Mesoscopic Free Energy Model for Shape Memory Alloys Under Complex Loading

Dissertation

zur Erlangung des Grades

des Doktors der Ingenieurwissenschaften

der Naturwissenschaftlich-Technischen Fakultät

der Universität des Saarlandes

von

Wesley Ballew

Saarbrücken

2021

Tag des Kolloquiums:	2021-07-01
Dekan:	Univ.-Prof. Dr. Rer. Nat. Jörn Walter
Berichterstatter/innen:	Univ.-Prof. Dr. Ing. Habil. Stefan Seelecke Univ.-Prof. Dr. Ing. Stefan Diebels
Vorsitz:	Univ.-Prof. Dr. Techn. Romanus Dyczij-Edlinger
Akad. Mitarbeiter/in:	Dott. Ing. Emanuele Grasso

Kurzfassung

In dieser Dissertation wird ein bestehendes Modell für einachsige Formgedächtnislegierungen (SMAs) modifiziert, um gleichzeitig zusätzliche Fähigkeiten und eine einfachere Formulierung zu erhalten; es wird dann als Plattform zur Entwicklung eines neuen Modells verwendet, das ein kombiniertes Axial- und Scherverhalten (2-D) aufweist.

Eine detaillierte Untersuchung des 1-D-SMA-Modells von Müller, Achenbach und Seelecke (M-A-S) identifiziert und behandelt mehrere Merkmale, die eine Herausforderung für eine multiaxiale Theorie darstellen. Zu diesen Entwicklungen gehören eine neue Technik zur Vorhersage der Wahrscheinlichkeit einer Phasenumwandlung, eine neue Formulierung der freien Energie, die die Martensit-Varianten voneinander entkoppelt, und eine quantitative Untersuchung der Latentwärme-Vorhersagen. Bevor mit der 2-D-Entwicklung begonnen wird, wird die neue Formulierung ausgenutzt, um zwei neue 1-D-Fähigkeiten hinzuzufügen, nämlich die Zug-Druck-Asymmetrie und die temperaturabhängige Hysterese, die bisher nicht im M-A-S-Modell berücksichtigt wurden. Diese beiden Merkmale werden zusammen mit einer Reihe anderer üblicher SMA-Effekte in einer Reihe von Simulationen demonstriert, die mit experimentellen Daten verglichen werden.

Die Entwicklung des 2D-Modells konzentriert sich auf einen einzigen Satz experimenteller Daten, bei dem NiTi-Drähte sowohl einer reinen Zug- als auch einer kombinierten Zug-Torsions-Belastung ausgesetzt wurden. Die Tests umfassten einen breiten Temperaturbereich und mehrere verschiedene nichtproportionale Belastungspfade. Die Experimente zeigten eine Reihe von charakteristischen Hystereseschleifen, und diese Theorie besagt, dass nur vier Martensitvarianten in zwei symmetrischen Paaren den gesamten Satz der beobachteten Reaktionen erklären können. Die Kinematik wird zunächst qualitativ begründet und dann werden die vier Varianten mit einer freien Energielandschaft auf einer normalen Dehnungs-Scher-Dehnungs-Domäne in Beziehung gesetzt. Das 2-D-Modell wird unter Verwendung derselben thermodynamischen Prinzipien wie die 1-D-Version abgeleitet, mit Ausnahme der kinetischen Gleichungen, die neu formuliert werden mussten, um der Tatsache Rechnung zu tragen, dass transformierende Varianten mehrere mögliche Ziele haben. Die Variantenauswahl wird nun von der lokalen Stabilität der aktuellen Phase, den lokalen Stabilitäten der Zielphasen und der mit jedem möglichen Pfad verbundenen Energiedifferenz bestimmt. Die Simulationen der 2-D-Experimente zeigen vielversprechende Ergebnisse, ohne das 1-D-Verhalten des Modells zu beeinträchtigen. Das Modell sagt das Zug-Torsionsverhalten über den gesamten Bereich der Temperatur/Axialspannungskombinationen genau voraus und erfasst auch die beiden Hauptmodi der axialen Reaktion. Die Größenordnungen der axialen Dehnungen sind in einigen Fällen ungenau, und dies bietet eine Gelegenheit für zukünftige Arbeiten, die hier vernachlässigten makroskopischen Feldeffekte zu untersuchen.

Abstract

In this dissertation an existing model for uniaxial shape memory alloys (SMAs) is modified to simultaneously have additional capabilities and a simpler formulation; it is then used as a platform to develop a new model capable of combined axial and shear (2-D) behavior.

A detailed examination of the 1-D SMA model by Müller, Achenbach, and Seelecke (M-A-S) identifies and addresses several features that present challenges to a multiaxial theory. These developments include a new technique to predict the probability of phase transformation, a new free energy formulation that decouples the martensite variants from one another, and a quantitative examination of the latent heat predictions. Before beginning the 2-D development, the new formulation is exploited to add two new 1-D capabilities, tension-compression asymmetry and temperature-dependent hysteresis, that were not previously accommodated by the M-A-S model. These two features, along with a number of other common SMA effects, are demonstrated in a series of simulations that are compared to experimental data.

The 2-D model development is focused on a single set of experimental data where NiTi wires were subjected to pure tension as well as combined tension-torsion loading. The tests included a wide range of temperatures and several different non-proportional loading paths. The experiments revealed a series of distinctive hysteresis loops and this theory proposes that only four martensite variants, in two symmetric pairs, can explain the entire set of responses observed. The kinematics are first justified in a qualitative manner and then the four variants are related to a free energy landscape on a normal strain-shear strain domain. The 2-D model is derived using the same thermodynamic principles as the 1-D version except for the kinetic equations, which had to be reformulated to account for the fact that transforming variants have multiple possible destinations. The variant selection is now governed by the local stability of the current phase, the local stabilities of the destination phases, and the energy difference associated with each possible path. Simulations of the 2-D experiments show promising results without compromising the 1-D behavior of the model. The model accurately predicts the tension-torsion behavior across the entire range of temperature/axial stress combinations and also captures the two major modes of axial response. The magnitudes of the axial strains are inaccurate in some cases and this presents an opportunity for future work to explore the macroscopic field effects that were neglected here.

To: Determination, persistence, and patience

Acknowledgements

I would like to express my gratitude to Prof. Stefan Seelecke for his patience and trust. The pursuit of this degree has spanned almost 15 years and continued through numerous life-altering events – losing my first engineering job, finding my second engineering job, marrying my wife, transferring my studies from NC State University to Universität des Saarlandes, the births of both of my children, the transition to my third engineering job, and finally a global pandemic. Each of these stages presented new challenges to a part-time academic pursuit and I do not know if Prof. Seelecke ever thought I would just let it go. Either way, the door was held open as long as I was willing to keep crawling toward it.

I must also thank my wife Lauren for all the times she took on extra duties, especially through the baby-years, to allow me to hide in my office and think about shape memory alloys and energy wells. My children do not even know that Daddy is "*still in school*".

Last, I would like to acknowledge the engineering communities where I have developed my professional career. My mentors and colleagues at Martin Marietta Composites, Mallett Technology, and Sierra Nevada Space Systems have encouraged and supported my studies along the way. A few words here, among many, have helped me stay motivated when a part-time PhD seemed like an impossible task.

"The PhD will yield to persistence."
- Dr. James Bosnik, Mallett Technology

"If your advisor does not like you, you will never graduate."
- Dr. Abed Khaskia, Mallett Technology

"If you hang around long enough, maybe they just give you the damn PhD."
- Greg Solomon, Martin Marietta Composites

Table of Contents

1 Introduction

1.1 Objective and Outline

The starting point for this dissertation is a material model for the 1-D, elastocaloric behavior of shape memory alloys from Müller, Achenbach, and Seelecke (Müller and Wilmanski (1981), Achenbach and Müller (1985), Achenbach (1989), Müller and Seelecke (2001)). This dissertation examines the model's underlying thermodynamics and kinetic theory and then contributes an updated version of the 1-D model that extends its capabilities as well as a new 2-D model that is applied to combined tension-shear loading. We begin with an overview of the document and then give a brief introduction to some common SMA modeling techniques.

In Chapter 2 we shift the focus to the Müller, Achenbach, and Seelecke (M-A-S) model. This begins with a discussion of its history as well as commentary on how it is different from and similar to other SMA modeling approaches. Next, we introduce the concept of a mesoscopic layer and its free energy as well as how they fit into the multiscale system. Equipped with only these concepts, before presenting the first equation, we show a series of figures that demonstrate how a multiwell free energy can produce several common SMA effects. The intuitive, visualizable nature of this model is one of its advantages among SMA theories that have a wide range of complexity and abstractness. The remainder of the chapter is devoted to the basic mathematical structure of the M-A-S model. This includes the Helmholtz free energy, Gibbs free energy, kinetics law, stress-strain equations, and energy balance equations.

Chapter 3 is the first of two in-depth examinations and is dedicated to the kinetics law. The kinetics equations use a statistical description of the mesoscopic system to predict phase change at the macroscopic scale. In this way they are the main connection between the mesoscopic and macroscopic systems. They are based on the calculation of transformation probabilities, which are derived from the Boltzmann equation. The transformation probabilities have been calculated a number of different ways between various applications of the M-A-S model, but each form is still based on the same parent equation. Beginning with the Boltzmann equation, we derive several of the different expressions and discuss their applications, advantages, and disadvantages. Based on this assessment, we next offer a new transformation probability equation that combines the benefits of two previous forms. This version reduces the number of equations required to operate the 1-D model and is well-suited for extension to 2-D. This chapter concludes with section devoted to the transformation rate parameters. The kinetic equations describe phase change as a time-dependent process and, since SMA transformations occur very rapidly, care must be taken to strike the proper balance between rapid transformations and numerical issues.

The second in-depth discussion concerns the model's thermodynamics in Chapter 4. First, we address the issue of how stress, strain, and temperature at the mesoscopic scale and macroscopic scale are related. This is an important groundwork to lay because many of the parameters used to define the mesoscopic free energy are obtained from macroscopic test data. After establishing this connection, we show the complete construction of the multiwell free energy from a simple set of experimentally obtained parameters. Next, we show how the free energy can be used to relate transformation stress to other thermodynamic properties such as chemical free energy, internal energy, entropy, and latent heat. Special attention is given to the irreversibility that is inherent in a multiwell free energy-system. Many SMA theories, that address the thermodynamics of phase change, assume reversibility when relating these quantities and this yields equations analogous to the Claussius-Clapeyron equation. We show that, due to the smoothness conditions on the M-A-S free energy, we can use the same

procedures to derive internal energy, entropy, and latent heat expressions that include irreversibility. To test these equations, we compare latent heat predictions to experimentally measure parameters as well as predictions using the traditional, reversible, equations. The chapter concludes with an additional exercise that relates the macroscopic stress-strain curve, the total entropy produced during the stress-strain cycle, and the latent heat difference between forward and reverse transformations.

In Chapter 5, we use the thermodynamic insights from Chapter 4 and the updated kinetics model from Chapter 3 to develop a new version of the 1-D model. This model has two additional capabilities: tension-compression asymmetry and temperature-dependent hysteresis. This is accomplished by reinterpreting the free energy such that each martensite variant can have a unique well and barrier

. The curvatures of the energy barriers (which dictate the hysteretic behavior) are also allowed to be functions of temperature. These generalizations involve some algebraic restructuring, the use of a dual set of material parameters, and also re-derivation of the caloric quantities. This chapter also includes a series of simulations that show tension-compression symmetry, temperature-dependent hysteresis, and a variety of other SMA effects. In each case the model is compared to experimental data and the results are discussed. For readers interested in implementing the model, the chapter concludes with user's guide including a list of parameters with descriptions and values along with a flowchart of required equations.

In Chapter 6, we use the 1-D model as the basis to formulate a new multidimensional SMA model. This requires the introduction of four martensite variants, a 2-D free energy landscape, and the assertion that symmetry between the variant pairs enables uniaxial tension, pure shear, or coupled behavior. Using principles from the modified free energy in Chapter 5, we formulate a new free energy that operates in terms of critical free energies as opposed to critical stresses or strains. This approach is more general and naturally accommodates multiaxial coupling. The new free energy and transformation criteria also necessitate the development of a novel set of transformation probabilities that utilize the stability of current phase to initiate transformation and the stability of the destination phases to govern transformation paths (i.e. variant selection). An extensive set of tension-torsion experimental data (Sittner et al. (2009)) is used to develop the model and to test its performance in uniaxial tension tests, uniaxial actuation tests, and numerous combined tension-torsion tests.

The dissertation concludes with a set of appendices containing supplemental material. This includes details overcoming numerical overflow issues for one of the transformation probability equations, step-by-step instructions to build the tension-compression asymmetric free energy, free energy plots for several different sets of experimental data, additional stress-strain calculations for the 2-D model, and a summary of tension-torsion simulations results from other SMA models.

1.2 SMA Modeling Overview

Shape memory alloys offer an enticing array of possibilities for engineering applications due to their unique properties, namely superelasticity and the shape memory effect. Over the last 60 years or so, this research field has grown steadily and the number of successful SMA-based devices in industry continues to increase. Despite this level of maturity as a technology, SMAs are still largely considered to be niche products and perceived (fairly or not) to be complex and difficult for design. This reputation is understandable, if not well-deserved, because the behavior we observe in laboratory specimens is the result of several different, interrelated mechanisms occurring simultaneously at multiple physical scales. The main way that researchers and engineers have navigated this challenging landscape is by restricting the scope of the experiments, theories, and devices to 1-D. The number of

SMA publications that focus on 1-D behavior vastly outweigh those considering two or more dimensions. Even under the narrowed scope of uniaxial behavior, SMAs are nonlinear, hysteretic, and thermo-mechanically coupled. The field of material models aiming to capture these effects is wide and includes a variety of approaches. In this section, rather than attempt to provide a comprehensive review of the SMA modeling, we categorize a handful of approaches that either contrast or have similarities to the M-A-S model. More general and comprehensive reviews can be found in Ziolkowski (2015), Barbarino et al. (2014), Lagoudas et al. (2006), Patoor et al. (2006), as well as Paiva and Savi (2006).

1.2.1 Phenomenological (Macromechanical) Models

One of the main distinguishing features between SMA material models is the physical meaning of the internal variables. They are mostly related to the presence of martensitic microstructures, but there is still room for wide variation from one interpretation to the next. A number of SMA theories describe martensite in a purely phenomenological manner, at a scale much larger than the crystal lattice (Boyd and Lagoudas (1996), Bo and Lagoudas (1999a,b), Lagoudas and Bo (1999), Bernardini and Pence (2002), Müller and Bruhns (2006), Zaki (2012), Morin et al. (2011), Zaki and Moumni (2007), Thamburaja (2010), Armattoe et al. (2014), Chemisky et al. (2011), Saleeb et al. (2011), Xiao (2014), Peultier et al. (2006), Zhou (2012), Auricchio and Sacco (1999), Masud et al. (1997), Auricchio et al. (1997), Auricchio and Lubliner (1997), Auricchio and Sacco (1997), Lubliner and Auricchio (1996), Arghavani et al. (2010), Auricchio et al. (2003), Auricchio and Sacco (2001), Bodaghi et al. (2013), Stebner and Brinson (2013), Christ and Reese (2009), Reese and Christ (2008), in addition to Helm and Haupt (2003)). These formulations assume effective transformation strains that represent a large number of arbitrarily oriented crystals expressing an arbitrary number of variants. Under this assumption, the transformation strain accumulates in any direction and is dictated by the state of stress via yield surfaces and flow rules. These models usually include a macroscopic free energy and adopt traditional continuum thermodynamics guidelines such as those set forth by Coleman and Noll (1963) and Coleman and Gurtin (1967). This framework dictates that the free energy must be convex and that all constitutive equations must be derived from partial derivatives of it. Moving forward, we will use the phrase *traditional continuum thermodynamics* to refer to this general approach. The plasticity framework is easily extended to multiple dimensions and additional internal variables are readily added to introduce secondary effects such as plastic strain and cyclic evolution. One drawback to this type of framework is that the free energies (and the other model quantities) lack a direct connection to phases; this means that SMA effects such as transformation reversal and hysteresis must be reintroduced through phase diagrams or other logical constructs.

1.2.2 Micromechanical Models

A contrasting approach to SMA modeling is motivated by micromechanics where internal variables represent specific crystallographic martensite variants (Patoor et al. (1989), Patoor et al. (1996), Entemeyer et al. (2000), Huang and Brinson (1998), Yu et al. (2012), Wang et al. (2007), Wang et al. (2008), and Richards et al. (2013)). Since each variant is based on a specific crystallographic configuration, it can be associated with a unique transformation strain tensor and phase fraction. The transformation strains are weighted by the phase fractions and superimposed to represent the total strain of a single SMA crystal. These formulations also typically follow continuum thermodynamics guidelines but in contrast to plasticity-type models, the partial derivatives yield stress-strain relationships and driving forces for individual phases. Interaction energies are often included to motivate favorable combinations of variants (i.e. twinning) and temperature-dependent (chemical) energies promote

temperature-induced transformations. This type of model can represent complex 3-D single crystal behavior that can be traced to specific micromechanical effects. These models are also often extended to polycrystalline applications, but this is usually an averaging or field solution based on many instances the single-crystal formulation.

1.2.3 Phase Field Models and Localization Effects

Phase field models describe a spatial domain that is divided into regions, each occupied by a single martensite (James and Kinderlehrer (1989), Ball et al. (1992), Abeyaratne and Knowles (1988), Abeyaratne and Knowles (1989), Abeyaratne and Knowles (1990), Abeyaratne and Knowles (1991), Leo et al. (1993)). The boundaries between the phases are dynamic and phase change is represented the growing or shrinking of various zones. This type of model represents a new category because it often uses a multiwell free energy to motivate the system's metastable nature. In free energy space, each well contains some amount of material and is separated from the other phases by energy barriers. When stress or temperature conditions push one phase out of its energy well (i.e. into an unstable configuration), transformation occurs via boundary motion. The collective response of the phase fields yields the SMA material behavior and maintains global equilibrium. If the representative phase field volume is assumed to be very small (micro or nanometer), then these models are well suited to predict effects like self-accommodation and variant selection. If the phase field domain is considered to be the entire macroscopic body, these models are capable of capturing transformation front motion of the Lüders band-type. A number of SMA traits are strongly related to macroscopic phase field effects and this continues to be an active area of research (Shaw and Kyriakides (1995), Shaw (2002), Shaw et al. (2008), Churchill et al. (2009a), Churchill et al. (2009b), Churchill et al. (2010), Dai and Wang (2010), Dhote et al. (2012), Dhote et al. (2013), Song et al. (2013)).

1.2.4 Multidimensional Models

Some devices, like stents and morphing structures (Rogers et al. (1999), Duerig et al, (2000), Migliavacca et al. (2002), Gong et al. (2003), Stoeckel et al. (2004), Kleinstreuer et al. (2008), Ballew and Seelecke (2009), Zaki (2012), Hartl and Lagoudas (2007), Barbarino et al. (2014)) utilize SMA behavior in multiple dimensions and push the material into multiple axes of stress and strain. Despite the potential of multidimensional SMA devices, the design and modeling of such devices is challenging for a number of reasons including 1) there are numerous variants (24 for NiTi for example), 2) these variants form in alternating microscopic layers that interact with one another, 3) the alternating layers form mesoscopic plate groups that further interact with each other and surrounding material, and 4) these features can initiate at multiple locations within a macroscopic specimen which can grow, shrink, or travel.

Most efforts target simple two-dimensional experiments such as biaxial tension/compression or combined tension and torsion (Bouvet et al. (2004), Arghavani et al. (2010), Saleeb et al. (2011), Sedlak et al. (2012), among others). Even in these simple systems the observed behavior can be complex and sometimes difficult to attribute to a specific martensitic effect. In 2009 a ESF S3T EUROCORES-sponsored program called *Roundrobin SMA Modeling* was completed with two main goals, 1) to publish a rich experimental data set for multidimensional SMA modelers to use and 2) to provide some modeling benchmarks for the field. The experimental data set examined the thermomechanical behavior of NiTi wires under a wide array of multiaxial and non-proportional loading scenarios. For modeling portion of the program several teams of researchers used their SMA models to simulate the experiments and then the results were published in a conference paper (Sittner et al (2009)). In this dissertation we make use of the data from this program and refer to it frequently. For brevity, we will refer to the

Roundrobin SMA Modeling program as simply the *RR* program and the data as *RR* data. The modeling efforts stemming from this program demonstrated a number of different approaches, levels of completeness, and degrees of success.

The model by Gibeau et al. (2009) uses stress-induced martensite as well as self-accommodated martensite internal variables. A macroscopic Helmholtz free energy contains terms related to internal energy, entropy, elastic energy, specific heat, and variant interaction. A Gibbs free energy is derived from the Helmholtz energy and it also has a phase dependent elastic modulus. This is an often-used technique to produce smooth, polycrystalline-like stress-strain curves as opposed to the trilinear, single-crystal type. The free energies are used in the traditional continuum thermodynamics manner to define driving forces for elastic strain, thermal strain, transformation strain, entropy, as well as the proportions of stress-induced and self-accommodated martensite. A yield function is used to define the threshold for transformation while a Claussius-Duhem inequality (along with several logical conditions) are used to supply flow rules. A specific choice of stress-invariant parameters is used to also enable tension-compression asymmetry – this is similar to Drucker-Prager methods which are common concrete and soil modeling. This model has reasonable accuracy in most of the uniaxial simulations and showed only a small handful of tension-torsion cases where the agreement was poor. These results, as well as several other *RR* program publications, are included in appendix to this dissertation (Chapter 10.6).

Rio et al. (2009) employs an elasto-hysteresis model where the total stress was composed of a hyperelastic component and an elastoplastic hysteresis component. The hyperelastic potential is based on strain invariants, relative volume variation, deviatoric deformation intensity, and an angle giving the direction of the deformation. The hysteresis is governed by an incremental model with discrete memory points. This approach is one of the more unique ones among the *RR* program participants (and SMA models in general) and showed promising correlation to the experimental data. This model showed excellent agreement in the uniaxial tension cases and also captured the torque-twist responses very well (see Chapter 10.6). The axial strain-twist curves were qualitatively similar to the experimental data, but the magnitudes were incorrect. This is also one of the main deficiencies of the model developed in this dissertation and will be discussed further in Chapter 6.

Chemisky and colleagues (2009) present a model where the inelastic strain is an average of the strain stemming from martensite transformation and from the formation of twinned martensite. A Gibbs free energy is written containing chemical, elastic, thermal strain, inelastic, and mixing energy terms. A set of dissipation potentials and evolution rules govern the progression of the internal variables and phase diagram-like method is used to identify the material parameters. When compared to experimental data, the predictions made using this model were accurate in the uniaxial simulations and mixed in the combined tension-torsion cases. The model did capture the major modes of torque-twist and axial strain-twist, but the magnitudes were only correct at certain temperatures and in certain strain ranges (see Chapter 10.6).

The Hartl et al. (2009) model assumes a total strain composed of an elastic component (with a stiffness that is phase fraction-dependent), a thermal component, a transformation component, and a plastic component. The formulation is of the plasticity-type, where the direction of the inelastic strain accumulation is governed by the stress deviator tensor (essentially the direction of maximum shear). The evolution equations use loading vs. unloading logic, a maximum achievable strain parameter that is stress dependent, as well as back-stress and transformation hardening terms. The start-stop criteria are based on driving forces which are also related to the

various terms from the evolution equations. This model was accurate in the *RR* uniaxial test cases but only showed results for a very limited set of multiaxial simulations (see Chapter 10.6).

Sedlak et al. (2009, 2012) present a model with material anisotropy, an R-phase, martensite reorientation, along with multi-axial loading. The total strain is segregated into elastic and inelastic parts, where the inelastic strain is based on a transformation strain tensor and a martensite phase fraction. A Helmholtz free energy is introduced with a mechanical and chemical component as well as penalty-type energy terms that are added to enforce system constraints. These penalty terms limit the inelastic strain to the deviatoric components and keep the phase fractions between zero and one. Dissipation functions are introduced as alternatives to yield functions and these govern the onset of stress-induced transformations, temperature-induced transformations, as well as martensite reorientation. The reorientation term has a temperature-dependent component which allows temperature dependent hysteresis. A set of thermodynamic forces along with minimization routines for the dissipation functions play the role of flow rules and dictate the evolution of the inelastic strains. This model showed good agreement to the experimental data in both the uniaxial tests as well as the tension-torsion tests (see Chapter 10.6).

The model by Bodaghi et al. (2013) uses self-accommodated martensite, oriented martensite, as well as preferred-direction martensite internal variables to describe the phase composition of the material. The strain is decomposed into elastic and inelastic components and the strain domain is made to be two dimensional, only considering a normal and a shear (torsional) component. A HFE function is written with elastic energy, internal energy, entropy, and transformation hardening terms. This is followed by a total dissipation inequality of the Claussius-Duhem type, which gets incorporated into evolution equations and optimization procedures. The Evolution equations and optimization conditions (Kuhn-Tucker-type) are derived from the free energy and dissipation expressions. These function like flow rules in traditional plasticity theory and then a set of limit functions are added that behave like yield surfaces. The results from this model were, like many other entries, accurate in the uniaxial cases but lacking in the tension-torsion simulations. The major modes of behavior were loosely represented, but the magnitudes and detailed shapes of the response curves were off in several cases (see Chapter 10.6).

In Auricchio et al. (2009, 2014), three internal variables are used to represent SMA phases: austenite, single-variant martensite, and multivariant martensite. A Helmholtz free energy is written as a function of total strain, temperature, phase fractions, and a transformation direction tensor. A chemical free energy is also written for each phase along with a set of interaction energies, which influence the behavior when multiple phases are present. Partial derivatives of the free energy are used to derive constitutive equations and thermodynamic forces. The driving force expressions also have logical terms based on Kuhn-Tucker conditions and the transformation direction is assumed to be the direction of the deviatoric stress vector. The driving forces, along with consistency parameters, limit functions, and optimization (Kuhn-Tucker) conditions, are used to derive a set of evolution equations for the internal variables. This model was accurate across the entire set of uniaxial cases including full and partial-loading tension tests and well as stress and strain-controlled thermal cycling tests. The tension-torsion cases exhibited mixed agreement to the experiments, but only a small sub-set of the data was shown (see Chapter 10.6).

In Chatziathanasiou et al. (2016), the authors present a plasticity type of model that considers forward transformation, reverse transformation, and reorientation. The transformation strain parameter represents the

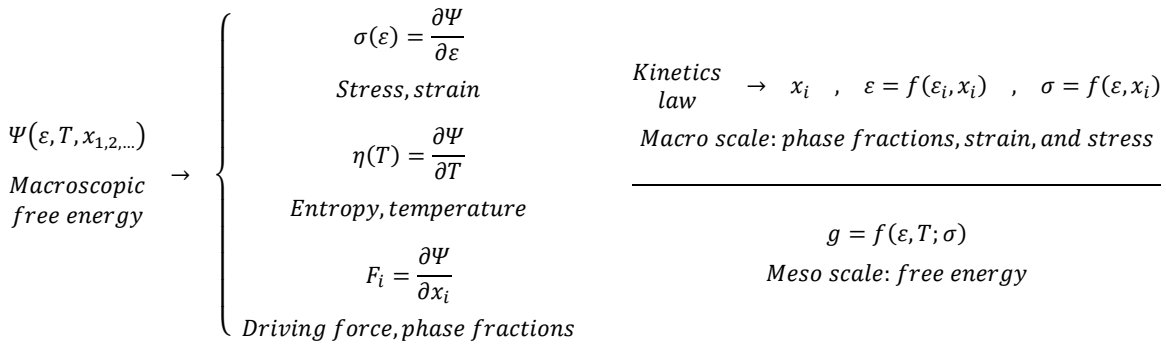
average strain in a martensite representative volume element (RVE), where the strain direction can change (reorient) and the phase fraction can grow or shrink. A rate of change for each internal variable is derived from a Gibbs free energy and a set of yield surfaces is used to define the activation criteria. A Claussius-Duhem expression is written to add some additional constraints to the system and maximum transformation strain parameters are used to allow transformations to complete. This model was only compared to a small set of uniaxial cases and no multiaxial tests.

2 SMA Model by Müller, Achenbach, and Seelecke

At this point, nearly 30 years after its inception, the 1-D SMA model by Müller, Achenbach, and Seelecke (which we will call the M-A-S model) has numerous variations, adaptations, and extensions. In this chapter we aim to merely introduce the model and give an overview of its fundamental features and operation. This includes a brief history, some example applications, a description of its multiscale motivation, and the basic equations required to operate the model.

The M-A-S model has overlap with both the micromechanical and phase field modeling techniques discussed in the previous chapter. The core model is derived from a nonconvex free energy, but the free energy represents a smaller (mesoscopic) scale than the rest of the model. This is a unique feature among SMA theories. The mesoscopic phase fields are not resolved but, instead, statistical physics techniques are used to motivate the macroscopic behavior. This type of model was originally motivated by the van der Waals gas equations (Müller (1979)) or Ginzburg-Landau theory for phase transformations (Falk (1980), Falk (1983)). The early works employed macroscopic potential energy landscapes, high levels of thermal activation, and various descriptions of the energy barriers. The discussions were mainly qualitative but showed the straightforward manner in which a multiwell potential can produce the signature stress plateaus and hysteresis exhibited by SMA material. The theory was furthered by Seelecke (2002) and Seelecke and Müller (2004) where a mesoscopic, temperature-dependent free energy was introduced. When applied to the low thermal activation case of an SMA solid, this model was able to accurately capture both stress and temperature-induced transformation as well as latent heat effects. Since then, the M-A-S model has been employed in a variety of SMA applications including: extension for polycrystallinity (Heintze and Seelecke (2008), Rizzello et al (2019)), thermal boundary condition effects (Furst et al. (2012), macroscopic field effects (Schmidt et al. (2016)), as well as simulations of an SMA cooling effects (Ullrich et al. (2014), Welsch et al. (2018)). Due to the generality of its thermal activation and mesoscopic free energy framework, the M-A-S model has also been adapted to represent other ferroic materials. Ferroelectric and magnetic SMA implementations of the M-A-S model can be found in Kim and Seelecke (2007), Krevet et al. (2008). Morrison et al. (2008a); Morrison et al. (2008b); Seelecke et al. (2005), Smith et al. (2003), and Smith et al. (2006). These formulations assign different physical meanings to the phases, but the use of nonconvex free energy and thermally activated kinetics are analogous to the original M-A-S model.

The M-A-S model deviates from traditional continuum thermodynamics models in two main ways. The first difference is related to the practice of deriving all constitutive equations directly from the free energy. Figure 1a illustrates a set of model equations structured this way.



a) Common continuum thermodynamics equation structure b) M-A-S model equation structure

Figure 1: Comparison of SMA equation structures

In the M-A-S framework, the free energy is addressed at a different scale than the constitutive equations and this is sketched in Figure 1b. The continuum thermodynamics rules are maintained at the mesoscopic scale, but a macroscopic free energy is simply not necessary. This allows the main stress and strain equations to take simple mixture rule forms and essentially leaves all the nonlinear effects to be governed by the kinetics law and mesoscopic system. The second deviation from the traditional continuum thermodynamics guidelines is the fact that that the M-A-S free energy is non-convex. Strict adherence to the arrangement in Figure 1a, with a non-convex free energy, would yield regions of zero or negative stiffness. In a more general view of the system, however, the free energy zones where $\frac{\partial \psi}{\partial \varepsilon}$ yields a negative number do not represent the stiffness of any material. These regions signify dynamic stages where the system is not governed by equilibrium conditions. Imagine the force vs. deflection curve for a buckling column or even a light switch. A free energy with multiple wells simply reflects a system with more than one stable configuration (metastable). SMA materials are just one of many examples of metastability and, in the M-AS model, the dynamic event is a phase change. The convex zones are governed by stress-strain equilibrium and the concave zones are governed by phase equilibrium; the two act together to maintain overall equilibrium.

Many of the models mentioned in Chapter 1 are arranged in similar manner to Figure 1a which means the free energies are macroscopic and convex. The consequence is that additional features like load-unload logic, phase diagrams, or additional potentials must be added to produce SMA effects such as forward and reverse transformation, stress and temperature-driven transformation, hysteresis, and latent heat. A multiwell free energy eliminates the need for logical constructions because it automatically contains the information necessary to yield SMA behavior. Segregating the free energy to the mesoscopic scale, as we will discuss in later chapters, is actually a convenient way to maintain thermodynamic consistency without neglecting its important multiwell nature.

Apart from these two distinctions, the M-A-S model has many of the common elements found in SMA constitutive models. An energy well is associated with each phase and the location of the well minimum designates a specific deformation state for that phase. In this way, the well locations play the role of the transformation strain vectors that are assigned to each variant in micromechanics models. Energy barriers separate each well, but the barriers grow and shrink as stress and temperature conditions distort the energy landscape. When a barrier disappears, transformation occurs because material jumps between wells. In this way, the relative heights of the energy barriers supply the activation (start/stop) conditions for the transformation kinetics. Start-stop conditions take many forms but are one of the expected building blocks in any SMA modeling circle. SMA models (both micromechanics-based and phenomenological) often use phase diagrams to govern transformation conditions; the

multi-well free energy is an analogous technique. When viewed on the strain-temperature plane, the stable (low energy) and unstable (high energy) zones essentially create the regions of a phase diagram. The multi-well free energy also has several parallels to plasticity and optimization based SMA models. The stable energy wells are elastic regions and each well has a boundary with the surrounding unstable zones. This boundary acts much like a yield surface – when stress pushes a layer beyond the boundary, inelastic strain (transformation) occurs. The paths of the layers into another, more favorable, well are analogous to flow rules which dictate the direction of plasticity-type transformation strains. The free energy drives the system toward a minimal Gibbs free energy (global equilibrium) because layers always choose a lower energy well if it becomes available. This rule is essentially an optimization scheme seeking a minimum condition on a multi-well landscape.

Now, after having some context for the M-A-S model's philosophical place among SMA modeling strategies, we shift the focus to the model itself. We begin by introducing the martensite internal variables and how they are incorporated into the overall system.

2.1 Multiscale Motivation

The M-A-S model is motivated by physics and three different scales of observation which are labeled in Figure 2 as *Micro*, *Meso*, and *Macro*.

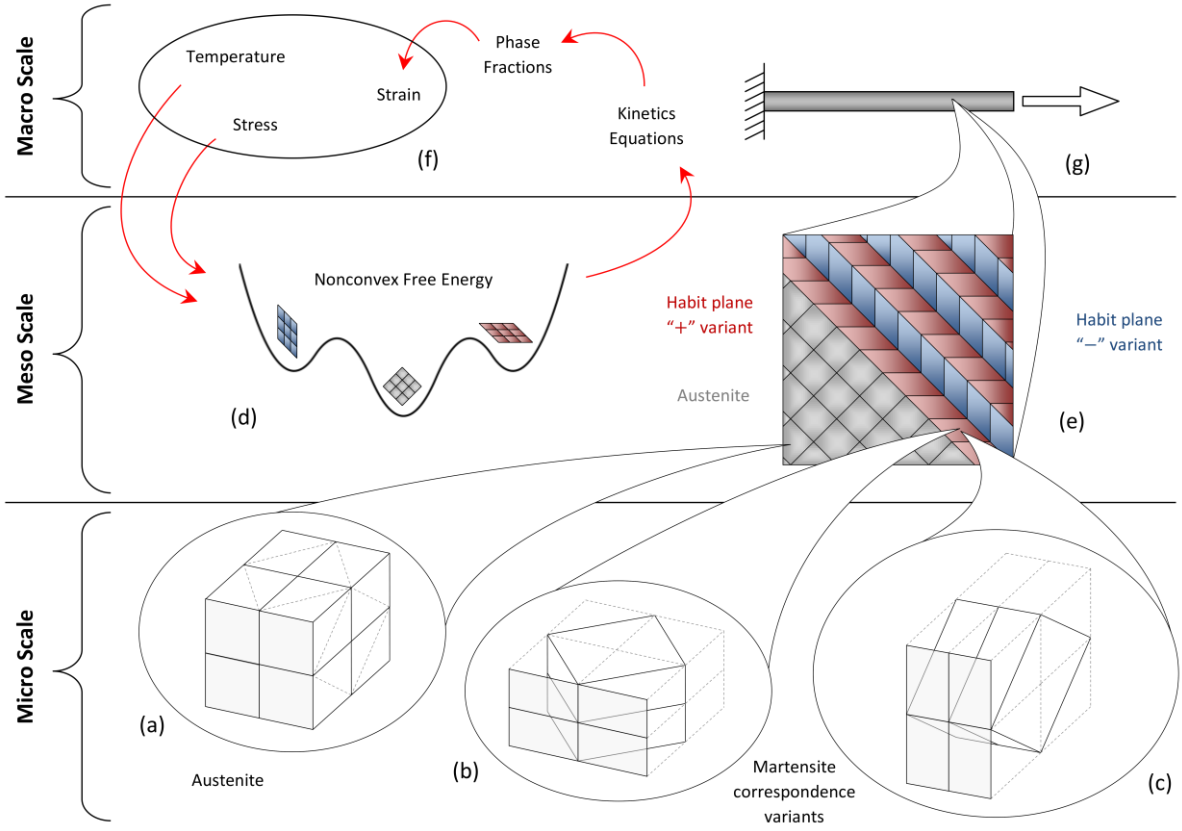


Figure 2: Micro, meso, and macro scale sketch

The model considers three phases: a parent austenite phase along with two martensite variants ($+$ and $-$). The phases are distinguished by lattice distortions at a μm scale of measurement, which we refer to as mesoscopic. The mesoscopic lattice distortions are assumed to be lattice invariant shears associated with habit plane-type martensite variants. These represent a higher level of organization than correspondence variants, which are shown at the microscopic scale and are not directly considered. After this point, we drop the correspondence vs. habit

plane distinction and refer to the mesoscopic habit plane variants as simply *variants* or *phases*. The variants in Figure 2 represent positive and negative shear deformations with respect to the same habit plane, which is illustrated by the line separating the austenite region from the martensite. The alternating pattern of $\pm$ variants represents self-accommodation, where martensite patches appear in such a way that the macroscopic material is not distorted. When more variants are considered, these self-accommodating patches have a variety of shapes and a good overview can be found in Ziolkowski (2015). The strips of material in Figure 2e (sometimes called lathes, plates, or layers) represent the fundamental material unit at the mesoscopic scale and we use the term *layer*. Each layer represents a small uniform zone of mesoscopic material that can exhibit one of the three phases. The arrangement of layers shown in the figure is for conceptual purposes only, however. The model does not require the knowledge or prediction of phase fields. Instead, we treat the mesoscopic system in a phenomenological manner and assume a large population of layers is governed by a multi-well free energy.

The free energy in Figure 2d has one energy well for each phase and, in this sketch, each well has an equal number of layers. This distribution represents the state of the mesoscopic phase field to the right where the amount of material in each phase is roughly equal. It should be noted that it is typically not the case that the parent phase and all of the martensite variants are present in equal proportion. The occurrence of each phase depends on temperature and stress and this is reflected in the free energy. The multi-well landscape distorts as stress and temperature conditions change and this dictates which phase populations grow or shrink. The horizontal axis of the free energy (sometimes called the order parameter) represents the deformation associated with each phase. The mesoscopic deformation is projected onto the axis of a 1-D macroscopic body (Figure 2g) where the "−" variant is associated with compressive strain and the "+" variant is associated with tensile strain.

The off-axis components of the deformation are neglected in the traditional M-A-S model and it is therefore restricted to 1-D applications. This is appropriate for most SMA devices which involve thin, wire-like bodies that are restricted to uniaxial modes of deformation. There are situations where SMA components experience more complicated stress-strain states and two of these are addressed in later chapters; Chapter 5 deals with SMAs in bending and Chapter 6 addresses SMAs in combined tension and torsion. The first four chapters though, are dedicated to examination and development of the 1-D model in the context of uniaxial applications.

The 1-D body in Figure 2g is thin and wire-like and has the scale of a typical tensile test specimen measured in mm or cm. The main, macroscopic model describes the point-wise stress-strain-temperature behavior of its material. A point-wise material model does not consider field variations and this is an important distinction for SMA materials. Even in wire-like bodies, macroscopic field variations play a major role in their behavior. This aspect of the system is neglected here, however, to keep the focus on the fundamental material law. Field effects can be addressed by implementing the fundamental material law within a larger field solution (Furst et al. (2012)) or including additional gradient-based free energy terms (Welsch et al. (2018)). This modeling choice is discussed further in Chapter 5 and 6 where models are compared to several sets of experimental data.

The red arrows in Figure 2f represent the mathematical flow of information in the model and show that, except for the free energy, all of the equations are macroscopic. Following the arrows in the counterclockwise direction, the macroscopic stress and temperature first influence the mesoscopic free energy. The free energy is used by the macroscopic system to calculate average strain for each phase and to drive the kinetic equations. The kinetic equations govern the phase fractions which are used, along with the phase strains, to calculate the total strain. This chain of equations relates the stress, strain, and temperature and the results is an SMA material model

that includes both stress and temperature-driven transformations, thermal and mechanical hysteresis, as well as thermomechanical coupling. The segregation of the multiwell free energy to the mesoscopic system allows a convenient way to keep the hysteretic and the caloric information that is inherently embedded in the landscape, yet avoid the complexities of defining stress-strain relationships in the negative slope regions. This topic is discussed in greater detail in Chapter 4.

2.2 Visualization of the Multiscale Operation

Here we leverage the visualizable nature of the multiwell free energy model. Before presenting the first equation, we can show how layers jumping between energy wells can produce most recognizable signature SMA effects.

2.2.1 Superelastic Behavior

Figure 3 illustrates how the macroscopic and mesoscopic systems act together to produce a typical quasi-static superelastic response. At the center of the figure is a macroscopic stress-strain curve for a superelastic load-unload cycle. At selected points during the cycle (states 1-6), snapshots of the mesoscopic system are also included.

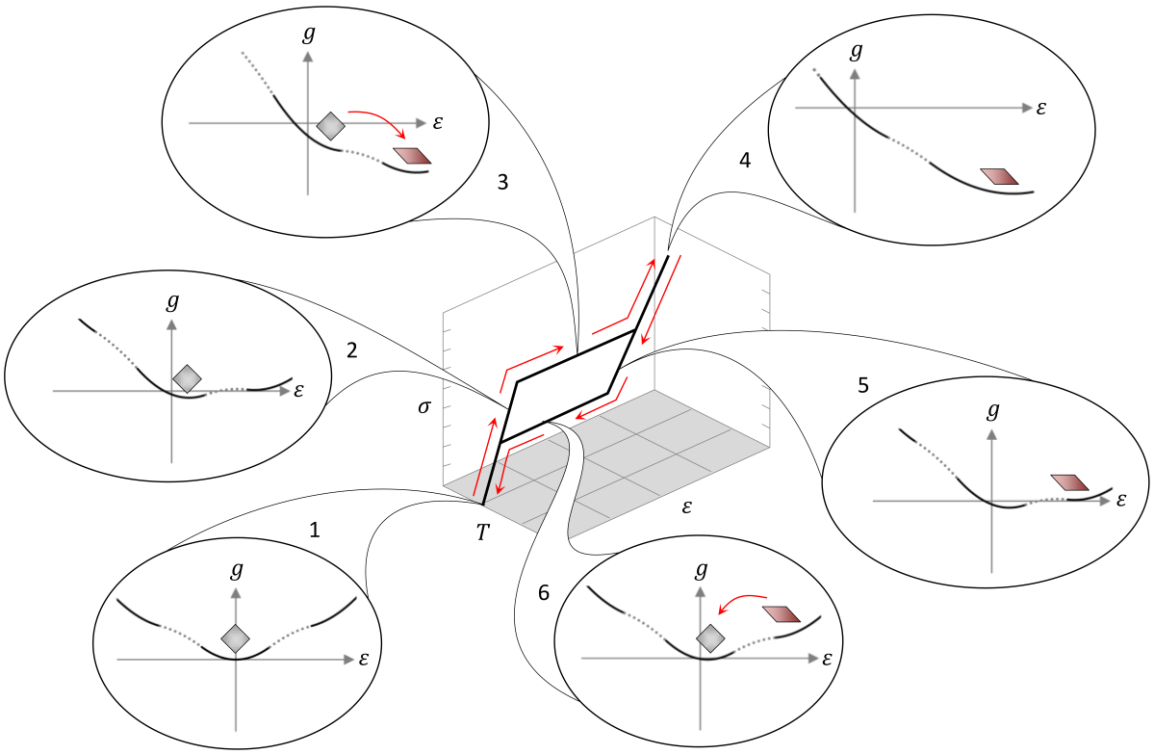


Figure 3: Multi-scale sketch showing super-elastic behavior

State 1 represents the initial, zero stress state. The only stable phase is austenite and all of the layers are in that energy well. State 2 shows a point on the tensile loading branch prior to transformation. The GFE landscape is being tilted to the right due to the tensile stress and the layers shift to the right following the well minimum. This increases the average strain but does not cause transformation. State 3 illustrates the instant where the transformation from austenite to tensile martensite occurs. In a stress-controlled situation this transformation happens rapidly whereas, in a strain-controlled experiment, it progresses to accommodate the applied strain; in either case the transformation mechanism is the same. When the austenite energy well loses convexity, its layers transform by jumping into the martensite well. This population shift causes the large change in macroscopic strain

with no change in stress. At state 4, the transformation has been completed and additional stress has been applied. The layers now follow the martensite well minimum and the macroscopic behavior is once again linear. State 5 represents an instant during unloading where the stress is lower than the stress required for forward transformation (state 3). In state 5 the austenite well has become stable once again and signifies a lower energy state, but transformation does not initiate because a barrier still remains between the wells. The stress must be reduced to a lower level (state 6) for the reverse transformation to initiate and this leaves a macroscopic hysteresis loop. The reverse transformation returns the macroscopic stress-strain path to the austenite branch and complete unloading returns both systems (meso and macro) to state 1.

