# ABSTRACT

BRAUN, THOMAS R. High Speed Model Implementation and Inversion Techniques for Smart Material Transducers. (Under the direction of Professor Ralph C. Smith).

Smart material transducers are utilized in wide range of applications, including nanopositioning, fluid pumps, high accuracy, high speed milling, objects, vibration control and/or suppression, and artificial muscles. They are attractive because the resulting devices are solid-state and often very compact. However, the coupling of field or temperature to mechanical deformation, which makes these materials effective transducers, also introduces hysteresis and time-dependent behaviors that must be accommodated in device designs and models before the full potential of compounds can be realized. In this dissertation, we present highly efficient modeling techniques to characterize hysteresis and constitutive nonlinearities in ferroelectric, ferromagnetic, and shape memory alloy compounds and model inversion techniques which permit subsequent linear control designs.

# HIGH SPEED MODEL IMPLEMENTATION AND INVERSION TECHNIQUES FOR SMART MATERIAL TRANSDUCERS

by
Thomas R. Braun

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

APPLIED MATHEMATICS

Raleigh, North Carolina
2007

APPROVED BY:

| | |
|---|---|
| Ralph C. Smith<br>Chair of Advisory Committee | Hien Tran |
| Pierre Gremaud | Stefan Seelecke |

## DEDICATION

To my wife Sarah and son Matthew, who have supported, encouraged, and tolerated me through this process.

# BIOGRAPHY

Thomas R. Braun received his Bachelor's degree in applied mathematics from Asbury College in May 2001. Although originally planning on continuing straight to graduate school, an offer from Los Alamos National Laboratory (LANL) convinced him to take some time away from formal studies. Tom spent three years in the space engineering group at LANL, working on signal processing algorithms for remote sensing. During this time, he met his wife Sarah at a Bible study, and they were married on August 2, 2003. The two left Los Alamos in the summer of 2004 to continue studies, Tom in applied mathematics at North Carolina State, and Sarah in elementary education at Campbell. On January 12, 2006, Tom and Sarah welcomed a new family member with the birth of their son, Matthew Ryan Braun.

Tom is a geek at heart who can frequently be seen tweaking computer code for one of several projects or maintaining various websites. Anything that combines mathematical and/or analytical skills with computer programming generally gets his interest. First and foremost, however, Tom is a believer in Christ and active with his local church and the broader Christian community.

## ACKNOWLEDGEMENTS

I would like to thank Professor Ralph Smith for his support and advice throughout this entire process, starting even before I arrived at North Carolina State as a student. His advice, direction, and support has been and continues to be invaluable. I'd also like to thank Professors Hien Tran, Pierre Gremaud, and Stefan Seelecke for their help and support on my committee, and all of my professors and fellow graduate students. The community here as a whole have made this a very friendly, welcoming place to learn and work.

This research has been supported primarily by the United States Department of Education GAANN Fellowship, in part through the NSF grant CMS-0201560 and in part by the Air Force Office of Scientific Research through the grant AFOSR-FA9550-04-1-0203. Special thanks also go to David Shafer and Todd Marcks for the the excellent way in which the fellowship was managed and coordinated here at North Carolina State University.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction to Smart Materials

Smart materials are increasingly considered as actuators and sensors for aerospace, aeronautic, industrial and automotive applications due to their unique transduction capabilities. Actuator capabilities are derived from the converse effect in which input fields produce deformations in the material whereas the direct effect, comprised of stress-induced fields (electric or magnetic materials) or temperature changes (shape memory alloys), provides the materials with sensor capabilities. Ferroelectric materials are presently employed in microphones, accelerometers, fluid pumps, nanopositioning stages, sonar transducers, vibration sensors and actuators, ultrasonic sources, inkjet printers, and camera focusing mechanisms. Ferromagnetic transducers are typically bulkier than their ferroelectric counterparts, due to the circuitry and housing required to produce magnetic fields, but they generally provide greater input forces. They are also presently being considered for a wide variety of applications including torque sensors, high speed, high accuracy milling, ultrasonic sources, sonar transduction, and vibration sensing and attenuation. Finally shape memory alloys provide much larger strains than their ferroelectric or ferromagnetic counterparts but typically cannot be actuated at high cycle rates. They are also employed in a wide variety of applications including biomedical wires and grabbers, pipe clamps, automated valves for fire control systems, cell phone antennas, and structural vibration suppression systems. Details regarding present and predicted applications can be found in [40].

However, the electromechanical, magnetomechanical and thermomechanical coupling mechanisms that endow these compounds with unique transducer capabilities also produce hysteresis and constitutive nonlinearities that must be accommodated in designs before the multifunctional material attributes can be fully realized. For certain applications, these nonlinear effects can be minimized by restricting input field levels; however, this limits the applicability of the materials. For ferroelectric devices, hysteresis can be partially mitigated through the use of charge or current controlled amplifiers [21–24]. However, this is often significantly more expensive then using voltage controlled amplifiers and current control is ineffective if maintaining a DC bias as required for many positioning mechanisms. A third alternative is to linearize hysteretic devices using feedback control. Whereas effective at lower frequencies, this technique becomes less

effective at higher frequencies if sampling rates are limited. Moreover, extensive linearization can limit control authority for prescribed tasks such as vibration attenuation or reference signal tracking.

To fully utilize the unique transduction capabilities of ferroelectric and ferromagnetic materials for high performance applications, it is thus necessary to construct models that are appropriate for material characterization, device designs, and control implementations. These objectives dictate that the models quantify hysteresis and constitutive nonlinearities in a manner that is sufficiently accurate to meet design specifications and sufficiently efficient to permit real-time implementation. Moreover, models often must be efficiently inverted to provide filters which approximately linearize transducer dynamics for subsequent linear control design.

Several classes of models for ferroelectric and ferromagnetic materials meet these criteria including homogenized energy models [1–5,26,40–42,45,46], Preisach models [14,29,30,37,47,50], and domain wall models [10,18,27,44]. Moreover, inverse models based on these techniques have been experimentally implemented, e.g. see [15, 16, 30–32, 47–49]. In this work, we focus on the development of highly efficient implementation and inversion techniques for the homogenized energy model for subsequent use in high speed design and control algorithms. We employ this technique due to its physical basis, which permits correlation of parameters with measured properties of data [40], and its capability to characterize a wide range of physical behavior including closure of biased minor loops in quasistatic regimes, thermal relaxation or creep, reversible behavior prior to switching, and temperature-dependent dynamics. Details regarding the framework including a discussion of how it compares and contrasts to Preisach models, can be found in [40].

Aspects of this work follow the development given in [3, 4] and extend previous work in a number of regards. We establish that the continuity of field-polarization, field-magnetization, or field-strain relations is maintained with approximation but also demonstrate numerical integration or quadrature stepsize requirements to achieve this continuity numerically. This directly impacts subsequent inversion algorithms. New quadrature techniques to improve the continuity of approximate relations for improved inversion and control implementation are then discussed. We develop look-up table and rational Chebyshev techniques to improve the implementation for certain applications as well as implementation techniques to characterize the strains generated by ferroelectric or ferromagnetic rods. Finally, we design highly efficient algorithms to construct inverse models that incorporate thermal relaxation for subsequent linear control design.

Chapter 2 introduces implementation algorithms for the nonlinear and hysteretic field-polarization/field-magnetization relations in ferroelectric and ferromagnetic compounds as well as strain relations that allow algorithm construction for displacements generated by rod transducers. The implementation of the inverse model is addressed and numerically demonstrated in Chapter 3. These modeling and inversion techniques are extended in Chapter 4 to incorporate ferroelectric or ferromagnetic regimes in which stress-dependent $90^o$ switching must be considered. Finally, the modeling and inversion techniques are extended to shape memory alloys in

Chapter 5.

## 1.1 Ferroelectric Compounds

Ferroelectric materials exhibit a coupling between electric field and strain. Although a wide variety of these materials exist, the most commonly employed are the ceramics lead zirconate titanate (PZT) and lead magnesium niobate (PMN) as well as the polymer polyvinylidene fluoride (PVDF). The crystalline structure for lead titanite ($PbTiO_3$), which is typical for the entire class of materials, is given in Figure 1.1. The materials are ferroelectric below a material dependent critical temperature known as the Curie temperature. Below this temperature, the titanate ion has an equilibrium that is off-center in the crystal. In general there are 6 equilibrium: up, down, left, right, in front, or in back. This creates a dipole; since this ion is off-center, the charge within the lattice structure is off-center as well. Thus, in Figure 1.1(a) there is a positive charge, or polarization, on the upper half of the crystal. Application of an electric field in this direction shifts the titanate ion up, increasing the polarization. As it shifts, the other ions in the lattice attempt to maintain an equilibrium with the titanate ion, and shift up as well. This elongates the crystal and produces a strain. If the electric field is removed, the crystal returns to its original configuration, as depicted in Figure 1.1(c). Application of an electric field in the opposite direction has a more profound impact. The titanate ion will move linearly with the field for a brief period, but then reaches a point where the energy required to maintain its current position is large enough that it crosses an energy barrier to a lower equilibrium. At this point, called the coercive point, it switches to this equilibrium instead. This is depicted in Figure 1.1(d). If the electric field is increased, the ion continues to move with the field, as shown in Figure 1.1(e). However, when the electric field is removed the titanate ion returns to the lower equilibrium position Figure 1.1(f). It takes application of a field in the positive direction

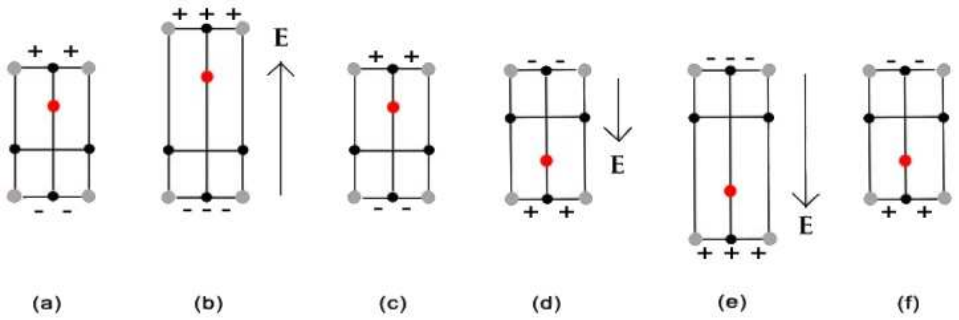



Figure 1.1: Diagram of the lattice structure of the piezoceramic $PbTiO_3$. The titanate ion, depicted in red, moves with the electric field and the entire lattice structure elongates as it does so. When applying a field in the opposite direction, the ion reaches a point where it switches to the equilibrium in the lower half of the lattice structure, and stays at this equilibrium when the field is removed.

to return the ion to its original position. Whereas the precise lattice structure for each material is different, similar mechanisms occur in all ferroelectric materials.

Any bulk material will contain a very large number of these lattice structures, possibly with different parameters to describe them. However, they form into homogenous regions, with the difference in lattice elements occurring in different regions. A plot of the field-polarization behavior for PZT is given in Figure 1.2(a). Note that whereas a steep transition occurs, due to different regions having different properties the transition is more gradual than that observed within a homogeneous region. However, the nonlinear hysteretic behavior of the material is immediately obvious. Another key feature is depicted in Figures 1.2(b) and (c). In this case, the input field is held constant for a period of time, and the polarization of the material continues to change. This behavior, called creep, is caused by the thermal relaxation properties of the material. From a modeling perspective, we see that any high-accuracy model of these materials must incorporate these thermal effects.

## 1.2 Ferromagnetic Compounds

Ferromagnetic materials respond to magnetic field and yield a magnetization and an inductance rather than a polarization. At the atomic level, magnetic moments are formed by the spin and angular momentum of electrons and these moments align themselves in a uniform direction. This creates a region or domain of uniformly aligned magnetic moments. However, at the macroscopic scale and in the absence of an applied field, the quantum forces which uniformly align the dipoles are less significant than the forces which would seek to minimize the magnetization by aligning the dipoles in opposite directions. To balance these effects, a macroscopic material exhibits many different domains, each of which is uniform internally but may be oriented differently than the domains around it. However, if a sufficient magnetic field is applied, the domains will rotate to align themselves with the magnetic field. This rotation yields a strain as well as a change in magnetization. When the field is removed, the domains may rotate some to reduce the strain but will tend to stay in their aligned orientation until a sufficient magnetic field is applied in the opposite orientation to rotate them the other way. For a bulk material, this yields a hysteresis loop like the one shown in Figure 1.3.

Note that at the macroscopic level both ferroelectric and ferromagnetic materials are characterized by regions of uniform polarization or magnetization which retain a remanence value after removal of an applied field and which require a sufficiently large field in the direction opposite their current direction to reorient them. Both also exhibit a bulk strain which is directly related to this polarization or magnetization. Additionally, ferromagnetic materials have similar thermal relaxation properties and are also subject to creep, as can be observed for a steel rod in Section 2.2, Figure 2.2. Thus, uniform models can be developed for both materials that describe the mesoscopic (region, domain level) and macroscopic behavior.

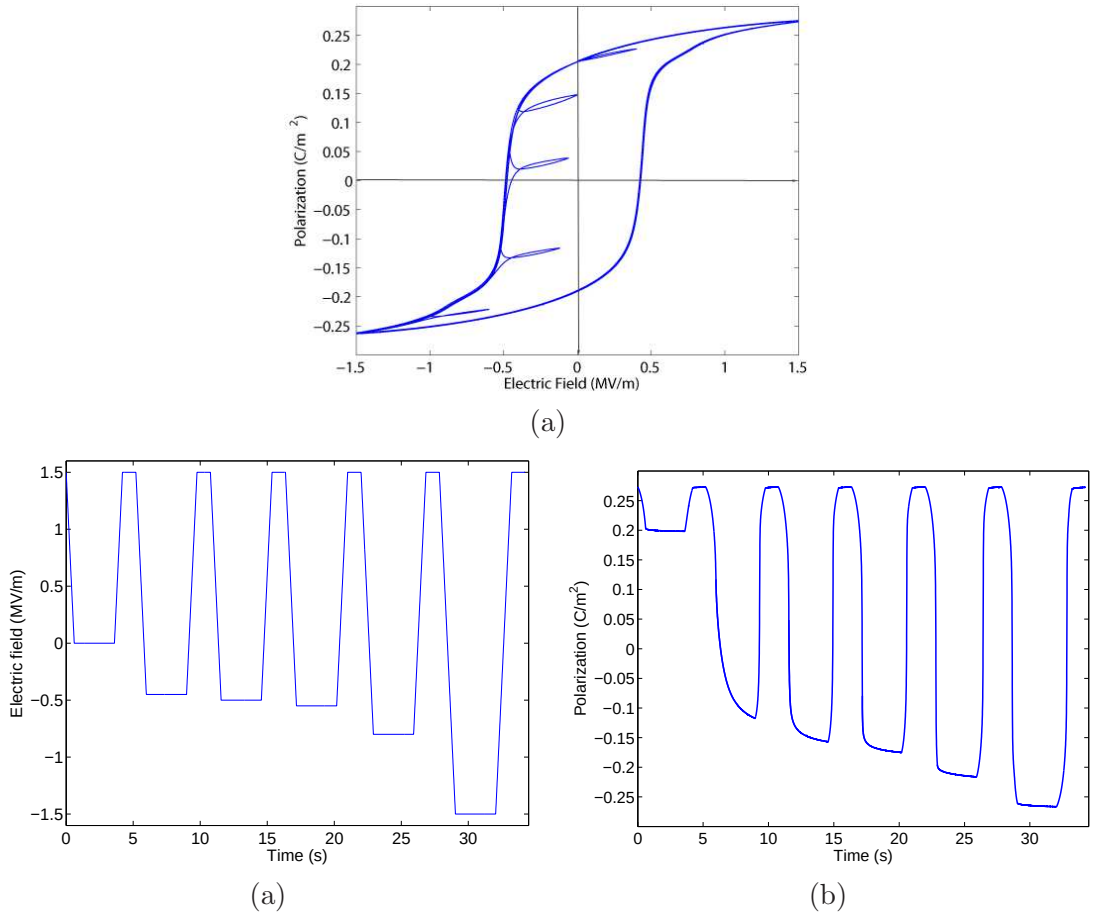



Figure 1.2: PZT data from [51] illustrating time-dependent material properties. (a) Nonclosure of field-polarization minor loops, and (b), (c) change in polarization for fixed input field values.

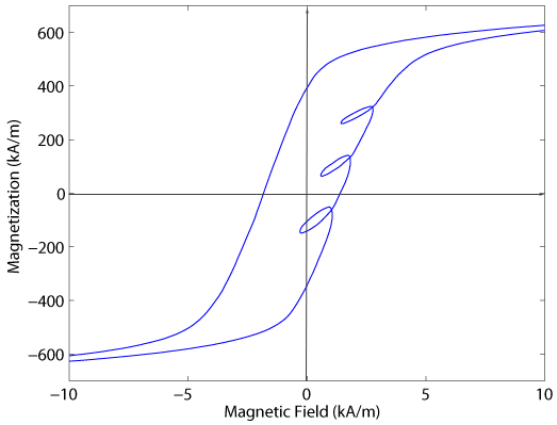



Figure 1.3: Magnetic field versus magnetization for a steel rod. Data taken from [5].

## 1.3 Shape Memory Alloy Compounds

Shape memory alloys (SMAs) exhibit additional effects which must be incorporated in constitutive models. SMAs are capable of recovering from much larger strains than ferroelectric or ferromagnetic actuators – for NiTi (Nickel Titanium), the most common shape memory alloy, recovery from up to 10% strains has been observed [25, 26, 38–40]. At higher temperatures, and in the absence of an applied stress, the materials exist in the highly symmetric austenite phase, whereas at lower temperatures, the materials exhibit one of several variants of a lower symmetry martensite phase. These phases are depicted in Figure 1.4 for the uniaxial case. Note that martensite phases are longer along one axis; thus when all crystals exhibit the same martensite variant the bulk material exhibits a strain. This gives two fundamental methods by which the SMA may be utilized. First, superelastic behavior may be obtained by applying a stress to a material in the austenite phase. The stress can force the material to transform into martensite which allows a much larger deformation than would otherwise be possible. Upon removal of the stress, the material will revert to the austenite state and recover its original shape. This fact is utilized to build vibration suppression systems for civil structures, cell phone antennas, and flexible wires for medical and aerospace applications [40]. The other method to utilize an SMA involves what is termed the shape memory effect. In this method, the material is machined into a desired shape while in its high-temperature austenite phase, and then allowed to cool to martensite. The shape of the martensite crystals allows a quasiplastic deformation to occur under an applied load. When the material is heated, however, it returns to austenite and the original undeformed shape is recovered. This fact has been utilized to create grippers, switches, relays, and pipe/wire couplers [40]. Shape memory alloys are also under investigation to modify jet engine chevron shape in flight to control noise and maximize fuel efficiency [7, 40]. Additionally, many of these alloys are biocompatible, and a variety of biomedical applications are being investigated [11, 12, 40].

For actuation, the shape memory effect is often employed to produce a desired strain. Transformation can be induced through Joule heating by application of a current. While this can force a transformation from martensite to austenite, it cannot induce austenite to transform back into

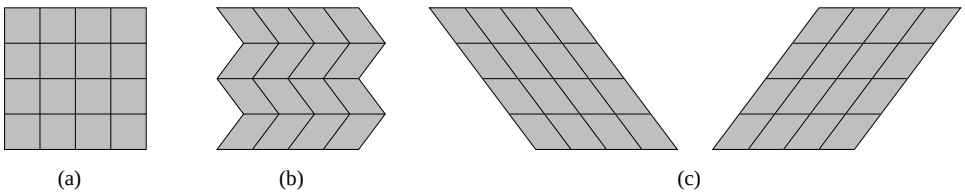
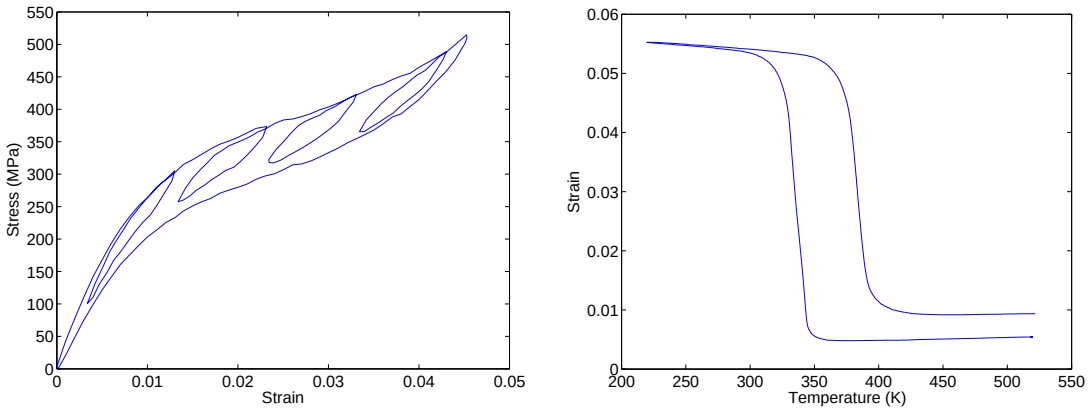



Figure 1.4: Representation of phases exhibited by shape memory alloy crystals, (a) the high-symmetry austenite phase, (b) martensite phases in the twinned configuration which preserves the same macroscopic shape as the austenite configuration possessed, and (c) detwinned martensite variants in which a macroscopic strain is manifested.

martensite. One must wait until the normal conductive, convective, or radiative processes dissipate enough heat to allow the material to transform on its own. This limits the frequency at which the SMA can be actuated. For this reason, SMA actuators are limited either to low-frequency applications or to microdevices where the ratio of surface area to volume is high. More recently magnetic shape memory alloys have been developed which combine the shape memory and superelastic effects with ferromagnetic behavior. This holds potential for large strain response at higher cycle rates. However, the construction and physical properties of these materials remains an active research area [13, 34] and the materials themselves are not treated here.

The strain of an SMA wire for a time varying stress but constant external temperature can be observed in Figure 1.5(a), while the strain for a fixed applied load at varying temperatures may be observed in Figure 1.5(b). A more detailed overview and numerous references on their application can be found in [25, 26, 38–40].




(a) (b)

Figure 1.5: Response of the shape memory alloy NiTi. (a) Stress versus strain under moderately fast loading [28] and (b) temperature versus strain for a constant load [17].

# Chapter 2

# Ferroelectric and Ferromagnetic Model Development

## 2.1 Homogenized Energy Model — Theoretical Development

We summarize here the homogenized energy model for ferroelectric and ferromagnetic materials [5, 40–42, 45, 46]. We note that this is a multiscale approach in which mesoscopic behavior is quantified via energy principles and a macroscopic model is subsequently constructed via stochastic homogenization techniques. The macroscopic ferroelectric model is

$$P(E;x_+) = \int_0^\infty \int_{-\infty}^\infty \nu_c(E_c)\nu_I(E_I)[\overline{P}(E+E_I;E_c,x_+)]\,dE_I\,dE_c \tag{2.1}$$

where $P$ denotes the polarization, $E$ is the electric field, $E_c$ denotes the coercive field value at which the dipoles change orientation, $E_I$ quantifies the interaction field due to dipole interactions, $\overline{P}$ is a mesoscale polarization relation, and $x_+$ is the internal material state. For magnetic materials the analogous model

$$M(H;x_+) = \int_0^\infty \int_{-\infty}^\infty \nu_c(H_c)\nu_I(H_I)[\overline{M}(H+H_I;H_c,x_+)]\,dH_I\,dH_c \tag{2.2}$$

is employed, where $H$ is magnetic field and $M$ is magnetization. The two model parameters $E_c$ and $E_I$ or $M_c$ and $M_I$ are assumed to vary throughout the material and have associated densities $\nu_c$ and $\nu_I$. The densities are constrained by the the physical relations:

$$\begin{aligned}
&1.\ \text{both } \nu_c \text{ and } \nu_I \text{ are bounded by decaying exponentials,}\\
&2.\ \nu_c \text{ is strictly positive,}\\
&3.\ \nu_I \text{ is symmetric about 0, and}\\
&4.\ \textstyle\int_0^\infty \nu_c(E_c)dE_c = 1, \qquad \int_{-\infty}^\infty \nu_I(E_I)dE_I = 1.
\end{aligned} \tag{2.3}$$

The integrals in (2.1) or (2.2) are solved numerically via quadrature relations; that is,

$$P(E;x_+) = \sum_{i=1}^{N_c}\sum_{j=1}^{N_I} \nu_c(E_c[i])\nu_I(E_I[j])w_c[i]w_I[j]\overline{P}(E+E_I[j];E_c[i];x_+[i,j]) \tag{2.4}$$

or the magnetic equivalent, where $w_c$ and $w_I$ give the quadrature weights. To simplify the subsequent discussion, we formulate equations solely in terms of the electric field and polarization, with the understanding that the model is unchanged if magnetic field and magnetization/inductance are instead employed.

The kernel $\overline{P}$ is modeled via energy principles. The mesoscopic Helmholtz energy is taken to be

$$\psi(P) = \begin{cases} \eta(P+P_R)^2/2, & P \le -P_I \\ \dfrac{\eta}{2}(P_I - P_R)\left(\dfrac{P^2}{P_I} - P_R\right), & |P| < P_I \\ \eta(P-P_R)^2/2, & P \ge P_I \end{cases} \tag{2.5}$$

where $P_I = P_R - E_c/\eta$ denotes the positive inflection point at which the dipoles switch orientation, $P_R$ is the local remanence polarization, and $\eta$ is the reciprocal slope $\frac{\partial E}{\partial P}$. This energy relation is plotted in Figure 2.1.

The Gibbs free energy

$$G = \psi - EP \tag{2.6}$$

balances this internal Helmholtz energy with the electrostatic energy; i.e., work performed by the applied external field. As detailed in [40], where the Legendre transform properties of the Gibbs energy are discussed, $G$ is a function of the independent variable $E$ and the dependent variable $P$.

For regimes in which thermal relaxation is negligible, direct minimization of (2.6) yields the kernel

$$\overline{P}(E+E_I;x_+) = \frac{E+E_I}{\eta} + 2P_R x_+ - P_R \tag{2.7}$$

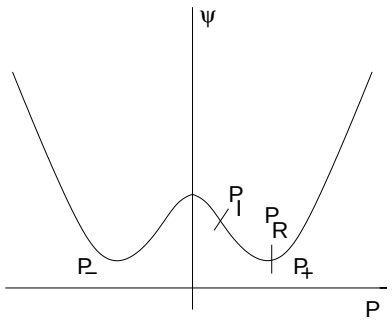



Figure 2.1: Helmholtz energy function (2.5).

where $x_+ = 1$ for positively oriented dipoles and $x_+ = 0$ for negatively oriented dipoles. For algorithm development, $x_+$ is set to 1 whenever $E + E_I > E_c$ and to 0 whenever $E + E_I < -E_c$.

Thermal activation is manifested by dipoles having sufficient thermal energy to switch states before a minima of the Gibbs energy is eliminated. To quantify this, the Gibbs energy is balanced with the relative thermal energy through Boltmann's relation

$$\mu(G) = C \exp\left(-\frac{GV}{kT}\right) \tag{2.8}$$

where $V$ is the mesoscopic reference volume, $k$ is Boltzmann's constant, $T$ is the material temperature in Kelvin, and $C$ is a constant chosen to ensure integration to unity. It is shown in [40, 41, 46] that the resulting kernel is

$$\overline{P} = x_+ \langle P_+ \rangle + (1 - x_+) \langle P_- \rangle \tag{2.9}$$

where $x_+ \in [0, 1]$ denotes the fraction of positively oriented dipoles and

$$\langle P_+ \rangle = \frac{\int_{P_I}^{\infty} P \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP}{\int_{P_I}^{\infty} \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP} \quad , \quad \langle P_- \rangle = \frac{\int_{-\infty}^{-P_I} P \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP}{\int_{-\infty}^{-P_I} \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP} \tag{2.10}$$

are the average polarizations associated with positive and negative dipole orientations. The evolution of dipole fractions is governed by the differential equation

$$\dot{x}_+ = -p_{+-} x_+ + p_{-+}(1 - x_+) \tag{2.11}$$

where the likelihoods $p_{+-}$ and $p_{-+}$ of a dipole switching from positive to negative, or vice-versa, are

$$p_{+-} = \frac{\exp\left(\frac{-G(E+E_I,P_I)V}{kT}\right)}{\tau(T) \int_{P_I}^{\infty} \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP} \qquad p_{-+} = \frac{\exp\left(\frac{-G(E+E_I,-P_I)V}{kT}\right)}{\tau(T) \int_{-\infty}^{-P_I} \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP} \tag{2.12}$$

and $\tau$ quantifies the material and temperature-dependent relaxation time. It is noted in [40] that whereas these likelihood relations are commonly employed in practice, they must be treated as approximate in a statistical sense. As an alternative, the formulations

$$p_{+-} = \frac{\int_{P_I}^{P_I+\epsilon} \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP}{\tau(T) \int_{P_I}^{\infty} \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP} \qquad p_{-+} = \frac{\int_{-P_I-\epsilon}^{-P_I} \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP}{\tau(T) \int_{-\infty}^{-P_I} \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP} \tag{2.13}$$

have also been employed. These prove advantageous for statistical analysis but are more computationally expensive than (2.12). Since our focus here is primarily on efficient model computation, we choose to focus on the former likelihood formulation.

