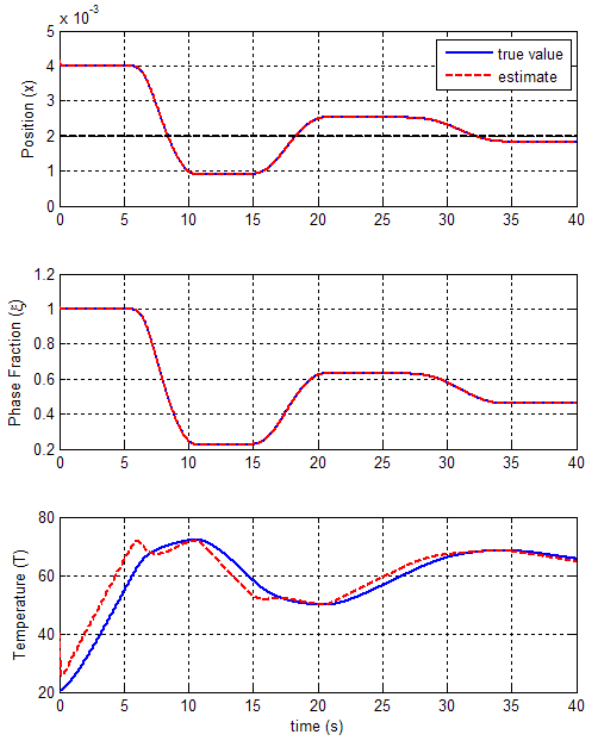


Figure 93 - Estimator using average properties accurately estimates phase fraction but does not accurately track the temperature due to ignored nonlinearities

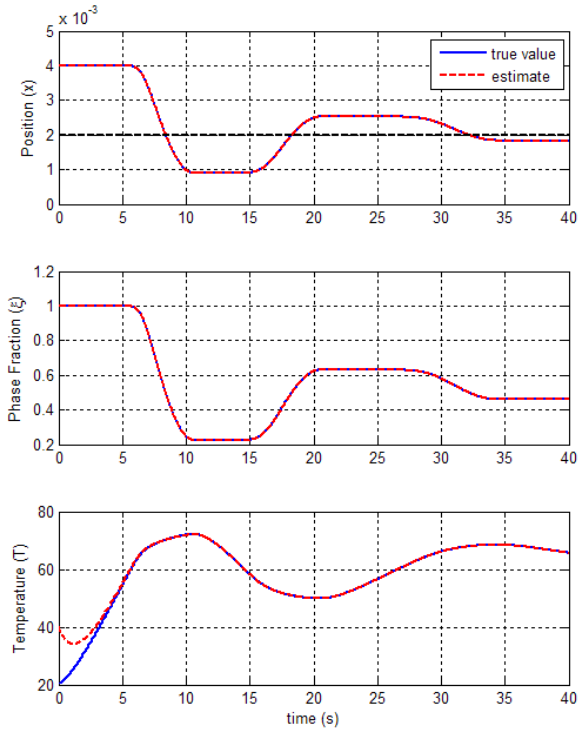


Figure 94 - Estimator using gain-scheduling still struggles when far from the transformation region, but quickly tracks all state variable during transients

Experiment

The linear state estimator based on (53) was implemented with a physical spring-mass-damper shown in Figure 95. The wire is attached to a lever on a shaft – the shaft has a rotary encoder and a large flywheel for a mass. An antagonist spring provides preload and a small spring force. The specifications are provided in Table 10. This system provides a good trial of the state estimator because the rotary encoder cannot be numerically differentiated to determine speed and acceleration without excessive filtering, but the estimator can provide smooth estimates. The physical properties of the mass and spring were confirmed by system identification of the open-loop system with a step input. This experiment confirms that even the constant-property estimator can provide acceptable estimates of states which cannot be physically measured. The estimator poles (of $A - LC$) were placed to provide fast settling but while allowing some overshoot.

Table 10 - Experimental device parameters

	Value	**Units**
Mass	0.13	kg
Spring Rate	18	N/m

Damping	0.915	Ns/m
SMA Diameter	0.1	mm
SMA Length	233	mm

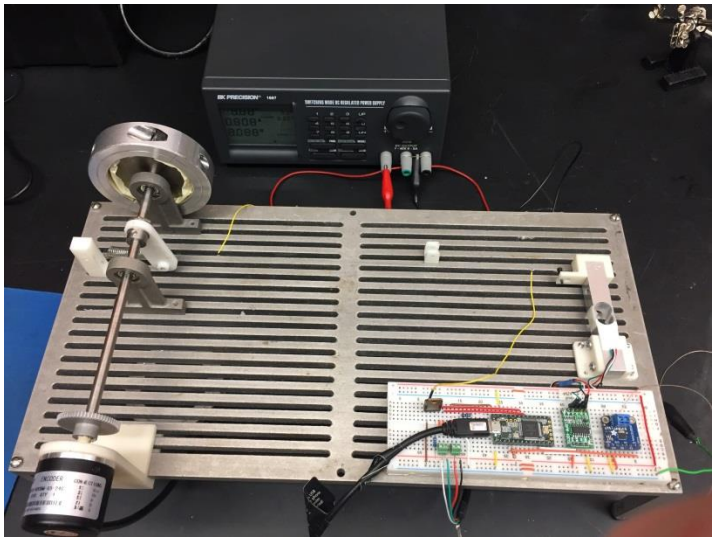


Figure 95 - Experimental spring-mass-damper testing platform. The fine wire runs from the white rocker on the left, to the white clip on the load cell at right.

First, simple PI controller was used to regulate to a single set-point from initial conditions with the wire relaxed at room temperature (Figure 96). After a short transient, the estimator quickly filters the position measurement and provides estimates of phase fraction and temperature. In practice, the temperature is impractical to measure and phase fraction is impossible to measure – so the estimates can only be compared to known acceptable values. After the initial transient, phase fraction is in the expected range (0.5 to 1.0) and temperature matches the transformation region of the wire. The system was simulated with the same parameters as in the experiment (Figure 97). The transients are response are closely replicated as the estimator begins to track after the first half second. The simulation overshoots the setpoint where the phsycal system appears overdamped; this error is likely due to the modelling simplification to assume the SMA wire as a spring rate but no damping (i.e. ignoring the super-elastic effect).

Another interesting aspect of the estimator is the poor tracking during transients. Consider the experimental estimator during a disturbance caused by blowing air over the wire (Figure 98). The increased air current cools the wire, so the PI controller reacts by heating the wire. Of course, the estimator does not know about the air-flow which effectively increases heat transfer

(in the model, constant h). Because the controller responds to the disturbance by increasing applied power, the estimator predicts that the temperature rises instead of falling. This demonstrates one of the limitations of estimators in general. In a more complex implementation, the h parameter might be estimated with a nonlinear estimator such as an extended Kalman filter (if the system were still observable).

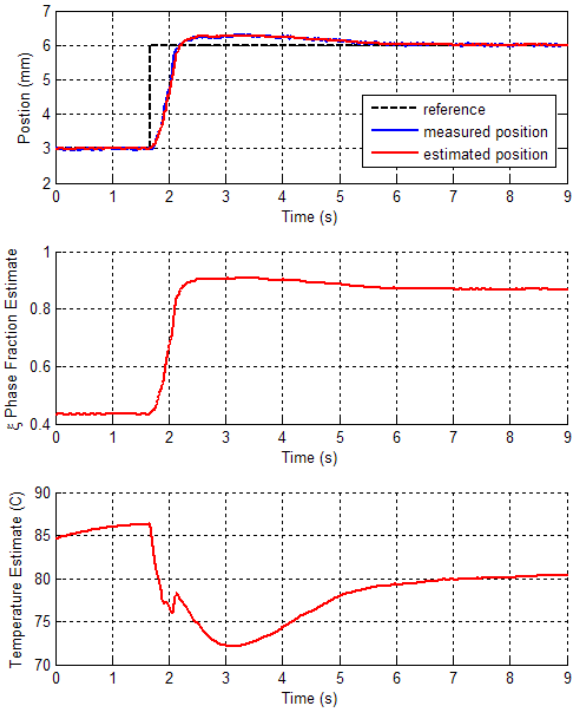


Figure 96 - Experimental results of the estimator. After initial transients (before 0.4 seconds) the estimator provides reasonable estimates of phase fraction and temperate. Steady state values are close to predicted values.

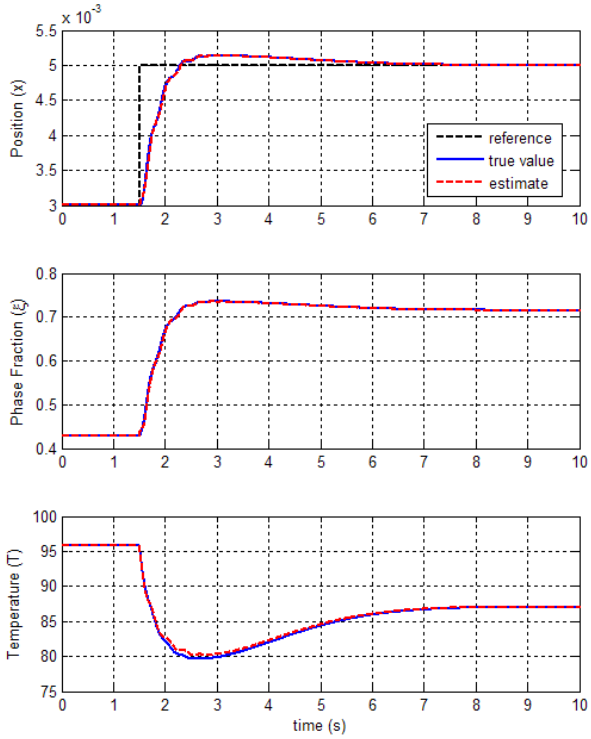


Figure 97 - Simulation with parameters matched to the experiment, shows reliable tracking once transformation begins (near 0.4 seconds).

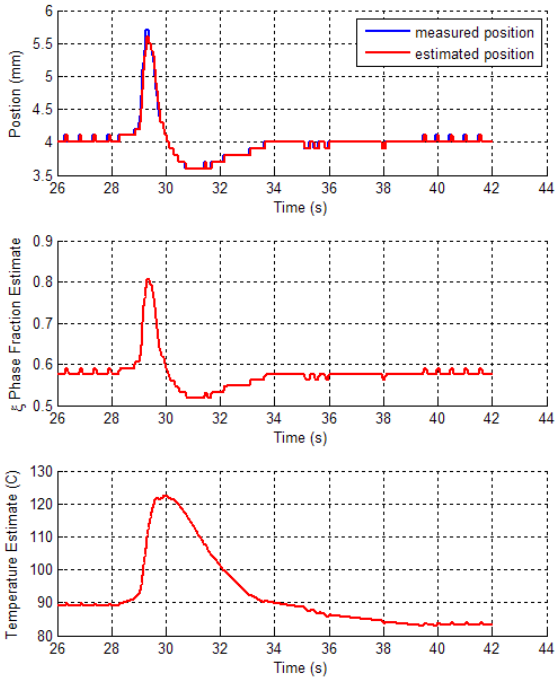


Figure 98 - Experimental response to a disturbance caused by blowing on the wire to cool it. The disturbance is not known to the estimator, so the temperature estimate rises since power is increased by the controller.

Conclusion

A LPV perspective of the SMA actuator system was developed from the complete non-linear SMA actuator model driving a spring-mass-damper plant. The model made the simplifying assumption that phase fraction change is driven primarily by temperature (shape-memory) and not stress (super-elasticity) in practical applications. This assumption was used to create the LPV state-space model, which is shown to be observable but not controllable. Two estimators were developed – one uses average properties in a constant state-space model, while the other allows the model to change as the parameters vary – a form of gain-scheduling applied to estimation. The gain-scheduled estimator is far superior, but it relies on an accurate model where $\partial\xi/\partial T$ can be determined. In practice, it is not possible to validate that the phase fraction and temperature were accurately estimated except by using their estimates in a feedback control scenario. However, it was shown in experiment that even the simple estimator can provide reasonable and fast estimates of the material phase fraction and temperature. The temperature estimate was easily disrupted by changes in ambient conditions. Future work should include estimation of these critical parameters ($\partial\xi/\partial T$ and h), perhaps requiring additional measurements (perhaps of electrical resistance) to ensure the system is observable.

Chapter 10
Linearizing Controllers

Nonlinear controllers can greatly increase the speed and accuracy of SMA actuator control. By creating new states which invert or offset nonlinearities, linear control tools can be applied to the nonlinear system. Two similar methods are employed here. First, Input-Output linearization is used to design a classical compensator that cancels the major nonlinearities. It was seen that biasing the controller to remove interaction with the heat transfer is the most significant improvement to the control law – cancellation of the nonlinear hysteresis played a smaller role in improving the closed loop response. Second, linear control techniques are applied to a state space model created using the Exact Velocity Linearization method. It is show that the EVL method is a convenient method for creating state-space equivalents to Input-Output linearizers. Because it is a state-space model, EVL has the added benefit that the controller can be designed using pole placement techniques. The EVL model was used to a robust, output-feedback servo controller using the estimated state variables as demonstrated previously.

Introduction

Analysis tools for linear systems provide a wide range of robust solutions to control and estimation problems [205, 206, 207]. Many nonlinear control and estimation techniques attempt to either frame the nonlinear problem in a format where the linear tools apply or to invert nonlinearities with the control such that the dynamics appear linear [208]. This chapter focuses on the latter method. As shown in Chapter 9, the nonlinearity is entirely contained in the hysteretic and nonlinear relationship between temperature and crystal phase fraction. The same model presented there is used here, with the validated assumption that temperature (the shape memory effect) has a much larger influence on phase change than stress (super-elastic effect). Both linear and nonlinear control methods for SMA actuators have been studied extensively in the literature [141, 142]. Unfortunately, most of the controllers seen in the literature have no basis in a model of the material; they consider the principles characteristics of the behavior (hysteresis, etc.), though a few draw meaningful conclusions from the model that improve control performance or accuracy (c.f. [212, 213, 149, 211]). Two modelling approaches, not

found in the SMA control literature, are employed here: Input-Output linearization [208] and Exact Velocity linearization [214, 215, 216]. The actuator is employed to control the spring-mass-damper system shown in Figure 99.

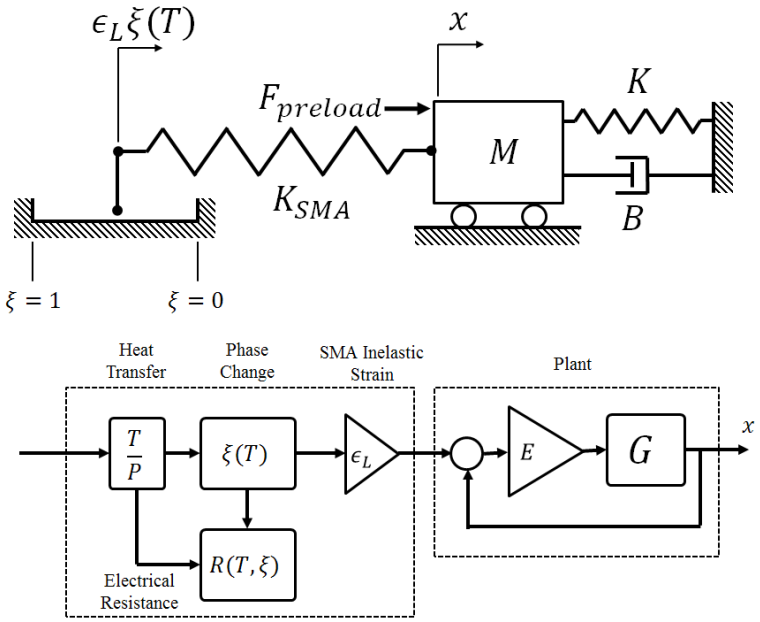


Figure 99 - The equivalent SMA actuator system controlling position of a spring-mass-damper plant

Input-Output Feedback Linearization

In many cases, system nonlinearities can be 'cancelled' between the input and output by using a control law which inverts the nonlinear components of the system behavior. The controller which does this is said to linearize the system from input to output since an input reference will be tracked with apparently linear response due to the controller inverting or offsetting nonlinearities. One reliable way to implement this concept is appropriately called the Input Output Linearization (IOL) technique [208]. The technique involves three steps for designing the controller:

1. Differentiate the system output (position measurement) until the input appears
2. Design a controller that cancels nonlinearities and introduces desired linear dynamics
3. Determine whether any hidden or 'internal' dynamics are unstable

Applying these steps will yield a model that mathematically achieves desired dynamics where the response appears linear – in practice this often requires full state measurement (or estimate), in this case even knowledge of the hidden state. The first I-O linearization step is performed as follows for the SMA system. Starting with position measurement y:

$$y = x$$
$$\dot{y} = \dot{x}$$
$$\ddot{y} = \ddot{x} = -\frac{K + K_{SMA}}{M}x - \frac{B}{M}\dot{x} + \frac{K_{SMA}L\epsilon_L}{M}\xi$$
$$\dddot{y} = \dddot{x} = -\frac{K + K_{SMA}}{M}\dot{x} - \frac{B}{M}\ddot{x} + \frac{K_{SMA}L\epsilon_L}{M}\dot{\xi}$$

(55)

The important revelation here is to rewrite the time derivative of ξ, using the chain rule of differentiation, as:

$$\dot{\xi} = \frac{\partial \xi}{\partial T}\frac{dT}{dt} = \frac{\partial \xi}{\partial T}\dot{T} = \frac{\partial \xi}{\partial T}\frac{1}{Q}(P - hA_s(T - T_\infty))$$
$$Q \equiv mc_p - m\Delta H\frac{\partial \xi}{\partial T}$$

(56)

There is a 'kinematic' relationship between ξ and $\frac{\partial \xi}{\partial T}$, as shown in a previous chapter, though it is difficult to estimate this quantity. Combining (55) and (56) yields:

$$\dddot{y} = -\frac{K + K_{SMA}}{M}\dot{x} - \frac{B}{M}\ddot{x} + \frac{K_{SMA}L\epsilon_L}{M}\frac{\partial \xi}{\partial T}\frac{1}{Q}(P - hA(T - T_\infty))$$

(57)

The controller that cancels the nonlinearity and hidden dynamics is comprised of a nonlinear inverting element α and offsetting element β:

$$P = \alpha u + \beta = \left(\frac{1}{\frac{\partial \xi}{\partial T}}\frac{K_{SMA} + K}{K_{SMA}e_L}Q\right)u + hA(T - T_\infty)$$

(58)

Which yields third order linearized dynamics from the new input u to output y. A useful perspective of this model is to consider that the control law schedules a gain (α) and an offset (β) to the control effort based on measured conditions. The accuracy of this method will be dependent on the ability to accurately cancel the nonlinearities and hidden dynamics – essentially only limited by the model and measurement accuracy.

A compensator can now be designed using classical control methods. Assuming the

linearization technique can be applied perfectly, the apparent transfer function from input u to output y is:

$$\frac{Y}{U} = \frac{K + K_{SMA}}{ms^3 + Bs^2 + (K + K_{SMA})s} = \left(\frac{1}{s}\right)\left(\frac{K + K_{SMA}}{ms^2 + bs + (K + K_{SMA})}\right) \quad (59)$$

If the desired control capability is step response with no overshoot, a simple proportional controller will suffice for demonstration ($u = -x$). Obviously the inversion was already chosen such that this system has a stable step response. As expected, the IOL method has eliminated the first order delay from temperature to phase change, resulting in an integrator in the transfer function.

Finally, to complete the I-O modelling process, the stability of the internal or 'zero' dynamics of this system should be considered. The concern is that our offset (β) could be mathematically cancelling a function which is not bounded, and so in practice the controller would break. This is not a concern with our system since the only hidden state is temperature, which we know has a first order, stable response. For completeness, this internal state can be considered – assuming the controller drives all the states it can control to zero, what dynamics remain? For our system

$$\dot{T} = \frac{1}{mc_p} P_{ss} - \frac{hA}{mc_p}(T + T_\infty)$$

So regardless of the final value of P, the internal dynamics stabilize to a constant value.