2.2.2 Shape Memory Behavior

Figure 4 shows a macroscopic stress-strain-temperature path for a demonstration of shape memory behavior. The initial state (1) represents zero stress and a low temperature where there is no convex austenite well. In this situation, layers equally occupy the – and + wells (self-accommodation). In state 2 a tensile stress has been applied but not enough to initiate transformation, so the macroscopic behavior is linear. State 3 shows the instant that transformation into a single martensite variant has just completed. The increasing stress caused the martensite – well to lose convexity and its layers jumped to the much lower martensite + well. The macroscopic result is, once again, a large change in strain with no change in stress.

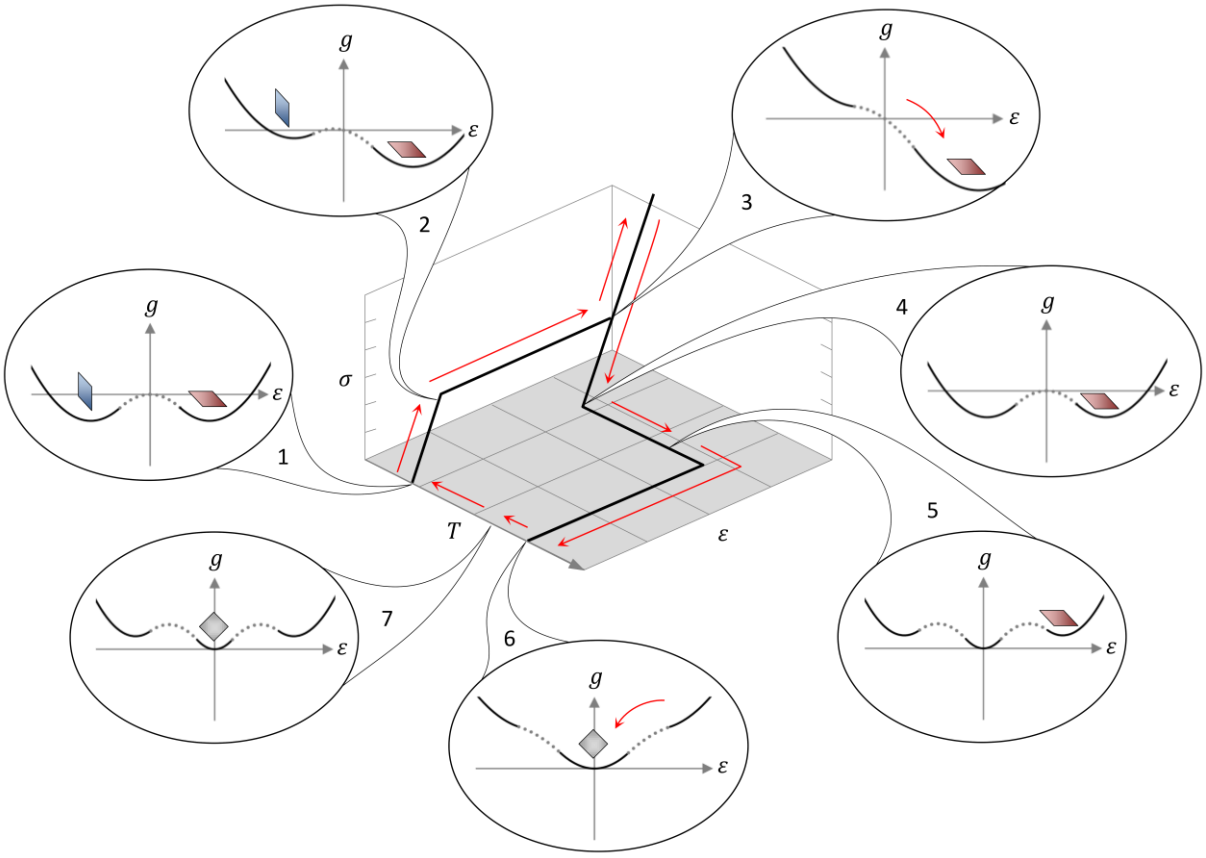


Figure 4: Multi-scale sketch showing shape memory behavior

Continued loading and unloading result in elastic behavior along the martensite branch because all layers remain in that phase. State 4 represents complete mechanical unloading. Unlike the superelastic case (Figure 6), there is no austenite well available so the layers remain trapped in the martensite + well. It should be noted that states 4 and 1 represent the same stress and temperature but, due to the difference in their mesoscopic populations,

state 4 shows a residual strain whereas state 1 does not. At state 5, stress free heat is being applied and the austenite well has reappeared. Even though the austenite well represents a lower energy, it is not until state 6, when the martensite well loses convexity, that the layers transform back to the central well. At the macroscopic scale, the result is a stress-free return to the original state (shape memory). The heating element is removed and the system is cooling when state 7 is shown. The austenite well will continue to become less convex as the temperature decreases till its convex space is zero. At this point the layers are free to populate the martensite wells and they do so equally returning us to the initial configuration (state 1).

2.2.3 Actuation Behavior

Figure 5 shows an actuation test-type of temperature-induced transformation cycle, where the specimen undergoes temperature cycles while under a constant tensile load. Here, the black line represents the stress-strain-temperature path of the material and the green arrows show the sequence. There are also two sets of lines, one in blue and one in red, that indicate what the superelastic stress-strain path would be at a low and high temperature respectively.

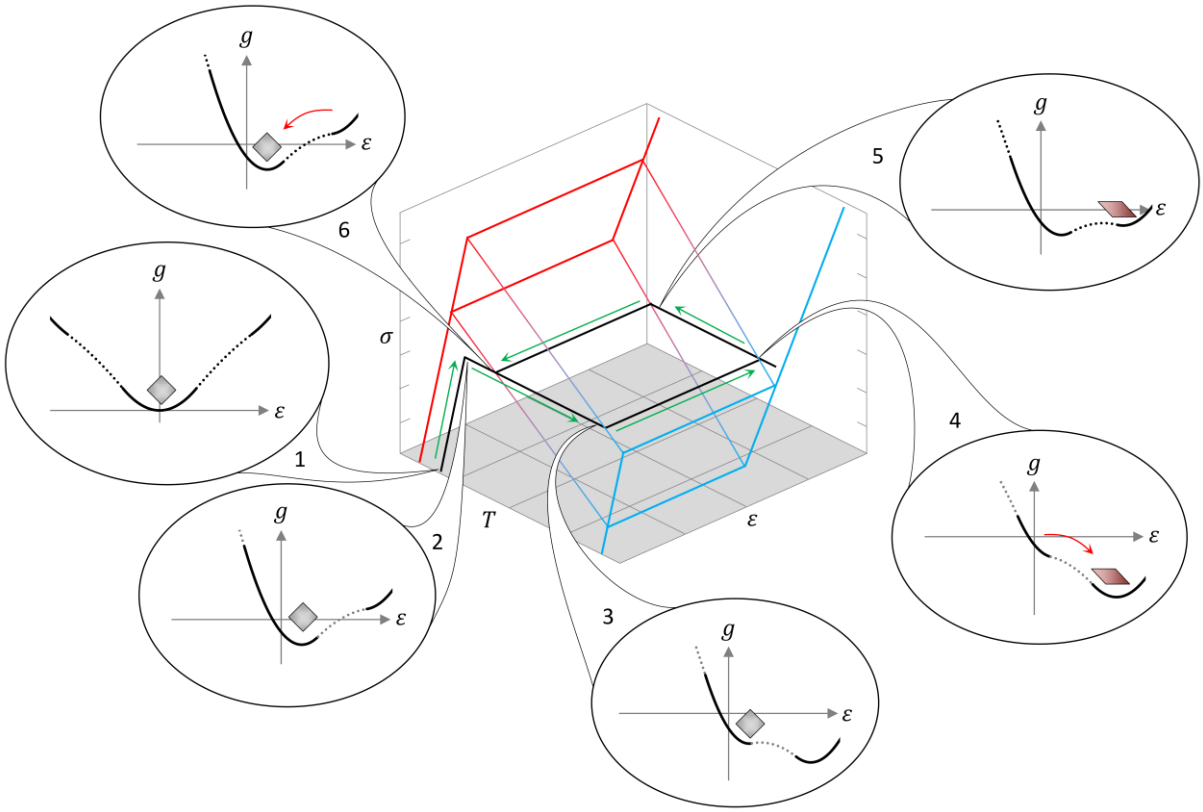


Figure 5: Multi-scale sketch showing actuation behavior

The two sets of lines are also connected by lines that transition from blue to red – this shows the evolution of the transformation stresses over the range of temperatures. These are for reference only and help to visualize the meaning of each transition point in the actuation cycle. In state 1 the material is at a superelastic temperature and only the austenite well is stable. At state 2, an axial load has been applied, but the stress is below either the forward or reverse transformation stress at that temperature. The stress is held constant and the temperature is decreased between states 2 and 3. As the temperature decreases the forward transformation does as well and state 3 shows the point where it is equal to the applied stress. The tensile martensite well becomes available and all of the layers transform, resulting in a large increase in strain with no change in stress. In state 4, the layers are once

again in mechanical equilibrium with the externally applied stress and remain at the same strain as the temperature decreases further, then increases again toward state 5. State 5 shows the moment just before another transformation initiates. The austenite well lowers at the temperature increases and the martensite layers have only a small barrier retaining them. The temperature increases a bit more and the reverse transformation stress becomes equal to the applies stress, the barrier disappears, and the layers jump back to the austenite well (state 6). The layers are once again in mechanical equilibrium and will rest unless the stress or temperature is changed. If the temperature cycles continue the actuation cycles will do the same as material perpetuates the sequence: 6, 3, 4, 5, 6, etc.

2.2.4 Strain Rate Dependence

Figure 6 shows a mechanical load-unload sequence with self -heating. Like Figure 5, the black likes and green arrows show the stress-strain-temperature path of the material while the blue, red, and grey lines are for reference.

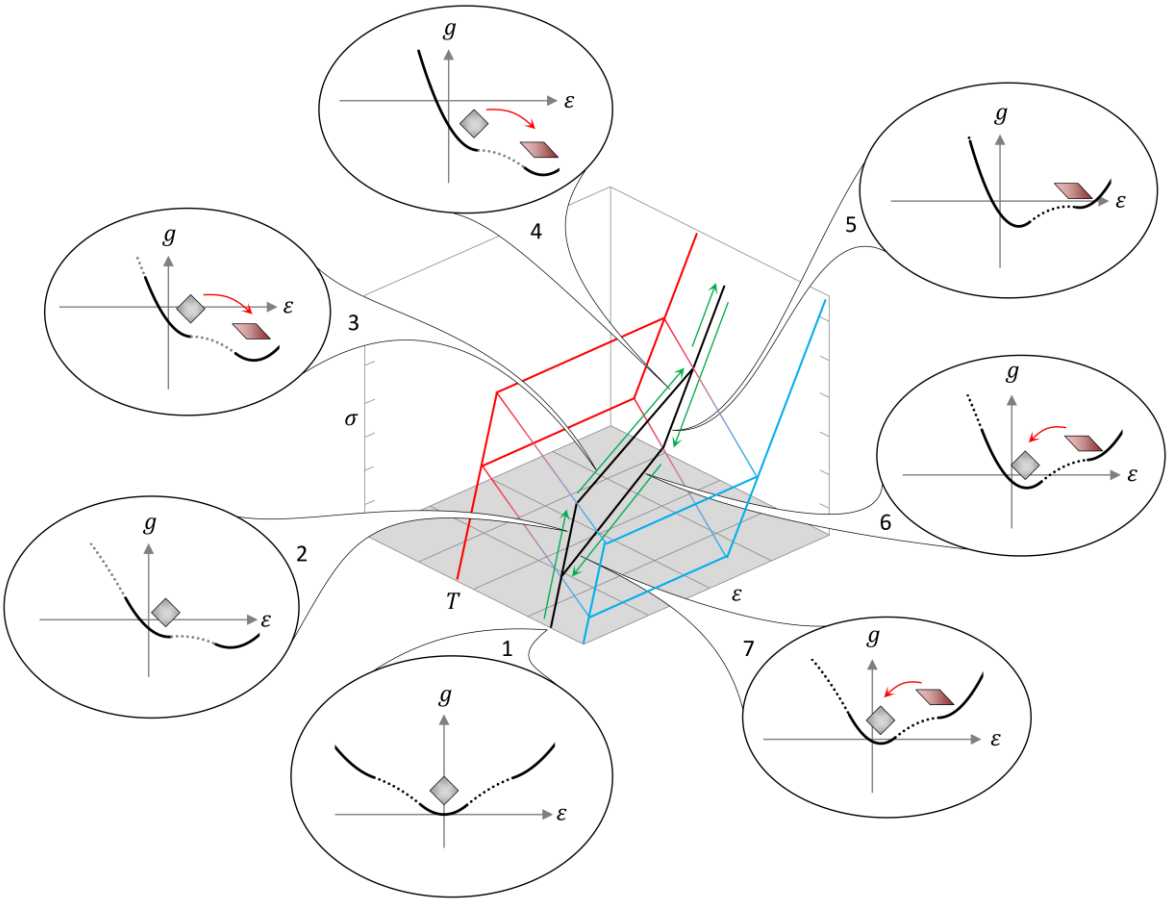


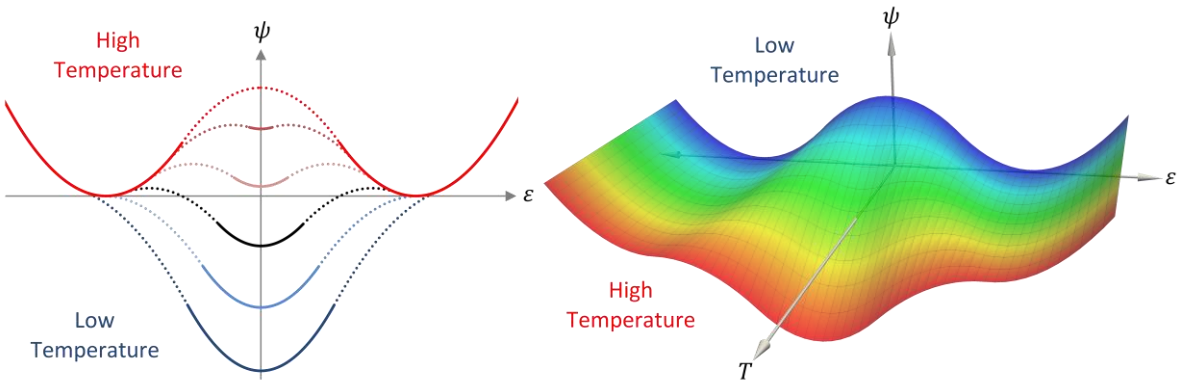
Figure 6: Multi-scale sketch showing transformation hardening and softening

The blue set of lines traces a quasistatic load-unload path at a low (but still superelastic) temperature while the red lines represent a higher temperature. The four lines that connect the corner points of the blue and red hysteresis loops show the continuous evolution of the hysteresis loop between the low and high temperatures. The two hysteresis loops and the four connecting lines should be visualized as forming a 3D box-like shape. The initial state (1) shows a free energy under zero stress where only the austenite well is stable. At state 2, a tensile stress has been applied, but the stress is less than the forward transformation stress and the layers remain trapped. After state 2, the stress continues to be increased and forward transformation begins. The forward transformation path (which progresses through states 3 and 4) represents a large change in strain, but also increases in stress and

temperature. The latent heat from the forward transformation increases the temperature of the specimen and this means that the stress must be continually increased to keep the transformation going. The mesoscopic free energies in states 3 and 4 look similar in that they each show both austenite and martensite layers, but the martensite well in state 4 is lower than state 3 because the stress in higher. At a point just beyond state 4, the transformation completes and the heating stops. The specimen behaves linearly through further loading and unloading. State 5 shows the point on the martensite unloading branch where the stress has almost reached the reverse transformation value; the austenite well is much lower but there is still a slight barrier locking the layers in the martensite phase. A small amount of additional unloading initiates the reverse transformation but, once again, the latent heat effects change the path; this time cooling the specimen. As the specimen cools, the stress must be continually lowered to progress the transformation as shown in states 6 and 7. The phase fractions change but, until the transformation is complete, there are layers in both the austenite and martensite wells. Once the martensite well is finally emptied, the self-cooling stops and the stress-strain-temperature path returns to the original branch and state 1.

2.3 Helmholtz Free Energy

The mesoscopic free energy governs the behavior of the layers that were shown in the previous series of figures jumping between energy wells and this dictates the hysteretic, nonlinear behavior of the material. The Helmholtz free energy is the foundation for the stress-strain laws, temperature-dependence, and caloric properties in the M-A-S model. The HFE is metastable, having two energy wells at high temperatures and three at low temperatures as shown in Figure 7.



(a) HFE vs. strain at various temperatures (b) Helmholtz free energy vs. temperature and strain

Figure 7: Helmholtz free energy over a range of temperatures

The energy landscape is constructed from a chain of parabolas and this is highlighted in Figure 7(a), where the solid lines represent the stable (convex) parabolas for each phase and the dotted lines represent the unstable (concave) energy barrier parabolas. Mathematically, the HFE is defined by the piecewise function ψ in Equation 1.

$$\psi(\varepsilon,T)=\begin{cases}\psi_-(\varepsilon) & -\infty<\varepsilon\le\varepsilon_{-a}(T) \\ \psi_-^B(\varepsilon,T) & \varepsilon_{-a}(T)<\varepsilon<\varepsilon_{a-}(T) \\ \psi_a(\varepsilon,T) & \varepsilon_{a-}(T)\le\varepsilon\le\varepsilon_{a+}(T) \\ \psi_+^B(\varepsilon,T) & \varepsilon_{a+}(T)<\varepsilon<\varepsilon_{+a}(T) \\ \psi_+(\varepsilon) & \varepsilon_{+a}(T)\le\varepsilon<\infty\end{cases} \quad where \quad \begin{aligned} &\psi_-(\varepsilon)=\frac{E_m}{2}(\varepsilon+\varepsilon_R)^2 \\ &\psi_-^B(\varepsilon,T)=A(T)\varepsilon^2-B(T)\varepsilon+C(T) \\ &\psi_a(\varepsilon,T)=\frac{E_a}{2}\varepsilon^2+\beta(T) \\ &\psi_+^B(\varepsilon,T)=A(T)\varepsilon^2+B(T)\varepsilon+C(T) \\ &\psi_+(\varepsilon)=\frac{E_m}{2}(\varepsilon-\varepsilon_R)^2\end{aligned} \tag{1}$$

Since we have a mesoscopic system and a macroscopic system, it is important to clarify first the meaning of the strain ε, stress σ, and temperature T that define the HFE. At the macroscopic scale, σ and ε are the engineering strain and stress in the system and are given by Equations 26 and 27. In the context of a mesoscopic equation, they represent the average strain and stress for all layers in a given phase. Since we deal only with the average mesoscopic values, we are relieved from dealing with the complex, local variations that are known to exist in mesoscopic phase fields. The temperature T is assumed to be experienced uniformly at all scales, but can also be considered an average parameter which neglects local variations.

The other parameters in Equation 1 can be identified by their subscripts and/or superscripts, which follow the convention shown below.

$\#_{-} \rightarrow$ *Quantity associated with the − phase or energy well*
$\#_{a} \rightarrow$ *Quantity associated with the a phase or energy well*
$\#_{+} \rightarrow$ *Quantity associated with the + phase or energy well*
$\#_{ij} \rightarrow$ *Quantity associated with the transformation from phase* $i \rightarrow j$ *(not tensor indices)*
$\#_{m} \rightarrow$ *Quantity associated both both or either martensite variant*
$\#_{-}^{B} \rightarrow$ *Quantity associated with the energy barrier on the − side of the a well*
$\#_{+}^{B} \rightarrow$ *Quantity associated with the energy barrier on the + side of the a well*
Super or subscripts not shown here have meaning specific to the individual parameter

Three of the parameters needed to build Equation 1 (E_a, E_m, and ε_R) are measured directly from tension test data. The remaining parameters (A, B, C, and β) along with the domain boundaries for each function (ε_{-a}, ε_{a-}, ε_{a+}, and ε_{+a}) are derived from the transformation stress, which is also estimated from tension test data. The complete set of material parameters required to construct the HFE (i.e. characterize the SMA material) is given in Table 1.

Parameter	Definition
E_a	Elastic modulus while in the austenite phase
E_m	Elastic modulus while in the martensite phase
ε_R	Residual strain (strain remaining after load-unload cycle while in the martensite temperature range)
T_H	The higher of the two test data temperatures
T_L	The lower of the two test data temperatures
σ_{a+}^{H}	Forward tensile transformation stress at T_H
σ_{+a}^{H}	Reverse tensile transformation stress at T_H
σ_{a+}^{L}	Forward tensile transformation stress at T_L
σ_{Δ}	Hysteresis width (Difference between forward and reverse transformation stress at T_H)

Table 1: Experimentally measured parameters required for the Helmholtz free energy

Additional detail is given about how to measure the material parameters in Chapter 4 where we also detail the construction of the HFE. The cornerstone quantity for deriving the other parameters is the forward transformation stress, which is a function of temperature given by Equation 2.

$$\sigma_{a+}(T) = max\left(\sigma_{a+}^{L} + \left(\frac{\sigma_{a+}^{H} - \sigma_{a+}^{L}}{T_H - T_L} \right)(T - T_L), 0 \right) \tag{2}$$

Equation 2 describes the forward transformation stress as a linear function of temperature, which is acceptably accurate and consistent with other SMA models. It is important to note, however, that it is actually an approximation for a slightly more complex function that is derived directly from the free energy in Chapter 4.

Once the forward transformation stress is known from Equation 2, the reverse transformation stress can be determined using the hysteresis width parameter.

$$\sigma_{+a}(T) = \sigma_{a+}(T) - \sigma_{\Delta} \qquad where \qquad \sigma_{\Delta} = \sigma_{a+}^{H} - \sigma_{+a}^{H} \tag{3}$$

Equation 3 is based on the assumption that the hysteresis width does not vary with temperature. This is true for most SMA materials but there are exceptions. This is discussed further in Chapter 5, where the model is extended to allow for temperature-dependent hysteresis. The forward and reverse transformation stresses allow definition of the critical strains $\varepsilon_{ij}(T)$, which designate the boundaries (inflection points) between the energy wells.

$$\varepsilon_{a+}(T) = \frac{\sigma_{a+}(T)}{E_a} \qquad , \qquad \varepsilon_{+a}(T) = \frac{\sigma_{+a}(T)}{E_m} + \varepsilon_R \tag{4,5}$$

These expressions are based on Equations 16-18, which more generally relate the stress experienced by the material to the average strain of the layers in each phase. The traditional M-A-S model assumes tension-compression symmetry, which means that the compressive stress-strain curve is identical to the tensile curve, except in the $(-\sigma, -\varepsilon)$ quadrant. Symmetry of the HFE allows the compressive critical strains and stresses to be written in terms of their analogous values in tension.

$$\varepsilon_{a-}(T) = -\varepsilon_{a+}(T) \quad , \quad \varepsilon_{-a}(T) = -\varepsilon_{+a}(T) \tag{6,7}$$

$$\sigma_{a-}(T) = -\sigma_{a+}(T) \quad , \quad \sigma_{-a}(T) = -\sigma_{+a}(T) \tag{8,9}$$

This assumption is irrelevant for tension-only applications and it is appropriate for 1-D applications where twinning needs to be captured, but there are situations where asymmetry is observed. This addressed in Chapter 5 where the model is extended to include tension-compression asymmetry (in addition to temperature-dependent hysteresis.

The chemical free energy (β) and the energy barrier coefficients (A, B, and C) can be expressed in terms of either the transformation stresses or strains but using a combination of the two yields more concise equations.

$$A(T) = \frac{\sigma_{\Delta}}{2\big(\varepsilon_{a+}(T) - \varepsilon_{+a}(T)\big)} \qquad B(T) = \frac{\sigma_{+a}(T)\varepsilon_{a+}(T) - \sigma_{a+}(T)\varepsilon_{+a}(T)}{\varepsilon_{a+}(T) - \varepsilon_{+a}(T)} \tag{10,11}$$

$$C(T) = \beta(T) + \frac{\varepsilon_{a+}(T)B(T)}{2} \qquad \beta(T) = -\frac{\sigma_{a+}(T)\varepsilon_{+a}(T) + \sigma_{+a}(T)\big(\varepsilon_R - \varepsilon_{a+}(T)\big)}{2} \tag{12,13}$$

Equations 10-13 are derived from free energy continuity assumptions in Chapter 4. Equation 13 is the final supporting expression required to build the HFE (i.e. fully define Equation 10). The HFE governs the stress-strain behavior and caloric properties of the material, but the Gibbs free energy motivates the kinetic (i.e. phase change) behavior.

2.4 Gibbs Free Energy

The GFE is defined by Legendre transformation of the HFE; $g \equiv \psi - \sigma\varepsilon$. The critical strains ε_{ij} also separate the individual GFE functions and the complete expression is given by Equation 14.

$$g(\varepsilon,T;\sigma)=\begin{cases} g_-(\varepsilon,T;\sigma) & -\infty<\varepsilon\le\varepsilon_{-a}(T) \\ g_-^B(\varepsilon,T;\sigma) & \varepsilon_{-a}(T)<\varepsilon<\varepsilon_{a-}(T) \\ g_a(\varepsilon,T;\sigma) & \varepsilon_{a-}(T)\le\varepsilon\le\varepsilon_{a+}(T) \\ g_+^B(\varepsilon,T;\sigma) & \varepsilon_{a+}(T)<\varepsilon<\varepsilon_{+a}(T) \\ g_+(\varepsilon,T;\sigma) & \varepsilon_{+a}(T)\le\varepsilon<\infty \end{cases} \quad where \quad \begin{aligned} g_-(\varepsilon,T;\sigma)&=\psi_-(\varepsilon)-\sigma\varepsilon \\ g_-^B(\varepsilon,T;\sigma)&=\psi_-^B(\varepsilon,T)-\sigma\varepsilon \\ g_a(\varepsilon,T;\sigma)&=\psi_a(\varepsilon,T)-\sigma\varepsilon \\ g_+^B(\varepsilon,T;\sigma)&=\psi_+^B(\varepsilon,T)-\sigma\varepsilon \\ g_+(\varepsilon,T;\sigma)&=\psi_+(\varepsilon)-\sigma\varepsilon \end{aligned} \tag{14}$$

In Equation 14 we use the notation $g(\varepsilon,T;\sigma)$ for GFE. This implies that strain and temperature establish the HFE and the stress, separated by a semicolon, is distinguished as an additional parameter that distorts the landscape into the GFE. Under equilibrium assumptions, GFE is often written as a function of temperature and stress alone because strain and stress are explicitly related. This is also true for the current model, when we refer to a specific location on the energy landscape. But, to define the full landscape, both stress and strain are independently required. Figure 8 shows how the GFE landscape is distorted by various combinations of stress and temperature (recall the warping free energies in Figures 3-6).

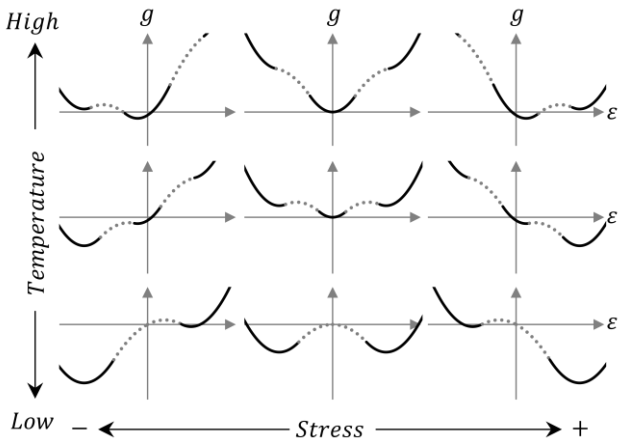


Figure 8: Gibbs free energy at various temperatures and stresses

Resolution of the full GFE landscape is not required for model operation. Instead, the GFE function only needs to be evaluated at selected locations which are shown in Figure 9.

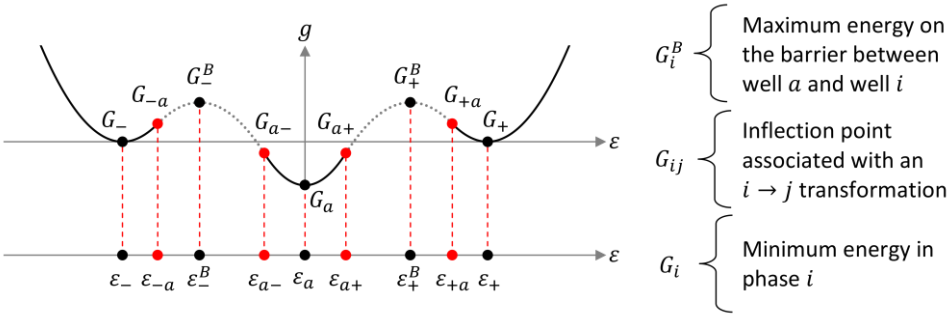


Figure 9: Locations on the GFE-strain landscape

The free energies G_{ij} are found by evaluating Equation 14 at the critical strains ε_{ij} (given by Equations 4-5).

$$G_{ij}(\sigma,T)=g_i\big(\varepsilon_{ij}(T),T;\sigma\big) \tag{15}$$

The extremum strains ε_i are given by Equations 16-18; these represent both the average strain for all of the layers in each phase i and algebraic minimum of each parabola g_i. Moving forward, we refer to these as the *average* strains or the *phase* strains.

$$\varepsilon_-(\sigma) = \frac{\sigma}{E_m} - \varepsilon_R \quad , \quad \varepsilon_a(\sigma) = \frac{\sigma}{E_a} \quad , \quad \varepsilon_+(\sigma) = \frac{\sigma}{E_m} + \varepsilon_R \tag{16,17,18}$$

The statistical mechanics basis for these expressions are detailed further in Chapter 3. The minimum energy within each phase (G_i) is then determined by evaluating Equation 14 at each strain ε_i.

$$G_-(\sigma,T) = g_-\big(min\big(\varepsilon_-(\sigma), \varepsilon_{-a}(T)\big), T; \sigma\big) \tag{19}$$

$$G_a(\sigma,T) = g_a(\varepsilon^*, T; \sigma) \qquad where \qquad \varepsilon^* = max\left(\varepsilon_{a-}(T), min\big(\varepsilon_a(\sigma), \varepsilon_{a+}(T)\big)\right) \tag{20}$$

$$G_+(\sigma,T) = g_+\big(max\big(\varepsilon_{+a}(T), \varepsilon_+(\sigma)\big), T; \sigma\big) \tag{21}$$

The min and max conditions in Equations 19-21 stem from the piecewise nature of the GFE. The algebraically determined minimum or maximum of a given energy well may fall outside of the admissible strain range for that well. The conditional nature of these expressions restricts the strains to the proper ranges.

The energy barrier peaks are found by maximizing barrier functions ($g_{\mp}^B$) within their respective domains. The extremum strains ε_-^B and ε_+^B are given by Equations 22-23.

$$\frac{\partial g_-^B(\varepsilon,\sigma)}{\partial \varepsilon} = 0 \quad \rightarrow \quad \begin{matrix} solve \\ for\ \varepsilon \end{matrix} \quad \rightarrow \quad \varepsilon_-^B(\sigma,T) = max\left(\varepsilon_{-a}(T), min\left(\frac{\sigma + B(T)}{2A(T)}, \varepsilon_{a-}(T)\right)\right) \tag{22}$$

$$\frac{\partial g_+^B(\varepsilon,\sigma)}{\partial \varepsilon} = 0 \quad \rightarrow \quad \begin{matrix} solve \\ for\ \varepsilon \end{matrix} \quad \rightarrow \quad \varepsilon_+^B(\sigma,T) = max\left(\varepsilon_{a+}(T), min\left(\frac{\sigma - B(T)}{2A(T)}, \varepsilon_{+a}(T)\right)\right) \tag{23}$$

This allows the maximum energy associated with each barrier to be evaluated via the main GFE function (Equation 14).

$$G_-^B(\sigma,T) = g(\varepsilon_-^B(\sigma,T), T; \sigma) \quad , \quad G_+^B(\sigma,T) = g(\varepsilon_+^B(\sigma,T), T; \sigma) \tag{24,25}$$

Once again, min and max conditions are necessary to maintain the admissible strain ranges for each GFE function. Now we shift the focus to the main, macroscopic model.

2.5 Macroscopic Model

The macroscopic model assumes a continuum mixture of phases with no consideration of phase fields, habit plane orientations, or phase boundaries. The mesoscopic system contains a large population of layers governed by the Helmholtz and Gibbs free energies but, at the macroscopic scale, it is sufficient to only know the fraction of material that is in each phase.

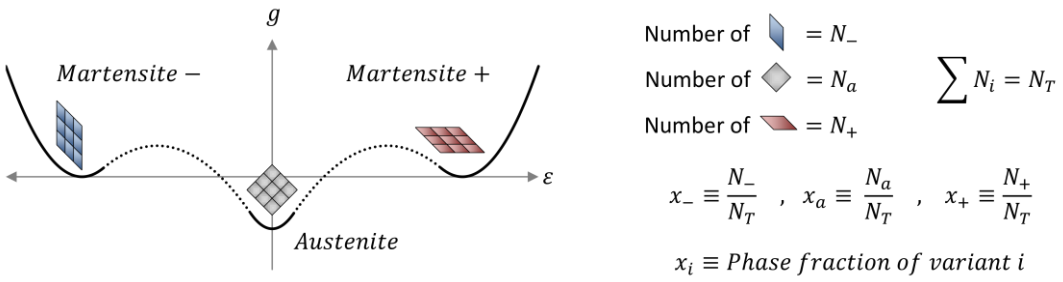


Figure 10: Groups of layers in each phase and phase fractions

The phase fractions, along with the average strain in each phase, are used to define total macroscopic strain in Equation 26.

$$\varepsilon = \varepsilon_- x_- + \varepsilon_a x_a + \varepsilon_+ x_+ \qquad where \tag{26}$$

$\varepsilon \rightarrow Total\ macroscopic\ strain$

$\varepsilon_i \rightarrow Average\ strain\ experienced\ by\ the\ layers\ in\ the\ i\ phase$

$x_i \rightarrow Phase\ (volume)\ fraction\ of\ layers\ in\ the\ i\ phase$

$i \rightarrow \in -, a, +\ (compressive\ martensite, austenite, tensile\ martensite)$

The average strains are calculated using Equations 16-18, which are functions stress. Equation 26 can also be used to write an expression relating stress, total strain, and the phase fractions by substituting 16-18 for the corresponding ε_i and then solving for stress.

$$\varepsilon = \left(\frac{\sigma}{E_m} - \varepsilon_R\right) x_- + \left(\frac{\sigma}{E_a}\right) x_a + \left(\frac{\sigma}{E_m} + \varepsilon_R\right) x_+ \quad \rightarrow \ solve\ for\ \sigma \ \rightarrow \quad \sigma(\varepsilon) = \frac{\varepsilon + \varepsilon_R (x_- - x_+)}{\frac{x_-}{E_m} + \frac{x_a}{E_a} + \frac{x_+}{E_m}} \tag{27}$$

Equation 27 is the main stress-strain equation for the model and is not derived from a free energy potential. The complexities of the free energy and the thermodynamic requirements are handled at the mesoscopic scale and this allows a simple rule of mixtures-type constitutive equation for the main model.

The phase fractions are governed by a kinetics law, which is given by Equations 28-30. This set of ordinary differential equations (ODE's) dictates the time rates of change of the phase fractions.

$$\dot{x}_- = x_a\, p_{a-} - x_-\, p_{-a} \qquad \dot{x}_+ = x_a\, p_{a+} - x_+\, p_{+a} \qquad x_a = 1 - x_- - x_+ \tag{28,29,30}$$

The p_{ij} parameters are based on thermal activation theory and the probability that a mesoscopic layer will transform from phase i to phase j. We refer to these as the transformation probabilities and they are calculated using Equation 31, which is based on the height of the GFE barriers.

$$p_{ij} = \frac{1}{\tau} e^{-\frac{1}{\gamma}\Delta G_{ij}} \qquad where \qquad \gamma = \frac{K_B T}{v} \qquad and \qquad \begin{aligned} \Delta G_{-a} &= G_-^B(\sigma, T) - G_-(\sigma, T) \\ \Delta G_{a-} &= G_-^B(\sigma, T) - G_a(\sigma, T) \\ \Delta G_{a+} &= G_+^B(\sigma, T) - G_a(\sigma, T) \\ \Delta G_{+a} &= G_+^B(\sigma, T) - G_+(\sigma, T) \end{aligned} \tag{31}$$

The energy values at the well minima and the tops of the barriers are given by Equations 19-21, 24, and 25. In the kinetic ODE's, the product $x_i \cdot p_{ij}$ yields the rate at which each transformation occurs. K_B is the Boltzmann constant and the parameters τ and v have physical interpretations from thermal activation theory (which is detailed in Chapter 3). For model usage, they are constants chosen to ensure that the transformation rates are ≈ instantaneous compared to the mechanical strain rates. Equation 31 can take several different forms, each offering benefits in certain scenarios. These different versions and their relationship to the original theory are the focus of Chapter 3, where a new formulation is also presented.

The thermomechanical coupling observed in an SMA material is attributed to two factors: the temperature-dependence of the critical transformation stress and the latent heat generated/absorbed during phase transformation. An energy balance equation must be added to the set of kinetic ODE's to fully account for thermal effects. Equation 32 considers heat loss to the environment due to convection and heat generated or absorbed during phase transformation.

$$\dot{T} = \frac{1}{\rho c}\left(\dot{x}_+ h_+ + \dot{x}_- h_- - \alpha h_c (T - T_0)\right) \qquad where \tag{32}$$

$\dot{T}$ → *Time derivative of temperature*
ρ → *Density*
c → *Specific heat*
α → *Surface are per unit volume complete a* $\rightarrow +$ *transformation*
h_c → *Convective heat transfer coefficient*
T_0 → *Ambient temperature*
h_- → *Latent heat per unit volume due to a complete a* $\rightarrow -$ *transformation*
h_+ → *Latent heat per unit volume due to a complete a* $\rightarrow +$ *transformation*

The latent heats are determined directly from the mesoscopic free energy, using Equation 33.

$$h_{\pm}(\sigma) = \frac{\sigma^2}{2} E_{\Delta} \pm \sigma \varepsilon_R + u_R \qquad where \qquad E_{\Delta} = \frac{1}{E_m} - \frac{1}{E_a} \tag{33}$$

The latent heat coefficients are functions of stress and mechanical material parameters which means that caloric measurements, such as those obtain from differential scanning calorimetry (DSC), are not required. This feature is explored in more detail in Chapter 4. The parameters α and h_c reflect details of system geometry (thin, wire-like bodies) and environment and are not specifically related to the material model. There are no temperature gradient (conduction) terms since development in this work is restricted uniform field problems.

The equations given in this section are sufficient to operate the basic M-A-S SMA model, which underpins the work in this dissertation. In the next two chapters, we take a closer look at the underlying theory and also begin to present new developments for the model. The two chapters are essentially broken up by which free energy quantity they focus on – the Gibbs free energy or the Helmholtz free energy. To understand which free energy influences which part of the model, it is helpful to think of operating the model as answering three questions:

1) *How much material is in each phase?* → *Kinetics model (Gibbs free energy)* → *Phase fractions are driven by stress and temperature*
2) *What is the strain of the material in each phase?* → *Constitutive equations (Helmholtz free energy)* → *Phase strains are driven by stress and temperature*
3) *How is temperature coupled to the mechanics?* → *Caloric properties and stress − dependence (Helmholtz free energy)* → *Latent heats are dictated by the HFE's multiwell nature and are influened by stress*

The next chapter is dedicated to question 1 as it explores the kinetics equations in several different forms leading to the development of a new version. This is followed by the *Thermodynamics* chapter which deals with questions 2 and 3, namely the stress-strain relationships and the thermal behavior.

3 Kinetics

The transformation probability equations are the main drivers for the kinetics model because they provide the start/stop conditions for phase transformation. In this chapter we review the way several versions of these equations were derived and this exercise leads us to a new version (published in Ballew and Seelecke (2019)). We begin with an introduction to the Boltzmann equation and thermal activation theory, then use these concepts to derive the most general form of the kinetics model. This formulation is applicable to a wide variety of thermal activation levels (i.e. gases, liquids, or solids) but is mathematically complex. Next, we show how the following different forms of the kinetics equations have been derived from the general form and applied to the M-A-S model:

- The *General* model → Uses low thermal activation to simply the original equations
- The *Energy Barrier* model → Low thermal activation limit to remove the probability integrals
- The *Driving Force* approximation → Uses the driving force to approximate the energy barriers
- The *Inflection Point* model → New formulation requiring fewer parameters and expressions

We conclude the chapter by briefly discussing mesoscopic inhomogeneity and presenting a detailed examination of the thermal activation parameters. Inhomogeneity, or polycrystallinity, is a commonly observed SMA effect and is accounted for in the M-A-S model through the kinetics equations. While not a primary focus point for this dissertation, it had to be included in order to model some of the validation experiments in Chapter 5. The thermal activation parameters are effectively transformation rate scaling variables; necessary to ensure that the transformation strain rates far exceed the strain rates encountered in macroscopic system. Different kinetics models show different types of sensitivity to the thermal activation parameters, so we present two different examples showing how to identify correct settings.

3.1 Introduction to Thermal Activation

Thermal activity refers to random atomic fluctuations that have kinetic energy proportional to temperature. The effect of these fluctuations depends on the phase and free energy characteristics of the material. In a solid material the atoms fluctuate with low energy compared to their inter-atomic attraction and are therefore locked into a fixed lattice arrangement. In a gas, the fluctuations easily overcome the attraction to neighboring material which leads to their amorphous consistency. When the opportunity for phase change, diffusion, chemical reaction, etc. is available, thermal activity can be related to their onset and progression. These fluctuations can be thought of as the sampling of alternate energy states in the vicinity of a current one. When a group of atoms is near the threshold between states, say two phases, a fraction of them will sample configurations beyond the threshold and thus change phase. In this type of process, both the amplitude of fluctuation and the height of the energy barrier affect the rate of transformation.

Typically, the number of atoms (or fraction of material) in a certain state is only quantified in a statistical sense. The Boltzmann equation relates the probability of finding material in a certain energy state to total number of possible energy states. For a continuous distribution of available energies, Equation 34 gives the probability f of finding material within an infinitesimal range of energy $U + dU$.

$$f = \frac{e^{-\frac{U+dU}{KT}}}{\int_{-\infty}^{\infty} e^{-\frac{U}{KT}} dU} \qquad where \qquad \begin{array}{l} U \rightarrow Energy\ (thermodynamic\ configuration) \\ K \rightarrow The\ Boltzmann\ contant \\ T \rightarrow Temperature \end{array} \tag{34}$$

The integral in the denominator represents the total energy of all possible states and the Boltzmann constant universally relates the entropy of a system to the total number of possible thermodynamic states. This equation can be used to predict the kinetics behavior of a process by examining the probability of finding particles at a threshold between states. This results in a rate equation similar to the Arrhenius equation used in chemical reaction and diffusion modeling. In the current model we use Gibbs free energy (GFE) to describe the landscape of thermodynamic states and in a 1-D stress-strain system. We begin here with several illustrations to introduce the essential concepts.

3.1.1 Single Phase Material

The upper row of plots in Figure 11 shows GFE (g) and the lower row shows probability density (f), both plotted against strain. The green dots in each GFE well represent particles that undergo fluctuations in strain and energy. Each f curve shows that particles are more likely to be found at lower energies.