## 2.2 Comparison of Response to Data

The homogenized energy model has been extensively validated against data from a wide range of actuators [5, 10, 40, 46]. Since our focus is computational rather than experimental, we briefly illustrate its performance by comparing it to steel, PLZT, and Terfenol-D actuators. Figures 2.2(b) and (c) show a fit to a cylindrical steel rod for the input in Figure 2.2(a), as reported in [5]. A fit to creep data on this same rod is given in Figures 2.2(e) and (f). Representative fits for PLZT and Terfenol-D data are given in Figure 2.3.

## 2.3 Homogenized Energy Model Implementation

The average polarization relations (2.10) and switching likelihood relations (2.12) are computationally expensive to compute, especially since they must be computed at each quadrature point and each time step. However, the choice of a piecewise quadratic Helmholtz energy, as opposed to a single fourth-order polynomial or other function, poses several advantages. One advantage is it allows these average polarizations and switching likelihoods to be expressed in a more computationally efficient form. To simplify notation, let $E_e = E + E_I$ denote the effective field. Consider the integral

$$\int_{P_I}^{\infty} \exp\left(\frac{-G(E_e, P)V}{kT}\right) dP.$$

By completing the square, this is equivalent to

$$\sqrt{\frac{\pi kT}{2V\eta}} \exp\left(\frac{V\eta}{2kT}\left(\frac{E_e^2}{\eta^2} + 2P_R\frac{E_e}{\eta}\right)\right) \operatorname{erfc}\left(\frac{V\eta}{2kT}\left(P_I - P_R - \frac{E_e}{\eta}\right)\right). \tag{2.14}$$

The likelihood of switching from positive to negative can thus be formulated as

$$p_{+-} = \frac{\sqrt{2V\eta}}{\tau(T)\sqrt{\pi kT}\operatorname{erfcx}\left(\frac{V\eta}{2kT}\left(P_I - P_R - \frac{E_e}{\eta}\right)\right)}. \tag{2.15}$$

Similarly, the fundamental theorem of calculus may be applied to $\int_{P_I}^{\infty} P \exp\left(\frac{-GV}{kT}\right) dP$ to obtain

$$\langle P+\rangle = \frac{\sqrt{2kT}}{\sqrt{\pi V\eta}\operatorname{erfcx}\left(\frac{V\eta}{2kT}\left(P_I - P_R - \frac{E_e}{\eta}\right)\right)} + \frac{E_e}{\eta} + P_R. \tag{2.16}$$

Similar approaches may be taken for the other average polarization and switching likelihood. Introducing the change of variables

$$z_+ = \sqrt{\frac{V}{2kT\eta}}\left(-E - E_I - E_c\right), \quad z_- = \sqrt{\frac{V}{2kT\eta}}\left(E + E_I - E_c\right), \tag{2.17}$$

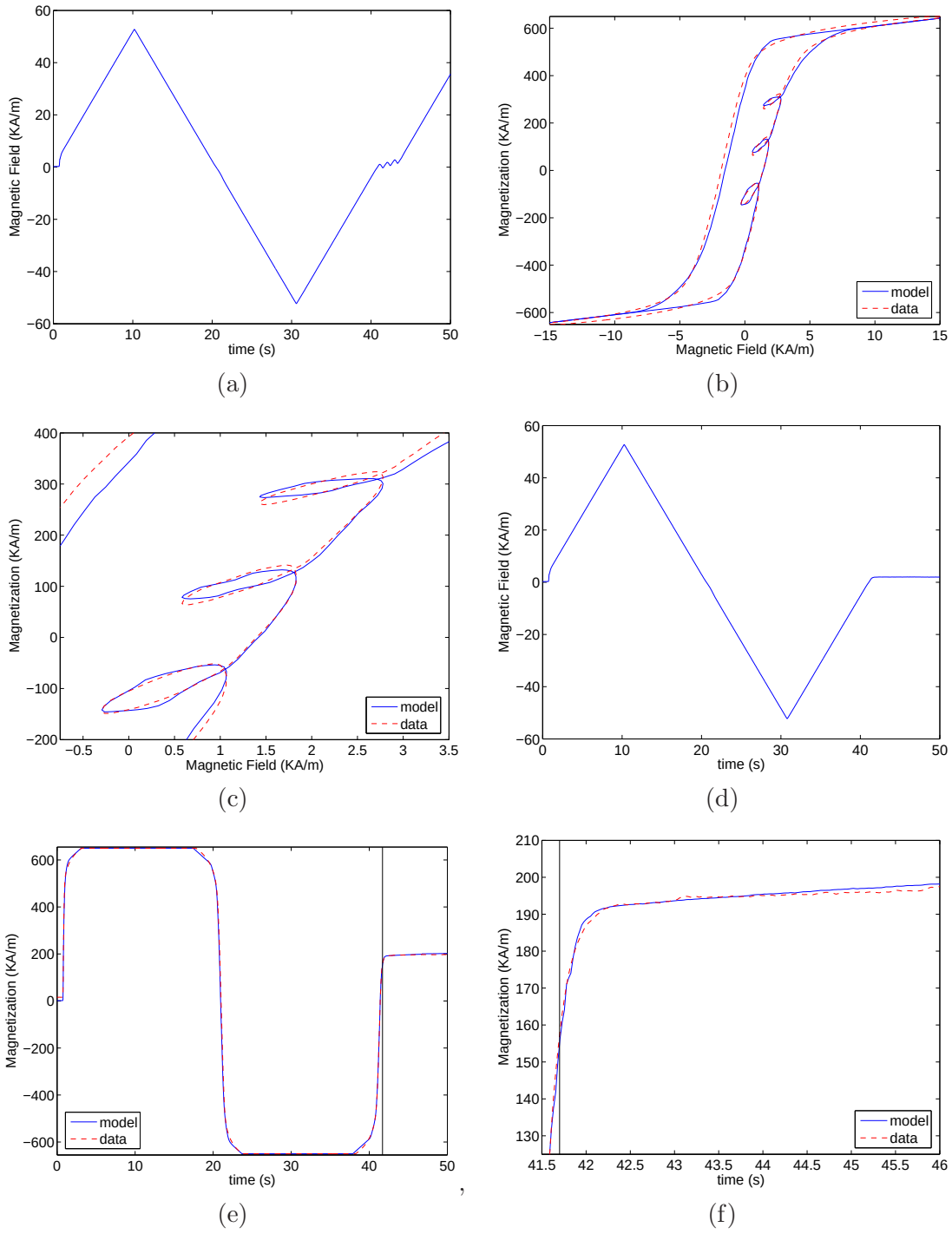
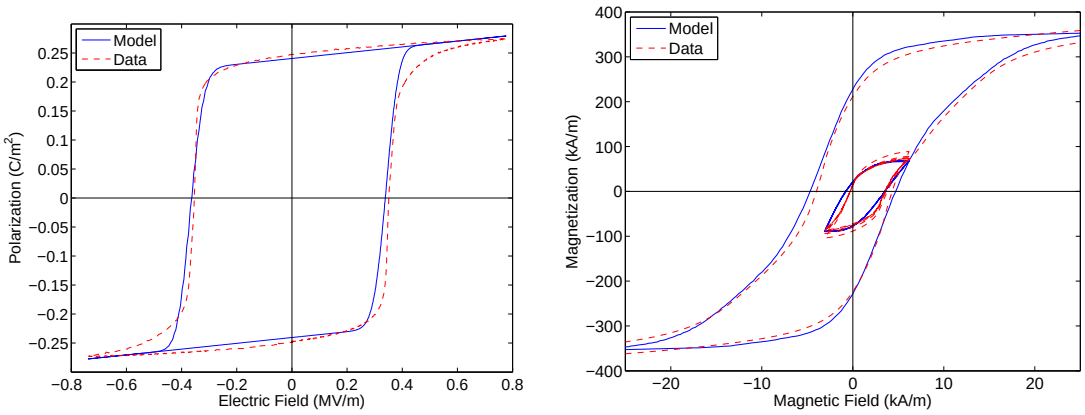



Figure 2.2: Comparison of model versus data for a cylindrical steel rod [5]. (a) shows the input field used in (b) and (c), which gives a sequence of three minor loops. (c) is a close-up of (b). The magnetic field seen in (d) gives the magnetization shown in (e) and (f). The field is held constant from the point delineated by the vertical black line. It is noted that the magnetization continues to creep up after this point.




(a) (b)

Figure 2.3: Comparison of model versus data for (a) PLZT and (b) Terfenol-D. PLZT data is taken from [20] whereas Terfenol-D data is courtesy of Marcelo Dapino at The Ohio State University.

yields the relations

$$\begin{aligned} p_{+-} &= \frac{1}{\tau(T)}\sqrt{\frac{2V\eta}{\pi kT}}\frac{1}{\operatorname{erfcx}(z_+)}, \qquad \langle P+\rangle = \sqrt{\frac{2kT}{\pi V\eta}}\frac{1}{\operatorname{erfcx}(z_+)} + \frac{E+E_I}{\eta} + P_R, \\ p_{-+} &= \frac{1}{\tau(T)}\sqrt{\frac{2V\eta}{\pi kT}}\frac{1}{\operatorname{erfcx}(z_-)}, \qquad \langle P-\rangle = -\sqrt{\frac{2kT}{\pi V\eta}}\frac{1}{\operatorname{erfcx}(z_-)} + \frac{E+E_I}{\eta} - P_R. \end{aligned} \tag{2.18}$$

Since the scaled complementary error functions appear in both switching likelihoods and average polarizations, the values need only be computed once instead of twice. To this end, define

$$pos = \frac{1}{\operatorname{erfcx}(z_+)}, \quad neg = \frac{1}{\operatorname{erfcx}(z_-)}. \tag{2.19}$$

The model can then be expressed as

$$P(E;x_+) = \frac{E}{\eta} - P_R + \sum_{i=1}^{N_c}\sum_{j=1}^{N_I}\overline{w}_c[i]\overline{w}_I[j]\left(x_+\left(pos+neg+k_1\right)-neg\right) \tag{2.20}$$

where

$$\overline{w}_c[i] = \sqrt{2kT/\pi V\eta}\nu_c(E_c[i])w_c[i], \qquad \overline{w}_I[j] = \nu_I(E_I[j])w_I[j], \qquad k_1 = P_R\sqrt{2\pi V\eta/kT}.$$

We note that the relations

$$\int_{-\infty}^{\infty}\int_{0}^{\infty}\nu_c(E_c)\nu_I(E_I)dE_cdE_I = 1, \qquad \int_{-\infty}^{\infty}\int_{0}^{\infty}\nu_c(E_c)\nu_I(E_I)E_IdE_cdE_I = 0$$

were used to simplify the equations resulting from the physical constraints (2.3). Similarly the evolution of dipole fractions is given by

$$\dot{x}_+ = \frac{1}{k_2}\left(-x_+(pos+neg)+neg\right) \tag{2.21}$$

where

$$k_2 = \tau(T)\sqrt{\frac{\pi kT}{2V\eta}}.$$

The differential equation (2.21) may be solved by any number of numerical differential equations routines. For example, using the backward Euler method, $x_+$ is computed at each time step as

$$x_+(t+\Delta t) = \frac{\frac{k_2}{\Delta t}x_+(t)+neg}{\frac{k_2}{\Delta t}+neg+pos}. \tag{2.22}$$

A more accurate, but also slightly more computationally intensive approximation, can be solved by assuming $E(t)$ is constant between $t$ and $t+\Delta t$; that is, approximate $E(t)$ by a step function, as is often the case in a digital system. In this case, the differential equation is a linear constant coefficient equation, and can be solved analytically to obtain

$$x_+(t+\Delta t) = \frac{neg}{neg+pos} + \left(x_+(t) - \frac{neg}{neg+pos}\right)\exp\left(-\frac{\Delta t}{k_2}(neg+pos)\right). \tag{2.23}$$

Whereas more accurate, the difference between the solution values provided by this formulation and (2.22) is typically not significant. For example, Figure 2.4 shows the results of both methods for an example input and parameter set (in this case, taken to match the ferroelectric material PLZT). As can be seen in the magnified version, the difference between the methods is minimal. For most applications, (2.22) is considered sufficiently accurate and we employ this approach due to its greater computational simplicity.

The formulations given here can significantly improve the run-time of the algorithm as compared with a more direct implementation of the equations in Section 2.1. However, in many cases better performance is still required. This leaves two options: reduce the number of quadrature points or consider additional approximation strategies.

## 2.4 Quadrature and the Continuity of P

It is observed that the computation time of the model is directly proportional to $N_c N_I$, the number of quadrature points. One may hope to improve computational performance by reducing the number of quadrature points. However, this exaggerates actual or apparent discontinuities in the output $P$.

By definition of the integral, the model (2.1) is continuous and differentiable. However, the model with quadrature (2.4) does not retain this smoothness. Since (2.4) is composed of

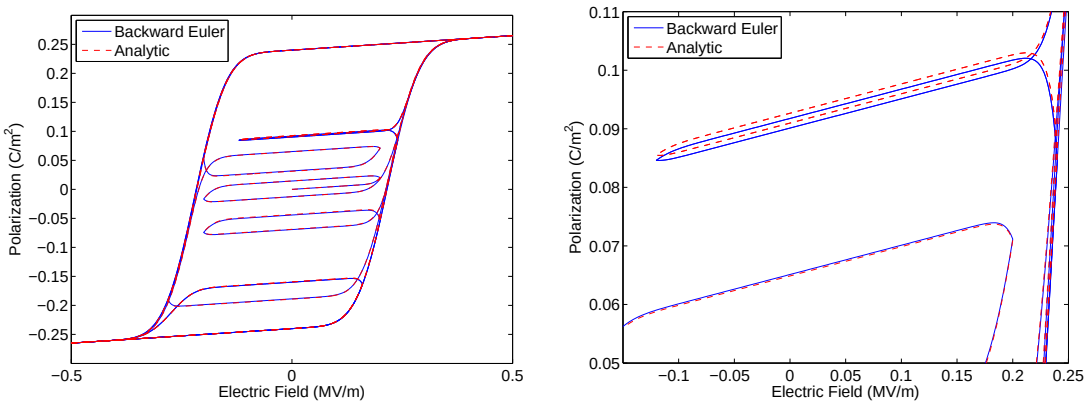



Figure 2.4: Comparison between the backward Euler approximation (2.22) and analytic approximation (2.23) for an example input, with parameters taken to match PLZT. (a) Entire hysteresis curve, and (b) magnified view.

finite sums, $P$ is smooth if and only if $\overline{P}$ is smooth for every quadrature point. Clearly for the negligible relaxation case (2.7), $\overline{P}$ is discontinuous, and its value changes by $2P_R$ at the coercive point $E + E_I = E_c$. Thus, $P$ contains $N_c N_I$ finite jumps, each of size

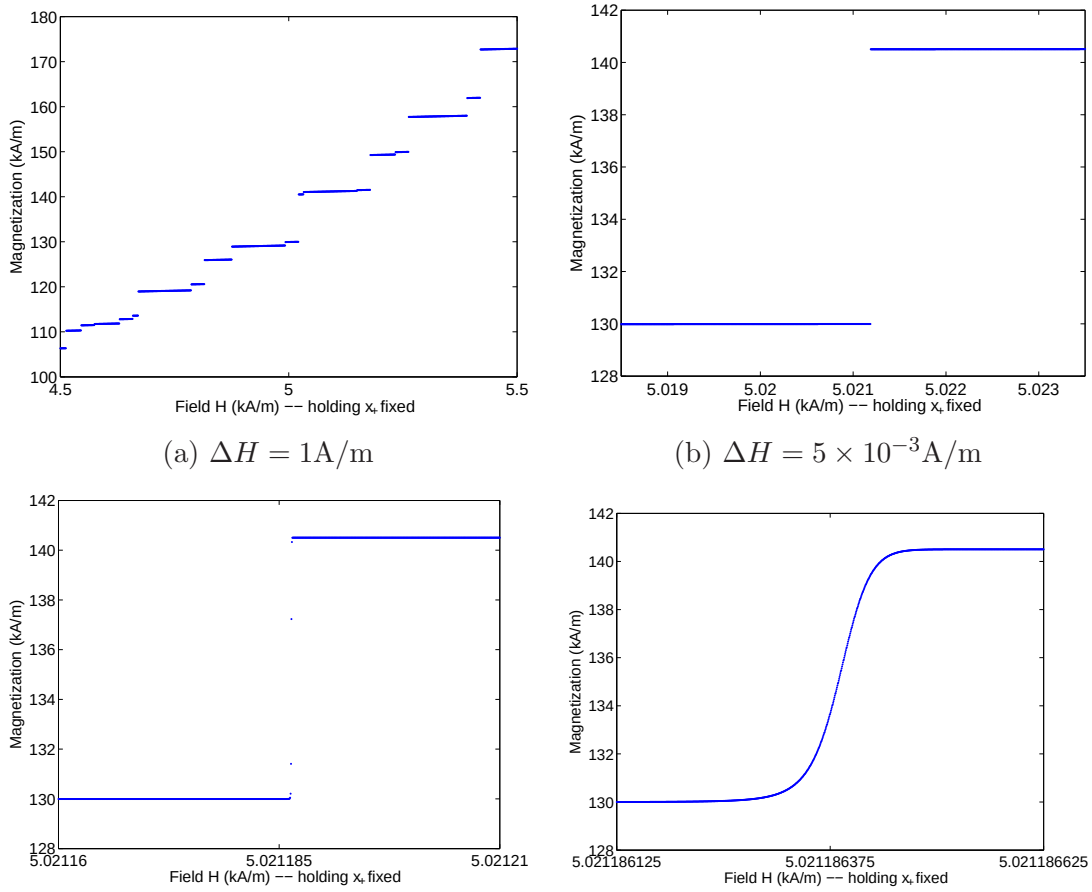
$$2P_R \nu_c(E_c[i])\nu_I(E_I[j])w_c[i]w_I[j], \quad i = 1 \dots N_c, j = 1 \dots N_I.$$

For the relaxation case, it can be shown that $P(E; x_+)$ is a continuous function. However, a plot of $P(E; x_+)$ (or rather $M(H; x_+)$ since the parameters come from a least-squares data fit to a Terfenol-D rod) shows that extremely small stepsizes may be required to resolve this continuity. Figure 2.5(a) shows a plot of $M(H; x_+)$ for various values of $H$ where the state $x_+$ was held fixed; i.e., each plotted sample was obtained by inputting the given $H$ to the model with the same value of $x_+$. The step between each input value of H was 1 A/m, and the function appears to be discontinuous. However, zooming in on one of these regions shows that the function is in fact continuous (see Figures 2.5(b) – 2.5(d)). Note, however, that it takes a $\Delta H$ approximately 10 orders of magnitude below $H$ to resolve the function in detail. It is unrealistic to expect such steep slopes to be resolved numerically. In practice, therefore, we must allow for the relaxation model to be discontinuous as well.

Model-based control applications require a smooth model response. As detailed in [40], the relaxation model limits to the negligible relaxation model as $\frac{V}{kT} \to \infty$. Since, the jump discontinuities in the negligible relaxation model are of height

$$2P_R \nu_c(E_c[i])\nu_I(E_I[j])w_c[i]w_I[j], \quad i = 1 \dots N_c, j = 1 \dots N_I,$$

we expect that for large $\frac{V}{kT}$ the "discontinuities" of the relaxation model will also be approxi-




(a) $\Delta H = 1\text{A/m}$

(b) $\Delta H = 5 \times 10^{-3}\text{A/m}$

(c) $\Delta H = 5 \times 10^{-5}$ A/m

(d) $\Delta H = 5 \times 10^{-7}\text{A/m}$

Figure 2.5: Plot of $M(H; x_+)$ for 6 intervals of composite 4-point Gaussian quadrature on each density. For each value of $H$, $x_+$ was held fixed; thus the plots represent a single timestep of the model for various input levels. The parameters in these plots were taken from a least-squares fit to Terfenol-D data.

mately this height. Thus, the actual or apparent discontinuities may be minimized by minimizing $\nu_c(E_c[i])\nu_I(E_I[j])w_c[i]w_I[j]$.

One way to increase smoothness is thus to increase the number of quadrature points. If the densities remain unchanged and the number of quadrature points is doubled, then $w_c[i]w_I[j]$ is halved for all $i, j$. This cuts the size of the jump in half. However, computation time is proportional to the number of quadrature points, so halving the height of the jumps in this way doubles the computation time. This is clearly undesirable. A better situation can be obtained if care is taken on how the quadrature is performed.

Since both $\nu_c$ and $\nu_I$ are bounded by decaying exponentials, quadrature may be performed on a finite interval and the tails of the distribution need not be considered. For accuracy and efficiency, quadrature is performed by a composite rule; the interval of consideration is divided into several smaller intervals and a quadrature rule is applied to each subinterval. Typically, all

the subintervals are taken to be equally spaced and this approach was used in the plots given previously. However, equally spaced subintervals are merely a convenience – there is no theoretical reason why one subinterval cannot be bigger than another. In particular, if the region is partitioned such that the integral over each subinterval is equal, then $\nu_c(E_c[i])\nu_I(E_I[j])w_c[i]w_I[j]$ will be on the same order of magnitude for all $i, j$. (Equality is not obtained because $\nu_c$ and $\nu_I$ vary on the interval and because for a general quadrature formula not all the weights are equal.) This has the effect of leveling out the jumps, decreasing the big ones while the small ones increase. The result is depicted in Figure 2.6.

The required number of quadrature points is often determined by accuracy and smoothness requirements. This approach can reduce the number of quadrature points needed for a given smoothness, resulting in significant computational savings.

## 2.5 Approximation Methods for Improved Computational Performance

Reduction of quadrature points alone is not always sufficient to achieve the desired performance of the relaxation model. Significant savings can also be obtained by introducing numerical approximations into the kernel itself. Exploration of the kernel shows that the majority of computation time is spent in the calculation of (2.19), and we thus focus on methods to approximate

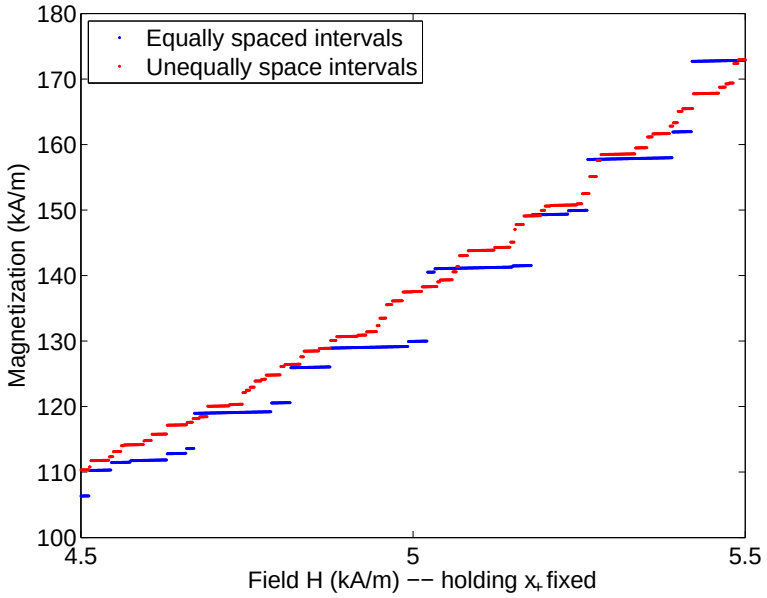



Figure 2.6: Comparison of $M(H; x_+)$ for 6 intervals of composite 4-point Gaussian quadrature on each distribution.

*pos* and *neg*. This is equivalent to approximating

$$f(z) = \frac{1}{\text{erfcx}(z)} \tag{2.24}$$

since both *pos* and *neg* require this function for different values of $z$. The rest of the model is employed in the manner detailed previously.

### 2.5.1 Look-up Tables

The simplest approach conceptually is to store $f(z)$ in a table for values of interest and approximate *pos* and *neg* by looking up the table value nearest $z_+$ and $z_-$, respectively. At the cost of memory, this reduces the complex erfcx calculation to a simple index calculation and memory access. The indexes into the lookup table may also be calculated efficiently. Let $h$ be the difference in $z$ between two elements of the table, and let $\ell$ be the $z$ value corresponding to the first (lowest) table entry. Then define

$$\tilde{E}_c[i] = \frac{E_c[i]}{h}\sqrt{\frac{V}{2kT\eta}} - \frac{\ell}{h}, \quad i = 1 \dots N_c, \qquad \tilde{E}_I[j] = \frac{E_I[j]}{h}\sqrt{\frac{V}{2kT\eta}}, \quad j = 1 \dots N_I,$$

$$\tilde{E} = \frac{E}{h}\frac{V}{2kT\eta}.$$

Most of these calculations can be performed offline. Only $\tilde{E}$ needs to be computed at each time step and this may be done outside the loops over the quadrature points. Once these values are obtained, the table index is obtained by rounding

$$pos\_index = -\tilde{E} - \tilde{E}_I[j] - \tilde{E}_c[i], \qquad neg\_index = \tilde{E} + \tilde{E}_I[j] - \tilde{E}_c[i] \tag{2.25}$$

to the nearest integer. These values must be checked against the table bounds. An index less than zero signifies a zero should be given instead of a value from the table, whereas an index greater than or equal to the table size implies that the linear approximation should be used instead. Note that the linear approximation for larger $z$ can be adjusted to take the table index (preferably without rounding) instead of the $z_+$ or $z_-$ as input as well, allowing $\tilde{E}_c$ and $\tilde{E}_I$ to replace the memory taken by the original quadrature points.

The lookup table approach has several advantages; it is conceptually straightforward, algorithm run-time is independent of table size (within architecture-dependent limits), and approximation error goes to zero as the table size increases. Even relatively small tables yield decent accuracy as illustrated in Table 2.3 for an example input field. Here it is seen that even 100 table elements produce a root-mean-square error less than approximately 0.25% of the remanence polarization, and at 10,000 table elements this error was reduced to approximately 0.002% of the remanence polarization. Note that 10,000 elements is less than 100 KB of memory when

using IEEE double precision floating point numbers. Finally, table values do not need to be recalculated if material properties (i.e., temperature) change. However, there are architectures where this moderate amount of memory is unattainable, or where memory access is too slow. For these cases, we alternatively consider polynomial approximations.

### 2.5.2 Rational Chebyshev Approximation

If the look-up table approximation is infeasible, the function $f(z)$ can instead by approximated by a polynomial or a rational function

$$R_{mk}(z) = \frac{p_m(z)}{q_k(z)} \tag{2.26}$$

where $p$ is a polynomial of degree $m$ and $q$ is a polynomial of degree $k$. Here $R_{mk}$ is chosen such that

$$\max w(z)\big|f(z) - R_{mk}(z)\big| \tag{2.27}$$

is minimized. This is referred to as a rational Chebyshev or minimax approximation, and is the approach used to implement erfc internally as well. While it can be shown that a unique minimix approximation exists for all $m$ and $k$, computation of these approximations in numerically intensive [36]. However, $f(z)$ does not depend on the material parameters or inputs. Thus, the minimax approximation can be calculated in advance and made available to the code for any choice of parameters. Methods to calculate these approximations using the Remez (also called Remes) algorithm are available in both *Mathematica* and *Maple*, alleviating the user from coding minimax algorithms themself.

Like the look-up table, the minimax approximation is applied on a bounded interval. Additionally, the weight function $w(z)$ must be specified. The obvious choices are $w(z) = 1$ for absolute error or $w(z) = 1/f(z)$ for relative error. For larger $z$, relative error is a better metric for controlling error. However, for small $z$ calculation of relative error is problematic since $f(z) \longrightarrow 0$. Thus, we let

$$w(z) = \min\left(1, \frac{\operatorname{erfc}(z)}{\exp(-z^2)}\right) \tag{2.28}$$

to use relative error for larger $z$ and absolute error for smaller $z$. The transition from absolute to relative error occurs at $z = 0$ and $w(z)$ is continuous.

We consider two methods to improve the accuracy of the approximation. One approach is to increase $m$ and/or $k$ while the other is to use multiple polynomials over smaller regions instead of a single polynomial over the whole region. Higher degree polynomials increase computational effort in floating point operations whereas adding more polynomials increases conditional statements and control logic. An optimal balance of these is architecture-dependent. As an example,

let $f(z)$ be approximated as

$$f(z) \approx \begin{cases} 0 & z < -3 \\ R_\ell(z) = \dfrac{p_{\ell 0} + p_{\ell 1} z + p_{\ell 2} z^2 + p_{\ell 3} z^3}{q_{\ell 0} + q_{\ell 1} z + q_{\ell 2} z^2} & -3 \le z \le -\frac{1}{2} \\ R_h(z) = \dfrac{p_{h0} + p_{h1} z + p_{h2} z^2 + p_{h3} z^3}{q_{h0} + q_{h1} z + q_{h2} z^2} & -\frac{1}{2} < z \le 25 \\ mz + b & z \ge 25. \end{cases} \tag{2.29}$$

This employs two minimax approximations with a degree 3 numerator and degree 2 denominator where the regions for each approximation were chosen so that the minimax error (2.27) is approximately the same. This holds the error (2.27) below $4 \times 10^{-4}$ for all $z < 25$. Since this accuracy is well above the limits imposed by IEEE double precision floating point numbers, further computational savings may be realized by transforming $R_\ell(z)$ and $R_h(z)$ into continued fraction form, namely

$$R_\ell(z) = a_{\ell 0 const} + a_{\ell 0 lin} z + \cfrac{b_{\ell 1}}{z + a_{\ell 1} + \cfrac{b_{\ell 2}}{z + a_{\ell 2}}}, \quad R_h(z) = a_{h0 const} + a_{h0 lin} z + \cfrac{b_{h1}}{z + a_{h1} + \cfrac{b_{h2}}{z + a_{h2}}}. \tag{2.30}$$

Note that for each of these evaluations only 1 multiply, 2 divides, and 5 additions are required. The values of the relevant coefficients are given in Table 2.1.