Simulation

In practice, the estimation of $\partial\xi/\partial T$ will cause concerns. However, two factors move in our favor. First, the temperature can be estimated fairly accurately (Chapter 9). Since $\partial\xi/\partial T$ enters most effectively as a control gain, we can look directly to the linear transfer function root locus to determine the effect of poorly modeling that gain when implementing the controller:

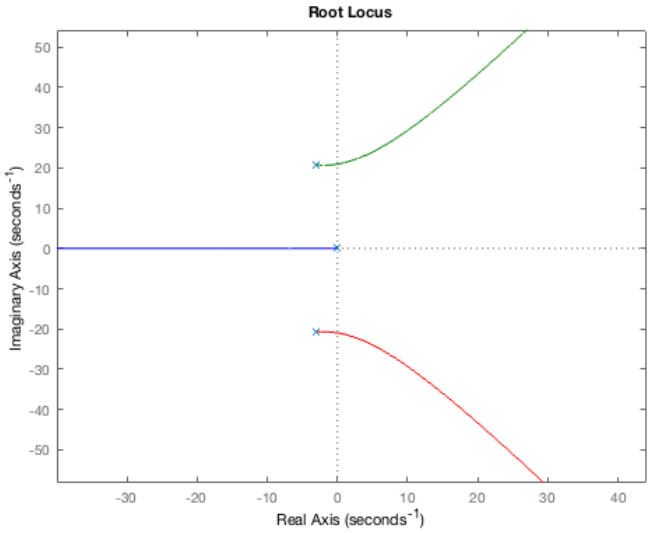


Figure 100 - Apparent linear root locus of the I-O linearized system with a proportional controller. If the nonlinearities are not well known, it has the effect of changing the closed-loop eigenvalues on this locus but does not affect the form of the locus itself

Essentially the closed-loop response is stable so long as the gain is not chosen too high. So even if there is an error, the controller will stabilize the system. Two approaches can be taken at this point; either estimate $\partial\xi/\partial T$ and continually update the controller so that it perfectly inverts the nonlinearity, or choose a constant $\partial\xi/\partial T$. Using the estimated values provides acceptable response, however the estimator introduces some oscillation since the control overshoots (Figure 101). This is because $\partial\xi/\partial T$ can become very small when far from the transformation region, leading to unreasonable control effort which disturbs the system if the estimated states are not perfectly tracked.

Fortunately the controller is successful even if the inverting term is not used. Setting $\partial\xi/\partial T$ to the constant value -0.05 provides the response shown in Figure 102. This shows that the largest contribution of the IOL technique was not the inversion of phase change, but feeding forward the temperature response lag. Input-output linearization reveals the power of 'cancelling out' the slow heat transfer.

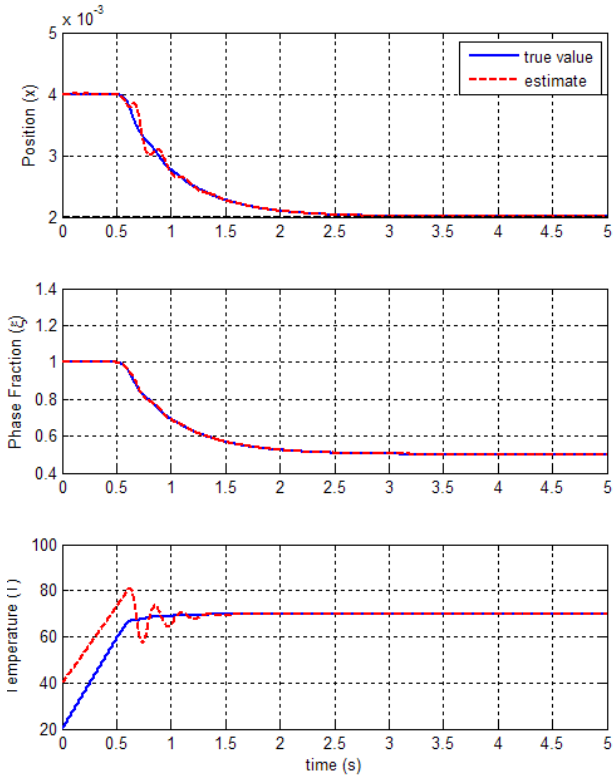


Figure 101 - IO Linearization using estimated coefficients introduces oscillation due to sensitive inversion when far from the transformation region.

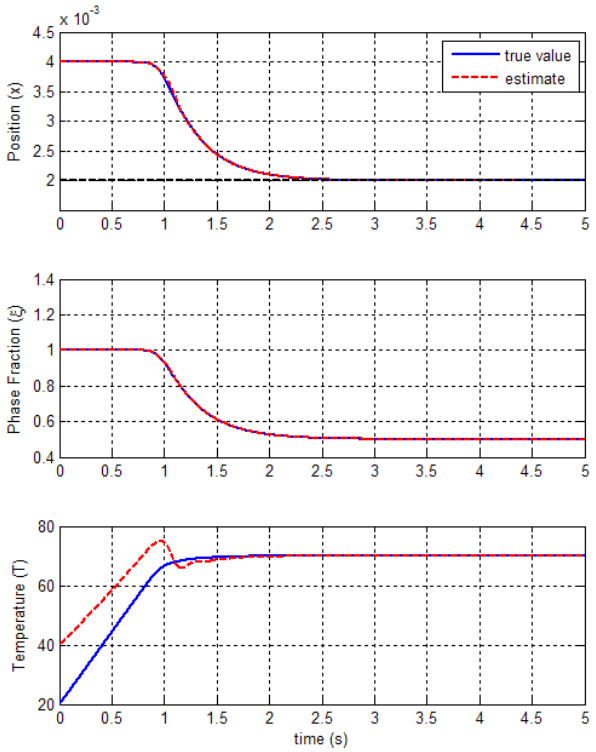


Figure 102 – I-O linearizing controller simulation using constant gains ($\partial\xi/\partial T = -.05$) acts like feedforward.

Perspective

The input-output linearization can be seen in a few lights. Most obviously, it provides an offset term that varies with the temperature state. One way to see this varying offset is that it acts to feedforward and cancel a temperature 'disturbance' in the model. Secondarily, the input-output method provides a varying gain. This can be seen as a 'scheduled' gain, which is common when treating the system as an LPV model when using state-space techniques. Unfortunately, the varying gain is very large when far from the transformation region so it affects controller stability when implemented. In the end, the offset is the more important term in providing quick stability. The ability to achieve zero steady state error (in the proportional controller) depends heavily on an accurate measurement or estimate of T, but does not require an accurate estimate of the gain.

Exact Velocity Linearization

The Exact Velocity Linearization is a method for both state-space and transfer function controller design which provides a linear system representation without introducing any simplification to the model [216, 214, 215]. The concept arises from Leith and Leithart who considered that the flaw in Taylor series expansion is that a linearization point must be chosen and adjusted to change the local linearization region.

The simplification of a nonlinear system using Taylor series assumes that only first order effects are significant in the operating region. In other words, it requires nonlinearities to be negligible during operation. Adjusting the 'operating region' by changing equilibrium points can extend the method for application to nonlinear control and estimation, but has limitations. Considering a general first order nonlinear differential equation:

$$\dot{x} = F\big(x(t), u(t)\big)$$

$$y = G(x, u)$$

Where $x(t)$ is the state vector, $u(t)$ is the input, and t is time. If this model is linearized about an equilibrium point using a Taylor series expansion, the resulting model is:

$$\dot{\delta x} = A\delta x + B\delta u$$

$$\delta y = C\delta x + D\delta u$$

Where:

$$A \equiv \nabla_x F(x,u)|_{x_0,u_0} \quad B \equiv \nabla_u F(x,u)|_{x_0,u_0} \quad C \equiv \nabla_x G(x,u)|_{x_0,u_0} \quad D \equiv \nabla_u G(x,u)|_{x_0,u_0}$$

$$\delta x \equiv x - x_0 \qquad \delta u \equiv u - u_0$$

This method essentially 'freezes' the system and assumes the state and input will stay very near the initial point such that only first order effects are significant. If the system is nonlinear compared to the range of δx and δu then the model is not accurate. In some cases, the model can be re-linearized in real time (as in Extended Kalman Filters, for instance). This leads to a large loss of information about the system behavior – unless the nonlinearities are very simple and the range of δx and δu are very small, these methods require design of new controllers for each operating point.

The Exact Velocity Linearization (EVL) method guarantees the ability to create an equivalent 'linear-parameter varying' model from any system, with no approximation error. This is achieved by replacing the differences δx and δu with true derivatives. The entire nonlinear state equations are differentiated once – the chain rule ensures this pulls a time derivative of each state out of any operator it might be hidden in. The EVL model creation process is simply:

$$\ddot{x} = A\dot{x} + B\dot{u}$$

$$\dot{y} = C\dot{x} + D\dot{u}$$

$$A \equiv \nabla_x F(x,u) \qquad B \equiv \nabla_u F(x,u) \qquad C \equiv \nabla_x G(x,u) \qquad D \equiv \nabla_u G(x,u)$$

Mathematically, this is just differentiation using the chain rule, and preserves all information about the original system besides initial conditions. However, we have the added concern that $A, B, C, \& D$ vary as the state changes – thus the resultant model is a 'Linear Parameter Varying' system. A controller can be designed based on these state matrices under different conditions, but general stability is not guaranteed even if stability is achieved for every 'snapshot' of the state. However, it has been shown that stability is guaranteed if the control law cancels (typically by inversion) any nonlinear dynamics directly such that the input-output response appears linear [216]. Because effective input-to-output linearization is our desire in most cases, EVL is an excellent tool for controller design. The second apparent problem with this method is that it uses time derivatives of all states; this causes concern since differentiation in practice introduces noise and other concerns [206]. Fortunately, it is nearly always possible to design a controller based on EVL then integrate it *analytically* such that no differentiation is required when the controller is implemented.

The EVL method is an effective tool to control the SMA system. Using the model developed in a previous chapter, the EVL model is found by simply differentiating the state equations and separating states from parameters:

$$\dddot{x} = -\frac{(K + K_{SMA})}{M}\dot{x} - \frac{M}{M}\ddot{x} + \frac{K_{SMA}L\epsilon_L}{m}\dot{\xi}$$

(60)

Not coincidentally, it was seen when examining the input-output linearization that:

$$\dddot{x} = -\frac{K + K_{SMA}}{M}\dot{x} - \frac{B}{M}\ddot{x} + \frac{K_{SMA}L\epsilon_L}{M}\frac{\partial\xi}{\partial T}\dot{T}$$

$$Q \equiv mc_p - m\Delta H\frac{\partial\xi}{\partial T}$$

(61)

Finally, we include the last state to complete the SMA model:

$$\ddot{T} = \frac{1}{Q}\left(\dot{P} - hA_s\dot{T}\right)$$

(62)

And can write the EVL state equations:

$$\begin{Bmatrix}\ddot{x}\\ \dddot{x}\\ \ddot{T}\end{Bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -\frac{K + K_{SMA}}{M} & -\frac{B}{M} & \frac{\partial\xi}{\partial T}\frac{K_{SMA}L\epsilon_L}{M} \\ 0 & 0 & -\frac{hA_s}{Q}\end{bmatrix}\begin{Bmatrix}\dot{x}\\ \ddot{x}\\ \dot{T}\end{Bmatrix} + \begin{bmatrix}0\\ 0\\ \frac{1}{Q}\end{bmatrix}\dot{P}$$

$$\dot{y} = \dot{x}$$

(63)

This model is very useful since it is already in the correct form to implement a robust servo controller (see Appendix) but, improving on the initial LPV model (Chapter 9), it is both fully controllable and observable! It has a pair of underdamped, stationary poles (due to the spring-mass-damper plant), and a single, real, stable pole that varies with $\partial\xi/\partial T$. The gain also varies.

A control law is now to be designed for this system. Either transfer function or state-space techniques can be used. So long as the control law cancels the nonlinearities, then a resulting stable linearized controller is guaranteed to be stable for the nonlinear system [216]. Here we use this model to create a robust servo control system – this is especially convenient for the EVL model since it is already in derivative form! As in a linear robust servo model, the integration is

analytical so the resultant control law is simple, although all nonlinearity in this EVL state-space must be cancelled to guarantee stability. The robust servo controller is designed by constructing a new state which includes error $(x - x_{des})$ as a state variable:

$$z \equiv \begin{Bmatrix} e \\ \dot{x} \end{Bmatrix}$$

$$\dot{z} = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -\frac{K + K_{SMA}}{M} & -\frac{B}{M} & \frac{\partial \xi}{\partial T}\frac{K_{SMA} L \epsilon_L}{M} \\ 0 & 0 & 0 & -\frac{hA_s}{Q} \end{bmatrix} z + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{Q} \end{bmatrix} \dot{P}$$

(64)

Because it is already differentiated, the EVL model allows us to quickly build a robust controller. It is clear where to cancel the nonlinearity (the gain on the third state will vary). The feedback gain matrix can be designed using any standard linear techniques (such as pole-placement) as desired. First consider the symmetric root locus of this system which will reveal the set of Linear Quadratic Regulator 'optimal' pole locations (Figure 103). The nonlinear term is set to $\partial \xi / \partial T = -.05$, a typical value, for plotting. The plot suggests that, for low gains, the first pair of closed-loop poles should be placed near the spring-mass-damper plant poles. The other two closed loop poles should be real, located between the heat transfer eigenvalue and the origin (integrator) pole that was introduced by the servo method.

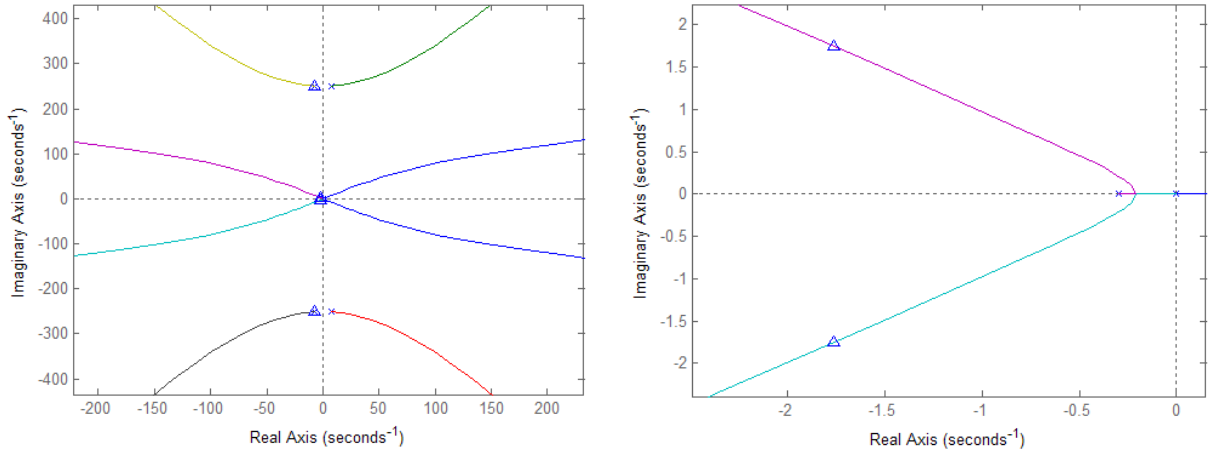


Figure 103 – Symmetric root locus with $\partial \xi / \partial T = -.05$. The selected closed loop poles (shown as triangles) were chosen with a $Q = [25e4\ 0\ \ 0\ 0]$ and $R = 1$

A controller was designed where the effort/speed compromise is based only on driving integral of error - $Q = diag(25e4\ 0\ 0\ 0)$ and $R = 1$. The wide disparity is related to the low strain of the

actuators and high stiffness of the SMA equivalent spring. The resulting controller has closed-loop poles that are underdamped as shown in Figure 103.

The controller was simulated using the same method as with the I-O Linearizer – the models from Chapters 3 and 6 represent the nonlinear plant, the estimator of Chapter 9 was employed to determine the states, and the EVL robust servo controller is employed for control. The controller is implemented by analytically integrating; the gains designed using the symmetric root locus are used, but the state feedback uses the state and the integral of error:

$$z_i = \int z \equiv \left\{ \iint \begin{matrix} e \\ x \end{matrix} \right\} \tag{65}$$

To account for the nonlinearity, the state-feedback gain matrix K must change as $\partial\xi/\partial T$ changes. The full range of variations for the four terms of K is shown in Figure 104. In the active transformation region ($\partial\xi/\partial T < -.05$) the control gains are consistent except for k_4 which enters the control effort through the temperature and is proportional to $\partial\xi/\partial T$. The controller is first implemented by recalculating gains based on the current value of the varying parameter, but holding the gain constant when it is very small so the inversion does not lead to excessive control effort. The result is shown in Figure 105. When outside the transformation region the controller is mediocre, but rapidly becomes effective once transformation starts (beyond 4 seconds).

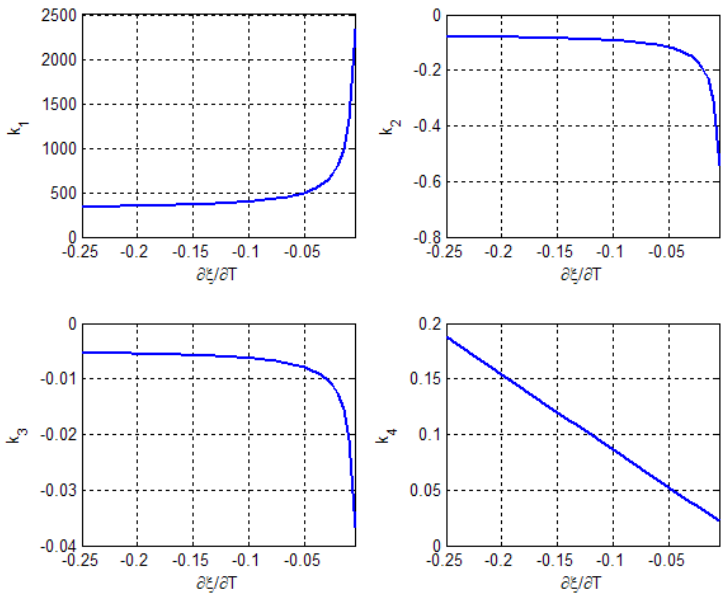


Figure 104 - The four components of the 'optimal' gain vector can vary. Most are near constant in the center of transformation except k_4

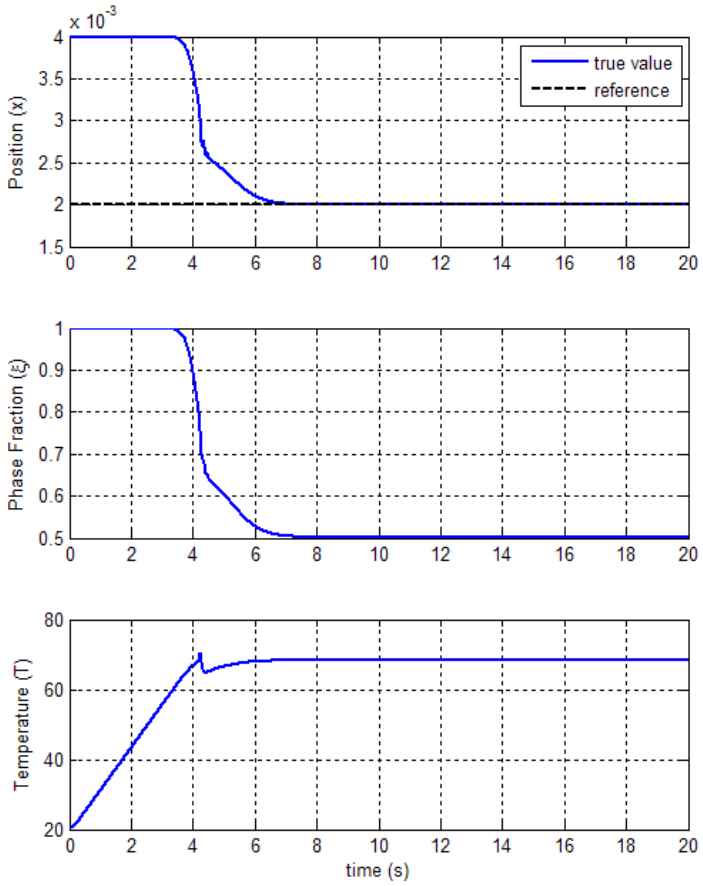


Figure 105 - Robust servo using gain-scheduling. The initial response is poor until the transformation region is reached.

Alternately, the controller can be designed using typical properties and leaving the gain K constant when the controller runs. The controller using the constant parameter model (a true linear control law) provides underdamped response with 26% overshoot (Figure 106). The controller reliably drives the steady state error to zero even with the poor modelling assumption introduced by holding the gain constant. It is not quite as fast as expected from the root locus due to the excessive overshoot.