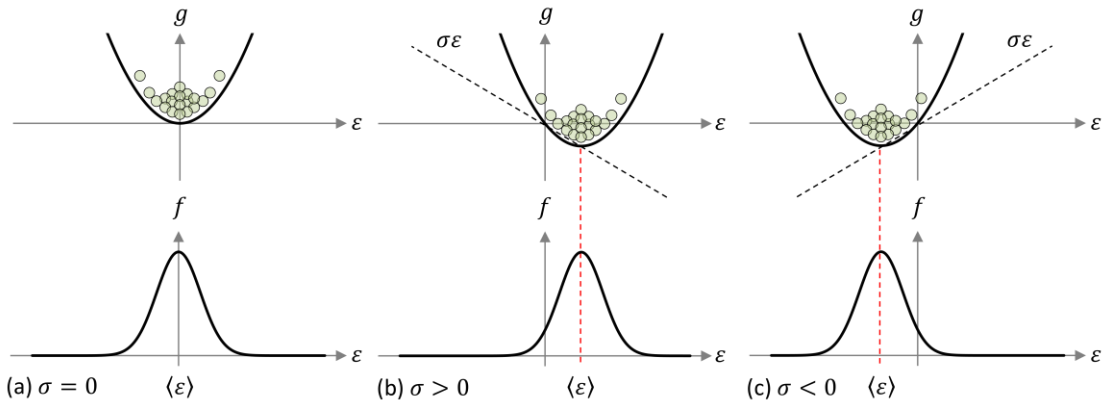


Figure 11: Gibbs free energy and probability distribution in a thermally active single-phase material

Figure 11(a) shows the system under zero stress, 11(b) represents a tensile stress, and 11(c) a compressive stress. The dashed lines are to visualize the influence of stress which is responsible for the right (b) or left (c) tilt in the GFE. Since the particles always cluster at the minimum energy, the mean strain ($\langle\varepsilon\rangle$) shifts to the right (b) or left (c) as the GFE distorts. For a given GFE function $g(\varepsilon, T; \sigma)$, the probability density can be written using the Boltzmann Equation.

$$f = \frac{e^{-\frac{1}{\gamma}g(\varepsilon,T;\sigma)}}{\int_{-\infty}^{\infty} e^{-\frac{1}{\gamma}g(\varepsilon,T;\sigma)}d\varepsilon} \quad where \quad \gamma = \frac{KT}{v} \quad and \quad \begin{matrix} \sigma \rightarrow Stress \\ v \rightarrow Activation\ volume \end{matrix} \tag{35}$$

It should also be noted that g, in Equation 35, is independently a function of both stress and strain. This is due to the fact that a given value of stress is not explicitly related to a single value of strain (as seen in Figure 11). The notation $(\varepsilon, T; \sigma)$ implies that ε and T specify a Helmholtz free energy value, but σ is an additional parameter that distorts the landscape and yields the GFE.

The width of each bell curve represents the level of thermal activity because large fluctuations (high thermal activity) would correspond to a large standard deviation in particle locations. Low thermal activation would be signified by small fluctuations and therefore a narrow distribution of particles. This property is dictated by the parameter γ which is derived from v, K, and T. The activation volume v sets the representative length scale of the system and this is necessary because the fluctuations are measured as strain. Imagine a representative volume of material containing only a few atoms; in this case the length scale would be small and the fluctuations would correspond to large strains. On the other hand, if the volume of material considered is much larger, the atomic fluctuations are very small compared to the length and would represent near-zero strains. In Figure 12, we show the zero-stress state from Figure 11 with three different levels of thermal activity.

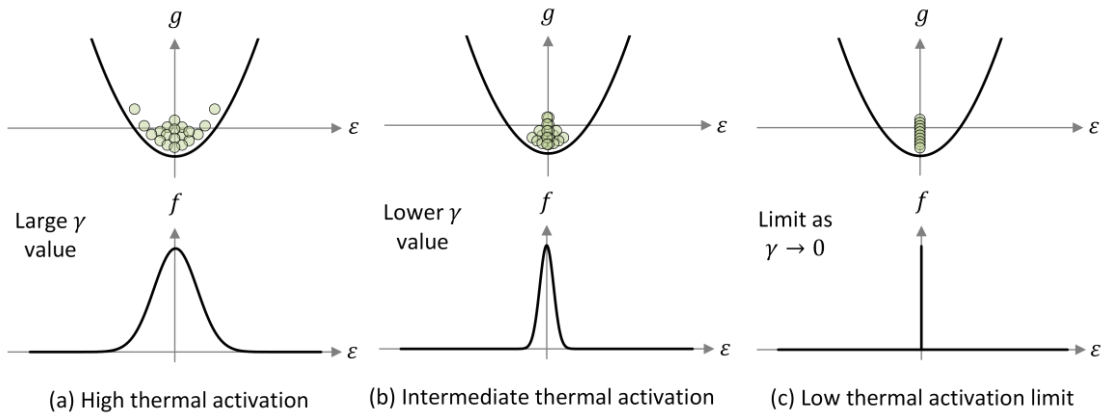


Figure 12: Gibbs free energy and probability density for different levels of thermal activation

All three plots in Figure 12 show the same GFE landscape and the same number of particles. The upper-most system has very high thermal activity shown by large fluctuations of the particles and a very wide Bell curve. In 12(b) the thermal activity is lower therefore the amplitude of the fluctuations has been reduced. 12(c) illustrates the limiting case where thermal activation is zero. In this case, all of the particles are concentrated at the energy well minimum and the probability density now has a Dirac delta shape. This represents a solid material where the atomic fluctuations are negligibly small and can be neglected. The probability density function given by Equation 35 is general and all three cases can be described by varying the parameter γ.

3.1.2 Multi-Phase Material

We now extend the concepts shown in Figure 11 to a system with multiple stable states and the opportunity for a process (switching between phases) to take place. Figure 13 shows a stress-strain system with two convex energy wells connected by a concave energy barrier.

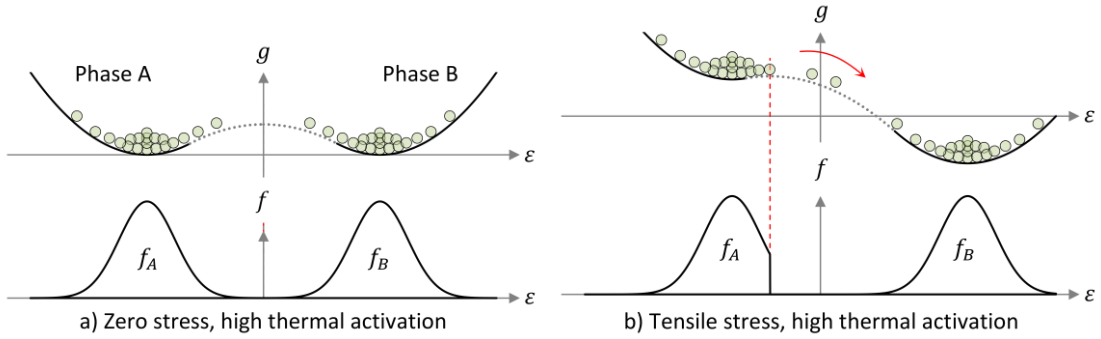


Figure 13: Gibbs free energy and phase change probability for a thermally active material

In 13a, the system is under zero stress and each of the energy wells represents an equally stable phase. Each phase contains its own population of particles that fluctuate about the corresponding local GFE minimum. This is reflected in the probability distribution plots where the peaks coincide with the g minima. Because they gravitate to the lowest energy state, any particles that fluctuate beyond the energy barrier will jump into the neighboring well (change phase). The phase change, however, is only seen in Figure 13(b) where the system is under a tensile stress. The GFE landscape has been tilted to the right and phase B represents a lower energy state. The strain associated with the barrier peak (ε_{cr}) has shifted to the left and this has decreased the convex range of strain associated with the A phase. Now, in phase A, the critical strain is much closer to the energy minimum and some fraction of the particles reach the threshold for transformation to phase B. The truncated probability densities

represent the fact that particles found beyond the barrier exit phase i and cannot be represented by f_i. The rates of transformation from $A \rightarrow B$ and $B \rightarrow A$ can be represented by evaluating the Boltzmann equation f_A and f_B, respectively, at the barrier strain.

$$rate_{AB} = \frac{1}{\tau} f_A(\varepsilon_{cr}) = \frac{1}{\tau} \frac{e^{-\frac{1}{\gamma} g(\varepsilon_{cr},T;\sigma)}}{\int_{-\infty}^{\varepsilon_{cr}} e^{-\frac{1}{\gamma} g(\varepsilon,T;\sigma)} d\varepsilon} \quad , \quad rate_{BA} = \frac{1}{\tau} f_B(\varepsilon_{cr}) = \frac{1}{\tau} \frac{e^{-\frac{1}{\gamma} g(\varepsilon_{cr},T;\sigma)}}{\int_{-\infty}^{\varepsilon_{cr}} e^{-\frac{1}{\gamma} g(\varepsilon,T;\sigma)} d\varepsilon} \tag{36,37}$$

In Equations 36 and 37, the terms $f_i(\varepsilon_{cr})$ represent the probability that a particle will have sufficient energy to exit phase i. τ is a sampling frequency which, mathematically, is a proportionality constant between $rate_{ij}$ and the probability $f_i(\varepsilon_{cr})$. In Figure 13, the particles transform from $A \rightarrow B$ because $f_A(\varepsilon_{cr})$ is now non-zero. On the other hand, the $B \rightarrow A$ transformation does not occur because the probability $f_B(\varepsilon_{cr})$ is extremely low.

Because of the effect shown in Figure 12, the transformation rate equations (36 and 37) can represent drastically different behaviors at different levels of thermal activation. In Figure 14 we show the same two-phase system as Figure 13 but with three different sets of probability density plots. The fluctuating particles are not shown here because the three different values of γ represent three different populations that could exist on the same GFE. The distribution of the particles is implied by the probability distributions as they were in Figures 11 and 12.

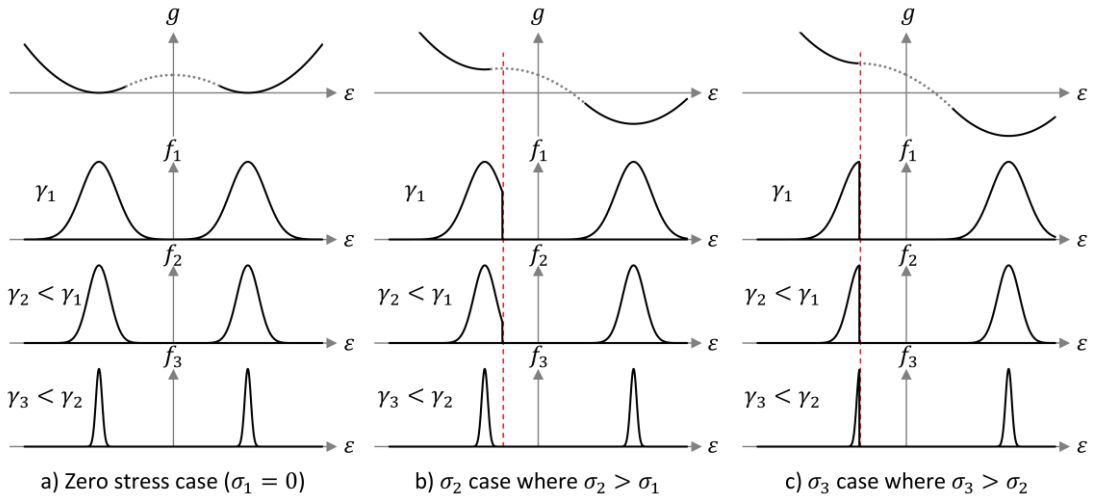


Figure 14: Transformation probability for different levels of stress and thermal activation

Each column in Figure 14 corresponds to the same GFE under a different level of tensile stress. The three rows of probability density plots reflect different levels of thermal activity. The basic mechanism in Equations 36 and 37 is that transformation begins when the non-zero part of the Bell curve collides with the critical strain and this is illustrated by Figure 14. In the γ_1 and γ_2 cases, the Bell curves have reached critical strain in frame (b) which means the transformations have begun. In the lowest thermal activation case (γ_3) however, the Bell curve has not yet reached the critical strain so the transformation has not initiated. In frame (c), all three Bell curves have reached the critical strain and all three cases represent active phase change. This is an important effect and allows this type of equation to reflect transformations in either high (gases) or low (solids) thermal activation materials. In the next section we begin to specify this framework to a solid SMA material with a description of the length scales considered as well as the physics and mathematical elements associated with each.

3.2 General Kinetics Model

The transformation kinetics equations (28-30) describe the rate of change of the phase fractions at the macroscopic scale. At the mesoscopic scale, transformation involves individual layers jumping from well to well, but the model does not require or track the behavior of any individual layer. Instead, we use the mesoscopic free energy and thermal activation theory to measure the probability that each possible transformation will occur. We begin by extending the two-well example from Figures 13 and 14 to the 1-D SMA system in Figure 15.

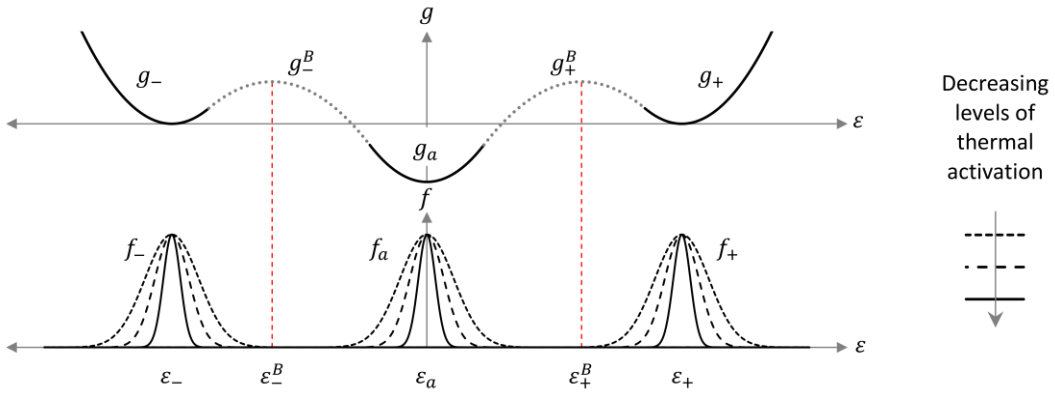


Figure 15: Probability distributions for a SMA stress-strain system

Each of the three energy wells g_-, g_a, and g_+ can have a population of randomly fluctuating layers that reflect each of the three phases $-$, a, and $+$. The lower set of plots shows a series of probability densities that correspond to each energy well and three different levels of thermal activation. Each function f_i represents the probability of finding a layer, within phase i, at a given strain and energy. According to the Boltzmann equation (in an analogous manner to Equation 35) the probability densities are given by Equations 38-40.

$$f_- = \frac{e^{-\frac{1}{\gamma}g(\varepsilon,T;\sigma)}}{\int_{-\infty}^{\varepsilon_-^B} e^{-\frac{1}{\gamma}g(\varepsilon,T;\sigma)}d\varepsilon} \qquad f_a = \frac{e^{-\frac{1}{\gamma}g(\varepsilon,T;\sigma)}}{\int_{\varepsilon_-^B}^{\varepsilon_+^B} e^{-\frac{1}{\gamma}g(\varepsilon,T;\sigma)}d\varepsilon} \qquad f_+ = \frac{e^{-\frac{1}{\gamma}g(\varepsilon,T;\sigma)}}{\int_{\varepsilon_+^B}^{\infty} e^{-\frac{1}{\gamma}g(\varepsilon,T;\sigma)}d\varepsilon} \tag{38-40}$$

The thermal activation parameter γ is defined in the same manner as in Equation 35 such that large and small values represent high and low thermal activation respectively. Equations 38-40 predict the probability of finding layers at a given strain but do not specifically predict transformation. In order to predict each rate of transformation, we examine the probability of finding a layer at each critical strain.

$$p_{a-}(\sigma) = \frac{1}{\tau} f_a(\varepsilon_-^B(\sigma,T),\sigma) \qquad p_{a+}(\sigma) = \frac{1}{\tau} f_a(\varepsilon_+^B(\sigma,T),\sigma) \tag{41,42}$$

$$p_{-a}(\sigma) = \frac{1}{\tau} f_-(\varepsilon_-^B(\sigma,T),\sigma) \qquad p_{+a}(\sigma) = \frac{1}{\tau} f_-(\varepsilon_+^B(\sigma,T),\sigma) \tag{43,44}$$

Equations 41 and 42 represent the rate of transformation from the a well into the $-$ and $+$ wells respectively whereas equations 43 and 44 represent the rate of transformation from the $-$ and $+$ wells into the a well. We do not consider transformation directly from the $-$ well to the $+$ well or vice versa. In the low temperature case where austenite is not stable, layers jump in one side then instantly out the other side of the a well. This constitutes martensite-to martensite transformation (sometimes called martensite reorientation). The parameter $1/\tau$ is another thermal activation parameter that, at the mesoscopic scale, represents the number of fluctuations per unit time. In this sense, the product of fluctuation rate and the probability that the motion will overcome the barrier yields the rate of layers actually jumping out of the well. At the macroscopic model scale, the parameter τ is a scaling parameter to assure that transformations occur rapidly compared to bulk material strain rates. The transformation

rates in Equations 41-44 rise and fall as the mesoscopic GFE distorts due to changes in stress and temperature. The rising and falling transformation rates then drive the kinetic equations, which are responsible for the macroscopic SMA behavior of the model.

Because the level of thermal activation affects the transformation probabilities (recall Figure 14), it also affects the macroscopic stress-strain behavior. We begin with the case of high thermal activation. Figure 16 shows the GFE, probability density, and stress-strain curve for an SMA material at three stages of tensile loading.

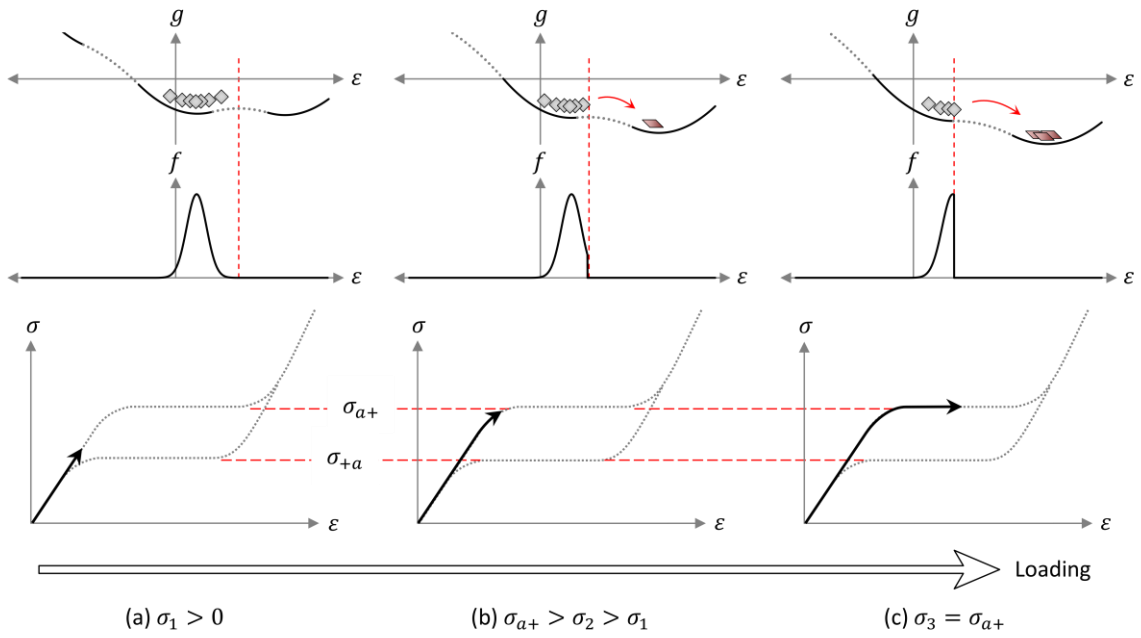


Figure 16: Transformation behavior for high thermal activation material, loading

In Figure 16(a), the tensile stress has caused the GFE to tilt to the right which lowers the martensite + well and raises the martensite − well. The austenite layers follow the energy well minimum which has shifted to the right of the zero-strain axis. The macroscopic stress-strain curve, which reflects the average strain of the layers, shows a tensile strain in response to the tensile stress. In Figure 16(b), the tensile stress has increased and the austenite well minimum has shifted further to the right. Also, the martensite + well is lower and the energy barrier peak has moved closer to the austenite well minimum. Now, even though the well minimum has not reached the critical strain, some of the fluctuating layers jump into the lower martensite + well. At the macroscopic scale, the onset of transformation results in a slight nonlinearity in the stress-strain curve even though the stress has not reached σ_{a+}. In Figure 16(c), the stress has reached the critical transformation value where the austenite well minimum and the barrier peak collide with one another. At this point, the highest probability corresponds to transformation and layers shift to the martensite + well at a much higher rate. The macroscopic result is a large change in the strain with very little change in stress (a stress plateau). Due to the wide distribution in layers however, the stress must be increased well beyond σ_{a+} to complete the transformation; this is implied by the dotted line tracing the path onto the fully martensite branch.

Figure 17 shows three stages (from right to left) during the unloading of the same system after fully transforming to the martensite + phase.

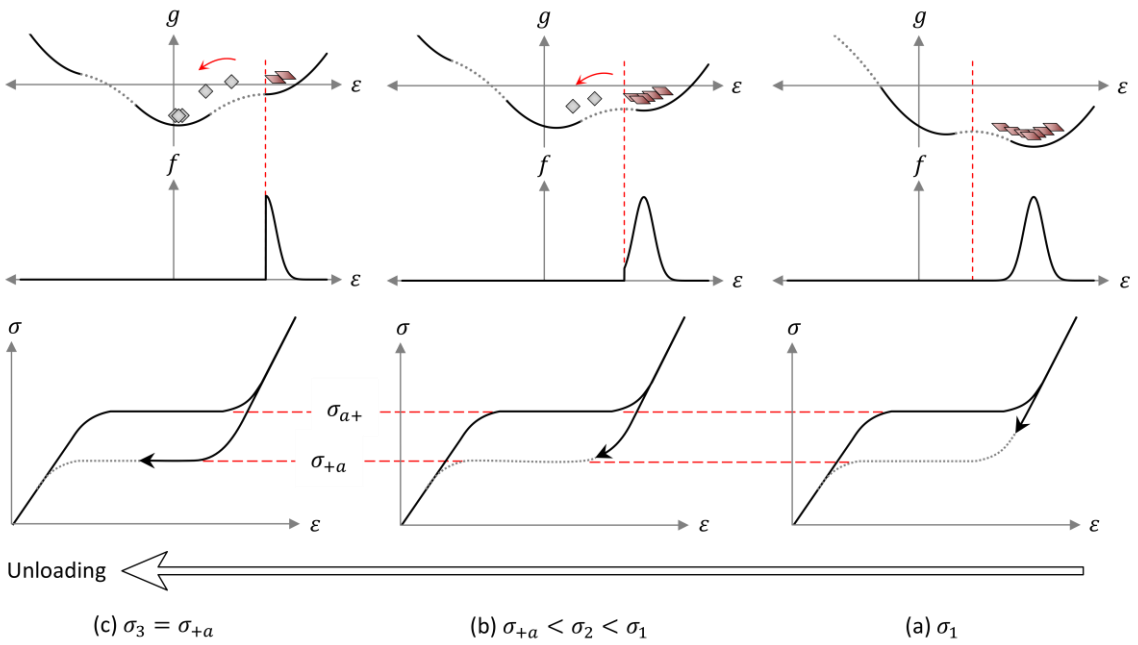


Figure 17: Transformation behavior for high thermal activation material, unloading

17(a) represents a state where the system has been unloaded to a stress value lower than the forward transformation stress σ_{a+}. The layers are still trapped in the martensite well because the barrier strain is still beyond even the highest fluctuations. In 17(b) the stress is much closer to the reverse transformation stress σ_{+a} and the critical strain is now much closer to the martensite well minimum. Some of the fluctuating layers are able to jump the barrier causing a nonlinearity in the macroscopic unloading curve. In 17(c) the critical stress has been reached and the probability is now high that layers will jump back into the austenite well. The macroscopic result is a rapid transformation back toward the linear austenite branch.

It should be noted here that the barrier strain reaches the austenite well at one stress level (Figure 16(c)) while it reaches the martensite well at a different stress level (Figure 17(c)). This is due to the curvature of the energy barriers and serves an important purpose in the model. It is by this mechanism that the macroscopic hysteretic behavior observed in SMA materials is embedded in the mesoscopic free energy function. This eliminates the need for load-unload logic or phase diagrams to produce a hysteretic response.

One additional distinction should be made here concerning the shape of the stress-strain curves shown in Figures 16 and 17. The smooth transitions into and out of the stress plateaus are not to be mistaken for a polycrystalline SMA response. The rounded nature of the superelastic stress-strain path is due to premature initiation and delayed completion of transformations. This is a rate dependent effect, even in the absence of latent heat effects, and represents inappropriate usage of the model. If the applied strain rate were varied the shape of the stress-strain path would be significantly different and this phenomenon is examined later in a set of simulations. A polycrystalline material has a smooth stress-strain curve due to field inhomogeneity, even at quasistatic strain rates – this effect is added later in the chapter.

Now we reproduce the previous two Figures in the low thermal activation limit case, i.e. for a solid SMA material. Figure 18 is analogous to Figure 16 and shows the loading branch of a superelastic SMA material.

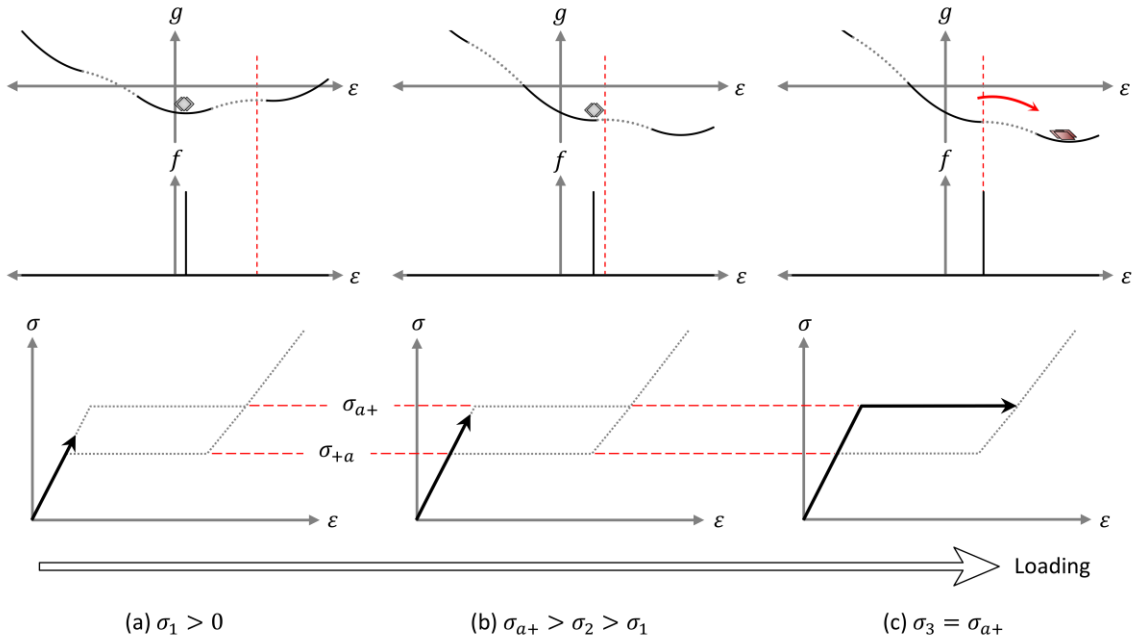


Figure 18: Transformation behavior for low thermal activation material, loading

In Figure 18(a), a tensile stress has been applied resulting in a GFE that is tilted to the right. The martensite + well is lower, the barrier strain has moved to the left, and the austenite well minimum has moved to the right. In 18(b), the stress has been increased to a value that is very close to the critical forward transformation stress σ_{a+} but no layers jump the barrier. They are tightly clustered at the well minimum and the macroscopic response remains linear up to a stress value equal to σ_{a+}. In 18c the stress has reached σ_{a+} and all of the layers suddenly jump to the neighboring well. The macroscopic stress-strain curve reflects this by a sudden and complete shift to the martensite branch (a perfectly flat stress plateau), which is representative of a single crystal SMA. The fundamental mechanism here is that transformation occurs when the well minimum coincides with a critical strain. The construction of the HFE ensures that this will happen at the critical stresses.

Figure 19 illustrates a sequence (right to left) of unloading steps after the system has reached a stress well level beyond that shown in Figure 18**(**c).

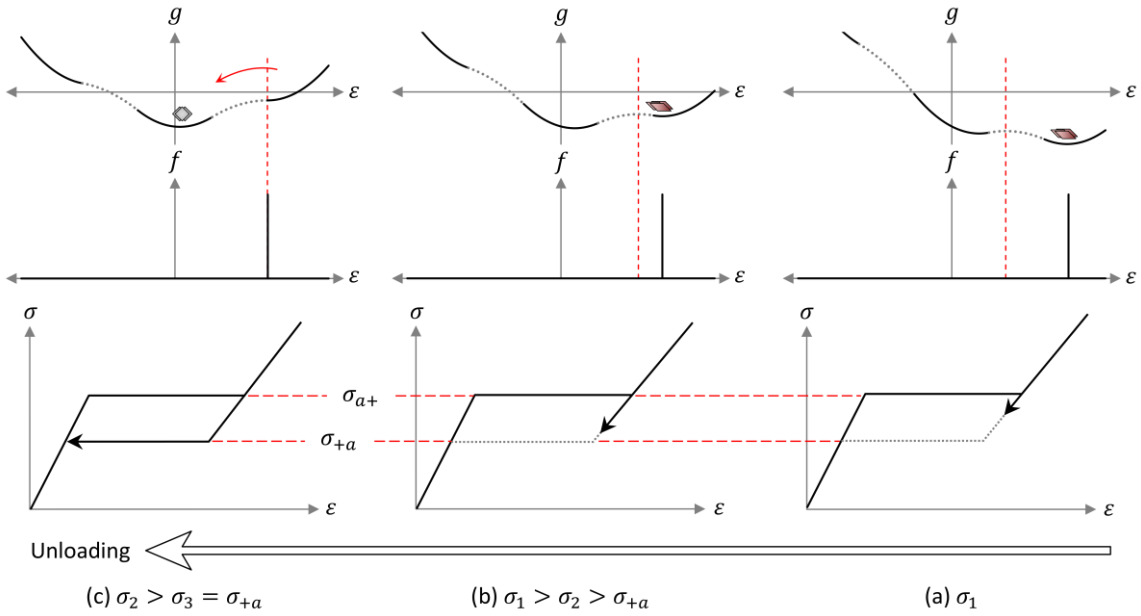


Figure 19: Transformation behavior for low thermal activation material, unloading

In Figure 19(a) the stress has been reduced to a value below that of the forward transformation stress σ_{a+}. For the layers in martensite well, a barrier still exists and they are unable to find the austenite well. In 19(b) the stress is lower but still slightly higher than the reverse transformation stress σ_{+a}. A small barrier remains, no layers transform, and macroscopic unloading path remains linear. The reverse transformation is shown in 19(c) where the stress equals σ_{+a}. Like the loading case, all of the layers simultaneously jump to the austenite well as soon as the critical strain reaches the martensite well minimum. The macroscopic consequence is another sudden, flat stress plateau leading back to the linear austenite branch.

Figures 16-19 illustrate how the general kinetics model can represent gradual or sudden transformations depending on thermal activation parameters. Since solid SMA materials represent low thermal activation and sudden transformations, we can introduce an alternate criterion for the onset of transformation. Figure 20 shows another pairing of a GFE and probability distribution for a low thermal activation material but now, three strain locations are marked in red.

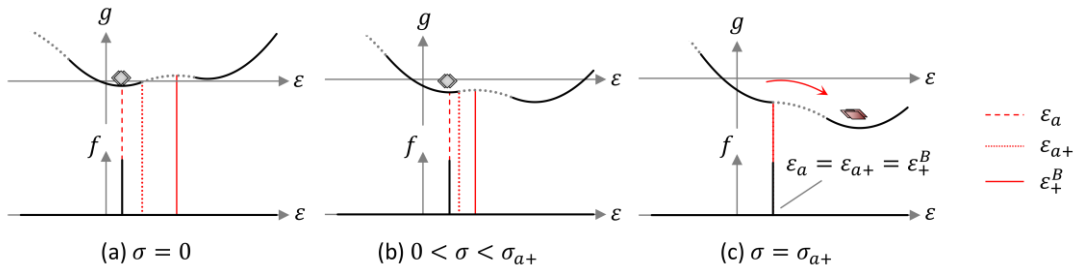


Figure 20: Equivalent transformation thresholds for low thermal activation material

Figures 20(a) and (b) show that, as the stress increases, the three red lines move closer to one another. In 20(c) we see that, at the instant when the critical stress is reached, the three lines coincide. This means that, for low thermal activation materials, the inflection points and barrier peaks collide with the well minimum at the same instant. Because of this, we can use the either the barrier peaks or the inflection points to signify the transformation thresholds. If we choose to use the inflection point criteria, Figure 15 is redrawn as Figure 21 and now there are four critical strains instead of three.

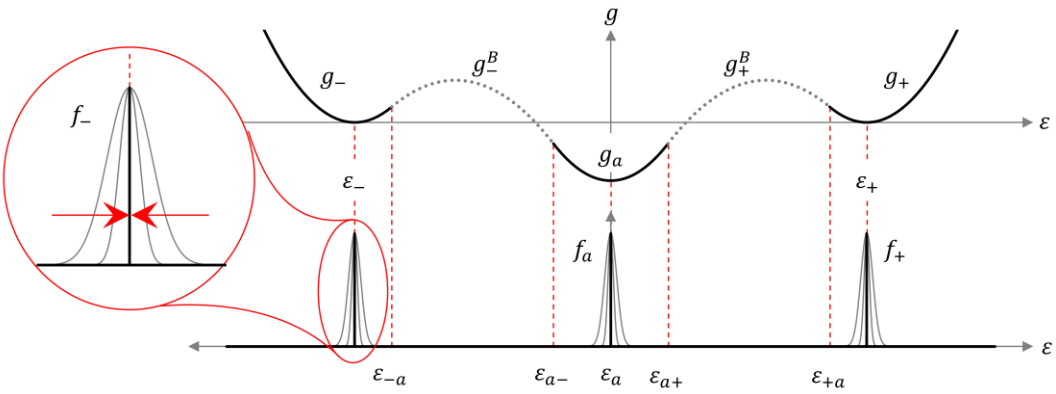


Figure 21: Gibbs free energy and low thermal activation probability distributions

Because this interpretation is restricted to low thermal activation materials, we the limiting case where the Bell shrink to Dirac delta functions. The inflection points now signify the limits of where we expect to find layers, so the probability density functions are adjusted accordingly.

$$f_- = \frac{e^{-\frac{1}{\gamma}g_-(\varepsilon,T;\sigma)}}{\int_{-\infty}^{\varepsilon_{-a}} e^{-\frac{1}{\gamma}g_-(\varepsilon,T;\sigma)}d\varepsilon} \qquad f_a = \frac{e^{-\frac{1}{\gamma}g_+(\varepsilon,T;\sigma)}}{\int_{\varepsilon_{-a}}^{\varepsilon_{a+}} e^{-\frac{1}{\gamma}g_a(\varepsilon,T;\sigma)}d\varepsilon} \qquad f_+ = \frac{e^{-\frac{1}{\gamma}g_+(\varepsilon,T;\sigma)}}{\int_{\varepsilon_{+a}}^{\infty} e^{-\frac{1}{\gamma}g_+(\varepsilon,T;\sigma)}d\varepsilon} \tag{45-47}$$

Equations 45-47 are appealing because they eliminate the piecewise evaluaton of the integrals as well as the need to define the barrier functions g_-^B and g_+^B. The transformation probabilities are obtained by evaluating the probability densities at each critical strain, so these expressions are updated to have the form of Equations 48-51.

$$\begin{aligned} p_{a-}(\sigma) &= \frac{1}{\tau} f_a(\varepsilon_{a-}(\sigma,T),\sigma) \qquad & p_{a+}(\sigma) &= \frac{1}{\tau} f_a(\varepsilon_{a+}(\sigma,T),\sigma) \\ p_{-a}(\sigma) &= \frac{1}{\tau} f_-(\varepsilon_{-a}(\sigma,T),\sigma) \qquad & p_{+a}(\sigma) &= \frac{1}{\tau} f_-(\varepsilon_{+a}(\sigma,T),\sigma) \end{aligned} \tag{48-51}$$

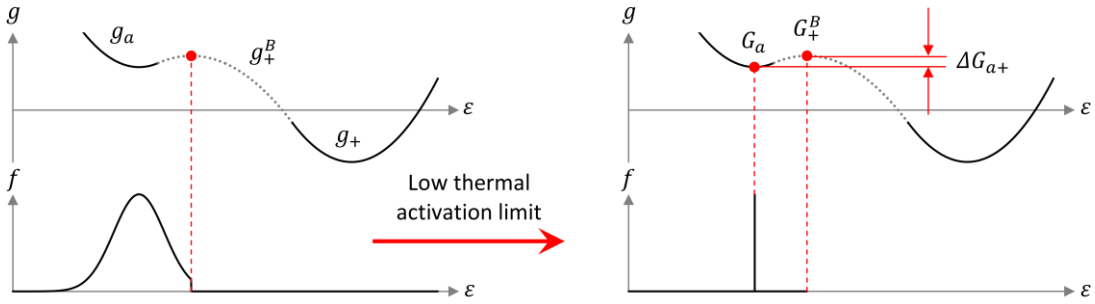
This version of the transformation probability eqautions have been sucessfully used in several SMA publications (Seelecke (2002), Seelecke and Müller (2004) , Smith et al. (2006), Heintze and Seelecke (2008), Ballew and Seelecke (2009), Furst et al. (2012)) and is also included in a comparison of kinetics models later this chapter. We refer to this formulation as the *General* kinetics model (even though it is a restricted case for Equations 41-44) because it still includes the probability integrals. This feature gives the *General* model unique rate insensitivity properties, and this is examined later in this chapter.

Despite these simplifications gained by using the inflection points as the transformation thresholds, equations 45-47 (and therefore Equations 48-51) still require special attention for use in a numerical simulation. Their evaluation requires several steps of algebraic manipulation, the use of special functions like $erf(z)$ and $erfx(z)$, as well as multiple logical constructs to prevent numerical overflow errors. Details on their evaluation and associated numerical issues are included in Chapter 10.1. In the next section we examine a newer technique to calculate the transformation probabilities that is also based on the low thermal activation assumption but offers a simpler set of equations.

3.3 Energy Barrier Kinetics Model

The *Energy Barrier* model begins back at Equations 41-44, where the barrier peaks are used for the transformation thresholds. If we take the mathematical limit of the probability expressions as the thermal activation parameter γ approaches zero, we restrict the model to low thermal activation materials but also obtain a much simpler expression. Since this model is intended for SMA solids, we accept this limitation and replace the

probability distributions f_i with their low thermal activation limit expressions. This is illustrated in Figure 22 where we show the transformation rate p_{a+} in the general form (left) and the low thermal activation limit (right).



$$p_{a+} = \frac{1}{\tau}\frac{e^{-\frac{1}{\gamma}g(\varepsilon_+^B,T;\sigma)}}{\int_{\varepsilon_-^B}^{\varepsilon_+^B} e^{-\frac{1}{\gamma}g(\varepsilon,T;\sigma)}d\varepsilon} \qquad \lim_{\gamma\to 0} p_{a+} = \frac{1}{\tau}\frac{e^{-\frac{1}{\gamma}g(\varepsilon_+^B,T;\sigma)}}{e^{-\frac{1}{\gamma}g(\varepsilon_a,T;\sigma)}} = \frac{1}{\tau}\frac{e^{-\frac{1}{\gamma}G_+^B}}{e^{-\frac{1}{\gamma}G_a}} = \frac{1}{\tau}e^{-\frac{1}{\gamma}(G_+^B-G_a)} = \frac{1}{\tau}e^{-\frac{1}{\gamma}\Delta G_{a+}}$$

Figure 22: Low thermal activation limit of the transformation probability using the barrier height as the threshold

When the Bell curve shrinks to zero width, the integral becomes a pointwise evaluation of the integrand. The result is an exponential expression based on the energy difference between the well minimum and the barrier maximum. The other three transformation probabilities are derived in the same manner yielding a set of four expressions based on relative energy.

$$p_{ij}(\sigma,T) = \frac{1}{\tau}e^{-\frac{1}{\gamma}\Delta G_{ij}(\sigma,T)} \qquad where \qquad \begin{cases} \Delta G_{-a} = G_-^B(\sigma,T) - G_-(\sigma,T) \\ \Delta G_{a-} = G_-^B(\sigma,T) - G_a(\sigma,T) \\ \Delta G_{a+} = G_+^B(\sigma,T) - G_a(\sigma,T) \\ \Delta G_{+a} = G_+^B(\sigma,T) - G_+(\sigma,T) \end{cases} \qquad \text{(31 } \textit{Repeated here}\text{)}$$

The above expression, the *Energy Barrier* formulation, was given in Chapter 2. It represents the basic concept (the ΔG) underpinning more recent applications of the M-A-S theory (Welsch et al. (2018)). When this formulation is used, the model yields the same results as the *General* model but is much simpler and bypasses the numerical overflow issues (discussed in Chapter 10.1). Despite these benefits, evaluation of Equation 31 depends on fully defined energy barrier functions and these are not always available.

3.4 Driving Force Approximation

The *Energy Barrier* formulation was successful on a 2D free energy, where the barrier energies where found at saddle points between bowl-shaped wells, but it proved to be insufficient when the model was extended to additional dimensions (9) in Kim and Seelecke (2007). To overcome this, a technique was developed where the barrier energies were approximated in terms of the driving forces. Driving force is the energy difference between two energy well minima. It represents the impetus for layers to move from one well to the other but, alone, does not consider what type of barrier may be resisting the transformation. The driving force from phase $i \to j$ is simply the minimum energy in phase i minus the minimum energy in phase j.

$$F_{ij}(\sigma,T) = G_i(\sigma,T) - G_j(\sigma,T) \qquad where \qquad ij \in -a, a-, a+, +a \qquad (52)$$

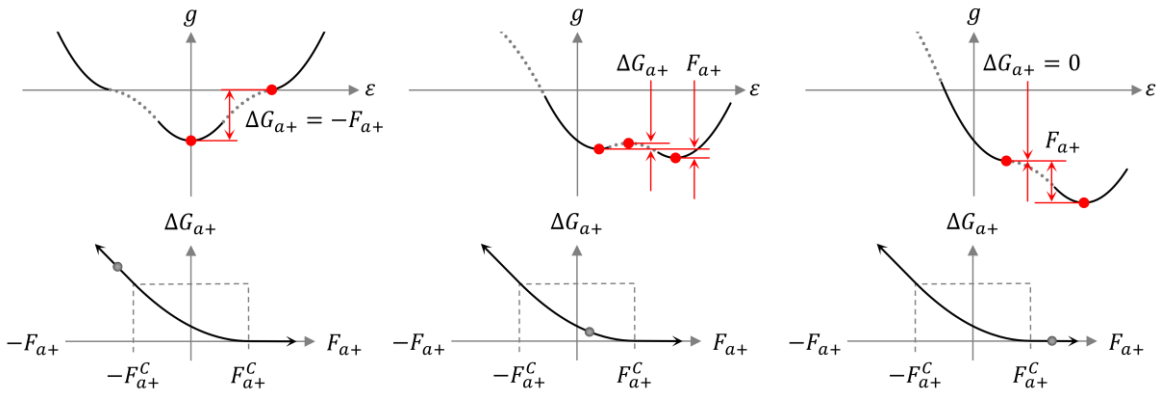
If we examine a plot of ΔG (from Equation 31) vs. driving force, we find that a quadratic, piecewise function (Equation 53) can be used to accurately approximate ΔG.

$$\Delta G_{ij}(F_{ij}) \approx \begin{cases} -F_{ij} & for \quad F_{ij} < -F_{ij}^C \\ \frac{1}{4}F_{ij}^C\left(1-\frac{F_{ij}}{F_{ij}^C}\right)^2 & for \quad -F_{ij}^C \geq F_{ij} \geq F_{ij}^C \quad where \quad ij \in -a, a-, a+, +a \\ 0 & for \quad F_{ij}^C < F_{ij} \end{cases} \tag{53}$$

The parameter F_{ij}^C represents the driving force (Equation 53) evaluated at each transformation (critical) stress; these are called the critical driving forces.

$$F_{ij}^C(T) = F_{ij}\left(\sigma_{ij}(T), T\right) \quad where \quad ij \in -a, a-, a+, +a \tag{54}$$

Figure 23 shows a GFE landscape under three different levels of stress where both the energy barrier height and the driving force are labeled.