### 2.5.3 Method Comparisons

Neither method is optimal for all situations. However, the look-up table approach is generally the most efficient in terms of run-times and accuracy. Tables 2.2 and 2.3 list the run times and errors for an example input. Parameter values were taken to match the ferroelectric material PLZT, and the temperature was assumed to remain constant. It can be seen that either the look-up table or polynomial approximations should almost always be preferred over the exact formulations. As expected, the look-up table approach outperforms the polynomial approach. The run time for various table sizes is not shown because as long as the table was kept small enough to fit in the processor's cache, the run-time was not significantly affected by the size of the table. In cases where the modest memory requirement cannot be achieved, the rational Chebyshev approximation should be used. This takes slightly longer and/or provides less accuracy than the look-up table approximation, but reduces the memory requirement to a handful of values. It is doubtful the memory will ever pose an issue on a conventional processor. Regions where it may prove unattainable are on embedded microprocessors and custom designed field programmable gate arrays (FPGAs) or application specific integrated circuits (ASICs). Rational Chebyshev approximations are a natural choice for these applications.

Table 2.1: Coefficients for the minimax approximation given in (2.29).

| | | | |
|---|---|---|---|
| $m$ | 1.771048244745112288944163714 | $b$ | 0.070532738046130371816086597 4 |
| $a_{\ell 0const}$ | 1.053053417788476733699618764 55 | $a_{\ell 0lin}$ | 0.141908467579507101523861449079 |
| $a_{\ell 1}$ | 0.652521450575033268529746994429 | $b_{\ell 1}$ | 2.050403340286826595097767105 01 |
| $a_{\ell 2}$ | -0.139674180341575164555785667908 | $b_{\ell 2}$ | 2.886593385496838808465136103 39 |
| $a_{h0const}$ | 0.056893441152762091532098222640 6 | $a_{h0lin}$ | 1.770706106689974595546470640 88 |
| $a_{h1}$ | -12.932727730329452231646847127 9 | $b_{h1}$ | 0.269387658595592656676548244377 |
| $a_{h2}$ | 16.415252643349734858384471326 5 | $b_{h2}$ | 216.983649137026313567944222991 |

Table 2.2: Time steps per second computed by full-precision, look-up table, and rational Chebyshev algorithms for three different computer architectures (averaged from 1 million time step). Polynomial approximation was performed using the continued fraction form (2.30) of the example polynomial given in (2.29). Look-up tables were computed for $-6 \leq z \leq 25$. In all cases, the model was run with 40 quadrature points for each distribution (1600 total).

| | Intel Xeon 1.7 GHz 256 KB L2 Cache | Intel Xeon (64 bit) 3.8 GHz 2 MB L2 Cache | Intel Pentium-M 2.16 GHz 2 MB L2 Cache |
|---|---|---|---|
| Negligible relaxation algorithm | 10,748 | 62,725 | 31,969 |
| Relaxation – full numerical precision | 429 | 1801 | 1276 |
| Relaxation – look-up table approximation | 5344 | 14,923 | 14,226 |
| Relaxation – polynomial approximation (2.29) | 4997 | 11,674 | 12,177 |

Table 2.3: Approximation error for an example input using parameters obtained for PLZT. Polynomial approximation was performed using the continued fraction form (2.30) of the example polynomial given in (2.29); look-up tables were computed for $-6 \leq z \leq 25$, and erfcx was implemented directly as exponential and error function evaluations, with the linear approximation utilized for large $z$. Forty quadrature points were given for each distribution (1600 total)

| | Peak error (% of $P_R$) | RMS error (% of $P_R$) |
|---|---|---|
| Look-up table (100 table elements) | 0.65860% | 0.25593% |
| Look-up table (1000 table elements) | 0.063883% | 0.0055925% |
| Look-up table (10,000 table elements) | 0.0020362% | 0.00019354% |
| Polynomial approximation (2.29) | 0.5377% | 0.1647% |

## 2.6 Displacement Model

Whereas polarization or inductance/magnetization can sometimes be measured, it is often necessary to formulate the model in terms of displacement instead of polarization. Any such extension requires knowledge of the geometry of the transducer. Rods or beams can be treated with a one-dimensional transducer model, whereas more general shapes require higher-dimension models. As detailed in [40], both nanopositioning stages employed in common atomic force microscope designs and magnetic transducers employed for high speed milling utilize ferroelectric and ferromagnetic rods that are clamped at one end and subject to damped restoring forces at the other. Hence we illustrate the model for a rod of length $\ell$ and cross sectional area $A$ as depicted in Figure 2.7. The end mass $m_\ell$ with damping $c_\ell$ and stiffness $k_\ell$ encompass adjacent transducer and plant dynamics.

It is shown in [40] that the constitutive relation for the induced stress is given by

$$\sigma = Y\varepsilon + C\dot{\varepsilon} - a_1(P - P_0) - a_2(P - P_0)^2 \tag{2.31}$$

where $Y$ is the Young's modulus of the transducer, $C$ is the Kelvin-Voight damping coefficient, $P_0$ is the point the material is poled around, $a_1$ and $a_2$ are constants of proportionality relating the induced polarization to mechanical force, and $P = P(E; x_+)$ is given in (2.4).

Due to the distributed nature of the device, partial differential equation models have been developed in [40, 43] to characterize the device dynamics. However, for a number of motivating applications, it has been shown that because forces and fields are uniform in space, strains and hence displacements are also uniform thus permitting the use of ordinary differential equation representations. In this case, the strain is given by

$$\varepsilon(t) = \frac{u(t)}{\ell} \tag{2.32}$$

where $u(t)$ denotes the displacement of the rod tip at time $t$.

A force balance yields

$$m\frac{d^2u}{dt^2} + c\frac{du}{dt} + ku = \alpha_1(P - P_0) + \alpha_2(P - P_0)^2 \tag{2.33}$$

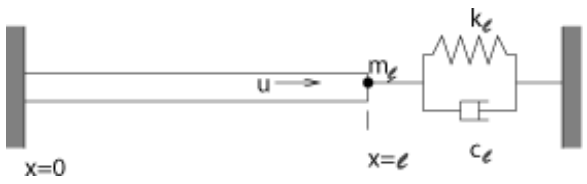



Figure 2.7: Prototypical rod used to construct overall model. The end mass $m_\ell$, stiffness $k_\ell$ and damping $c_\ell$ represent adjacent dynamics of the transducer and plant.

where $\rho$ is the density of the rod and

$$m = \rho A\ell + m_\ell, \qquad c = \frac{CA}{\ell} + c_\ell, \qquad k = \frac{YA}{\ell} + k_\ell,$$

$$\alpha_1 = Aa_1, \qquad \alpha_2 = Aa_2.$$

Further details as well as models for additional geometries may be found in [40, 43].

The solution to this differential equation can be obtained with a variety of numerical methods. However, variable-step methods require that $P$ be known at arbitrary times, which in turn requires that the homogenized energy model be solved at arbitrary times. This is typically not feasible in practice. We thus focus on fixed time-step methods. Of these, the most accurate approximation for a given stepsize may be obtained by assuming that $P$ is fixed between time steps so (2.33) can be solved analytically. This is an approximation, since thermal effects dictate $P$ will vary with time even if $E$ is constant; however, all other fixed step methods implicitly make similar assumptions and include additional approximation error as well.

The approximation of (2.33) depends on the relation of $\frac{c^2}{m^2}$ to $\frac{4k}{m}$. First, to simplify notation, let

$$v = \frac{a_1}{m}\left(P - P_0\right) + \frac{a_2}{m}\left(P - P_0\right)^2. \tag{2.34}$$

Here $v$ is assumed to be a constant, although the value of the constant changes at each time step. Additionally, let

$$\tilde{k} = \frac{k}{m}, \qquad \tilde{c} = \frac{c}{m}.$$

If $\tilde{c}^2 > 4\tilde{k}$ then the solution is given by

$$u(t) = b_1 \exp\left(-\frac{1}{2}\left(\tilde{c} + \sqrt{\tilde{c}^2 - 4\tilde{k}}\right)t\right) + b_2 \exp\left(-\frac{1}{2}\left(\tilde{c} - \sqrt{\tilde{c}^2 - 4\tilde{k}}\right)t\right). \tag{2.35}$$

Again, to simplify notation let $z = \sqrt{\tilde{c}^2 - 4\tilde{k}}$. Letting $u(0) = u_0$ and $\dot{u}(0) = u_1$ gives

$$b_1 = \frac{1}{2}\left(1 - \frac{\tilde{c}}{z}\right)\left(u_0 - \frac{v}{\tilde{k}}\right) - \frac{1}{z}u_1, \qquad b_2 = \frac{1}{2}\left(1 + \frac{\tilde{c}}{z}\right)\left(u_0 - \frac{v}{\tilde{k}}\right) + \frac{1}{z}u_1. \tag{2.36}$$

Substitution into (2.35) and collection of like terms yields

$$\begin{aligned} u(t) = {} & \frac{1}{2}\left(\left(1 - \frac{\tilde{c}}{z}\right)\exp\left(-\frac{1}{2}\left(\tilde{c} + z\right)t\right) + \left(1 + \frac{\tilde{c}}{z}\right)\exp\left(-\frac{1}{2}\left(\tilde{c} - z\right)t\right)\right)\left(u_0 - \frac{v}{\tilde{k}}\right) \\ & + \frac{1}{z}\left(\exp\left(-\frac{1}{2}\left(\tilde{c} - z\right)t\right) - \exp\left(-\frac{1}{2}\left(\tilde{c} + z\right)t\right)\right)u_1 + \frac{v}{\tilde{k}}. \end{aligned} \tag{2.37}$$

If we assume the step $\Delta t$ is fixed, then we can define the six constants

$$\begin{aligned}
d_1 &= \frac{1}{2}\left(\left(1-\frac{\tilde{c}}{z}\right)\exp\left(-\frac{1}{2}\left(\tilde{c}+z\right)\Delta t\right)+\left(1+\frac{\tilde{c}}{z}\right)\exp\left(-\frac{1}{2}\left(\tilde{c}-z\right)\Delta t\right)\right),\\
d_2 &= \frac{1}{z}\left(\exp\left(-\frac{1}{2}\left(\tilde{c}-z\right)\Delta t\right)-\exp\left(-\frac{1}{2}\left(\tilde{c}+z\right)\Delta t\right)\right),\\
d_3 &= \frac{1}{k}\left(1-d_1\right),\\
d_4 &= \frac{1}{4}\left(\frac{\tilde{c}^2}{z}-z\right)\left(\exp\left(-\frac{1}{2}\left(\tilde{c}+z\right)\Delta t\right)-\exp\left(-\frac{1}{2}\left(\tilde{c}-z\right)\Delta t\right)\right),\\
d_5 &= \frac{1}{2}\left(\left(1+\frac{\tilde{c}}{z}\right)\exp\left(-\frac{1}{2}\left(\tilde{c}+z\right)\Delta t\right)+\left(1-\frac{\tilde{c}}{z}\right)\exp\left(-\frac{1}{2}\left(\tilde{c}-z\right)\Delta t\right)\right),\\
d_6 &= -\frac{d_4}{\tilde{k}}.
\end{aligned} \tag{2.38}$$

In terms of these constants, the solution of the rod model for $\tilde{c}^2 > 4\tilde{k}$ is approximated by

$$\begin{aligned}
u(t+\Delta t) &= d_1 u(t) + d_2 \dot{u}(t) + d_3 v(t+\Delta t),\\
\dot{u}(t+\Delta t) &= d_4 u(t) + d_5 \dot{u}(t) + d_6 v(t+\Delta t).
\end{aligned} \tag{2.39}$$

A similar approach can be taken for the other two cases. If $\tilde{c}^2 = 4\tilde{k}$, we again obtain (2.39) except with the constants given by

$$\begin{aligned}
d_1 &= \left(1+\frac{\tilde{c}}{2}\Delta t\right)\exp\left(-\frac{\tilde{c}}{2}\Delta t\right), & d_4 &= -k\Delta t\exp\left(-\frac{\tilde{c}}{2}\Delta t\right),\\
d_2 &= \Delta t\exp\left(-\frac{\tilde{c}}{2}\Delta t\right), & d_5 &= \left(1-\frac{\tilde{c}}{2}\Delta t\right)\exp\left(-\frac{\tilde{c}}{2}\Delta t\right),\\
d_3 &= \frac{1}{k}\left(1-d_1\right), & d_6 &= -\frac{d_4}{\tilde{k}}.
\end{aligned} \tag{2.40}$$

Finally, if $\tilde{c}^2 < 4\tilde{k}$, the values for the constants are

$$\begin{aligned}
d_1 &= \exp\left(-\frac{\tilde{c}}{2}t\right)\left(\cos\left(\frac{z}{2}\Delta t\right)+\frac{\tilde{c}}{z}\sin\left(\frac{z}{2}\Delta t\right)\right), & d_2 &= \frac{2}{z}\exp\left(-\frac{\tilde{c}}{2}t\right)\sin\left(\frac{z}{2}\Delta t\right),\\
d_3 &= \frac{1}{k}\left(1-d_1\right), & d_4 &= -\frac{1}{2}\left(z+\frac{\tilde{c}^2}{z}\right)\exp\left(-\frac{\tilde{c}}{2}t\right)\sin\left(\frac{z}{2}\Delta t\right),\\
d_5 &= \exp\left(-\frac{\tilde{c}}{2}t\right)\left(\cos\left(\frac{z}{2}\Delta t\right)-\frac{\tilde{c}}{z}\sin\left(\frac{z}{2}\Delta t\right)\right), & d_6 &= -\frac{d_4}{\tilde{k}}.
\end{aligned} \tag{2.41}$$

Thus, the rod model costs only six multiplies and four additions per iteration, plus three multiplications and three additions for the calculation of $v$. That is, the rod model computation is essentially free when compared to the computational cost of the homogenized energy model.

# Chapter 3

# Inverse Model

As discussed in Chapter 1, one method to accurately control a smart material actuator involves the use of an inverse model or compensator in the manner depicted in Figure 3.1. Linearization in this manner allows simple control techniques including PI/PID and LQR designs to be utilized over the entire range of the actuator. Given a specified displacement, we wish to determine the field necessary to bring the actuator to that value. Such inverse compensators have previously been explored for the negligible relaxation model in [16, 40]. Here we consider a method that works for both the relaxation and negligible relaxation models.

The rod model can be inverted analytically. Solving (2.39) for $v$ yields

$$v(t) = \frac{1}{d_3}u(t) - \frac{d_1}{d_3}u(t-\Delta t) - \frac{d_2}{d_3}\dot{u}(t-\Delta t). \tag{3.1}$$

Once $v$ has been determined, the required polarization can be determined from the quadratic formula; namely

$$\widehat{P}(t) = P_0 - \frac{\alpha_1}{2\alpha_2} \pm \sqrt{\left(\frac{\alpha_1}{2\alpha_2}\right)^2 + \frac{mv(t)}{a_2}} \tag{3.2}$$

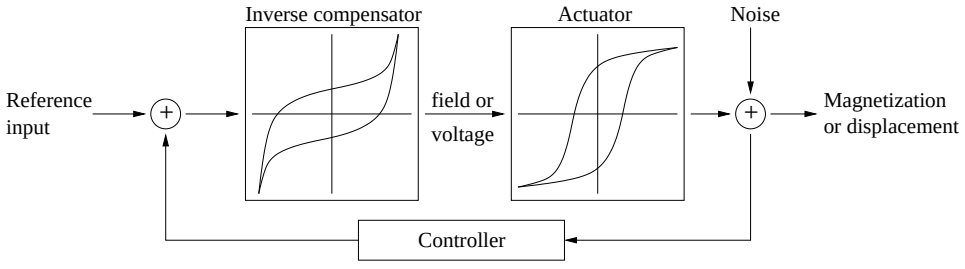



Figure 3.1: Control configuration utilizing an inverse model and a linear control law.

as long as $\alpha_2 \neq 0$. If $\alpha_2$ is 0, the equation is linear and the solution is

$$\widehat{P}(t) = P_0 + \frac{m}{\alpha_1} v(t). \tag{3.3}$$

Note that for $\alpha_2 \neq 0$, either polarization value is acceptable, as both give equivalent displacements. However, to minimize the effects of modeling error, a single branch should always be utilized.

A definition of the inverse homogenized energy model is the following: given any valid state $x_+$ and any specified $\widehat{P}$ within the operating range of the material, determine $E$ such that $P = \widehat{P}$. Unlike the rod model, however, it is not feasible to invert (2.4) analytically. Thus, we reformulate the problem as numerical root finding, namely determining the value $E$ such that for a given $x_+$ and $\widehat{P}$,

$$P(E; x_+) - \widehat{P} = 0. \tag{3.4}$$

We note that this function is monotone in $E$ (holding $x_+$ fixed). However, as discussed in Section 2.4, the function may contain a finite number of simple jumps.

## 3.1 Discontinuous Root Finding

If (3.4) was continuously differentiable, the secant method would provide nearly optimal convergence to the solution in terms of function evaluations [19]. However, for a discontinuous function it may not converge. The bisection method will converge, but often this convergence is slow. Methods exist that combine bisection with interpolation to attempt to improve this rate; see [6] for example. However, we observe better performance by exploiting the monotone nature of the function to directly improve the convergence of the bisection method.

The bisection method, or any other method which utilizes it, requires both that the root be bounded and that the function change signs within the bounds. For a monotone function, these requirements are the same. Theoretical bounds on $E$ may be computed by considering the case when all dipoles have switched. In these cases (all dipoles positive or all negative), $E$ can be determined analytically. The required value of $E$ is by necessity between these extreme cases, namely

$$\eta(\widehat{P} - P_R) \leq E \leq \eta(\widehat{P} + P_R). \tag{3.5}$$

However, direct application of the bisection method with these bounds yields slow convergence because the bounds are large.

Instead, we employ the secant method as a way to calculate tighter bounds. The secant method does not require bounds but instead uses approximate derivatives to converge to the root. Since the function is not smooth, this may not converge. However, the monotone nature of the function dictates that the secant method will always step in the correct direction. Thus, on each iteration either the secant method does not step far enough, in which case a better

approximation to $E$ is obtained, or it steps too far. In this latter case, the iterations have crossed over the root value and provide a bound on the root. We therefore apply the secant method to (3.4) first. If there is sufficient relaxation or the value $\widehat{P}$ does not lie near any discontinuities, we obtain rapid convergence. If the secant method fails to converge, the secant iterations themselves are utilized to determine the bound on the root which can then be given to bisection to find the actual root. These bounds are typically much tighter than (3.5), which improves the convergence of the bisection method. For the purposes of our algorithm, failure to converge in the secant method is defined as

$$|P(E_i; x_+) - \widehat{P}| \geq |P(E_{i-2}; x_+) - \widehat{P}|, \quad \text{for all } i \geq 3 \tag{3.6}$$

where $E_i$ is the value of $E$ determined by the ith iteration of the secant method. Such a definition is a heuristic to detect failure quickly without penalizing a single poor step.

The secant method requires an initial approximation to the derivative to compute its first iteration. One such approximation is $1/\eta$ which is the derivative if no dipoles are switching. The actual derivative may be much larger than this, but will never be smaller. Another approximation involves using the previous two time steps of the model in a manner analogous to how the secant method itself approximates derivatives. This approximation is simply $\Delta P/\Delta E$, where $\Delta P$ is the difference in $P(E; x_+)$ for the previous two time steps and $\Delta E$ is the corresponding differences in $E$. The latter approximation is deemed more accurate and is utilized when the previous two time steps are known. When this is not the case, $1/\eta$ is utilized as the initial derivative. The computed derivative value should also be checked against this lower limit on each iteration of the secant method, to trap some errors that may arise with the subtraction of nearly equal numbers.

## 3.2 Validation and Performance

The performance of the inverse model is largely dependent on the material parameters being utilized. More relaxation, smaller variances in $\nu_c$ and $\nu_I$, and slowly changing reference inputs generally yield better performance than small relaxation, highly variable distributions, or rapidly varying inputs. As an example, consider the inverse model results in Figure 3.2 for the stepped polarization shown in Figure 3.2(a). The parameters are specified as those estimated for PLZT to obtain the model fit in Figure 2.3(a). Ten intervals with 4-point Gaussian quadrature are utilized for each distribution and the time step is $\Delta t = 0.001$. As can be seen, the inverse model accounts for the relaxation in the device, decreasing the electric field as appropriate to hold the polarization constant. The error between the specified and actual values is small – five orders of magnitude below the remanence polarization.

The results in Figure 3.2 were computed with a small tolerance for the secant and bisection methods. Accuracy and effort for various tolerance levels are compiled in Table 3.1. Note that the algorithm makes efficient use of computation; reasonable accuracy levels may be obtained

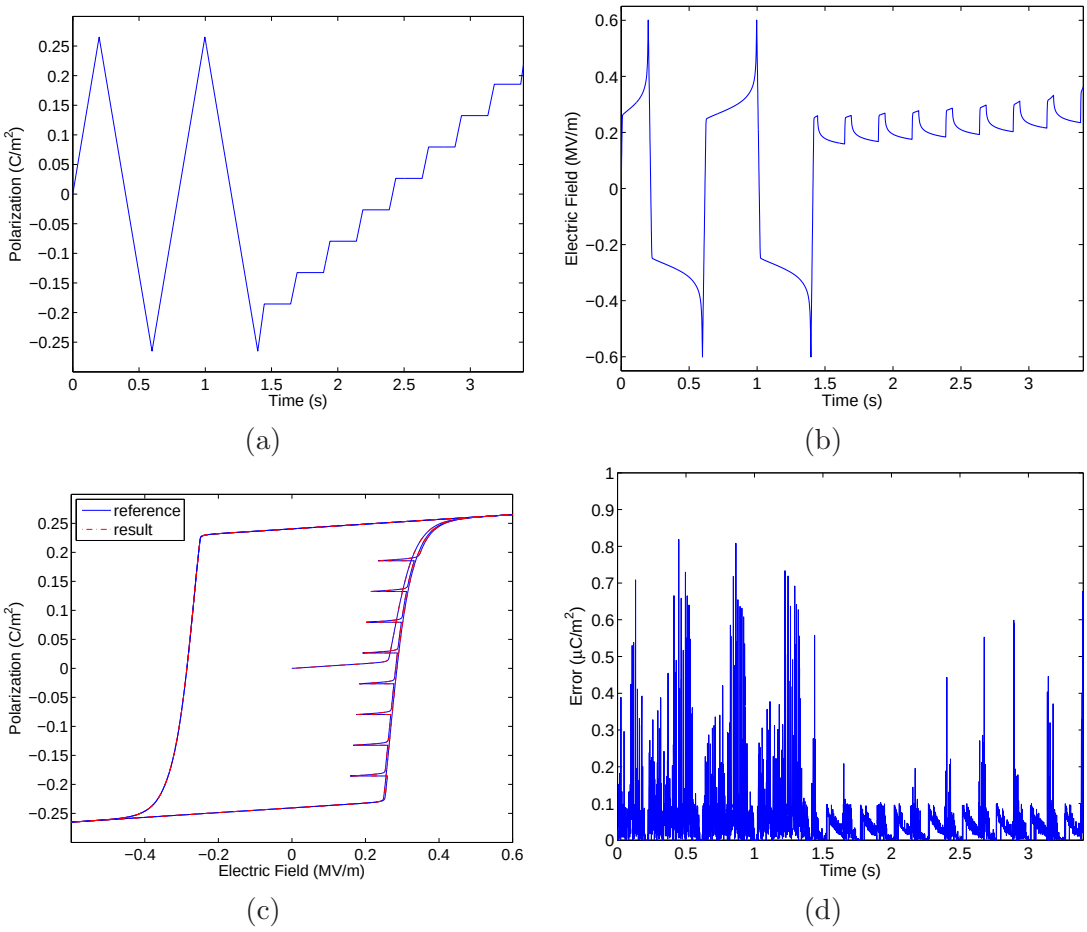


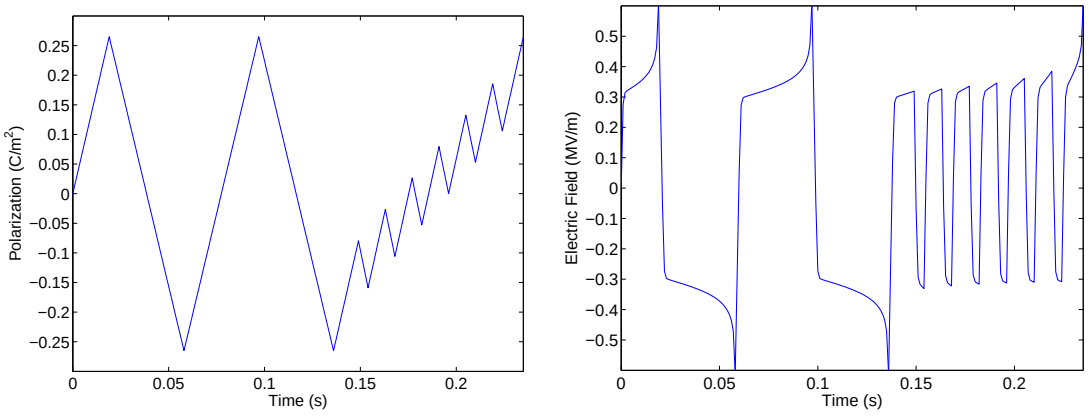

Figure 3.2: (a) Reference polarization and (b) electric field $E$ given by the inverse model with parameters estimated for PLZT. (c) Comparison of the resulting polarization $P$ and the reference polarization $\widehat{P}$. (d) Absolute error $|\widehat{P} - P|$.

Table 3.1: Effort to compute the inverse model in terms of average and maximum number of function evaluations for various error tolerances, using parameters for PLZT and the input given in Figure 3.2(a). † denotes that due to discontinuities, some time steps did not converge within the specified tolerance, and only those that did are averaged. In these cases, the maximum number of function evaluations was the user-specified maximum.

| Tolerance | Average Function Evaluations | Maximum Function Evaluations |
|---|---|---|
| $P_R \times 10^{-2} = 0.0024$ | 1.0855 | 3 |
| $P_R \times 10^{-3} = 0.00024$ | 1.2667 | 4 |
| $P_R \times 10^{-4} = 2.4 \times 10^{-5}$ | 2.0157 | 5 |
| $P_R \times 10^{-5} = 2.4 \times 10^{-6}$ | 3.5423 | 9 |
| $P_R \times 10^{-6} = 2.4 \times 10^{-7}$ | 4.8066† | user-defined |

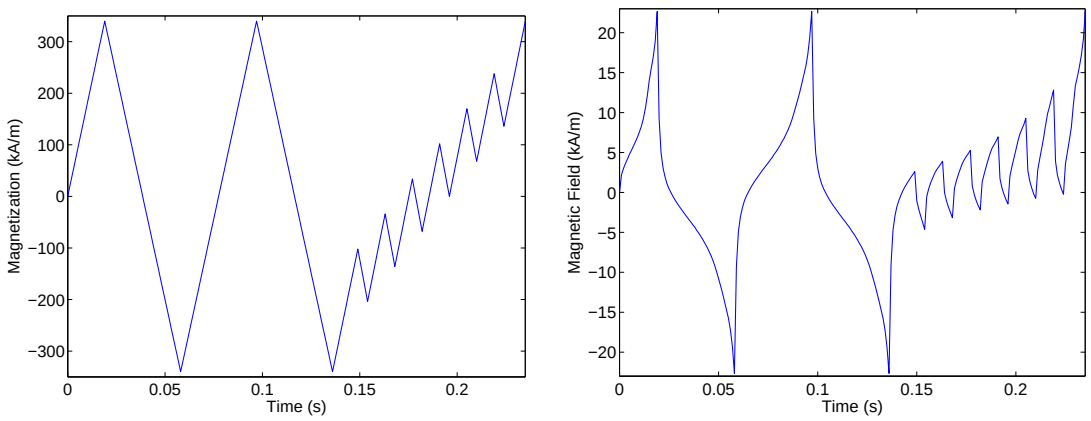
with only 2 function evaluations (evaluations of the forward homogenized energy model) per iteration on average, or 5 in the worst case. However, this efficiency is dependent on the input. Figure 3.3(a) shows an input polarization which changes at roughly an order of magnitude higher rate. Using the same material parameters and time step on this input yields the results in Figure 3.3(b) and Table 3.2. The same accuracy level as before requires almost 5 function evaluations on average and 16 in the worst case. While increased, this is still easily attainable. The level of effort also depends on the material being modeled. Scaling Figure 3.3(a) to an appropriate level and applying parameters for a Terfenol-D rod yields the results in Figure 3.4 and Table 3.3. The Terfenol-D parameters require greater work for the same level of accuracy, but relatively speaking they are able to achieve a greater accuracy than that achieved in the PLZT case.

Finally, we note that these observations apply directly to the case when displacement is utilized as input. Figure 3.5 and Table 3.4 show the accuracy and level of effort for the Terfenol-D parameters with strain (displacement) as an input. The tolerance for the root finding algorithm was also specified in terms of displacement in this case, with $P(E; x_+)$ input to the rod model on each iteration of secant or bisection to determine the error. We note that the inverse compensator provides an accurate way to control the device while maintaining a reasonable number of computations.




(a) (b)

Figure 3.3: (a) Reference polarization and (b) electric field $E$ given by the inverse model with parameters estimated for PLZT.




(a) (b)

Figure 3.4: (a) Reference magnetization and (b) magnetic field $H$ given by the inverse model with parameters estimated for Terfenol-D.