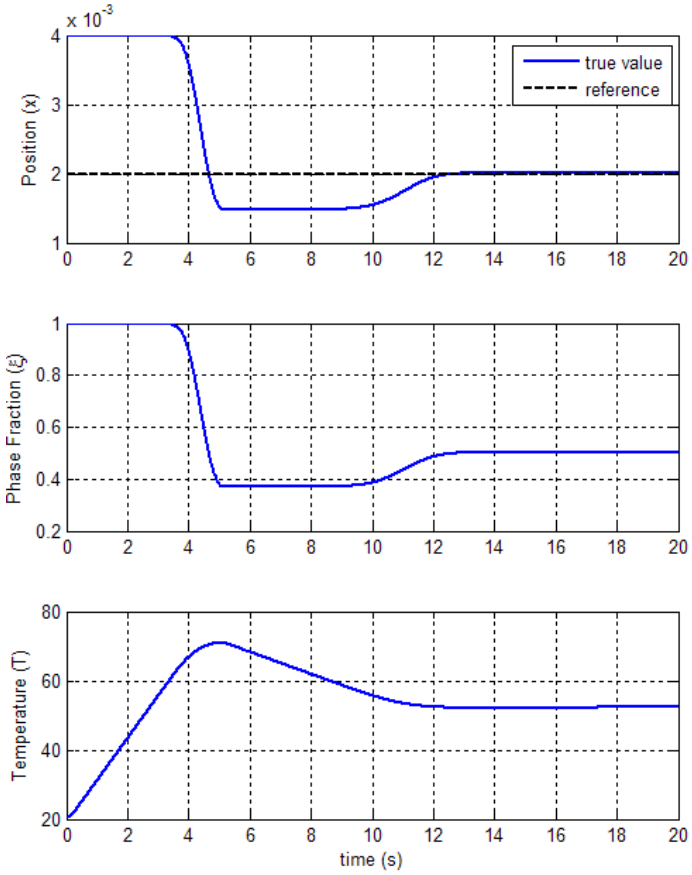


Figure 106 - The robust servo controller using a static gain, designed using the EVL servo model

Conclusion

Two control methods have been applied to the SMA actuator system. The IOL method was shown to provide quick control primarily due to feeding-forward and eliminating a lag due to temperature rise time. The EVL modelling approach was shown to easily transform nonlinear systems into reduced order LPV systems. The EVL model was used to design a robust servo controller. The robust servo controller was not found to be reliable as it depends too heavily on quickly varying states which are known to be difficult to estimate. The two methods do have some similarities. Consider the transfer function that can be constructed from the EVL state-space:

$$\frac{sX}{sP} = \left(\frac{sT}{sP}\right)\left(\frac{sX}{sT}\right) = \left(\frac{1}{Qs + hA_s}\right)\left(\frac{\frac{\partial \xi}{\partial T} K_{SMA} L \epsilon_L}{Ms^2 + Bs + (K + K_{SMA})}\right)$$

(66)

This system is similar to the model found using Input-Output linearization. In fact, the model is essentially identical, the only difference is that with EVL we created a State-Feedback Controller which controls the temperature directly, whereas with IOL we ‘cancelled’ the temperature as possible. The EVL model therefore has a first order pole in place of the integrator seen in the IOL model. In practice, both methods require state measurements or estimates to be implemented, since the performance is directly tied to the accurate estimation of the varying parameters. A control method which does not require a model is presented in Chapter 11.

Chapter 11

Sliding Mode Control when Input Power is Limited

Both the linear and nonlinear controllers considered previously are able to accurately control the SMA actuator system, but in many cases are critically dependent on knowledge of the state and the material behavior. Sliding-model control provides a method that is robust even without a detailed model of the system. SMC is particularly easy to implement with SMA actuators because they are heated electrically using PWM: The PWM signal can be directly replaced by the SMC switching effort when the sample rate is sufficiently fast. However, the method cannot be applied without some design consideration since the electrical power input saturates and cooling speed is limited by SMA actuator size in practice. It is important to choose a sliding surface that is slow enough that the control effort to reach and maintain it. The limitations are maximum electrical power available (for heating), the difference between the SMA temperature and room ambient temperature, the expected range of all state variable, and uncertainty in the model used to design the controller. A controller designed within these limits leads directly to a maximum sliding mode 'bandwidth' for which the controller is guaranteed to be stable. Experiments confirm that sliding mode control is an excellent technique for SMA actuators.

Introduction

Sliding mode control provides a powerful and robust method to control very nonlinear systems which might have un-modelled dynamics or uncommon complexities like hysteresis or friction [208, 217, 218, 219]. In short, sliding mode control switches the control input very quickly between fully forward and fully reversed based on special measured conditions which ensure this fast switching will stabilize the system. However, the power of the method is only available when the physical system provides opportune conditions. The most critical such condition is that control effort (the push or pull on the system) can be switched very quickly without exciting the system in undesired ways. This can be achieved if the switching is very fast compared to the plant bandwidth. For shape memory actuators, the primary input is electric current which can be switched at speeds from 1 kHz to beyond 10 kHz using common electrical components such as MOSFETs or other transistors (see Chapter 6). These speeds are far beyond

the typical 2Hz to 10Hz thermal bandwidth of SMA actuators, so the switching itself provides no disturbance to the actuator. In practice, a sliding mode controller for SMA will turn the electrical power fully on to heat the material, or fully off to allow it to cool. The mathematical design of a sliding mode controller focuses on determining conditions for switching between these two inputs (rather than designing a continuously varying control effort). More specifically, a sliding mode control law designs a 'sliding surface' in the phase plane with our desired dynamics of the plant, applies the effort expected to keep the system in this plane (the equivalent control), then switches between positive and negative control effort to force the system to remain on this plane (the switching control). The control designer must satisfy two conditions:

1. The sliding surface is stable and of sufficient order to create a reachable sliding surface
2. The sliding surface has a bandwidth that does not exceed the ability of the control effort to reach and maintain the sliding surface.

Curiously, after these conditions are met there is very little need for a model of the plant for this controller to perform well. This provides a huge benefit in SMA where unmeasured conditions such as ambient temperature can change the behavior of the plant drastically and frequently. This chapter provides a background on sliding mode controller design, a method to determine stable bandwidth limits when controlling the LPV SMA model, and application examples for a physical SMA system. The practical intention of this control method is to provide robust control with little need for a model, so the switching control is more important than the equivalent control. Also, because the control effort can so easily be switched, this chapter does not consider 'soft' switching (such as sigmoid functions instead of sign functions) nor add intentional hysteresis to the controller – there is plenty of *that* in the physical plant.

Because the SMA devices are typically heated using electric current, sliding mode control has proven to be a reliable control method since fast switching does not disrupt the device. Simple implementation can be effective, and requires very little knowledge of the material behavior to implement at fair speeds [220, 221, 142, 152]. Sliding mode control has been used for SMA actuators in several cases: Song uses a first order sliding surface to control a slow jet nozzle outlet and other devices [222, 223, 159, 224]. Only Ashrafiuon has used a model to help design the controller [211, 225]. To implement the controller, the bandwidth of the sliding surface must be matched to the capabilities of the actuator. The sliding surface must be based on models or simulation since power input and ambient cooling speed are limited in implementation and

should drive the selection of an attractive sliding surface. A solution to this problem has not been studied previously, although the importance of ensuring attractive sliding surfaces was noted by Crews and Smith [226].

Sliding Mode Control Review

Sliding mode control is a form of variable structure control in which a system is driven onto a chosen stable dynamic equation (a 'surface' in the phase plane). Rather than driving the error to zero with a control law, the controller pushes the system onto a trajectory which is inherently stable. The sliding mode controller can implicitly cancel nonlinearities, poor models, and uncertainties in the system so long as the maximum control effort is sufficient, even without knowledge of these disturbances. To achieve this incredible ability requires a nonlinear control input which switches (infinitely quickly) between maximum and minimum inputs. This is not a problem for actuators which can accept fast input switching; it's especially easy to implement in systems which already accept pulse-width modulated input signals (electric motors, Ohmic heating devices like SMA actuators, etc.). Comprehensive design for a sliding mode control in these devices is performed in three steps: selecting a sliding surface, choosing a soft switching function (if needed), and validating that the controller can provide the equivalent input which would track the desired trajectory if the system and uncertainty amplitudes were known.

Selecting a Sliding Surface

The sliding surface is a stable dynamic equation. If the controller is used for an n^{th} order system, the sliding surface must be of order $n-1$ [218]. A reliable method to select the surface is:

$$\sigma(x) = (s+\lambda)^{n-1}$$

The controller will drive the system to behave, dynamically, like this equation by driving the system such that $\sigma = 0$. This differential equation with repeated Real roots is a good choice for hysteretic shape memory alloys or other cases where overshoot leads to long settle times. The bandwidth of the sliding surface is determined by λ. Given unlimited control effort, λ can be very fast compared to the system open-loop dynamics. However, in practice λ will be limited by the maximum control effort and by the switching speed, though switching speed is not considered here (c.f. [219]).

The Switching Control

A sliding mode control is implemented by switching the control effort between maximum and minimum limits to drive the system toward the sliding surface. The control effort is switched based on the current phase-plane location relative sliding surface:

$$u_{switch} = if \begin{Bmatrix} \sigma < 0 \\ \sigma > 0 \end{Bmatrix} then \begin{Bmatrix} U_{min} \\ -U_{max} \end{Bmatrix}$$

This is more easily written:

$$u_{switch} = -U * sign(\sigma)$$

In many systems, it is not possible to instantly switch the control effort. In these cases, the signum (sign) function can be replaced with a softer switch such as a sigmoid or tanh function. This change will cause steady state errors in the controller. For SMA devices, soft switching is not necessary because the electric heating current can be switched on and off in nanoseconds using Field Effect Transistors, for instance.

For demonstration, an unstable second order system was controlled using a hard switching law. The dynamic system is second order, governed by:

$$\ddot{x} - 2\zeta\omega_n\dot{x} + \omega_n^2 x = u$$

In the demonstration, $\omega_n = 1$ and $\zeta = 70\%$. The sliding surface is chosen to be first order, since the plant is second order:

$$\sigma = \dot{x} + \lambda x$$

For this demonstration, the sliding surface eigenvalue is $\lambda = 2$. If the controller is successful, the trajectory of the system will remain on $\sigma = 0$, and the dynamics will be first order:

$$\dot{x} = -\lambda x$$

The controller is implemented assuming the maximum control effort is $U = 8$, results for the initial condition $x_0 = 0, \dot{x}_0 = 0$ are shown in Figure 107. Looking to the left plot, the position is shown quickly driven to the origin in $\sim$2 seconds, consistent with the selected sliding mode time constant. The red line shows the difference between the sliding surface desired trajectory and the actual trajectory – the sliding mode is reached in approximate 0.2 seconds. Another way to see the system reaching then maintaining the sliding mode is shown in the lower right of Figure 107 where the phase plane is shown. The controller initially applies full effort in one direction to drive the system toward the sliding mode (shown in red). Once the sliding mode is reached, the

control effort switches rapidly to keep the system on the sliding mode – which is does successfully. Looking now to the top right figure, this switching control effort is very obvious as the controller switches rapidly between maximum and minimum value while on the sliding mode. The other lines on this chart (u_{lp} and u_{eq}) will be discussed in the following section.

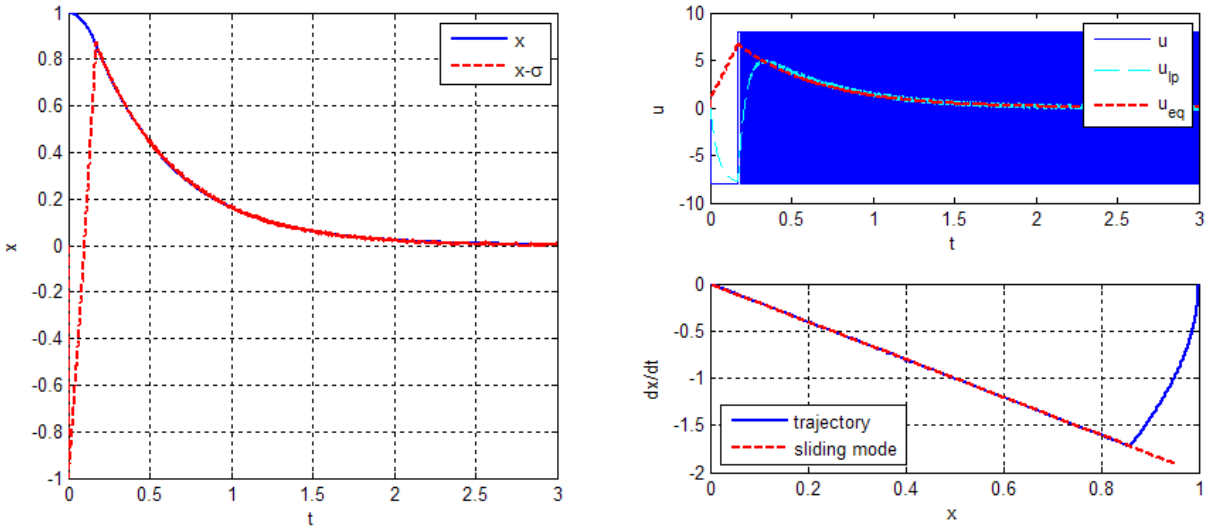


Figure 107 - Sliding mode control of an unstable second order system. The controlled position quickly tracks the sliding mode – position vs time is shown on the left, the phase plane is shown in the lower right. Control effort is shown in the top right – while on the sliding mode, the control effort switches rapidly between maximum and minimum values.

The Equivalent Control

Why does such a simple switching control law work? The stability of the controller can be analyzed in detail using a Lyapunov argument to ensure that the controller drives the system toward the sliding mode, and that it is able to maintain the sliding mode once it is reached. First, consider that the system is momentarily on the sliding surface – can the control effort keep the system on the sliding? Consider the class of nonlinear systems which can be modelled by:

$$\dot{x} = f(x) + g(x)u + \eta$$

Where f and g are potentially nonlinear functions and η contains all unknown disturbances. The equivalent control effort u_{eq} which would ensure $\dot{\sigma} = 0$ such that the system does not leave the sliding mode once it is reached, is found by solving:

$$0 = \dot{\sigma} = (\nabla_x \sigma)\dot{x} = (\nabla_x \sigma)\big(f(x) + g(x)u_{eq}\big)$$

Secondly, determine whether this controller can guarantee that the sliding surface is reached at all. The so called 'reachability condition' considers the controller which ensures:

$$\sigma\dot{\sigma} < 0$$

(67)

In general, the result will include all known linear and nonlinear plant dynamics, as well as an estimate of the maximum uncertainty or disturbances that might be encountered. It is often easy to design a linear controller to augment the switching control if the plant model is accurate, since a continuous control law can ensure reachability for the system except uncertainties or modeling errors.

The critical result of this analysis is to determine whether the control effort is sufficient to reach and maintain the sliding mode. Note that it is not necessary to implement the equivalent control, though in some systems that could reduce chatter or better ensure the sliding system is reached. Consider the switching control alone: The switching does not disturb the system because the switching is fast compared to the plant bandwidth. In other words, only the low frequency component of the control effort actually affects the system. Consider the control effort to have both low-pass u_{lp} and high-pass u_{hp} components:

$$u = u_{lp} + \left(u - u_{lp}\right) = u_{lp} + u_{hp}$$

The result is that, given sufficient control effort, the u_{lp} component of the switching control will provide perfect equivalent effort. Now consider the reaching condition (67) with the switching control:

$$0 > \sigma\dot{\sigma} = \sigma\left(u_{eq} + \eta - U * sign(\sigma)\right)$$

The controller which ensures reachability simply must have more effort (U) available than the maximum equivalent control and the maximum of any unknowns.

Consider the same unstable 2nd order system example from the previous discussion. The equivalent control which would maintain the sliding mode if the model was perfect is:

$$0 = \dot{\sigma} = \ddot{x} + \lambda\dot{x} = \lambda\dot{x} - \frac{k}{m}x - \frac{b}{m}\dot{x} + u_{eq}$$

$$u_{eq} = \left(\frac{b}{m} - \lambda\right)\dot{x} + \frac{k}{m}x$$

Because the model used in the controller is exact, the control effort required for reachability is easily determined based on initial conditions. Looking back to Figure 107, the comparison between this analytical equivalent control and the low pass component of the actual controller are shown in the top right plot in red and cyan, respectively. It is clear that, after a delay during

reaching, the controller provides the equivalent control effort and can therefore maintain the sliding mode. If the maximum control effort was reduced to $U = 2.8$, for instance, the system has a long delay before reducing speed enough to allow the controller to track the needed equivalent control effort (Figure 108). The same error could be introduced by increasing the initial condition magnitude, adding uncertainties, increasing sliding surface bandwidth, etc. This could be improved by adding the equivalent control effort to the switching control or by augmenting the switching control with a linear control law, anything which pushes the system into a region where the switching is effective.

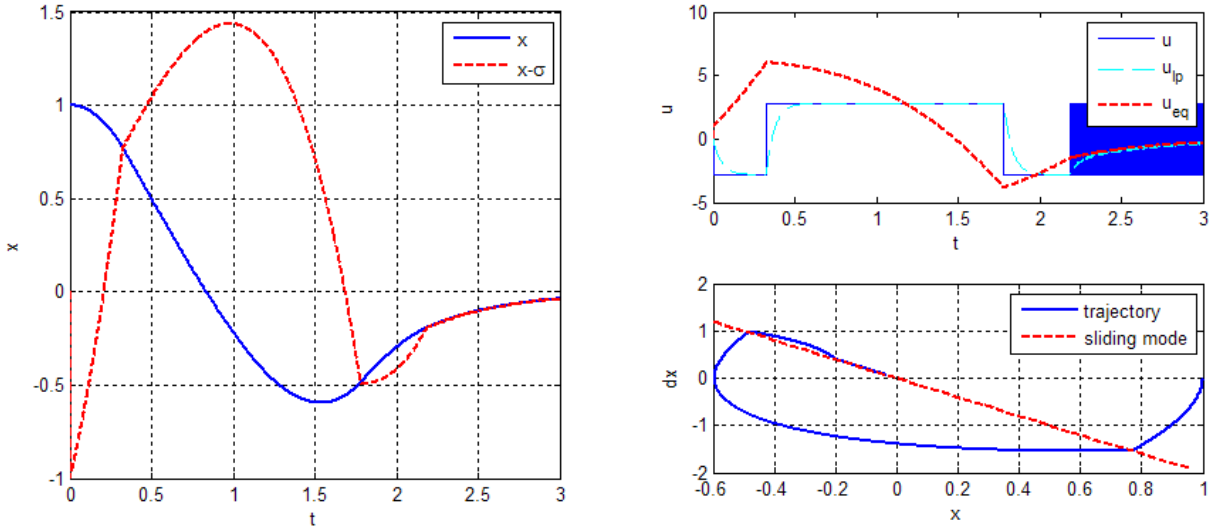


Figure 108 - control where the maximum input effort is barely sufficient to reach the sliding model

The main difficulty when designing a reliable sliding controller is balancing the maximum control effort against known system dynamics, modelling uncertainties, disturbance magnitudes, and sliding bandwidth. In most implementations, the control effort limits are fixed, the maximum range of the state is known, and the uncertainties are not known precisely. A sliding mode should be chosen of order $n - 1$, and an appropriate switch chosen. The remaining design parameter available is the sliding mode bandwidth to balance everything else.

Design for the SMA System

Consider the SMA system described in Chapter 9:

$$\dot{T} = \frac{P}{Q} - \frac{hA_s}{Q}(T - T_\infty)$$

$$Q \equiv mc_p - m\Delta H \frac{\partial \xi}{\partial T}$$

$$M\ddot{x} + B\dot{x} + (K + K_{sma})x = K_{sma} L \epsilon_L \xi$$

$$\xi = f(T)$$

To simplify the design process, consider the design as if the temperature is the control input, and the heat transfer equation is a filter through which the switching input *P* passes:

$$M\dddot{x} + B\ddot{x} + (K + K_{sma})\dot{x} = K_{sma} L \epsilon_L \dot{\xi} = K_{sma} L \epsilon_L \frac{\partial \xi}{\partial T} \dot{T}$$

(68)

Which can be written, since the first order temperature dynamics are known, as:

$$M\dddot{x} + B\ddot{x} + (K + K_{sma})\dot{x} = \frac{K_{sma} L \epsilon_L}{Q} \frac{\partial \xi}{\partial T} \left(P - hA_s(T - T_\infty)\right)$$

(69)

The sliding surface is chosen to be second order because the system is third order. The sliding surface is critically damped so that the actuator does not overshoot during regulation, which would cause slow settling and oscillation due to hysteresis.

$$\sigma = \ddot{x} + 2\lambda\dot{x} + \lambda^2 x$$

The equivalent control law is designed considering:

$$\dot{\sigma} = \dddot{x} + 2\lambda\ddot{x} + \lambda^2\dot{x} = 0$$

By inserting the dynamics known from (68), the equivalent control can be derived:

$$\frac{K_{sma} L \epsilon_L}{MQ} \frac{\partial \xi}{\partial T} \dot{T} = \left(\frac{B}{M} - 2\lambda\right)\ddot{x} + \left(\frac{(K + K_{sma})}{M} - \lambda^2\right)\dot{x}$$

$$\frac{K_{sma} L \epsilon_L}{MQ} \frac{\partial \xi}{\partial T} \left(P_{eq} - hA_s(T - T_\infty)\right) = \left(\frac{B}{M} - 2\lambda\right)\ddot{x} + \left(\frac{(K + K_{sma})}{M} - \lambda^2\right)\dot{x}$$

$$P_{eq} = \frac{\left(\left(\frac{B}{M} - 2\lambda\right)\ddot{x} + \left(\frac{(K + K_{sma})}{M} - \lambda^2\right)\dot{x}\right)}{\frac{K_{sma} L \epsilon_L}{MQ} \frac{\partial \xi}{\partial T}} + hA_s(T - T_\infty)$$

(70)

And the switching control is implemented:

$$P_{switch} = if \begin{Bmatrix} \sigma < 0 \\ \sigma > 0 \end{Bmatrix} then \begin{Bmatrix} P_{max} \\ 0 \end{Bmatrix}$$

This is not an ideal scenario because the electrical input can only heat the wire, while ambient air cools it. The heating can quickly contract the wire but cooling and extension is limited. The result is that tracking when the actuator is contracting can be faster than when extending and cooling.

Simulation

This switching controller was implemented to control the standard demonstration of SMA actuator and spring-mass-damper plant. The maximum input power is 3 W, and cooling is limited by the ambient air at 20℃. The sliding surface is a critically damped second order ODE, with both poles located at $\lambda = -10$. First, the controller regulates the system to a position in the mid-range of actuator motion (Figure 109). The difference between the equivalent input P_{eq} and the low-pass filtered actual input is very close – errors are likely due to model simplifications that were made when designing the controller, not errors in the control itself.