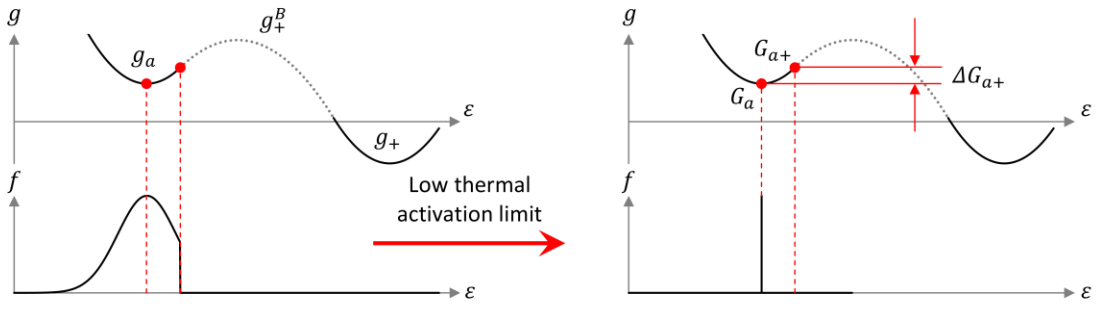
(a) $\sigma = 0, \Delta G = -F$ (b) $0 < \sigma < \sigma_{a+}, \Delta G = f(F)$ (c) $\sigma = \sigma_{a+}, \Delta G = 0$

Figure 23: The driving force approximation of the energy barrier heights

Using Equation 53 to approximate the energy barrier heights in Equation 31 is referred to as the *Driving Force* technique. Even though it was developed in a ferroelectric context, the *Driving Force* model showed promise as a candidate for extension to of the ferroelastic model to multiple dimensions. It approximates the energy barriers and this is important for multi-dimensional energy landscapes, but has a drawback in the case of temperature-dependent hysteresis. For some combinations of material parameters and temperature-dependent hysteresis, the quadratic $\Delta G(F)$ function yields incorrect values which means it is limited to constant hysteresis cases. In the next section, we introduce a new version of the transformation probability expression that overcomes this limitation, offers several other advantages, and is steppingstone toward the 2-D model in Chapter 6.

3.5 Inflection Point Kinetics Model

Here, we combine the assumptions that were used to write two previous formulations to further refine the transformation probably calculations. Figure 24 shows transformation probability p_{a+} with general form (Equations 48-51) on the left and the low thermal activation limit on the right. Now, the inflection point is used to define the critical strain, not the energy barrier peak.



$$p_{a+} = \frac{1}{\tau}\frac{e^{-\frac{1}{\gamma}g(\varepsilon_{a+},T;\sigma)}}{\int_{\varepsilon_{a-}}^{\varepsilon_{a+}} e^{-\frac{1}{\gamma}g(\varepsilon,T;\sigma)}d\varepsilon} \qquad \lim_{\gamma\to 0} p_{a+} = \frac{1}{\tau}\frac{e^{-\frac{1}{\gamma}g(\varepsilon_{a+},T;\sigma)}}{e^{-\frac{1}{\gamma}g(\varepsilon_a,T;\sigma)}} = \frac{1}{\tau}\frac{e^{-\frac{1}{\gamma}G_{a+}}}{e^{-\frac{1}{\gamma}G_a}} = \frac{1}{\tau}e^{-\frac{1}{\gamma}(G_{a+}-G_a)} = \frac{1}{\tau}e^{-\frac{1}{\gamma}\Delta G_{a+}}$$

Figure 24: Low thermal activation limit of the transformation probability using the inflection point as the threshold

Again, the integral reduces to an exponential expression based on relative energy, but now the relative energy is measured from the well minimum to the edge of the convex well (i.e. the inflection point). To calculate the transformation rates in this manner we use Equation 55.

$$p_{ij}(\sigma,T) = \frac{1}{\tau}e^{-\frac{1}{\gamma}\Delta G_{ij}(\sigma,T)} \qquad where \qquad \begin{cases} \Delta G_{-a} = G_{-a}(\sigma,T) - G_-(\sigma,T) \\ \Delta G_{a-} = G_{a-}(\sigma,T) - G_a(\sigma,T) \\ \Delta G_{a+} = G_{a+}(\sigma,T) - G_a(\sigma,T) \\ \Delta G_{+a} = G_{+a}(\sigma,T) - G_+(\sigma,T) \end{cases} \tag{55}$$

Now, both energy values evaluated for each p_{ij} lie on the same energy function g_i and formulation of the barrier energy functions (g_-^B and g_+^B) is no longer required. This reduces the total number of equations required to operate the model and does not affect the model behavior as long as the low thermal activation assumption applies (i.e. the model is being applied to SMA solids). This form of the kinetics model relies on the fact that the minimum energy G_i and inflection point energy G_{ij} coincide when critical stress σ_{ij} is reached. Figure 25 illustrates this for a superelastic tensile load-unload cycle.

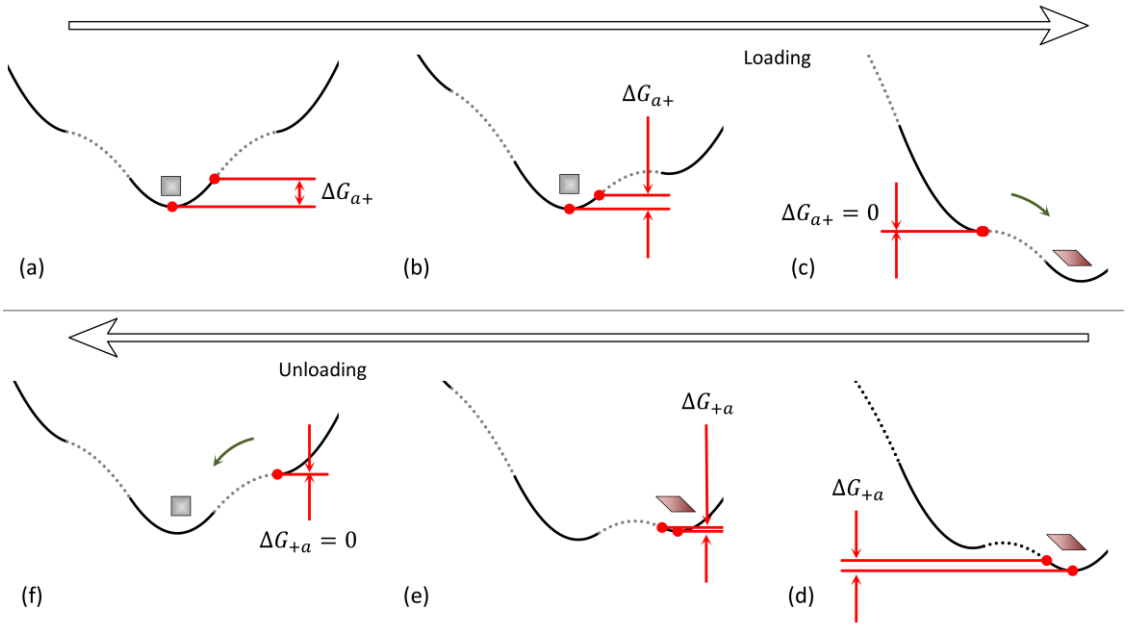


Figure 25: Inflection point ΔG during a forward and reverse superelastic transformation

As the initially austenite system is loaded in Figure 25(a-c), the energy difference between the well minimum and its right-most inflection point shrinks. When the critical stress is reached in 25(c), the well minimum and the inflection point become coincident and the ΔG value reaches zero. This causes the a layers to immediately jump into the martensite $+$ well. Noting that the lower row reads from right to left, we see that the system is being unloaded in 25(d) and (e) but the layers remain trapped in the martensite well. It is not until the reverse critical stress is reached in 25(f) that the layers can return to the central austenite well. Using this transformation mechanism, the hysteretic single crystal behavior in Figure 19 is reproduced here.

Figure 26 shows the same transformation mechanism during a temperature-induced transformation cycle.

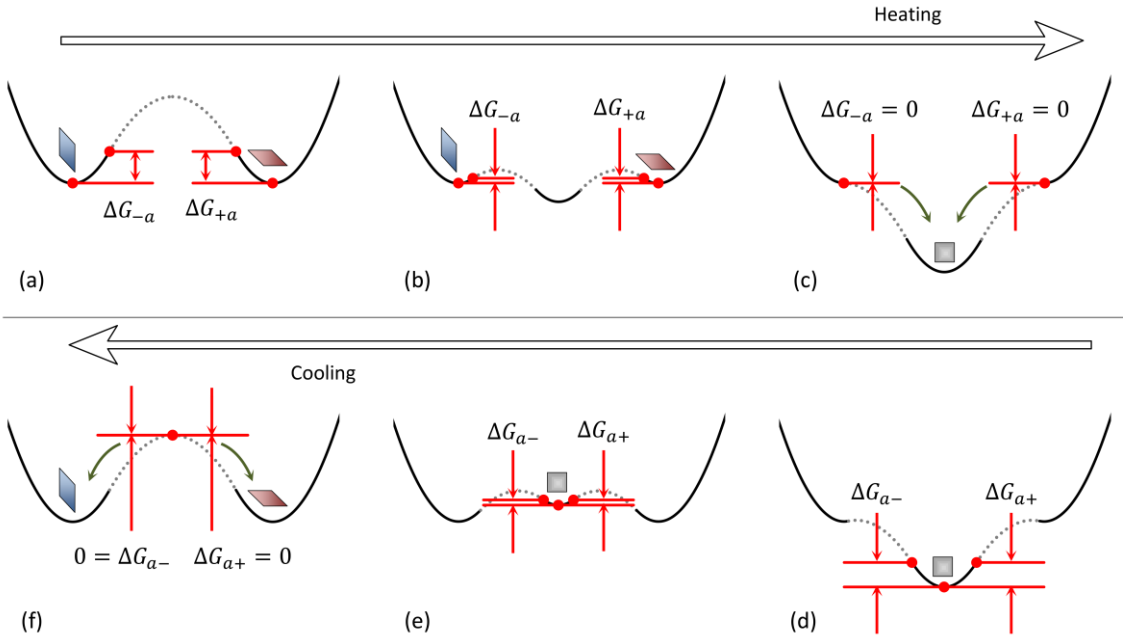


Figure 26: Inflection point ΔG leading to a temperature induced transformation

The upper row (from left to right) shows that, as the central a well descends, the $-a$ and the $+a$ barriers diminish to zero and this activates the switch from twinned martensite to austenite. The reverse transformation is shown in the lower row (from right to left) where the system is cooled and the a well climbs to the point of losing convexity. When both ΔG's are equal to zero the layers to find the two martensite wells in equal proportions (i.e. twinning). In the previous two figures, only selected $\Delta G's$ were shown. In model operation, all four values ($\Delta G_{ij\in -a,a-,a+,+a}$) are continually tracked and all four transformation probabilities ($p_{ij\in -a,a-,a+,+a}$) are continually calculated. The consequence is that the transformation rates rise and fall allowing the model to capture hysteretic and history behavior of a single crystal SMA material.

Figure 27 illustrates why this formulation, in the case of low thermal activation, is equivalent to the energy barrier-based version.

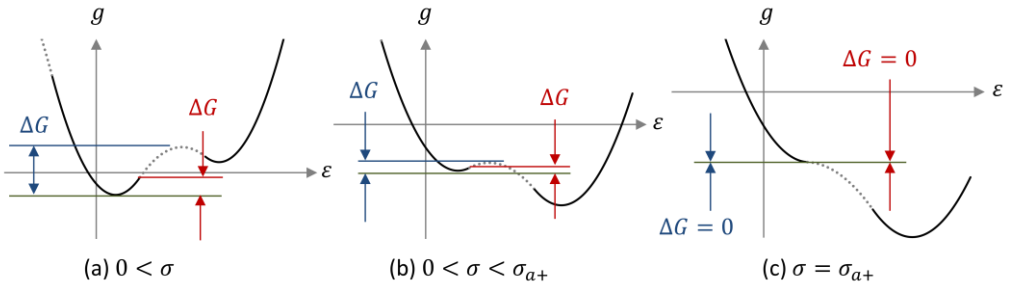


Figure 27: Gibbs free energy landscape showing that the energy barrier ΔG (blue) and that the inflection point ΔG (red) reach zero simultaneously

The concept for the inflection point ΔG model was motivated by the needs of a multidimensional formulation (Chapter 6) but it is also very useful in 1-D applications. This technique is used in Chapter 5, where the 1-D model is extended to accommodate temperature-dependent hysteresis and tension-compression asymmetry.

3.6 Mesoscopic Inhomogeneity

The M-A-S model fundamentally represents a single-crystal SMA material, where the transformations occur at a single stress or temperature value. The single crystal-type of response is really an ideal, limiting case of how actual SMA specimens behave. Depending on the amount of texture in a specimen's microstructure and its loading history, the transformations can also occur gradually over and range of stress or temperature. This is commonly called polycrystalline or inhomogeneous behavior. There are a number of ways to add polycrystalline behavior to a single-crystal-based SMA model, but it can be conveniently added to the M-A-S model via the kinetics equations. Essentially, the transformation stress parameter is made to be a function of the phase fractions in addition to temperature (this is detailed in Heintz and Seelecke (2008) and further refined in Rizzello et al. (2019)). Physically, this reflects the fact that some mesoscopic layers will transform more easily than others and this the transformation threshold changes as the phase fractions evolve. While this is not a primary focus of this dissertation, inhomogeneity effects were necessary to model some of the experiments used to validate the model in Chapter 5. For this reason, we add an empirical adaptation of the technique from Heintz and Seelecke (2008) to the kinetics model here.

The effects of mesoscopic inhomogeneity are added by introducing the concept of a representative lattice element which adds a phase fraction-dependent term to the transformation stresses.

$$\sigma_{a\pm}^{eff}(T,x_{\pm}) = \sigma_{a\pm}(T) \pm \sigma_v^{\pm}(x_{\pm}) \qquad \sigma_{\pm a}^{eff}(T,x_{\pm}) = \sigma_{\pm a}(T) \pm \sigma_v^{\pm}(x_{\pm}) \tag{56,57}$$

$$where \qquad \sigma_v^{\pm}(x) = \sigma_l^{\pm}\left(x-\frac{1}{2}\right) + \begin{cases} -\sigma_b^{\pm}(1-2x)^{e_b^{\pm}} & if\ x < \frac{1}{2} \\ +\sigma_c^{\pm}(2x-1)^{e_c^{\pm}} & if\ x \geq \frac{1}{2} \end{cases}$$

The parameter σ_{ij}^{eff} represents the transformation threshold of the most favorable layer. To implement this option in the model, the transformation stress $\sigma_{ij}(T)$ is replaced by the effective transformation stress $\sigma_{ij}^{eff}(T,x)$ in all equations used to calculate the ΔG values. In Chapter 5 the *Model Implementation* section, where we give a user's guide style summary of the essential model equations, the formulation assumes that the inhomogeneity option is active. It should also be noted that the inhomogeneity equations (56 and 57) do not need to be situationally added to or removed from the model. If the parameters $\sigma_l^{\pm}$, $\sigma_b^{\pm}$, and $e_b^{\pm}$ are set equal to zero, the single crystal model is recovered. The exponential expression in Equations 56 and 57 is a phenomenologically

adapted formulation that produces similar results to the more rigorous forms. Figure 28 shows σ_{a+}^{eff} and σ_{a+} vs. the martensite + phase fraction.

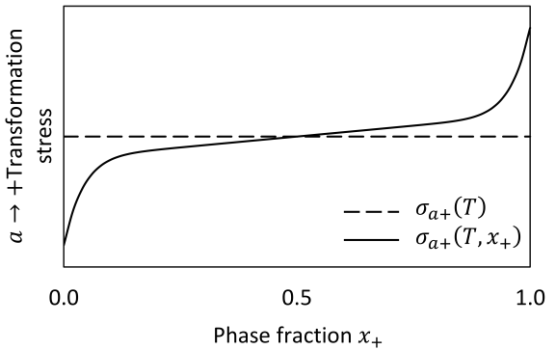


Figure 28: Representative layer transformation stress

Low values x_+ correspond to low values of σ_{a+}^{eff} and this represents the small fraction of layers that, due to local stress fields, transform at macroscopic stresses far below the average value σ_{a+}. The opposite is true at high values of x_+; when the transformation is nearly complete, the last few layers require stress much greater than σ_{a+} to overcome their barriers due to local stress fields. We present several simulations using the polycrystalline version of the new model in the following section.

3.7 Comparison of Kinetics Models

A single crystal SMA, under isothermal conditions, is expected to have a stress-strain response such as those shown in Figures 3 and 4. Examples of this would be extremely slow strain rates, a low ambient temperature, a thin wire, etc. In this situation, the characteristic shape of the stress-strain path is independent of strain rate because the underlying microstructures transform very rapidly. From a modeling perspective, when latent heats are not considered, the same results should be obtained for any applied strain rate. The M-A-S model, using any of the kinetics formulations, achieves this goal but only if proper values are chosen for thermal activation parameters (τ and γ). For this reason, it is important to understand that the general kinetics model and the ΔG kinetics models have somewhat different relationships with (sensitivities to) τ and γ. In this section we show the behavior of each model and discuss the underlying mechanisms as each converges to the proper rate independent response.

Figure 29 shows the rate of transformation (phase $i \rightarrow j$) vs. stress for the general kinetics model. This figure applies to any of the transformation rates $p_{ij \in -a,a-,a+,+a}$ and shows that the rate increases in the vicinity of the critical stress σ_{ij} for any level of thermal activation.

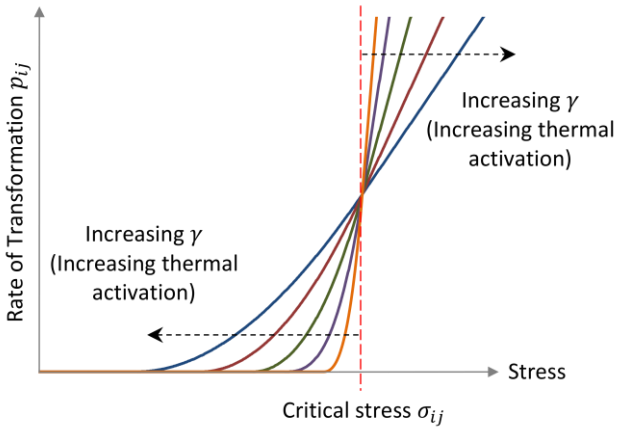


Figure 29: Rate of transformation vs. stress for the general kinetics model

We see that p_{ij} has a smooth ramp characteristic with a final slope that depends on γ. The early onset of transformation from Figures 16(a) and 17(a) can be explained here as the higher thermal activation curves begin to rise earlier with respect to the critical stress. Another important feature of the general form is that the rate of transformation continues to rise after the critical stress has been exceeded. This mechanism encourages strain rate independence in the macroscopic response because the transformation rates can increase to arbitrarily high levels. Figure 29 also shows how the level of thermal activation affects this feature. The high thermal activation cases start transforming earlier but accelerate at a lower rate and the opposite is true for low thermal activation. When γ is small, the rate of transformation remains at zero till the stress has nearly reached the critical stress then climbs immediately to a huge value. This rapid rate of change in phase fractions allows the model to yield the same stress response for any rate of applied strain – i.e. rate-independent behavior. The sampling rate $1/\tau$ also affects the rate of transformation p_{ij} but plays a minor role for the general kinetics model. This term is a linear scale factor for the rate of transformation but does not change the overal characteristic of the curves in Figure 29.

Sampling rate τ is more critical for the ΔG kinetics model, where the ramping nature of the transformation rate is no longer present. In contrast to the ramping curves in Figure 29, the ΔG model has a step-like characteristic and this is due to its exponential form. The rate of transformation jumps to a maximum when the stress exceeds the critical stress and remains locked there as shown in Figure 30.

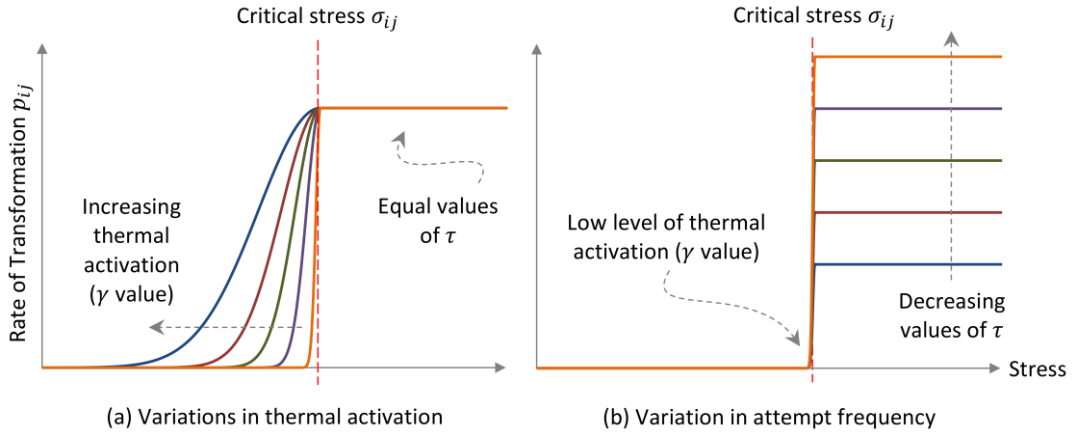


Figure 30: Rate of transformation vs. stress for the ΔG kinetics model

Figure 30 shows the rate of transformation vs. stress in two situations. The left-most plot (a) shows the sensitivity to the thermal activation parameter γ while the sampling time τ is held constant. The plot on the right (b) represents a constant value of γ while τ is varied. In Equations 31 and 55, the ΔG term becomes locked at zero

when the critical stress has been exceeded therefore the transformation rate freezes at a value of $\frac{1}{\tau}e^0$. The result is that the τ parameter alone sets the maximum rate of transformation and this is shown in Figure 30(b). The γ parameter, on the other hand, dictates the suddenness of the transformation, i.e. the sharpness of the step in Figure 30(a).

The simplest way to find the proper values for the thermal activation parameters is to run a series of simulations of a load-unload cycle at different strain rates. When two significantly different strain rates produce indistinguishable stress-strain plots (again, latent heat effects are turned off), then isothermal rate independence has been achieved. Higher values of τ and lower values of γ lead to better performance (i.e. lower sensitivity to strain rate), but they cannot be set arbitrarily high and low repsectively. Very low/high values of τ/γ result in very rapid transformations. This is desireable but is limited by the numerical integration scheme's ability to handle stiff equations. Fortunately, a single value for each parameter can ensure rate independene over several decades of strain rate without causing numerical issues. Next, we show example sensitivity studies demonstrating how to establish thermal activation parameters for both the general model and a ΔG-type of model.

Figure 31 shows results from 45 simulations of a strain-controlled load-unload cycle using the general model. Each row of plots in Figure 31 represents a value of γ, each column designates a strain rate $\dot{\varepsilon}$, and each individual plot (a-i) contains five curves corresponding to the five values of τ.

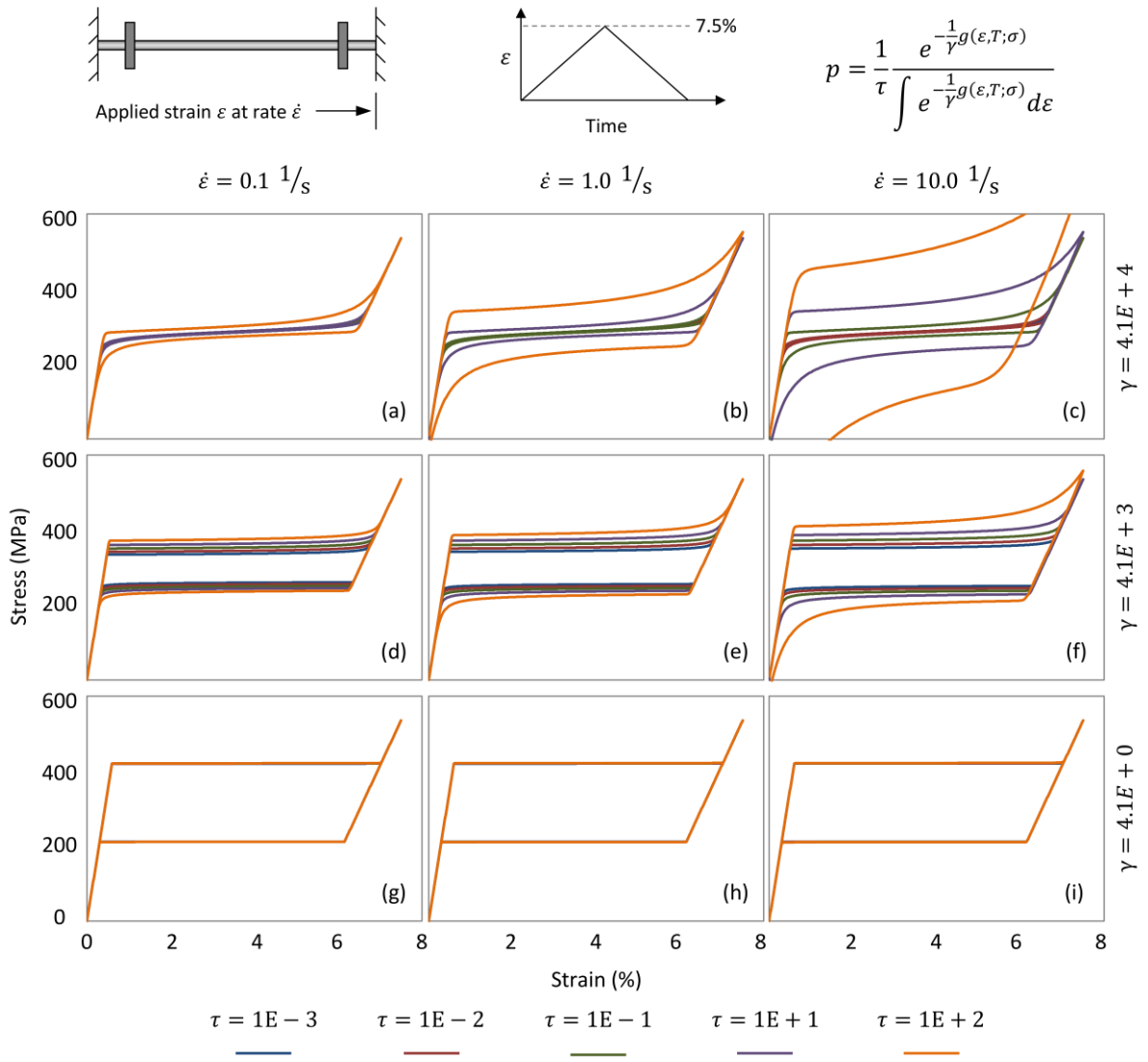


Figure 31: General kinetics model with various thermal activation parameters and strain rates

The series of curves shows competition between the effects from Figure 29 and the applied strain rate. The rounded shape of the curves as they enter the stress plateaus in Figure 31(a-f) is due to the premature initiation of transformation associated with high levels of thermal activation. The smooth exits from the stress plateaus, which are most promenant for large values of τ, are the result of insufficient transformation rate. The large τ reduces the rate of change of the phase fraction x_+ and this causes stress to increase during transformation. The stress rise during forward transformation is due to a back-log of layers trying to get out of the austenite phase and the reverse effect is also seen during unloading.

The hysteresis in the responses stems from a combination of two different phenomena – free energy barriers and strain rate effects. In the cases representing high thermal activation 31(a-c), the hysteresis is due completely to strain rate effects. The large thermal fluctations make the system unaware of the energy barriers and this results in a Maxwell-like stress-strain response for the lower values of τ. For larger values of τ, the rates of transformation are lower and the strain-rate induced hysteresis appears. A stress plateau only forms when the rate of transformation exceeds the rate of applied strain. Scaling the rate of transformation up or down by ${}^{1}/_{\tau}$ either decreases or increases the amount of time required for a plateau to form. This delay in plateau formation appears

in the stress-strain response as a hysteresis loop. Because it is a rate effect, the variation in hystresis width differs for different strain rates (compare 31(a) to (b) and (c)).

In Figure 31(d-f) we see two effects due to the reduced thermal activation. The first is the appearance of a hysteresis loop at even the lowest strain rate (d). The layers now sense the barriers enough to distinguish between forward and reverse thresholds. The second effect is a reduced sensitivity to τ and this is most apparent when comparing (c) and (f). In (c), the $\tau = 1 \cdot 10^{-2}$ case yielded a path that was highly distorted due to untransformed layers. In (f), for the same value of τ, the response is much more consistent with the other curves. The lower value of γ allows the transformation rate to increase more rapidly, thereby compensating for the slowing influence of large τ's.

31(g-i) represents appropriate choices of thermal activation parameters as well as a correctly shaped and rate-independent response. When low thermal activation parameters are assigned, the premature transformation no longer occurs. The system precisely obeys the forward and reverse transformation stress thresholds resulting in a consistent hysteretic response. Furthermore, in all 15 cases (31(g-i)), no strain rate-induced hysteresis is encountered. For low levels of thermal activation, the transformation ramps immediately to a rate that is essentially instantaneous compared to the applied strain rate. Mathematically, this eliminates the widening hysteresis observed in 31(a-f) because no back-log of untransformed materials appears. This is also physically representative of the isothermal rate independent behavior of actual SMA tensile specimens. At critical stress or temperature, the change in microstructures occurs at a rate that is essentially instantaneous when compared to rate of bulk deformation.

Next we focus on the ΔG model. As shown in Figure 30, this form shares some features of its general counterpart but also has some key differences. In Figure 32 we see another set of 45 simulations that are identical to those in Figure 31 except we use the ΔG form of the kinetics model (Equation 55) and a different set of values for τ.

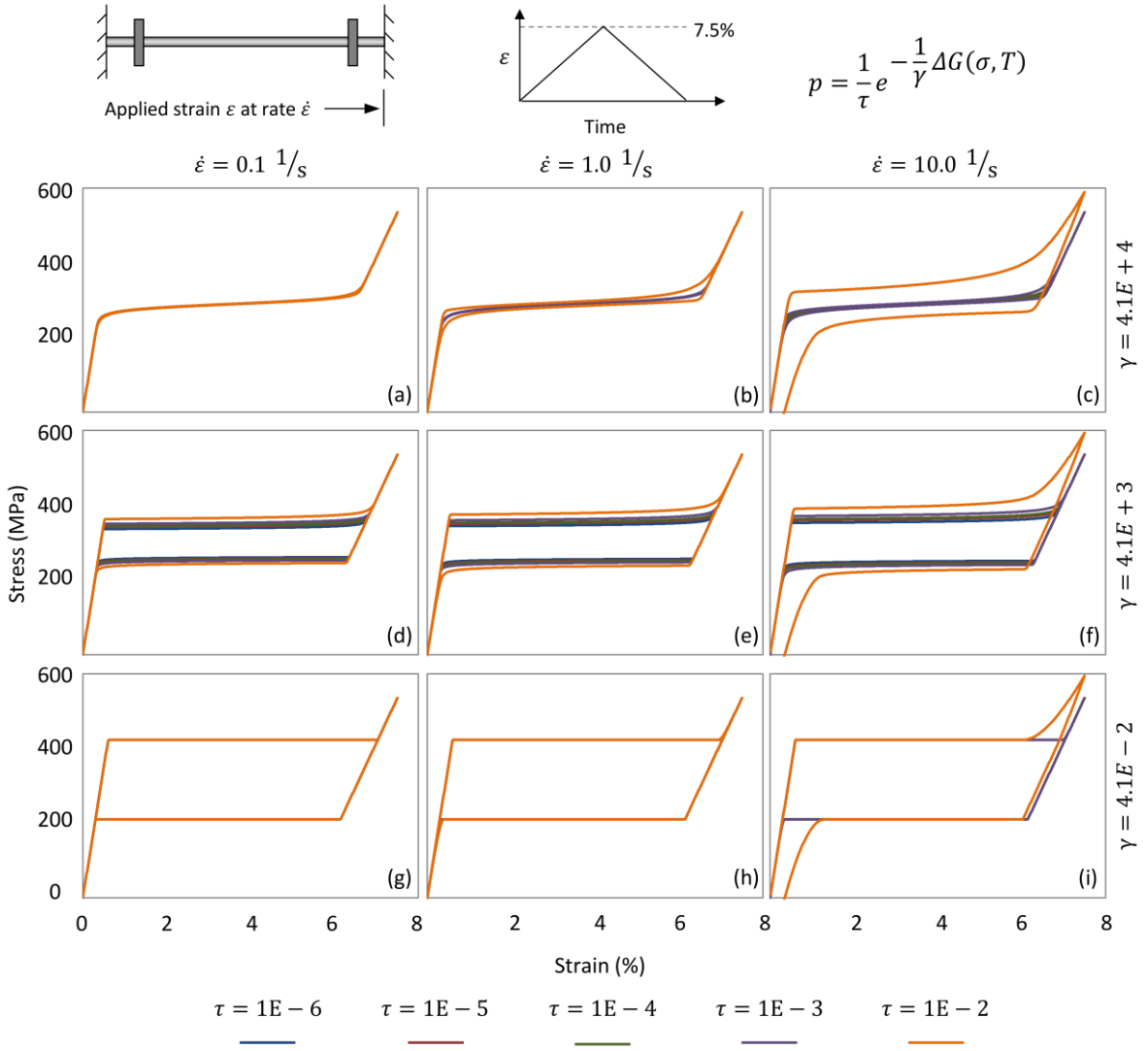


Figure 32: ΔG kinetics model with various thermal activation parameters and strain rates

The most apparent trend in Figure 32 is the variation in hysteresis width as the value of γ changes. This effect is shared with the general model and stems from the premature onset of transformation when thermal activation is unrealistically high (a-f). For the low thermal activation cases (32(g-i)) the transformations initiate precisely when the threshold stress is reached.

The responses in Figure 32 are also affected by the applied strain rate but in a different manner than the curves in Figure 31. Recalling Figure 30, the ΔG kinetics models are associated with transformation rates that climb to a max value then freeze there. The inability to accelerate the transformation makes these models much more sensitive to the parameter τ. For this reason, a much lower set of τ values had to be used for the simulations in Figure 32 to show convergence. The sensitivity to τ is most visible for the cases with the largest value ($\tau = 1 \cdot 10^{-2}$) where both strain rate hysteresis (32(c)) and incomplete transformation (32(c, f, and i)) result. When the maximum transformation rate is fixed, the applied strain rate dictates a minimum threshold for the successful transformation. In a sense, the ΔG model has a pass-fail type of relationship with τ. If τ is low enough, the rate of transformation is high and the model creates well-formed stress plateaus. On the other hand, if τ is too large, the transformation lags and stress plateaus are unable to form. In fact, there is only a small range of τ where a poorly formed plateau can be created (one example is shown by $\tau = 1 \cdot 10^{-2}$). Any larger values result in a stress-strain

curve that does not even resemble an SMA response and were not included in the study. The proper selection of τ requires that the rate of transformation exceed any strain rate to be encountered in the system. Analytically, the value can be assumed to be near-zero resulting in an approximately infinite rate of transformation but in a numerical simulation a finite value must be assigned. If the maximum strain rate $\dot{\varepsilon}_{max}$ is known, an expression for the maximum value of τ can be written (Equation 58) based on the amount of strain associated with a full forward transformation.

$$\tau_{max} = \frac{\dot{\varepsilon}_{max}}{\varepsilon_R + \frac{\sigma_{a+}(T)}{E_m} - \varepsilon_{a+}(T)} \tag{58}$$

A value that is one or two orders of magnitude lower than τ_{max} typically yields an accurate and rate-independent response without encountering overly stiff kinetic equations and numerical integration issues.

We conclude this section with a set of simulations comparing the two types of ΔG model (the general model is also included for reference). Figure 33 shows six sets of simulations of a full tension-compression cycle at a superelastic temperature. The three columns represent different combinations of thermal activation parameters and the two rows represent two different strain rates.

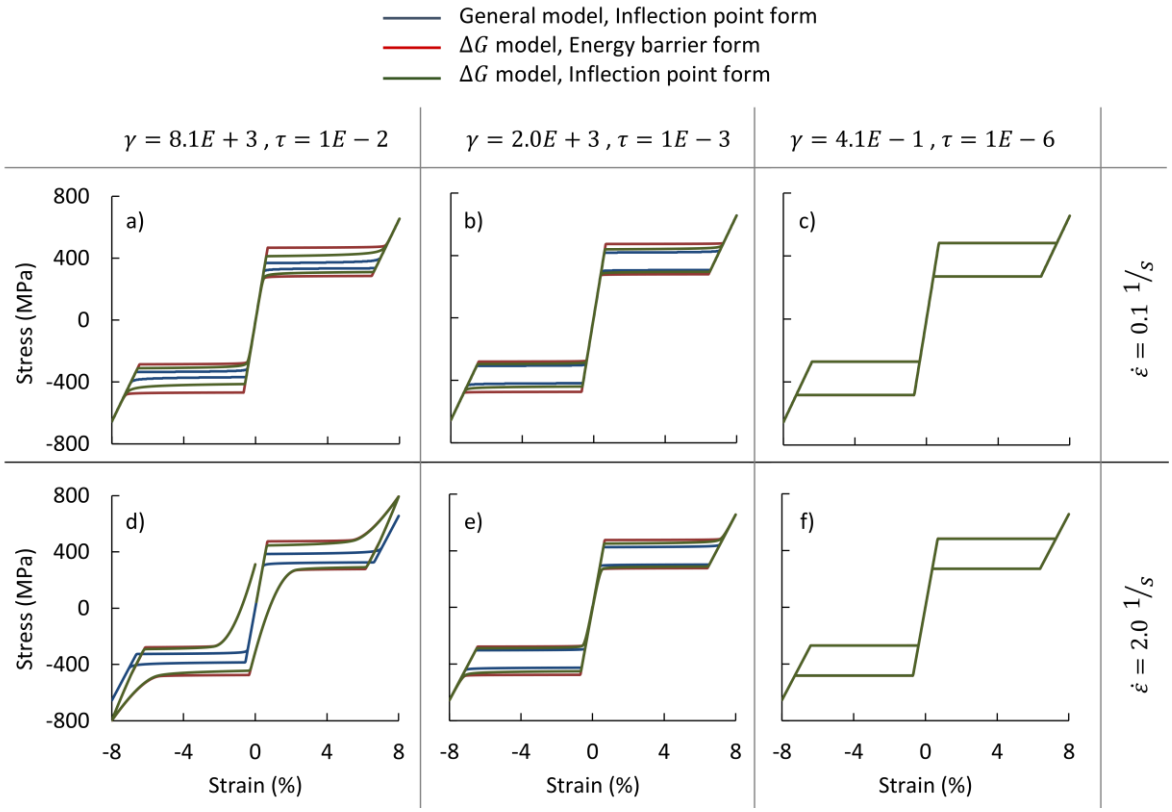


Figure 33: comparison of three forms of the kinetics model

Cases 33(a) and (d) represent high levels of thermal activation along with low attempt frequency. 33(a) shows that all three models yield different stress-strain paths because the transformations occur at different stress levels. 33(d) shows that the deviation is exacerbated when the applied strain rate increases, due to the large value of τ. The rate of transformation, especially in the two ΔG models, depends on the attempt frequency τ and a value too large results in artificial and incorrect strain rate dependence. Cases 33(b) and (e) show both lower thermal activation and higher attempt frequency. The deviation in transformation stress is reduced and the strain rate sensitivity has essentially been removed. The right-most column (c and f) shows a converged set of γ and τ values

where, at either strain rate, all three models exhibit indistinguishable responses. This means that, if care is taken to choose appropriate values for thermal activation parameters, any of the kinetics models presented in this chapter can be used and the same SMA behavior will be produced. The reasons to choose one vs. the other are based ease of implementation and opportunity to include additional effects as is done Chapters 5 and 6.

After a detailed examination of the kinetic theory, this chapter equipped us with a new set of transformation probability equations. The target for a closer look is the model's thermodynamic framework, which encompasses the free energy, caloric properties, and irreversibility.

4 Thermodynamics

In this chapter we show how the Helmholtz free energy is constructed and this naturally exposes how other thermodynamic quantities (such as internal energy, entropy, latent heat, etc.) are woven into the model. We begin by establishing the relationships between stress, strain, and free energy at both the mesoscopic and macroscopic scales. This is an important concept because SMA models constitute a variety of multiscale arrangements. Next, we show how to measure the essential material constants from experimental data – this is a straightforward process after establishing the relationship between the scales. The stress-strain relationships and material constants provide the tools to construct the free energy and we walk through this process, discussing the underlying physics and assumptions along the way. The end result is the Helmholtz free energy equation that was given in Chapter 2.

The equations that support the free energy show us that the transformation stress is one of the main building blocks. The next section is devoted to the transformation stress equation where we show that it can be derived directly from the free energy. This exercise also reveals several other caloric properties of the material, which are then used to derive the latent heat coefficients. The latent heat expressions contain information related to irreversibility and this is unique among SMA theories. As such, we examine the latent heat formulation in more detail by relating it to other SMA theories as well as comparing predictions to experimental data.

We conclude the chapter by examining the irreversibility from the perspective of the stress-strain curve. SMA modelers sometimes write Clausius-Duhem-like inequalities as a platform to add entropy production terms such as dissipation potentials to the system. The M-A-S model does not contain any explicit entropy-production terms because the irreversibility is inherently contained in the multiwell free energy. We demonstrate this by showing that the entropy derived from a macroscopic stress-strain cycle can be written in terms of the free energy-derived latent heats.

4.1 Relating Stress, Strain, and Free Energy

The mesoscopic Helmholtz free energy is constructed from macroscopic tension test data. Before discussing the connection between stress-strain behavior at the two scales, we consider the mesoscopic layers in a given phase. The differential internal energy du_i for phase i is given by In Equation 59.

$$du_i = Tds_i + \sigma d\varepsilon_i \tag{59}$$

All of the quantities represent the average, or most probable, value for all of the layers in phase i. No subscript is assigned to temperature or stress because these quantities are, on average, equal for all phases. We subtract $d(Ts_i)$ from both sides and recognize that $d(Ts_i) = s_i dT + Tds_i$ which allows us to cancel two terms.

$$du_i - d(Ts_i) = Tds_i + \sigma d\varepsilon_i - d(Ts_i) \tag{60}$$

$$d(u_i - Ts_i) = \cancel{Tds_i} + \sigma d\varepsilon_i - s_i dT - \cancel{Tds_i} \tag{61}$$

The Helmholtz free energy is defined as $\psi = u - Ts$ so we substitute ψ to write a differential form of the Helmholtz free energy for phase i.

$$d\psi_i = \sigma d\varepsilon_i - s_i dT \qquad yields \qquad \left.\frac{\partial \psi_i}{\partial \varepsilon_i}\right|_{T=const} = \sigma \qquad and \qquad \left.\frac{\partial \psi_i}{\partial T}\right|_{\varepsilon=const} = -s_i \tag{62,63}$$

Equation 62 relates the stress and strain to the free energy at the mesoscopic scale, but the test data used to build the free energy is macroscopic. Next, we need to relate the mesoscopic stress-strain behavior to a macroscopic test specimen. This is accomplished by recalling Equation 26, which defines the macroscopic strain as the weighted average of the strains in each phase (ε_i). This means that, when all of the mesoscopic layers are in the same phase, the macroscopic strain is equal to the mesoscopic strain.

$$\begin{aligned}
&All\ of\ the\ layers\ are\ the\ martensite-phase && \rightarrow && x_- = 1, x_a = x_+ = 0 && \rightarrow && \varepsilon = \varepsilon_- \\
&All\ of\ the\ layers\ are\ the\ austenite\ phase && \rightarrow && x_- = 0, x_a = 1, x_+ = 0 && \rightarrow && \varepsilon = \varepsilon_a \\
&All\ of\ the\ layers\ are\ the\ martensite+phase && \rightarrow && x_- = x_a = 0, x_+ = 1 && \rightarrow && \varepsilon = \varepsilon_+
\end{aligned}$$

Figure 34 shows the stress-strain response for a typical (single-crystal-like) SMA material along with the corresponding mesoscopic Helmholtz free energy. This illustration shows that the three linear branches of the stress-strain curve represent regions where the material exhibits only one phase, i.e. where the mesoscopic and macroscopic strains are equivalent.

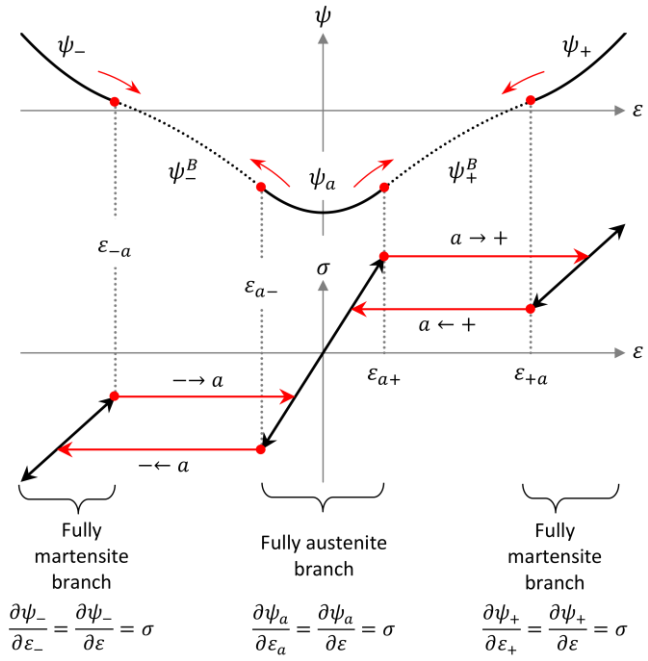


Figure 34: Mesoscopic HFE and macroscopic stress-strain behavior

The black segments on the macroscopic stress-strain plot correspond to regions when the material is in a single phase and the red segments represent phase change. The black stress-strain segments have arrows on both ends because the specimen can be loaded or unloaded on these branches and the response will remain linear because the phase fractions will not change. The red stress-strain segments have arrows on only one side because they are one-way branches, either forward transformation ($a \rightarrow \pm$) or reverse transformation ($\pm\rightarrow a$). Each linear stress-strain branch corresponds to a convex energy well and they are related by the Equation 62. No stress-strain relationship is required for the red segments. When phase change is underway, the layers in each well are at

thermal and mechanical equilibrium (i.e. temperature and stress remain constant) and the kinetics equations drive the phase fractions toward phase equilibrium.