Table 3.2: Effort to compute the inverse model in terms of average and maximum number of function evaluations for various error tolerances, using parameters for PLZT and the input given in Figure 3.3(a). † denotes that due to discontinuities, some time steps did not converge within the specified tolerance, and only those that did are averaged. In these cases, the maximum number of function evaluations was the user-specified maximum.

| Tolerance | Average Function Evaluations | Maximum Function Evaluations |
|---|---|---|
| $P_R \times 10^{-2} = 0.0024$ | 2.8776 | 10 |
| $P_R \times 10^{-3} = 0.00024$ | 3.9286 | 13 |
| $P_R \times 10^{-4} = 2.4 \times 10^{-5}$ | 4.8027 | 16 |
| $P_R \times 10^{-5} = 2.4 \times 10^{-6}$ | $7.9252^{\dagger}$ | user-defined |
| $P_R \times 10^{-6} = 2.4 \times 10^{-7}$ | $9.3187^{\dagger}$ | user-defined |

Table 3.3: Effort to compute the inverse model in terms of average and maximum number of function evaluations for various error tolerances, using parameters for Terfenol-D and the input given in Figure 3.4(a).

| Tolerance | Average Function Evaluations | Maximum Function Evaluations |
|---|---|---|
| $M_R \times 10^{-2} = 3400$ | 2.8673 | 9 |
| $M_R \times 10^{-3} = 340$ | 5.4354 | 16 |
| $M_R \times 10^{-4} = 34.0$ | 10.7143 | 22 |
| $M_R \times 10^{-5} = 3.40$ | 14.2993 | 25 |
| $M_R \times 10^{-6} = 0.340$ | 17.2279 | 29 |

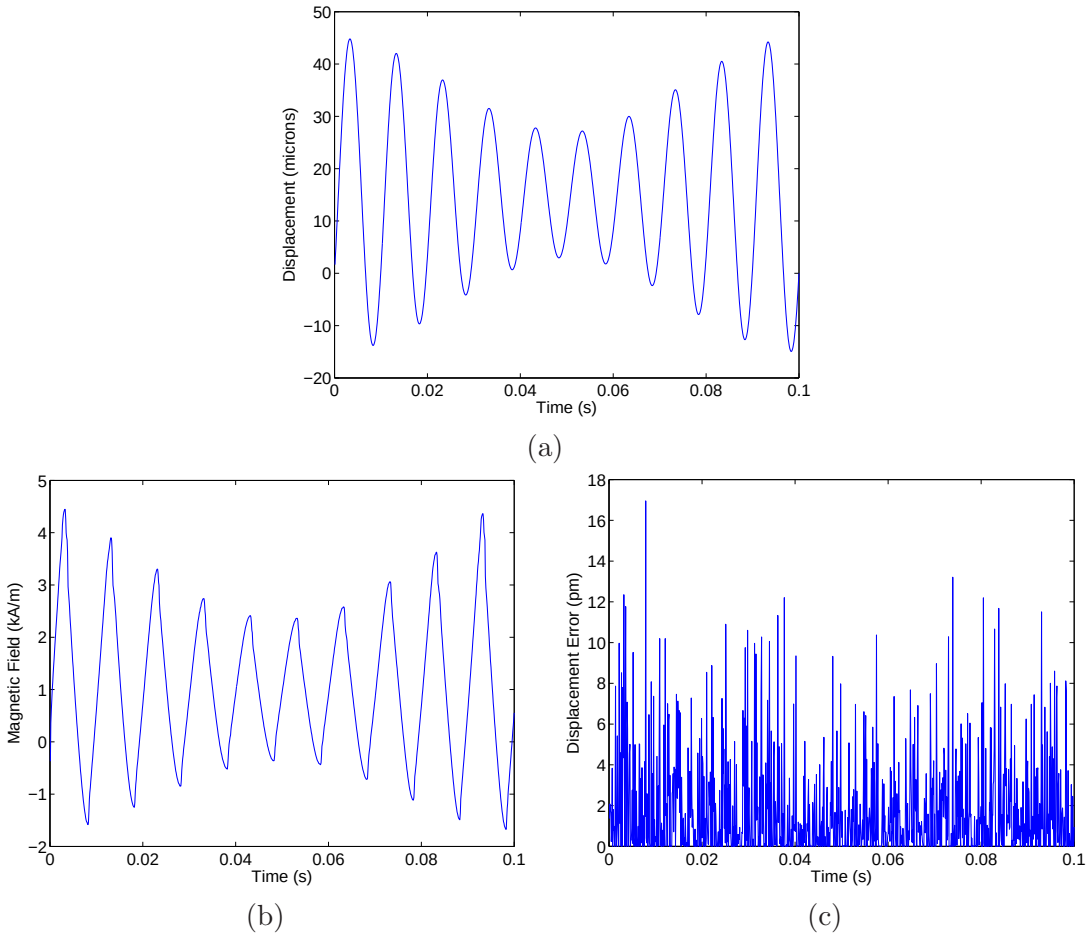



Figure 3.5: Results for the combined inverse model with parameters estimated for a Terfenol-D rod. (a) Specified displacement $\widehat{u}$ and (b) magnetic field $H$ provided by the inverse algorithms. (c) Absolute error $|\widehat{u} - u|$ for displacement $u$ given by the rod model.

Table 3.4: Effort to compute the inverse model in terms of average and maximum number of function evaluations utilizing parameters estimated for a Terfenol-D rod and the reference displacement given in Figure 3.5(a).

| Error tolerance | Average model evaluations | Maximum model evaluations |
|---|---|---|
| 1 micron | 2.1240 | 8 |
| 100 nm | 2.2880 | 9 |
| 10 nm | 6.5170 | 19 |
| 1 nm | 11.6500 | 22 |
| 100 pm | 14.7040 | 25 |
| 10 pm | 17.6930 | 29 |

# Chapter 4

# Stress-dependent 90° Switching

The previous discussion is based on the assumption that dipoles or moments are aligned either in the direction of the field or diametrically opposite to it – this characterizes the internal energy associated with field-induced 180° switching. As illustrated in Figure 4.1, the application of stresses perpendicular to the poling direction additionally produces 90° switching. Figure 4.2 shows that this phenomenon can be significant by comparing polarization data for PLZT at two different stress levels [20]. It is illustrated in [1, 2] that whereas the strict incorporation of 90° switching requires a 2-D or 3-D energy landscape, reasonable approximations can be obtained for a number of applications with 1-D Helmholtz relations whose minima correspond with 180° and 90° dipole positions. We summarize this latter case.

## 4.1 Homogenized Energy Model – Theoretical Development

The homogenized energy model for stress-dependent materials is

$$P(E,\sigma;x_+,x_-) = \int_0^\infty \int_{-\infty}^\infty [\overline{P}(E+E_I,\sigma;E_c;x_+,x_-)](t)\nu_c(E_c)\nu_I(E_I)dE_I dE_c, \tag{4.1}$$

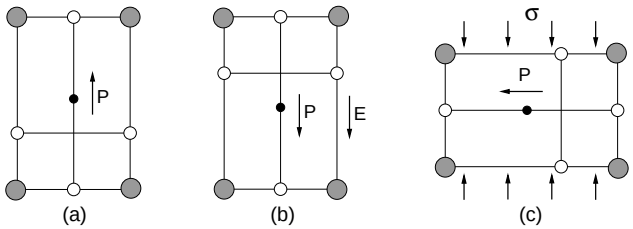



Figure 4.1: Possible polarization states of a single crystal including (a) the state with no applied field, (b) the 180° switch which occurs with a sufficiently strong negative field, and (c) a 90° switch induced by stress.

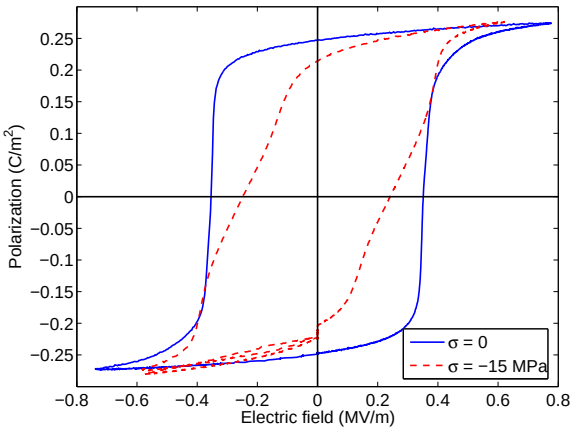



Figure 4.2: PLZT data showing the effect of stress on polarization [20].

or the magnetic analogue, where $\sigma$ denote applied stress and both $x_+$ and $x_-$ are now dipole fractions of positively and negatively oriented dipoles respectively. The fraction of 90° orientated dipoles follows from these by $x_{90} = 1 - x_+ - x_-$. The densities are subject to the same constraints given in Chapter 2, and the same caveats apply to their continuity and integration.

### 4.1.1 Kernel Development

Following the development in [1, 2], the Helmholtz energy quantifying the internal energy associated with the three dipole states is taken to be

$$\psi_p(P) = \begin{cases} \eta(P + P_R)^2/2, & P \le -P_I \\ \eta_1(P + P_m)^2/2 + \beta, & -P_I < P < -P_{90I} \\ \eta_2 P^2/2 + \Delta, & |P| < P_{90I} \\ \eta_1(P - P_m)^2/2 + \beta, & P_{90I} < P < P_I \\ \eta(P - P_R)^2/2, & P \le -P_I \end{cases} \tag{4.2}$$

where

$$P_m = \frac{\eta(P_I - P_R)P_{90I} - \eta_2 P_{90I} P_I}{\eta(P_I - P_R) - \eta 2 P_{90I}}, \qquad \eta_1 = \eta \frac{P_I - P_R}{P_I - P_m},$$

$$\beta = \frac{\eta}{2}(P_I - P_R)^2 - \frac{\eta_1}{2}(P_I - P_m)^2, \qquad \Delta = \frac{\eta_1}{2}(P_{90I} - P_m)^2 + \beta - \frac{\eta_2}{2}P_{90I}^2.$$

The parameters $P_{90I}$ and $\eta_2$ are analogous to the inflection point $P_I$ and reciprocal slope $\eta$, except they apply to dipoles that are normal to the applied field (i.e., in a 90° orientation). Note that due to the one dimensional context, we consider all normal directions as the same state. This distinguishes this from a true 2-D approach but greatly simplifies the computation.

The energy relation (4.2) incorporates the internal energy associated with dipole switching but neglects electromechanical coupling and elastic contributions. Assuming negligible stress, these terms are also negligible and as such were excluded from the discussion in Chapter 2; however, when the stress is not negligible these terms must also be included. When constructing an electromechanical coupling component for the Helmholtz energy, we consider operation about a poled state $P_0$ – for many applications $P_0$ is the positive or negative macroscopic remanence polarization. The electromechanical coupling energy is then taken to be

$$\begin{aligned}\psi_{es}(P,\varepsilon) &= -q\varepsilon(P-P_0)^2 \\ &= -a\varepsilon P - q\varepsilon P^2 - \widehat{q}\varepsilon,\end{aligned} \tag{4.3}$$

where $q$ is the electrostrictive coupling coefficient, $a = -2P_0q$ is the piezoelectric coupling coefficient, and $\varepsilon$ is the strain. The elastic energy for uniaxial regimes is

$$\psi_{el}(\varepsilon) = \frac{1}{2}Y\varepsilon^2 \tag{4.4}$$

where $Y$ is the Young's modulus of the material. The total Helmholtz energy is thus

$$\psi(P,\varepsilon) = \psi_p(P) + \psi_{es}(P,\varepsilon) + \psi_{el}(\varepsilon). \tag{4.5}$$

The associated Gibb's energy is

$$G = \psi - EP - \sigma\varepsilon \tag{4.6}$$

where $\sigma$ is the stress. As detailed in [1, 2], enforcement of the equilibrium condition $\frac{\partial G}{\partial \varepsilon} = 0$ allows the strain to be written as

$$\begin{aligned}\varepsilon &= Y^{-1}(\sigma + aP + qP^2 + \widehat{q}) \\ &= Y^{-1}(\sigma + aP + qP^2) + \varepsilon_R\end{aligned} \tag{4.7}$$

where $\varepsilon_R$ is a remanent strain. Since we are typically interested in strains generated with respect to the polarized or biased state, we take $\varepsilon_R = 0$ is subsequent discussions. Substitution of (4.7) into (4.6), in light of this assumption, yields a Gibb's formulation posed solely in terms of the input field and stress.

Due to the definition (4.2) in the Helmholtz energy, 4 switching points occur: from either positive or negative to 90° and from 90° to either 180° orientation. We let $E_c$ denote the field at which the polarization switches to a positive orientation and specify the transition to a ninety degree orientation through the independent parameter $P_I$. These two quantities are easier to approximate from data under low to intermediate stress levels than the coercive point or inflection point for 90° orientations. From these we derive the input electric field levels at

which the dipole switches to 90° orientations occur, namely

$$\begin{aligned} E_{+90} &= \frac{-1}{Y}(2q^2P_I^3 + 3aqP_I^2 + (a^2 - \eta Y + 2q\sigma)P_I + a\sigma + \eta Y P_R), \\ E_{-90} &= \frac{1}{Y}(2q^2P_I^3 - 3aqP_I^2 + (a^2 - \eta Y + 2q\sigma)P_I - a\sigma - \eta Y P_R). \end{aligned} \tag{4.8}$$

Note that this implies 180° switches will always be symmetric whereas if linear electromechanical coupling is significant, the 90° switches may not be symmetric. This was chosen to match behavior observed in stressed ferroelectric materials.

Additionally, note that when constructing the macroscopic model (4.1), $E_c$ and $E_I$ are assumed to be manifestations of underlying densities. The inflection point $P_I$ is not treated as a stochastic variable, but rather as a constant. One could argue theoretically that this too should be distributed. The practical consequences of this, however, prohibit use of a such a formulation. With two densities to integrate, the run time is proportional to the square of of the number of quadrature intervals chosen for each density. If a third density were to be added, the run-time would be proportional to the cube of the the number of quadrature intervals, which becomes unacceptably slow even for reasonable number of intervals. It also greatly increases the complexity of estimating parameters for the model, both because it increases the numbers of parameters to estimate and because a large amount of independent data is needed to ensure any parameters estimated are physical and not merely "curve-fits" to the data. Because of these difficulties, the model is approximated by letting $P_I$ be a constant. However, other choices are possible; one could choose to distribute $P_I$ and leave $E_c$ fixed, or one could distribute the difference in coercive points in a matter analogous to the methods used for shape memory alloys. Each of these should be understood as various methods of approximating the physical case and each give similar results.

For regimes in which thermal activation is negligible, the kernel $\overline{P}$ is obtained from the condition $\frac{\partial G}{\partial P} = 0$. This may be obtained numerically for each of the cases $P \geq P_I$, $P \leq -P_I$ and $|P| \leq P_{90I}$. While it is possible that more than one real root may exist, only one is typically within a valid range for $P$. As such, it is not difficult to numerically determine the correct solution.

To incorporate thermal activation, we balance (4.6) with relative thermal energy via the Boltzmann relation. This yields three average polarization relations

$$\begin{gathered} \langle P_+ \rangle = \frac{\int_{P_I}^{\infty} P \exp\left(\frac{-G(E,P)V}{kT}\right) dP}{\int_{P_I}^{\infty} \exp\left(\frac{-G(E,P)V}{kT}\right) dP}, \qquad \langle P_- \rangle = \frac{\int_{-\infty}^{-P_I} P \exp\left(\frac{-G(E,P)V}{kT}\right) dP}{\int_{-\infty}^{-P_I} \exp\left(\frac{-G(E,P)V}{kT}\right) dP}, \\ \langle P_{90} \rangle = \frac{\int_{-P_{90I}}^{P_{90I}} P \exp\left(\frac{-G(E,P)V}{kT}\right) dP}{\int_{-P_{90I}}^{P_{90I}} \exp\left(\frac{-G(E,P)V}{kT}\right) dP} \end{gathered} \tag{4.9}$$

corresponding to the three dipole states. This resulting kernel is

$$\overline{P} = x_+ \langle P_+ \rangle + x_- \langle P_- \rangle + (1 - x_+ - x_-) \langle P_{90} \rangle . \tag{4.10}$$

To quantify the evolution of the phase fractions $x_+$ and $x_-$, it is necessary to specify the likelihoods that dipoles switch between various states. In this formulation, we assume that dipoles can only switch between adjacent states; i.e., a negative dipole must switch to 90° and not directly to positive. This behavior is based on the behavior observed in Figure 4.3 for the $\sigma = -15$ MPa data which undergoes a 90° switch between the 180° states. We note that the intermediate 90° switch can be extremely rapid so that the model also accurately characterizes the $\sigma = 0$ data for which the effect of 90° is minimal – see Figure 4.3 in Section 4.2.

This simplifying assumption yields the likelihood relations

$$\begin{aligned}
p_{+90} &= \frac{\exp\left(\frac{-G(E+E_I,P_I)V}{kT}\right) dP}{\tau(T) \int_{P_I}^{\infty} \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP}, \qquad & p_{-90} &= \frac{\exp\left(\frac{-G(E+E_I,-P_I)V}{kT}\right) dP}{\tau(T) \int_{-\infty}^{-P_I} \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP}, \\
p_{90+} &= \frac{\exp\left(\frac{-G(E+E_I,P_{90I})V}{kT}\right) dP}{\tau(T) \int_{-P_{90I}}^{P_{90I}} \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP}, \qquad & p_{90-} &= \frac{\exp\left(\frac{-G(E+E_I,-P_{90I})V}{kT}\right) dP}{\tau(T) \int_{-P_{90I}}^{P_{90I}} \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP}
\end{aligned} \tag{4.11}$$

or

$$\begin{aligned}
p_{+90} &= \frac{\int_{P_I}^{P_I+\epsilon} \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP}{\tau(T) \int_{P_I}^{\infty} \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP}, \qquad & p_{-90} &= \frac{\int_{-P_I-\epsilon}^{-P_I} \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP}{\tau(T) \int_{-\infty}^{-P_I} \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP}, \\
p_{90+} &= \frac{\int_{P_{90I}-\epsilon}^{P_{90I}} \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP}{\tau(T) \int_{-P_{90I}}^{P_{90I}} \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP}, \qquad & p_{90-} &= \frac{\int_{-P_{90I}}^{-P_{90I}+\epsilon} \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP}{\tau(T) \int_{-P_{90I}}^{P_{90I}} \exp\left(\frac{-G(E+E_I,P)V}{kT}\right) dP}
\end{aligned} \tag{4.12}$$

where the first is understood to be an approximate in the statistical sense but is commonly employed for its more efficient computation, and the latter is to be preferred for statistical analysis. Given these, the dipole fractions evolve according to the differential equation

$$\begin{bmatrix} \dot{x}_- \\ x_+ \end{bmatrix} = \begin{bmatrix} -p_{-90} - p_{90-} & -p_{90-} \\ -p_{90+} & -p_{+90} - p_{90+} \end{bmatrix} \begin{bmatrix} x_- \\ x_+ \end{bmatrix} + \begin{bmatrix} p_{90-} \\ p_{90+} \end{bmatrix} . \tag{4.13}$$

Again, a backward Euler method typically provides sufficient accuracy while minimizing computation. We note that unlike the stress-free case, the integrals in (4.9), (4.11) and (4.12) cannot be reduced to error functions and exponentials and hence must be approximated with a quadrature method.

## 4.2 Comparison of Model Response to Data

The formulation of the homogenized energy model given in the previous section is able to accurately characterize stress-dependent effects over a wide range of stress inputs. This is illustrated for a PLZT sample compressed at 0, 6, 10, and 15 MPa in Figure 4.3. A single set of parameters is used in all four plots, with only the applied stress level modified in each. The change in curvature exhibited by the data with $\sigma = -15$ MPa illustrates that the manifestation of 90° switching becomes increasingly prominent as prestress levels increase. Note that this change in curvature is matched in the model. More details on the experimental setup may be found in [20].

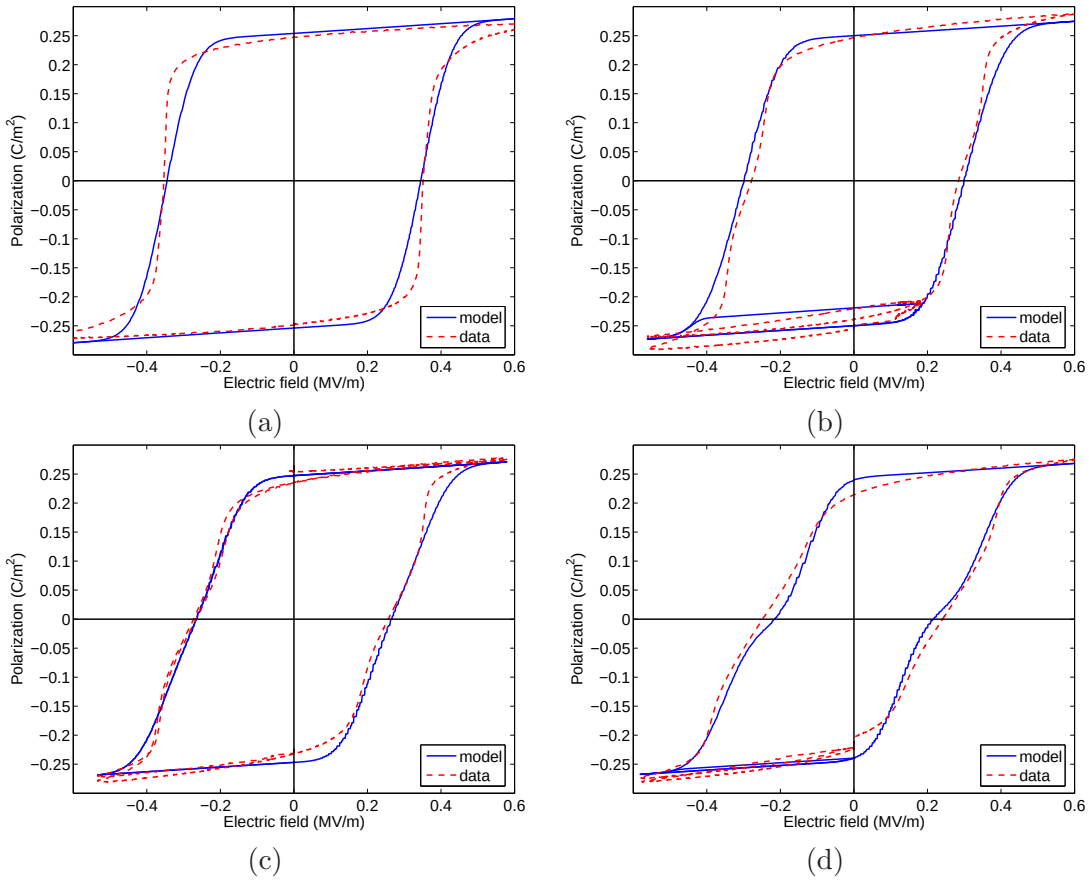



Figure 4.3: Comparison of $E - P$ curves from the negligible relaxation model with PLZT data from [20] for compressive prestresses of (a) $\sigma_0 = 0$ MPa, (b) $\sigma_0 = -6$ MPa, (c) $\sigma_0 = -10$ MPa and (d) $\sigma_0 = -15$ MPa.

## 4.3 Look-up Tables for Faster Computation

### 4.3.1 Negligible Relaxation Algorithm

For the negligible relaxation case, there are no likelihoods to compute. However, minimization of the Gibb's energy (4.6) requires computing the zeros of a cubic polynomial. Given the piecewise definition of $\psi_P$, this effectively yields three different polynomials to minimize corresponding to the three possible dipole orientations. As reported in [4], direct calculation of these minima increases the computation time required for the negligible relaxation model to over an order of magnitude more than was required in the two-state case. This can be mitigated by performing the minimization in advance and storing the results in a look-up table. The real-time portion of the code need only access the closest table value or possibly linearly interpolate between table values. Although this increases the memory required, implementation of this approximation ran almost as fast as the two-state case (11 seconds instead of 9 for 120,000 timesteps and 80 quadrature points on each distribution) [4].

The memory requirement here is increased even over the relaxation case for the two-state model. Instead of a single table, three tables are needed, one for each case. Additionally, in general both applied field $E$ and stress $\sigma$ are independent inputs. This requires that each table contain two dimensions, one for $E_e = E + E_I$ and one for $\sigma$. Table bounds are set by bounding these inputs either based on operating conditions or by material limits. For example, for any material there is a certain stress beyond which the material breaks; it makes no sense to store stresses larger than this in the table. Some memory may be conserved by noting that for values of $E_e < -E_c[N_c]$, the minima corresponding to positive orientations need not be computed and stored because the dipole will have always switched to a different orientation. Likewise negative orientations need not be computed for $E_e > E_c[N_c]$, and the ninety degree orientation does not need computed for values at either extreme.

If memory requirements prove problematic, polynomial interpolants/least squares approximations may be performed here. However, since the data now contains two dimensions, calculation of approximations in this manner is numerically intensive. These calculations may be performed offline; however, the look-up tables or polynomial approximations are dependent on the material parameters in this formulation. Thus, polynomial interpolants must be recalculated if the material parameters change. As such, this approach should only be taken when parameters may be held fixed for an extended period.

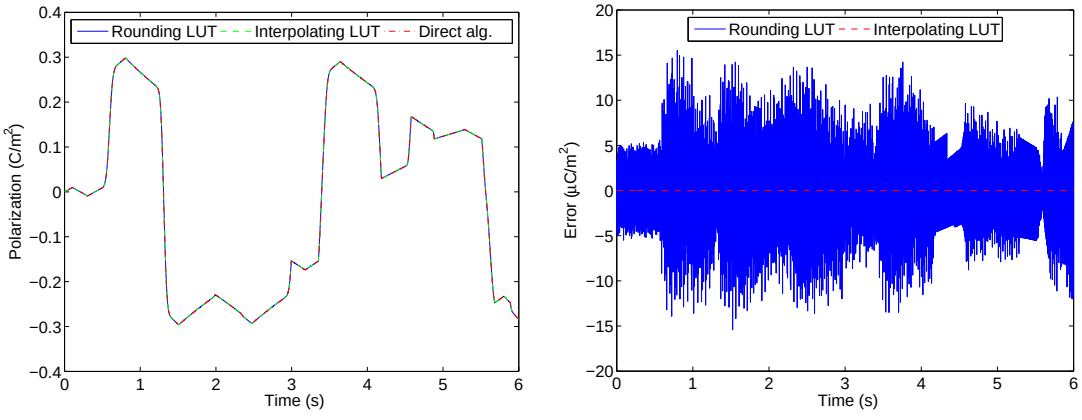
The increased memory requirement over the two-state formulation may pose a concern for some embedded applications, but reasonable accuracy is obtainable with moderately small tables. Figure 4.4 compares the direct implementation of the negligible relaxation, stress-dependent model with implementations employing this look-up table approximation, and Table 4.1 gives the root mean square error for various size tables. The smallest look-up table size given in the Table 4.1 requires only 3/8 of a megabyte of storage for all three tables but gives a root mean square error almost four orders of magnitude below the remanence polarization

Table 4.1: Comparison of the direct and look-up table implementations of the stress-dependent, negligible relaxation algorithm. Run times are in seconds and errors are root mean square values expressed as percentages of the saturation polarizations (0.3 C/m$^2$). The material parameters were chosen to match PLZT and run-times are for 120,000 time steps on a Pentium IV Xeon 1.7 GHz running Linux.

| | | | Rounding | | Interpolating | |
|---|---|---|---|---|---|---|
| Field points | Stress points | Setup time | Run-time | error | Run-time | error |
| 125 | 125 | 0.95 | 10.36 | 0.01258 | 10.61 | $9.26 \times 10^{-5}$ |
| 250 | 125 | 1.52 | 10.45 | 0.00922 | 10.44 | $2.53 \times 10^{-5}$ |
| 500 | 125 | 3.04 | 10.34 | 0.00907 | 10.45 | $8.71 \times 10^{-6}$ |
| 125 | 250 | 1.84 | 10.42 | 0.01000 | 10.45 | $9.03 \times 10^{-5}$ |
| 250 | 250 | 3.67 | 10.38 | 0.00514 | 10.40 | $2.30 \times 10^{-5}$ |
| 500 | 250 | 7.40 | 10.49 | 0.00487 | 10.70 | $6.26 \times 10^{-6}$ |
| 125 | 500 | 4.89 | 10.60 | 0.00922 | 10.66 | $8.97 \times 10^{-5}$ |
| 250 | 500 | 9.79 | 10.50 | 0.00342 | 10.71 | $2.24 \times 10^{-5}$ |
| 500 | 500 | 19.71 | 10.56 | 0.00298 | 10.63 | $5.67 \times 10^{-6}$ |

with rounding to the nearest table element or approximately six orders of magnitude below the remanence polarization with linear interpolation. Even the largest listed look-up table table size, which is almost 20 times more accurate, requires only 6 MB of storage. Also note that the time required to compute the look-up table (listed as setup time in Table 4.1) is directly proportional to the size of the table, but within certain architecture dependent limits the portion which must be run in real-time is independent of table size.

Figure 4.4 and Table 4.1 also demonstrate that a simple and computationally efficient linear interpolation can increase the accuracy of the approximation by two to three orders of magnitude.




(a) (b)

Figure 4.4: Comparisons of the (a) polarization and (b) error in polarization given by a direct implementation of the stress-dependent, negligible relaxation model and an implementation based on look-up tables (LUT) with and without interpolating between the table elements.