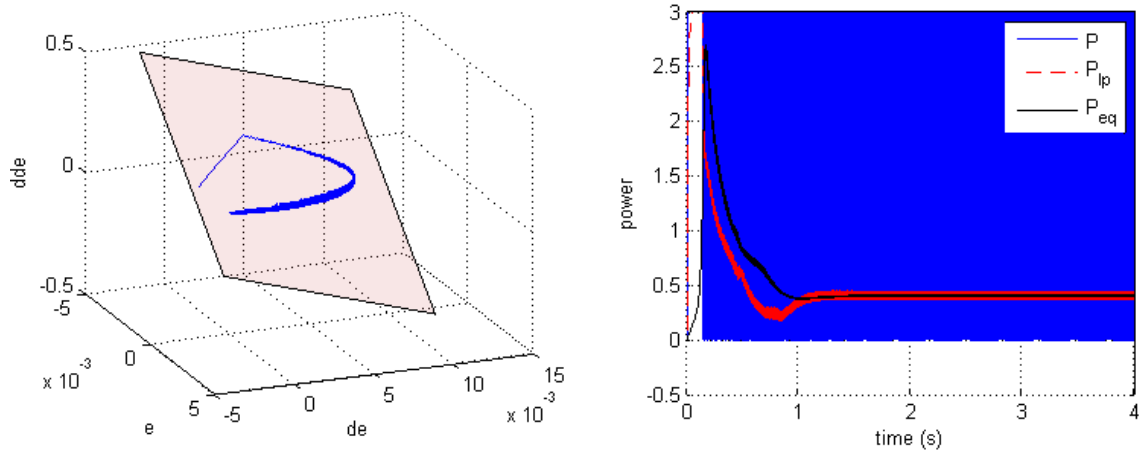


Figure 109 – (LEFT) The trajectory approaching and maintaining the sliding surface is shown. (RIGHT) The control effort is sufficient, in this case, to nearly provide the same input as the ideal equivalent controller.

A similar simulation with the same constraints was used to regulate the system to two different positions in sequence (Figure 110). The first reference is approached mostly by heating the wire, while the second requires cooling to reach the reference. Because of this, the controller is able to provide the equivalent control sufficiently when driving toward the first (heating) position. However, it cannot quickly track during cooling because this speed is limited to the heat transfer to the air – this is seen as a gap in control activity from 2 to near 2.6 seconds. Overall, the controller performs excellently compared to the continuous controllers of previous

chapters. The settle time is near 1 second for both heating and cooling, and no overshoot causes long oscillations caused by hysteresis.

For comparison, the simulation was repeated with the only change being that input power was limited to 1 Watt (Figure 111). In this case, it is clear that the sliding mode was not reached very quickly during the heating nor cooling periods. The balance between available electrical power and sliding bandwidth was not sufficient, as shown by the long delay before control activity when the simulations is first started (from 0 to 1 second) and when the reference is changed (2 to 3 seconds).

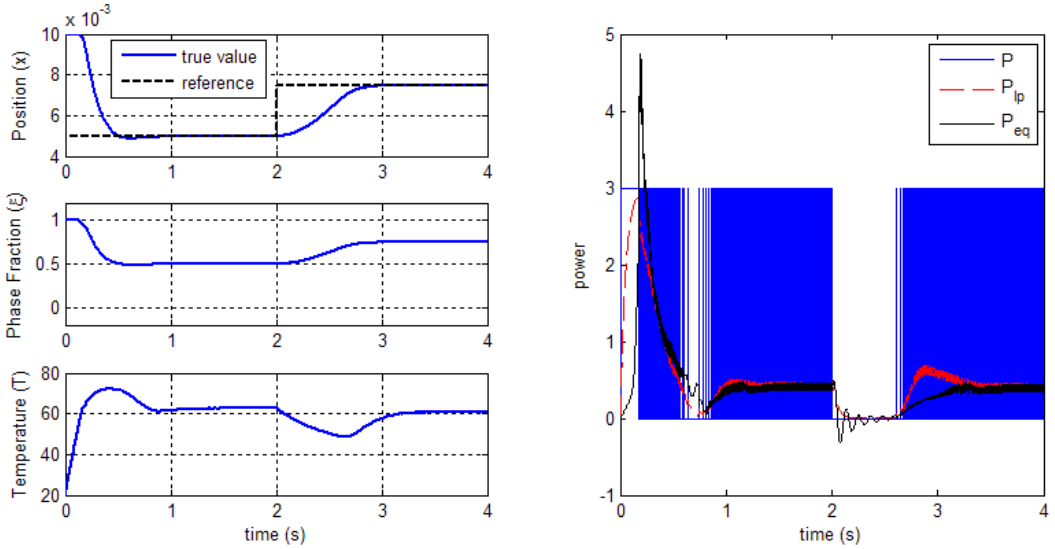


Figure 110 - Sliding mode control regulates to two positions. The control effort is high enough that the heating effort needed is easily achieved, but cooling speed cannot quickly reach the sliding mode.

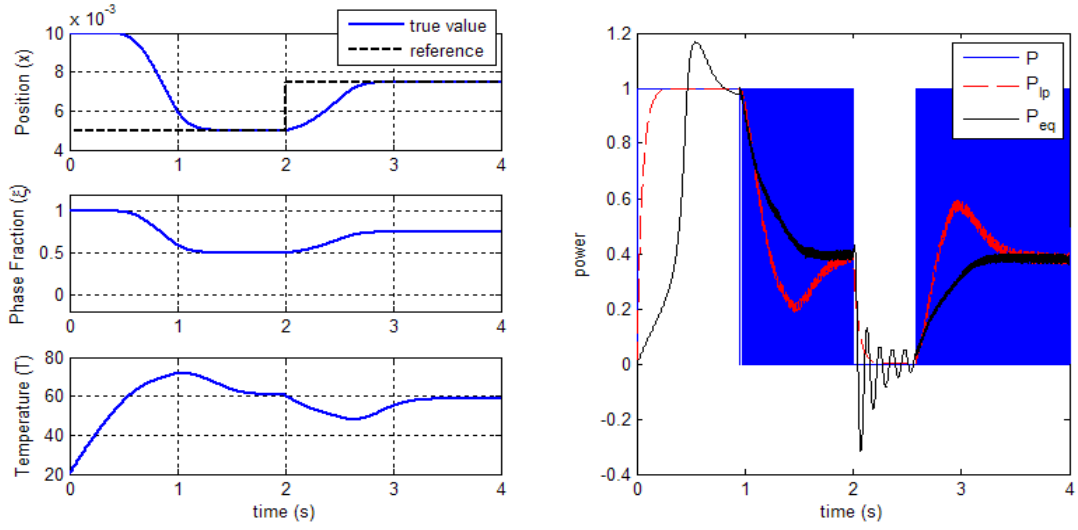


Figure 111 - Sliding mode control regulates to the same two positions of Figure 110. This time, the control effort is too high for a one Watt power source to maintain the desired sliding mode even when heating.

Experiment

An experiment was performed to validate the sliding mode controller. In addition to demonstrating the sliding-mode control performance, this experiment shows that the 2nd order sliding surface was an appropriate choice for this system. An SMA wire was attached to a physical spring-mass-damper as used for the estimator in Chapter 9. The sliding-mode controller was implemented with the same critically damped sliding surface, with eigenvalues at $\lambda = -10$. The maximum power in practice was limited by a 6 V supply, which provides peak power near 1 Watt. A step reference was reliably tracked with this controller, with no overshoot and minimal steady-state error (Figure 112). To reduce the noise in the error derivatives needed when computing the sliding surface value, the sample rate was reduced to 200 Hz – this is much fast than the SMA bandwidth so the controller works, and provides accurate tracking with a noisy position measurement. The problem arises since double-differentiation of most common sensors will lead to significant noise which directly disrupts the sliding surface calculation. A first reaction to this problem might be to reduce the sliding surface order. However, as expected, this greatly disrupts the control accuracy and leads to oscillation around the reference (Figure 113).

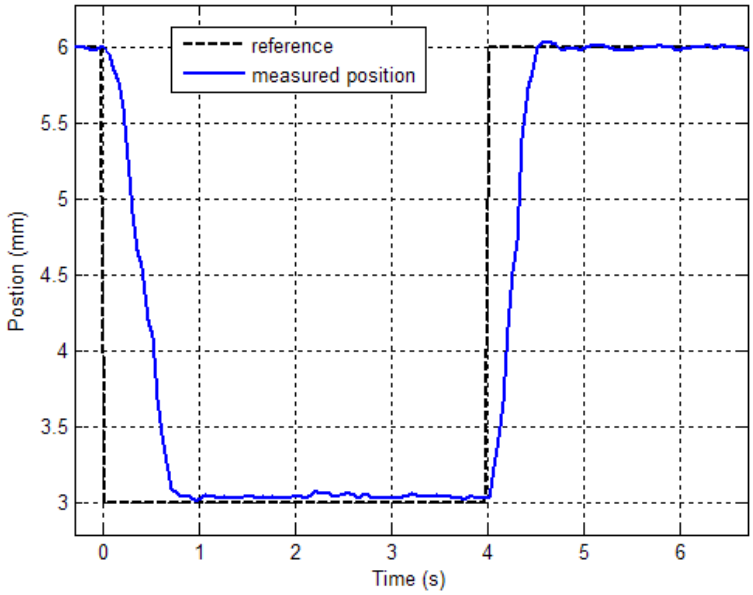


Figure 112 - Control using a second order sliding mode is fast and reliable (Compare to Figure 110)

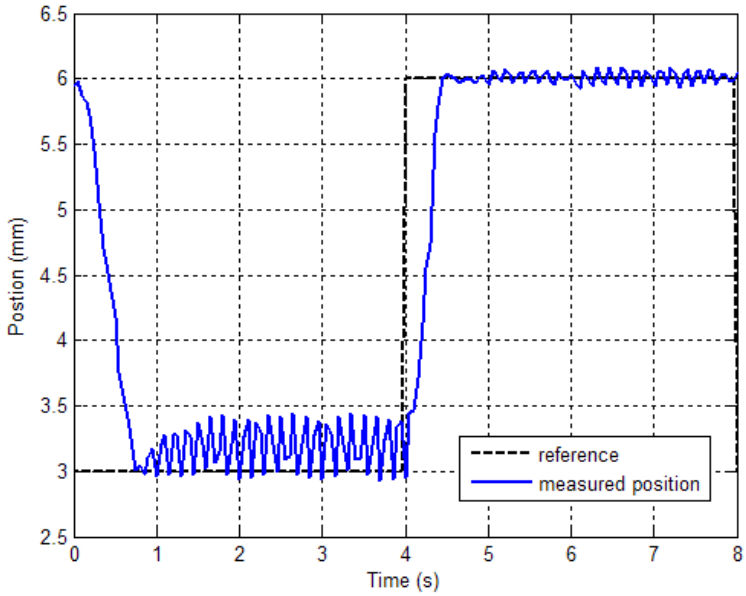


Figure 113 - Control using a first order sliding surface leads to limit cycle oscillations

Conclusion

Sliding mode control was reviewed and employed for control of an SMA actuator. It was shown that a critically damped second order sliding surface is required, and that the following variables interact in selection of the sliding surface bandwidth:

- Available control effort
- Range of the state
- Accuracy of the system model
- Range of potential disturbances

The controller is easily implemented in the SMA system since fast switching is easily achieved with solid-state electrical switching as with MOSFETs. This controller is very robust to unknown models - the SMA actuator has many unknown parameters which are difficult to estimate and which may change over time. This control method has been used extensively by the author; the excellent capabilities will be made clear as the method is used to control all the experiments and demonstrations shown in Chapter 12. There are very few drawbacks to the sliding mode controller for SMA. One consideration is that the acceleration of the system must be determined for the second order control law – this requires double differentiation of a position sensor such as the dual-measurement sensing method shown in Chapter 7. In this case, it may be better to create a filter or otherwise create smooth version of the states needed in the sliding surface equation.

Chapter 12

Conclusion

The mechanical, electrical, and controls designs presented in previous chapters were applied to several mechanical systems which explore and demonstrate that these ideas can be synthesized and do indeed enable shape memory alloy actuators for robotics. The devices created during this research range from small RC servo mechanisms to robots with multiple degrees of freedom. Conclusions are drawn about the success of this thesis, and potential directions for future work are discussed.

Micro-Servo

A small PCB demonstrates the dual-measurement self-sensing technique, sliding mode control, and the great potential of SMA to reduce cost and complexity of existing RC servos. The speed, range, and torque are comparable to an equal weight RC servo. The total cost is less than a dollar, even in prototype quantities.

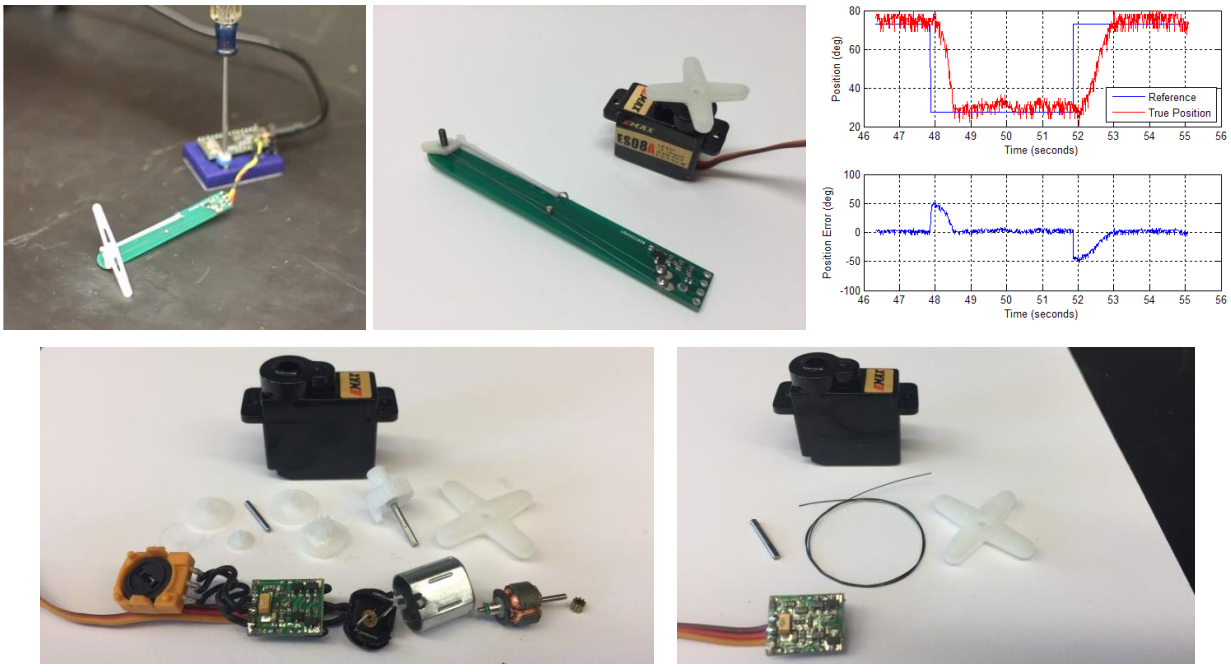


Figure 114 - RC Servo on a PCB shows the simplicity and strength of SMA devices with self-sensing

Bowden Tube

This is another demonstration which shows the potential to replace RC servos. The Bowden tube provides a technique to package a long length of SMA wire in a way that is not intrusive. Curiously, appropriate design of the tube dimensions increases the heat transfer rate – making this device faster than the bare wire in ambient air [5].

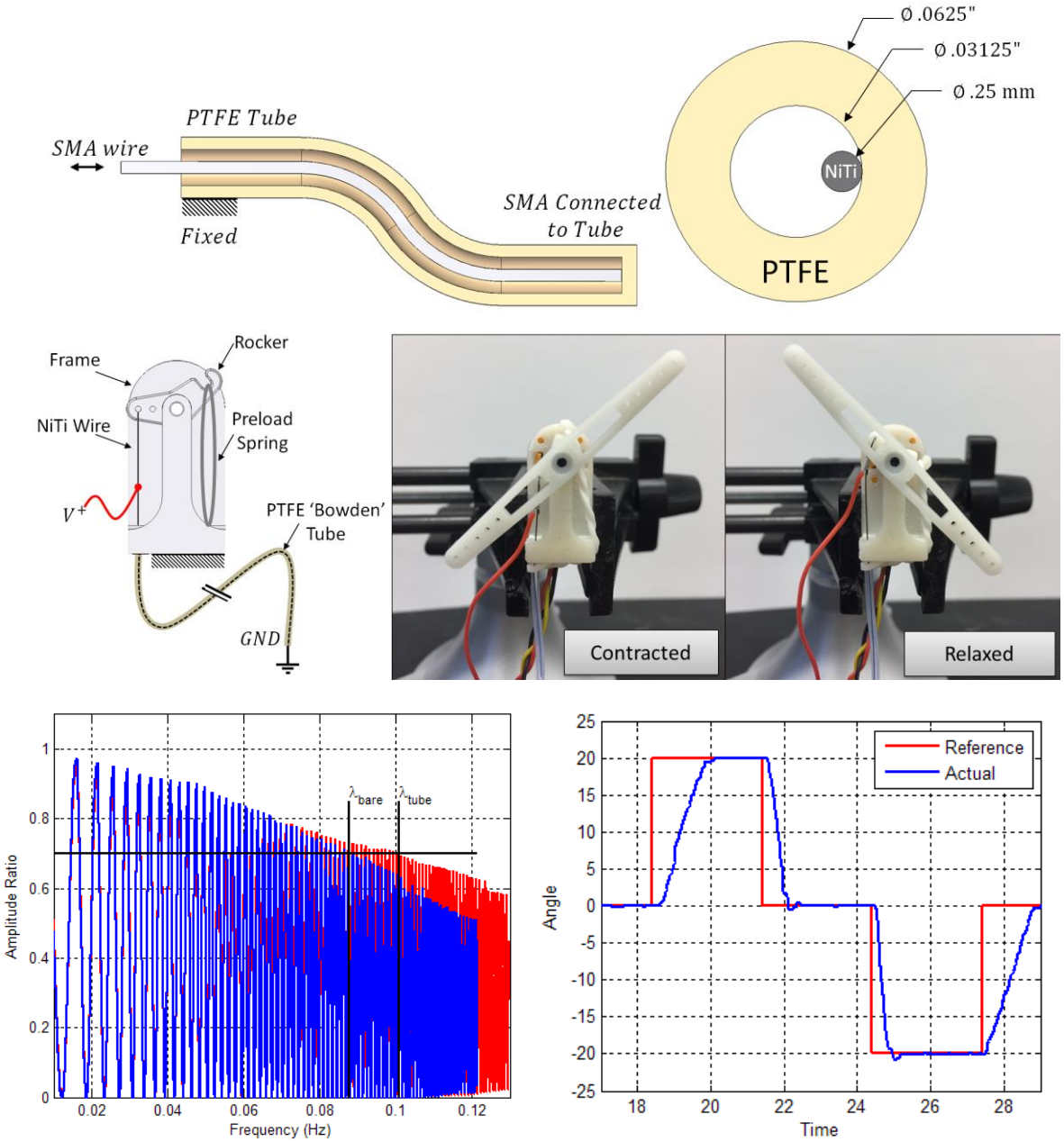
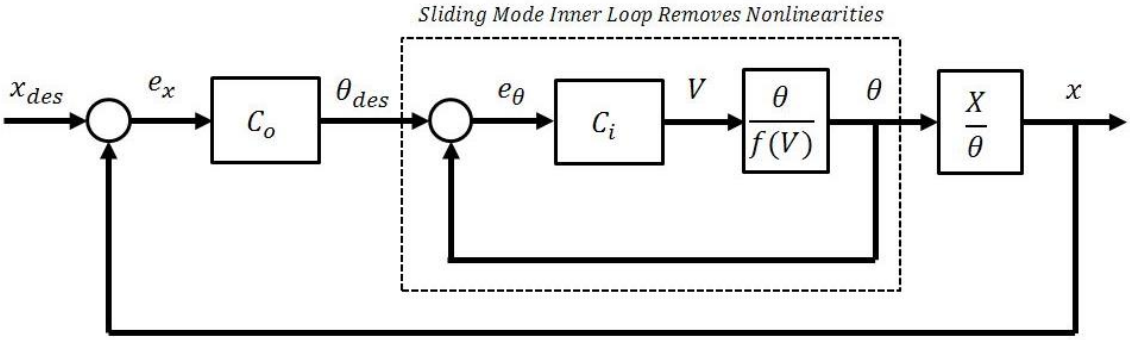
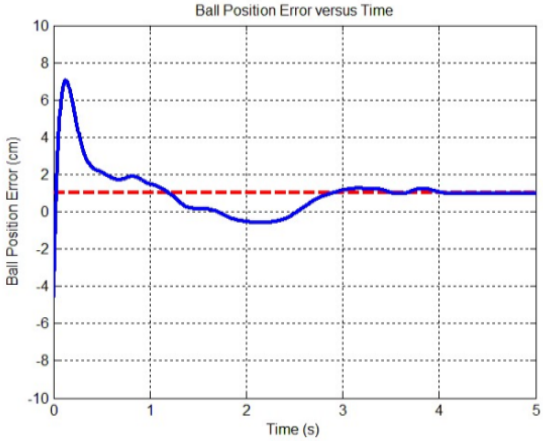


Figure 115 - A rotary actuator with a Bowden tube which allows a long length of SMA wire to be packaged in an unobtrusive manner

Ball-on-Beam

The ball-on-beam is a classic demonstration for controls classes – it provides an unstable, 4th order plant which is difficult to stabilize and impressive when successful. In this case, the problem was solved by using a cascaded controller, with a sliding mode inner loop for the SMA to control beam angle, and a PID outer loop where the ball position is controlled. This system is robust to disturbances, but is always 'twitchy' in ambient air since small air currents can quickly cool the SMA wire [4].