The main take-away from Figure 34 is that we can use standard techniques (i.e. $\frac{\partial \psi}{\partial \varepsilon}=\sigma$) to relate macroscopic stress-strain data to the mesoscopic free energy function, as long as we restrict ourselves to the linear branches. Fortunately, all of the material parameters required to build the free energy are measured from the linear, single-phase, branches of the curve. Figure 35 shows tensile test data for NiTi specimens from Chang et al. (2006), material parameter measurements, and the corresponding free energies in a similar manner to Figure 34.

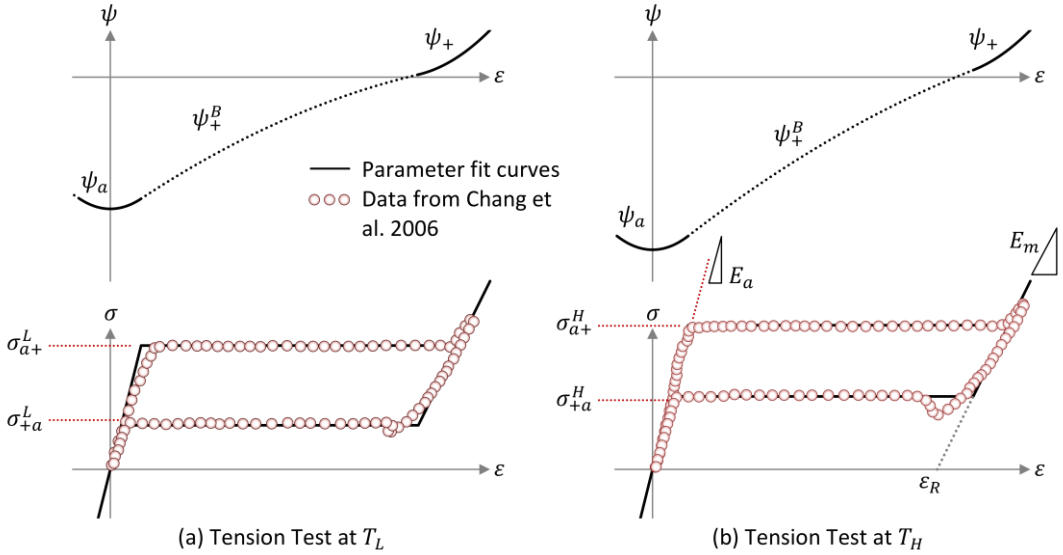


Figure 35: Tension test data at two temperatures and corresponding HFE's

Figure 35 shows two tensile test data sets, each at a different temperature (T_L and T_H). The material parameters measured in the figure are the same ones given in Table 1 (Chapter 2) and represent the essential variables required to build the mesoscopic free energy. Now we have the essential tools to show the complete construction of the Helmholtz free energy function (Equation 1).

4.2 Construction of the Helmholtz Free Energy

The construction of the free energy begins with the concept of a metastable, i.e. multiwell landscape. We use a chain of parabolas as shown in Figure 36(a) but this is not the only method. Another commonly cited multiwell free energy, the Landau-Ginzburg formulation by Falk (1980 and 1983), is shown in 36(b).

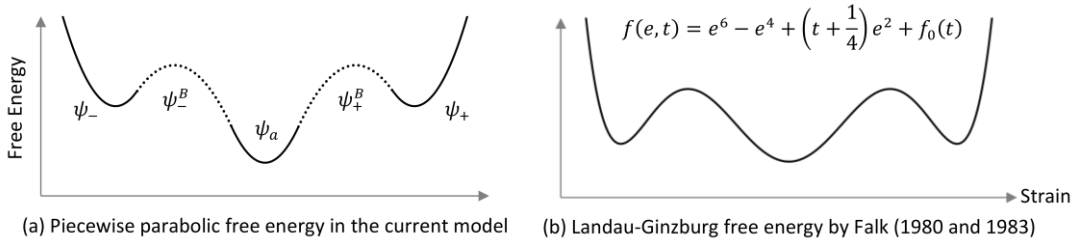


Figure 36: Sketch of a Landau-Ginzburg and the current model's free energy

Using a train of parabolas is convenient for a number of reasons. First, they produce linear stress-strain responses within each phase which is expected in a single crystal material. Furthermore, the stresses and strains associated with the inflection points, and therefore the transformations, can be defined as explicit functions of

temperature; this is a common assumption in SMA theory. Lastly, each of the phases has a distinct energy function with a simple, recognizable form.

To build the piecewise free energy in 36(a), we start with three convex energy wells and assume that each is composed of a strain-dependent (mechanical) part and a temperature-dependent (chemical) part as sketched in Figure 37.

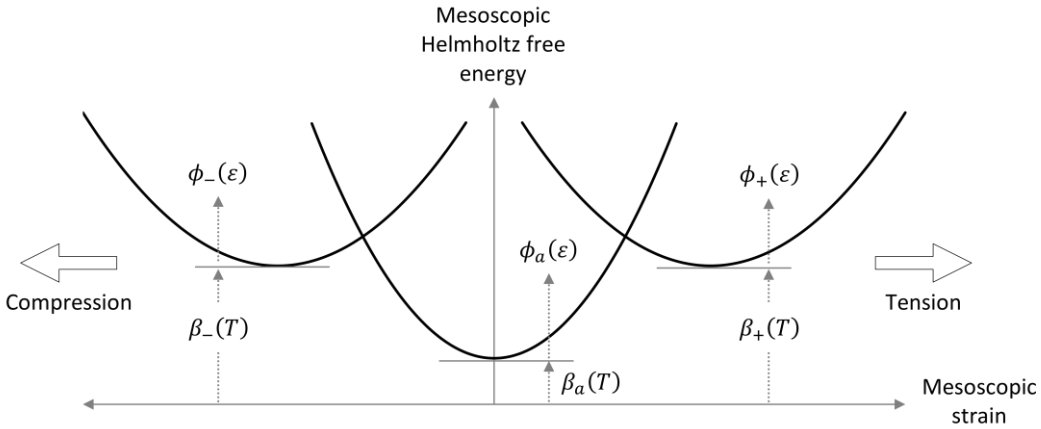


Figure 37: Mechanical and chemical free energies for each phase

The three free energies are given by Equations 64-66.

$$\hat{\psi}_-(\varepsilon,T) = \phi_-(\varepsilon) + \beta_-(T) \quad (64) \qquad \hat{\psi}_a(\varepsilon,T) = \phi_a(\varepsilon) + \beta_a(T) \quad (65) \qquad \hat{\psi}_+(\varepsilon,T) = \phi_+(\varepsilon) + \beta_+(T) \quad (66)$$

We assume the entire HFE landscape, which will include all three functions $\hat{\psi}_i$, is symmetric about the zero-strain axis. This means that the resulting SMA model will exhibit tension-compression symmetry and this is expected in a 1-D single crystal model, were the two variants represent symmetric twins. The tension-compression asymmetry observed in some SMA experiments is understood to be the result of more complex local stress states and variant combinations. This effect is addressed in Chapter 5, where the model is extended to account for tension-compression asymmetry.

The only difference between the two martensite energy functions is the sign of their strain offsets (one is offset to the compressive side and one to the tensile side). Considering that the chemical free energies of the two martensite wells are equal we use a common parameter "m", which signifies unspecified martensite.

$$\beta_-(T) = \beta_+(T) = \beta_m(T) \quad (67)$$

The energies shown in Figure 37 represent true, absolute energies and this requires definition of a zero-energy state. Knowledge of the zero energy is not necessary, so we choose to use a relative form of the HFE. The relative energy ψ is defined by subtracting the martensite chemical free energy (CFE) from the total energy $\hat{\psi}$ for each phase.

$$\psi_-(\varepsilon) = \hat{\psi}_-(\varepsilon,T) - \beta_m(T) = \phi_-(\varepsilon) \quad (68)$$

$$\psi_a = \hat{\psi}_a(\varepsilon,T) - \beta_m(T) = \phi_a(\varepsilon) + \beta(T) \qquad where \qquad \beta(T) = \beta_a(T) - \beta_m(T) \quad (69)$$

$$\psi_+(\varepsilon) = \hat{\psi}_+(\varepsilon,T) - \beta_m(T) = \phi_+(\varepsilon) \quad (70)$$

Now, only the austenite energy retains a chemical free energy term and we adopt the parameter β with no subscript. Figure 38 shows the relative free energies where the martensite wells are fixed and the austenite well moves up or down with temperature.

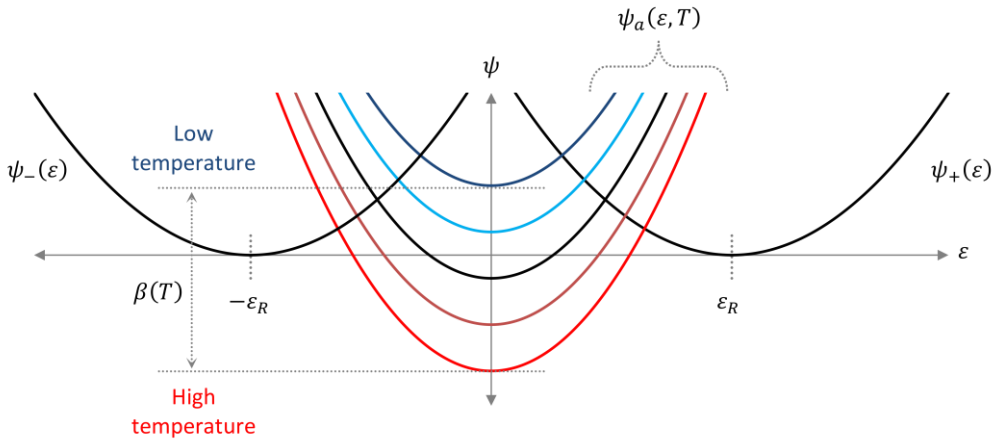


Figure 38: Relative Helmholtz free energies at various temperatures

From this point forward, we only consider the relative energies drop the designation *relative*. The strains associated with the minimum martensite energies are designated by $-\varepsilon_R$ and ε_R. ε_R is the magnitude of the residual strain and can be obtained directly from a stress strain curve as shown in Figure 35. Now, considering each energy well's vertical and horizontal offset, the parabolas can be given the following forms.

$$\psi_-(\varepsilon) = A_m(\varepsilon + \varepsilon_R)^2 \quad (71) \qquad \psi_a(\varepsilon) = A_a\varepsilon^2 + \beta(T) \quad (72) \qquad \psi_+(\varepsilon) = A_m(\varepsilon - \varepsilon_R)^2 \quad (73)$$

From Figure 35, we also know that each of the functions ψ_i is associated with a linear elastic branch of the stress-strain curve where the elastic modulus can be measured. Equation 62 relates the elastic modulus to the free energy functions and this allow us to define the A_m and A_a.

$$\frac{\partial^2\psi_a}{\partial\varepsilon^2} = E_a \quad \rightarrow \quad A_a = \frac{E_a}{2} \tag{74}$$

$$\frac{\partial^2\psi_+}{\partial\varepsilon^2} = E_m \quad \rightarrow \quad A_m = \frac{E_m}{2} \tag{75}$$

The three free energies are now written as follows.

$$\psi_-(\varepsilon) = \frac{E_m}{2}(\varepsilon + \varepsilon_R)^2 \quad (76) \qquad \psi_a(\varepsilon) = \frac{E_a}{2}\varepsilon^2 + \beta(T) \quad (77) \qquad \psi_+(\varepsilon) = \frac{E_m}{2}(\varepsilon - \varepsilon_R)^2 \quad (78)$$

Despite the fact that the chemical free energy has yet to be defined, Equations 71-73 can be used (via Equation 62) to define general stress-strain relationships for the layers in each phase.

$$\frac{\partial\psi_-}{\partial\varepsilon_-} = \sigma = E_m(\varepsilon_- + \varepsilon_R) \quad \rightarrow \quad \varepsilon_-(\sigma) = \frac{\sigma}{E_m} - \varepsilon_R \tag{16 *Repeated here*}$$

$$\frac{\partial\psi_a}{\partial\varepsilon_a} = \sigma = E_a\varepsilon_a \quad \rightarrow \quad \varepsilon_a(\sigma) = \frac{\sigma}{E_a} \tag{17 *Repeated here*}$$

$$\frac{\partial\psi_+}{\partial\varepsilon_+} = \sigma = E_m(\varepsilon_+ - \varepsilon_R) \quad \rightarrow \quad \varepsilon_+(\sigma) = \frac{\sigma}{E_m} + \varepsilon_R \tag{18 *Repeated here*}$$

Equations 16-18 were initially presented in Chapter 2 and represent a unique stress-strain relationship for each phase – they relate the average stress to the average strain for the layers in each particular energy well. At this point we only have one unknown parameter, the chemical free energy $\beta(T)$, but we still have to connect the energy wells. We do this by introducing concave energy barriers ψ_-^B and ψ_+^B as shown in Figure 39.

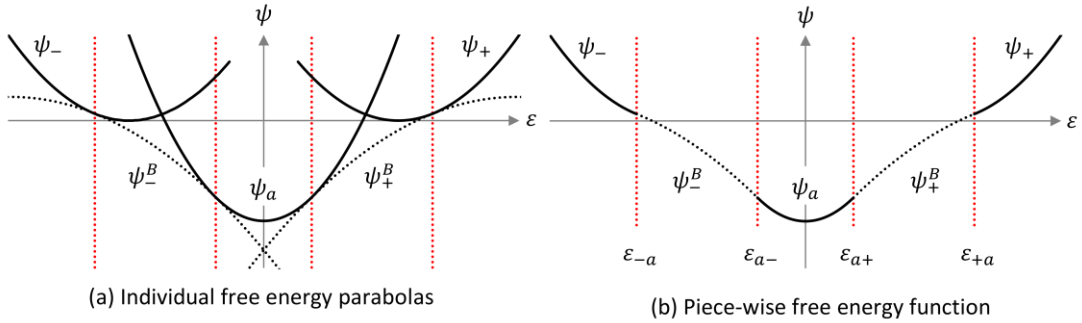


Figure 39: Piece-wise construction of Helmholtz free energy

Figure 39(a) shows the geometric arrangement of the five individual parabolas but they form the piecewise continuous landscape in 39(b). The five parabolas are connected at four strain locations and these are the critical strains, which were given by Equations 4-5. The energy barriers are also parabolic and are written as follows, where A, B, and C are added to β in the list of coefficients that need to be determined.

$$\psi_-^B(\varepsilon,T) = A(T)\varepsilon^2 - B(T)\varepsilon + C(T) \tag{79}$$

$$\psi_+^B(\varepsilon,T) = A(T)\varepsilon^2 + B(T)\varepsilon + C(T) \tag{80}$$

The fundamental principle in defining the energy barriers is a requirement that the HFE landscape be C1 continuous at all temperatures. This allows a set of four equations relating A, B, C, and β to be written in terms of $\varepsilon_{ij}(T)$ (which are functions of the transformation stresses $\sigma_{ij}(T)$). Figure 40 shows the energy barrier ψ_+^B smoothly connecting the ψ_a well to the ψ_+ well at three different temperatures.

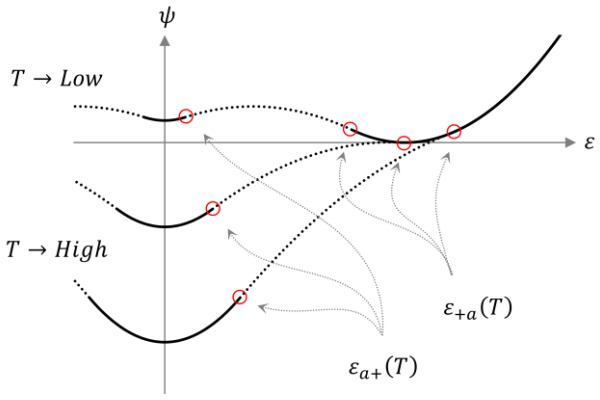


Figure 40: Strain locations for C1 continuity conditions on the HFE

The continuity conditions at the critical strain locations yield the following four equations.

$$\psi_a\big(\varepsilon_{a+}(T)\big) = \psi_+^B(\varepsilon_{a+}(T),T) \qquad \psi_+^B(\varepsilon_{+a}(T),T) = \psi_+(\varepsilon_{+a}(T),T) \tag{81,82}$$

$$\frac{\partial\psi_a}{\partial\varepsilon}\big(\varepsilon_{a+}(T)\big) = \frac{\partial\psi_+^B}{\partial\varepsilon}(\varepsilon_{a+}(T),T) \qquad \frac{\partial\psi_+^B}{\partial\varepsilon}(\varepsilon_{+a}(T),T) = \frac{\partial\psi_+}{\partial\varepsilon}(\varepsilon_{+a}(T),T) \tag{83,84}$$

Simultaneous solution of Equations 81-84 and some algebraic manipulation yield four independent equations which are all functions of temperature.

$$A(T) = \frac{\sigma_\Delta}{2\left(\varepsilon_{a+}(T) - \varepsilon_{+a}(T)\right)} \qquad B(T) = \frac{\sigma_{+a}(T)\varepsilon_{a+}(T) - \sigma_{a+}(T)\varepsilon_{+a}(T)}{\varepsilon_{a+}(T) - \varepsilon_{+a}(T)}$$

$$C(T) = \beta(T) + \frac{\varepsilon_{a+}(T)B(T)}{2} \qquad \beta(T) = -\frac{\sigma_{a+}(T)\varepsilon_{+a}(T) + \sigma_{+a}(T)\left(\varepsilon_R - \varepsilon_{a+}(T)\right)}{2} \tag{10-13}$$

Repeated here

It should be noted that Equations 10-13 contain the parameters σ_Δ, σ_{+a}, and ε_{ij} even though they are all derived from the forward transformation stress $\sigma_{a+}(T)$. These quantities can be written completely in terms of σ_{a+}, but the forms given here yield the most concise expressions. Equations 10-13 also represent the final expressions required to define the Helmholtz free energy (Equation 1). Because the transformation stress $\sigma_{a+}(T)$ plays such an important role in the free enery, it deserves some further exploration. The linear $\sigma_{a+}(T)$ (Equation 2) is common in SMA modeling but, in the next section, we derive it directly from the free energy instead of assuming its form.

4.3 Temperature Dependence

In this section, we use the mesoscopic Helmholtz free energy function to derive a relationship between the transformation stress and temperature. This exercise reveals some valuable information about how the transformation stress, which is can easily be thought of as a purely mechanical parameter, is closely related to the caloric properties of the material. We begin by with a useful diagram relating several mesoscopic thermodynamic potentials in Figure 41.

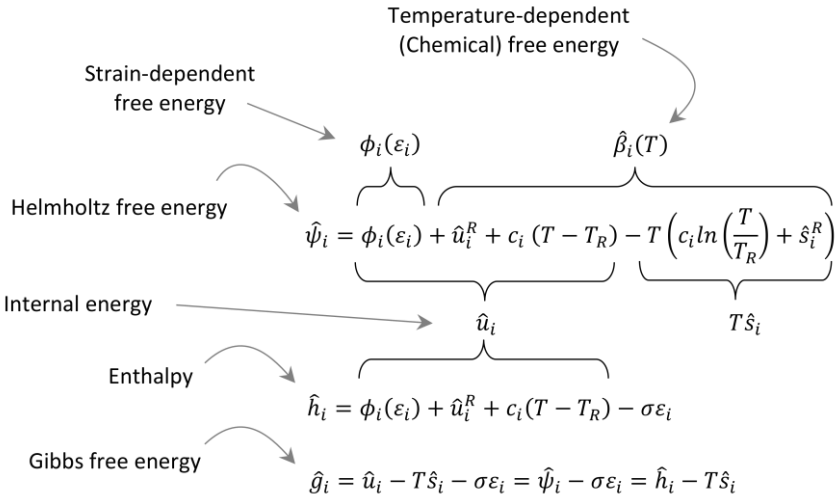


Figure 41: Mesoscopic thermodynamic potentials

This sketch shows and relationships between the thermodynamic potentials for material in phase i where: u_i^R is an internal energy constant, s_i^R is an entropy constant, c_i is the specific heat, and T_R is a reference temperature for the specific heat coefficient. We also show that Helmholtz free energy can be segregated into internal energy and entropy parts or strain-dependent and temperature-dependent parts. The temperature-dependent component, often called the chemical free energy, is given by Equation 85.

$$\beta_i(T) = u_i^R + c_i(T - T_R) - Tc_i ln\left(\frac{T}{T_R}\right) - Ts_i^R \tag{85}$$

This equation represents an absolute energy quantity whereas the $\beta(T)$ in the previous section (Equation 12) was defined as a relative energy between the austenite and martensite phases. To relate these two equations, we

must write a relative form of Equation 85. First, based on the symmetry of the HFE we regard the internal energy and entropy constants of the two martensite phases as equivalent and use the subscript m in place of – or +.

$$u_-^R = u_+^R = u_m^R \qquad s_-^R = s_+^R = s_m^R \tag{86,87}$$

We also follow the convention used for β and define relative quantities for internal energy and entropy constants with no phase specification.

$$u_a^R - u_m^R = u^R \qquad s_a^R - s_m^R = s^R \tag{88,89}$$

The R superscript is maintained to designate that these quantities are reference constants to distinguish them from their total counter parts u and s. Finally, we assume that the specific heats are equal for all three phases ($c_- = c_a = c_+$) and arrive at a linear relationship between $\beta(T)$ and temperature.

$$\beta(T) = \beta_a(T) - \beta_m(T) = u^R - s^R T \tag{90}$$

To relate the $\beta(T)$ from Equation 12 to Equation 90, we first rewrite Equation 12 completely in terms of the forward transformation stress $\sigma_{+a}(T)$.

$$\beta(T) = -\frac{\sigma_{a+}(T)\varepsilon_{+a}(T) + \sigma_{+a}(T)\big(\varepsilon_R - \varepsilon_{a+}(T)\big)}{2} \qquad \rightarrow \qquad Substitute\ \varepsilon_{ij}(T) = f(\sigma_{ij}(T))$$

$$\beta(T) = -\frac{1}{2}\left(\sigma_{a+}(T)\left(\frac{\sigma_{+a}(T)}{E_m} + \varepsilon_R\right) + \sigma_{+a}(T)\left(\varepsilon_R - \frac{\sigma_{a+}(T)}{E_a}\right)\right) \qquad \rightarrow \qquad Substitute\ \sigma_{+a}(T) = f(\sigma_{a+}(T))$$

$$\beta(T) = \frac{1}{2}\left(\frac{1}{E_a} - \frac{1}{E_m}\right)\sigma_{a+}(T)^2 + \frac{\sigma_\Delta}{2}\left(\frac{1}{E_m} - \frac{1}{E_a}\right)\sigma_{a+}(T) - \varepsilon_R\sigma_{a+}(T) + \frac{\sigma_\Delta\varepsilon_R}{2} \tag{91}$$

Equation 91 is quadratic in the space of transformation stress σ_{a+} and we rewrite it using temporary coefficients C_0, C_1, and C_2 to aid in the clarity as well as brevity of expressions.

$$\beta\big(\sigma_{a+}(T)\big) = C_2\sigma_{a+}(T)^2 + C_1\sigma_{a+}(T) + C_0 \qquad where \tag{92}$$

$$C_0 = \frac{\sigma_\Delta\varepsilon_R}{2} \qquad C_1 = \frac{\sigma_\Delta}{2}\left(\frac{1}{E_m} - \frac{1}{E_a}\right) - \varepsilon_R \qquad C_2 = \frac{1}{2}\left(\frac{1}{E_a} - \frac{1}{E_m}\right) = -\frac{E_\Delta}{2}$$

The parameter E_Δ represents the difference in stiffness between the phases, appears throughout the chemical free energy-related (i.e. caloric) expressions, and is given by Equation 93.

$$E_\Delta = \frac{1}{E_m} - \frac{1}{E_a} \tag{93}$$

Now we have two separate expressions for $\beta(T)$ (Equations 90 and 92) along with three unknown quantities: u^R, s^R, and $\sigma_{a+}(T)$. To proceed, we refer back to the experimental data (Figure 35) and recall that we know the transformation stress at two temperatures (T_L and T_H) regardless of the form $\sigma_{a+}(T)$ ends up taking.

$$\sigma_{a+}(T_L) = \sigma_{a+}^L \qquad \sigma_{a+}(T_H) = \sigma_{a+}^H \tag{94,95}$$

This means we can define the chemical free energy $\beta(T)$ at those two temperatures as well.

$$\beta_L = \beta\big(\sigma_{a+}(T_L)\big) \qquad \rightarrow \qquad \beta_L = C_2{\sigma_{a+}^L}^2 + C_1\sigma_{a+}^L + C_0 \tag{96}$$

$$\beta_H = \beta\big(\sigma_{a+}(T_H)\big) \qquad \rightarrow \qquad \beta_H = C_2{\sigma_{a+}^H}^2 + C_1\sigma_{a+}^H + C_0 \tag{97}$$

Now, we can write two instances of Equation 90, one at each temperature, and this gives us two expressions for the internal energy constant u^R.

$$\beta_L = u^R - s^R T_L \quad \rightarrow \quad u^R = \beta_L + s^R T_L \tag{98}$$

$$\beta_H = u^R - s^R T_H \quad \rightarrow \quad u^R = \beta_H + s^R T_H \tag{99}$$

Since u^R is a constant, we can equate the right-hand sides of Equations 98 and 99 and solve for the internal energy constant s^R.

$$\beta_L + s^R T_L = \beta_H + s^R T_H \quad \rightarrow \quad s^R T_H - s^R T_L = \beta_L - \beta_H \quad \rightarrow \quad s^R = \frac{\beta_L - \beta_H}{T_H - T_L} = -\frac{d\beta}{dT} \tag{100}$$

Equation 100 implies that the entropy constant is the rate of change of the chemical free energy with respect to temperature and this is expected from Equation 62 ($\frac{\partial \psi}{\partial T} = -s$). Knowing s^R, we can return to either Equation 98 or 99 to solve for u^R. With the internal energy and entropy constants defined we can now solve for the transformation stress by substituting Equation 90 for $\beta(T)$ in Equation 92, moving all terms to the right-hand side, and using the quadratic formula.

$$\beta(T) = u^R - s^R T = C_2 \sigma_{a+}(T)^2 + C_1 \sigma_{a+}(T) + C_0 \quad \rightarrow \quad 0 = C_2 \sigma_{a+}(T)^2 + C_1 \sigma_{a+}(T) + C_0 - u^R + s^R T$$

$$\sigma_{a+}(T) = \frac{-C_1 \pm \sqrt{{C_1}^2 - 4C_2(C_0 - u^R + s^R T)}}{2C_2} \tag{101}$$

The forward transformation stress always positive or zero therefore the solution with a positive square root term, which yields negative values, is disregarded. The negative square root term and a minimum condition give us an expression for the forward transformation stress as function of temperature.

$$\sigma_{a+}(T) = max\left(\frac{-C_1 - \sqrt{{C_1}^2 - 4C_2(C_0 - u^R + s^R T)}}{2C_2}, 0 \right) \tag{102}$$

If equation 102 is examined graphically it can be seen that the quadratic nature of the transformation stress is only observed across a temperature domain that is very large (0-3,000K) compared to the typical operating ranges of SMA tests and devices (200-400K). To show this, we again use the experimental tension test data from Chang et al. (2006). Figure 42(a) shows both the quadratic and linear form of the transformation stress plotted across a large range of temperatures. Figure 42(b), however, shows the same curves over a temperature scale that represents the typical operating range of SMA devices.

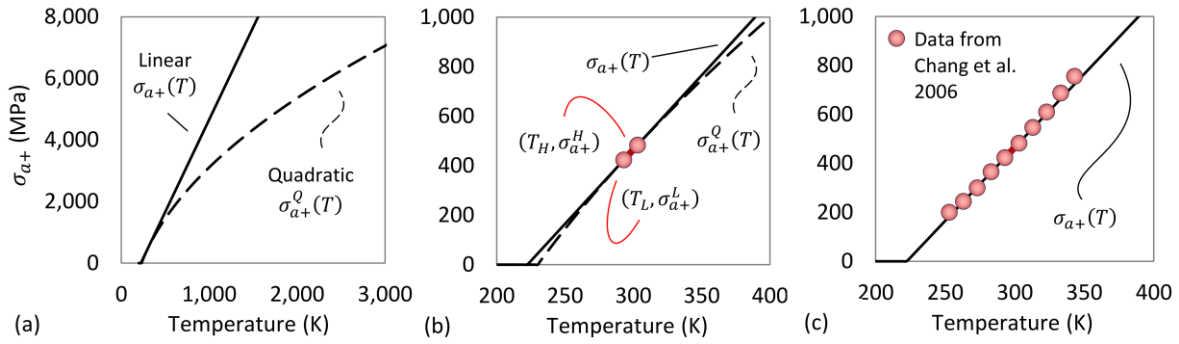


Figure 42: Quadratic and linear transformation stress plots

Figure 42(c) shows the actual transformation stresses measured from the experiments and we can see that the variation is very linear. Even though the quadratic function produces values that are nearly linear and yields indistinguishable model results, we use the simpler form given by Equation 2 for model operation.

$$\sigma_{a+}(T) = max\left(\sigma_{a+}^{L} + \left(\frac{\sigma_{a+}^{H} - \sigma_{a+}^{L}}{T_H - T_L}\right)(T - T_L), 0\right) \quad \text{(2 } \textit{Repeated here}\text{)}$$

In addition to highlighting additional relationships between the transformation stress and the free energy, this exercise also yields expressions for the internal energy and entropy constants. These are important quantities that will be used later in chapter to derive the latent heat parameters. Before proceeding to the latent heats and further manipulating the caloric parameters, a subtle but important note should be made about the decision to use the linear transformation stress vs the quadratic one.

4.3.1 Consequences of Using the Linear vs. the Quadratic Transformation Stress

When considered purely from the perspective of predicting the transformation stress, Equations 2 and 102 are interchangeable. But the choice to use the linear vs. quadratic expression has a few additional consequences related to the caloric properties and the smoothness of the free energy.

First, we recall that the expression relating the transformation stress to the chemical free energy is derived from the HFE continuity conditions (Equations 81-84). Furthermore, the HFE barrier functions connect smoothly to the energy well functions at the critical strains, which are functions of the transformation stress. If the quadratic expression for $\sigma_{a+}(T)$ is used, it must be maintained for all of expressions used to build the HFE. This includes the calculation of $\sigma_{+a}(T)$ (Equation 3), $\varepsilon_{a+}(T)$ and $\varepsilon_{+a}(T)$ (Equations 4-5), $A(T)$, $B(T)$, and $C(T)$ (Equations 10-12), as well as $\beta(T)$ (Equation 13). This will allow the free energy parabolas to connect smoothly as intended and the chemical free energy $\beta(T)$ will remain linear with temperature as assumed in Equation 90.

If the linear version $\sigma_{a+}(T)$ used, it must also be consistently be used throughout the set of HFE functions. The shapes of the energy barriers will change slightly, but the continuity conditions will hold and the five parabolas will still connect smoothly. This choice, however, forces us to relax the assumption that $\beta(T)$ is linear with temperature. When $\sigma_{+a}(T)$ becomes linear, $\beta(T)$ becomes quadratic. Fortunately, we find ourselves in the same situation that was shown in Figure 42. The difference between the linear and quadratic β's are only significant over a very large range of temperatures and it is reasonable to say that, when the linear transformation stress is used, that Equation 90 simply becomes an approximate relationship.

$$\beta(T) \approx u^R - s^R T \qquad \leftarrow \qquad When\ the\ linear\ \sigma_{a+}(T)\ is\ used \tag{103}$$

It should finally be noted, additionally, that the expressions for the internal energy and entropy constants (Equations 98, 99, and 100) are correct, regardless of which transformation stress equation is used. This is because they are derived from the two temperatures where the experimental transformation stresses are measured and treated as material constants. At these two temperatures, the linear and quadratic expressions intersect (as shown in Figure 42(b)) and the derived parameters are equivalent. These distinctions play a minor role in the overall model behavior, but it is important to maintain consistency between the equation sets because they are used further in the next section where we derive the latent expressions.

4.4 Caloric Properties

In this section we continue to manipulate the HFE and caloric parameters to derive the latent heat coefficients. This is a useful exercise for several reasons. The most salient benefit is that the only experimental data required to do this is the set of stress-strain curves that were used to build the free energy. Stress-strain data is typically more readily available than DSC data. This derivation is also insightful because it shows a unique feature of the M-A-S model, namely the way irreversibility is included in the latent heats. To highlight this, we

first neglect irreversibility and derive the latent heats in a form that is found in several other SMA models. Then, we include the irreversibility terms and repeat the derivation to arrive at a different expression. We conclude the section by testing the accuracy of the theory with comparisons to experimental measurements.

4.4.1 Reversible Case

The latent heat derivation begins with an expression for the relative minimum GFE between the austenite phase and either martensite phase.

$$F = G_a - G_+ \tag{104}$$

We refer to F as the *driving force* and it represents the impetus for layers to jump into a lower energy well. G_a and $G_\pm$ are were shown in Figure 9 and given by Equations 19-21. Substituting Equations 16-18 for $\varepsilon_i(\sigma)$ and some rearranging, Equation 104 yields the following expression for the driving force.

$$F = \frac{\sigma^2}{2}E_\Delta + \sigma\varepsilon_R + \beta(T) \tag{105}$$

In this example, we assume an SMA material that is reversible ($\sigma_\Delta = 0$) and where the elastic moduli are equal for all phases ($E_m = E_a \rightarrow E_\Delta = 0$). Under these circumstances, the driving force (Equation 105) is written as follows.

$$F = \sigma\varepsilon_R + \beta(T) \tag{106}$$

Substituting Equation 90 for $\beta(T)$ yields an expression in terms of stress, transformation strain, internal energy constant and the entropy constant.

$$F = \sigma\varepsilon_R + u^R - s^R T \tag{107}$$

We can also write Equation 104 in terms of relative enthalpy, entropy, and temperature by assuming that each G has the form $\hat{g} = \hat{h} - \widehat{sT}$.

$$F = G_a - G_+ = \hat{h}_a - \hat{h}_+ - (\hat{s}_a - \hat{s}_+)T \tag{108}$$

We recognize the entropy constant from Equation 89 and define relative enthalpy (h_+) in a similar manner.

$$F = h_+(\sigma) - s^R T \qquad where \qquad h_+(\sigma) = \hat{h}_a - \hat{h}_+ \tag{109}$$

Now, we substitute for driving force F from Equation 107 and solve for enthalpy.

$$\sigma\varepsilon_R + u^R - s^R T = h_+(\sigma) - s^R T \qquad \rightarrow \qquad h_+(\sigma) = \sigma\varepsilon_R + u^R \tag{110}$$

Equation 109 is a function of stress, residual strain (which is equal to transformation strain in the simplified case), and the relative internal energy constant. This expression has a stress-dependent component as well as a chemical component and is similar to the expressions found in Ortin et al. (1992), Wollants et al. (1993), and Shaw (2000). Continuing, we see that several other several other parameters have simpler forms in the reversible case. The chemical free energy expression (Equation 92) is also simplified as follows.

$$\beta(T) = C_2\sigma_{a+}(T)^2 + C_1\sigma_{a+}(T) + C_0 \quad \rightarrow \quad \begin{matrix} C_0 = \frac{\sigma_\Delta\varepsilon_R}{2} = 0 & C_2 = \frac{E_\Delta}{2} = 0 \\ C_1 = \frac{\sigma_\Delta}{2}\left(\frac{1}{E_m} - \frac{1}{E_a}\right) - \varepsilon_R = -\varepsilon_R \end{matrix} \quad \rightarrow \quad \beta(T) = -\varepsilon_R\sigma_{a+}(T) \tag{111}$$

The internal energy and entropy can be written as shown here.

$$s_R = \frac{\beta_L - \beta_H}{T_H - T_L} = \frac{-\varepsilon_R \sigma_{a+}^L + \varepsilon_R \sigma_{a+}^H}{T_H - T_L} = -\frac{d\sigma_{a+}}{dT}\varepsilon_R \qquad , \qquad u_R = \beta(T) + s_R T = s_R T - \varepsilon_R \sigma_{a+}(T) \tag{112,113}$$

Returning to Equation 109, we can substitute for s^R and express the driving force using Equation 114.

$$F = h_+(\sigma) - T\frac{d\sigma_{a+}}{dT}\varepsilon_R \tag{114}$$

Equations 109 and 114 give the driving force as a linear function of temperature where the enthalpy represents a stress-dependent vertical axis-intercept. A plot of diving force vs. temperature, as shown in Figure 43, illustrates the interrelationships between each of the parameters in Equations 106, 107 and 114.

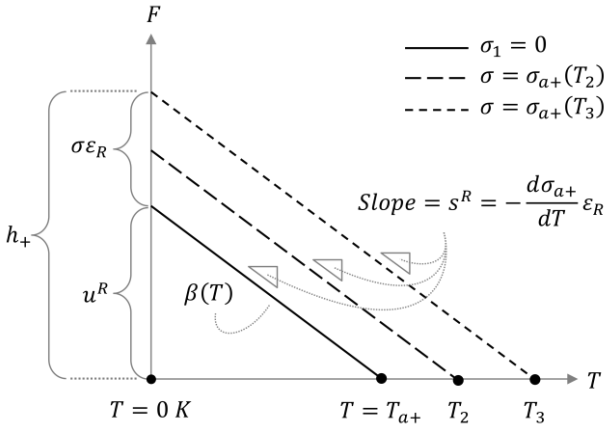


Figure 43: Driving force vs. temperature showing latent heat and CFE components

Because of Equation, 114, we can project F back to the zero-temperature axis to examine latent heat. Comparing the driving force curves at zero and nonzero stress distinguishes the stress free (chemical) latent heat from the stress-dependent (mechanical) component. We see that the chemical component of the latent heat is the internal energy constant, which is also the zero-temperature intercept of the chemical free energy. We also see that the zero-stress driving force line reaches zero at temperature T_{am}, which corresponds to a stress-free $a \rightarrow +$ transformation ($\sigma_{a+}(T_{am}) = 0$). This stems from a feature that is unique to the reversible case, where zero driving force is synchronized with the transformation stress (at any temperature above T_{am}). This can be shown mathematically by substituting Equation 111 for β in Equation 106.

$$F = \sigma\varepsilon_R + \beta(T) = \sigma\varepsilon_R - \varepsilon_R\sigma_{a+}(T) \qquad \rightarrow \qquad F = \varepsilon_R\big(\sigma - \sigma_{a+}(T)\big) \tag{115}$$

From Equation 115 we see that the driving force is equal to zero when the stress is equal to the transformation stress. This effect can also be observed in GFE as shown by Figure 44.

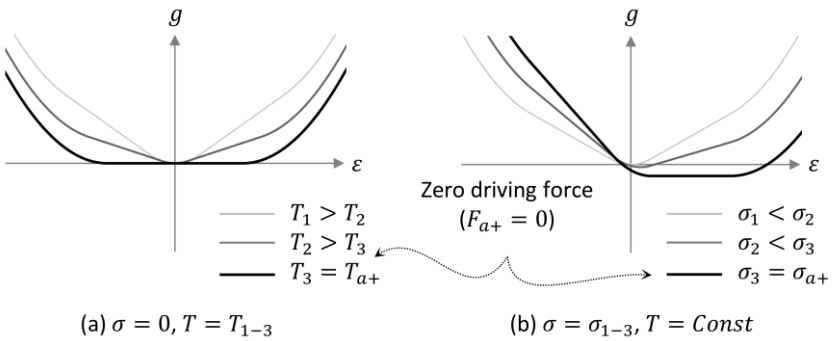


Figure 44: GFE vs. temperature (a) and stress (b) showing zero driving force with no energy barriers (reversible)

In Figure 44(a) we see the GFE of a reversible SMA under zero stress and at three different temperatures, including the martensite start temperature T_{am}. When the temperature reaches T_{am}, the transformation stresses (σ_{ij}) and the driving forces between the phases (F_{ij}) are zero. Figure 44(b) shows a constant temperature and three different levels of stress. Once again, when the stress is equal to transformation stress σ_{a+} (which is equal to σ_{+a}), the driving forces between the two phases are zero. In both a and b, the zero-driving force configuration represents a situation where there is no distinction between the phases. Layers in 44(a) would have equal probabilities of being in all three phases and in 44(b) they would equally distribute between austenite and tensile martensite. Both figures reflect the Maxwell definition of reversibility, that a small change in environment could reverse the direction of transformation.

If we specify Equation 114 to the conditions where the stress is equal to the transformation stress, it can be rearranged to take the form of the Clausius-Clapeyron equation.

$$F = 0 \quad \rightarrow \quad h_+\big(\sigma_{a+}(T)\big) = T\frac{d\sigma_{a+}}{dT}\varepsilon_R \tag{116}$$

This equation offers a convenient connection between caloric (h_+) and mechanical ($\frac{d\sigma_{a+}}{dT}$, ε_R) material parameters but is limited to the current, simplified, case where σ_Δ and $E_\Delta = 0$. Actual SMA materials are not reversible and usually do not have phases with equal stiffnesses.

4.4.2 General Case

Now, we repeat the same procedure for the general case where the irreversibility (σ_Δ) and the difference in stiffness between the phases (E_Δ) are maintained. Beginning with Equation 105, we substitute Equation 90 which expresses β as a function of the relative internal energy and entropy constants.

$$F = \frac{\sigma^2}{2}E_\Delta + \sigma\varepsilon_R + u^R - s^R T \tag{117}$$

Now, we substitute the right-hand side of Equation 109 for the driving force to get an expression for enthalpy that now includes an additional term.

$$h_+(\sigma) - Ts^R = F_{a+} = \frac{\sigma^2}{2}E_\Delta + \sigma\varepsilon_R + u^R - Ts^R \qquad \rightarrow \qquad h_+(\sigma) = \frac{\sigma^2}{2}E_\Delta + \sigma\varepsilon_R + u^R \tag{118}$$

Equation 118 calculates the martensite + latent heat coefficient that is used in the macroscopic energy balance (Equation 32). This expression includes additional irreversibility terms that the simplified form (Equation 110) does not (note that u^R, given by Equations 98 or 99, also contains additional terms). The expression for the martensite – latent heat coefficient is derived in a similar manner and is given by Equation 119.

$$h_-(\sigma) - Ts^R = F_{a-} = \frac{\sigma^2}{2}E_\Delta - \sigma\varepsilon_R + u^R - Ts^R \qquad \rightarrow \qquad h_-(\sigma) = \frac{\sigma^2}{2}E_\Delta - \sigma\varepsilon_R + u^R \tag{119}$$

Equations 118 and 119 were shown in Chapter 2 (bundled together in Equation 33) – these expressions were originally introduced in Seelecke and Müller (2004). Here, they have been derived in a different manner in order to show their consistency with other SMA models (in the simplified case) and for closer examination of the full, irreversible, form. For comparison, we now write and expression that is analogous to Equation 114 for the driving force.

$$F = h_+(\sigma) - T\frac{d\sigma_{a+}}{dT}\varepsilon_R - T\begin{pmatrix}Additional\ \sigma_\Delta^\pm \\ and\ E_\Delta^\pm\ terms\end{pmatrix} \tag{120}$$

Only the first three terms are shown for Equation 120 because the full expression is lengthy and not useful apart from the current exercise. Here, we use this expression and coefficients obtained from Chang et al. (2006) to produce a plot similar to Figure 43, showing the driving force vs. temperature.

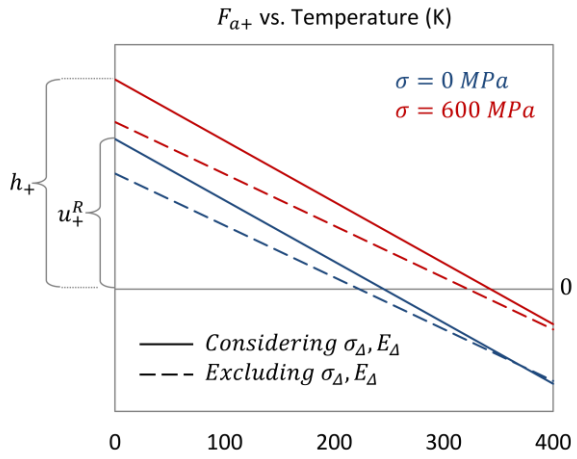


Figure 45: Driving force vs. temperature plot for both the simplified and general case

Figure 45 shows two sets of curves, one including and one excluding the σ_Δ and E_Δ coefficients. For the sets that *excluded* σ_Δ and E_Δ, we simply averaged the forward and reverse transformations stresses to get a single value and did the same for the two elastic modulus values. This allows us to compare the simplified case to the irreversible case where everything else is equal. The plot shows that the irreversibility terms influence both the driving force and the latent heat. The stress-free cases show that roughly 20% of the relative internal energy $u_\pm^R$ (and therefore the CFE) is due to the irreversibility and stiffness-difference parameters. The difference is less pronounced, but the stress-dependent component of the latent heat also increases in the general case due to the additional term in Equation 118. The additional terms in Equation 112 (the relative entropy constant) are illustrated by the differences in slope between the simplified and general cases. Another import difference between the simplified and general cases is the meaning of zero-driving force. The general case considers energy barriers which means that a driving force of zero no longer represents conditions for transformation as illustrated by Figure 46.