Since the tables are two-dimensional, interpolation was performed on one dimension at a time. Allowing $E_1$ and $\sigma_1$ to represent the closest table value below $E_e$ and $\sigma$, and $E_2$, $\sigma_2$ to be the values just above the desired value, this can be written in one equation as

$$\begin{aligned}\langle P\rangle =& P_{11}\left(1-\frac{E_e-E_1}{E_{step}}\right)\left(1-\frac{\sigma-\sigma_1}{\sigma_{step}}\right)+P_{12}\frac{E_e-E_1}{E_{step}}\left(1-\frac{\sigma-\sigma_1}{\sigma_{step}}\right)\\ &+P_{21}\left(1-\frac{E_e-E_1}{E_{step}}\right)\frac{\sigma-\sigma_1}{\sigma_{step}}+P_{22}\frac{(E_e-E_1)(\sigma-\sigma_1)}{E_{step}\sigma_{step}}.\end{aligned} \tag{4.14}$$

Here $\langle P\rangle = \langle P_-\rangle$, $\langle P_{90}\rangle$ or $\langle P_+\rangle$, values from the table are given by $P_{11} = \langle P\rangle(E_1,\sigma_1)$, $P_{12} = \langle P\rangle(E_1,\sigma_2)$, etc., $E_{step}$ is the distance between two field values in the table, and $\sigma_{step}$ the distance between stress values. This interpolation calculation did not significantly increase the run-time of the algorithm on the tested architecture (Intel Pentium IV Xeon 1.7 GHz running Linux). The proximity in run-times for the non-interpolated and interpolated cases appears to be partially an architecture dependent feature, and other architecture may give a slightly larger disparity between the run-times. Nevertheless, interpolation remains an efficient operation and the increase in accuracy it allows suggests it should be employed in most situations that use this formulation of the model.

### 4.3.2 Thermal Relaxation Algorithm

Similar approaches are possible with the relaxation case, except that the memory required to store the look-up tables as well as the time needed to compute the tables greatly increases. Since symmetry may not be maintained for some parameter choices, this necessitates seven tables – three for average polarizations and four for switching likelihoods. Further, the coercive point enters the formulations instead of being a value checked separately. This requires each table have a dimension for effective field, another for stress, and a third for coercive field. Since the coercive field points are known *a priori* the elements in the table can be chosen as the quadrature points for the distribution over $E_c$. This choice introduces no approximation error for the coercive field component of the table and typically decreases the look-up table size as well. Nevertheless, the table is still three-dimensional and thus the memory required is typically in the tens to hundreds of megabytes. This is easily obtained on modern desktop or laptop computers but likely not possible in microprocessor or embedded applications. If achievable in available memory, the look-up table does provide over three orders of magnitude improvement in run-time over a direct implementation in the real-time portion of the computations. In fact, excluding the time needed to calculate the look-up tables, the run-time is less than 25% slower than the two-state relaxation algorithm [4]. As before, polynomial approximations could be employed here as well, but since they are typically numerically intensive to compute and must be recalculated when the material parameters change, this is only advisable when material parameters are fixed over an extended period of time.

Table 4.2: Comparison of the direct and look-up table for the stress-dependent, relaxation algorithm, with a varying input stress and 80 quadrature points for each density. Errors are root mean square values expressed as percentages of the saturation polarizations (0.3 $C/m^2$). Run-times are for 600 time steps on a Pentium IV Xeon 1.7 GHz running Linux.

| Field points | Stress points | Setup time (s) | Run-time (s) | RMS Error (%) |
|---|---|---|---|---|
| 25 | 25 | 107.20 | 0.28 | 2.42363 |
| 50 | 25 | 173.97 | 0.28 | 1.47033 |
| 100 | 25 | 307.18 | 0.29 | 1.16744 |
| 25 | 50 | 215.55 | 0.28 | 2.52963 |
| 50 | 50 | 343.12 | 0.28 | 1.21354 |
| 100 | 50 | 615.48 | 0.29 | 1.06211 |
| 25 | 100 | 426.51 | 0.28 | 2.53719 |
| 50 | 100 | 693.13 | 0.28 | 1.21641 |
| 100 | 100 | 1225.09 | 0.29 | 1.06611 |

Figure 4.5 compares the look-up table approximation to a direct formulation of the model, and Table 4.2 gives run-times and accuracy for various table sizes. Note that even with the small tables employed here the time needed to calculate the look-up tables is significant. However, even these small tables achieve a root mean square error of approximately 1% to 2% of the remanence polarization. Thus, look-up tables present an option for real-time model implementation and control when both stress-dependent 90° switching and thermal relaxation must be considered, but because of the memory required this formulation should only be used when necessary and not treated as a general model for all situations.

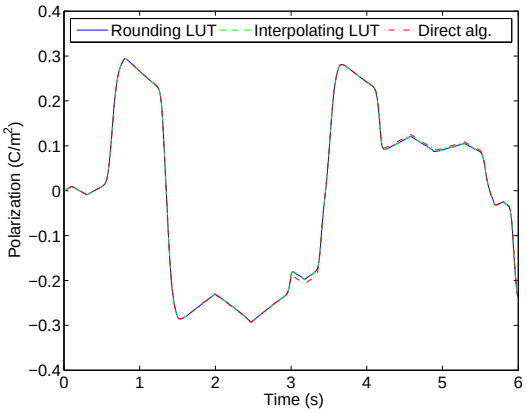



Figure 4.5: Comparisons of the polarization given by a direct implementation of the stress-dependent relaxation model and an implementation based on look-up tables.

## 4.4 Inverse Model

Application of the inverse model to the the stress-dependent 90° switching formulation requires clear specification of the model input and output. The forward model requires that both stress and field be specified. In most applications typically only one of these is controllable. Thus, the inverse model will require both polarization/magnetization and stress as inputs and produce field as its product, or it will require polarization/magnetization and field as input and return the required stress level. In either case, the inverse model may be extended to take take displacement values as input instead of polarization as described in Chapter 3.

Regardless of whether stress or field is the known quantity, the inverse model is formulated as

$$P(E, \sigma; x_+, x_-) - \widehat{P} = 0. \tag{4.15}$$

where $P(E, \sigma; x_+, x_-)$ is the homogenized energy model (4.1) and $\widehat{P}$ is the desired polarization. We focus on the typical case where the stress $\sigma$ is known and the inverse model seeks to ascertain the required field level. The formulation which seeks to determine stress proceeds analogously except that varying stress alone will only increase or decrease the percentage of dipoles in the 90° orientation and as such cannot effectively change the sign of the polarization. Thus, the polarization given to the inverse model in this case should always be kept consistent with the known input field to ensure a solution exists.

We assume the initial material state is known and updated as we proceed through the inverse model, just as we did in Chapter 3. With this assumption (4.15) is a scalar function of $E$ and may be solved with secant iterations followed by bisection, as it was in Chapter 3. Using a known but varying stress and the material parameters determined from a least-squares fit to PLZT this yields the results depicted in Figure 4.6 and Table 4.3. Note the the actual or apparent discontinuities introduced by the quadrature formulation greatly limit the accuracy for this particular choice of parameters, and that as expected this effect is much more pronounced

Table 4.3: Effort to compute the inverse model in terms of average and maximum number of function evaluations for various error tolerances, using parameters for PLZT and the input given in Figure 4.6(a) and (b). † denotes that due to discontinuities, some time steps did not converge within the specified tolerance, and only those that did are averaged. In these cases, the maximum number of function evaluations was the user-specified maximum.

| | Negligible Relaxation Function Evaluations | | Thermal Relaxation Function Evaluations | |
|---|---|---|---|---|
| Tolerance | Average | Maximum | Average | Maximum |
| $P_R \times 10^{-2} = 0.0024$ | $1.6660^{\dagger}$ | user-defined | 1.1850 | 2 |
| $P_R \times 10^{-3} = 0.00024$ | $2.8296^{\dagger}$ | user-defined | 1.3294 | 2 |
| $P_R \times 10^{-4} = 2.4 \times 10^{-5}$ | $4.7263^{\dagger}$ | user-defined | $2.0284^{\dagger}$ | user-defined |

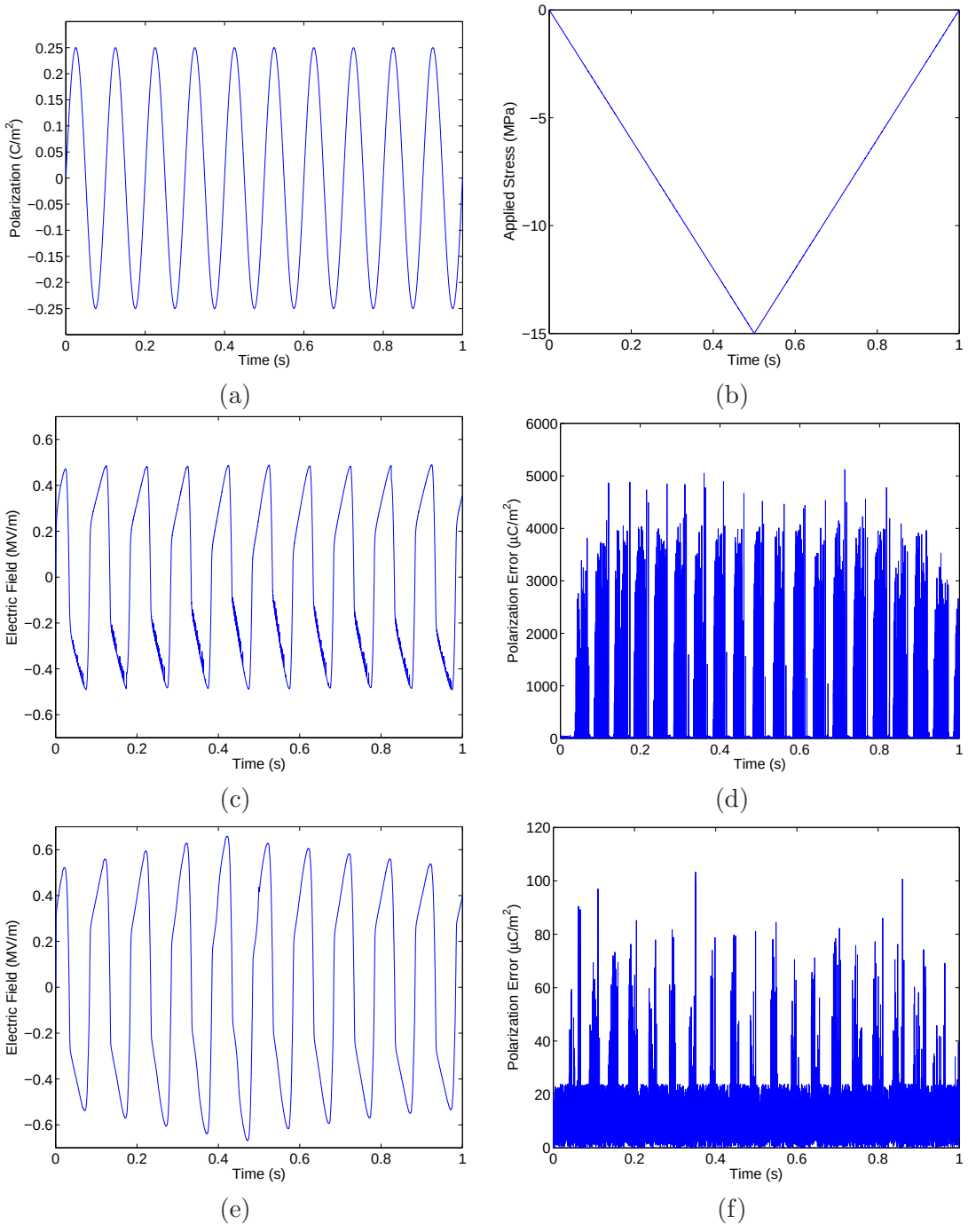



Figure 4.6: Results of the inverse model for the input polarization (a) and stress (b) with $\Delta t = 1 \times 10^{-4}$. The resulting electric field and error in polarization for the negligible relaxation model are given in (c) and (d) respectively, while the corresponding results for the relaxation model are given in (e) and (f). Material parameters were chosen to match PLZT and the tolerance for the root-finding method was set to $2.4 \times 10^{-5}$.

in the negligible relaxation model. Nevertheless, the given accuracy is sufficient for many control applications.

# Chapter 5

# Shape Memory Alloys

As mentioned in the introduction, shape memory alloy (SMA) actuators are significantly more complex to model than their ferroelectric or ferromagnetic counterparts. At low temperatures, the material is composed of lattice elements in one or more martensite variants. In general there are 24 martensite variants. However, we will consider the one-dimensional case in which there are only two, which we will refer to as positive and negative. The two different martensite variants give two different ways the lattice structure may be skewed. The variant which corresponds to an elongation along of the axis of interest as stress increases will always be referred to here as positive. At moderately high temperatures, the material will prefer a highly symmetric austenite orientation, and only under application of a sufficient load will any martensite variants remain. As the temperature continues to increase, the material will reach a point where austenite is the only stable variant regardless of stress level. Some SMA materials will also briefly exhibit an intermediate rhombohedral phase during transformation; however, this phase is not considered herein.

Shape memory alloy actuators typically employ the shape memory effect with a constant or varying applied load to produce a strain. The applied load stretches the SMA as all crystal elements transition to the same martensite state, and heating the actuator causes the SMA to transform to its unstretched austenite state. While the ferroelectric and ferromagnetic actuator response was temperature-dependent, we were able to eliminate temperature from consideration by assuming it was constant. Such an assumption is not valid for SMA transducers. This holds even for SMA devices employed in the pseudoelastic regime where temperature is not externally altered. The transformation from martensite to austenite absorbs energy in the form of heat, while transformation to austenite releases heat energy. Depending upon the rate of transformation and the heat exchange with the surrounding environment, this absorbtion or release of energy can significantly alter the temperature of the material [39]. This addition adds significant complexity to the model as compared to the ferroelectric or ferromagnetic case. In the next section, we will introduce the homogenized energy model for these materials as developed in [25, 26, 38–40]. We will then explore approximation methods for real-time implementation

and investigate inverse formulations similar to those developed in Chapter 3.

## 5.1 Homogenized Energy Model – Theoretical Development

We consider first the case of SMA that exhibit nearly uniform crystal structure throughout the bulk material. In this case, the macroscopic model equals the mesoscopic model or kernel, i.e.,

$$\varepsilon(\sigma, T; x_+; x_-) = \overline{\varepsilon}(\sigma, T; \sigma_c, x_+, x_-) + \lambda_A(1 - x_+ - x_-) + \lambda_M(x_+ + x_-) \tag{5.1}$$

for some constant value of $\sigma_c$. Such high-grade materials are difficult and expensive to obtain, however.

Commercial grade SMAs exhibit multiple layers, each possessing different parameters. This gives a homogenized energy model very similar to those previously developed, namely

$$\begin{aligned}\varepsilon(\sigma, T; x_+, x_-) = \int_0^\infty \int_{-\infty}^\infty \big(\overline{\varepsilon}(\sigma + \sigma_I, T; x_+, x_-) + \lambda_A(1 - x_+ - x_-) \\ + \lambda_M(x_+ + x_-)\big)\nu_c(\sigma_c)\nu_I(\sigma_I)d\sigma_I d\sigma_c.\end{aligned} \tag{5.2}$$

In either case, $\varepsilon$ is the strain, $\sigma$ is the applied stress, $T$ is the material temperature, $x_+$ is the fraction of lattice elements in the martensite plus orientation, $x_-$ is the fraction of lattice elements in the martensite minus orientation, $\sigma_I$ gives the local interaction stress, and $\lambda_A$, $\lambda_M$ are the coefficients of thermal expansion for austenite and martensite, respectively. The parameter $\sigma_c$ describes the difference in the coercive stress $\sigma_A$ where austenite transforms to martensite and the coercive stress $\sigma_M$ where martensite transforms back to austenite; i.e.,

$$\sigma_c = \sigma_A - \sigma_M. \tag{5.3}$$

The parameters $\sigma_c$ and $\sigma_I$ are allowed to vary across lattice elements and are quantified by the densities $\nu_c$ and $\nu_I$, in an analogous manner to $E_c$ and $E_I$ for ferroelectric materials. As before, $\nu_c$ and $\nu_I$ are constrained by the the physical relations:

1. both $\nu_c$ and $\nu_I$ are bounded by decaying exponentials,
2. $\nu_c$ is strictly positive,
3. $\nu_I$ is symmetric about 0, and
4. $\int_0^\infty \nu_c(\sigma_c)d\sigma_c = 1, \qquad \int_{-\infty}^\infty \nu_I(\sigma_I)d\sigma_I = 1.$

(5.4)

The integrals in (5.2) are solved numerically via quadrature, yielding the numerical model

$$\varepsilon(\sigma, T; x_+, x_-) = \sum_{i=1}^{N_c}\sum_{j=1}^{N_I} \overline{\varepsilon}(\sigma + \sigma_I[j], T[i,j]; x_+[i,j], x_-[i,j])\nu_c(\sigma_c[i])\nu_I(\sigma_I[j]). \tag{5.5}$$

The same caveats regarding the continuity and smoothness of the ferroelectric or ferromagnetic

model with quadrature given in Section 2.4 apply to (5.5) as well. This is especially important for the inverse formulation given in Section 5.5.

### 5.1.1 Kernel Development

The kernel $\overline{\varepsilon}$ is given by balancing the Gibb's free energy with thermal energies. For SMA materials, the Helmholtz energy is specified by

$$\psi(\varepsilon,T)=\begin{cases}\dfrac{Y_M}{2}\left(\varepsilon+\varepsilon_T\right)^2 & \varepsilon\le-\varepsilon_M\\ -\dfrac{Y_0(T)}{2}\left(\varepsilon+\varepsilon_0(T)\right)^2+\psi_0(T) & -\varepsilon_M(T)<\varepsilon<-\varepsilon_A(T)\\ \dfrac{Y_A}{2}\varepsilon^2+\Delta\beta(T) & |\varepsilon|\le\varepsilon_A(T)\\ -\dfrac{Y_0(T)}{2}\left(\varepsilon-\varepsilon_0(T)\right)^2+\psi_0(T) & \varepsilon_A(T)<\varepsilon<\varepsilon_M(T)\\ \dfrac{Y_M}{2}\left(\varepsilon-\varepsilon_T\right)^2 & \varepsilon\ge\varepsilon_M\end{cases} \tag{5.6}$$

where $Y_A$ and $Y_M$ are the Young's moduli for the austenite and martensite phases, respectively, $\varepsilon_T$ is the remanent strain for positive martensite variants, $\varepsilon_A$ and $\varepsilon_M$ are the inflection points where the concavity of the energy changes (the points where switching occurs), and

$$Y_0(T)=\frac{Y_A\varepsilon_A(T)-Y_M(\varepsilon_M(T)-\varepsilon_T)}{\varepsilon_M(T)-\varepsilon_A(T)},$$

$$\varepsilon_0(T)=\varepsilon_A(T)\frac{Y_A\varepsilon_M(T)-Y_M(\varepsilon_M(T)-\varepsilon_T)}{Y_A\varepsilon_A(T)-Y_M(\varepsilon_M(T)-\varepsilon_T)},$$

$$\psi_0(T)=-\frac{1}{2}\frac{Y_M(\varepsilon_M(T)-\varepsilon_T)\left(Y_A\varepsilon_A(T)\varepsilon_T+2\Delta\beta(T)\right)}{Y_A\varepsilon_A(T)-Y_M(\varepsilon_M(T)-\varepsilon_T)}.$$

Without loss of generality, the martensite minima of $\psi$ have been shifted to zero so that the austenite minimum has height

$$\Delta\beta(T)=\beta_A(T)-\beta_M(T) \tag{5.7}$$

where $\beta_A$ and $\beta_M$ represent the chemical free energies of austenite and martensite, respectively. Enforcing continuity for $\psi$ gives an explicit relation for $\Delta\beta$, namely

$$\Delta\beta(T)=-\frac{1}{2}\left(Y_A\varepsilon_A(T)\varepsilon_M(T)+Y_M(\varepsilon_M(T)-\varepsilon_T)(\varepsilon_T-\varepsilon_A(T))\right). \tag{5.8}$$

The Helmholtz energy is depicted in Figure 5.1. Note that for sufficiently large temperatures the martensite minima are eliminated and the only stable state is austenite. Further note that for sufficiently low temperatures there is no minimum at $\varepsilon=0$. In this case austenite is unstable and switching occurs directly between the two martensite variants. The Helmholtz energy for

this case can be simplified to

$$\psi(\varepsilon,T)=\begin{cases}\dfrac{Y_M}{2}\left(\varepsilon+\varepsilon_T\right)^2 & \varepsilon\le-\varepsilon_M\\ \Delta\beta(T)\left(1-\dfrac{1}{\varepsilon_M(T)\varepsilon_T}\varepsilon^2\right) & |\varepsilon|\le\varepsilon_A(T)\\ \dfrac{Y_M}{2}\left(\varepsilon-\varepsilon_T\right)^2 & \varepsilon\le-\varepsilon_M\end{cases} \tag{5.9}$$

and the relation of inflection points to chemical free energy is

$$\Delta\beta(T)=\frac{1}{2}Y_M\varepsilon_T(\varepsilon_T-\varepsilon_M(T)). \tag{5.10}$$

This case is also depicted in Figure 5.1. In either case, the Gibb's energy

$$G=\psi-\sigma\varepsilon \tag{5.11}$$

balances the internal Helmholtz energy with the work performed. As detailed in [40], where the Legendre transform properties of the Gibb's energy are discussed, $G$ is a function of the independent variables $\sigma$ and $T$ and the dependent variable $\varepsilon$. More details on the development of this energy may be found in [25, 26, 38–40].

To utilize the Gibb's energy, it is necessary to determine the inflection points from the given model parameters. Just as $P_I$ was related to $E_c$, $\varepsilon_A$ and $\varepsilon_M$ can be related to the coercive stresses by

$$\sigma_A(T)=Y_A\varepsilon_A(T),\quad \sigma_M(T)=Y_M(\varepsilon_M(T)-\varepsilon_T). \tag{5.12}$$

However, the model parameters give only the difference $\sigma_c(T)=\sigma_A(T)-\sigma_M(T)$. It is thus necessary to determine one of the coercive stresses by other means. This is achieved by employing (5.8) or (5.10) with a direct calculation of the chemical free energy. Given the internal energies

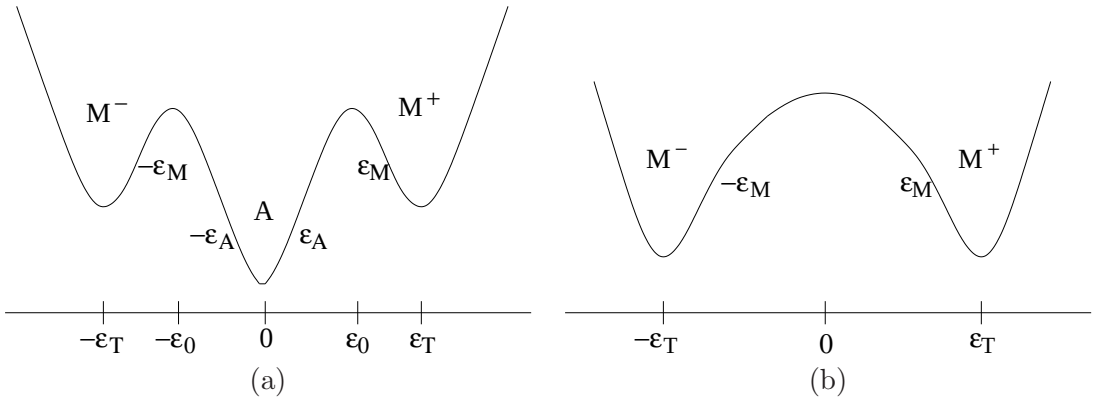



Figure 5.1: Helmholtz energies in shape memory alloys for (a) temperatures where the austenite phase is stable and (b) temperatures where only martensite is stable.

$u_A$ and $u_M$ of austenite and martensite variants at the reference temperature $T_R$, the chemical free energies can be formulated as

$$\begin{aligned}\beta_A(T) &= c_A(T - T_R) + u_A - Tc_A \ln\left(\frac{T}{T_R}\right) + \eta_A, \\ \beta_M(T) &= c_M(T - T_R) + u_M - Tc_M \ln\left(\frac{T}{T_R}\right) + \eta_M\end{aligned} \tag{5.13}$$

where $c_A$, $c_M$ are the specific heat capacities of austenite and martensite, respectively, and $\eta_A$, $\eta_M$ are the austenite and martensite entropies at $T_R$ [25, 26, 38–40]. The difference in chemical free energies is thus given as

$$\Delta\beta(T) = \Delta u - \Delta\eta T + \Delta c\left(T - T_R - T\ln\left(\frac{T}{T_R}\right)\right) \tag{5.14}$$

where $\Delta u = u_A - u_M$, $\Delta\eta = \eta_A - \eta_M$ and $\Delta c = c_A - c_M$. Equating (5.8) and (5.14) and applying (5.12) yields the relation

$$\sigma_A(T) = \frac{\sigma_c}{2} - \frac{Y_M}{1 - Y_R}\varepsilon_T + \sqrt{\frac{\sigma_c^2}{4} + \frac{Y_M}{1 - Y_R}\left(\frac{Y_M}{1 - Y_R}\varepsilon_T^2 - 2\Delta\beta(T)\right)} \tag{5.15}$$

where $Y_R = Y_M/Y_A$. The radicand is non-negative as long as $\Delta\beta(T) \leq Y_M\varepsilon_T^2/2$. This will be enforced by the model. The coercive stress for switching from a martensite variant is

$$\sigma_M(T) = \sigma_A(T) - \sigma_c. \tag{5.16}$$

Given the coercive stresses, the necessary condition $\frac{\partial G}{\partial \varepsilon} = 0$ gives the negligible relaxation model. With relaxation, the Gibb's energy (5.11) is balanced with thermal energy through the Boltzmann relation. This yields the expected strains

$$\begin{aligned}\langle\varepsilon_+\rangle(\sigma_e, T) &= \frac{\int_{\varepsilon_M}^{\infty} \varepsilon \exp\left(\frac{-G(\sigma_e,T;\varepsilon)V}{kT}\right) d\varepsilon}{\int_{\varepsilon_M}^{\infty} \exp\left(\frac{-G(\sigma_e,T;\varepsilon)V}{kT}\right) d\varepsilon}, \quad \langle\varepsilon_-\rangle(\sigma_e, T) = \frac{\int_{-\infty}^{-\varepsilon_M} \varepsilon \exp\left(\frac{-G(\sigma_e,T;\varepsilon)V}{kT}\right) d\varepsilon}{\int_{-\infty}^{-\varepsilon_M} \exp\left(\frac{-G(\sigma_e,T;\varepsilon)V}{kT}\right) d\varepsilon}, \\ \langle\varepsilon_A\rangle(\sigma_e, T) &= \frac{\int_{-\varepsilon_A}^{\varepsilon_A} \varepsilon \exp\left(\frac{-G(\sigma_e,T;\varepsilon)V}{kT}\right) d\varepsilon}{\int_{-\varepsilon_A}^{\varepsilon_A} \exp\left(\frac{-G(\sigma_e,T;\varepsilon)V}{kT}\right) d\varepsilon}\end{aligned} \tag{5.17}$$

of the martensite plus, martensite minus, and austenite phases, respectively. Here $k$ is Boltzmann's constant, $V$ is the volume of the homogeneous layer being modeled, and $\sigma_e = \sigma + \sigma_I$ is the effective stress. The kernel is

$$\overline{\varepsilon}(\sigma + \sigma_I, T; x_+, x_-) = x_+\langle\varepsilon_+\rangle + x_-\langle\varepsilon_-\rangle + (1 - x_+ - x_-)\langle\varepsilon_A\rangle \tag{5.18}$$

where conservation gives the austenite phase fraction as $1 - x_+ - x_-$. For temperatures where the austenite well is stable, we assume that martensite will always switch to austenite and do not consider switching directly from one martensite variant to the other. As shown in [25,26,38–40], with this assumption the phase fractions evolve according to the differential equations

$$\begin{aligned} \dot{x}_-(t) &= -(p_- + p_{A-})x_-(t) - p_{A-}x_+(t) + p_{A-} \\ \dot{x}_+(t) &= -(p_+ + p_{A+})x_+(t) - p_{A+}x_-(t) + p_{A+} \end{aligned} \tag{5.19}$$

where $p_-$ is the likelihood martensite minus will switch to austenite, $p_+$ is the likelihood martensite plus will switch to austenite, $p_{A-}$ is the likelihood austenite will switch to martensite minus, and $p_{A+}$ is the likelihood of austenite switching to martensite plus. For temperatures at which the austenite well is unstable (where the coercive stress $\sigma_A$ given in (5.15) is not positive) all austenite will switch to one of the martensite variants. Once this occurs, the phase fraction of austenite is 0, the phase fraction of martensite minus is $x_- = 1 - x_+$ by conservation, and $x_+$ evolves according to the differential equation

$$\dot{x}_+(t) = -(p_+ + p_-)x_+ + p_- \tag{5.20}$$

where $p_+$ is the likelihood of martensite plus switching to martensite minus and $p_-$ is the likelihood of martensite minus becoming martensite plus. In either case the switching likelihoods are given by

$$\begin{aligned} p_+(\sigma_e, T) &= \frac{\exp\left(\frac{-G(\sigma_e,\varepsilon_M,T)V}{kT}\right)}{\tau \int_{\varepsilon_M}^{\infty} \exp\left(\frac{-G(\sigma_e,\varepsilon,T)V}{kT}\right) d\varepsilon}, \quad p_-(\sigma_e, T) = \frac{\exp\left(\frac{-G(\sigma_e,-\varepsilon_M,T)V}{kT}\right)}{\tau \int_{-\infty}^{-\varepsilon_M} \exp\left(\frac{-G(\sigma_e,\varepsilon,T)V}{kT}\right) d\varepsilon}, \\ p_{A+}(\sigma_e, T) &= \frac{\exp\left(\frac{-G(\sigma_e,\varepsilon_A,T)V}{kT}\right)}{\tau \int_{-\varepsilon_A}^{\varepsilon_A} \exp\left(\frac{-G(\sigma_e,\varepsilon,T)V}{kT}\right) d\varepsilon}, \quad p_{A-}(\sigma_e, T) = \frac{\exp\left(\frac{-G(\sigma_e,-\varepsilon_A,T)V}{kT}\right)}{\tau \int_{-\varepsilon_A}^{\varepsilon_A} \exp\left(\frac{-G(\sigma_e,\varepsilon,T)V}{kT}\right) d\varepsilon} \end{aligned} \tag{5.21}$$

where $\tau$ gives the material-dependent relaxation time quantifying the rate at which dipoles attempt to switch orientations. Instead, the formulation

$$\begin{aligned} p_+(\sigma_e, T) &= \frac{\int_{\varepsilon_M}^{\varepsilon_M+\delta} \exp\left(\frac{-G(\sigma_e,\varepsilon,T)V}{kT}\right) d\varepsilon}{\int_{\varepsilon_M}^{\infty} \exp\left(\frac{-G(\sigma_e,\varepsilon,T)V}{kT}\right) d\varepsilon}, \quad p_-(\sigma_e, T) = \frac{\int_{-\varepsilon_M-\delta}^{varepsilon_M} \exp\left(\frac{-G(\sigma_e,\varepsilon,T)V}{kT}\right) d\varepsilon}{\int_{-\infty}^{-\varepsilon_M} \exp\left(\frac{-G(\sigma_e,\varepsilon,T)V}{kT}\right) d\varepsilon}, \\ p_{A+}(\sigma_e, T) &= \frac{\int_{\varepsilon_A-\delta}^{\varepsilon_A} \exp\left(\frac{-G(\sigma_e,\varepsilon,T)V}{kT}\right) d\varepsilon}{\int_{-\varepsilon_A}^{\varepsilon_A} \exp\left(\frac{-G(\sigma_e,\varepsilon,T)V}{kT}\right) d\varepsilon}, \quad p_{A-}(\sigma_e, T) = \frac{\int_{-\varepsilon_A}^{-\varepsilon_A+\delta} \exp\left(\frac{-G(\sigma_e,\varepsilon,T)V}{kT}\right) d\varepsilon}{\int_{-\varepsilon_A}^{\varepsilon_A} \exp\left(\frac{-G(\sigma_e,\varepsilon,T)V}{kT}\right) d\varepsilon} \end{aligned} \tag{5.22}$$

where $\delta$ is a small fixed constant and both $V$ and $\tau$ depend on $\delta$ may be employed for statistical

analysis. However, as we did for the ferroelectric and ferromagnetic materials we choose to employ the first formulation for its greater computational simplicity.