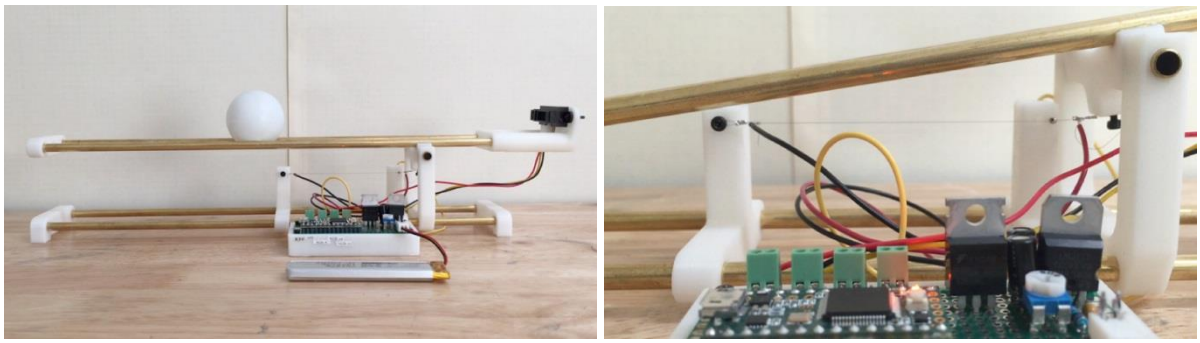


Figure 116 - A Ball-on-beam demonstration shows that the SMA can provide fast and reactive motion

Hexapod

An 18 degree-of-freedom (DoF) robot was built from 18 identical servos (similar to those shown in Figure 114). The robot stands only 3 inches tall. This machine demonstrates the ability to package many SMA actuators in a small space. Controlled by a central PWM driver system, all 18 DoF are coordinated to create walking and crawling gaits.

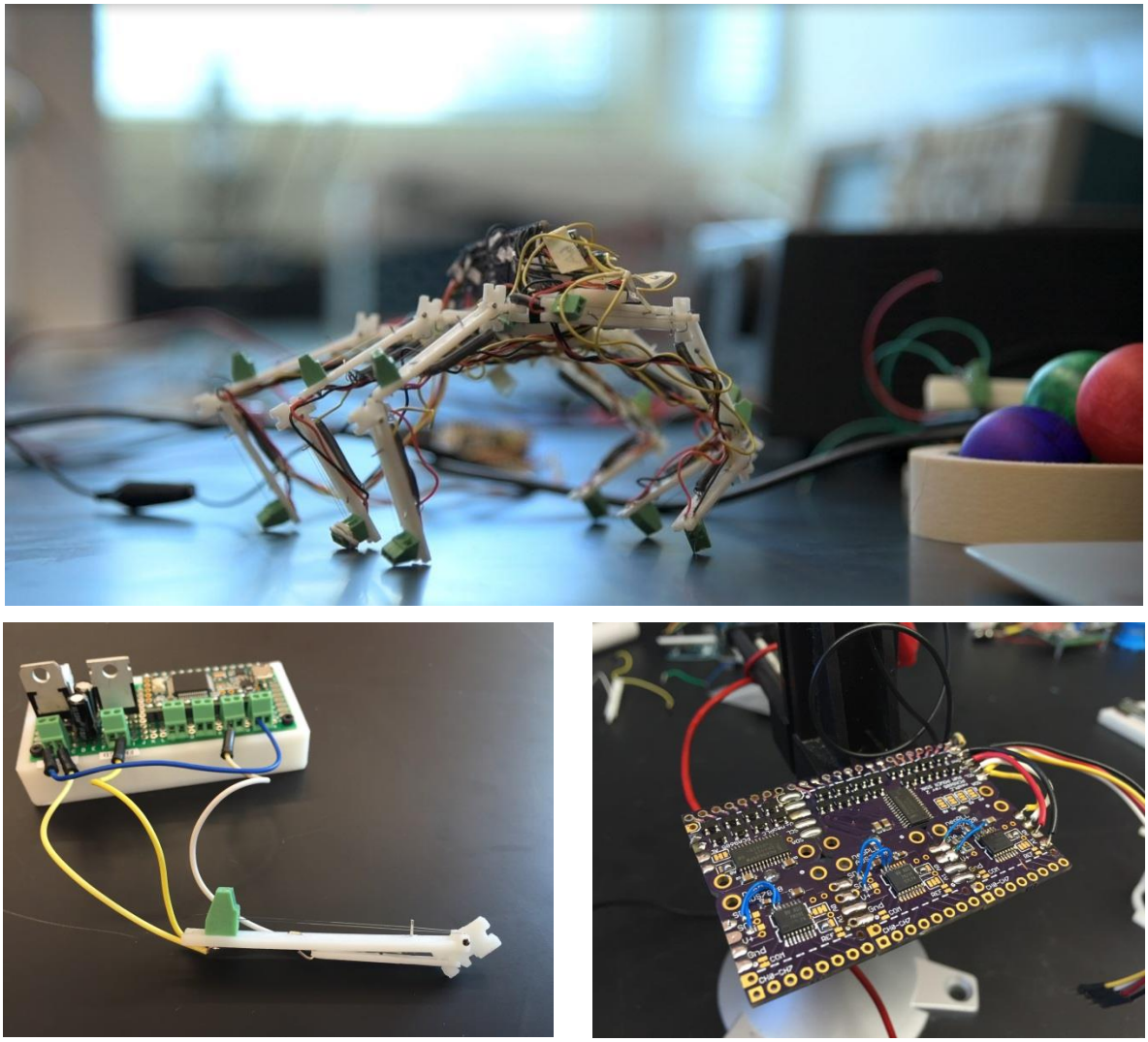


Figure 117 - A robotic hexapod with 18 degrees-of-freedom. All DoF are created from identical rotary actuators.

Ultralight Arm

A lightweight arm was designed as an appendage for a small space-craft such as a cube satellite. The SMA wire is completely shielded inside the arm's tubes, which would prevent radiation from quickly cooling the wires. Each joint is designed to ensure the wire will not fatigue due to curvature. The arm uses a motion-tracking camera for feedback control without angular position sensors on the joints. This provides an interesting control problem since the kinematics are not known to the controller.

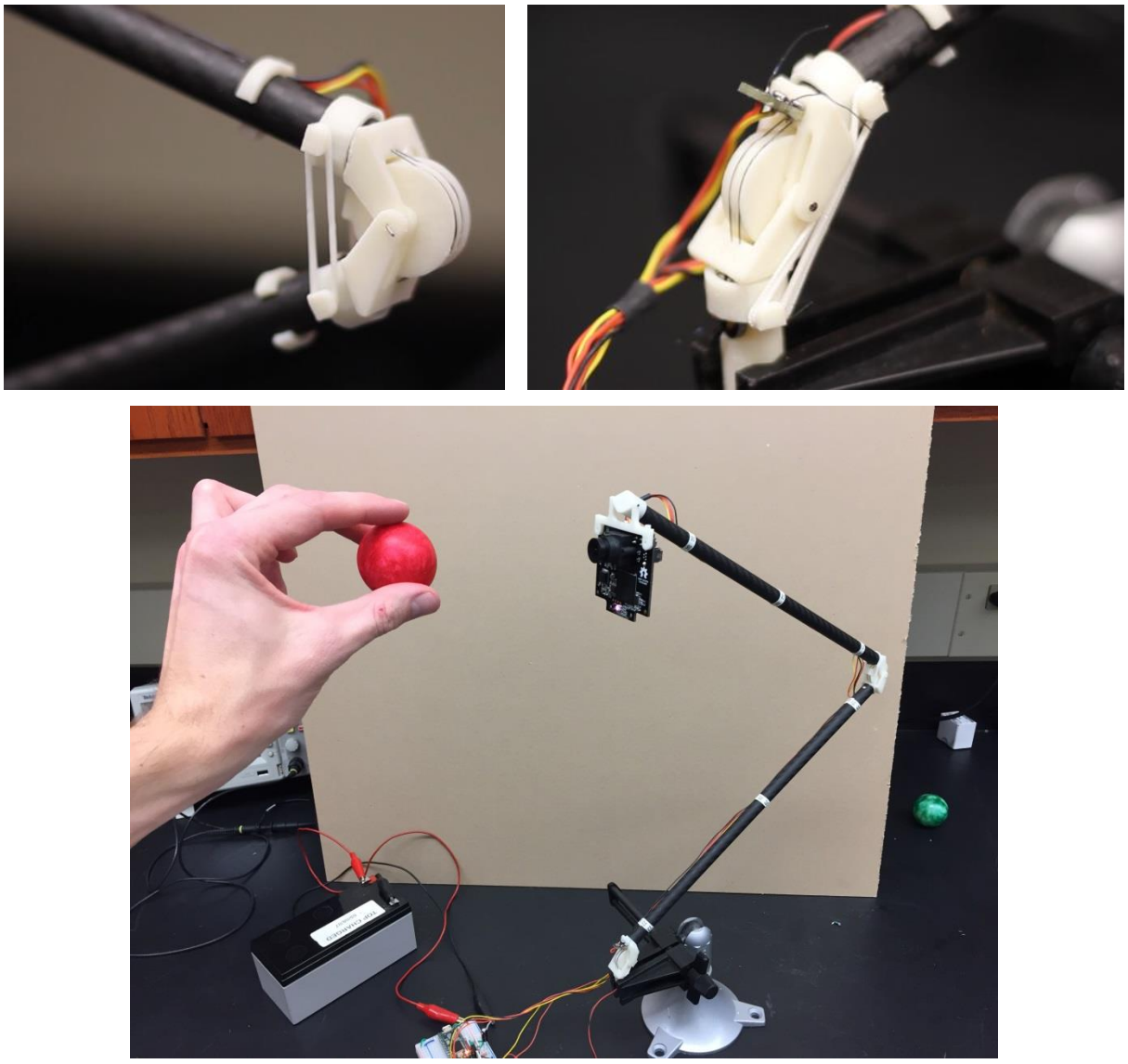


Figure 118 - A 2-DoF, ultralight robotic arm for a cube-satellite with a color-tracking camera. The system provides an interesting control problem since the joint angles are not measured, only the position of the camera relative to the ball.

Dexterous Robot

An 8-DoF robot with two arms was created to demonstrate SMA actuators in a 2/3 scale humanoid torso. Rotary potentiometers measure the angle of each joint, and machine vision is used to track the ball location. The strength, speed, and silence of the SMA actuators are displayed by balancing a ball on a flat paddle.

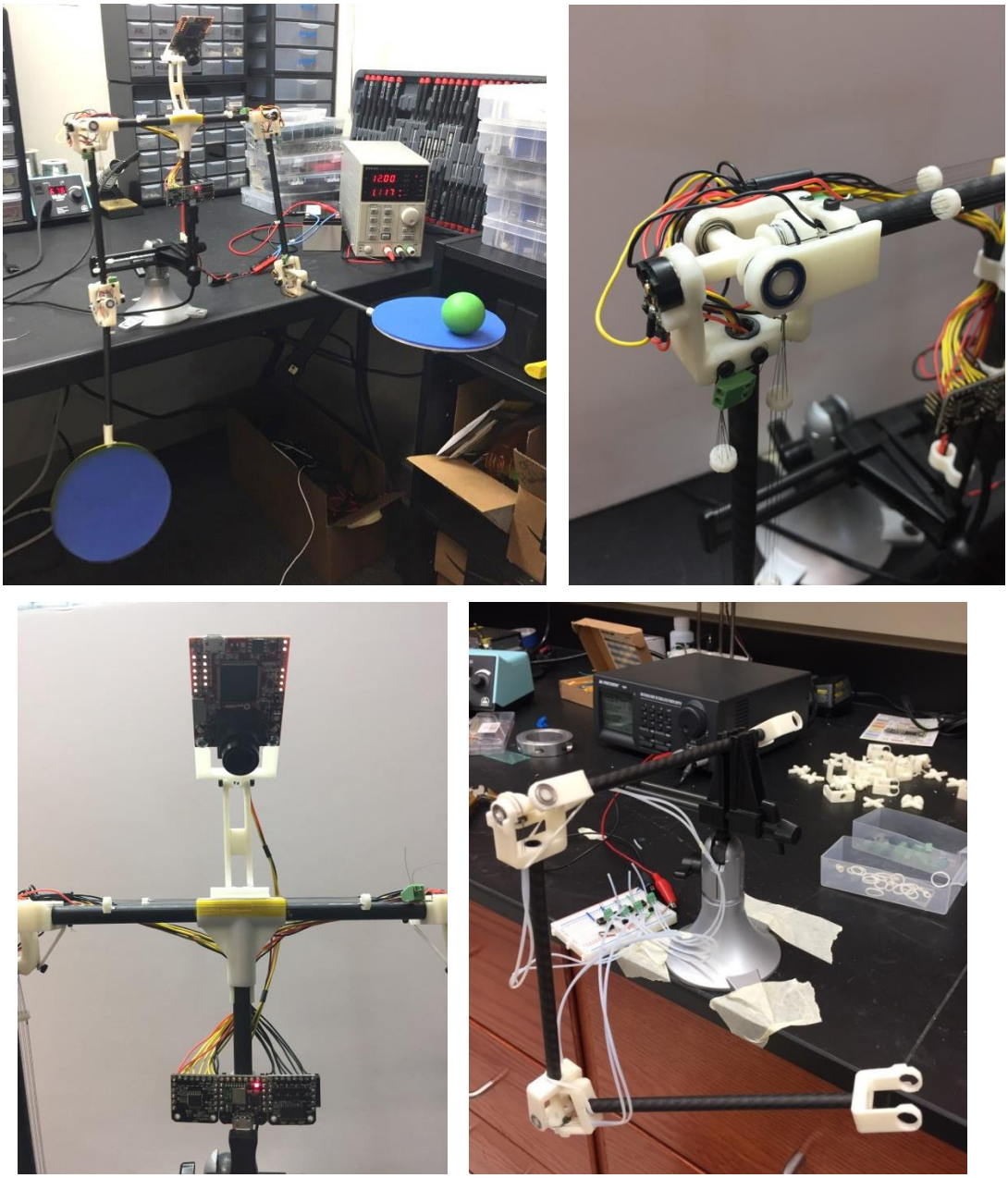


Figure 119 - A dexterous 8-DoF robot demonstrates that complex robotic devices can be built from SMA at low cost. The bottom right image shows an embodiment using Bowden tubes instead of exposed wire.

Thesis Conclusion

This thesis is, admittedly, not the first time SMA actuators have been studied. However, the great bulk of research has been in the material science of SMA, and no one has actually shown that SMA can be used as reliable robotic actuators. A critical part of the author's plan for this thesis is that concepts and designs generated during the research be implemented in physical hardware. Without physical implementation, this work would not provide any real proof that this has been achieved. The mechanisms studied range from simple straight wire devices, to complex robots. The simple devices are built up into machines which demonstrate all the key features of SMA which make them better, in some applications, than any other artificial muscle technology. In the introduction, an outline of the good and bad features was shown. The comparison is shown again in Table 11. While some unfortunate aspects of SMA like low energy efficiency will always be present (until new alloys are discovered), most of the obstacles to enabling SMA actuators for robotics have been overcome in the course of this research. The low strain and low speed has been overcome by packaging fine SMA wires in new ways like Bowden tubes and parallel arrays which increases heat transfer. The dual measurement technique provides reliable estimation of wire length with no sensors. The nonlinear behavior in actuators was modelled in a new way that ensures it can be simulated accurately. The hysteresis and nonlinear response was overcome with improved estimators and non-linear control methods based on the new models.

Table 11 - The good, bad, and ugly (but solvable) features of shape memory alloy actuators

The Good	The Bad	The Ugly*
High specific strength	Energy efficiency (3-5%)	Low strain ***Solved by mechanical design***
High specific energy (work)	Life-time (100k-1M cycles)	Sensors bigger than wire ***Dual measurement technique***
Silent operation		Nonlinear and hysteretic ***Improved control & estimation***
Excellent DoF/$		Difficult training/processing ***Studied heat treatment process***
Excellent DoF/m^2		Bandwidth vs size ***Solved by mechanical design***

* solved in this thesis

Where does the research stand now? The major challenges to implementing any SMA actuator are now overcome. It is time to focus in the other direction: promote the beneficial

aspects and make use of the incredible potential. The most obvious direction is to move toward machines with more degrees of freedom in a smaller package. Prosthetic hands, robotic insects, smart fabrics: any device where the size and number of degrees of freedom has ruined the potential for other designs. There are opportunities to study the damping and sensing properties of super-elastic SMA for smart materials and potentially for armor. In the direction of improved materials, the known alloys which are better than NiTi should be studied; NiTiCu can provide lower hysteresis and NiTiFe can provide higher strength – perhaps these benefits outweigh the increased costs of these alloys. In robotics, shape memory alloys are, of course, not the only technology which could disrupt the robotics industry – pneumatics, shape memory polymers, and others have some potential. But with the order of magnitude advantage on cost and strength, SMA will remain the dominant candidate for the next decade or more.

~~~
~~~

References

[1] T. R. Lambert, A. Gurley and D. Beale, "Heat Transfer Modelling and Bandwidth Determination of SMA Acuators in Robotics Applications," in *Early Career Technical Conference*, Birmingham, 2016.

[2] T. R. Lambert, A. Gurley, K. Kubik, D. Beale and R. Broughton, "Numerical Heat Transfer Modelling of SMA Actuators and Controller Comparison," in *ASME Conf. on Smart Materials, Adaptive Structures, and Intelligent Systems*, Snowbird, 2017.

[3] A. Gurley, D. Beale, R. T. Lambert and R. Broughton, "Robust Self Sensing in NiTi Actuators Using a Dual Measurement Technique," in *ASME 2016 Conf. on Smart Materials, Adaptive Structures and Intelligent Systems*, Stowe, 2016.

[4] A. Gurley, T. R. Lambert, D. Beale and R. Broughton, "Dual Measurement Self-Sensing Technique of NiTi Actuators for use in Robust Control," *Smart Materials and Structures,* (in review).

[5] A. Gurley and et al, "Bowden Tube NiTi Actuators with Linear Parameter Varying Model and Sliding Mode Control," in *Proc. ASME 2017 Conf. on Smart Materials, adaptive Structures, and Intelligent Systems*, Snowbird, 2017.

[6] T. R. Lambert, A. Gurley, D. Beale and R. Broughton, "SMA Actuator Material Model with Self-Sensing and Sliding Mode Control; Experiment and Multibody Dynamics Model," *J Smart Materials and Structures*, vol. 26, no. 3, 2017.

[7] L. Brinson and M. Huang, "Simplifications and Comparisons of Shape Memory Alloy Constitutive Models," *Journal of Intelligent Material Systems and Structures*, vol. 7, pp. 108-114, 1996.

[8] A. Paivo and M. Savi, "An Overview of Constitutive Models for Shape Memory Alloys," *mathematical Problems in Engineering,* vol. 2006, pp. 1-30, 2006.

[9] K. Tanaka, "a thermomechanical sketch of shape memory effect: one-dimensional tensile behavior," *Res Mechanica,* vol. 2, no. 3, pp. 59-72, 1986.

[10] H. Prahled and I. Chopra, "Comparative Evaluation of Shape Memory Alloy Constitutive Models with Experimental Data," *Journal of Intelligent Material Systems and Structures,* vol. 12, pp. 383-395, 2001.

[11] H. Sayyaadi, M. R. Zakerzadeh and H. Salehi, "A comparative analysis of some one-dimensional shape memory alloy constitutive models based on experimental tests," *Scientia Iranica B,* vol. 19, no. 2, pp. 249-257, 2012.

[12] L. Brinson, "One-dimensional constitutive behavior of shape memory alloys: thermomechanical derivation with non-constant material functions and redefined martensite internal variable," *Journal of Intelligent Materials, Systems, and Structures,* vol. 4, pp. 229-242, 1993.

[13] A. Banerjee, "Simulation of shape memory alloy wire actuator behaviour under arbitrary thermo-mechanical loading," *Smart Materials and Structures,* vol. 21, pp. 1-13, 2012.

[14] L. Li, Q. Li and F. Zhang, "One-Dimensional Constitutive Model of Shape Memory Alloy with an Empirical Kinetics Equations," *Journal of Metallurgy, Hindawi Publishing Corporation,* vol. 2011, no. 563413, pp. 1-14, 2011.