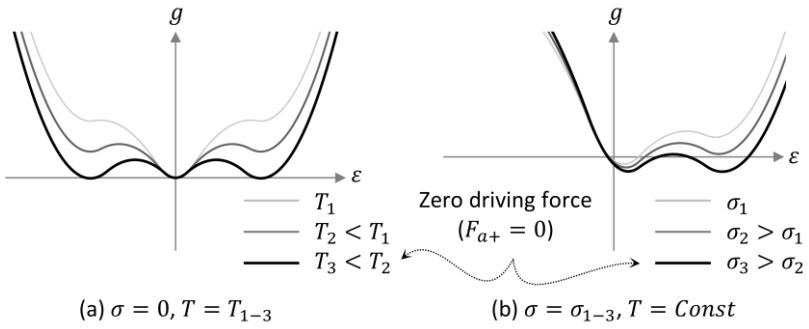


Figure 46: GFE vs. temperature (a) and stress (b) showing zero driving force with energy barriers (irreversible)

In Figure 46(a), the temperature decreases to the point where the driving forces between the phases are all equal to zero but, due to the presence of energy barriers, layers in any of the wells would remain trapped there. The decrease in temperature would have to be continued to eliminate the barriers and allow transformation. 46(b) shows a series of stress states at the same temperature. The highest stress yields zero driving force between

austenite and tensile martensite but, once again, transformation does not occur and the stress would have to be pushed beyond the Maxwell stress to free layers from their local energy wells.

If we continue the procedure from the simplified case, we can still use Equation 120 and the $F = 0$ condition to isolate the latent heat in an expression motivated by the Clausius-Clapeyron equation. The zero-driving force state (the Maxwell stress) is now described by quadratic function of temperature derived from Equation 117, but can be reasonably approximated as the average of the forward and reverse transformation stresses. This allows us to eliminate F from the expression, but we see that the additional terms are still present and the traditionally simple form of the Clausius-Clapeyron equation is not recovered.

$$h_+(avg(\sigma_{a+}(T), \sigma_{+a}(T))) = T\frac{d\sigma_{a+}}{dT}\varepsilon_R + T\begin{pmatrix}Additional\ \sigma_\Delta \\ and\ E_\Delta\ terms\end{pmatrix} \tag{121}$$

For the model operation, Equation 118 is better choice to calculate the latent heat coefficients because it applies to any stress state and does not require calculation of the zero-driving force stress.

4.4.3 Comparison to Experimental Data

SMA modelers who are considering thermomechanical coupling often rely on DSC measurements to assign values to latent heat parameters. Several expressions in this section can be used to define latent coefficients from mechanically measured parameters. This is appealing because they could be used in cases where calorimetric measurements are not available. To evaluate the accuracy of these calculations, we show a small sample of latent heat values obtained from DSC measurements in Table 2 (top row) along with predictions using Equation 118 (second row). We also include values calculated from the simplified case (Equation 116) in the third row for comparison purposes.

Data source	Chang et al 2006	Schmidt et al 2015	
		Untrained	Trained
DSC latent heat measurement $\left(J/g\right)$	11.0 $(a \to m)$ 15.0 $(m \to a)$	12.3 $(a \to m)$ 11.8 $(m \to a)$	9.1 $(a \to m)$ 10.0 $(m \to a)$
$h_+(\sigma = 0)$	14.1	12.5	7.5
$\frac{d\sigma}{dT} \cdot \varepsilon_R \cdot T$	11.7	7.3	4.6

Table 2: Latent heat calculations vs. DSC measurements

The three data sets were chosen because they included tension test data at multiple temperatures along with the DSC measurements, but several details should be discussed before comparing results. The Chang et al. 2006 data set reflects NiTi and the Schmidt et al. 2015 values correspond to NiTiCuV, which is an SMA material studied for use in elastocaloric cooling devices (Schmidt et al. (2015a), Schmidt et al. (2015b), Schmidt et al. (2016a), Schmidt et al. (2016b), Schmidt et al. (2016c)). It should also be pointed out that the $a \to m$ transformation in Chang at al. 2006 was signified by a two-stage process indicating an intermediate R-phase; the first step released 5 $\frac{J}{g}$ and the second step 6 $\frac{J}{g}$. The current model regards the transformation from austenite to martensite as a single-step processes. The thermodynamic calculations compare quantities before and after complete transformation and are not affected by the presence of intermediate configurations. In this light, the latent heat value of 11. 0 $\frac{J}{g}$ shown in the table is the sum of the two measured values – i.e. the total latent heat released between the initial and final states. Two data sets are shown under Schmidt et al. 2015 because the

material was tested in its as-received condition (untrained) and the again after being subjected to 100 tension cycles (trained) to stabilize the behavior. The NiTi material in Chang et al. 2006 was only tested in its untrained state. The experimental measurement row contains two values per experiment because different values were measured for $a \rightarrow m$ and $m \rightarrow a$ transformations. We speculate that these differences may be due to multi-variant effects that would only be observed in stress-free transformations. In the second and third row, only a single value is shown for each experiment because neither formula yields different latent heats for stress-free forward and reverse transformations. We regard each set of experimental values as a range where we expect the corresponding calculation to fall. The Clausius-Clapeyron calculation in the third row has a temperature parameter and one might be tempted to use this to predict two different stress-free values, one at T_{a+} and one at T_{+a}. This would be inconsistent with the reversibility assumptions that were used to derive the equation, so we use the average of the two temperatures in each calculation.

Comparing the values in the table, the first observation is that all three $h_+(\sigma = 0)$ values are reasonable – within 22% of the average of the two experimental values (one is within 3%). In the Chang et al. 2006 case, both the h_+ and the Clausius-Clapeyron calculation yield values that fall between the two experimental measurements and thus no argument can be made that one is better than the other. In both of the Schmidt et al. 2015 cases however, the h_+ value is notably closer to the experimental results. One possible explanation is that the NiTiCuV and NiTi have similar stiffnesses and transformation stresses, but NiTiCuV has a smaller transformation strain (2.5%). In the case where the transformation strain value is lower, the σ_Δ and E_Δ terms in the second row have a larger influence and cause a greater deviation from the third-row calculations. In all three cases, the latent heat properties calculated from stress-strain data could reasonably be used if DSC data were not available.

4.5 Entropy Production

A number of SMA models write a Clausius-Planck (or Clausius-Duhem) inequality as a platform to isolate and add entropy production terms (dissipation potentials) to the system free energy. The entropy producing terms are typically those associated with inelastic deformations, which include most of the noteworthy SMA effects. The current model is based on a mesoscopic free energy that does not contain any explicit entropy-production terms. In this section we show that entropy production can be calculated from the current model's stress-strain equation in a manner that is consistent with the Clausius-Planck equation as well as other model quantities.

The sketch in Figure 47 represents a thermodynamic process where a superelastic SMA wire is loaded and unloaded in a strain-controlled manner. The strain is applied at a rate that is constant and sufficiently slow such that any heat is released /absorbed to/from the environment with no change in temperature.

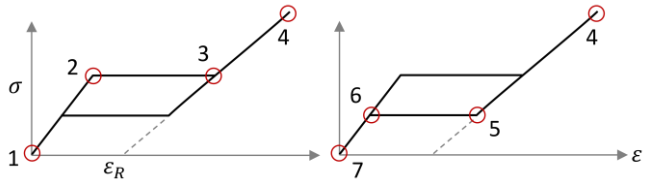


Figure 47: Superelastic isothermal SMA load-unload process

We examine the entropy production in terms of the Clausius-Planck equation which is a form of the Clausius-Duhem equation accounting for the fact that heat always flows from hot to cold (Holzapfel (2000)). For a macroscopic material point in a 1-D solid, the rate of entropy production per unit volume is given by Equation 122.

$$D_{int}(t) = \sigma\dot{\varepsilon} - \dot{\Psi} - s\dot{T} \geq 0 \qquad where \tag{122}$$

$D_{int}(t) \rightarrow Rate\ of\ internal\ disspation\ (entropy\ production)$

$\dot{\Psi} \rightarrow Rate\ of\ change\ of\ Helmholtz\ free\ energy$

$\sigma\dot{\varepsilon} \rightarrow Stress\ power$

$s\dot{T} \rightarrow Rate\ of\ entropy\ production\ due\ to\ heat\ transfer$

To find the total entropy production (D_{int}^{Tot}) for the process shown in Figure 47 we integrate from time t_1 to time t_7. Note that the $s\dot{T}$ term has been dropped because the process is isothermal.

$$D_{int}^{Tot} = \int_{t_1}^{t_7} D_{int}\, dt = \sum_{n}^{1..6} D_{int}^{n,n+1} = \sum_{n}^{1..6} \int_{t_n}^{t_{n+1}} \left(\sigma\dot{\varepsilon} - \dot{\Psi}\right) dt \tag{123}$$

The uppercase Ψ represents the macroscopic HFE per unit volume and this is calculated in a similar manner to the macroscopic strain, as a weighted average of the mesoscopic energy densities.

$$\Psi(\sigma, T, x_i) = \sum_{i\in -,a,+} \psi_{\mathrm{i}}(\varepsilon_{\mathrm{i}}(\sigma), T) x_{\mathrm{i}} \tag{124}$$

$\Psi \rightarrow Macroscopic\ HFE\ per\ unit\ volume$

$\psi_i \rightarrow Mesoscopic\ HFE\ per\ unit\ volume\ for\ phase\ i$

$\varepsilon_i(\sigma) \rightarrow Average\ strain\ for\ phase\ i\ (function\ of\ stress\ \sigma)$

$x_i \rightarrow Macroscopic\ fraction\ of\ material\ in\ phase\ i\ (phase\ fractions)$

We rewrite Equation 124 by inserting Equations 16-18 for $\varepsilon_i(\sigma)$ and Equation 1 for $\psi_i(\varepsilon, T)$. Collecting terms and simplifying we arrive at an expression for the total HFE in terms of stress, temperature, and phase fractions.

$$\Psi(\sigma, T, x_i) = \frac{\sigma^2}{2}\left(\frac{x_-}{E_m} + \frac{x_a}{E_a} + \frac{x_+}{E_m}\right) + \beta(T)x_a \tag{125}$$

The time derivative of the Ψ is expanded as follows and selected terms that will be used in the evaluation of Equation 123 are expanded (according to Equation 125) for reference.

$$\dot{\Psi} = \frac{\partial\Psi}{\partial\sigma}\dot{\sigma} + \frac{\partial\Psi}{\partial T}\dot{T} + \frac{\partial\Psi}{\partial x_-}\dot{x}_- + \frac{\partial\Psi}{\partial x_a}\dot{x}_a + \frac{\partial\Psi}{\partial x_+}\dot{x}_+ \qquad where \tag{126}$$

$$\frac{\partial\Psi}{\partial\sigma} = \sigma\left(\frac{x_-}{E_m} + \frac{x_a}{E_a} + \frac{x_+}{E_m}\right) \quad , \quad \frac{\partial\Psi}{\partial x_a} = \frac{\sigma^2}{2E_a} + \beta(T) \quad , \quad \frac{\partial\Psi}{\partial x_+} = \frac{\sigma^2}{2E_m} \quad , \quad \dot{T} = \dot{x}_- = 0$$

Note that this loading cycle is assumed to be isothermal and only involves tensile loading therefore temperature T and the compressive martensite phase fraction x_- never change. The time derivative of the macroscopic strain is given by Equation 127 where the mesoscopic strains are regarded as functions of stress according to Equations 16-18. Once again, selected terms that will be used in the evaluation of Equation 123 are also included for reference.

$$\dot{\varepsilon} = \dot{\varepsilon}_-(\sigma)x_- + \dot{\varepsilon}_a(\sigma)x_a + \dot{\varepsilon}_+(\sigma)x_+ + \varepsilon_-(\sigma)\dot{x}_- + \varepsilon_a(\sigma)\dot{x}_a + \varepsilon_+(\sigma)\dot{x}_+ \qquad where \tag{127}$$

$$\varepsilon_a(\sigma) = \frac{\sigma}{E_a} \quad , \quad \varepsilon_+(\sigma) = \frac{\sigma}{E_m} + \varepsilon_R \quad , \quad x_- = 0 \quad , \quad \dot{\varepsilon}_a(\sigma) = \frac{\dot{\sigma}}{E_a} \quad , \quad \dot{\varepsilon}_+(\sigma) = \frac{\dot{\sigma}}{E_m} \quad , \quad \dot{x}_- = 0$$

Having established Equations 124, 125, 126, and 127 for reference, we turn our attention to solving Equation 123 for D_{int}^{Tot}. We begin with the first segment of time, $t_1 \rightarrow t_2$. Since this loading occurs at a superelastic temperature, we know that the austenite phase fraction (x_a) is equal to one through the entire step. Based on this, we use Equations 127 and 126 to define $\dot{\varepsilon}$ and $\dot{\Psi}$ during time segment 1.

$$\dot{\varepsilon} = \dot{\varepsilon}_a(\sigma)x_a = \frac{\dot{\sigma}}{E_a} \quad , \quad \Psi = \frac{\partial \Psi}{\partial \sigma}\dot{\sigma} = \sigma\left(\frac{x_a}{E_a}\right)\dot{\sigma} = \sigma\frac{\dot{\sigma}}{E_a}$$

Substituting these into the first integral from Equation 123, we that the entropy production for the first loading branch is equal to zero. This is expected because, prior to transformation, the material behaves in a linear elastic manner.

$$D_{int}^{1,2} = \int_{t_1}^{t_2} D_{int}\, dt = \int_{t_1}^{t_2} \left(\sigma\dot{\varepsilon} - \Psi\right) dt = \int_{t_1}^{t_2} \left(\sigma\frac{\dot{\sigma}}{E_a} - \sigma\frac{\dot{\sigma}}{E_a}\right) dt = 0 \tag{128}$$

We forgo the solutions of Equation 123 over the other elastic steps in Figure 47 and accept the following as true.

$$t_1 \to t_2 \quad \left\{\begin{matrix} Linear\ elastic \ , \ x_a = 1 \\ No\ phase\ change \end{matrix}\right\} \ \to D_{int} = 0 \qquad t_4 \to t_5 \quad \left\{\begin{matrix} Linear\ elastic \ , \ x_+ = 1 \\ No\ phase\ change \end{matrix}\right\} \ \to D_{int} = 0$$

$$t_2 \to t_3 \quad \left\{\begin{matrix} \sigma = \sigma_{a+} \ , \ x_a = 1 \to 0 \\ x_+ = 0 \to 1 \end{matrix}\right\} \ \to D_{int} \neq 0 \qquad t_5 \to t_6 \quad \left\{\begin{matrix} \sigma = \sigma_{+a} \ , \ x_a = 0 \to 1 \\ x_+ = 1 \to 0 \end{matrix}\right\} \ \to D_{int} \neq 0$$

$$t_3 \to t_4 \quad \left\{\begin{matrix} Linear\ elastic \ , \ x_+ = 1 \\ No\ phase\ change \end{matrix}\right\} \ \to D_{int} = 0 \qquad t_6 \to t_7 \quad \left\{\begin{matrix} Linear\ elastic \ , \ x_a = 1 \\ No\ phase\ change \end{matrix}\right\} \ \to D_{int} = 0$$

Equation 123 now has only two non-zero terms which correspond to steps $t_2 \to t_3$ and $t_5 \to t_6$. We also know that the stress is constant during both of these steps so we can rewrite the expression as shown here.

$$D_{int}^{Tot} = D_{int}^{2,3} + D_{int}^{5,6} = \int_{t_2}^{t_3} \left(\sigma_{a+}\dot{\varepsilon} - \Psi\right) dt + \int_{t_5}^{t_6} \left(\sigma_{+a}\dot{\varepsilon} - \Psi\right) dt \tag{129}$$

We begin with the first term, D_{int}^{23}. The time derivative of ε is written according to Equation 127 where the value of stress is equal to σ_{a+}.

$$\dot{\varepsilon} = \varepsilon_a(\sigma)\dot{x}_a + \varepsilon_+(\sigma)\dot{x}_+ = \frac{\sigma_{a+}}{E_a}\dot{x}_a + \frac{\sigma_{a+}}{E_m}\dot{x}_+ + \varepsilon_R\dot{x}_+ \tag{130}$$

We know that the macroscopic strain is applied at a constant rate and that stress is constant during this step. The phase fractions evolve at a constant rate during this step to keep the right-hand side of Equation 130 equal to the left. We also know that x_a progresses from $1 \to 0$ and that x_+ progresses from $0 \to 1$ so we can write expressions for their rates of change.

$$\dot{x}_a = \frac{-1}{\Delta t_{23}} \qquad , \qquad \dot{x}_+ = \frac{1}{\Delta t_{23}} \tag{131,132}$$

Δt_{23} represents the amount of time between t_2 and t_3. We insert Equations 131 and 132 into Equation 130, simplify, and arrive at a new expression for $\dot{\varepsilon}$.

$$\dot{\varepsilon} = \frac{1}{\Delta t_{23}}\left(\varepsilon_R + \sigma_{a+}\left(\frac{1}{E_m} - \frac{1}{E_a}\right)\right) \tag{133}$$

We now focus on the Ψ term in the expression for D_{int}^{23} (Equation 129). We expand Ψ according to Equation 126, recall that stress is equal to σ_{a+}, and make use of the expressions for $\dot{x}_a$ and $\dot{x}_+$ (Equations 131 and 132).

$$\Psi = \frac{\partial \Psi}{\partial x_a}\left(\frac{-1}{\Delta t_{23}}\right) + \frac{\partial \Psi}{\partial x_+}\left(\frac{1}{\Delta t_{23}}\right) = \frac{1}{\Delta t_{23}}\left(-\frac{\sigma^2}{2E_a} - \beta(T) + \frac{\sigma^2}{2E_m}\right) \tag{134}$$

Equations 133 and 134 complete the definition of the integrand in D_{int}^{23} (Equation 129) and we see that none of the quantities vary in time. We solve the integral, substitute the expressions for $\dot{\varepsilon}$ and Ψ, and simplify to arrive at a final expression for the total entropy produced during step $t_2 \to t_3$.

$$D_{int}^{2,3} = \int_{t_2}^{t_3} \left(\sigma_{a+}\dot{\varepsilon} - \dot{\Psi}\right) dt = \left(\sigma_{a+}\dot{\varepsilon} - \dot{\Psi}\right)\Delta t_{23} = \frac{{\sigma_{a+}}^2}{2}\left(\frac{1}{E_m} - \frac{1}{E_a}\right) + \sigma_{a+}\varepsilon_R + \beta(T) \tag{135}$$

We now turn our attention to the $D_{int}^{5,6}$ term in Equation 129. During step $5 \rightarrow 6$ we know the value of stress and the rates of change of the phase fractions.

$$\sigma = \sigma_{+a} \quad , \quad \dot{x}_a = \frac{1}{\Delta t_{56}} \quad , \quad \dot{x}_+ = \frac{-1}{\Delta t_{56}}$$

We use these three values and follow a similar procedure to the calculation of $D_{int}^{2,3}$ (Equations 130-135) to obtain the entropy production term $D_{int}^{5,6}$.

$$D_{int}^{56} = -\frac{{\sigma_{+a}}^2}{2}\left(\frac{1}{E_m} - \frac{1}{E_a}\right) - \sigma_{+a}\varepsilon_R - \beta(T) \tag{136}$$

Now we substitute $D_{int}^{2,3}$ (Equation 135) and D_{int}^{56} (Equation 136) into Equation 129 to get the total entropy production for the load-unload cycle shown in Figure 47.

$$D_{int}^{Tot} = D_{int}^{23} + D_{int}^{56} = = \frac{{\sigma_{a+}}^2 - {\sigma_{+a}}^2}{2}\left(\frac{1}{E_m} - \frac{1}{E_a}\right) + \sigma_{a+}\varepsilon_R - \sigma_{+a}\varepsilon_R = \frac{{\sigma_{a+}}^2 - {\sigma_{+a}}^2}{2}\left(\frac{1}{E_m} - \frac{1}{E_a}\right) + \sigma_{\Delta}\varepsilon_R \tag{137}$$

Examination of Equation 137 shows that the entropy production is a function of the difference between the forward and reverse transformation stresses, the disparity between the elastic moduli of the two phases, and the residual strain. These properties are dictated by the shape of the multiwell HFE landscape and, in this sense, the entropy production is inherently contained in the HFE and does not have to be specified as a separate potential.

The entropy production calculated here, from the Clausius-Planck equation (122), is often called internal dissipation. The system sketched in Figure 47, however, does not consider any effects such as friction, damping, or viscoelasticity which are often regarded as dissipative in the sense that strain energy is converted (lost) to heat. The nonzero entropy in Equation 137 indicates that different amounts of energy were stored by the wire during loading than were released during unloading. This difference in energy is visually apparent by the presence of a hysteresis loop and should also be equal to its area. We turn to the latent heat calculations to account for this exchange between strain energy and heat.

Recalling Equation 118, we know that the latent heat associated with phase transformation is dependent on stress in addition to the internal energy difference between the phases. This requires that, because the forward and reverse transformations occur at different levels of stress, different amounts of heat be exchanged during loading and unloading. Here we show that the entropy production calculated in Equation 137 can also be calculated as the difference between latent heats during forward and reverse transformation. The latent heat released during a forward transformation is found by evaluating Equation 118 with $\sigma = \sigma_{a+}$ and the heat absorbed during reverse transformation uses the same equation with $\sigma = \sigma_{+a}$. The difference between these two quantities (recalling that $\sigma_{a+} - \sigma_{+a} = \sigma_{\Delta}$) simplifies to the same expression as Equation 137, the entropy production.

$$h_{\Delta} = h_+(\sigma_{a+}) - h_+(\sigma_{+a}) = = \frac{{\sigma_{a+}}^2 - {\sigma_{+a}}^2}{2}\left(\frac{1}{E_m} - \frac{1}{E_a}\right) + \sigma_{\Delta}\varepsilon_R = D_{int}^{Tot} \tag{138}$$

This calculation for h_{Δ} indicates that more heat is released during loading than is absorbed during unloading. It is expected that this value be equal to the dissipation calculated by the Clausius-Planck equation, which represents the strain energy lost during each cycle. Often times, the entropy production is calculated simply as the area inside the hysteresis loop. We also include this exercise and show that it produces the same result as Equation

137 and 138. Figure 48 shows the calculation of the hysteresis area from a stress-strain curve using parameters in Figure 48.

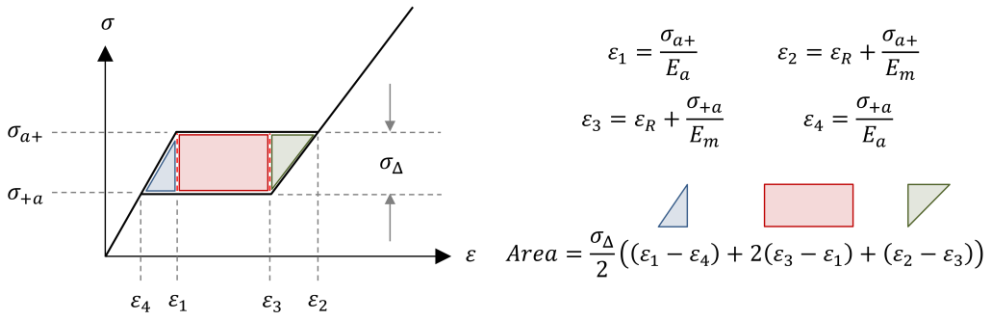


Figure 48: Geometric calculation of hysteresis area

After substituting the expressions for strain locations ε_{1-4}, we see that the area simplifies to the expected form that is equivalent to Equations 137 and 138.

$$Area = \frac{\sigma_\Delta}{2}\left(\left(\frac{1}{E_m} - \frac{1}{E_a}\right)(\sigma_{a+} + \sigma_{+a}) + 2\varepsilon_R\right) = \frac{{\sigma_{a+}}^2 - {\sigma_{+a}}^2}{2}\left(\frac{1}{E_m} - \frac{1}{E_a}\right) + \sigma_\Delta \varepsilon_R = D_{int}^{Tot} = h_\Delta$$

It should be reiterated here that the Helmholtz free energy is the basis for all three of these quantities yet contains no explicit information about entropy, hysteresis, or latent heat. This highlights one of the advantages of a multiwell free energy; the energy barriers between the wells inherently contain the necessary system information for all of these effects. Now, after in-depth examinations of the kinetic theory and the underlying free energy, we begin to undertake reformulations of the model to extend its capabilities.

5 New 1-D Model with Extended Capabilities

In this chapter we present a new version of the 1-D model that can have tension-compression symmetry and temperature-dependent hysteresis. This is accomplished by reinterpreting the free energy such that each martensite variant can have a unique well and barrier. The curvatures of the energy barriers (which dictate the hysteretic behavior) are also allowed to be functions of temperature. These generalizations involve some algebraic restructuring, the use of a dual set of material parameters, and also re-derivation of the caloric quantities. The general structure of the new model remains similar to the original (Chapter 2). Some of the equations will be repeated here, but this is only to have a complete description of the new model in one chapter – the discussion will be focused what its different.

The larger goal of updating the free energy is to enable the addition of more variants in Chapter 6, but tension-compression asymmetry is a convenient 1D application for this concept. Experimental observations often show stress-strain curves for the same SMA material that are notably different in a tension test vs. a compression test. The tensile responses tend to have lower transformation stress magnitudes, flatter stress plateaus, and larger transformation strains compared to compressive test results. Electron microscopy and neutron diffraction observations of tension and compression test specimens have shown differences in martensite plate and dislocation formation as well as different variant compositions (Liu et al. (1998), Sittner et al. (2004)). At a larger scale, thermal imaging techniques demonstrate that localization and transformation front motion are not observed in compression-dominated material (Reedlunn et al. (2014)). Many SMA applications such as actuators and vibration isolators employ only tension; in which case a tension-only or tension-compression symmetric model is sufficient.

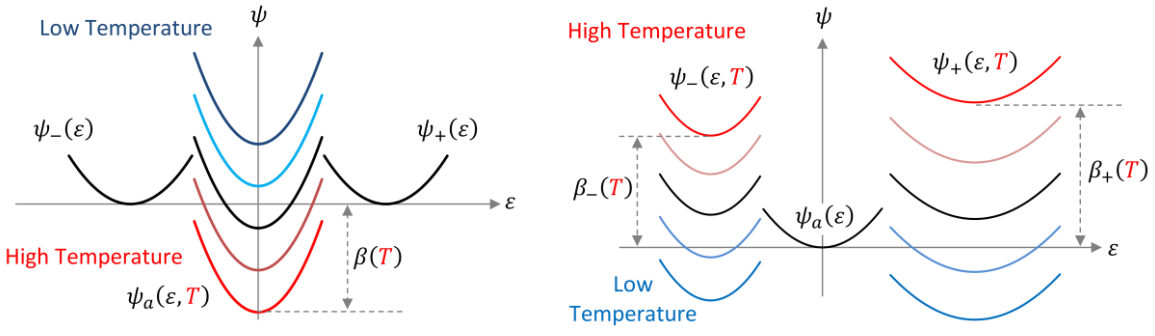
In some devices however, the SMA members behave in bending and distinguishing between the tensile and compressive responses could aid in modeling and design. In the *Simulations* section, we compare the results of a simulation to experimental data from Reedlunn et al., 2014 where NiTi tubes were tested in both tension and compression. We also use this data as the basis for several plots and diagrams in the *Mesoscopic Free Energy* section to show the shape of the energy landscape in the in the more general asymmetric case.

During the reformulation of the free energy and subsequent rederivation of the caloric properties, we also allow the hysteresis width parameter to be temperature-dependent and this is based on experimental observations. Once in the superelastic range, most SMA materials exhibit stress and temperature hysteresis loops with relatively constant widths. In some cases, however, stress hysteresis widths are observed that vary with temperature and temperature hysteresis widths are observed that vary with stress. Several publications report experiments showing stress-dependence in the temperature hysteresis width (Xiao (2014), Nascimento et al. (2009), Hamilton et al. (2004), Ortin et al. (1992)). In 2009, Sittner and a group of researchers reported an extensive set of experiments on NiTi wires under various combinations of tension and torsion. This set of experiments is the basis for the 2-D Model later on, but also contains very distinct examples of stress hysteresis loops that vary with temperature. This was also observed for TiNiCuCo specimens in Welsch et al. (2018) where the M-A-S model was used to simulate the experiments. In this case, a temperature-dependent hysteresis was included but in a manner that was limited to superelastic temperatures and a constant latent heat coefficient. In this chapter, we treat the temperature-dependent hysteresis in a more thorough manner where the formulation is valid through the entire temperature range.

To include both of these effects, tension-compression asymmetry and temperature-dependent hysteresis, we first rebuild the free energy.

5.1 Helmholtz Free Energy

The traditional M-A-S model begins with a set of three free energy parabolas where it is assumed that the two martensite wells have equal chemical free energies, strain offsets, and curvatures. Since the martensite chemical free energies are equal, they are set as the reference zero-energy and the austenite free energy is rewritten as a relative quantity. This is shown in Figure 49a, where the central austenite wells moves up and down with temperature.



(a) Energy well parabolas, Traditional M-A-S model (b) Energy well parabolas, Updated model

Figure 49: Free energy parabolas for showing different reference zero-energies

The updated model must accommodate asymmetry between the martensite wells and this leads us to reverse which well is considered to be the reference zero. In Figure 49b, the austenite well is vertically fixed and the two martensite wells move up and down with temperature. We see also that the two wells assume different heights at

the same temperature (color), they do not have the same strain, and they have different curvatures. This represents the case of a tension-compression asymmetric material. The Helmholtz free energy corresponding to the new model and Figure 49b is given by Equation 139.

$$\psi(\varepsilon,T)=\begin{cases}\psi_-(\varepsilon,T) & -\infty<\varepsilon\le\varepsilon_{-a}(T)\\ \psi_-^B(\varepsilon,T) & \varepsilon_{-a}(T)<\varepsilon<\varepsilon_{a-}(T)\\ \psi_a(\varepsilon) & \varepsilon_{a-}(T)\le\varepsilon\le\varepsilon_{a+}(T)\\ \psi_+^B(\varepsilon,T) & \varepsilon_{a+}(T)<\varepsilon<\varepsilon_{+a}(T)\\ \psi_+(\varepsilon,T) & \varepsilon_{+a}(T)\le\varepsilon<\infty\end{cases}\quad where\quad \begin{aligned}&\psi_\pm(\varepsilon,T)=\frac{E_\pm}{2}\left(\varepsilon-\varepsilon_R^\pm\right)^2+\beta_\pm(T)\\ &\psi_\pm^B(\varepsilon,T)=A^\pm(T)\varepsilon^2+B^\pm(T)\varepsilon+C^\pm(T)\\ &\psi_a(\varepsilon)=\frac{E_a}{2}\varepsilon^2\end{aligned} \tag{139}$$

It should first be noted that the energy barriers are not shown in Figure 49 and that the energy barrier terms in Equation 139 are shown in a lighter color. One of the key features of the updated model is that it uses the *Inflection Point* kinetics model from Chapter 3, which means that the energy barriers never have to be defined. They are included for the same of completeness, and are shown in several upcoming figures, but their expressions are reserved for Appendix 10.2. Continuing, there are three main differences between Equation 139 and the original expression (Equation 1). First, the chemical free energy terms (the β's) now belong to the martensite functions because the austenite chemical free energy is now the reference zero. The second difference is that the martensite and energy barrier coefficients now make a distinction between $+$ and $-$. Previously, they either used an m subscript for nondescript martensite or no subscript at all. The last difference is that the two martensite free energies as well as the energy barrier equations have the exact same form. Because the coefficients are now associated with either positive or negative side of the free energy, they can carry signs that are consistent with the quantities they represent. For example, the compressive residual strain parameter (ε_R^-) will have a negative value because it reflects a compressive strain. This allows both of the mechanical energy terms in the martensite functions to have the $\frac{E}{2}(\varepsilon-\varepsilon_R)^2$ form.

An example of the full free energy (using experimental data from Reedlunn et al. (2014)) is shown in Figure 50.

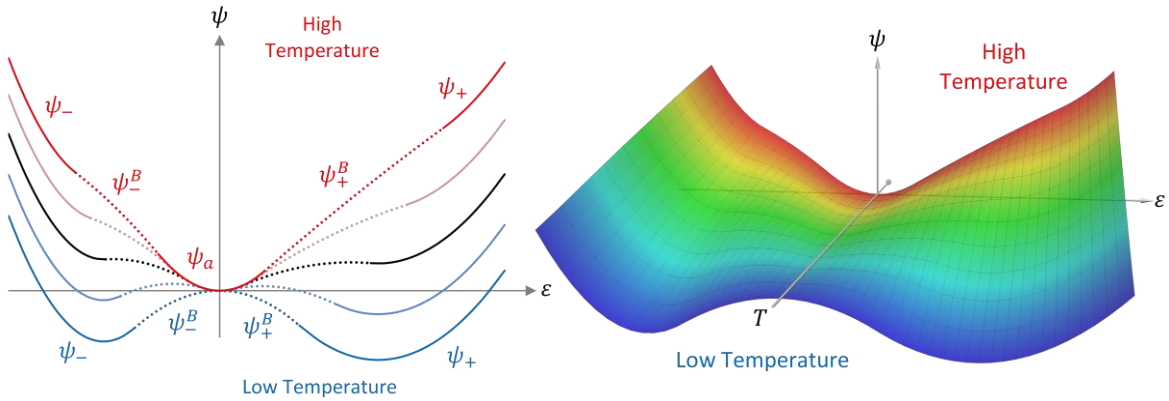


Figure 50: (a) HFE vs. strain at five temperatures, (b) HFE vs. strain and temperature

Figure 50a shows how the curves, even with the asymmetry, still converge to form a two-well landscape at low temperatures. Figure 50b shows the same information on a continuous strain-temperature domain – we can see that the entire landscape is still smooth, despite the varying curvatures in the wells. We have to make a few assumptions to ensure that this happens, and these will be discussed next as we outline the material parameter identification.

5.1.1 Material Parameters

The process for defining the material parameters is similar to that in Chapter 4 but there are now two sets of parameters (tensile and compressive) as well as a few additional relationships that must be established. Previously, two stress-strain experiments were required, one at a high temperature (T_H) and one at a slightly lower temperature (T_L). Now, a third set of test data is required; a compressive stress-strain curve. This data can be at either T_L or T_H but, for this chapter, we assume it is at T_H. Figure 51 shows an idealized set of stress-strain curves at temperature T_H, the measured parameters, and the corresponding Helmholtz free energy.

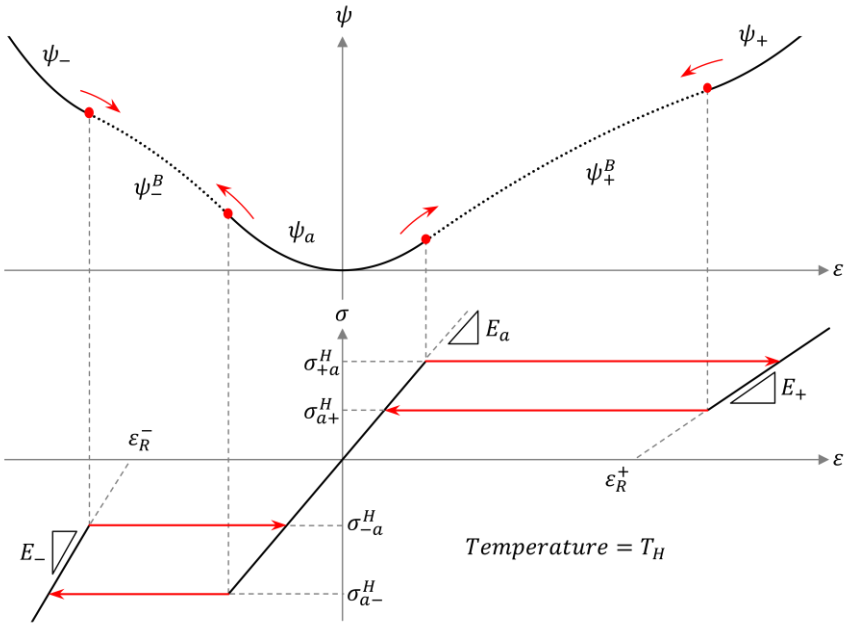


Figure 51: Tension stress-strain curves, compression stress-strain curves, measured parameters, and the free energy at temperature T_H

We see how the asymmetry in the stress-strain curves is translated to asymmetry in the free energy. From this data set, which is at temperature T_H, the following parameters are measured:

$E_i \rightarrow$ *Elastic modulus for each phase* $i \in -, a, +$

$\varepsilon_R^{\pm} \rightarrow$ *Residual strain for the* $+$ *and* $-$ *phases*

$\sigma_{a+}^H \rightarrow \sigma_{a+}$ *measured at temperature* T_H

$\sigma_{+a}^H \rightarrow \sigma_{+a}$ *measured at temperature* T_H

$\sigma_{a-}^H \rightarrow \sigma_{a-}$ *measured at temperature* T_H

$\sigma_{-a}^H \rightarrow \sigma_{-a}$ *measured at temperature* T_H

Tensile tests and compression tests will usually have different setups for metallic materials and will, therefore, be two distinct sets of data. Experiments show us, however, that the linear austenite branches in tension and compression tests have similar slopes and this is expected from elasticity theory when no buckling is present. This means that the tension and compression data can be assembled into a single effective stress-strain data set if they are taken at the same temperature. An example of this is shown in Figure 53 using the data from Reedlunn et al. (2014).

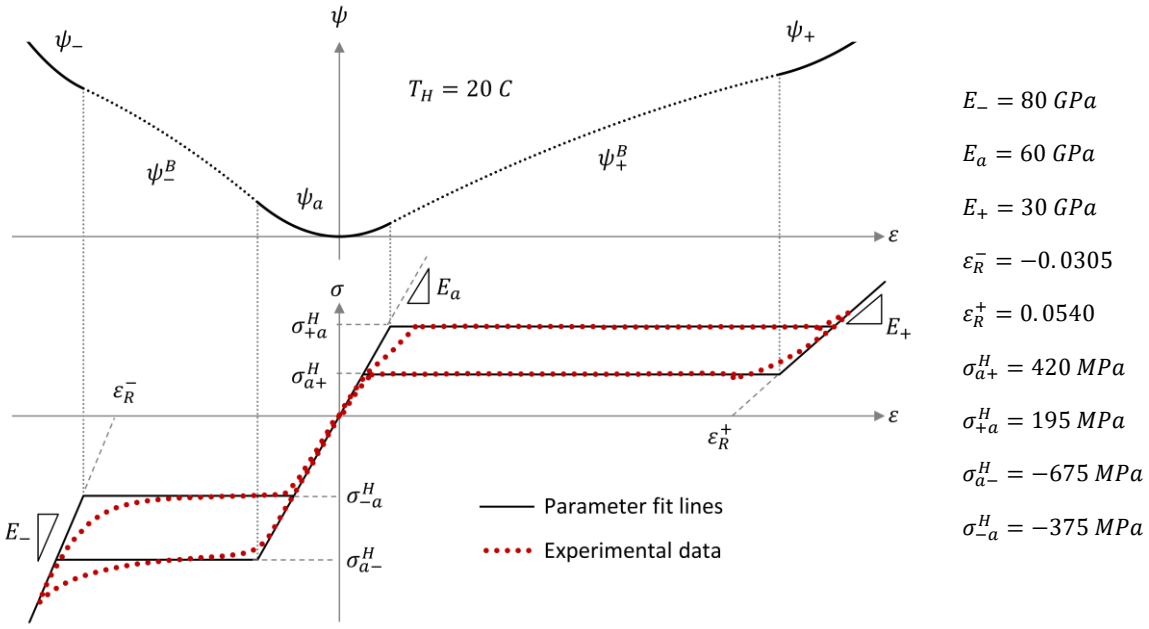


Figure 52: Example parameter measurements using tension and compression data from Reedlunn et al. (2014)

The nonlinearity in the tensile loading branch, just prior to forward transformation, is due to the presence of an intermediate R phase, which is not considered in this model. The second test is conducted at temperature T_L, which is slightly lower than T_H. This data set only needs to include tensile curves and only the transformation stresses need to be measured.

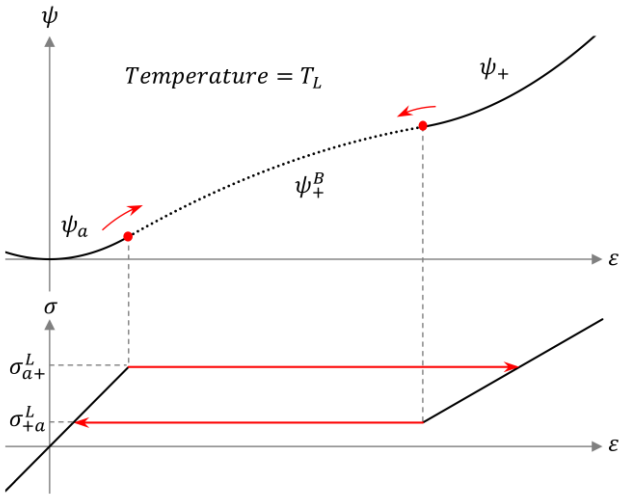


Figure 53: Tension stress-strain curves, compression stress-strain curves, measured parameters, and the free energy at temperature T_L

$$\sigma_{a+}^L \rightarrow \sigma_{a+}\ measured\ at\ temperature\ T_L$$

$$\sigma_{+a}^L \rightarrow \sigma_{+a}\ measured\ at\ temperature\ T_L$$

We should point out here that if you know (or can estimate) the temperature-rates of change of the transformation stresses ($\frac{d\sigma_{a+}}{dT}$ and $\frac{d\sigma_{+a}}{dT}$), the low temperature test can be neglected. The material constants σ_{a+}^L and

σ_{+a}^{L} can instead be derived at some chosen T_L using the T_H values along with the $\frac{d\sigma}{dT}$ values. This is what was done with the Reedlunn et al. (2014) data because the experimental data was only given at one temperature.

The 11 parameters listed above represent the complete set of material constants needed to operate the model for a given SMA material – whether it has tension-compression asymmetry or not and whether it has temperature-dependent hysteresis or not. In cases where tension-compression asymmetry is not required, the following assignments can be made.

$$E_- = E_+ \qquad \varepsilon_R^- = -\varepsilon_R^- \qquad \sigma_{a-}^{H} = -\sigma_{a+}^{H} \qquad \sigma_{-a}^{H} = -\sigma_{+a}^{H}$$

In cases where the hysteresis width is assumed to constant, the reverse transformation stress equations (149 and 150) can be replaced with expressions based on the forward values.

$$\sigma_{+a}(T) = \sigma_{a+}(T) - \sigma_{\Delta}^{+}(T_H) \qquad \sigma_{-a}(T) = \sigma_{a-}(T) + \sigma_{\Delta}^{-}(T_H)$$

The material parameters defined here are considered to be constants and are next used to define the transformation stress-temperature relationships.

5.1.2 Transformation Stress

First, we focus on the tensile side of the free energy and use σ_{a+}^{H}, σ_{a+}^{L}, σ_{+a}^{H} and σ_{+a}^{L} to define the Clausius-Clapeyron-like parameters, $\frac{d\sigma_{a+}}{dT}$ and $\frac{d\sigma_{+a}}{dT}$, which represent linear rates of change with respect to temperature.

$$\frac{d\sigma_{a+}}{dT} = \frac{\sigma_{a+}^{H} - \sigma_{a+}^{L}}{T_H - T_L} \qquad , \qquad \frac{d\sigma_{+a}}{dT} = \frac{\sigma_{+a}^{H} - \sigma_{+a}^{L}}{T_H - T_L} \tag{140}$$

The two quantities are defined independently to allow temperature-dependent hysteresis. These two parameters can be used to approximate the two tensile transformation stresses as linear functions of temperature.

$$\sigma_{a+}(T) = \sigma_{+a}^{H} + \frac{d\sigma_{+a}}{dT}(T - T_H) \qquad , \qquad \sigma_{+a}(T) = \sigma_{a+}^{H} + \frac{d\sigma_{a+}}{dT}(T - T_H) \tag{141,142}$$

Next, we use Equation 140 to extrapolate the temperature where the tensile forward transformation stress ($\sigma_{a+}(T)$) reaches zero. We call this critical temperature T_{a+} or the *martensite start* temperature because martensite will form at this temperature under stress-free conditions (this effect is shown in Figure 4).

$$\sigma_{a+}(T_{a+}) = 0 \qquad \rightarrow \qquad T_{a+} = T_H - \frac{\sigma_{a+}^{H}}{\frac{d\sigma_{a+}}{dT}} \tag{143}$$

We assume all martensite variants have the same start temperature and thus we can also define the critical temperature for the compressive variant from T_{a+}.

$$T_{a-} = T_{a+} = T_{am} \tag{144}$$

Physically, this reflects the fact that the austenite well becomes unstable at a certain temperature and, at that point, all of the martensite wells become available. Since the two parameters are equal, we use a common symbol T_{am} where the m represents martensite without specifying a variant. Figure 54 shows the HFE and stress-strain curve at temperature T_{am}.