Similarities between this model and the ferroelectric/ferromagnetic model should be apparent. At low temperatures, there are two variants of martensite and other than temperature dependence the model is identical to the ferroelectric/ferromagnetic model in Chapter 2. At higher temperatures where austenite is stable, the model if very similar to the ferroelectric/ferromagnetic model with 90° switching. However, the Helmholtz energy in the SMA case does not include any fourth order terms of the order parameter ($P$ for ferroelectrics and $\varepsilon$ for SMAs). As will be shown in Section 5.3 this allows the likelihoods and average polarizations to be written as error functions, which greatly improves computationally efficiency and minimizes error. In both temperature ranges, though, the necessity to determine the material temperature and then ascertain the coercive stresses for this temperature introduces significant additional complexity.

### 5.1.2 Thermal Evolution

The internal temperature of the SMA actuator is affected by heat transfer with the surrounding environment, energy absorbed or released by phase transformations, and application of a current to the actuator (Joule heating). As developed in [25, 26, 38–40], we will model these effects via a simplified energy balance.

As shown in [39], heat generation or absorbtion during phase transformations can be significant and is a source of material self-heating or self-cooling. The specific enthalpies of austenite and martensite phases are

$$\begin{aligned} h_A &= \frac{-\sigma_e^2}{2Y_A} + \Delta\beta(T(t)) + T(t)\left(c_A \ln\left(\frac{T(t)}{T_R}\right) + \eta_A\right) \\ h_{M+} &= \frac{-\sigma_e^2}{2Y_M} - \sigma_e\varepsilon_T + T(t)\left(c_M \ln\left(\frac{T(t)}{T_R}\right) + \eta_M\right) \\ h_{M-} &= \frac{-\sigma_e^2}{2Y_M} + \sigma_e\varepsilon_T + T(t)\left(c_M \ln\left(\frac{T(t)}{T_R}\right) + \eta_M\right) \end{aligned} \tag{5.23}$$

where $c_A$, $c_M$ are the density dependent specific heats of austenite and martensite and $\eta_A$, $\eta_M$ are the constant entropies measured at temperature $T_R$. The effect of transformation enthalpy on the material temperature can be modeled via the differential equation

$$(c_A x_A + c_M(x_+ + x_-))\dot{T}(t) = -h_A\dot{x}_A - h_{M+}\dot{x}_+ - h_{M-}\dot{x}_- \tag{5.24}$$

where $x_A$ is the phase fraction of austenite. The conservation equations $x_A = 1 - x_+ - x_-$ and

$\dot{x}_A = -\dot{x}_+ - \dot{x}_-$ simplify this to

$$\begin{aligned}(c_A - \Delta c(x_+ + x_-))\dot{T}(t) &= -(h_A - h_{M-})\dot{x}_- - (h_A - h_+)\dot{x}_+ \\ &= -\left(\frac{\sigma_e^2}{2}\left(\frac{1}{Y_M} - \frac{1}{Y_A}\right) + \Delta c(T(t) - T_R) + \Delta u\right)(\dot{x}_+ + \dot{x}_-) \\ &\quad + \sigma_e \varepsilon_T (\dot{x}_- - \dot{x}_+)\,.\end{aligned} \tag{5.25}$$

Whereas this can always be applied, we note that for regimes where austenite is unstable the simpler relation

$$(c_A - \Delta c(x_+ + x_-))\dot{T}(t) = -2\sigma_e \varepsilon_T \dot{x}_+ \tag{5.26}$$

may be employed instead.

To incorporate heat transfer with the environment, we assume that the internal temperature can be effectively approximated by modeling heat transfer at the surface of the actuator. Given this assumption, the effect of heat transfer is incorporated by

$$(c_A - \Delta c(x_+ + x_-))\dot{T}(t) = -h_c \Omega (T(t) - T_E(t)) - \epsilon \sigma_e^B \Omega (T^4(t) - T_E^4(t)) \tag{5.27}$$

where $h_c$ is the average convection heat transfer coefficient, $\Omega$ is the ratio of surface area to volume, $T_E(t)$ is the external or ambient temperature in Kelvin, $\epsilon$ is the surface emissivity, and $\sigma_e^B$ is the Stefan-Boltzmann constant. The first term on the right hand side of (5.27) approximates convective heat exchange, whereas the second term models radiative heat exchange. The convection coefficient $h_c$ is weakly temperature-dependent in general, but in this discussion we approximate it by a constant. Similarly, $\Omega$ is weakly strain dependent in general but is treated as a constant. These approximations are justified in [25, 26].

Finally, one principle method of controlling an SMA actuator is through application of a current, which changes the internal temperature of the actuator through resistive Joule heating. The effect of Joule heating on temperature is given by

$$(c_A - \Delta c(x_+ + x_-))\dot{T}(t) = \frac{I^2}{\zeta^2}\left((\rho_A^e + \lambda_A^e T(t))(1 - x_+ - x_-) + (\rho_M^e + \lambda_M^e T(t))(x_+ + x_-)\right) \tag{5.28}$$

where $I$ is the current, $\zeta$ is the cross-sectional area of the current path, and $\rho_A^e$, $\rho_M^e$, $\lambda_A^e$ and $\lambda_M^e$ are the constant and temperature dependent coefficients of resistivity for austenite and martensite. The geometric parameter $\zeta$ is also approximated as a constant, as justified in [25,26].

The overall temperature change in the actuator is given by the superposition of enthalpy changes, heat-transfer with the surrounding environment, and Joule heating. Combining (5.25),

(5.27) and (5.28) gives the thermal evolution equation

$$\begin{aligned}\dot{T}(t) =& \frac{1}{(c_A - \Delta c(x_+ + x_-))}\Bigg( - \left( \frac{\sigma_e^2}{2}\left( \frac{1}{Y_M} - \frac{1}{Y_A} \right) + \Delta c(T(t) - T_R) + \Delta u \right)(\dot{x}_+ + \dot{x}_-) \\ & + \sigma_e \varepsilon_T (\dot{x}_- - \dot{x}_+) - h_c \Omega (T(t) - T_E(t)) - \epsilon \sigma_e^B \Omega (T^4(t) - T_E^4(t)) \\ & + \frac{I^2}{\zeta^2}\big( (\rho_A^e + \lambda_A^e T(t))(1 - x_+ - x_-) + (\rho_M^e + \lambda_M^e T(t))(x_+ + x_-) \big) \Bigg).\end{aligned} \tag{5.29}$$

Coupled with (5.19), this gives a system of three first-order differential equations that must be solved at each timestep and each quadrature point. A large, rapid temperature change occurs when lattice elements switch phase, resulting in a numerically stiff system. This must be taken into account when solving the differential equations.

## 5.2 Comparison of Model Response to Data

The homogenized energy model for SMA materials has been extensively validated against experimental data. Since our focus is computational, we offer here only a couple comparisons to illustrate its effectiveness, and refer the reader to [25, 26, 38–40] for more details.

A comparison of the the model to the stress-strain response of NiTi [28] may be observed in Figure 5.2(b). The external temperature was held at a constant 293 K for this test, and the applied stress was varied as given in Figure 5.2(a). The temperature of the SMA actuator was not measured in the data set, but the spatially averaged temperature as a function of time predicted by the model is shown in Figure 5.2(c). Note that for this rate of loading, the model predicts a change of slightly over 5 degrees in the average internal temperature of the SMA as a minor loop is traversed. A comparison to temperature versus strain data [17] may be found in Figure 5.3(b). In this case the NiTi sample was placed in a thermal chamber under a constant load of 200 MPa and the ambient temperature was varied as given in Figure 5.3(a). Note that due to the high temperature employed in this experiment and the fact that the sample had just been annealed, the SMA exhibited a plastic deformation through the cycle. Such deformations are not included in the model, as most actuators are operated in regimes where plastic deformations can be avoided or minimized. However, the model does accurately characterize the phase transformation present in the data due to the temperature change. As shown in the figure, for this slow temperature gradient, the actuator temperature follows the ambient temperature very closely, although a small deviation can be observed when the phase transformations occur. A faster temperature gradient would lead to a more pronounced difference.

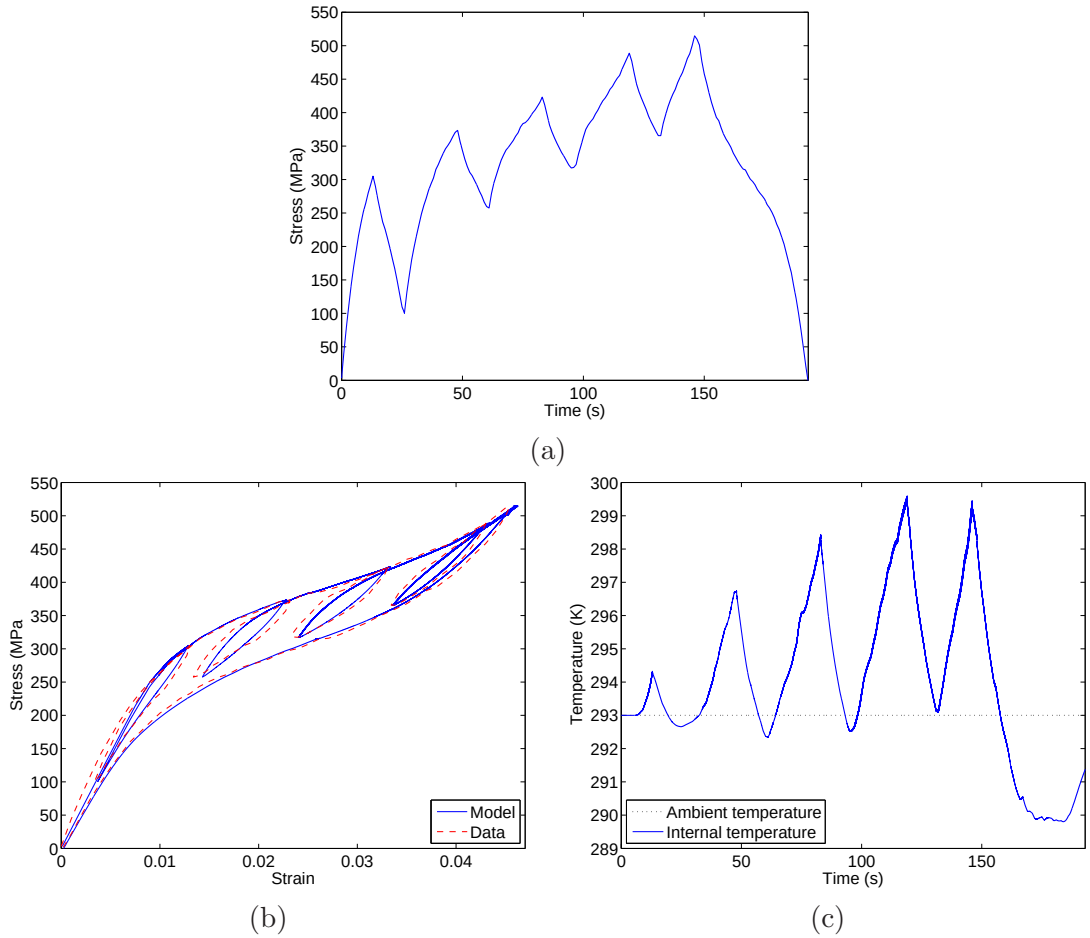
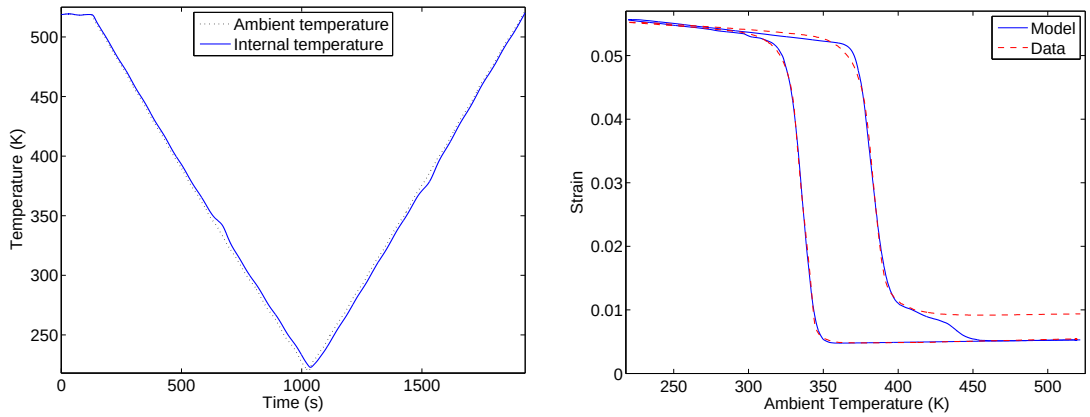



Figure 5.2: (b) Comparison of SMA model response to stress-strain data [28] for the applied stress given in (a) and a constant ambient temperature of 293 K. (c) Internal temperature (spatial average) predicted by the model.

## 5.3 Homogenized Energy Model Implementation

Many of the same techniques utilized to simplify the computation of the homogenized energy model for ferroelectric and ferromagnetic materials in Sections 2.3 and 2.5 can also be applied to the model for shape memory alloys. First, the switching likelihoods and average strains in the relaxation algorithm can be transformed into exponential and error functions. Letting

$$\begin{aligned} z_+ &= \sqrt{\frac{V}{2Y_M kT}}(\sigma_M - \sigma_e), \quad z_{A+} = \sqrt{\frac{V}{2Y_A kT}}(\sigma_A - \sigma_e), \\ z_- &= \sqrt{\frac{V}{2Y_M kT}}(\sigma_M + \sigma_e), \quad z_{A-} = \sqrt{\frac{V}{2Y_A kT}}(-\sigma_A - \sigma_e), \end{aligned} \tag{5.30}$$




(a) (b)

Figure 5.3: Comparison of SMA model response to temperature-strain data [17] for the dynamic external temperature given in (a) and a constant applied stress of 200 MPa. The internal temperature (spatially averaged) of the material predicted by the model is also given in (a), while both the measured and predicted strain are given in (b).

the switching likelihoods are given by

$$\begin{aligned} p_+(\sigma_e,T) &= \frac{1}{\tau}\sqrt{\frac{2Y_MV}{\pi kT}}\frac{1}{\operatorname{erfcx}(z_+)}, \quad p_{A+}(\sigma_e,T) = \frac{1}{\tau}\sqrt{\frac{2Y_AV}{\pi kT}}\frac{\exp(-z_{A+}^2)}{\operatorname{erfc}(z_{A-})-\operatorname{erfc}(z_{A+})}, \\ p_-(\sigma_e,T) &= \frac{1}{\tau}\sqrt{\frac{2Y_MV}{\pi kT}}\frac{1}{\operatorname{erfcx}(z_-)}, \quad p_{A-}(\sigma_e,T) = \frac{1}{\tau}\sqrt{\frac{2Y_AV}{\pi kT}}\frac{\exp(-z_{A-}^2)}{\operatorname{erfc}(z_{A-})-\operatorname{erfc}(z_{A+})} \end{aligned} \tag{5.31}$$

and the average strains are

$$\begin{aligned} \langle\varepsilon_+\rangle(\sigma_e,T) &= \sqrt{\frac{2kT}{\pi Y_MV}}\frac{1}{\operatorname{erfcx}(z_+)}+\frac{\sigma_e}{Y_M}+\varepsilon_T, \\ \langle\varepsilon_-\rangle(\sigma_e,T) &= -\sqrt{\frac{2kT}{\pi Y_MV}}\frac{1}{\operatorname{erfcx}(z_-)}+\frac{\sigma_e}{Y_M}-\varepsilon_T, \\ \langle\varepsilon_A\rangle(\sigma,T) &= \sqrt{\frac{2kT}{\pi Y_AV}}\frac{\exp(-z_{A-}^2)-\exp(-z_{A+}^2)}{\operatorname{erfc}(z_{A-})-\operatorname{erfc}(z_{A+})}+\frac{\sigma_e}{Y_A}. \end{aligned} \tag{5.32}$$

However, the average strain for austenite may suffer significant numerical error due to the subtraction of nearly equal numbers. For this reason, we instead approximate the average strains by

$$\langle\varepsilon_+\rangle(\sigma_e,T)\approx\frac{\sigma_e}{Y_M}+\varepsilon_T, \quad \langle\varepsilon_-\rangle(\sigma_e,T)\approx\frac{\sigma_e}{Y_M}-\varepsilon_T, \quad \langle\varepsilon_A\rangle(\sigma_e,T)\approx\frac{\sigma_e}{Y_A}. \tag{5.33}$$

These are the average strains given when relaxation is negligible and are good approximates for the relaxation case as well. Relaxation still enters the formulation through the likelihoods.

The likelihoods can also result in division by zero due to numerical approximation and large error from the subtraction of nearly equal numbers. The same methods utilized for the ferroelectric/ferromagnetic model (linear approximation for large $z_-$, $z_+$ and replacing $1/\text{erfcx}(z)$ with 0 for small $z = z_-, z_+$) may be utilized for $p_+$ and $p_-$. For the likelihoods of switching out of austenite, first note that

$$z_{A+} - z_{A-} = \sqrt{\frac{2V}{Y_A kT}}\sigma_A > 0 \tag{5.34}$$

as long as austenite is stable. Since erfc is monotone decreasing, in theory division by zero never occurs. However, $\sigma_A \longrightarrow 0$ as the internal temperature approaches the critical temperature where austenite becomes unstable. In finite precision arithmetic, therefore, it is possible that $\sigma_A > 0$ but $z_{A-} = z_{A+}$. To mitigate this issue, we specify that austenite is stable only for $\sigma_A \geq \overline{\delta}$, where $\overline{\delta}$ is a small positive constant. However, significant numerical error may still occur due to the asymptotic nature of erfc. Since $\text{erfc}(z) \approx 0$ for large $z$ while $\text{erfc}(z) \approx 2$ for small $z$, floating point resolution is maximized if positive operands are given to erfc. Since

$$\text{erfc}(z_{A-}) - \text{erfc}(z_{A+}) = 2 - \text{erfc}(-zA-) - (2 - \text{erfc}(-z_{A+})) = \text{erfc}(-z_{A+}) - \text{erfc}(-z_{A-}) \tag{5.35}$$

this can always be achieved. While helpful, this does not mitigate the problem caused by the asymptotic decay of erfc. However, $\text{erfcx}(z)$ rises approximately linearly as $z$ increases and thus if the operands $z_{A-}$ and $z_{A+}$ are positive and distinct, $\text{erfcx}(z_{A-}) - \text{erfcx}(z_{A+}) \neq 0$. We therefore formulate the likelihoods of switching out of austenite as

$$p_{A+}(\sigma_e, T) = \begin{cases} \frac{1}{\tau}\sqrt{\frac{2Y_A V}{\pi kT}}\frac{1}{\exp(z_{A+}^2 - z_{A-}^2)\text{erfcx}(z_{A-}) - \text{erfcx}(z_{A+})} & \text{if } z_{A-} > 0 \\ \frac{1}{\tau}\sqrt{\frac{2Y_A V}{\pi kT}}\frac{1}{\text{erfcx}(-z_{A+}) - \exp(z_{A+}^2 - z_{A-}^2)\text{erfcx}(-z_{A-})} & \text{if } z_{A+} < 0 \\ \frac{1}{\tau}\sqrt{\frac{2Y_A V}{\pi kT}}\frac{\exp(-z_{A+}^2)}{\text{erfc}(z_{A-}) - \text{erfc}(z_{A+})} & \text{otherwise,} \end{cases} \tag{5.36}$$

$$p_{A-}(\sigma_e, T) = \begin{cases} \frac{1}{\tau}\sqrt{\frac{2Y_A V}{\pi kT}}\frac{1}{\text{erfcx}}(z_{A-}) - \exp(z_{A-}^2 - z_{A+}^2)\text{erfcx}(z_{A+}) & \text{if } z_{A-} > 0 \\ \frac{1}{\tau}\sqrt{\frac{2Y_A V}{\pi kT}}\frac{1}{\exp(z_{A-}^2 - z_{A+}^2)\text{erfcx}(-z_{A+}) - \text{erfcx}(-z_{A-})} & \text{if } z_{A+} < 0 \\ \frac{1}{\tau}\sqrt{\frac{2Y_A V}{\pi kT}}\frac{\exp(-z_{A-}^2)}{\text{erfc}(z_{A-}) - \text{erfc}(z_{A+})} & \text{otherwise.} \end{cases} \tag{5.37}$$

Note that if both $z_{A-}$ and $z_{A+}$ are negative, the additive inverse of the operands is used with erfcx, thereby avoiding error due to the asymptotic nature of erfc, whereas if both are positive erfcx is utilized on the operands directly. Coupled with $\overline{\delta}$, this ensures the likelihoods for

austenite may be computed accurately.

Whereas the switching likelihoods are always positive, in theory they are not bounded above. For the ferroelectric/ferromagnetic model there was a single differential equation and this was not an issue. However, a system of equations must now be solved. While the full system including thermal evolution is nonlinear, if we assume a constant internal temperature the system (5.19) is linear for a given stress $\sigma$ and can be written in matrix form as

$$\begin{bmatrix} \dot{x}_- \\ \dot{x}_+ \end{bmatrix} = \begin{bmatrix} -(p_- + p_{A-}) & -p_{A-} \\ -p_{A+} & -(p_+ + p_{A+}) \end{bmatrix} \begin{bmatrix} x_- \\ x_+ \end{bmatrix} + \begin{bmatrix} p_{A-} \\ p_{A+} \end{bmatrix}. \tag{5.38}$$

Because the switching likelihoods may grow unboundedly, the resulting coefficient matrix can be ill-conditioned. Inclusion of thermal evolution does not alleviate this issue; it only obfuscates the nature of the problem. To mitigate the ill-conditioning, we choose to set an upper bound for the likelihoods. The sole use of the switching likelihoods is to determine the evolution of the phase fractions, and as such a likelihood of 1000 has almost the same effect as a likelihood of 10,000 or 1,000,000. For a suitable upper bound, therefore, no significant accuracy is lost in the saturation. Similarly, likelihoods very close to zero can cause ill-conditioning. Instead, we force very small likelihoods to zero and explicitly deal with the possibility of a zero likelihood when solving the phase evolution equations.

The system of differential equations for the phase and temperature evolution is stiff, requiring a stiff or unconditionally stable method to solve it numerically. One approach is to use a stiff, variable-step method such as those supplied by the MATLAB routines `ode15s` and `ode23s`. While this may prove the method of choice for offline analysis/calculation where computation time is not a factor, these methods have limited or no applicability in sampled or real-time environments. They require that the input stress, ambient temperature, and applied current be know for all time and not just at sampled points. Even when these are known, the amount of computation time required is highly variable, and may not meet real-time requirements. For control applications, therefore, we instead choose to approximate the differential equations so that we can apply an unconditionally stable fixed-step method. This ensures that the requirements of a real-time environment may be achieved. Since the equations are approximate, however, the step size $\Delta t$ will need to be relatively small compared to the actuation rate to ensure accuracy.