[15] J.-H. Chung, J.-S. Heo and J.-J. Lee, "implementation strategy for the dual transformation region in the Brinson SMA constitutive model," *Smart Materials and Structures,* vol. 16, pp. N1-N5, 2007.

[16] S. Furst, J. Crews and S. Seelecke, "Stress, strain, and resistance behavior of two opposing shape memory alloy actuator wires for resitsancebased self-sensing applications," *Intelligent Material Systems and Structures,* vol. 24, no. 16, pp. 1951-1968, 2013.

[17] J. G. Boyd and D. C. Lagoudas, "A Thermodynamical constitutive model fo shape memory materials. Part 1. The monolithic shape memory alloy," *International Journal of Plasticity,* vol. 12, no. 6, pp. 805-842, 1996.

[18] J. G. Boyd and D. C. Lagoudas, "A Thermodynamical Constitutive Model for Shape Memory Materials. Part II. The SMA Composite Material," *International Journal of Plasticity,* vol. 12, no. 7, pp. 843-873, 1996.

[19] D. Lagoudas, D. Hartl, Y. Chemisky, M. Luciano and P. Popov, "Constitutive model for the numerical analysis of phase transformation in polycrystalline shape memory alloys," *International Journal of Plasticity,* vol. 32, no. 33, pp. 155-183, 2012.

[20] A. Duval, M. Haboussi and T. Ben Zineb, "Modelling of localization and propogation of phase transformation in superelastic SMA by a gradient nonlocal approach".

[21] E. A. Pieczyska, H. Tobushi and K. Kulasinski, "Development of transformation bands in TiNi SMA for various stress and strain rates studied by a fast and sensitive infrared

camera," *Smart Materials and Structures,* vol. 22, pp. 1-8, 2013.

[22] P. Schlosser, D. Favier, H. Louche and L. Orgeas, "Experimental characterization of NiTi SMAs thermomechanical behaviour using temperature and strain full-field measurements".

[23] S. Daly, G. Ravichandran and K. Bhattacharya, "Stress-induced martensitic phase transfomation in thin sheets of Nitinol," *Acta Materialia,* vol. 55, pp. 3593-3600, 2007.

[24] B. Reedlunn, S. Daly and J. Shaw, "Superelastic Shape Memory Alloy Cables: Part I - Isothermal Tension Experiments," *International Journal of Solids and Structures,* vol. 50, no. 20-21, pp. 3009-3026, 2013.

[25] P. Junker, "An accurate, fast and stable material model for shape memory alloys," *Smart Materials and Structures,* vol. 23, pp. 1-15, 2014.

[26] F. Auricchio and E. Sacco, "A temperature-dependent beam for shape-memory alloys: constitutive modelling, finite-element implementation and numerical simulations," *Computer Methods in Applied Mechanics and Engineering,* vol. 174, pp. 171-190, 1999.

[27] X. Gao, R. Qiao and L. C. Brinson, "Phase diagram kinetics for shape memory alloys: a robust finite element implementation," *Smart Materials and Structures,* vol. 16, pp. 2102-2115, 2007.

[28] C. Liang and C. Rogers, "One-dimensional thermomechanical constitutive relations for shape memory materials," *Journal of Intelligent Materials, Systems, and Structures,* vol. 1, pp. 207-234, 1990.

[29] L. Brinson and M. Panico, "Comments to the paper "Differential and integrated form consistency in 1-D phenomenological models for shape memory alloy constitutive behavior"," *International Journal of Solids and Structures,* vol. 46, pp. 217-220, 2009.

[30] K. Ikuta, M. Tsukamoto and S. Hirose, "Mathematical Model and Experimental Verification of Shape Memory Alloy for Designing Micro Actuator," in *Micro Electro Mechanical Systems (MEMS '91) Proceedings. An investigation of Micro Structures, Sensors, Actuators, Machines, and Robots IEEE,* Nara, 1991.

[31] Y. Ivshin and T. Pence, "A Thermomechanical Model for a one Variant Shape Memory Material," *Journal of Intelligent Material Systems and Strructures,* vol. 5, pp. 455-473, 1994.

[32] K. Boubaker, "A Model for Conjoint Shape Memory and Pseudo-Elastic Effects during Martensitic Transformation," *International Scholarly Research Network - Metallurgy,* vol.

2012, pp. 1-5, 2012.

[33] K. Tanaka and T. Mori, "The hardening of crystals by non-deforming particles and fibers," *Acta Metallurgica,* vol. 18, no. 8, pp. 931-941, 1970.

[34] M. Elahinia and M. Ahmadian, "An enhanced SMA phenomenological model: I. The shortcomings of the existing models," *Smart Materials and Structures,* vol. 14, pp. 1297-1308, 2005.

[35] M. Elahinia and M. Ahmadian, "An enhanced SMA phenomenological model: II. The experimental study," *Smart Materials and Structures,* vol. 14, pp. 1309-1319, 2005.

[36] N. Zotov, V. Marzynkevitsch and E. Mittemeijer, "Evaluation of kinetic equations describing the martensite-austenite phase transformation in NiTi shape memory alloys," *Journal of Alloys and Compounds,* vol. 616, pp. 385-393, 2014.

[37] B. Jaber, S. Mehrez and O. Ghazouani, "A 1D constitutive model for shape memory alloy using strain and temperature as control variables and including martensite reorientation and asymmetric behaviors," *Smart Materials and Structures,* vol. 23, pp. 1-10, 2014.

[38] K. Malukhin and K. Ehmann, "A model of the kinetics of the temperature-induced phase transformation in NiTi alloys and its experimental verification," *Journal of intelligent Material Systems and Structures,* vol. 23, no. 1, pp. 35-44, 2011.

[39] C. Liang and C. A. Rogers, "Design of Shape Memory Alloy Actuators," *Journal of Mechanical Design,* vol. 114, pp. 223-230, 1992.

[40] Pisces Conservation, "Growth II," 18 December 2010. [Online]. Available: http://www.pisces-conservation.com/growthhelp/index.html?richards_curve.htm. [Accessed 22 April 2015].

[41] D. Madill and D. Wang, "Modeling and L2-Stability of a Shape Memory Alloy Position Control System," *IEEE Transactions on Control Systems Technology,* vol. 6, no. 4, pp. 473-481, 1998.

[42] S. Zhu and Y. Zhang, "A thermomechanical constitutive model for superelastic SMA wire with strain-rate dependence," *Smart Materials and Structures,* vol. 16, pp. 1696-1707, 2007.

[43] J. Jayender, R. Patel, S. Nikumb and M. Ostojic, "Modeling and Control of Shape Memory Alloy Actuators," *IEEE Transactions on Control Systems Technology,* vol. 16, no. 2, pp.

279-287, 2008.

[44] O. Heintze, A Computationally Efficient Free Energy Model for Shape Memory Alloys - Experiments and Theory, Raleigh: North Carolina State University, 2004.

[45] S. Furst and S. Seelecke, "Modeling and experimental characterization of the stress, strain, and resistance of shape memory alloy actuator wire with controlled power input," *Journal of Intelligent Material Systems and Structures,* vol. 23, no. 11, pp. 1233-1247, 2012.

[46] O. Heintze and S. Seelecke, "A coupled thermomechanical model for shape memory alloys - From single crystal to polycrystal," *Materials Science and Engineering A,* vol. 2008, pp. 389-394, 2008.

[47] S. Seelecke, "Modeling the dynamic behavior of shape memory alloys," *International Journal of Non-Linear Mechanics,* vol. 37, pp. 1363-1374, 2002.

[48] S. Seelecke and I. Muller, "Shape memory alloy actuators in smart structures: Modeling and simulation," *ASME Applied Mechanics Review,* vol. 57, no. 1, pp. 23-46, 2004.

[49] V. Kumar, M. Morjaria and S. Mukherjee, "Numerical Integration of Some Stiff Constitutive Models of Inelastic Deformation," *Journal of Engineering Materials and Technology,* vol. 102, pp. 92-96, 1980.

[50] V. Banthia and S. Mukherjee, "On an Improved Time Integration Scheme for Stiff Constitutive Models of Inelastic Deformation," *Journal of Engineering Materials and Technology,* vol. 107, pp. 282-285, 1985.

[51] Dynalloy, Inc., "Technical Characteristics of Flexinol Actuator Wires," [Online]. Available: http://www.dynalloy.com/TechSheets.php. [Accessed 22 September 2014].

[52] T. Duerig, "Nitinol Ms vs Ni content," 2 October 2009. [Online]. Available: https://commons.wikimedia.org/wiki/File:Nitinol_Ms_vs_Ni_content.jpg. [Accessed 6 July 2016].

[53] D. Goldstein, R. Jones and R. Sery, "Method of Modifying the Transformation Temperature Range of TiNi Base Shape Memory Alloys". United States of America Patent 4283233, 11 August 1981.

[54] H. Sadiq, M. B. Wong, R. Al-Mahaidi and X. L. Zhao, "The effects of heat treatment on the recovery stresses of shape memory alloys," *Smart Materials and Structures,* vol. 19, pp. 1-7, 2010.

[55] W. Huang, "Shape Memory Alloys and their Application to Actuators for Deployable Structures," Peterhouse, Cambridge, 1998.

[56] J. Uchil, K. Mohanchandra, K. Mahesh and K. Ganesh Kumara, "Thermal and electrical characterization of R-phase dependence on heat-treat temperature in Nitinol," *Physica B,* vol. 253, pp. 83-89, 1998.

[57] J. Uchil, K. Ganesh Kumara and K. Mahesh, "Effect of thermal cycling on R-phase stability in a NiTi shape memory alloy," *Materials Science and Engineering A,* vol. 332, pp. 25-28, 2002.

[58] J. Uchil and et al, "Study of critical dependence of stable phases in Nitinol on heat treatment using electrical resistivity probe," *Materials Science and Engineering A,* vol. 251, pp. 58-63, 1998.

[59] S. Kim, E. Hawkes, K. Cho, M. Jolda, J. Foley and R. Wood, "Micro artificial muscle fiber using NiTi spring for soft robotics," in *The 2009 IEEE/RSJ Internatoinal Conference on Intelligent Robots and Systems,* St. Louis, 2009.

[60] N. B. Morgan and M. Broadley, "Taking the art out of Smart! - Forming processes and durability issues for the applciatoin of NiTi shape memory alloys in medical devices," in *Materials and Processes for Medical Devices Conference 2003,* Anaheim, 2004.

[61] N. Lewis, A. York and S. Seelecke, "Experimental charcterization of self-sensing SMA actuators under controlled convective cooling," *Smart Materials and Structures,* vol. 22, pp. 1-10, 2013.

[62] A. Sofla, D. Elzey and H. Wadley, "Cyclic degradation of antagonistic shape," *Smart Materials and Structures ,* vol. 17, pp. 1-6, 2008.

[63] X. D. Wu, J. S. Wu and Z. Wang, "The variation of electrical resistance of near stoichiometric NiTi during thermo-mechanic procedures," *Smart Materials and Structures,* vol. 8, pp. 574-578, 1999.

[64] S. Padula, "Thermomechanical methodology for stabilizing shape memory alloy (SMA) response". United States of America Patent 8409372, 2 April 2013.

[65] M. H. Wu, "Fabrication of Nitinol Materials and Components," in *Proceedings of the International Conference on Shape Memory and Superelastic Technologies*, Kunming, 2001.

[66] H. Y. Luo and E. W. Abel, "A comparison of methods for the training of NiTi two-way shape memory alloy," *Smart Materials and Structures,* vol. 16, pp. 2543-2549, 2007.

[67] H.-C. Kim, Y.-I. Yoo and J.-J. Lee, "Two-way shape memory effect induced by repetitive compressive loading cycles," *Smart Materials and Structures,* vol. 18, pp. 1-10, 2009.

[68] G. Kang and D. Song, "Review on structural fatigue of NiTi shape memory alloys: Pure mechanical and thermo-mechanical ones," *Theoretical and Applied Mechanics Letters,* vol. 5, pp. 245-254, 2015.

[69] "Thermomechanical cyclic loading and fatigue life characterization of nickel rich NiTi shape-memory alloy actuators," *Behavior and Mechanics of Multifunctional and Composite Materials,* vol. 6929, pp. 1-11, 2008.

[70] C. Chluba and et al, "Ultralow-fatigue shape memory alloy films," *Science,* vol. 348, no. 6238, pp. 1004-1007, 2015.

[71] S. W. Robertson, A. R. Pelton and R. O. Ritchie, "Mechanical fatigue and fracture of Nitinol," *International Materials Reviews,* vol. 57, no. 1, pp. 1-36, 2012.

[72] H. Tobushi, Y. Ohashi, T. Hori and H. Yamamoto, "Cyclic Deformation of TiNi Shape-memory Alloy Helical Spring," *Experimental Mechanics,* pp. 304-308, 1992.

[73] E. Hornbogen, "Review of Thermo-mechanical fatigue of shape memory alloys," *Materials Science,* vol. 39, pp. 385-399, 2004.

[74] D. Lagoudas, "Fatigue Life and Dynamic Respnse of SMA Actuators," US Air Force, Arlington, 2004.

[75] D. Lagoudas and O. Bertacchini, "Surface Crack Development in Transformation Induced Fatigue of SMA Actuators," in *Residual Stress and Its Effects on Fatigue and Fracture*, Alexandroupolis, Springer, 2006, pp. 209-222.

[76] O. Bertacchini and et al, "Thermomechanical cyclic loading and fatigue life characerization of nickel rich NiTi shape-memory alloy actuators," in *Proceedings of SPIE 6929, Behavior and Mechanics of Multifunctional and Composite Materials 2008*, 2008.

[77] J. Schick, "Transformatoin Induced Fatigue of Ni-Rich Shape Memory Alloy Actuators," Texas A&M University, College Station, 2009.

[78] TOKI CORPORATION, "BMF Catalog," [Online]. Available: http://www.toki.co.jp/biometal/download/downloadfiles/BMF_Catalog_140217.pdf.

[Accessed 14 March 2016].

[79] A. Wick, X. Gong, J. Fino, J. Sheriff and A. Pelton, "Bending Fatigue Characteristics of Nitinol," in *Proceedings of SMST-2004*, Baden-Baden, 2006.

[80] M. M. Patel, "Characterizing Fatigue Response of Nickel-Titanium Alloys by Rotary Beam Testing," *Journal of ASTM International,* vol. 4, no. 6, pp. 1-11, 2007.

[81] R. Gilbertson and C. de Miranda, Muscle WIres Project Book, 3rd Revised Edition, Mondo-Tronics, 2000.

[82] S. Shabalovskaya, J. Anderegg and J. Van Humbeeck, "Critical overview of Nitinol surfaces and their modifications for medical applications," *Acta Biomaterialia,* vol. 4, pp. 447-467, 2008.

[83] J. Kawakita, M. Stratmann and A. W. Hassel, "High Voltage Pulse Anodizing of a NiTi Shape Memory Alloy," *Journal of the Electrochemical Society,* vol. 154, no. 6, pp. C294-C298, 2007.

[84] L. Zhu, C. Trepanier, A. R. Pelton and J. Fino, "Oxidation of Nitinol and Its Effect on Corrosion Resistance," in *Medical Device Materials: Proceedings of the Materials and Processes for Medical Devices Conference,* Anaheim, 2003.

[85] S. Shabalovskaya, G. Rondelli, J. Anderegg, B. Simpson and S. Budko, "Effect of Chemical Etching and Aging in Boiling Water on the Corrosion REsitsance of Nitinol Wires with Black Oxide Resulting from Manufacturing Process," *Journal of Biomedical Materials Research Part B: Applied Biomaterials,* vol. 66B, pp. 331-340, 2003.

[86] C. L. Chu, C. Y. Chung and P. K. Chu, "Surface oxidaton of NiTi shape memory alloy in a boiling aqeous solution containing hydrogen peroxide," *Materials Science and Engineering A,* vol. 417, pp. 104-109, 2006.

[87] J. Gerald, "Nitinol heater elements". United States of America Patent 6410886, 25 June 2002.

[88] N. Sheiko and et al, "PEEK (poyether-ether-ketone)-coated nitinol wire: Film stability for biocompatibility applications," *Applied Surface Science,* vol. 389, pp. 651-665, 2016.

[89] V. Zatsiorsky and B. Prilutsky, Biomechanics of Skeletal Muscles, Champaign: Human Kinetics, 2012.

[90] D. Stoeckel, "Shape Memory Actuators for Automotive Applications," *Materials and Design,* vol. 11, no. 6, pp. 302-307, 1990.

[91] J. Jami, M. Leary, A. Subic and M. Gibson, "A review of shape memory alloy research, applications, and opportunities," *Materials and Design,* vol. 56, pp. 1078-1113, 2014.

[92] N. Morgan, "Mdeical shape memory alloy applications - the market and its products," *Materials Science and Engineering A,* vol. 378, pp. 16-23, 2004.

[93] C. Mavroidis, C. Pfeiffer and M. Mosley, "Conventional Actuators, Shape Memory Alloys, and Electrorheological Fluids," *Automation, Miniature Robotics and Sensors for Nondestructive Testing and Evaluation,* vol. 4, pp. 1-26, 2000.

[94] H. S. Tzou, H. -J. Lee and S. M. Arnold, "Smart Materials, Precision Sensors/Actuators, Smart Structures, and Structronic Systems," *Mechanics of Advanced Materials and Structures,* vol. 11, pp. 367-393, 2004.

[95] J. Jani, M. Leary, A. Subic and M. Gibson, "A review of shape memory alloy research, applications, and opportunities," *Materials and Design,* vol. 56, pp. 1078-1113, 2014.

[96] I. Chopra, "Review of State of Art of Smart Structures and Integrated Systems," *AIAA Journal,* vol. 40, no. 11, pp. 2145-2187, 2002.

[97] S. Toru, "Fast and Accurate Position Control of Shape Memory Alloy Actuators," Universite Paris, Paris, 2008.

[98] W. Buehler and R. Wiley, "Nickel-Base Alloys". United States of America Patent 3174851, 23 March 1965.

[99] Y. Bar-Cohen, "Artificial Muscles using Electoactive Polymers (EAP): Capabilities, Challenges and Potential," Jet Propulsion Laboratory, National Aeronautics and Space Administration, Pasadena, 2005.

[100] C. Haines, R. Baughman and e. al, "Artificial Muscles from Fishing Line and Sewing Thread," *Science,* vol. 343, pp. 868-872, 2014.

[101] J. E. Huber, N. A. Fleck and M. F. Ashby, "The selection of mechanical actuators based on performance indices," *Proceedings of the Royal Society,* vol. 453, pp. 2185-2205, 1997.

[102] J. M. Hollerbach, I. W. Hunter and J. Ballantyne, "A Comparative Analysis of Actuator Technologies for Robotics," in *The Robotics Review 2*, vol. 2, Cambridge, MIT Press,

1992, pp. 299-342.

[103] D. Hartl and D. Lagoudas, "Aerospace Applications of Shape Memory Alloys," *Proceedings of the Institue of Mechanical Engineers, Part G. Journal of Aerospace Engineering,* vol. 221, no. 4, pp. 535-552, 2007.

[104] W. Coral, C. Rossi, J. Colorado, D. Lemus and A. Barrientos, "SMA-Based Msucle-Like Actuation in Biologically Inspired Robots: A State of the Art Review," in *Smart Actuation and Sensing Systems - Recent Advances and Future Challenges,* InTech, 2012, pp. 53-82.

[105] M. Wu and L. M. Schetky, "Industrial Applications for Shape Memory Alloys," in *Proceedings of the International Conference on Shape Memory and Superelasticity*, Pacific Grove, 2000.

[106] J. Van Humbeeck, "Non-medical applications of shape memory alloys," *Materials Science and Engineering A,* vol. 273, no. 275, pp. 134-148, 1999.

[107] F. Calkins and J. Mabe, "Shape Memory Alloy Based Morphing Aerostructures," *Journal of Mechanical Design,* vol. 132, pp. 111012:1-7, 2010.

[108] A. Nespoli and et al, "The high potential of shape memory alloys in developing miniature mechanical devices: A review on shape memory alloy mini actuators," *Sensors and Actuators A: Physical,* vol. 158, pp. 149-160, 2010.

[109] M. Sreekumar and et al, "Critical Review of current trends in shape memory alloy actuators for intelligent robots," *Industrial Robots: An International Journal,* vol. 34, no. 4, pp. 285-294, 2007.

[110] B. Stirbu, "Shape Memory Alloy Actuators in Robotics: A Review," Royal Military Academy, 2015.

[111] C. Liu, H. Qin and P. T. Mather, "Review of progress in shape memory polymers," *Journal of Materials Chemistry,* vol. 17, pp. 1543-1558, 2007.

[112] J. Ueda and C. Stefanini, "Workshop on Biologically Inspired Actuation," in *IEEE International Conference on Robotics and Automation*, Shanghai, 2011.

[113] A. Nespoli and et al, "The high potential of shape memory alloys in developing miniature mechanical devices: A review on shape memory alloy mini-actuators," *Sensors and Actuators A: Physical,* vol. 158, pp. 149-160, 2010.