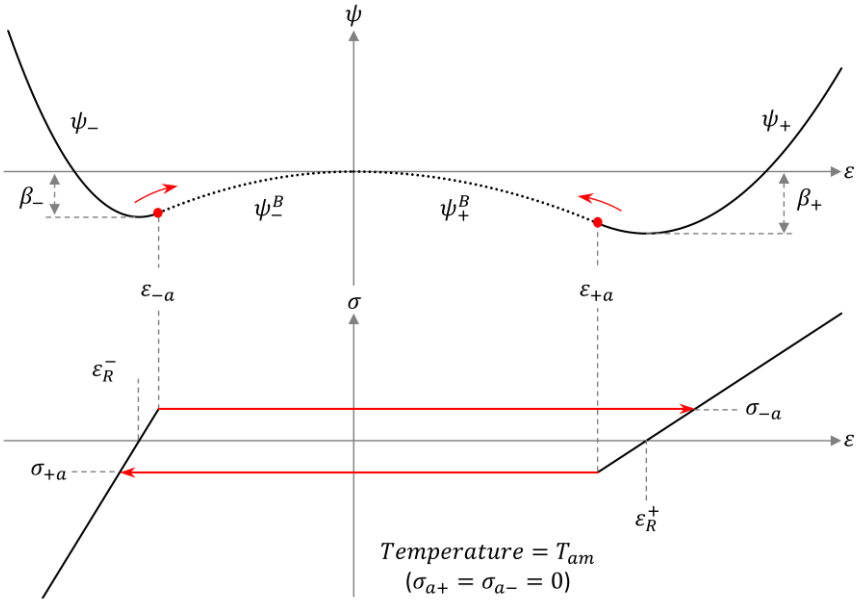


Figure 54: Mesoscopic HFE and macroscopic stress-strain response at the martensite start temperature

Here, the HFE has assumed a two-well configuration and the transformations constitute martensite to martensite phase changes, which result in a single tension-compression hysteresis loop. The low temperature configuration also makes the martensite wells easier to see (because they have become convex). The CFE's of the two martensite wells (β_- and β_+) are labeled here but not in Figure 51 for this reason. Since all martensite variants become available under zero stress and at temperature T_{am}, we can write $\sigma_{a-}(T_{am}) = 0$ and this allows us to the define the compressive forward transformation stress, σ_{a-}, as a function of temperature.

$$\sigma_{a-}(T) = \sigma_{a-}^H + \frac{d\sigma_{a-}}{dT}(T - T_H) \qquad where \qquad \frac{d\sigma_{a-}}{dT} = \frac{\sigma_{a-}^H - 0}{T_H - T_{am}} \approx \frac{\sigma_{a-}^H - \sigma_{a-}^L}{T_H - T_L} \tag{145}$$

Equation 145 shows two different expressions for $\frac{d\sigma_{a-}}{dT}$. It is a mathematical requirement of the model that both the tensile and compressive forward transformation stresses must reach zero at the same temperature, T_{am}. Because of this, the first expression supersedes the second one. It is expected that both determinations of $\frac{d\sigma_{a-}}{dT}$ yield similar values. The reverse transformation stresses can have different values at T_{am} thus the expression for $\sigma_{-a}(T)$ is written directly from the experimental parameters.

$$\sigma_{-a}(T) = \sigma_{-a}^H + \frac{d\sigma_{-a}}{dT}(T - T_H) \qquad where \qquad \frac{d\sigma_{-a}}{dT} = \frac{\sigma_{-a}^H - \sigma_{-a}^L}{T_H - T_L} \tag{146}$$

The temperature T_{am} is the lowest temperature considered by the model. At temperatures lower than T_{am}, the free energy remains locked in the same the single hysteresis loop configuration shown in Figure 54. This is enforced by adding the minimum temperature condition to the transformation stress equations (147-150). This prevents the forward transformation stresses from becoming negative (which would be physically meaningless) and also causes the reverse transformation stresses to halt at T_{am}. This reflects experimental observations where SMA stress-strain responses begin to stagnate as it reaches the single hysteresis-loop configuration (Chang et al., 2006, Sittner et al., 2009). A second limiting condition must be placed on the two reverse transformation stresses to address cases where the hysteresis width is temperature dependent. We assume the hysteresis width can vary linearly with temperature, but do not allow either reverse transformation stress to reach a larger magnitude than

the corresponding forward value. This would invert the hysteresis loop and yield negative entropy production, violating the second law of thermodynamics. We address this by placing upper limits on the two reverse transformation stresses. It should be noted that, for the two temperature-dependent hysteresis data sets examined in this chapter (Sittner et al. (2009), Welsch et al. (2017)), this limit was not necessary. The temperatures where the forward and reverse transformation stresses would have intersected one another were much higher than the temperatures considered. The four transformation stress expressions are given in their final conditional forms by Equations 147-150.

$$\sigma_{a+}(T) = \sigma_{a+}^{H} + \frac{d\sigma_{a+}}{dT}(T^* - T_H) \qquad \sigma_{+a}(T) = min\left(\sigma_{+a}^{H} + \frac{d\sigma_{+a}}{dT}(T^* - T_H), \sigma_{a+}(T)\right) \tag{147,148}$$

$$\sigma_{a-}(T) = \sigma_{a-}^{H} + \frac{d\sigma_{a-}}{dT}(T^* - T_H) \qquad \sigma_{-a}(T) = max\left(\sigma_{a-}(T), \sigma_{-a}^{H} + \frac{d\sigma_{-a}}{dT}(T^* - T_H)\right) \tag{149,150}$$

$$where \qquad T^* = max(T, T_{am})$$

Next, we discuss thermal hysteresis. Two transformation temperatures are typically associated with SMA materials. The martensite start temperature (T_m) is given by Equation 144 and is based on the motion that the transformation stresses reach zero at that temperature. We can also define the temperatures at which the martensite wells become unstable and austenite begins to form under stress-free conditions by identifying the points at which $\sigma_{+a}(T)$ and $\sigma_{-a}(T)$ equal zero.

$$0 = \sigma_{\pm a}(T_{\pm a}) \quad \rightarrow \quad T_{\pm a} = T_L - \frac{\sigma_{a\pm}^{L} + \sigma_{\Delta}^{\pm}}{\frac{d\sigma_{\pm a}}{dT}} \tag{151}$$

This temperature is often called the austenite start temperature, but we maintain two different parameters because it is possible to construct a TC-asymmetric HFE where the tensile and compressive martensite wells become unstable at different temperatures. The complete set of transformation stress equations allows us to also define the critical transformation strains. There are defined in a similar manner to the previous model (Equations 4 and 5), but now all four are written independently.

$$\varepsilon_{a\pm}(T) = \frac{\sigma_{a\pm}(T)}{E_a} \qquad , \qquad \varepsilon_{\pm a}(T) = \frac{\sigma_{\pm a}(T)}{E_{\pm}} + \varepsilon_R^{\pm} \tag{152,153}$$

Now, having established the relationship between the transformation stresses and temperature, we proceed to the chemical free energies.

5.1.3 Chemical Free Energy

The only remaining parameters required to build the Helmholtz free energy wells are the martensite chemical free energies (β_- and $\beta_\pm$). We begin by defining a few intermediate parameters. σ_Δ^+ and σ_Δ^-, given by Equations 154, represent the hysteresis widths for the tensile and compressive curves respectively. These equations further characterize the material and help to shorten other, downstream expressions.

$$\sigma_{\Delta}^{\pm}(T) = max\left(\sigma_{\Delta}^{\pm}(T_H) + \frac{d\sigma_{\Delta}^{\pm}}{dT}(T - T_H), 0\right) \qquad where \qquad \begin{cases} \frac{d\sigma_{\Delta}^{\pm}}{dT} = \frac{\sigma_{\Delta}^{\pm}(T_H) - \sigma_{\Delta}^{\pm}(T_L)}{T_H - T_L} \\ \sigma_{\Delta}^{\pm}(T_H) = \sigma_{a\pm}^{H} - \sigma_{\pm a}^{H} \\ \sigma_{\Delta}^{\pm}(T_L) = \sigma_{a\pm}^{L} - \sigma_{\pm a}^{L} \end{cases} \tag{154}$$

Equation 154 yields the same values as subtracting the reverse transformation stresses from the corresponding forward values (Equations 149 147-150), but the linearly varying hysteresis width is convenient and sometimes necessary as a stand-alone parameter. We also must now define two different parameters to reflect the stiffness disparities between austenite and the two martensites.

$$E_{\Delta}^{\pm} = \frac{1}{E_{\pm}} - \frac{1}{E_a} \tag{155}$$

The chemical free energies are now given by Equation 156, which has a similar derivation to the original version (Equation 92).

$$\beta_{\pm}\left(\sigma_{a\pm}(T)\right) = C_2^{\pm}\sigma_{a\pm}(T)^2 + C_1^{\pm}(T)\sigma_{a\pm}(T) + C_0^{\pm}(T) \qquad where \qquad \begin{aligned} C_0^{\pm}(T) &= -\frac{\sigma_{\Delta}^{\pm}(T)\varepsilon_R^{\pm}}{2} \\ C_1^{\pm}(T) &= \varepsilon_R^{\pm} - \frac{\sigma_{\Delta}^{\pm}(T)E_{\Delta}^{\pm}}{2} \\ C_2^{\pm} &= \frac{E_{\Delta}^{\pm}}{2} \end{aligned} \tag{156}$$

The only differences between Equation 156 and Equation 92 are that there are two sets of parameters ($+$ and $-$) and that the σ_{Δ}'s are temperatrue-dependent. Equation 156 allows the chemical free energies to be defined as functions of forward transformation stresses, which are defined for all temperatures. This means that the Helmholtz free energy is defined sufficiently to operate the model and now we proceed to the Gibbs free energy.

5.2 Gibbs Free Energy

The Gibbs free energy is defined in same manner as the traditional model, from the Helmholtz free energy. This means the GFE is also asymmetric, but it functions in the same way – stress distorts the landscape which causes certain phases to switch between stable and unstable.

$$g(\varepsilon,T;\sigma) = \begin{cases} g_-(\varepsilon,T;\sigma) & -\infty < \varepsilon \le \varepsilon_{-a}(T) \\ g_{\pm}^{B}(\varepsilon,T;\sigma) & \varepsilon_{-a}(T) < \varepsilon < \varepsilon_{a-}(T) \\ g_a(\varepsilon;\sigma) & \varepsilon_{a-}(T) \le \varepsilon \le \varepsilon_{a+}(T) \\ g_{+}^{B}(\varepsilon,T;\sigma) & \varepsilon_{a+}(T) < \varepsilon < \varepsilon_{+a}(T) \\ g_+(\varepsilon,T;\sigma) & \varepsilon_{+a}(T) \le \varepsilon < \infty \end{cases} \quad , \quad \begin{aligned} g_{\pm}(\varepsilon,T;\sigma) &= \frac{E_{\pm}}{2}\left(\varepsilon - \varepsilon_R^{\pm}\right)^2 + \beta_{\pm}(T) - \sigma\varepsilon \\ g_{\pm}^{B}(\varepsilon,T;\sigma) &= A^{\pm}(T)\varepsilon^2 + B^{\pm}(T)\varepsilon + C^{\pm}(T) - \sigma\varepsilon \\ g_a(\varepsilon;\sigma) &= \frac{E_a}{2}\varepsilon^2 - \sigma\varepsilon \end{aligned} \tag{157}$$

Once again, the energy barrier expressions are grey because they are not actually required to operate the model. Like the previous model, the Gibbs free energy is only evaluated at specified locations but now there are fewer – only 7. Furthermore, all of the locations are in the convex energy wells as shown in Figure 55.

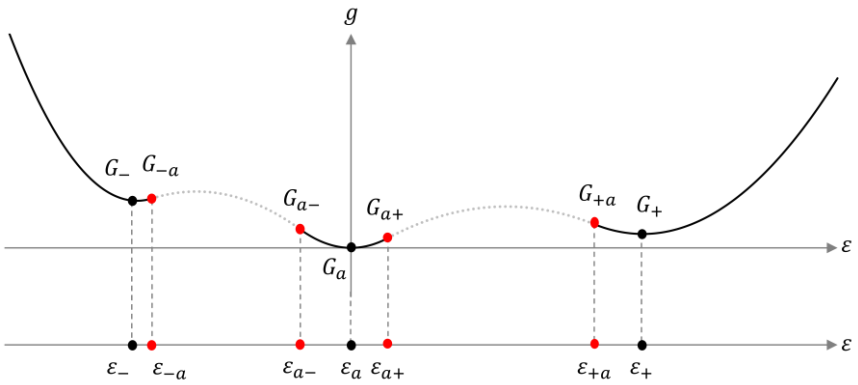


Figure 55: Locations on the GFE-strain landscape

This figure is similar to Figure 9 except that the landscape is asymmetric and the barrier energies are not shown (because they are not needed). The expressions to find the free energies G_{ij} are identical to those in the previous formulation, which were given by Equation 15, except now the critical strains (ε_{ij}) are found using Equations 152 and 153.

$$G_{ij}(\sigma,T) = g_i\big(\varepsilon_{ij}(T),T;\sigma\big) \qquad \text{(15) } \textit{Repeated here}$$

Because the residual strains now carry the proper sign, the extremum strains ε_i also have a slightly different form from the original model (Equations 16-18).

$$\varepsilon_{\pm}(\sigma) = \frac{\sigma}{E_m^{\pm}} + \varepsilon_R^{\pm} \qquad , \qquad \varepsilon_a(\sigma) = \frac{\sigma}{E_a} \qquad \text{(158, 159)}$$

And, finally, the minimum energy within each phase (G_i) is calculated in the same manner as the traditional model – the expressions (19, 20, 21) are repeated here for clarity and completeness.

$$G_{-}(\sigma,T) = g_{-}\big(min\big(\varepsilon_{-}(\sigma),\varepsilon_{-a}(T)\big),T;\sigma\big) \qquad (19)$$

$$G_a(\sigma,T) = g_a(\varepsilon^*,T;\sigma) \qquad where \qquad \varepsilon^* = max\left(\varepsilon_{a-}(T),min\big(\varepsilon_a(\sigma),\varepsilon_{a+}(T)\big)\right) \qquad (20)$$

$$G_{+}(\sigma,T) = g_{+}\big(max\big(\varepsilon_{+a}(T),\varepsilon_{+}(\sigma)\big),T;\sigma\big) \qquad (21)$$

At this point, the only mesoscopic parameters remaining to operate the model are the latent heat coefficients. Next, we calculate them using the chemical free energies.

5.3 Caloric Properties

The caloric properties are defined using the same basic procedure as the traditional model in Chapter 4, which begins with the assumed linear variation of the chemical free energy with temperature. This is expressed by Equation 160, where the tensile and compressive variant now have independent sets of internal energy and entropy constants.

$$\hat{\beta}_a(T) - \hat{\beta}_{\pm}(T) = -\beta_{\pm}(T) = u_{\pm}^R - s_{\pm}^R T \qquad where \qquad \begin{matrix} u_{\pm}^R = \hat{u}_a^R - \hat{u}_{\pm}^R \\ s_{\pm}^R = \hat{s}_a^R - \hat{s}_{\pm}^R \end{matrix} \qquad (160)$$

The subtraction order in Equation 160 is maintained from the previous formulation (Equation 90) even though intuition says to subtract the value of the reference *zero* energy from the other values. The only consequence, however, to reversing the order is that some parameters will switch sign. We use the $austenite\ minus\ marteniste$ order in Equation 160 because it yields latent heat coefficients with signs that are consistent with the energy balance e (i.e. forward transformation corresponds to positive heat generation and vice versa). Now, as discussed in the Chapter 2, we choose to make a compromise by using the linear transformation stress equations (Equations 147-150). This decision means that that the chemical free energy is now only approximately linear so Equation 160 should be re-written as Equation 161.

$$\beta_{\pm}(T) \cong s_{\pm}^R T - u_{\pm}^R \qquad (161)$$

However, the previous examination also tells us that Equation 160 is still correct at the two experimental temperatures T_L and T_H because the linear and quadratic transformation stresses intersect. This means we can write the following four expressions without concern about linear vs. quadratic transformation stress.

$$\beta_-(T_L) = \beta_-^L = s_-^R T_L - u_-^R \qquad \beta_+(T_L) = \beta_+^L = s_+^R T_L - u_+^R$$

$$\beta_-(T_H) = \beta_-^H = s_-^R T_H - u_-^R \qquad \beta_+(T_H) = \beta_+^H = s_+^R T_H - u_+^R$$

(162,163) (164,165)

We can also use Equation 156, and the experimentally measured transformation stresses, to computer the values of the chemical free energies at temperatures T_L and T_H.

$$\beta_-^L = C_2^- {\sigma_{a-}^L}^2 + C_1^-(T_L)\sigma_{a-}^L + C_0^-(T_L) \qquad \beta_+^L = C_2^+ {\sigma_{a+}^L}^2 + C_1^+(T_L)\sigma_{a+}^L + C_0^+(T_L)$$

$$\beta_-^H = C_2^- {\sigma_{a-}^H}^2 + C_1^-(T_H)\sigma_{a-}^H + C_0^-(T_H) \qquad \beta_+^H = C_2^+ {\sigma_{a+}^H}^2 + C_1^+(T_H)\sigma_{a+}^H + C_0^+(T_H)$$

(166,167) (168,169)

Since, the four values of β and now known, we can use Equations 162-165 to calculate the two entropy constants.

$$\beta_-^L = s_-^R T_L - u_-^R \qquad \beta_-^H = s_-^R T_H - u_-^R \qquad \beta_+^L = s_+^R T_L - u_+^R \qquad \beta_+^H = s_+^R T_H - u_+^R$$

$$\downarrow \qquad\qquad \downarrow$$

$$s_-^R = \frac{\beta_-^H - \beta_-^L}{T_H - T_L} = \frac{d\beta_-}{dT} \qquad s_+^R = \frac{\beta_+^H - \beta_+^L}{T_H - T_L} = \frac{d\beta_+}{dT}$$

(170) (171)

Now, we can use the entropy constants and Equations 162-165 to define the internal energy constants at either temperature T_L or T_H, we choose T_H.

the following four expressions without concern about linear vs. quadratic transformation stress.

$$u_-^R = s_-^R T_H - \beta_-^H \qquad u_+^R = s_+^R T_H - \beta_+^H$$

(172) (173)

Using the same procedure as in Chapter 4, we can write the enthalpy for each martensite phase in terms of stress, material constants, internal energy constants, and entropy constants.

$$h_-(\sigma) = \frac{\sigma^2}{2} E_\Delta^- + \sigma\varepsilon_R^- + u_-^R \qquad h_+(\sigma) = \frac{\sigma^2}{2} E_\Delta^+ + \sigma\varepsilon_R^+ + u_+^R$$

(174) (175)

Equations 174 and 175 complete the set of parameters that are derived from the mesoscopic free energies and we now proceed to the macroscopic model.

5.4 Macroscopic Model Equations

In order to allow tension-compression asymmetry and hysteresis widths that do not have constant width, the mesoscopic free energy had to be updated. Now that the mesoscopic updates are complete, the macroscopic model does not need significant changes. In fact, the only differences in the macro-scale equations are due to the notational differences and the fact that the residual strains now carry proper signs. The total strain is assumed to be a weighted average of the strains experienced by each phase and we use the same Equation as the original model (26).

$$\varepsilon = \varepsilon_- x_- + \varepsilon_a x_a + \varepsilon_+ x_+ \qquad (26)$$

The average strains within each phase are now calculated using Equations 158 and 159, which were given in this chapter's Gibbs free energy discussion. Once again, we can use Equations 158 and 159 to rearrange the total strain and get an expression for the macroscopic stress.

$$\varepsilon = \left(\frac{\sigma}{E_m^-} + \varepsilon_R^-\right) x_- + \left(\frac{\sigma}{E_a}\right) x_a + \left(\frac{\sigma}{E_m^+} + \varepsilon_R^+\right) x_+ \qquad \rightarrow \qquad \sigma(\varepsilon) = \frac{\varepsilon - \varepsilon_R^- x_- - \varepsilon_R^+ x_+}{\frac{x_-}{E_m^-} + \frac{x_a}{E_a} + \frac{x_+}{E_m^+}} \tag{176}$$

The phase fractions are governed by the same set of kinetic ODE's (Equations 28-30).

$$\dot{x}_- = x_a\, p_{a-} - x_-\, p_{-a} \qquad \dot{x}_+ = x_a\, p_{a+} - x_+\, p_{+a} \qquad x_a = 1 - x_- - x_+ \tag{28,29,30}$$

The transformation probabilities are computed using the *Inflection Point* technique (Equation 55) that was derived in Chapter 3, where the ΔG represents the difference between the minimum energy and the neighboring inflection point energy.

$$p_{ij}(\sigma, T) = \frac{1}{\tau} e^{-\frac{1}{\gamma}\Delta G_{ij}(\sigma,T)} \qquad where \qquad \begin{cases} \Delta G_{-a} = G_{-a}(\sigma,T) - G_-(\sigma,T) \\ \Delta G_{a-} = G_{a-}(\sigma,T) - G_a(\sigma,T) \\ \Delta G_{a+} = G_{a+}(\sigma,T) - G_a(\sigma,T) \\ \Delta G_{+a} = G_{+a}(\sigma,T) - G_+(\sigma,T) \end{cases} \tag{55}$$

The thermal activation parameters, τ and γ, have the same meaning and should be handled in the same manner as described in Chapter 3. The G_i and G_{ij} parameters are discussed in the Gibbs free energy discussion from this chapter, but they use the same expressions as the original model (15 and 19-21). The final macroscopic equation is the energy balance and this has the same form as the original version (32) except that the latent heats are now calculated using Equations 174 and 175.

$$\dot{T} = \frac{1}{\rho c}\left(\dot{x}_+ h_+ + \dot{x}_- h_- - \alpha h_c (T - T_0)\right) \tag{32}$$

The energy balance is added to the set of kinetic ODE's given by Equations 28-30 and adds thermomechanical coupling to the model. This enable effects such as transformation hardening, where the transforamation plateaus become inclined when the material is loaded quickly. We demonstrate this effect in the next section, along with several others, where the new model's performance is showcased and compared to experimental data.

5.5 Simulations

Here we have chosen a variety of experimental data sets that show different types of SMA behavior. For each set, we compare the model's prediction to the measured result and discuss the correlation.

5.5.1 Quasi-Static Uniaxial Tension Tests

A series of experiments on single crystal NiTi wires was presented by Chang and coworkers in 2006. The work includes a set of quasi-static strain-controlled tension tests over a range of temperatures as well as a set at room temperature using various strain rates. We first use data from the quasi-static tests to obtain material properties and to validate the basic isothermal performance of the current model. Figure 56 shows experimental data (black lines) at two selected temperatures a set of straight segments (red) fit by visual inspection to obtain the material parameters.

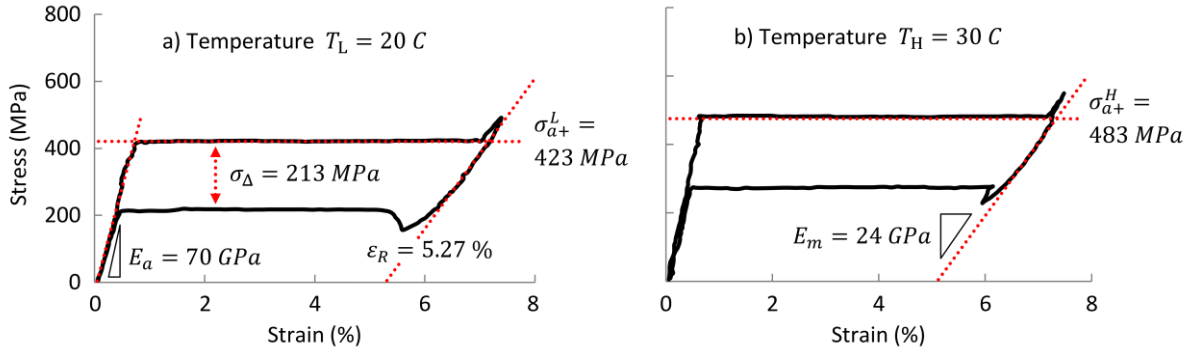


Figure 56: Material parameters obtained from experimental data in Chang et al. (2006)

Figure (57) shows the complete set of quasi-static tension tests from Chang et al. (2006) in black along with simulation results in red.

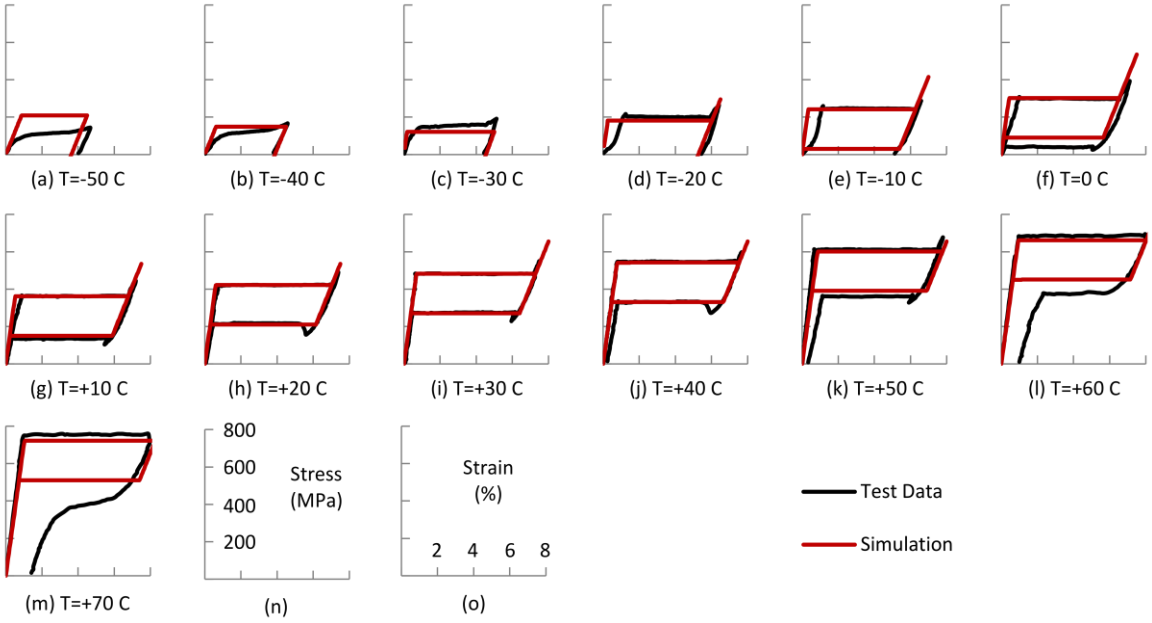


Figure 57: Simulation vs. experiment, quasi-static cases from Chang et al. (2006)

The model and the experimental data agree well for the temperatures -30C through +50C (Figure 56(c-k)). At -50 and -40C, the model assumes the microstructure is entirely made up of twinned martensite and that the transformation shown represents detwinning. The experimental responses for these two specimens reflect a smooth entry into the stress plateau, which is not seen at the other temperatures. This indicates the presence of a more complex set of microstructures that are not captured by the model. At the temperatures greater than +40C (Figure 56(k-m)), the experiments show that the stress-strain curve does not return to the original austenite branch after unloading. This is due to permanent plastic slip that occurs simultaneously to transformation at very high temperatures. Reflecting the plastic response of the material is outside the scope of the current model, which is intended for the design of SMA actuators (Seelecke, 2002) and thermal devices (Ullrich et al., 2014) where high-cycle usage is expected. Figure 57 shows that, at typical SMA operating temperatures, the current model accurately reflects both the pseudoplastic and superelastic responses.

5.5.2 Uniaxial Tension Tests at Various Strain rates

Chang and coworkers (2006) also conducted seven experiments with a similar setup to those shown in Figure 56, except the srain rate was varied. In these experiments, the ambient temperature was between 20 and 22 C, the

wire diameter was 0. 762 mm, and the gauge length was 33. 3 mm. The specimens were extended at a fixed rate of strian, paused, then unloaded at the same rate. The pause allowed the wire to return to room temperature so that both the loading and unloading curves represent the same initial temperature. The stress-strain results from each experiment are shown here in Figure 58.

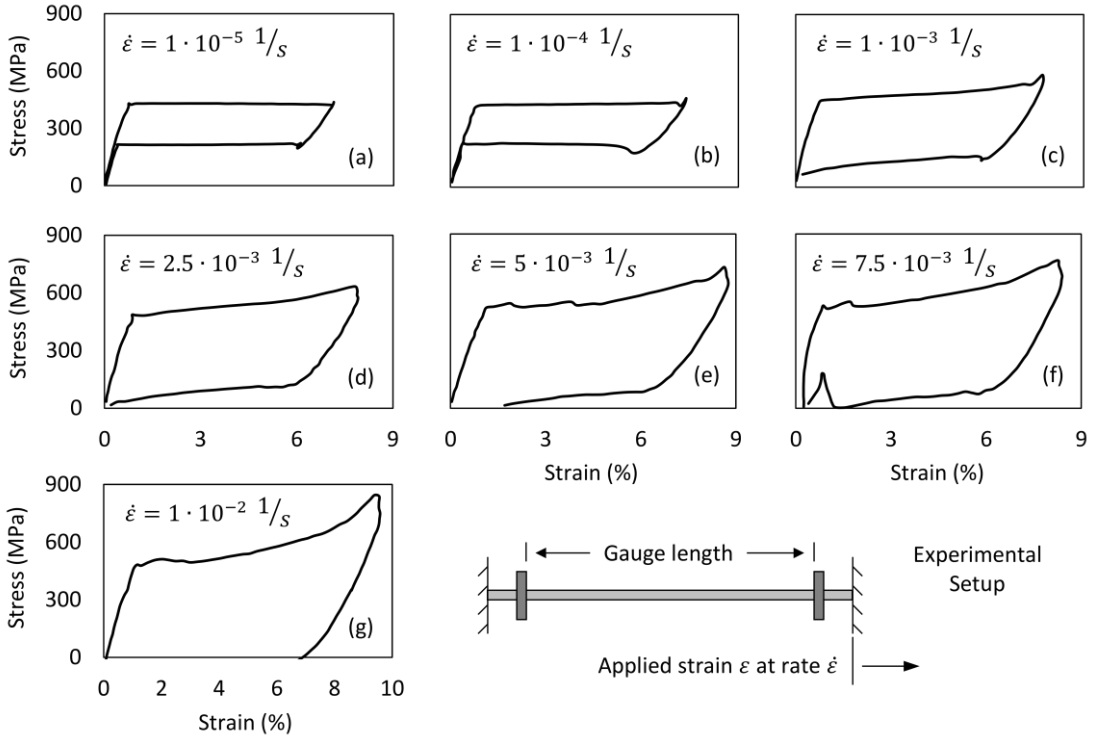


Figure 58: Tension test experiments at various strain rates from Chang et al. (2006)

The current model was used to simulate the cases shown in Figure 58(a-d). The three highest strain rate tests induced large amounts of residual strain and are not considered. The model does reasonably predict the transformation hardnening at strain rates beyond 10^{-2}, but the comparison to the experimental data in Figure 58(e-g) is misleading without consideration of plastic effects.

It should also be pointed out here that the plots in Figure 58 show engineering stress and engineering strain, both of which reflect an averaged response over the entire guage length. However, it is well known that even single crystal SMA wires exhibit inhomogeneous behavior. This was one of the focus points of the original paper where it was shown that transformations began in localized zones and then traveled along the length of the wire. The strains, strian rates, temperatures, and stresses were higher in the nucleation regions than in other portions of the wire. These localization effects must be considered when viewing the average stress-strain response of the specimens, where the curves reflect some combination of active zones and inactive zones.

The simulation cases each represent a single material point (or a uniform wire of arbitrary length) with no consideration of localization effects. The goal is to examine the fundamental material behavior comapred to the experimental stress-strain results. The thermal environment, on a per unit length basis, was approximated by using an ambient temperature of 21 C and using the same wire diameter (0. 762 mm) to calculate the surface area per unit volume in the energy balance equation (32). A convection coefficient of 90 $^{W}/_{m^2K}$ was measured for the experimental environment but a reduced value of 30 $^{W}/_{m^2K}$ was used in the simulations. Since the local transformation zones (which are much hotter) provide most of the nonlinearity in the experimental curves, the thermal environment in the simulations had to be scaled for quantitative comparison. This adjustment increased

the temperature of the simulated wires bringing them closer to the temperatures in the active zones of the experimental specimens. The wire temperatues in the simulations were still lower than the peak temperatures reported in the experiments, but more closely represent the average reponses. The results of the simulations along with the experimental data are shown in Figure 59.

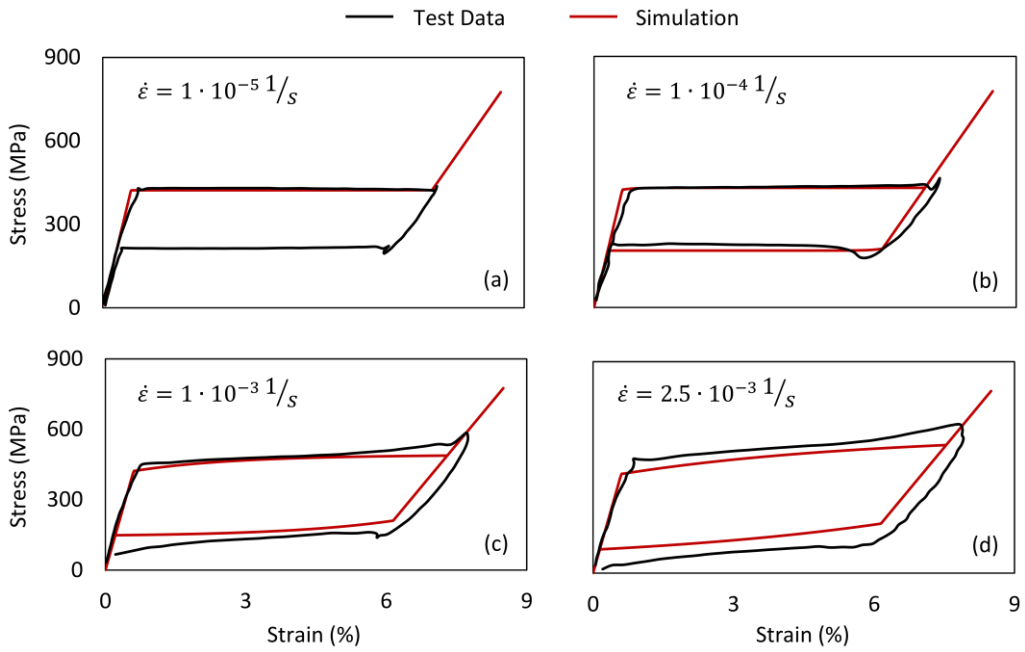


Figure 59: Model predictions compared to experimental data

The sensitivity of the model to the strain rate, in the scaled thermal environment, agrees very well with the experiments. Both the experimental data and the simulation results show a competition between convective heat transfer and latent heat generation/absorption. In Figures 59(a) and (b), the applied strain rates are low and therefore the latent heat effects are minimal. Any heat generated during forward transformation or absorbed during reverse transformation is quickly accounted for by the surrounding air. The temperature remains relatively constant in these cases and stress plateaus remain relatively flat. In Figures 59(c) and (d), the latent heating that occurs during forward transformation exceeds the cooling capacity of the surrounding environment so the temperature of the wire increases. This causes the transformation stress to increase along with the temperature which leads to the formation of a slope in the transformation branch of the stress-strain curve. As the strain rate increases (compare Figure 59(c) to (d)), more heat is generated and the slope becomes steeper.

Recalling that the tests (and simulations) were allowed to cool back to room temperature prior to unloading, the same behavior occurs in unloading except the roles are reversed. The latent heating is now negative and cools the wire while the environment supplies heat attempting to keep the wire at ambient temperature. At the slow strain rates, the environment is able to supply heat quickly enough and the wire temperature remains constant. At the higher strain rates, the wires cool and cause the plateau stress to decrease during reverse transformation.

5.5.3 Single Crystal Behavior - Stress-Strain and Actuation Responses

Figure 60 shows a set of superelastic tension tests on a single crystal SMA (CuZnAl) from Fu et al. (1993).

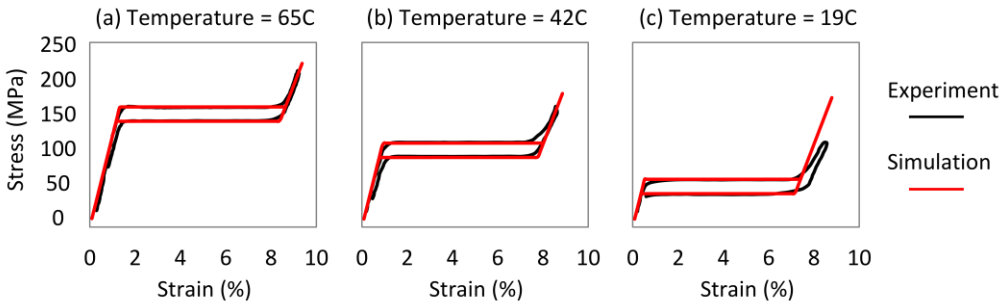


Figure 60: Superelastic tension of single-crystal SMA, simulations compared to test data from Fu et al. (1993)

The test data shows very flat stress plateaus and roughly constant hysteresis width. The model agrees well with the experimental data with most observable inconsistency being the 19C case, where the simulation slightly under-predicts the transformation strain. The general shape of the 19 C martensite branch also differs from the other two temperatures and we assume this to be an experimental anomaly. Also, the transitions into and out of the transformation plateaus are slightly rounded whereas the simulation curves have sharp corners. The simulations here represent the limiting case of a perfect single crystal, which is an engineering idealization of the type of response in Figure 60. The polycrystalline option in could be used to capture the curvature, but the results are well captured without it. Figure 61 shows a second set of actuation experiments from the same paper (Fu et al. (1993)) on the same CuZnAl material.

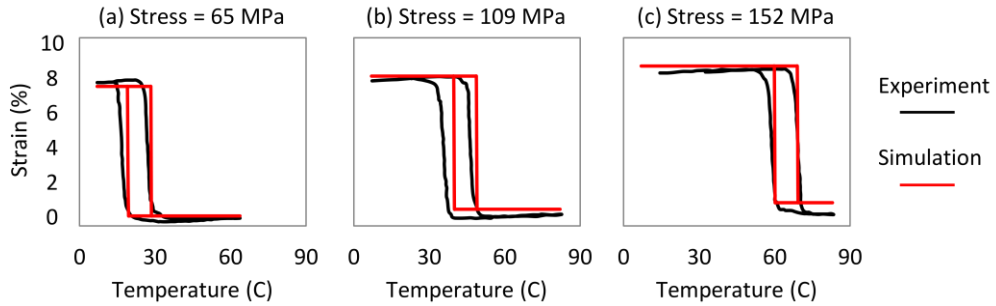


Figure 61: Thermal actuation of single-crystal SMA, simulations compared to test data from Fu et al. (1993)

The tension tests from Figure 60 were used to define the free energy and model parameters, then the same settings were used to simulate the actuations tests shown here. The experimental curves have, once again, somewhat rounded transitions as well as temperature plateaus that are not perfectly flat whereas the simulations show perfect rectangle hysteresis loops. The overall response is well represented but the disagreement is most pronounced in the 109 MPa case. There is also some disagreement between the high-temperature strain predictions that should be addressed. At high temperatures, the material is in the austenite phase and the strain is expected to scale linearly with applied stress, as is observed in the simulation data. The experimental strains, however, do not appear to follow this trend – the final strains values are relatively stagnant. This implies that some offset may have been present in the experimental data that was not accounted for in the simulations.

5.5.4 Relaxation Test

To verify the proper setting of the thermal activation parameters in the context of the new model, we show an alternate scenario to test and validate rate-independence. Here, we use a quasistatic relaxation test to examine the time-domain behavior of the kinetic equations. Figure 62a shows a quasi-static tension test on CuZnAl that

contains a sequence of pauses (Müller and Seelecke (2001)). The two zoomed in boxes show very slight discontinuities, but the macroscopic curve is sufficiently flat to regard the response as relaxation-free.

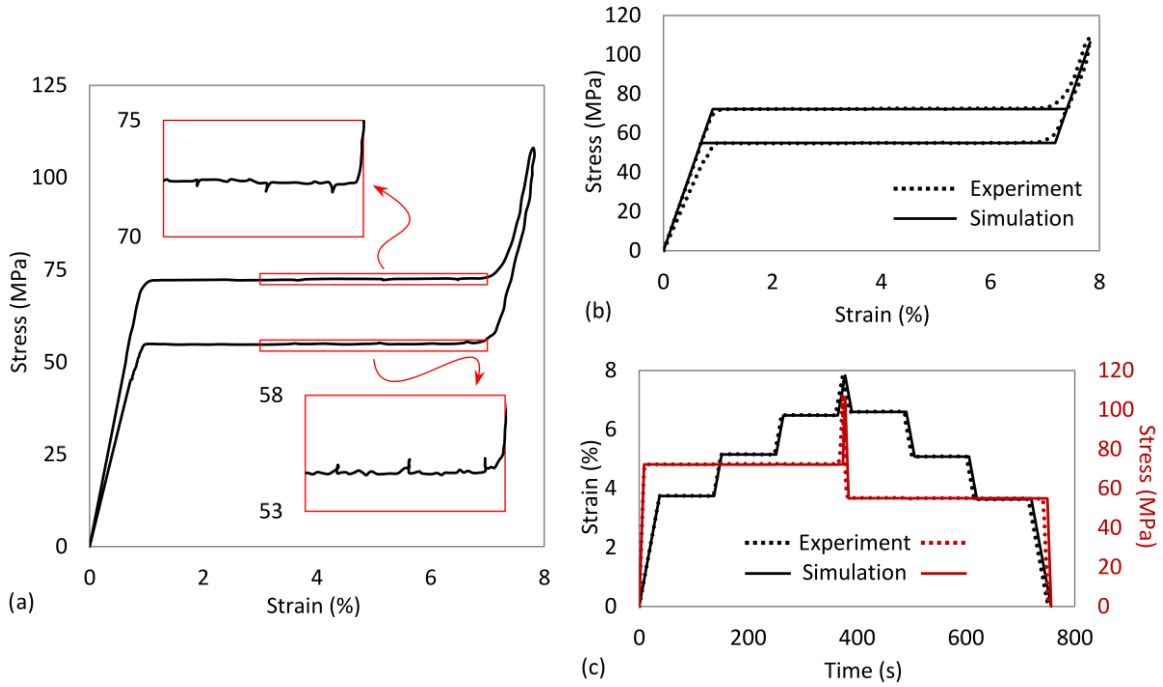


Figure 62: (a) Experimental data from Seelecke and Müller, 2001 (b-c) Simulations with properly chosen kinetic parameters

62b and c show the same experimental data along with the results of a simulation where the pauses were also included in the applied strain sequence. The simulations match the overall response well and, like the experiment, do not show any relaxation (reduction in stress) during the pauses. This indicates that the kinetic model is operating correctly and that the thermal activation parameters have been properly set. To give some additional insight into the effect of the thermal activation, Figure 63d and e show a second simulation where overly large values of τ and γ were used.

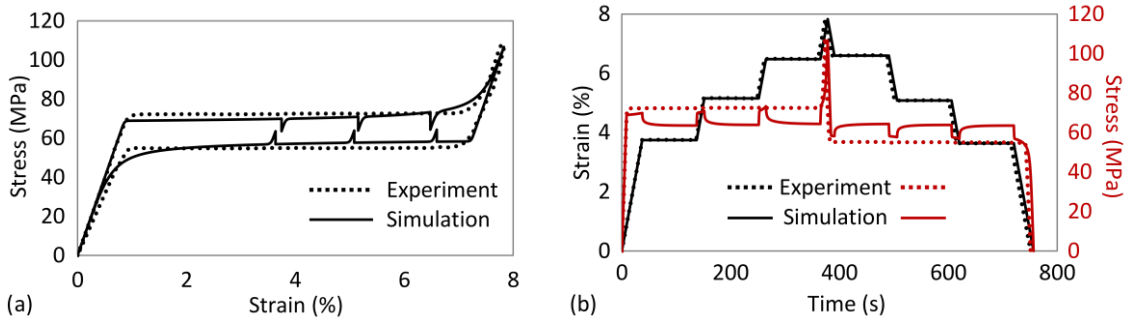


Figure 63: Relaxation simulations with improperly chosen kinetic parameters

On one hand, the choice of parameters seems to improve the agreement by adding smooth transitions between the linear branches and the stress plateaus (63d). The pause points however, show the artificial rate effects associated with this choice. The transformation progresses too slowly and, when the loading is paused, the backlog of untransformed layers catches up and this causes the stress to relax back toward the Maxwell line. A reversible, Maxwell-type, curve is the quasistatic limit of the high thermal activation case because the fluctuating layers are not sensitive to the energy barriers. Each time the applied strain restarts, some of the slowly transforming layers fall behind and the artificial hysteresis loop reforms. If the pauses were not included in Figure 63d, the stress-

strain agreement would falsely appear to be an improvement over the response in 62b. To prevent incorrect choices for τ and γ, the model should be tested in a relaxation case like the one shown here or by verifying that the model yields the same stress-strain curves under different strain rates.