An approximate for the unknown temperature at time $t + \Delta t$ is

$$T(t + \Delta t) \approx T(t). \tag{5.39}$$

While not valid for large $\Delta t$, the error in the approximate goes to zero as $\Delta t$ limits to zero; thus, it is valid for sufficiently small $\Delta t$. If this approximate temperature is used to determine the coercive stresses and switching likelihoods, (5.19) decouples from the thermal evolution equation (5.29). If we also approximate the input stress as a step function so that it is constant between timesteps, then a single timestep of the phase evolution equations yields a linear, autonomous

system of two equations. This may be solved with backward Euler, and may also be solved analytically. For temperatures where austenite is stable, the backward Euler solution is given by

$$\begin{aligned} x_-(t+\Delta t) &= \frac{1}{det}\left(x_-(t)(1+\Delta t p_+ + \Delta t p_{A+}) - \Delta t x_+(t) p_{A-} + (\Delta t)^2 p_{A-}(1+p_+)\right) \\ x_+(t+\Delta t) &= \frac{1}{det}\left(x_+(t)(1+\Delta t p_- + \Delta t p_{A-}) - \Delta t x_-(t) p_{A+} + (\Delta t)^2 p_{A+}(1+p_-)\right) \end{aligned} \tag{5.40}$$

where

$$det = 1 + \Delta t(p_- + p_+ + p_{A-} + p_{A+}) + (\Delta t)^2(p_- p_+ + p_- p_{A+} + p_+ p_{A-}),$$

whereas if austenite is not stable

$$\begin{aligned} x_+(t+\Delta t) &= \frac{x_+(t) + \Delta t p_-}{1 + \Delta t(p_+ + p_-)} \\ x_-(t+\Delta t) &= 1 - x_+(t+\Delta t). \end{aligned} \tag{5.41}$$

Note that all the switching likelihoods are nonnegative, while $\Delta t$ is strictly positive. Thus, division by zero cannot occur in this formulation. The pseudoanalytic solution, i.e. the analytic solution to the approximated differential equations, is similar, except that for this method division by zero must be explicitly avoided. If austenite is stable, the update for the phase fractions is given by the pseudocode in Figure 5.4, whereas if austenite is unstable the phase fraction update is

$$\begin{aligned} x_+(t+\Delta t) &= \begin{cases} x_+(t), & \text{if } p_- + p_+ = 0 \\ \dfrac{p_-}{p_- + p_+} + \left(x_+(t) - \dfrac{p_-}{p_- + p_+}\right)\exp(-\Delta t(p_- + p_+)), & \text{otherwise} \end{cases} \\ x_-(t+\Delta t) &= 1 - x_+(t+\Delta t). \end{aligned} \tag{5.42}$$

Once the phase fractions at time $t + \Delta t$ have been determined, their derivatives at time $t + \Delta t$ may be calculated and treated as parameters to the thermal equation (5.29). Although (5.19) gives the analytic value for the derivative at time $t+\Delta t$, use of these in (5.29) may yield a temperature gradient that is much too large if the derivatives have increased substantially from $t$ to $t + \Delta t$. Although theoretically less accurate, the backward differences

$$\dot{x}_+(t+\Delta t) = \frac{x_+(t+\Delta t) - x_+(t)}{\Delta t}, \quad \dot{x}_-(t+\Delta t) = \frac{x_-(t+\Delta t) - x_-(t)}{\Delta t} \tag{5.43}$$

are a better choice since they give an average derivative over $(t, t+\Delta t)$. The term for radiative heat exchange is the only nonlinear term in the resulting thermal equation. We apply the approximation (5.39) to this portion of the thermal evolution equation, which yields the heat

if $(p_{A-} == 0)$

$x_-(t+\Delta t) = x_-(t)\exp(-\Delta t p_-)$;

if $(p_{A+} == 0)$

$x_+(t+\Delta t) = x_+(t)\exp(-\Delta t p_+)$;

else if $(p_- - p_+ - p_{A+} == 0)$

$$x_+(t+\Delta t) = x_+(t)\exp(-\Delta t(p_+ + p_{A+})) - \Delta t x_-(t)\exp(-\Delta t p_-) + (1-\exp(-\Delta t(p_+ + p_{A+})))\frac{p_{A+}}{p_+ + p_{A+}};$$

else

$$x_+(t+\Delta t) = x_+(t)\exp(-\Delta t(p_+ + p_{A+})) + x_-(t)\frac{p_{A+}}{p_- - p_+ - p_{A+}}\big(\exp(-\Delta t p_-) - \exp(-\Delta t(p_+ + p_{A+}))\big) + (1-\exp(-\Delta t(p_+ + p_{A+})))\frac{p_{A+}}{p_+ + p_{A+}};$$

end if

else if $(p_{A+} == 0)$

$x_+(t+\Delta t) = x_+(t)\exp(-\Delta t p_+)$;

if $(p_+ - p_- - p_{A-} == 0)$

$$x_-(t+\Delta t) = x_-(t)\exp(-\Delta t(p_- + p_{A-})) - \Delta t x_+(t)\exp(-\Delta t p_+) + (1-\exp(-\Delta t(p_- + p_{A-})))\frac{p_{A-}}{p_- + p_{A-}};$$

else

$$x_-(t+\Delta t) = x_-(t)\exp(-\Delta t(p_- + p_{A-})) + x_+(t)\frac{p_{A-}}{p_+ - p_- - p_{A-}}\big(\exp(-\Delta t p_+) - \exp(-\Delta t(p_- + p_{A-}))\big) + (1-\exp(-\Delta t(p_- + p_{A-})))\frac{p_{A-}}{p_- + p_{A-}};$$

end if

else

if $((p_- == 0)$ and $(p_+ == 0))$

$c_1 = 1$;

$c_2 = 0$;

else

$$c_1 = \frac{p_+ p_{A-}}{p_- p_+ + p_+ p_{A-} + p_- p_{A+}};$$

$$c_2 = \frac{p_- p_{A=}}{p_- p_+ + p_+ p_{A-} + p_- p_{A+}};$$

end if

$z = \sqrt{(\Delta t(p_- + p_{A-} - p_+ - p_{A+}))^2 + 4p_{A-}p_{A+}}$

$$k_1 = -(\Delta t(p_- + p_{A-} - p_+ - p_{A+}) + z)\frac{c_1 - x_-(t)}{2z} - p_{A-}\frac{x_2 - x_+(t)}{z};$$

$$k_2 = (\Delta t(p_- + p_{A-} - p_+ - p_{A+}) - z)\frac{c_1 - x_-(t)}{2z} + p_{A-}\frac{x_2 - x_+(t)}{z};$$

$$x_-(t+\Delta t) = k_1\exp\left(-\frac{\Delta t}{2}(p_- + p_{A-} + p_+ + p_{A+}) + \frac{z}{2}\right) + k_2\exp\left(-\frac{\Delta t}{2}(p_- + p_{A-} + p_+ + p_{A+}) - \frac{z}{2}\right) + c_1;$$

$$x_+(t+\Delta t) = -\frac{k_1\Delta_t(p_- + p_{A-} - p_+ - p_{A+}) - k_1 z}{2\Delta t p_{A-}}\exp\left(-\frac{\Delta t}{2}(p_- + p_{A-} + p_+ + p_{A+}) + \frac{z}{2}\right) - \frac{k_2\Delta_t(p_- + p_{A-} - p_+ - p_{A+}) + k_2 z}{2\Delta t p_{A-}}\exp\left(-\frac{\Delta t}{2}(p_- + p_{A-} + p_+ + p_{A+}) - \frac{z}{2}\right) + c_2;$$

Figure 5.4: Pseudocode to update the phase fractions for regimes where austenite is stable using the pseudoanalytic method.

transfer portion of the thermal evolution as

$$(c_A - \Delta c(x_+ + x_-))\dot{T}(t) = -\big(h_c + \epsilon\sigma^B(\widetilde{T} - T_E(t))(\widetilde{T} + T_E^2(t))\big)\Omega(T(t) - T_E(t)) \tag{5.44}$$

where $\widetilde{T}$ is the approximation given by (5.39). If we then approximate the input current and external temperature by step functions as we did for the stress, the solution to (5.29) utilizing backward Euler is

$$T(t + \Delta t) = \frac{1}{1 - \Delta t k_1}(T(t) + \Delta t k_2) \tag{5.45}$$

and the analytic solution to the approximated equation is

$$T(t + \Delta t) = \left(T_0 + \frac{k_2}{k_1}\right)\exp(k_1 t) - \frac{k_2}{k_1}. \tag{5.46}$$

In either case, $k_1$ and $k_2$ are given by

$$\begin{aligned} k_1 =& \frac{1}{c_A - \Delta c(x_+ + x_-)}\Bigg(\Delta c(\dot{x}_+ + \dot{x}_-) - (h_c + \epsilon\sigma^B(T(t) - T_E(t))(T(t) + T_E(t))^2)\Omega \\ &+ \frac{I^2}{\zeta^2}(\lambda_A^e(1 - x_+ - x_-) + \lambda_M^e(x_+ + x_-))\Bigg), \end{aligned} \tag{5.47}$$

$$\begin{aligned} k_2 =& \frac{1}{c_A - \Delta c(x_+ + x_-)}\Bigg(\left(-\frac{\sigma_e^2}{2}\left(\frac{1}{Y_M} - \frac{1}{Y_A}\right) - \Delta c T_R + \Delta u\right)(\dot{x}_- + \dot{x}_+) \\ &+ \sigma_e\varepsilon_T(\dot{x}_- - \dot{x}_+) + (h_c + \epsilon\sigma^B(T(t) - T_E(t))(T(t) + T_E(t))^2)\Omega T_E \\ &+ \frac{I^2}{\zeta^2}(\rho_A^e(1 - x_+ - x_-) + \rho_M^e(x_+ + x_-))\Bigg). \end{aligned} \tag{5.48}$$

It should be noted that in forced convection applications, radiative heat transfer is typically one to three orders of magnitude smaller than convective transfer [25, 26]. Thus, in spite of the cubic error term in approximating the radiative transfer with (5.39), the overall error introduced by the approximation is negligible. For regimes where radiative heat transfer is a predominant factor in thermal evolution, (5.39) may still be employed but the requirements on $\Delta t$ to ensure a valid approximation will be more stringent.

The model response for the stress input in Figure 5.5, no applied current, and constant ambient temperature of 293 K with the backward Euler, pseudoanalytic and MATLAB `ode15s` methods can be observed in Figure 5.6. Note that, as expected, the pseudoanlytic solution typically produces a more accurate result. However, the differences between the methods are minor. Run times for the results shown are shown in Table 5.1. One can see that on this particular machine, the backward Euler algorithm was almost 20% faster than the pseudoanalytic method. On different computers backward Euler was found to be as much as 50% faster than the pseudoanalytic method, as shown in Table 5.2. Both methods were substantially faster than `ode15s`.

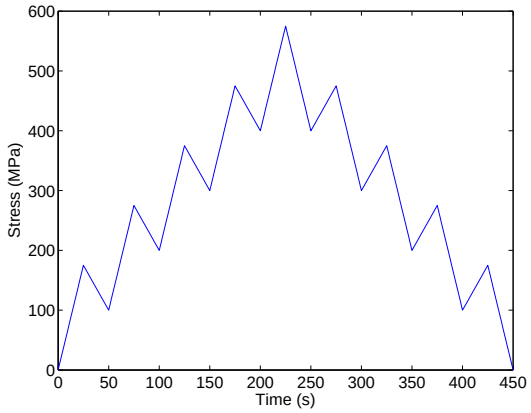



Figure 5.5: Applied stress employed for the results in Figure 5.6 and Tables 5.1 and 5.2.

Table 5.1: Run times in seconds for the backward euler, pseudoanalytic, and `ode15s` approximates depicted in Figure 5.6. All run times were computed on an Intel 64 bit Xeon 3.8 GHz processor with 2 MB of L2 cache running linux. Code was compiled with gcc.

| | Backward euler | pseudoanalytic | `ode15s` |
|---|---|---|---|
| variable-step | | | 77436.85 |
| $\Delta t = 0.1$ | 4.30 | 5.04 | |
| $\Delta t = 0.01$ | 43.24 | 51.13 | |
| $\Delta t = 0.001$ | 431.69 | 513.44 | |

Table 5.2: Timesteps per second for the SMA model employing the backward Euler and analytic approximates on three different computer architectures. Cases are given where look-up table approximates were employed and where full double-precision floating point values were utilized. In each case, the model was run with 40 quadrature points for each distribution (1600 total), and the model code was compiled with gcc.

| | Intel Xeon 1.7 GHz 256 KB L2 Cache | Intel Xeon (64 bit) 3.8 GHz 2 MB L2 Cache | Intel Pentium-M 2.16 GHz 2 MB L2 Cache |
|---|---|---|---|
| Backward Euler (full precision) | 206 | 1040 | 665 |
| Pseudoanalytic (full precision) | 136 | 880 | 495 |
| Backward Euler (look-up table) | 242 | 1063 | 972 |
| Pseudoanalytic (look-up table) | 149 | 899 | 640 |

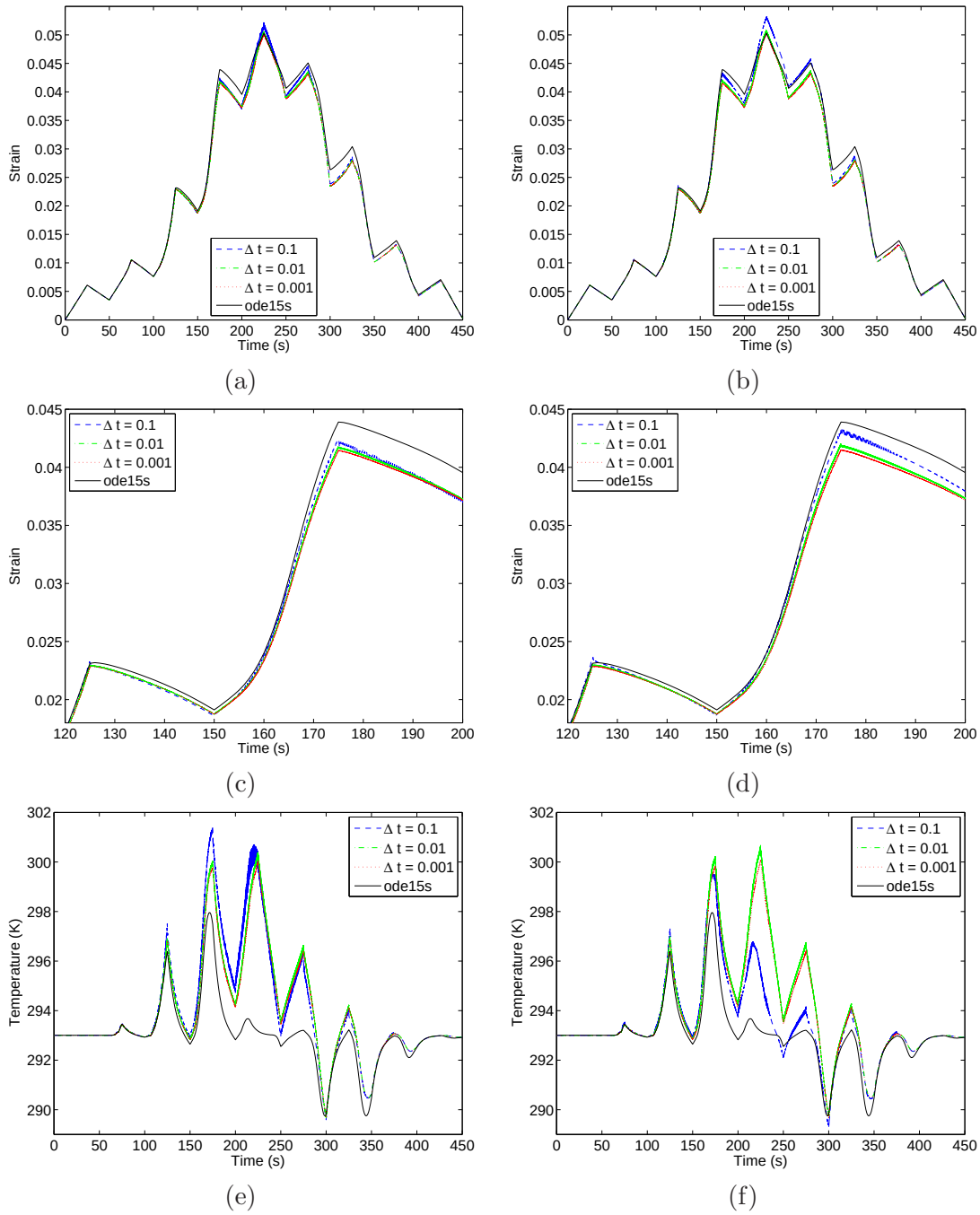



Figure 5.6: Comparison of the model results for the (a), (c), and (e) backward Euler algorithm and (b), (d), and (f) pseudoanalytic method for the stress input given in Figure 5.5. (a) and (b) give the resulting stress (magnified in (c) and (d)) while (e) and (f) give the spatially averaged internal actuator temperature.

Note that whereas the run times of this model are much slower than the ferroelectric or ferromagnetic model, the maximum actuation rate for SMA devices is also much lower. MEMS devices have achieved actuation rates up to approximately 100 Hz, while macroscale SMA actuators are typically operated below a couple of Hertz. Thus, relative to the actuation rate of the device being described, the sma model performs comparable or outperforms the ferroelectric/ ferromagnetic model.

## 5.4 Function Approximations in the SMA model

Given the order of magnitude improvement in speed obtained by approximating the error functions in the ferroelectric/ferromagnetic materials, it is reasonable to ask if applying a similar approach to the SMA model could yield significantly faster results. Table 5.2 shows the results of approximating the error functions and exponentials by faster approximations based on look-up tables. Unlike the ferroelectric/ferromagnetic case, where the majority of the run-time was spent in the computation of error functions, the work required to compute the coercive stresses and update the internal temperature is also a significant portion of the overall computational effort in the SMA model. For this reason, the improvements are more modest in the SMA case, giving at best a 50% improvement in speed for the backward Euler approximations and 30% improvement for the pseudoanalytic method. While not revolutionary, this improvement can be worth the extra memory usage and development cost for some applications, and we briefly outline the approximation process here.

The SMA model contains error functions (both erfcx and erfc), exponentials, square roots, and a logarithm. The square root functions typically do not require approximation. Some processors provide built-in instructions for square root; those that don't typically use Cody and Waite's approximation [9], which is quite efficient. Look-up table and rational Chebyshev approximations for 1/erfcx are given in Section 2.5. Approximation of erfc proceeds similarly. We choose bounds $(z_{low}, z_{high})$ over which to approximate and set $\mathrm{erfc}(z) = 2$ for $z < z_{low}$ and $\mathrm{erfc}(z) = 0$ for $z > z_{high}$. Within these bounds, a look-up table of values may be utilized to approximate the function. As before, accuracy is a function of table size and run-time is (within architecture dependent limits) independent of table size. A few thousand table values should be sufficient for most applications. However, if the required memory is not available, rational Chebyshev approximations may again be utilized. These are given in [8] for a variety of polynomial sizes/precisions. Note, however, that the polynomials in [8] should be converted to continued fraction form for efficiency if the floating point accuracy of the machine allows.

The exponential terms in the likelihoods and average polarizations can yield operands that are large in magnitude; thus, the exponential function cannot be directly approximated with a look-up table as the error functions were. Instead, we start with the approach given by [9]. Let

$z = n\ln(2) + g$ where $|g| \leq \ln(2)/2$. Then

$$\exp(z) = \exp(g)2^n. \tag{5.49}$$

Since $g$ is bounded, we need only approximate exp between $-\ln(2)/2$ and $\ln(2)/2$. The $2^n$ term is essentially free in binary arithmetic. In FORTRAN, the function ADX allows the exponent of a floating point number to be augmented, so if $e_g \approx \exp(g)$ then ADX($e_g$,n) = $e_g 2^n$. In C/C++, no such function exists, but similar results may be obtained through creative use of C's union structure. Figure 5.7 gives an example C langauge macro which gives $2^n$ for an integer $n$ and requires only 32 bit integers. A slightly simpler implementation can be obtained if 64 bit integers are available. Note that the IEEE 754 floating point standard is assumed for the code given in the figure.

If the values of $\ln(2)$ and $1/\ln(2)$ are stored as constants, $n$ and $g$ may be computed with one multiply and one subtraction each as $n = \text{round}(z/\ln(2))$ and $g = z - n\ln(2)$. Cody and Waite [9] develop sophisticated rational polynomial approximations to avoid wobbling precision when calculating $\exp(g)$. However, since we are only concerned with an approximate, we will employ a look-up table or direct rational polynomial approximation (converted to continued fraction form) of $\exp(g)$. Note that with a look-up table computation of the exponential in this manner involves only 4 paired multiplies and adds plus one bit shift. Two if statements should be added to this to check for numerical overflow and underflow.

It should be noted, however, that the approximate exponential developed in the preceding paragraph should only be employed when computing the switching likelihoods. If a backward Euler algorithm is being utilized to solve the differential equations, these are the only exponentials that appear. When employing the pseudoanalytic method, however, several additional

```
typedef union {
  double readback;
  struct {
#ifdef LITTLE_ENDIAN
  int32_t junk, exponent;
#else
  int32_t exponent, junk;
#endif
  } set;
} setexp_t;
static setexp_t settmp;
#define TWO_N(n) (settmp.set.junk = 0, settmp.set.exponent = \
                  (((int)n + 1023) << 20), settmp.readback)
```

Figure 5.7: Example C language macro to efficiently compute $2^n$, where $n$ is an integer.

exponential terms arise. The model result is very sensitive to error in the calculation of these exponentials, and as such the analytic approximations should use a full precision exponential at all times.

Similarly problems arise when attempting to approximate the logarithm in the calculation for $\sigma_A$. It's operand is $\frac{T}{T_A}$, which may easily be bounded by physical properties of the material. However, if the approximation is not monotone increasing, the approximated coercive stresses may not be a monotone function of temperature, which can significantly degrade the performance of the inverse models given in the next section. The results in Table 5.2 were calculated with a full precision logarithm, although a look-up table with interpolation could also have been employed.

Some applications/engineers will see the gains given by these approximations as worthwhile, while others will prefer the ease and smaller memory footprint of the functions supplied by the compiler. For gate array design, where these functions are not provided and floating point operations are likely not available, the given approximations provide a framework for implementing the functions needed by the model. More details specific to fixed-point arithmetic may be found in [9], although where appropriate gate-array designs should also consider CORDIC-based methods.

## 5.5 Inverse Model

To construct the inverse model for shape memory alloys, we first must specify the quantity to be determined, as was the case in Section 4.4. The SMA model has three inputs: stress, ambient temperature, and current. We will assume that only one of these inputs can be controlled, while the others are known but uncontrollable. Specifically, we consider the case when stress is controllable and the two cases when current is controllable. Outside of a thermal chamber, ambient temperature may be known but typically cannot be controlled; thus, the case where ambient temperature is the controlled parameter is not considered. Variations which allow for more than one input to be controlled simultaneously are possible, and can be extrapolated from the methods given here. However, physical actuators in which both current (or voltage) and applied stress can be controlled are rare, and as such we do not focus on these cases.

### 5.5.1 Stress Control

The inverse model described in Chapter 3 may be applied to the SMA case without significant modification as long as stress is the quantity to be determined. In this case, the root-finding problem solved at each timestep is given by

$$\varepsilon(\sigma; T_E, I, x_+, x_-, T) - \widehat{\varepsilon} = 0 \tag{5.50}$$

where $\widehat{\varepsilon}$ is the desired bulk material strain and all quantities except $\sigma$ are known. The function is scalar and monotone in $\sigma$ but, as in the ferroelectric/ferromagnetic case, may possess a finite number of jump discontinuities due to the quadrature approximation (see Section 2.4). Theoretical bounds

$$\min(Y_M(\widehat{\varepsilon} - \varepsilon_T), Y_A\widehat{\varepsilon}) \leq \sigma \leq \max(Y_M(\widehat{\varepsilon} - \varepsilon_T), Y_A\widehat{\varepsilon}) \tag{5.51}$$

may be ascertained by assuming all layers have uniform state. As before, we will solve this by first attempting the secant method. If this fails, we switch to bisection, utilizing the bounds given by the secant method to improve the convergence of bisection. Results are given in Figure 5.8 and Table 5.3 for an example input. Since the SMA model requires that $\Delta t$ be small compared to the actuation rate, the number of function evaluations needed per timestep is typically small. For this example, accuracy within 1 micro-strain is obtained with 2.7 function evaluations per timestep on average and 10 in the worst case scenario, while 1 nano-strain accuracy requires slightly over 4 function evaluations per timestep on average and 21 function evaluations in the worst-case scenario.

### 5.5.2 Current Control

The situation is more complicated when current is the quantity to be determined. Application of a current increases the internal temperature of the actuator through resistive (Joule) heating, which in turn changes the coercive stresses and induces switching. However, current cannot cool the actuator, and once heated the actuator must be allowed to cool at the rate dictated by the convection and radiative heat transfer coefficients. Thus, unlike the stress case it is possible that a requested strain cannot be achieved given the state of the actuator, no matter what current is applied. Additionally, the response for a given timestep may not be a unimodal function of current; the strain increases as martensite minus switches to austenite but decreases as marten-

Table 5.3: Effort to compute the inverse SMA model in terms of average and maximum number of function evaluations for various error tolerances, treating stress as the unknown to be determined and utilizing the input strain in Figure 5.8(a) with $\Delta t = 0.01$.

| Tolerance | Average Function Evaluations | Maximum Function Evaluations |
|---|---|---|
| $1 \times 10^{-3}$ | 1.0135 | 4 |
| $1 \times 10^{-4}$ | 1.1318 | 6 |
| $1 \times 10^{-5}$ | 1.4688 | 5 |
| $1 \times 10^{-6}$ | 2.7015 | 10 |
| $1 \times 10^{-7}$ | 3.3307 | 15 |
| $1 \times 10^{-8}$ | 3.6763 | 18 |
| $1 \times 10^{-9}$ | 4.1293 | 21 |

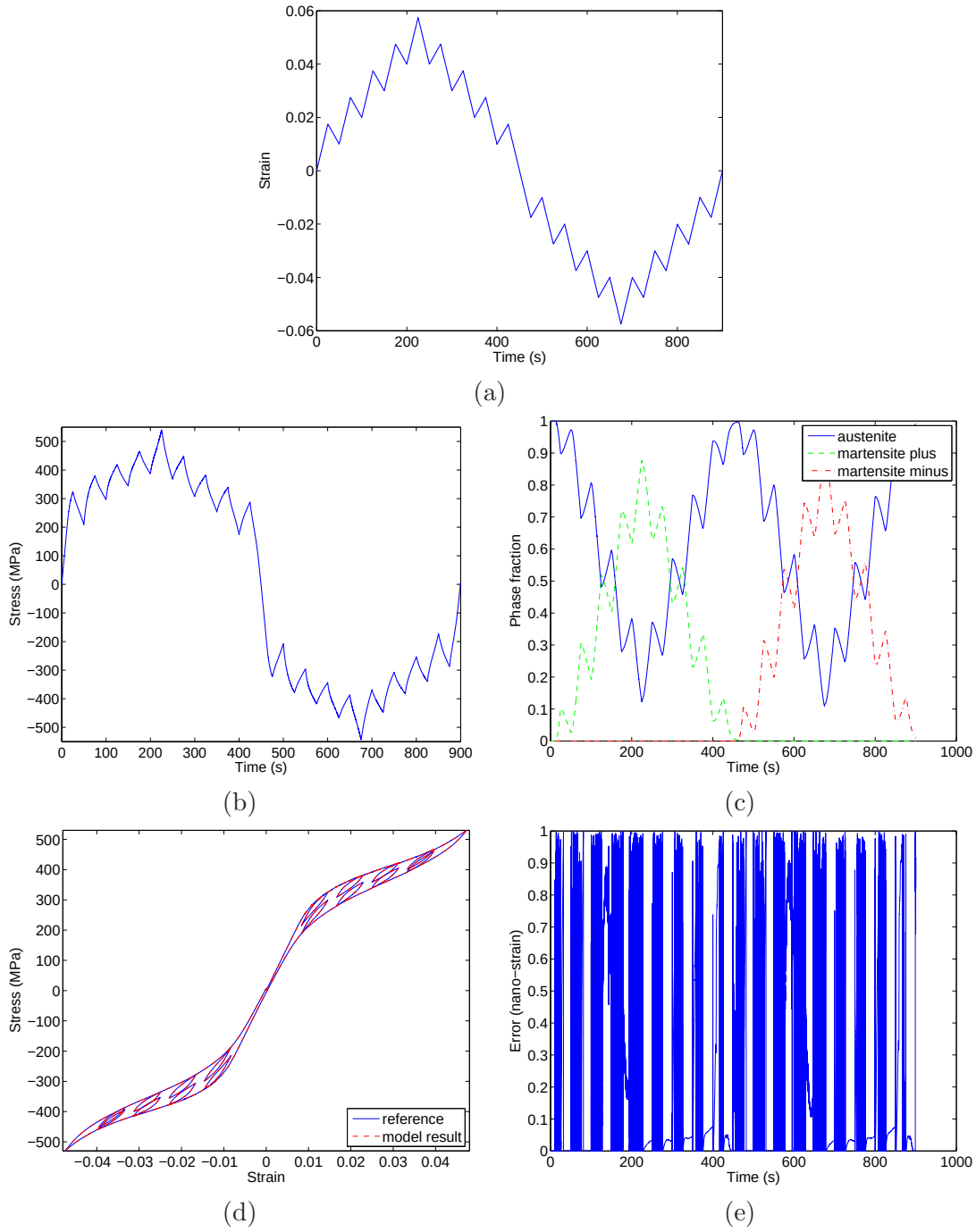



Figure 5.8: Results of the inverse SMA model to determine stress with no applied current, a constant ambient temperature of 293 Kelvin, and a strain tolerance of $1 \times 10^{-9}$. The (a) input strain $\widehat{\varepsilon}$ produced the needed stress $\sigma$ (b). The phase fractions $x_+$, $x_-$, $x_A$ predicted by the model for this stress can be observed in (c), while (d) gives the hysteresis plot and (e) gives the absolute error $|\varepsilon - \widehat{\varepsilon}|$ between requested reference strain and the strain given by the model.

site plus switches to austenite. However, to achieve actuation, a stress must be applied to the SMA (even if the stress level cannot be controlled) and it is rare for the stress to switch between compressive and tensile. Thus, we will assume that the material is either under compression and has no layers in the martensite plus phase, or is under tension and has no layers in the martensite minus phase. With this assumption, the compressive case is unimodal. The tension case is not unimodal as heating induces transformation to austenite which decreases the strain while heating also increases thermal expansion, but while layers are switching it displays a decreasing trend.

Although a slightly simpler algorithm may be proposed for the compression case, since it is unimodal, we develop here an algorithm that works equally for both cases. While we include thermal expansion, we assume that we are only interested in actuation due to phase transformation. For example, there may be two currents which yield the same strain, one by giving half austenite and half martensite, and the other by transforming the material to all austenite and then heating it until thermal expansion yields the desired strain. While the option with all austenite may give the desired strain and may (depending on the material state) be the only solution at the current timestep, over the long term this option degrades the maximum actuation rate. Once heated to this level, a much longer time must elapse for the actuator to cool sufficiently to be actuated again. Additionally, there is a risk of changing the microstructure of the SMA if it is heated too high. For these reasons, the inverse algorithm proposed here is designed to ignore this case, and only return results in the regime where phase transformations are occurring.

When considered with the approximate model from Section 5.3, the choice to primarily consider phase transformations dictates that the current at the present timestep cannot be obtained by considering only a single timestep, even if the material state is known. The phase fraction evolution (5.19) is decoupled from the thermal evolution (5.29), and as such a change in the applied current has no effect on the phase fractions predicted by the model until the following timestep. The model may be reformulated so that a change in current has an immediate effect on phase fractions. Even if this is performed, however, consideration of only a single timestep can yield a current which is clearly suboptimal over a larger time period. To illustrate, suppose the SMA is under tension and the desired strain at timestep 1 is significantly smaller than that given by the actuator with no applied current. Considering only timestep 1, the optimal approach is to apply a large current in order to significantly heat the actuator and induce a rapid phase transformation. However, with relaxation the large heat gradient from timestep 1 will continue to induce phase transformations to austenite even if the current source is turned off and the temperature is allowed to slowly decrease again. If the requested strain at timestep 2 is similar to that at timestep 1, overshoot occurs. Depending on the cooling properties of the material, it can take several timesteps without an applied current for the overshoot to be corrected. As soon as the temperature drops sufficiently that the strain with no applied current is greater than the requested strain, another large current pulse is applied to correct the strain in a single timestep.