[114] S. Seok, C. Onal, K.-J. Cho, R. Wood, D. Rus and S. Kim, "Meshworm: a Peristaltic Soft Robot With Antagonistic Nickel Titanium Coil Actuators," *IEEE/ASME Transactions on Mechatronics*, 2012.

[115] A. Nespoli, C. Biffi, R. Casati, F. Passaretti, A. Tuissi and E. Villa, "New Developments on Mini/Micro Shape Memory Actuators," InTech, 2012.

[116] J. Swensen and A. Dollar, "Optimization of Parallel Spring Antagonists for Nitinol Shape Memory Alloy Actuators," in *IEEE International Conference on Robotics and Automation (ICRA)*, Hong Kong, 2014.

[117] C. Haberland, M. Elahinia, J. Walker, H. Meier and J. Frenzel, "On the development of high quality NiTi shape memory and pseudoelastic parts by additive manufacturing," *Smart Materials and Structures*, vol. 23, pp. 1-13, 2014.

[118] D. Grant, Accurate and Rapid Control of Shape Memory Alloy Actuators, Montreal: McGill University, 1999.

[119] D. Grant and V. Hayward, "Variable Structure Control of Shape Memory Alloy Actuators," in *IEEE International Conference on Robotics and Automation*, Nagoya, 1995.

[120] J. Anadon, "Large Force Shape Memory Alloy Linear Actuator," University of Florida, Gainesville, 2002.

[121] A. Konstantinos and A. Tzes, "Design of an Innovative Prosthetic Hand with Compact Shape Memory Alloy Actuators," in *21st Mediterranean Conf. on Control and Automation (MED)*, Crete, 2013.

[122] A. Villoslada and et al, "High-displacement fast-cooling flexible Shape Memory Alloy actuator: application to an anthropomorphic robotic hand," in *14th IEEE-RAS Int. Conf. on Humanoid Robots (Humanoids)*, Madrid, 2014.

[123] D. Josephine Sevarani Ruth, K. Dhanalakshmi and S. Sunjai Nakshatharan, "Bidirectional angular control of an integrated sensor/actuator shape memory alloy based system," *Measurement*, vol. 69, pp. 210-221, 2015.

[124] K. Saito and et al, "Miniaturized Rotary Actuators Using Shape Memory Alloy for Insect-Type MEMS Microrobot," *Micromachines*, vol. 7, no. 58, pp. 1-11, 2016.

[125] M. Ho and J. Desai, "Modeling, Characterization and Control of Antagonist SMA Springs for use in a Neurosurgical Robot," in *IEEE International Conference on Robotics and*

Automation (ICRA), Karlsruhe, 2013.

[126] S.-B. Choi, "Position control of a single-link mechanism activated by shape memory alloy springs: experimental results," *Smart Materials and Structures,* vol. 15, pp. 51-58, 2006.

[127] G. Song, "Design and control of a Nitinol wire actuated rotary servo," *Smart Materials and Structures,* vol. 16, pp. 1796-1801, 2007.

[128] C. Y. Liu and W. Liao, "A Snake Robot Using Shape Memory Alloys," in *Proc. IEEE Int. Conf. on Robotics and Biomimetics,* China, 2004.

[129] D. Mandru, I. Lungu, S. Noveanu and O. Tatar, "Shape Memory Alloy Wires as Actuators for a Minirobot," in *IEEE Int. Conf. on Automation Quality and Testing Robots*, Cluj-Napoca, 2010.

[130] K. Yang and C. L. Gu, "A novel robot hand with embedded shape memory alloy actuators," *Proceedings of the Institution of Mechanical Engineers - Part C - Journal of Mechanical Engineering Science,* vol. 216, pp. 737-745, 2002.

[131] S. Furst, R. Hangekar and S. Seelecke, "Practical Implementation of Resistance Feedback Measurement for Position Control of a Flexible Smart Inhaler Nozzle," in *Proc. ASME 2010 Conf. on Smart Materials, Adaptive Structures and Intelligent Systems*, Philadelphia, 2010.

[132] S.-H. Song and et al, "35 Hz shape memory alloy actuator with bending-twisting mode," *Scientific Reports,* vol. 6, no. 21118, pp. 1-13, 2016.

[133] B. Konh, N. Datla and P. Hutapea, "Feasibility of Shape Memory Alloy Wire Actuation for an Active Steerable Cannula," *ASME J. Medical Devices,* vol. 9, pp. 1-11, 2015.

[134] G. Song, B. Kelly and B. Agrawal, "Active position control of a shape memory alloy wire actuated composite beam," *Smart Materials and Structures,* vol. 9, pp. 711-716, 2000.

[135] V. Bundhoo, E. Haslam, B. Birch and E. Park, "A shape memory alloy-based tenodn-driven actuation system for biomimetic artificial fingers, part I: design and evaluation," *Robotica,* pp. 1-16, 2008.

[136] K. Cho, J. Rosemarin and H. Asada, "Design of Vast DOF Artificial Muscle Actuators with a Cellular Array Structure and its Application to a Five-Fingered Robotic Hand," in *Proceedings of the 2006 IEEE International Conference on Robotics and Automation*, Orlando, 2006.

[137] K. O'Toole, "A Methodology Towards Comprehensive Evaluation of Shape Memory Alloy Actuators for Prosthetic Finger Design," Dublin Institute of Technology, Dublin, 2011.

[138] F. Simone, A. York and S. Seelecke, "Design and Fabrication of a Three-Finger Prosthetic Hand using SMA muscle wires," in *Proceedings of SPIE 9429 Bioinspiration, Biomimetics, and Bioreplication*, 2015.

[139] G. Gilardi, E. Haslam, V. Bundhoo and E. Park, "A shape memory alloy-based tendon-driven actuation system for biomimetic artificial fingers, part II: modelling and control," *Robotica,* vol. 28, pp. 675-687, 2010.

[140] K.-J. Cho and H. Asada, "Multi-Axis SMA Actuator Array for Driving Anthropomorphic Robot Hand," in *Proc. of the 2005 IEEE Int. Conf. on Robotics and Automation*, Barcelona, 2005.

[141] M. Elahinia, E. Esfahani and S. Wang, "Control of SMA Systems: Review of the State of the Art," in *Shape Memory Alloys: Manufacture, Properties and Applications*, Nova, 2010, pp. 49-68.

[142] M. Sreekumar, M. Singaperumal, T. Nagarajan, M. Zoppi and R. Molfino, "Recent advances in nonlinear control technologies for shape memory alloy actuators," *Journal of Zhejiang University SCIENCE A,* vol. 8, no. 5, pp. 818-829, 2007.

[143] Y. H. Teh and R. Featherstone, "Frequency response analysis of shape memory alloy actuators," in *Proceedings of the International Conference on Smart Materials and Nanotechnology in Engineering*, Harbin, 2007.

[144] R. B. Gorbet, K. A. Morris and R. C. Chau, "Mechanism of bandwidth improvement in passively cooled SMA position actuators," *Smart Materials and Structures,* vol. 18, pp. 1-9, 2009.

[145] T. Evdaimon, M. Sfakiotakis and D. Tsakiris, "A Closed-Loop Position Control Scheme for SMA-Actuated Joints," in *22nd Mediterranean Conference on Control and Automation (MED)*, Palermo, 2014.

[146] Y. H. Teh and R. Featherstone, "An Architecture for Fast and Accurate Control of Shape Memory Alloy Actuators," *The International Journal of Robotics Research,* vol. 27, pp. 595-611, 2008.

[147] Y. H. Teh, "Fast, Accurate Force and Position Control of Shape Memory Alloy Actuators," ANU College of Engineering and Computer Science, Canberra, 2008.

[148] M. Zakerzadeh and H. Sayyaadi, "Precise position control of shape memory alloy actuator using inverse hysteresis model and model reference adaptive control system," *Mechatronics,* vol. 23, pp. 1150-1162, 2013.

[149] J. Jayender, R. Patel, S. Nikumb and M. Ostojic, "Modeling and Control of Shape Memory Alloy Actuators," *IEEE Transactions on Control Systems Technology,* vol. 16, no. 2, pp. 279-287, 2008.

[150] E. Ayvali and J. Desai, "Pulse width modulation-based temperature tracking for feedback control of a shape memory alloy actuator," *Journal of Intelligent Material Systems,* vol. 25, no. 6, pp. 720-730, 2014.

[151] D. Grant and V. Hayward, "Constrained Force Control of Shape Memory Alloy Actuators," in *Proceedings of the 2000 IEEE International Conference on Robotics and Automation*, San Francisco, 2000.

[152] M. Elahinia and H. Ashrafiuon, "Nonlinear Control of a Shape Memory Alloy Actuated Manipulator," *Journal of Vibration and Acoustics (Transactions of the ASME),* vol. 124, pp. 566-575, 2002.

[153] E. Ayvali and J. Desai, "Towads a Dsicretely Actuated Steerable Cannula," in *2012 IEEE International Conference on Robotics and Automation*, Saint Paul, 2012.

[154] G. Song, V. Chaudhry and C. Batur, "A Neural Network Inverse Model for a Shape Memory Alloy Wire Actuator," *Journal of Intelligent Material Systems and Structures,* vol. 14, pp. 371-377, 2003.

[155] A. Kazuhiko, S. Aramaki and K. Yanagisawa, "Feedback Linearization for SMA (Shape Memory Alloy)," in *Society of Instrument and Control Engineers*, Sapporo, 1995.

[156] R. Damle, R. Lashlee, V. Rao and F. Kern, "Indentification and robust control of smart structures using artificial neural networks," *Smart Materials and Structures,* vol. 3, pp. 35-46, 1994.

[157] M. Elahinia, T. M. Seigler, D. J. Leo and M. Ahmadian, "Nonlinear Stress-based Control of a Rotary SMA-actuated Manipulator," *Intelligent Material Systems and Structures,* vol. 15, pp. 495-508, 2004.

[158] K. K. Ahn and B. K. Nguyen, "Position Control of Shape Memory Alloy Actuators Using Self Tuning Fuzzy PID Controller," *International Journal of Control, Automation, and Systems,* vol. 4, no. 6, pp. 756-762, 2006.

[159] G. Song, V. Chaudhry and C. Batur, "Precision tracking control of shape memory alloy actuators using neural networks and a sliding-mode based robust controller," *Smart Materials and Structures,* vol. 12, pp. 223-231, 2003.

[160] G. Song and N. Ma, "Robust control of a shape memory alloy wire actuated flap," *Smart Materials and Structures,* vol. 16, pp. N51-N57, 2007.

[161] E. Shameli, A. Alasty and H. Salaarieh, "Stability analysis and nonlinear cotnrol of a miniature shape memory alloy actuator for precise applications," *Mechatronics,* vol. 15, pp. 471-486, 2005.

[162] K. Ikuta, M. Tsukamoto and S. Hirose, "Shape Memory Alloy Servo Actuator System with Electric Resistance Feedback and Application for Active Endoscope," in *Proceedings of 1988 International Conference on Robotics and Automation*, Philadelphia, 1988.

[163] V. Novak, P. Sittner, G. Dayananda, B. Braz-Fernandes and K. Mahesh, "Electrical resistance variation of NiTi shape memory alloy wires in thermomechanical tests: Experiments and simulation," *Materials Science and Engineering A,* Vols. 481-482, pp. 127-133, 2008.

[164] H. Song, E. Kubica and R. Gorbet, "Resistance Modelling of SMA Wire Actuators," in *International Workshop at the Smart Materials, Structures & NDT in Aerospace*, Montreal, 2011.

[165] V. Antonucci, G. Faiella, M. Giordano, F. Mennella and L. Nicolais, "Electrical resistivity study and characterization during NiTi phase transformation," *thermochimica acta,* vol. 462, pp. 64-69, 2007.

[166] J.-J. Zhang, Y.-H. Yin and J.-Y. Zhu, "Electrical Resistivity-Based Study of Self-Sensing Properties for Shape Memory Alloy-Actuated Artificial Muscle," *Sensors,* vol. 13, pp. 12958-12974, 2013.

[167] D. Cui, G. Song and H. Li, "Modeling of the electrical resistance of shape memory alloy wires," *Smart Materials and Structures,* vol. 19, pp. 1-8, 2010.

[168] M. Pozzi and G. Airoldi, "The electrical transport properties of shape memory alloys," *Materials Science and Engineering A,* vol. A273, no. 275, pp. 300-304, 1999.

[169] Z. He, K. Gall and L. Brinson, "Use of Electrical Resistance Testing to Redefine the Transformatin Kinetcs and Phase Diagram for Shape Memory Alloys," *Metallurgical and Material Transactions A,* vol. 37A, pp. 579-587, 2006.

[170] H. Kim, Y. Han, D.-y. Lee, J.-I. Ha and K.-J. Cho, "Sensorless displacement estimation of a shape memory alloy coil spring actuator using inductance," *Smart Materials and Structures,* vol. 22, pp. 1-8, 2013.

[171] C.-C. Lan and C.-H. Fan, "An accurate self-sensing method for the control of shape memory alloy actuated flexures," *Sensors and Actuators A,* vol. 163, pp. 323-332, 2010.

[172] T.-M. Wang, Z.-Y. Shi, D. Liu, C. Ma and Z.-H. Zhang, "An accurately Controlled Antagonist Shape Memory Alloy Actuator with Self Sensing," *Sensors,* vol. 12, pp. 7682-7700, 2012.

[173] R. Josephine, N. Sunjai and K. Dhanalakshmi, "Differential resistance feedback control of a self-sensing shape memory alloy actuated system," *International Society of Automation Transactions,* vol. 53, pp. 289-297, 2014.

[174] N. Ma, G. Song and H.-J. Lee, "Position control of shape memory alloy actuators with internal electrical resistance feedack using neural networks," *Smart Materials and Structures,* vol. 13, pp. 777-783, 2004.

[175] M. Dib, R. Gorbet, E. Kubica, X. Gao, A. Browne and N. Johnson, "Adaptive SMA actuator priming using resistance feedback," *Smart Materials and Structures,* vol. 20, pp. 1-14, 2011.

[176] R. Featherstone and Y. H. Teh, "Improving the Speed of Shape Memory Alloy Actuators by Faster Electrical Heating," *Experimental Robotics IX, Springer Tracts in Advanced Robotics,* vol. 21, pp. 67-76, 2006.

[177] M. Elahinia, M. Ahmadian and H. Ashrafiuon, "Design of a Kalman filter for rotary shape memory alloy actuators," *Smart Materials and Structures,* vol. 13, pp. 691-697, 2004.

[178] M. Elahinia, "Application of the extended Kalman filter to control of a shape memory alloy arm," *Smart Materials and Structures,* vol. 15, pp. 1370-1384, 2006.

[179] R. Gorbet and D. Wang, "A Dissipativity Approach to Stability of a Shape Memory Alloy Position Control System," *IEEE Transactions on Control Systems Technology,* vol. 6, no. 4, pp. 554-562, 1998.

[180] R. B. Gorbet, D. Wang and K. A. Morris, "Preisach Model Indentification of a Two-Wire SMA Actuator," in *Proceedings of the 1998 IEEE International Conference on Robotics and Automation*, Leuven, 1998.

[181] H. Li, C.-X. Mao and J.-P. Ou, "Strain Self-Sensing Property and Strain Rate Dependent Constitutive Model of Austenitic Shape Memory Alloy - Experiment and Theory," *Journal of Materials in Civil Engineering ASCE,* vol. 17, no. 6, pp. 676-685, 2005.

[182] J. Crews and R. Smith, "Quantification of parameter and model uncertainty for shape memory alloy bending actuators," *J. Intelligent Material Systems and Structures,* vol. 25, no. 2, pp. 229-245, 2014.

[183] M. Thrasher and et al, "Efficiency analysis of shape memory alloy actuators," *Smart Materials and Structures,* vol. 3, pp. 226-234, 1994.

[184] H. Yin, Y. He and Q. Sun, "Effect of deformation frequency on temperature and stress oscillations in cyclic phase transition of NiTi shape memory alloy," *J. Mechanics and Physics of Solids,* vol. 67, pp. 100-128, 2014.

[185] X. Xie and et al, "Observation on rate-dependent cyclic transformation domain of super-elastic NiTi shape memory alloy," *Material Science & Engineering A,* vol. 671, pp. 32-47, 2016.

[186] C. Wells and D. Chan, "High-Side Voltage-to-Current (V-I) Converter," Texas Instruments, June 2013. [Online]. [Accessed 16 May 2017].

[187] S. Schaal, C. Atkeson and S. Vijayakumar, "Real-time robot learning with locally weighted statistical learning," in *lntemational Conference on Robotics & Automation*, San Francisco, CA, 2000.

[188] T. B. Moeslund, A. Hilton and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Computer Vision and Image Understanding,* vol. 104, no. 2-3, pp. 90-126, 2006.

[189] R. R. Ma and A. M. Dollar, "On Dexterity and Dexterous Manipulation," in *International Conference on Advanced Robotics*, Tallinn, Estonia, 2011.

[190] W. A. Lewinger, M. S. Branicky and R. D. Quinn, "Insect-inspired, Actively Compliant Hexapod Capable of Object Manipulation," in *Climbing and Walking Robots*, Berlin, 2006.

[191] C.-C. Lan and C.-H. Fan, "An accurate self-sensing method for the control of shape memory alloy actuated flexures," *Sensors and Actutators A,* vol. 163, pp. 323-332, 2010.

[192] T.-M. Wang, Z.-Y. Shi, D. Liu, C. Ma and Z.-H. Zhang, "An Accurately Controlled Antagonist Shape Memory Alloy Actuator with Self Sensing," *Sensors,* vol. 12, pp. 7682-

7700, 2012.

[193] Z. He, K. Gall and L. Brinson, "Use of Electrical Resistance Testing to Redefine the Transformation Kinetics and Phase Diagram for Shape Memory Alloys," *Metallurgical and Material Transactions A,* vol. 37A, pp. 579-587, 2006.

[194] J. Zhang and Y. Yin, "SMA-based bionic integration design of self-sensor-actuator-structure for artificial skeletal muscle," *Sensors and Actuators A: Physical,* vol. 181, pp. 94-102, 2012.

[195] C.-C. Lan, C.-M. Lin and C.-H. Fan, "A Self-Sensing Microgripper Module With Wide Handling Ranges," *IEEE/ASME Transactions on Mechatronics,* vol. 16, no. 1, pp. 141-150, 2011.

[196] S.-H. Liu, T.-S. Huang and J.-Y. Yen, "Tracking Control of Shape-Memory-Alloy Actuators Based on Self-Sensing Feedback and Inverse Hysteresis Compensation," *Sensors,* vol. 10, pp. 112-127, 2010.

[197] T. Georges, V. Brailovski and P. Terriault, "Characterization and design of antagonistic shape memory alloy actuators," *Smart Materials and Structures,* vol. 21, pp. 1-8, 2012.

[198] M. T. a. S. H. K. Ikuta, "Shape Memory Alloy Servo Actuator System with Electric Resistance Feedback and Application for Active Endoscope," in *Proceedings of 1988 International Conference on Robotics and Automation*, Philadelphia, 1988.

[199] K. Ikuta, M. Tsukamoto and S. Hirose, "Mathematical Model and Experimental Verification of Shape Memory Alloy for Designing Micro Actuator," in *Micro Electro Mechanical Systems*, Nara, 1991.

[200] Dynalloy, Inc., "Technical Characteristics of Flexinol Actuator Wires," Irvine, CA.

[201] J. Ditman, L. Bergman and T.-C. Tsao, "The design of extended bandwidth shape memory alloy actuators," *Journal of Intelligent Material Systems and Structures,* vol. 7, no. 6, pp. 635-645, 1996.

[202] S. Nakshatharan, D. Kaliaperumal and D. Ruth, "Effect of stress on bandwidth of antagonistic shape memory alloy actuators," *Journal of Intelligent Material Systems and Structures,* vol. 7, no. 2, pp. 153-165, 2016.

[203] C. V. G. a. C. B. Song, "A Neural Network Inverse Model for a Shape Memory Alloy Wire Actuator," *Journal of Intelligent Material Systems and Structures,* vol. 14, pp. 371-

377, 2003.

[204] J. H. C. a. S. S. S. J. Furst, "Numerical and experimental analysis of inhomogeneities in SMA wires induced by thermal boundary conditions," *Continuum Mechanical Thermodynamics*, vol. 10, no. 24, p. 497, 2012.

[205] G. Franklin, D. Powell and A. Emani-Naeini, Feedback Control of Dynamic Systems, Upper Saddle River: Pearson Education, 2006.

[206] G. Franklin, D. Powell and M. Workman, Digital Control of Dynamic Systems, Menlo Park: Addison Wesley Longman, Inc., 1998.

[207] C.-T. Chen, Linear System Theory and Design, New York: Oxford University Press, 2013.

[208] J.-J. Slotine and W. Li, Applied Nonlinear Control, Englewood Cliffs: Prentice Hall, 1991.

[209] J. Shamma, "An Overview of LPV Systems," in *Control of Linear Parameter Varying Systems with Applications*, New York, Springer, 2012, pp. 3-26.

[210] C. Mehendale and K. Grigoriadis, "Hysteresis Compensation Using LPV Gain-scheduling," in *Proc. 2004 American Control Conf.*, Boston, 2004.