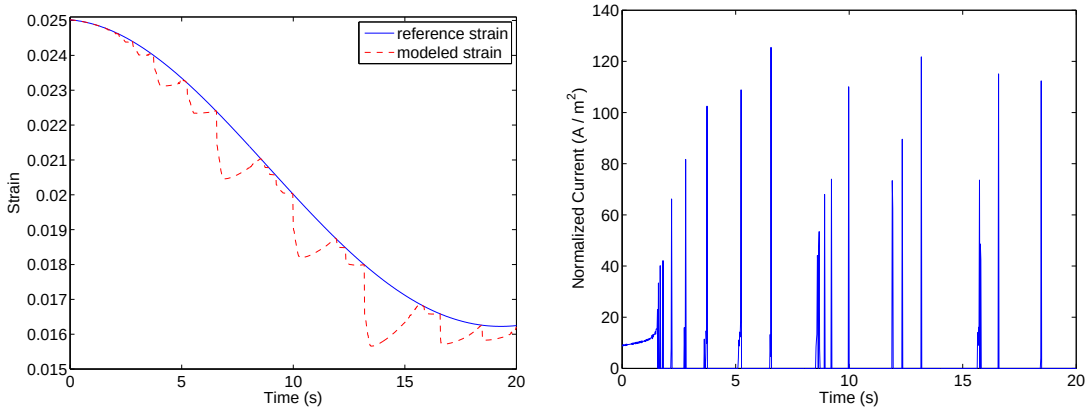
This again yields overshoot that may take several timesteps to correct. The resulting strain oscillates at and below the requested strain value and in the worst-case scenario the applied current is a sequence of impulses. We illustrate this complication in Figure 5.9.

While this problem cannot be completely mitigated within the constraints of a real-time setting, it can be substantially lessened by running the inverse model for current at two different timesteps. Let $\Delta t_1$ be the rate at which the requested strain and resulting current are updated and let $\Delta t_2 = \Delta t_1/n$ be the rate at which the the forward model described in Section 5.3 will be calculated, where $n \geq 2$ is a positive integer. In other words, the strain $\widehat{\varepsilon}$ and current $I$ will only be updated every $n$ model evaluations. Since $n \geq 2$, the model does not need to be reformulated, as a change in phase fractions due to applied current is observed before the next time the current needs to be updated. We choose $\Delta t_2$ so that the approximations in the forward model are sufficiently accurate. If $\Delta t_1$ is chosen sufficiently greater than the relaxation time constant of the material, the overshoot problem does not occur. However, smaller $\Delta t_1$ gives better tracking resolution since the current is updated more frequently. The optimal value depends on both the material and the input strain, but as shown below reasonable values will yield decent results even if some overshoot or tracking delay occurs.

The inverse problem is thus solved every $\Delta t_1$ seconds, where a function evaluations is actually $n$ evaluations of the forward model. The needed current is obtained by solving the root finding problem

$$\varepsilon(I; T_E; \sigma; x_+; x_-; T) - \widehat{\varepsilon} = 0. \tag{5.52}$$

Because the strain is not necessarily a unimodal function of current, the secant method may not step in the correct direction. We therefore cannot rely on the secant iterations to yield a




(a) (b)

Figure 5.9: Example showing the overshoot that occurs when the inverse model current is computed by only considering a single timestep of the forward model. The reference strain and strain given by the model with computed current are given in (a), while (b) gives the computed current $I/\zeta$.

bound on the root, and it is essential to provide external bounds. Only the square of the current appears in (5.29), and thus a negative current yields the same results as a positive one. Without loss of generality, we focus only on positive currents and set the lower bound of $I$ to zero. Two possibilities exist to determine the upper bound. First, we can consider the phase fraction of austenite (averaged over all layers) as a function of current and increase the current until we find a value which yields at least 99% of the material in the austenite phase. Since the phase fraction of austenite is monotone increasing with current, our standard secant/bisection approach may be utilized to solve this problem. So that cooling may occur quickly if needed at a later time, we use the minimum such current, and not just any value which yields at least 99% austenite. Numerically, this may be approximated by finding a current which yields between 99% and 99.2% of austenite. The second approach is similar: consider a user-specified maximum current, determined by properties of the actuator or the control circuitry, and if this value yields more than 99.2% austenite, use secant and bisection to back away from this value until the phase fraction is between 99% and 99.2%. The solution to the root-finding problem is required to be between these low and high values to be considered valid. If the requested strain is larger in absolute value than that given with no applied current, zero current should be returned as this is the closest possible state. If, however, the strain is less in absolute value than the value given by utilizing the maximum current, the maximum current should be returned as this gives the closest possible state. If the strain is between the values given by zero current and the maximum possible current, then barring any discontinuities introduced by quadrature (see Section 2.4) at least one solution to 5.52 exists for a current between the low and high values, and any such solutions is deemed acceptable. Note that the percentages used in this discussion are arbitrary, one could just as well find a maximum current that yields between 99.5% and 99.6% austenite, etc. Similar results will be obtained as long as the percentage is chosen large enough that no significant actuation potential is lost but less than 100% so that the value is still achievable if numerical error is present.

Although secant is less likely to converge here than in the case where stress is the quantity to determine (or the ferroelectric/ferromagnetic case), it often still performs adequately and we take the same approach here as we have previously: we try secant until the specified error tolerance is achieved or the error of the current iteration has not improved from two iterations ago; if the tolerance has still not been achieved the secant iterations together with the predetermined bounds are utilized to compute a (hopefully) tight bound on the root for the bisection method. The secant iterations may be bootstrapped by using either zero or the maximum current as the initial iterate and computing the initial derivative approximation from the evaluations of the model at zero current and the maximum current (which have already been computed). However, when the computed current from the previous timestep falls within the bounds, this frequently provides a better starting point.

Results of this approach with $\Delta t_1 = 1$ and $\Delta t_2 = 0.05$ are given in Figure 5.10 and Table 5.4 for an actuator under a tensile load of 200 MPa. Results for the same $\Delta t$ values but with a

compressive load of 200 MPa are given in Figure 5.11 and Table 5.5. Note that some overshoot still occurs, but that the results are reasonable for most control applications. In cases where the requested strain is not achievable, the algorithm returns the closest possible value, as depicted in Figure 5.12.

It should be noted that the different resistivity properties of austenite and martensite allow resistivity based feedback control to be effectively employed. Given the actuation rate of SMAs compared with the rate resistance measurements can be taken, this may often prove a better approach than reliance on an inverse compensator. However, in cases where resistance measurements are not feasible or prove inadequate, the inverse compensator provides another alternative to position the device.

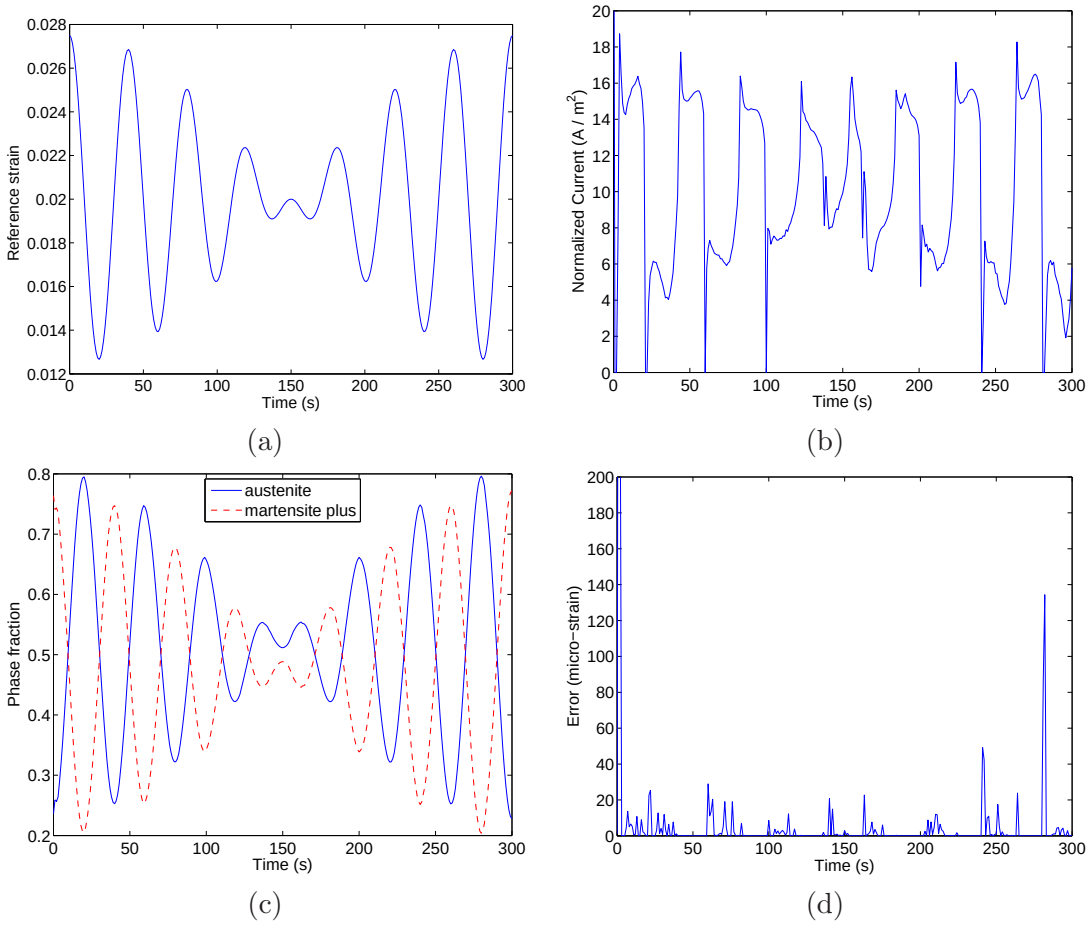



Figure 5.10: Results of the inverse SMA model with $\Delta t_1 = 1$, $\Delta t_2 = 0.05$, and $\sigma = 200$, treating current as the unknown value to determine. The (a) reference strain $\widehat{\varepsilon}$ gives the resulting (b) current $I/\zeta$, (c) phase fractions $x_+$, $x_A$ and (d) strain error $|\varepsilon - \widehat{\varepsilon}|$.

Table 5.4: Effort to compute the inverse SMA model in terms of average and maximum number of function evaluations for various error tolerances, treating current as the unknown to be determined and utilizing the input strain in Figure 5.10(a), $\Delta t_1 = 1$, $\Delta t_2 = 0.05$, and $\sigma = 200$. † denotes that due to discontinuities, some time steps did not converge within the specified tolerance, and only those that did are averaged. In these cases, the maximum number of function evaluations was the user-specified maximum.

| Tolerance | Average Function Evaluations | Maximum Function Evaluations |
|---|---|---|
| $1 \times 10^{-3}$ | 3.1096 | 9 |
| $1 \times 10^{-4}$ | 4.3355 | 12 |
| $1 \times 10^{-5}$ | 9.3478† | user-defined |
| $1 \times 10^{-6}$ | 14.8110† | user-defined |

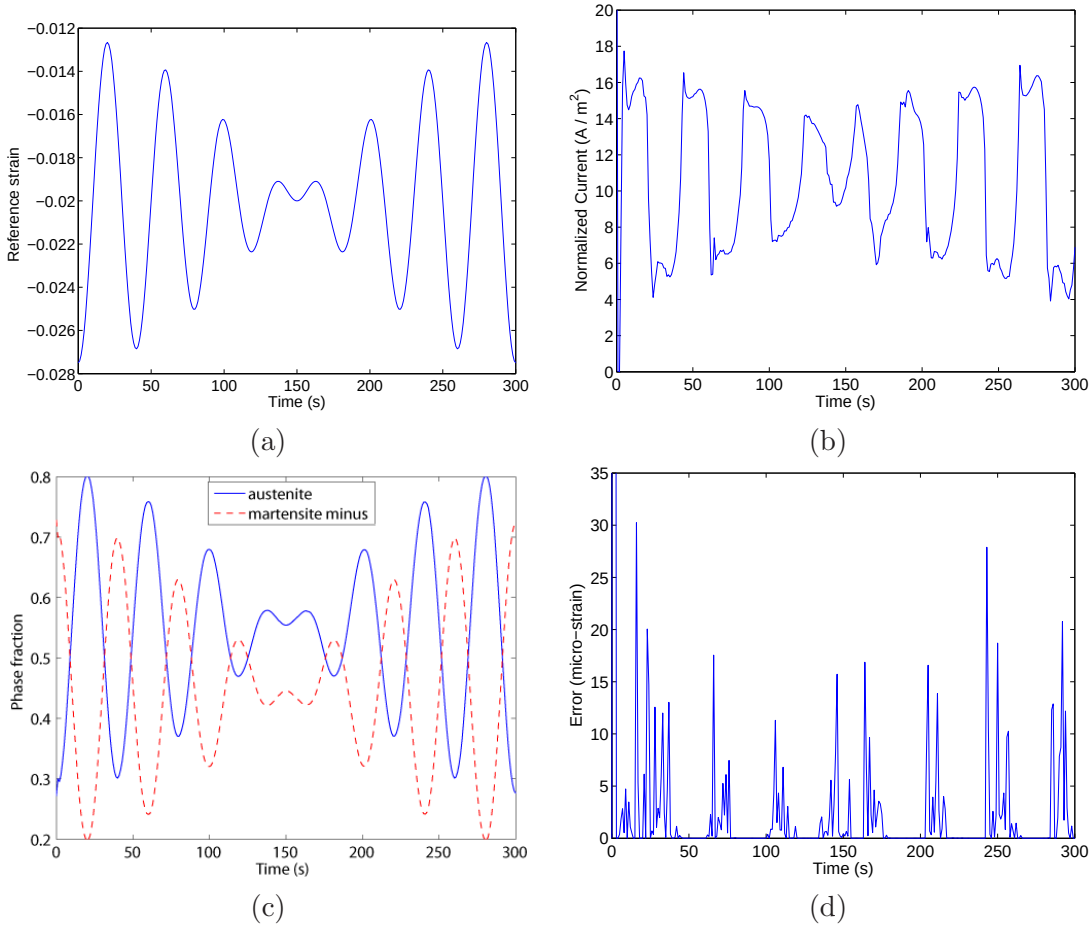
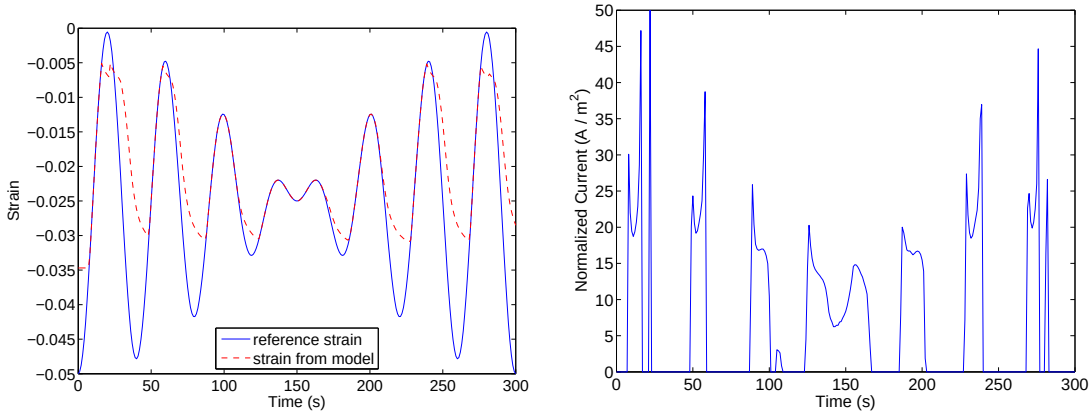



Figure 5.11: Results of the inverse SMA model with $\Delta t_1 = 1$, $\Delta t_2 = 0.05$, and $\sigma = -200$, treating current as the unknown value to determine. The (a) reference strain $\widehat{\varepsilon}$ gives the resulting (b) current $I/\zeta$, (c) phase fractions $x_-$, $x_A$ and (d) strain error $|\varepsilon - \widehat{\varepsilon}|$.

Table 5.5: Effort to compute the inverse SMA model in terms of average and maximum number of function evaluations for various error tolerances, treating current as the unknown to be determined and utilizing the input strain in Figure 5.11(a), $\Delta t_1 = 1$, $\Delta t_2 = 0.05$, and $\sigma = -200$. † denotes that due to discontinuities, some time steps did not converge within the specified tolerance, and only those that did are averaged. In these cases, the maximum number of function evaluations was the user-specified maximum.

| Tolerance | Average Function Evaluations | Maximum Function Evaluations |
|---|---|---|
| $1 \times 10^{-3}$ | 3.2791 | 9 |
| $1 \times 10^{-4}$ | 4.7841 | 14 |
| $1 \times 10^{-5}$ | 9.9733† | user-defined |
| $1 \times 10^{-6}$ | 15.7543† | user-defined |




(a) (b)

Figure 5.12: Results of the inverse model where current is the unknown quantity to be determined and the reference strain at times lies outside the values achievable by the actuator. (a) Reference strain $\widehat{\varepsilon}$ versus the strain $\varepsilon$ given by the model, and (b) computed current $I/\zeta$.

# Chapter 6

# Concluding Remarks

For applications that employ ferroelectric, ferromagnetic, or shape memory alloy actuators, the hysteretic and time-dependent behavior of the devices poses significant control problems. A completely accurate model, even if attainable, is of little value if it cannot be computed fast enough to specify the input level to the actuator. On the other hand, any attempt to control the actuator without a model describing the device must rely on significant feedback gains and a controller that updates significantly faster than the actuation frequency. This imposes an artificial limit on the actuation rate of the actuator while also introducing a greater sensitivity to noise. Classically, however, there have not been alternatives as accurate models could not be computed quickly enough.

Focusing on the homogenized energy model for its accuracy and physical motivation in energy principles, we have shown that it is possible to employ a model in a computationally efficient fashion even when thermal effects such as creep must be considered. We employed this model for ferroelectric and ferromagnetic materials both with the assumption of an approximately constant applied load and with a varying load. We also illustrate the model for shape memory alloys (SMAs). We investigated ways to maintain accuracy while reducing the number of quadrature points required by the model formulations by partitioning, thereby decreasing the computation time. We then showed that even with 1600 quadrature points, the models for ferroelectric or ferromagnetic devices could achieve rates over 14,000 timesteps per second on conventional computers, and that the model for shape memory alloys actuators could achieve a rate of over 600 timesteps per second (which given the slower actuation rates capable in SMA actuators is still significant). Finally, we introduced novel algorithms to invert these models, allowing input needed to drive the actuated to be computed from the desired polarization, magnetization, or strain. Although the speed of the inverse algorithm depends on the required accuracy, moderately high accuracy could be attained in a real-time environment within 1/20 of the speed of the forward model.

Note that these speeds are not theoretical upper limits. They represent those attained on current desktop and laptop computers with single-threaded code. The model described herein

parallelizes naturally as each timestep represents the weighted superposition of hundreds to thousands of independent evaluations of the model kernel. With the increasing prevalence of multi-core and/or multiprocessor machines, the run-time of these algorithms could likely be reduced by a factor of three or four by employing multi-threading. This potential speed-up will increase as quad-core processors become available. However, the physical/mechanical layout of actuators typically prefers a controller that is much more compact than can be provided by a desktop or laptop computer. For these regimes, field programmable gate array (FPGA) or application specific integrated circuit (ASIC) designs should be considered. While it remains an active research area, these may allow for performance gains of over an order of magnitude for these models.

Finally, we note that control methods for use with these actuators remains an open area of research. However, our research is not focussed on developing new control methodologies; the concepts of model-based control and inverse compensation are well-known in the control literature. It has only recently been possible to efficiently apply these methodologies due to computational limitations. The improvements given here, as well as future improvements in the computational performance of the models, will increasingly allow the rich control theory that has already been developed to be applied to smart material transducers.

# Bibliography

[1] B.L. Ball, R.C. Smith, S. Kim, and S. Seelecke, "A ferroelectric switching model for lead zirconate titanate (PZT)," Smart Structures and Materials 2005, Proceedings of the SPIE, Volume 5757, pp. 1-9, 2005.

[2] B.L. Ball, R.C. Smith, S. Kim, and S. Seelecke, "A stress-dependent hysteresis model for ferroelectric materials," *Journal of Intelligent Material Systems and Structures*, 18(1), pp. 69-88, 2007.

[3] T.R. Braun and R.C. Smith, "High speed model implementation and inversion techniques for ferroelectric and ferromagnetic transducers," CRSC Technical Report CRSC-TR07-10, *Journal of Intelligent Material Systems and Structures*, submitted.

[4] T.R. Braun and R.C. Smith, "Efficient implementation algorithms for homogenized energy models," *Continuum Mechanics and Thermodynamics*, 18(3-4), pp. 137-155, 2006.

[5] T.R. Braun, R.C. Smith, and M.J. Dapino, "Experimental validation of a homogenized energy model for magnetic after-effects," *Applied Physics Letters*, 88(12), 2006.

[6] R.P. Brent, *Algorithms for Minimization without Derivatives*, Prentice-Hall, Englewood Cliffs, N.J., 1973.

[7] F.T. Calkins and G.W. Butler, "Subsonic jet noise reduction variable geometry chevron," Proceedings of the 42nd AIAA Aerospace Sciences Meeting and Exhibit, 2004.

[8] W.J. Cody, "Rational chebyshev approximations for the error function," *Mathematics of Computation*, 23(107), pp. 631-637, 1969.

[9] W.J. Cody and W. Waite, *Software Manual for the Elementary Functions,* Prentice Hall, Englewood Cliffs, NJ, 1880.

[10] M.J. Dapino, R.C. Smith, L.E. Faidley, and A.B. Flatau, "A coupled structure-magnetic model for magnetostrictive transducers," *Journal of Intelligent Material Systems and Structures*, 11(2), pp. 134-152, 2000.

[11] T.W. Duerig, "The use of superelasticity in modern medicine," *MRS Bulletin*, 27(2), pp. 101-104, 2002.

[12] T.W. Duerig, A. Pelton and D. Stöckel, "An overview of nitinol medical applications," *Material Science and Engineering: A, Structural Materials: Properties, Microstructure and Processing*, Volumes 273-275, pp. 149-160, 1999.

[13] J. Enkovaara, A. Ayuela, L. Nordström and R.M. Nieminen, "Magnetic anisotropy in Ni2MnGa," *Physical Review B*, 65(13), 2002.

[14] P. Ge and M. Jouaneh, "Modeling hysteresis in piezoceramic actuators," *Precision Engineering*, 17, pp. 211-221, 1995.

[15] P. Ge and M. Jouaneh, "Tracking control of a piezoceramic actuator," *IEEE Transactions on Control Systems Technology*, 4(3), pp. 206-209, 1996.

[16] A.G. Hatch, R.C. Smith, T. De, and M.V. Salapaka, "Construction and experimental implementation of a model-based inverse filter to attenuate hysteresis in ferroelectric transducers," *IEEE Transactions on Control Systems Technology*, 14(6), pp. 1058-1069, 2006.

[17] A. Ishida, A. Takei, et. al., "Shape memory behavior of Ti-Ni thin films annealed at various temperatures," *Materials Research Society Symposium Proceedings*, 360, pp. 381-385, 1995.

[18] D.C. Jiles and D.L. Atherton, "Theory of ferromagnetic hysteresis", *Magnetism and Magnetic Materials*, 61, pp. 1265-1281, 1984.

[19] C.T. Kelley, *Iterative Methods for Linear and Nonlinear Equations,* SIAM, Philadelphia, 1995.

[20] C.S. Lynch, "The effect of uniaxial stress on the electro-mechanical response of 8/65/35 PLZT," *Acta Materialia*, 44(10), pp. 4137-4148, 1996.

[21] J.A. Main and E. Garcia, "Design impact of piezoelectric actuator nonlinearities," *Journal of Guidance, Control and Dynamics*, 20(2), pp. 327-332, 1997.

[22] J.A. Main and E. Garcia, "Piezoelectric stack actuators and control system design: strategies and pitfalls," *Journal of Guidance, Control, and Dynamics*, 20(3), pp. 479-485, 1997.

[23] J.A. Main, E. Garcia and D.V. Newton, "Precison position control of piezoelectric actuatorss using charge feedback," *Journal of Guidance, Control, and Dynamics*, 18(5), pp. 1068-73, 1995.

[24] J.A. Main, D. Newton, L. Massengil and E. Garcia, "Efficient power amplifiers for piezoelectric applications," *Smart Structures and Materials*, 5(6), pp. 766-775, 1996.

[25] J.E. Massad, "Macroscopic models for shape memory alloy characterization and design," Doctoral dissertation, North Carolina State University, 2003.

[26] J.E. Massad and R.C. Smith, "A homogenized free energy model for hysteresis in thin-film shape memory alloys," *Thin Solid Films*, 489(1-2), pp. 266-290, 2005.

[27] J.E. Massad and R.C. Smith, "A domain wall model for hysteresis in ferroelastic materials," *Journal of Intelligent Material Systems and Structures*, 14(7), pp. 455-471, 2003.

[28] Y. Matsuzaki, K. Funami and H. Naito, "Inner loops of pseudoelastic hysteresis of shape memory alloys: Preisach approach," Smart Structures and Materials 2002, Proceedings of the SPIE, 4699, pp. 355-364, 2002.

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

[30] S. Mittal and C-H. Meng, "Hysteresis compensation in electromagnetic actuators through Preisach model inversion," *IEEE/ASME Transactions on Mechatronics*, 5(4), pp. 394-409, 2000.

[31] J.M. Nealis and R.C. Smith, "$\mathcal{H}_\infty$ control design for a magnetostrictive transducer," Proceedings of the 42nd IEEE Conference on Decision and Control, pp. 1801-1806, 2003.

[32] J.M. Nealis and R.C. Smith, "Model-based robust control design for a magnetostrictive transducer operating in hysterestic and nonlinear regimes," *IEEE Transactions on Control Systems Technology*, 15(1), pp. 22-39, 2007.

[33] W.S. Oates, P. Evans, R.C. Smith, and M.J. Dapino, "Experimental implementation of a nonlinear control method for magnetostrictive transducers," Proceedings of the 46th IEEE conference on Decision and Control, to appear.

[34] K. Oikawa, T. Ota, T. Ohmori, Y. Tanaka, H. Morito, A. Fujita, R. Kainuma, K. Fukamichi, and K. Ishida, "Magnetic and martensitic phase transitions in ferromagnetic NiGaFe shape memory alloys," *Applied Physics Letters*, 81(27), pp. 5201-5203, 2002.

[35] W.H. Press, S.A. Teukolsky, W.T. Vetterline, B.P. Flanner, *Numerical Recipes in C,* 2nd Edition, Cambridge University Press, New York, 1992.

[36] A. Ralston, "Rational Chebyshev approximation," *Mathematical Methods for Digital Computers*, Vol. 2, John Wiley and Sons, New York, pp. 268-284, 1967.

[37] G. Robert, D. Damjanovic and N. Setter, "Preisach modeling of piezoelectric nonlinearity in ferroelectric ceramics," *Journal of Applied Physics*, 89(9), pp. 5067-5074, 2001.

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

[39] S. Seelecke and I. Müller, "Shape memory alloy actuators in smart materials – modeling and simulation," *ASME Applied Mechanics Reviews*, 57(1), pp. 23-46, 2004.

[40] R.C. Smith, *Smart Material Systems: Model Development*, SIAM, Philadelphia, 2005.

[41] R.C. Smith, M.J. Dapino, T.R. Braun and A.P. Mortensen, "A Homogenized energy framework for ferromagnetic hysteresis," *IEEE Transactions on Magnetics*, 42(7), pp. 1474-1769, 2006.

[42] R.C. Smith, M.J. Dapino and S. Seelecke, "A free energy model for hysteresis in magnetostrictive transducers," *Journal of Applied Physics*, 93(1), pp.458-466, 2003.

[43] R.C. Smith, A.G. Hatch, T. De, M.V. Salapaka, R.C.H. del Rosario, and J.K. Raye,"Model Development for Atomic Force Microscope Stage Mechanisms," *SIAM Journal on Applied Mathematics*, 66(6), pp. 1998-2026, 2006.

[44] R.C. Smith and J.E. Massad, "A unified methodology for modeling hysteresis in ferroic materials," Proceedings of the 2001 ASME Design Engineering Technical Conferences and Computer and Information in Engineering Conference, Vol 6, Pt B, pp. 1389-1398, 2001.

[45] R.C. Smith, S. Seelecke, M.J. Dapino and Z. Ounaies, "A unified framework for modeling hysteresis in ferroic materials," *Journal of the Mechanics and Physics of Solids*, 54(1), pp. 46-85, 2006.

[46] R.C. Smith, S. Seelecke, Z. Ounaies and J. Smith, "A free energy model for hysteresis in ferroelectric materials," *Journal of Intelligent Material Systems and Structures*, 14(11), pp. 719-739, 2003.

[47] G. Song, J. Zhao, X. Zhou and J.A. Abreu-Garci, "Tracking control of a piezoceramic actuator with hysteresis compensation using inverse Preisach model,", *IEEE/ASME Transactions on Mechatronics*, 10(2), pp. 198-209, 2005.

[48] X. Tan and J.S. Baras, "Modeling and control of hysteresis in magnetostrictive actuators," *Automatica*, 40(9), pp. 1469-1480, 2004.

[49] X. Tan, R. Venkataraman and P.S. Krishnaprasad, "Control of hysteresis: Theory and experimental results," Smart Structures and Materials 2001, Proceedings of the SPIE, Volume 4326, pp. 101-112, 2001.

[50] G. Webb, A. Kurdila and D.C. Lagoudas, "Adaptive hysteresis model for model reference control with actuator hysteresis," *Journal of Guidance, Control and Dynamics*, 23(3), pp. 459-465, 2000.

[51] A. York, "Experimental Investigation of Rate-Dependent Inner Hysteresis Loops in PZT," Master's thesis, North Carolina State University, 2004.