[211] H. Ashrafiuon and V. R. Jala, "Sliding Mode Control of Mechanical Systems Actuated by Shape Memory Alloy," *J. Dynamic Systems, Measurement, and Control*, vol. 131, pp. 1-6, 2009.

[212] A. Kilicarslan, K. Grigoriadis and G. Song, "Compensation of hysteresis in a shape memory alloy wire system using linear parameter-varying gain scheduling control," *IET Control Theory and Applications*, vol. 8, no. 17, pp. 1875-1885, 2014.

[213] C. Mehendale and K. Grigoriadis, "Hysteresis Compensation Using LPV Gain-scheduling," in *Proc. 2004 American Control Conference*, Boston, 2004.

[214] D. J. Leith and W. E. Leithead, "Gain-scheduled and nonlinear systems: dynamic analysis by velocity-based linearization families," *International Journal of Control*, vol. 70, no. 2, pp. 289-317, 1998.

[215] D. J. Leith and W. E. Leithead, "Modelling, Analysis & Design by Velocity-based Linearisation Families," Department of Electronic & Electrical Engineering, University of Strathclyde, Glasgow, 1998.

[216] M. Halas, M. Huba and K. Zakova, "The Exact Velocity Linearization Method," in *A*

Proceedings volume from the 2nd IFAC Conference, Bratislava, 2003.

[217] J. Hung, W. Gao and J. Hung, "Variable Structure Control: A Survey," *IEEE trans. on Industrial Electronics*, vol. 40, no. 1, pp. 2-22, 1993.

[218] Universita' Degli Studi Di Cagliari, "A Quick Introduction to Sliding Mode Control and its Applications," [Online]. Available: http://www.diee.unica.it/~eusai/BOSIO/. [Accessed 5 June 2017].

[219] C. Edwards and S. Spurgeon, Sliding Mode Control: Theory and Applications, Bristol: Taylor & Francis Inc., 1998.

[220] D. Grant, "Accurate and rapid control of shape memory alloy actuators," Centre for Intelligent Machines, McGill University, Montreal, 1999.

[221] D. Grant and V. Hayward, "Variable Structure Control of Shape Memory Alloy Actuators," in *IEEE International Conference on Robotics and Automation*, Nagoya, 1995.

[222] G. Song, N. Ma, H.-J. Lee and S. Arnold, "Design and control of a proof-of-concept variable area exhaust nozzle using shape-memory alloy actuators," *Smart Materials and Structures*, vol. 16, pp. 1342-1347, 2007.

[223] G. Song, B. Kelly and B. Agrawal, "Active position control of a shape memory alloy wire actuated composite beam," *Smart Materials and Structures*, vol. 9, pp. 711-716, 2000.

[224] G. Song, "Design and control of a Nitinol wire actuated rotary servo," *Smart Materials and Structures*, vol. 16, pp. 1796-1801, 2007.

[225] H. Ashrafiuon and J. V. Reddy, "Model-Based Sliding Mode Control for a Robot with SMA Actuators," in *Proc. ASME Design Engineering Technical Conf.*, Long Beach, 2005.

[226] J. Crews and R. Smith, "Quantification of parameter and model uncertainty for shape memory alloy bending actuators," *J. Intelligent Material Systems*, vol. 25, no. 2, pp. 229-245, 2013.

[227] C. Moler and C. Van Loan, "Nineteen Dubious Ways to Compute the Matrix Exponential, Twenty-Five Years Later," *SIAM Review*, vol. 45, no. 1, pp. 1-46, 2003.

[228] C. Van Loan, "Computing Integrals Involving the Matrix Exponential," *IEEE Transactions on Automatic Control*, vol. 23, no. 3, pp. 395-404, 1978.

[229] C.-T. Chen, Linear System Theory and Analysis, 4th Ed., New York: Oxford University

Press, 2013.

[230] J. Martins, P. Sturdza and J. Alonso, "The Complex-Step Derivative Approximation," *ACM Transactions on Mathematical Software,* vol. 29, no. 3, pp. 245-262, 2003.

[231] C.-T. Chen, Linear System Theory and Design, New York: Oxford University Press, 2013.

Appendix A: Control and Estimation Review

The control of linear state-space systems is well-organized now and made easy by computer pole-placement tools such as MATLAB. The easiest way to implement controllers that track a changing step reference, 'servo' controllers, is by imbedding an integrator into the model and placing the poles of that new model. For notational consistency and process clarity, the design of general linear output-feedback is presented, leveraging computer aided methods without detailed explanation of their algorithms. The concepts of observability and controllability are discussed. An important extension of this method is addition of integral action which increases the system type, eliminates steady state error to step changes in reference, and simplifies the input of reference values. It is shown that the Robust Servo technique is a powerful technique for higher order system control. Linearization methods are shown, as well as conversion between continuous and discrete representations. Finally, estimators to determine the entire state from the measured outputs are introduced.

Deterministic/Stochastic State-Space Models

The state-space form of differential equations provides an effective structure for implementing the navigation filter. A general state-space model with known deterministic inputs, stochastic disturbances, and stochastic measurement errors consists of the following equations:

$$\dot{x} = A_c x + B_c u + B_w w$$

$$y = C_c x + D_c u + D_v v$$

The elements are defined:

x – state vector $\in \mathbb{R}^{nx1}$
$\dot{x}$ - time derivative of the state $\in \mathbb{R}^{nx1}$
u - vector of known inputs $\in \mathbb{R}^{mx1}$
y - vector of measurements $\in \mathbb{R}^{kx1}$
w - noise (disturbance) on the state derivative $\in \mathbb{R}^{lx1}$
v - noise on the measurement of the state $\in \mathbb{R}^{jx1}$
A_c - the state-transition matrix $\in \mathbb{R}^{nxn}$
B_c - input matrix $\in \mathbb{R}^{nxm}$
B_w - a transformation on disturbance inputs $\in \mathbb{R}^{nxl}$
C_c - the measurement matrix $\in \mathbb{R}^{kxn}$
D_c - the feed-forward matrix $\in \mathbb{R}^{kxm}$
D_v - a transformation on the measurement noise $\in \mathbb{R}^{kxj}$

Typically some components of the above matrices are not needed, and the model reduces to

$$\dot{x} = A_c x + B_c u + B_w w$$
$$y = C_c x + v$$

Where the subscript $(*)_c$ denotes this is a continuous-time model. Also, usually the assumption is made that the noise can be approximated by the normal distribution, so

$$w \sim N(0, Q_c)$$
$$v \sim N(0, R_c)$$

Where Q_c and R_c are the continuous-time disturbance covariance, and continuous-time measurement noise covariance, respectively.

Many sampled data systems and methods (including the Kalman filter) are traditionally specified using the discrete state space form which uses difference equations instead of differential equations

$$x_{k+1} = A_d x_k + B_d u + w_k$$
$$y_k = C_d x_k + v_k$$
$$w_k \sim N(0, Q_d)$$
$$v_k \sim N(0, R_d)$$

where subsequent k sample times are separated by a non-zero step in time of length Δt. Note that the state transition now solves for the state a step from now, not the derivative of the state. The transformation from continuous-time to discrete-time equations is the solution to the differential equations after a period of Δt seconds has elapsed. It is usually assumed that the input is essentially constant during this interval (it may actually be constant in physical control systems), and thus the equations can be transformed using the following:

$$A_d = e^{A_c \Delta t}$$
$$B_d = \int_0^{\Delta t} e^{A_c \tau} d\tau \, B = A_c^{-1}(A_d - I)B_c$$
$$Q_d = \int_0^{\Delta t} e^{A_c \tau} B_w Q_c B_w^T e^{A_c^T \tau} d\tau$$
$$R_d = R_c / \Delta t$$
$$C_d = C_c$$

Note that the exponentials are matrix exponentials, which are not simply the element-wise exponent. The matrix exponential has been given much research attention, and several solution

processes have been developed [227]. One method is the Taylor series expansion, which truncated provides some approximate solutions to the above transformations. For sufficiently small Δt:

$$A_d \approx (I + A_c \Delta t)$$
$$B_d \approx B_c \Delta t$$
$$Q_d \approx B_w Q_c B_w^T \Delta t$$

Using a program such as MATLAB, routines for computing the matrix exponential efficiently have already been implemented. Van Loan published the following tricks for computing the discrete system matrices (these are exact solutions if the matrix exponential is computed correctly) [228]:

$$\begin{bmatrix} A_d & B_d \\ \emptyset & I \end{bmatrix} = expm\left(\begin{bmatrix} A_c & B_c \\ \emptyset & \emptyset \end{bmatrix} \Delta t\right)$$
$$\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = expm\left(\begin{bmatrix} -A_c & B_w Q_c B_w^T \\ \emptyset & A_c^T \end{bmatrix} \Delta t\right)$$
$$A_d = C_{22}^T$$
$$Q_d = C_{22}^T C_{12}$$

Linearization of Nonlinear State Equations

No physical system is truly linear. Nonlinear equations for a system can be arranged in the form of a system of first-order derivatives:

$$\dot{x} = f(x, u, w, t)$$
$$y = h(x, v, t)$$

Many systems are linear within a range near their operating point, and can be adequately modeled by linearized equations using the methods shown below [229]. By linearizing about the current state estimate, a linear model can be created and updated as the state varies through time. First, an 'operating point' is chosen, and the linear approximation is formed by taking the first term of the Taylor series expansion

$$\delta x = x - x_0$$
$$\delta u = u - u_0$$
$$f(x, u, t) = f(x, u, w, t)|_{x_0, u_0, w_0} + \frac{\partial f}{\partial x}|_{x_0, u_0, w_0} \delta x + \frac{\partial f}{\partial u}|_{x_0, u_0, w_0} \delta u + \frac{\partial f}{\partial w}|_{x_0, u_0, 0} w + \cdots$$

$$h(x,u,t) = h(x,u,v,t)|_{x_0,u_0,v_0} + \frac{\partial h}{\partial x}|_{x_0,u_0,v_0}\delta x + \cdots + v$$

The matrix equivalent to the partial derivatives above is the Jacobian matrix. The linearized state model can be arranged into a linearized state space form

$$\dot{\delta x} = F\delta x + G\delta u + G_w w$$

$$y = H\delta x + v$$

$$F_{i,j} = \frac{\partial f_i}{\partial x_j} \qquad G_{i,j} = \frac{\partial f_i}{\partial u_j}$$

$$G_{w\,i,j} = \frac{\partial f_i}{\partial w_j} \qquad H_{i,j} = \frac{\partial h_i}{\partial x_j}$$

In navigation, the equations of motion employed are non-linear with known derivatives (i.e. partial derivatives which have a symbolic solution). In these cases, the linearized state matrices can be solved directly. In many cases the partial derivatives are not known and must be solved symbolically. An approximate solution to a partial derivative can be found by determining the slope of the function at the operating point by sampling two values at that point. For example:

$$f = f(x_1, x_2, x_3, u_1, u_2)$$

$$\frac{\partial f}{\partial x_2} \approx \frac{f(x_1, x_2 + \epsilon, x_3, u_1, u_2) - f\left(x_{1_0}, x_{2_0}, x_{3_0}, u_{1_0}, u_{2_0}\right)}{\epsilon}$$

In the example ϵ is an arbitrarily small value, and the approximate solution approaches the exact solution as ϵ approaches 0. While this method works in theory, when implemented in a computer (using finite-precision math) the use of very small ϵ will cause errors in the computation due to subtractive cancelation. This can be remedied by the use of the complex step differentiation, in programming languages that support complex numbers, by the following approximation (using the example from above) [230]

$$\frac{\partial f}{\partial x_2} \approx \frac{\Im\left(f(x_1, x_2 + \epsilon i, x_3, u_1, u_2)\right)}{\epsilon}$$

$$\Im(*) - the\ imaginary\ part\ of\ (*)$$

In the present work, the nonlinear components consist mostly of trigonometric functions and the differentiation can be solved directly.

Controllability

The ability to drive the state to any desired location (equivalently, the ability to select all closed loop eigenvalues as desired) is determined using the controllability matrix; if the controllability matrix is full rank, the system is known to be controllable [231]. This matrix is formed for an n^{th} order system

$$\mathcal{C} \equiv [B\ AB\ \dots A^{n-1}B]$$

The choice of measurement $y = Cx$ has no effect on controllability. The controllability can be determined using the following line of MATLAB code, which returns a logical 0 or 1

```
>> length(A) == rank(ctrb(A,B))
```

State-Feedback Controller

Assuming the control law has access to all states (i.e. knows their value at all times), a feedback control gain is designed such that the input u is a function of the current state

$$u = -K\hat{x}$$

Assuming accurate models and measurements, this provides closed loop dynamics of

$$\dot{x} = Ax + Bu = Ax - BKx = (A - BK)x$$

The eigenvalues and stability of $(A - BK)$ determines the behavior of the closed loop system. Because the system is Controllable we know we can place the closed loop controller eigenvalues anywhere we want by appropriate choice of K. The desired eigenvalues are chosen and gain vector K is practically determined using a computer-aided tool (`>>place()` in MATLAB). Control eigenvalues can be chosen using many techniques, such as root locus, to achieve desired transient characteristics (such as settle time, rise time, and overshoot). The choice of closed loop poles is somewhat arbitrary theoretically, but practically there are speed limits based on the ability of the actuating input u (and by the sample rate if it is slow compared to system dynamics). Real-life actuators have limits to their output amplitude (saturation), and may have special considerations such as electrical applications such as where the input cannot be negative. An 'optimal' trade-off between the time response and the control effort can be solved using the 'Linear Quadratic Regulator' technique. The optimality is completely dependent on the

designer's choice of importance between input effort and system response dynamics. Since this is essentially subjective anyway in many cases, the rigorous application is neglected here. However, we learn from the method that reasonable closed loop poles can be seen using the symmetric root locus of the open-loop system. A symmetric root locus can be generated in MATLAB

```
>> rlocus(p*ss(A,B,C,D)*ss(-A,-B,-C,-D))
```

Where p is chosen as 1 or -1, whichever keeps the root locus asymptotes away from the Imaginary axis. Closed loop poles are selected by considering the locus in the left-half-plane only and placing the control poles near these loci. Where the locus is closer to the natural open-loop poles, the control effort is low and response is slow. The farther from the origin, response time is faster but control effort increases to accomplish this. Of course the designer can also completely ignore this and place the close loop poles to achieve other objectives such as low overshoot, for instance. One of the more useful results of the symmetric root locus is it often reveals that faster control can be accomplished with less effort than control which is slower than the open-loop poles; this is not intuitive.

A more objectively optimal method which will not be applied here is the Model Predictive Control technique that allows consideration of the control saturation and other nonlinearities in the controller design, at a cost of high online computational costs.

Robust Servo Control

It is known from the Internal Model Principle that controllers can only track a reference that contains dynamics of equal or lower Type than the system to be controlled. In practice, this generally leads to the requirement that a controller include at least one integrator in its dynamics so that it can track a step change in reference without error. A method to incorporate this with state-space controllers is to embed the integrator in the system model and to drive the integral of error to zero – in this regard the method looks like a regulator not a tracking controller which makes implementation easier. Define

$$\mathrm{e} \equiv r - y$$

where r is the desired (reference) output. Create a new state vector

$$z \equiv \begin{Bmatrix} \mathrm{e} \\ \dot{x} \end{Bmatrix}$$

$$\ddot{x} = A\dot{x} + B\dot{u} + B_w\dot{w}$$
$$\dot{y} = C\dot{x} + \dot{v}$$

For constant disturbances (w, v) this reduces to the modified state-space with new model A, B

$$\dot{z} = \begin{bmatrix} 0 & -C \\ \emptyset & A \end{bmatrix} z + \begin{bmatrix} 0 \\ B \end{bmatrix} \dot{u} = \mathrm{A}z + \mathrm{B}\dot{u}$$

This new model can regulate with zero steady state error even with changing reference output and with constant disturbance of both state and output. The controller is now designed using pole-placement for the pair (A, B), assuming they are controllable (typically (A, B) is controllable if (A, B) is). The robust controller can be combined with the linear state estimator for output feedback. Note that the control law $\dot{u}$ will be based on $-K\dot{x}$. This integration is analytical; only the output y is actually integrated and the rest of the state (or estimate) is used directly.

Observability

In a dual manner to the controllability, observability determines our ability to determine the entire state (equivalently, to place all estimator eigenvalues as desired) from the output (measurement) [231]. The observability matrix is constructed similarly to the controllability matrix – we hope that it also has rank n

$$\mathcal{O} \equiv [C \;\; CA \;\; CA^{n-1}]^T$$

The choice of input Bu has no effect on observability. The observability can be determined using the following line of MATLAB code, which returns a logical 0 or 1

```
>> length(A) == rank(obsv(A,C))
```

System designers often have a choice of measurement based the sensors that are used, so it is worth exploring the effect this choice has on observability. In general, it is desirable to measure the highest differential state (e.g. position instead of speed, temperature instead of temperature rate, etc.).

State Estimator (AKA Luenberger Observer)

We tested that the system is observable – assuming it is, then knowledge of that single state variable can be used to estimate the entire state. The estimator is formed by creating a new state

e – the error between the actual state and the estimated state – and creating a virtual controller to drive the error to 0. Mathematically, we calculate an estimated state $\hat{x}$ using our best approximation of the plant state-space, and an additional gain designed to drive the error between the estimate and the true output to zero.

$$\dot{\hat{x}} = \bar{A}\hat{x} + \bar{B}u + L(y - \hat{y}), \quad \hat{y} = \bar{C}\hat{x}$$
$$e \equiv x - \hat{x}$$
$$\dot{e} = \dot{x} - \dot{\hat{x}}$$

If the assumed model $(\bar{A}, \bar{B}, \bar{C})$ and the exact system (A, B, C) coincide, then the dynamics of the error are determined by

$$\dot{e} = (A - LC)e$$

Another view of this technique is to break the estimator into two steps: a prediction and correction that both occur at each time step. This is only meaningful for the discrete case, where the discrete model

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + L_D\left(y_k - \bar{C}\hat{x}_{k|k-1}\right) = (I - L_D\bar{C})\hat{x}_{k|k-1} + L_D y$$
$$\hat{x}_{k+1|k} = \bar{A}_D\hat{x}_{k|k} + \bar{B}_D u_k$$

This shows more clearly how $\hat{x}$ is essentially a weighted average of the model prediction and the actual measurement. In a Kalman Filter, the 'weight' L is chosen based on known error variances of the model and measurement (w, v). In a Luenberger observer, the weighting is chosen based on desired dynamics of the estimator, similar to methods commonly used with e.g. low-pass filters. Note that a Luenberger observer could be designed using an optimization method similar to Linear Quadratic Regulator – this is called a Linear Quadratic Estimator. If the weights of the optimization were chosen based on expected error and disturbance statistical properties (w, v), then the solution would collapse to the Kalman filter with constant gain (only without the advantage of estimate covariance prediction)!

We choose estimator poles by designing the 'closed loop state transition matrix' $(A - LC)$ which will drive the error asymptotically to zero. This can be done using numerical methods (such as `>>place()` in MATLAB). In the absence of well-characterized model and measurement errors, we choose desired poles of the estimator to be significantly faster than the controller dynamics such that they do not 'interfere' with the time response of the system. One

amazing thing about this method is it will generate the proper eigenvalues with very little restriction since the 'effort' to drive the state error to zero is purely mathematical and involves no physical application of control effort. Practically, the estimator poles are chosen as a compromise between estimation speed and filtering of noisy measurements. If the measurements are very precise, then very fast estimator poles can be chosen. If the model is very precise and measurements are inaccurate or noisy, the model can be trusted to predict the state better than the measurements by using slow estimator poles.

Output-Feedback Controller

In the State-feedback control design, knowledge of the entire state was required to design the control gain. With the state estimator, a control law can now be designed that relies on all states, using the state estimate for the controller. This provides closed loop dynamics:

$$\dot{x} = Ax - BK\hat{x} = Ax - BKx + BKe$$

Assuming that the estimator has provided accurate knowledge of the state (that is, e has been driven to zero), the response will simply be

$$\dot{x} = (A - BK)x$$

Thus the closed loop response (eventually!) looks the same as designed. These poles should be much slower than the estimator poles; otherwise the response of the two will interact, typically in a slow (longer settling time) manner. The choice of estimator and controller poles is completely independent (the Separation Principle), but the two can certainly interact in the time response.

To test the performance of the estimator and controller, it is useful to combine the estimator and state feedback control into one system. We can do that by creating a new state vector z that includes the original state x and the estimate error e (the italicized e is estimate error, while e was defined previously as the tracking error in the controller.)

$$z \equiv \begin{Bmatrix} x \\ e \end{Bmatrix}$$

$$\dot{z} = \begin{Bmatrix} \dot{x} \\ \dot{e} \end{Bmatrix} = \begin{bmatrix} A - BK & BK \\ \emptyset & A - LC \end{bmatrix} \begin{Bmatrix} x \\ e \end{Bmatrix} = A'z$$

Notice that the eigenvalues of A', the augmented state transition matrix, are the eigenvalues that were chosen in the pole placement process. This (now stable) system can be simulated for any initial conditions. To complete the short notes on optimization, it should be noted that if the controller and estimator are designed using LQR and LQE, respectively, then this resulting output-feedback control system is called a Linear Quadratic Gaussian.