5.5.5 Polycrystallinity

Next, we examine several different cases of polycrystalline SMAs. Figure 64 shows a set of simulations compared to the results from a tension experiment on a polycrystalline NiTiCuV specimen from Schmidt et al. (2016b).

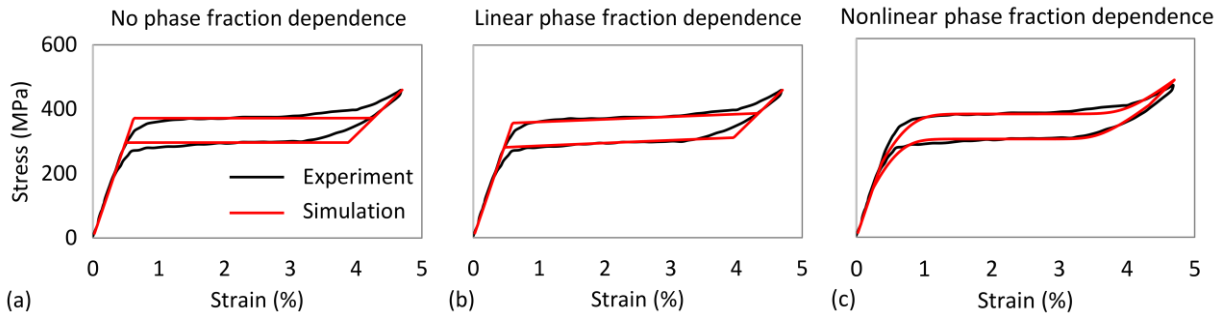


Figure 64: Simulations using various values of polycrystallinity parameters

64a-c all show the same experimental curve but each shows a simulation with different inhomogeneity parameters (from Equation 56). 64a shows the single crystal case, where the parameters σ_l, σ_b, and σ_c are equal to zero. 64b shows the linear polycrystallinity case, where only the linear coefficient σ_l has been assigned. In 64c, the exponential terms are activated to better capture the curve in the early and late stages of transformation. All three simulation cases give reasonable qualitative agreement and show that the mesoscopic inhomogeneity option allows a wide range of response characteristics.

Next we examine several more experiments from the same work that also involve latent heat effects. The experimental curve from Figure 64 was the quasistatic case from a set of tension tests that involved seven different strain rates. The entire set of stress-strain responses is shown in Figure 65.

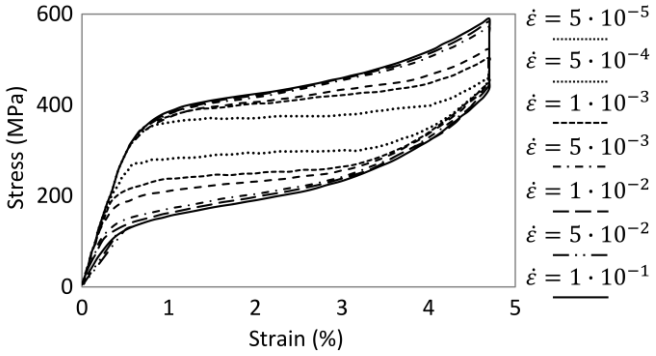


Figure 65: Tension test data at various strain rates from Schmidt et al. (2016b)

The $\dot{\varepsilon} = 5 \cdot 10^{-5}$ curve was used to define the base material properties and the first set of transformation stresses. The transformation stresses must be defined at two temperatures and, in this publication, a second set of stress-strain curves was not given. The authors did list a value for the parameter $\frac{\partial \sigma_{a+}}{\partial T}$ which, along with the assumption of constant hysteresis width and tension-compression symmetry, allowed us to fully define the material. The published convection coefficient and the actual wire dimensions were used to populate the energy balance parameters, but field effects were neglected. Figure 66 shows the results from point-wise simulations intended to represent the average specimen response.

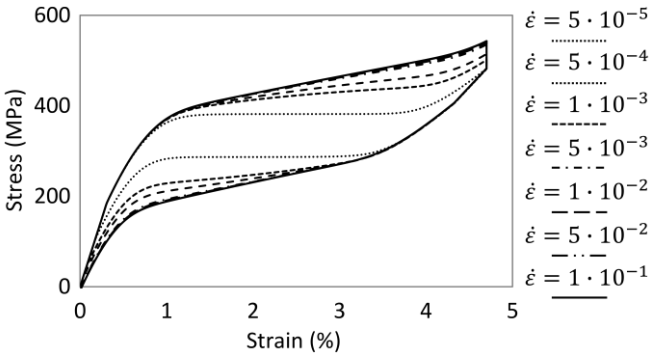


Figure 66: Simulations of tension tests from Schmidt et al. 2016b

The simulations show the same overall characteristics, but they under-predict the peak stress. Despite this shortcoming, several important aspects of the curves are well represented. The two highest strain rates stagnate in both data sets because the increasing strain rates lead to shorter heating events, which eventually become unable to further increase the temperature of the wire. The simulation cases also show a stress drop during the pause between loading and unloading, which is observed in the experiments. This decline in stress is due to untransformed austenite, which gets the opportunity to make the jump to martensite as the specimen cools and the critical stress decreases. Because the models under-predict the maximum stress, the decrease stress is also less than the observed value. To better compare the model results to the experiments, Figure 67 shows the highest strain rate curves from the Figures 65 and 66.

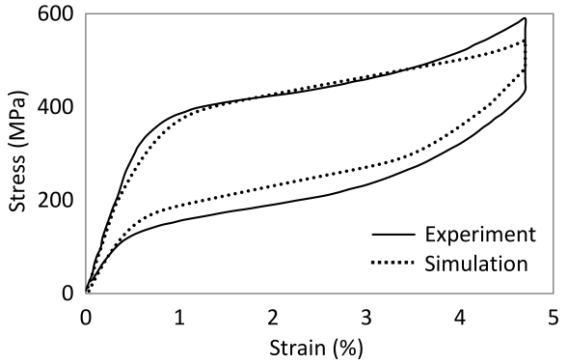


Figure 67: Simulation to experiment comparison for the highest strain rate ($\varepsilon = 1 \dot{\cdot} 10^{-1}$) case in Schmidt et al., 2016b

The simulation agrees well with the test data during the central portion of the forward transformation, but underpredicts the stress in the early and late stages. At the end of the forward transformation, the experimental curve continues to steepen while the simulation response is nearly linear. The experimental curve also experiences are greater decrease in stress during the pause between loading and unloading. The two curves maintain this offset throughout the entire unloading. Apart from the ≈20 MPa offset, the simulation matches the overall characteristic of the experimental unloading curve. We assume that much of the disagreement is an artifact of the more complex, spatially varying, thermal response of the ribbon. As the multiple transformation sites grow, their nonlinear influence on the specimen response becomes more dominant. The simulations, which neglect macroscopic field effects, do not capture this. The goal here is to examine the fundamental (non-spatial) self-heating behavior of the model using the new kinetics and inhomogeneity equations. Figures 66 and 67 show that the model is (a) operating as expected and (b) can predict behavior that is at least qualitatively correct without consideration of macroscopic field effects.

5.5.6 Tension-Compression Asymmetry

In Reedlunn et al. (2014), the authors examined the tension, compression, and bending behavior of set of NiTi tubes. The use of tube specimens and special fixturing allowed clear observation of the compressive response, which exhibited significant asymmetry compared to the tensile behavior. Measurement of strain fields in the specimens revealed that localization effects were present in tension but not in compression. This disparity lead to flat, single crystal-like, stress plateaus in tension while the hysteresis loop in compression exhibited the rounded and sloped shape typically associated with a textured SMA. Figure 68 shows experimental data from a quasistatic pair of tension and compression tests at room temperature along with simulation results.

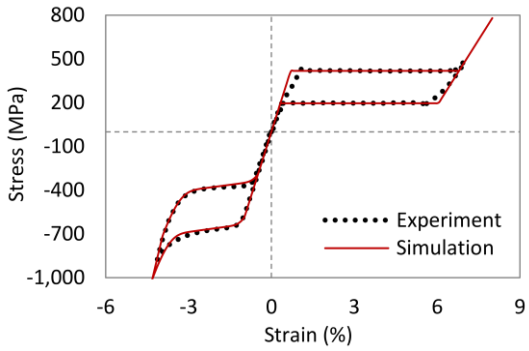


Figure 68: Simulation of tension and compression tests from Reedlunn et al. 2014

The first observation is that the model captures the both the single crystal-like tensile response and the polycrystalline-like compressive response in the same simulation. The tensile loading branch shows slight nonlinearity due to the presence of an intermediate phase that is not considered by the model, but the deviation is minor. The model also predicts a sudden, sharp transition into the reverse tensile stress plateau whereas the experimental curve has a slight curvature. We chose to neglect this effect in the simulations by setting the tensile local inhomogeneity parameters equal to zero. In tension tests such as this one, even though localization motivates the characteristic shape of the curve, the single crystal version of the model represents the behavior well. For the compressive response to be captured, both the linear and the exponential inhomogeneity terms had to be employed.

5.5.7 Temperature Dependent Hysteresis

The new formulation of the HFE also allows for temperature-dependent hysteresis width. To demonstrate this effect, we use the tension tests from the *Roundrobin SMA Modeling* program (Sittner et al. (2009)). NiTi wires were extended, partially unloaded, fully extended, then unloaded at quasistatic rates (i.e. isothermal) and these results are shown in Figure 69.

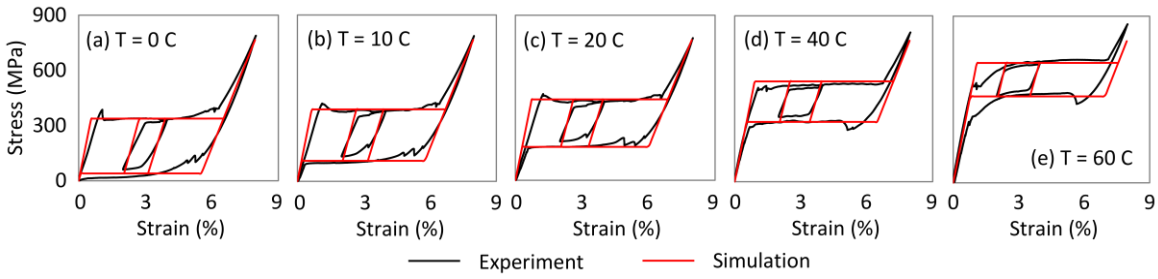


Figure 69: Simulation of tension tests from Sittner et al. (2009)

Apart from the local irregularities in each curve, both the partial and full hysteresis loops are well represented by the model. Comparing Figures 69a to 69e shows that the hysteresis width decreased by more than half. In some cases, a widening hysteresis loop can be explained by transitioning into a martensite to martensite loop, but this

normally occurs after the reverse transformation stress crosses the zero-stress line. The curves here are clearly superelastic so martensite to martensite transitions is not the explanation. All of the experimental stress-strain curves exhibit stress spikes at the beginning of transformation, which stem from macroscopic localization effects. It may also be the case that the variation in the hysteresis width is related to macroscopic effects like nucleation, but this is only speculation. Nucleation was neglected in the simulations, but the temperature dependent hysteresis in Figure 69 is too drastic to ignore. The reformulated free energy allows this effect to be captured without any other changes to the model.

5.5.8 Martensite Reorientation

The inner loops in Figure 69 represent partial transformations at superelastic temperatures. A similar phenomenon is observed when the model is cycled at a low temperature (below T_{am}), except the layers now distribute themselves between the two martensite wells. An example of this is shown in Figure 70, using the same model parameters as were used in Figure 69.

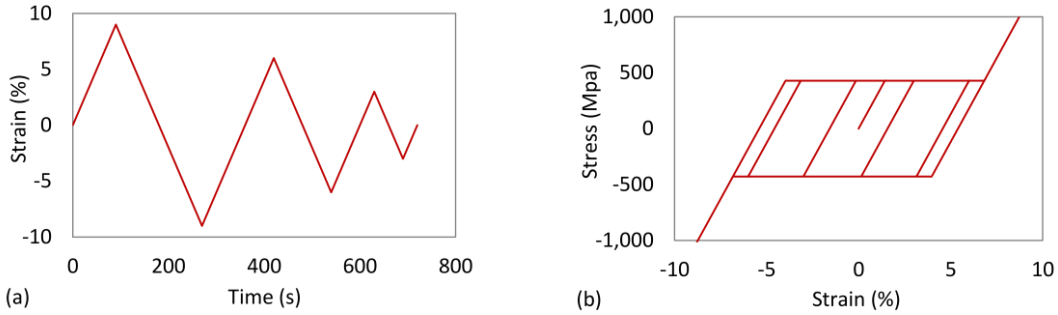


Figure 70: Martensite reorientation demonstrated in a low temperature tension-compression cycle

This behavior is a two-variant example of martensite reorientation, where an SMA material changes phase fractions without expressing (passing through) the parent austenite phase. Here, the layers jump back and forth between the two martensite energy wells shown in Figure 71 which causes the macroscopic response to assume the single hysteresis loop configuration. The inner loops are generated from a similar effect as in the superelastic case; the partially transformed material behaves linearly it reaches the reverse stress plateau.

5.6 Implementation Guide for the Updated 1-D Model

This chapter included several re-derivations and the manipulation of a few equations for the purposes of developing the updated model. Because of this, the final set of essential equations one would need to operate the model may not be completely clear. To address this, we include this chapter to summarize the essential parameters and equations needed to implement the model. Additionally, since several different SMA materials and experimental data sets were used in the *Simulations* section, we include multiple values in the material parameter table – one for each data set.

5.6.1 Material Parameters

Table 3 provides a summary of the material parameter symbols and is a companion to Table 4, which gives values on the following page.

Parameter	Definition
E_a	Elastic modulus of the austenite phase
E_+, E_-	Elastic modulus of the martensite phases ($+ \rightarrow$ tension, $- \rightarrow$ compression)
ε_R^+	Residual strain for the martensite $+$ phase (tension)

ε_R^-	Residual strain for the martensite – phase (compression)
T_H	The higher of the two test data temperatures
T_L	The lower of the two test data temperatures
σ_{a+}^H	Forward tensile transformation stress at T_H
σ_{+a}^H	Reverse tensile transformation stress at T_H
σ_{a-}^H	Forward compressive transformation stress at T_H
σ_{-a}^H	Reverse compressive transformation stress at T_H
σ_{a+}^L	Forward tensile transformation stress at T_L
σ_{+a}^L	Reverse tensile transformation stress at T_L
σ_b^+, σ_b^-	Sets the plateau stress variation as you enter and exit the austenite branch ($+ \rightarrow$ tension, $- \rightarrow$ compression)
σ_c^+, σ_c^-	Sets the plateau stress variation as you enter and exit the austenite branches ($+ \rightarrow$ tension, $- \rightarrow$ compression)
e_b^+, e_b^-	Sets the $\sigma - \varepsilon$ curvature as you enter and exit the austenite branch ($+ \rightarrow$ tension, $- \rightarrow$ compression)
e_c^+, e_c^-	Sets the $\sigma - \varepsilon$ curvature as you enter and exit the austenite branch ($+ \rightarrow$ tension, $- \rightarrow$ compression)
σ_l^+, σ_l^-	Sets slope of the stress plateau at the mid-transformation point ($+ \rightarrow$ tension, $- \rightarrow$ compression)
v	Activation volume, Thermal activation constant to ensure transformations occur rapidly
τ	Attempt frequency, Thermal activation constant to ensure transformations occur rapidly
K	Boltzmann constant
h_c	Convective heat transfer coefficient
c	Specific heat
ρ	Density
α	Surface area per unit volume

Table 3: Summary of material parameter definitions

5.6.2 Material Parameter Values

Several sets of simulations were presented in this paper, each corresponding to different experimental data sets. Table 4 shows a summary of the material parameter values used for each case. The table also included authors of the paper from which the data was extracted, the material used, as well as the figure numbers where the data and simulations are shown. Material parameters that have different tension and compression values are listed in the same row. If the same value was used for both (symmetry), only one value is listed; for tension-compression asymmetry cases, two values are shown. Also, some cases have thermal parameters listed as "-" and this indicates that the energy balance was not included in the simulations.

Source	Fu et al. (1993)	Müller and Seelecke (2001)	Schmidt et al. (2016b)	Reedlunn et al. (2014)	Sittner et al. (2009)	
Material	CuZnAl	CuZnAl	NiTiCuV	NiTi	NiTi	
Figure Number	60, 61	62, 63	64, 65, 67	68	69	
Parameter	**Value**	**Value**	**Value**	**Value**	**Value**	**Units**
E_a	12	8	60	60	60	GPa
E_+, E_-	8	8	20	30, 80	30, 30	GPa
ε_R^+	6.6	6.5	2.4	5.4	5.4	%
ε_R^-	$-\varepsilon_R^+$	$-\varepsilon_R^+$	$-\varepsilon_R^+$	–3.05	$-\varepsilon_R^+$	%
T_H	65	30	30	22	60	C
T_L	19	20	20	12	–30	C
σ_{a+}^H	163	94.3	445	420	650	MPa
σ_{+a}^H	143	76.3	365	195	465	MPa
σ_{a-}^H	$-\sigma_{a+}^H$	$-\sigma_{a+}^H$	$-\sigma_{a+}^H$	–675	$-\sigma_{a+}^H$	MPa
σ_{-a}^H	$-\sigma_{+a}^H$	$-\sigma_{+a}^H$	$-\sigma_{+a}^H$	–375.5	$-\sigma_{+a}^H$	MPa
σ_{a+}^L	65	72.62	375	360	185	MPa
σ_{+a}^L	45	54.62	295	130	–175	MPa
σ_b^+, σ_b^-	0	0	178, 0	0, 50	0	MPa
σ_c^+, σ_c^-	0	0	178, 0	0, 300	0	MPa
e_b^+, e_b^-	1	1	10, 0	12, 12	1	$Unitless$
e_c^+, e_c^-	1	1	10, 0	12, 6	1	$Unitless$
σ_l^+, σ_l^-	0	00	0	0, 75	0	MPa
v	$1.0 \cdot 10^{-21}$	$1.0 \cdot 10^{-21}$	$1.0 \cdot 10^{-21}$	$1.0 \cdot 10^{-21}$	$1.0 \cdot 10^{-21}$	m^3
τ	0.001	0.001	0.001	0.001	0.001	$1/s$
K	$1.38 \cdot 10^{-23}$	$1.38 \cdot 10^{-23}$	$1.38 \cdot 10^{-23}$	$1.38 \cdot 10^{-23}$	$1.38 \cdot 10^{-23}$	J/kg
h_c	-	-	15.0	-	-	$W/m^2 \cdot K$
c	-	-	450	-	-	$J/kg \cdot K$
ρ	-	-	7,340	-	-	kg/m^3
α	-	-	3,867	-	-	m^2/m^3

Table 4: Material parameter values used for all simulation cases that were compared to experimental data

5.6.3 Equations

Figure 72 is a flowchart of the equations required to implement the model. Since the main paper contains several derivations with intermediate equations, this table collects all of the essential equations in their final, ready to be coded, form. Execution of the model requires numerical integration of the kinetic ODE's and the figure here represents a single, strain-controlled, solution step. The *Inputs* box contains the known initial conditions and the equations sets 1-11 are the functions required to fully define (and numerically solve) the ODE's. The *Outputs* box contains the updated parameters after the solution/time step. The supporting expressions (sets 1-11) are also evaluated once more after solution to get updated values for parameters like stress or latent heat.

Numerical integration from t_1 to $t_1 + \Delta t$

Inputs (at $t = t_1$): $t, \Delta t, \varepsilon, \Delta\varepsilon$; σ, T, x_-, x_+

$$\begin{Bmatrix} \dot{x}_- \\ \dot{x}_+ \\ \dot{T} \\ 1 \end{Bmatrix} = \begin{Bmatrix} x_a\, p_{a-}(\sigma,T) - x_-\, p_{-a}(\sigma,T) \\ x_a\, p_{a+}(\sigma,T) - x_+\, p_{+a}(\sigma,T) \\ \frac{1}{\rho c}\left(\dot{x}_+ h_+ + \dot{x}_- h_- - \alpha h_c (T - T_0)\right) \\ x_- + x_a + x_+ \end{Bmatrix}$$

Outputs (at $t = t_1 + \Delta t$): x_-, x_+, T

Functions required to define the right-hand side of the ODE's

1) Transformation stress temperature-dependence parameters

$$\frac{d\sigma_{a+}}{dT} = \frac{\sigma_{a+}^H - \sigma_{a+}^L}{T_H - T_L} \qquad \frac{d\sigma_{+a}}{dT} = \frac{\sigma_{+a}^H - \sigma_{+a}^L}{T_H - T_L} \qquad T_{am} = T_H - \frac{\sigma_{a+}^H}{\frac{d\sigma_{a+}}{dT}} \qquad \frac{d\sigma_{a-}}{dT} = \frac{\sigma_{a-}^H - 0}{T_H - T_{am}} \qquad \frac{d\sigma_{-a}}{dT} = \frac{\sigma_{-a}^H - \sigma_{-a}^L}{T_H - T_L}$$

2) Transformation stresses

$$T^* = max(T, T_{am})$$

$$\sigma_{a+}(T) = \sigma_{a+}^H + \frac{d\sigma_{a+}}{dT}(T^* - T_H) \qquad \sigma_{+a}(T) = min\left(\sigma_{+a}^H + \frac{d\sigma_{+a}}{dT}(T^* - T_H), \sigma_{a+}(T)\right)$$

$$\sigma_{a-}(T) = \sigma_{a-}^H + \frac{d\sigma_{a-}}{dT}(T^* - T_H) \qquad \sigma_{-a}(T) = max\left(\sigma_{a-}(T), \sigma_{-a}^H + \frac{d\sigma_{-a}}{dT}(T^* - T_H)\right)$$

3) Effective transformation stress considering inhomogeneity

$$\sigma_{a\pm}^{eff}(T, x_\pm) = \sigma_{a\pm}(T) \pm \sigma_v^\pm(x_\pm)$$

$$\sigma_{\pm a}^{eff}(T, x_\pm) = \sigma_{\pm a}(T) \pm \sigma_v^\pm(x_\pm)$$

$$\sigma_v^\pm(x) = \sigma_l^\pm\left(x - \frac{1}{2}\right) + \begin{cases} -\sigma_b^\pm (1-2x)^{e_b^\pm} & if\ x < 1/2 \\ +\sigma_c^\pm (2x-1)^{e_c^\pm} & if\ x \geq 1/2 \end{cases}$$

4) Critical strains

$$\varepsilon_{a\pm}(T, x_\pm) = \frac{\sigma_{a\pm}^{eff}(T, x_\pm)}{E_a}$$

$$\varepsilon_{\pm a}(T, x_\pm) = \frac{\sigma_{\pm a}^{eff}(T, x_\pm)}{E_\pm} + \varepsilon_R^\pm$$

5) Current stress value

$$\sigma(\varepsilon, x_+, x_-) = \frac{\varepsilon - \varepsilon_R^+ x_+ - \varepsilon_R^- x_-}{\frac{1 - x_- - x_+}{E_a} + \frac{x_+}{E_+} + \frac{x_-}{E_-}}$$

6) Average strain for each phase

$$\varepsilon_a(\sigma) = \frac{\sigma}{E_a} \qquad \varepsilon_\pm(\sigma) = \frac{\sigma}{E_\pm} + \varepsilon_R^\pm$$

7) Caloric parameters and chemical free energies

$$\beta_\pm^L = C_2^\pm {\sigma_{a\pm}^L}^2 + C_1^\pm(T_L)\sigma_{a\pm}^L + C_0^\pm(T_L)$$

$$\beta_\pm^H = C_2^\pm {\sigma_{a\pm}^H}^2 + C_1^\pm(T_H)\sigma_{a\pm}^H + C_0^\pm(T_H)$$

$$C_0^\pm(T) = -\frac{\left(\sigma_{a\pm}^H - \sigma_{\pm a}^H\right)\varepsilon_R^\pm}{2} \qquad C_1^\pm(T) = \varepsilon_R^\pm - \frac{\left(\sigma_{a\pm}^H - \sigma_{\pm a}^H\right)E_\Delta^\pm}{2} \qquad C_2^\pm = \frac{E_\Delta^\pm}{2} \qquad E_\Delta^\pm = \frac{1}{E_\pm} - \frac{1}{E_a}$$

$$s_\pm^R = \frac{\beta_\pm^H - \beta_\pm^L}{T_H - T_L} = \frac{d\beta_\pm}{dT} \qquad u_\pm^R = s_\pm^R T_H - \beta_\pm^H \qquad \beta_\pm(T) = s_\pm^R T - u_\pm^R \qquad h_\pm(\sigma) = \frac{\sigma^2}{2}E_\Delta^\pm + \sigma\varepsilon_R^\pm + u_\pm^R$$

8) Gibbs free energy functions

$$g_a(\varepsilon;\sigma) = \frac{E_a}{2}\varepsilon^2 - \sigma\varepsilon$$

$$g_\pm(\varepsilon, T;\sigma) = \frac{E_\pm}{2}\left(\varepsilon - \varepsilon_R^\pm\right)^2 + \beta_\pm(T) - \sigma\varepsilon$$

9) Gibbs free energy minima and inflection point values

$$G_-(\sigma, T, x_\pm) = g_-\left(min\left(\varepsilon_-(\sigma), \varepsilon_{-a}(T, x_\pm)\right);\sigma\right)$$

$$G_+(\sigma, T, x_\pm) = g_+\left(max\left(\varepsilon_{+a}(T, x_\pm), \varepsilon_+(\sigma)\right);\sigma\right)$$

$$G_a(\sigma, T, x_\pm) = g_a\left(max\left(\varepsilon_{a-}(T, x_\pm), min\left(\varepsilon_a(\sigma), \varepsilon_{a+}(T, x_\pm)\right)\right), T;\sigma\right)$$

$$G_{ij}(\sigma, T, x_\pm) = g_i\left(\varepsilon_{ij}(T, x_\pm), T;\sigma\right) \qquad ij \in -a\ ,\ a-\ ,\ a+\ ,\ +a$$

10) Relative energies

$$\Delta G_{\pm a}(\sigma, T, x_\pm) = G_{\pm a}(\sigma, T, x_\pm) - G_\pm(\sigma, T, x_\pm)$$

$$\Delta G_{a\pm}(\sigma, T, x_\pm) = G_{a\pm}(\sigma, T, x_\pm) - G_a(\sigma, T, x_\pm)$$

11) Transformation probabilities

$$p_{ij}(\sigma, T, x_\pm) = \frac{1}{\tau}e^{-\frac{1}{\gamma}\Delta G_{ij}(\sigma, T, x_\pm)} \qquad ij \in -a\ ,\ a-\ ,\ a+\ ,\ +a$$

Figure 72: Flow chart of essential equations for model implementation

6 2-D Model

6.1 Chapter Overview

In this chapter, we extend the model to a multiaxial case where the SMA material experiences both normal and shear transformations. To accomplish this, we construct a free energy on a 2-D, normal strain-shear strain domain. This allows the identification of four mesoscopic martensite variants and a model that can be applied to any combination of axial and shear loading. The chapter begins with a summary of the experimental data that motivates the formulation and is also used later to test/validate the model. We introduce the model by first

proposing a mesoscopic arrangement of four martensitic variants that could result in the macroscopic deformations observed in the experiments. Identification of specific mesoscopic phases allows us to construct mesoscopic Helmholtz and Gibbs free energies. In this section we also define material parameters, latent heat properties, as well as the transformation criteria which are new and unique to the 2-D model. After construction of the 2-D free energy, we use it to derive the macroscopic stress-strain equations and the kinetics law. The kinetics law is based on the 1-D version presented in the *Kinetics* chapter but has been newly formulated to accommodate the multiple transformations paths possible in a 2-D free energy landscape. We conclude the chapter by showing simulations of a variety of experiments and discussing the model's performance.

6.2 Experimental Motivation

The *RR* experimental and modeling program was discussed in the introduction section, but mainly from the perspective of modeling approaches. The central piece of this effort was an extensive set of experiments that examined NiTi wire specimens under numerous temperature, tensile stress, and shear stress conditions. We will use the results of these experiments to develop and test the 2-D model so we begin with a summary of the data.

6.2.1 Roundrobin SMA Experiments

The *RR* data set is broken into three major sets, with different subdivisions in each. Table 5 shows a list of the experiments within each group along with identification numbers (left column) that are assigned here for concise reference back to a particular test or group.

Data Set ID	Description of the experiment
Set 1	Uniaxial tension tests, Includes 5 sets of experiments
Set 1. 1	Wire training prior to the tension tests, 100 load-unload cycles
Set 1. 2	Tension tests with a full load-unload cycle, At 7 temperatures
Set 1. 3	Tension tests with a partial load-unload cycle, At 7 temperatures
Set 1. 4	Cooling-heating cycles under constant stress, 7 stress values
Set 1. 5	Cooling-heating cycles at a fixed strain, 6 locations on the $\sigma - \varepsilon$ curve
Set 2	Combined tension and torsion tests
Set 2. 1	Axial stress = 70 MPa
Set 2. 1. 1	Wire training prior to the tension-torsion test, Torsion cycles at 50 C
Set 2. 1. 2	Austenite reset – One heating cycle after training
Set 2. 1. 3	Tension-torsion tests, Temperatures from -30 C to 50 C
Set 2. 2	Axial stress = 132 MPa
Set 2. 2. 1	Wire training prior to the tension-torsion test, Torsion cycles at 50 C
Set 2. 2. 2	Austenite reset – One heating cycle after training
Set 2. 2. 3	Tension-torsion tests, Temperatures from -15 C to 50 C
Set 2. 3	Axial stress = 194 MPa
Set 2. 3. 1	Wire training prior to the tension-torsion test, Torsion cycles at 50 C
Set 2. 3. 2	Austenite reset – One heating cycle after training
Set 2. 3. 3	Tension-torsion tests, Temperatures from -10 C to 50 C
Set 2. 4	Axial stress = 255 MPa
Set 2. 4. 1	Wire training prior to the tension-torsion test, Torsion cycles at 50 C
Set 2. 4. 2	Austenite reset – One heating cycle after training
Set 2. 4. 3	Tension-torsion tests, Temperatures from 5 C to 50 C
Set 2. 5	Axial stress = 317 MPa

Set 2. 5. 1	Wire training prior to the tension-torsion test, Torsion cycles at 50 C
Set 2. 5. 2	Austenite reset – One heating cycle after training
Set 2. 5. 3	Tension-torsion tests, Temperatures from 20 C to 50 C
Set 2. 6	Axial stress = 379 MPa
Set 2. 6. 1	Wire training prior to the tension-torsion test, Torsion cycles at 50 C
Set 2. 6. 2	Austenite reset – One heating cycle after training
Set 2. 6. 3	Tension-torsion tests, Temperatures from 20 C to 50 C
Set 3	*Constant torque tests*
Set 3. #	*Axial stress = 70, 132, 194, 255, 317 MPa for # = 1, 2, 3, 4, 5 respectively*
Set 3. #. 1	*Wire training prior to the test*
Set 3. #. 2	*Temperature cycles under constant axial stress and torque*

Table 5: List of experiments from the *Roundrobin SMA Modeling* program Sittner et al. (2009)

The third data set, the *Constant torque* tests, is listed here for completeness but is not specifically addressed or included in the model testing. The behavior observed in *Set 3* is within the capabilities of the model but does not require any features that are not well displayed by modeling *Set 2* (the tension-torsion tests). Therefore, only the uniaxial tension tests (*Set 1*) and tension-torsion tests (*Set 2*) will b used to develop and test the 2-D model.

Uniaxial Tension Tests

The first set of uniaxial tension tests (Set 2. 1) were conducted as follows:

1) The wire was trained with 100 tension cycles at 22 C
2) The specimen was heated or cooled uniformly to a set test temperature
3) The specimen was loaded and unloaded through a full, strain-controlled, tension cycle at a quasistatic rate

Figure 73 shows individual plots for each specimen.

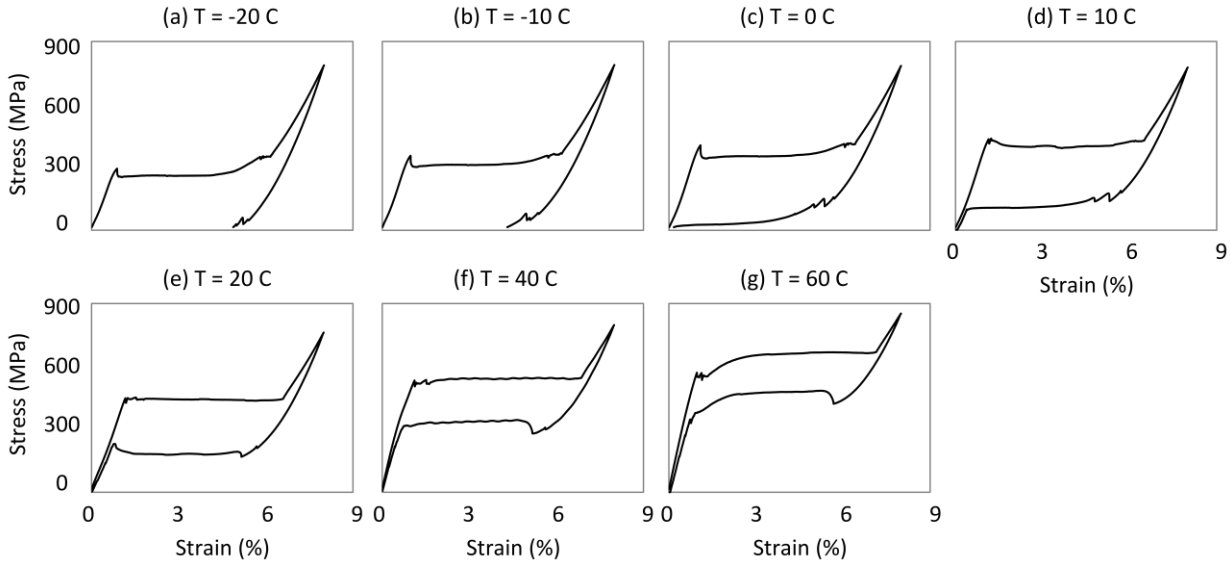


Figure 73: *RR* data set 1. 2: Uniaxial tension tests from the *RR* data set

It is important to note that these curves show signs of macroscopic localization effects, implying that grip effects influenced their shape along with the basic material response. The first of these observations is that, either during forward or reverse transformation, the responses have small stress peaks (spikes) as they enter or exit transformation. These appear when a small zone (or set of zones) begins to transform first and causes a small

change in the load cell reading before the transformation begins to propagate along the wire. The second localization effect is the flatness of the stress plateaus. Since these wires were trained prior to the tests, the curves are not expected to have the perfectly flat plateaus associated with virgin, single crystal specimens. However, some of the curves (73e and f for example) have very flat transformation zones. In a trained specimen, flat stress plateaus can also be caused by grip effects. The transformation begins at the ends of the wire and the overall stress-strain response reflects the transformation traveling along the wire as much as the fundamental material behavior.

The next set of uniaxial tests was similar to the first except it included a partial unload cycle. The tests were conducted in the following sequence:

1) Wire training – 100 tension cycles at 22 C
2) Allow the specimen to uniformly reach the test temperature
3) Load to ≈4% strain, which is near the middle of the forward transformation plateau
4) Unload enough to reach (or nearly reach) the reverse plateau
5) Reload beyond the point of complete transformation (≈8% strain)
6) Unload to zero load (not necessarily zero strain)The specimen was loaded and unloaded through a full, strain-controlled, tension cycle at a quasistatic rate

This data is shown in Figure 74.

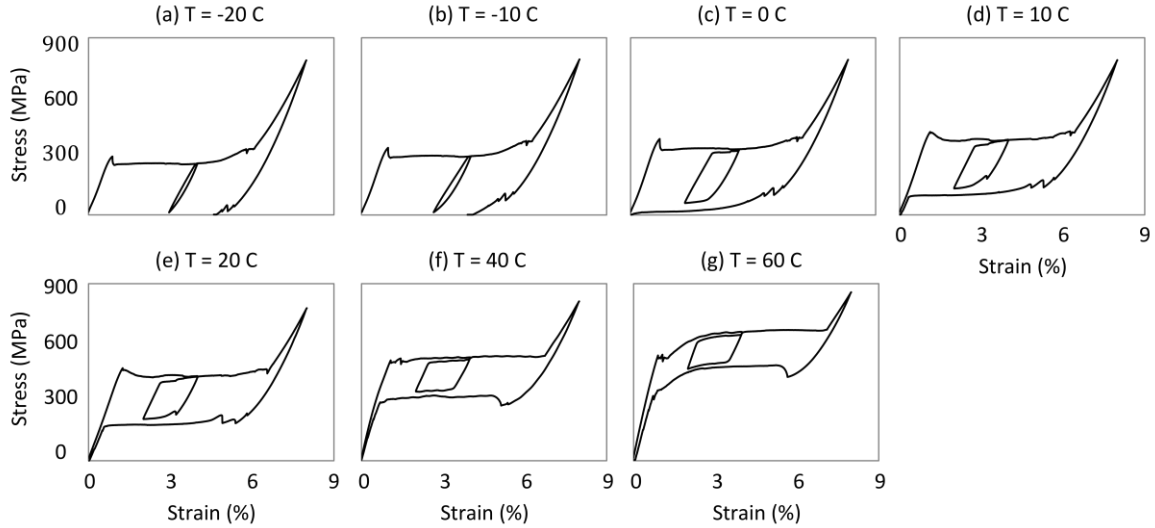


Figure 74: *RR* data set 1. 3: Uniaxial tension tests with partial load-unload cycles, Individual plots

Note here that the responses have very similar stress localization effects when compared to the specimens that did not have a partial cycle (Figure 73). It is not specifically stated in the documentation from the *RR* program (Sittner et al., 2009) that the same specimens from data set 1. 2 were then used for the 1. 3 experiments, but the similarities in the curves imply that this is the case.

Uniaxial Thermal Cycling Tests

In the next set of tests, the wire was tested in actuation mode under various constant axial stresses. The uniaxial thermal cycling tests were conducted as follows:

1) Wire training – 100 tension cycles at 22 C
2) Apply axial stress (hanging mass on the wire)
3) Heat specimen to a uniform highly austenitic temperature (70 or 100 C)
4) Cycle the temperature between ≈ -20 and 45 C (Heating/cooling rate ≈ 0. 1 $^{deg}/_{s}$)

The results from these experiments is shown in in Figure 75. The most outstanding observation is that there is a distinct stress threshold (between 300 and 400 MPa) below which actuation does not occur. This implies that, even at very low temperatures, there is an energy barrier keeping material in the austenite phase. Physically, this is likely due to mesoscopic inhomogeneity. Local stress and strain fields keep the austenite stable and the external load somewhere between 300 and 400 MPa is required to overcome them.

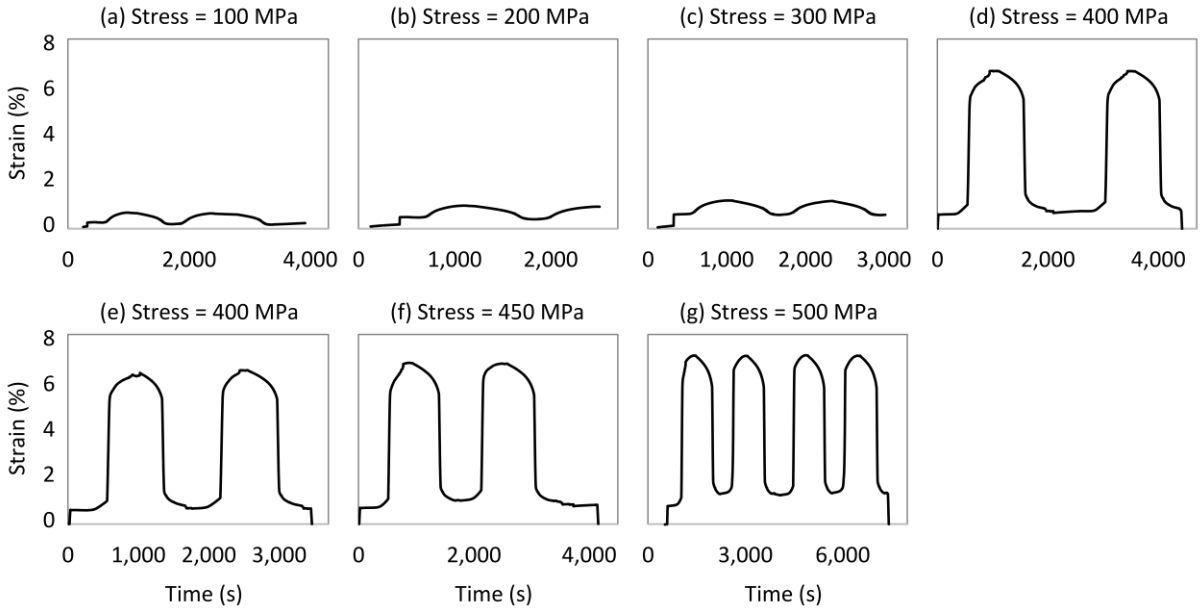


Figure 75: RR data set 1. 4: Cooling-heating cycles under constant axial stress, Individual specimen plots

The inhomogeneity effects are also seen in Figure 75 in the rounded shape of the curves as they enter and exit transformation in 75d-g. At the lower temperatures, there are bumps in the curves where the temperature increases. Phase transformation may account for some small part of this behavior, but it is mostly the linear thermal expansion of the wire.

Next, in experimental data set 1. 5, wires were extended to a specified strain and held there while undergoing thermal cycles.

1) Wire training – 100 tension cycles at 22 C
2) Heat specimen to a uniform 22 C (which is a superelastic temperature)
3) Extend the wire to a specified strain then hold (the fixed strain could be on the load or unload branch)
4) Cycle the temperature between ≈ 22 and 100 C (Heating/cooling rate ≈ 0. 15 $^{deg}/_{s}$)

In Figure 76 we see that three of the specimens were stopped on the forward transformation plateau and three were stopped on the reverse plateau but, in all cases, the overall response to the temperature cycles is the same. Each specimen exerts a load on the grip that is proportional to the increasing temperature and the result is the stress spike shown in each of the stress-strain curves. When each specimen in stopped, some of the material is suspended in a martensite configuration due to a combination of stress and local energy barriers. When the temperature is increased, the stress required to maintain this metastable martensite state also increases. At the point where the externally applied stress is no longer sufficient, the martensite layers return to austenite which causes a large increase in stress. When the temperature is lowered back to the original temperature, the metastable martensite state once again becomes available and the layers return to it because it minimizes the stress (and GFE) in the wire.

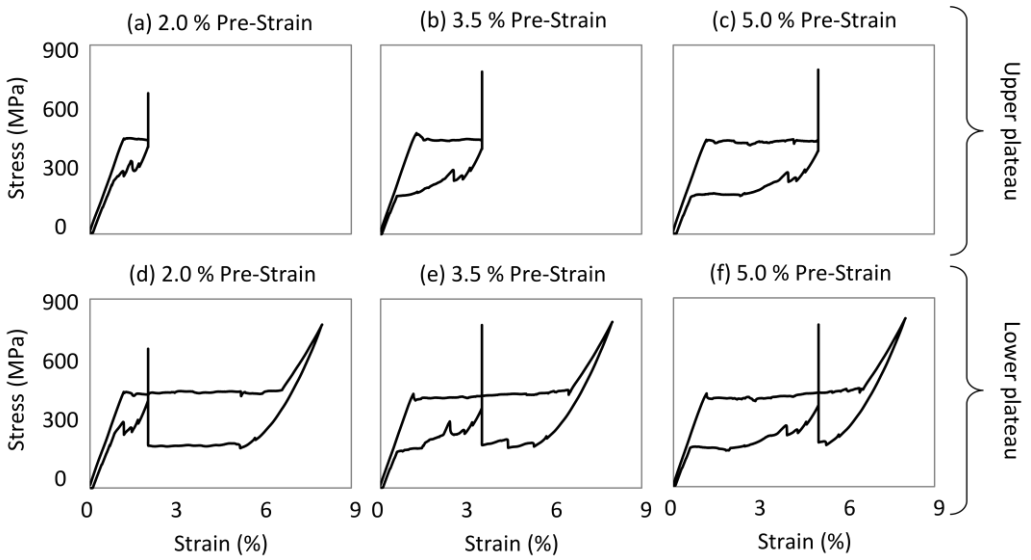


Figure 76: RR data set 1. 5: Cooling-heating cycles at a fixed strain, Individual specimen plots

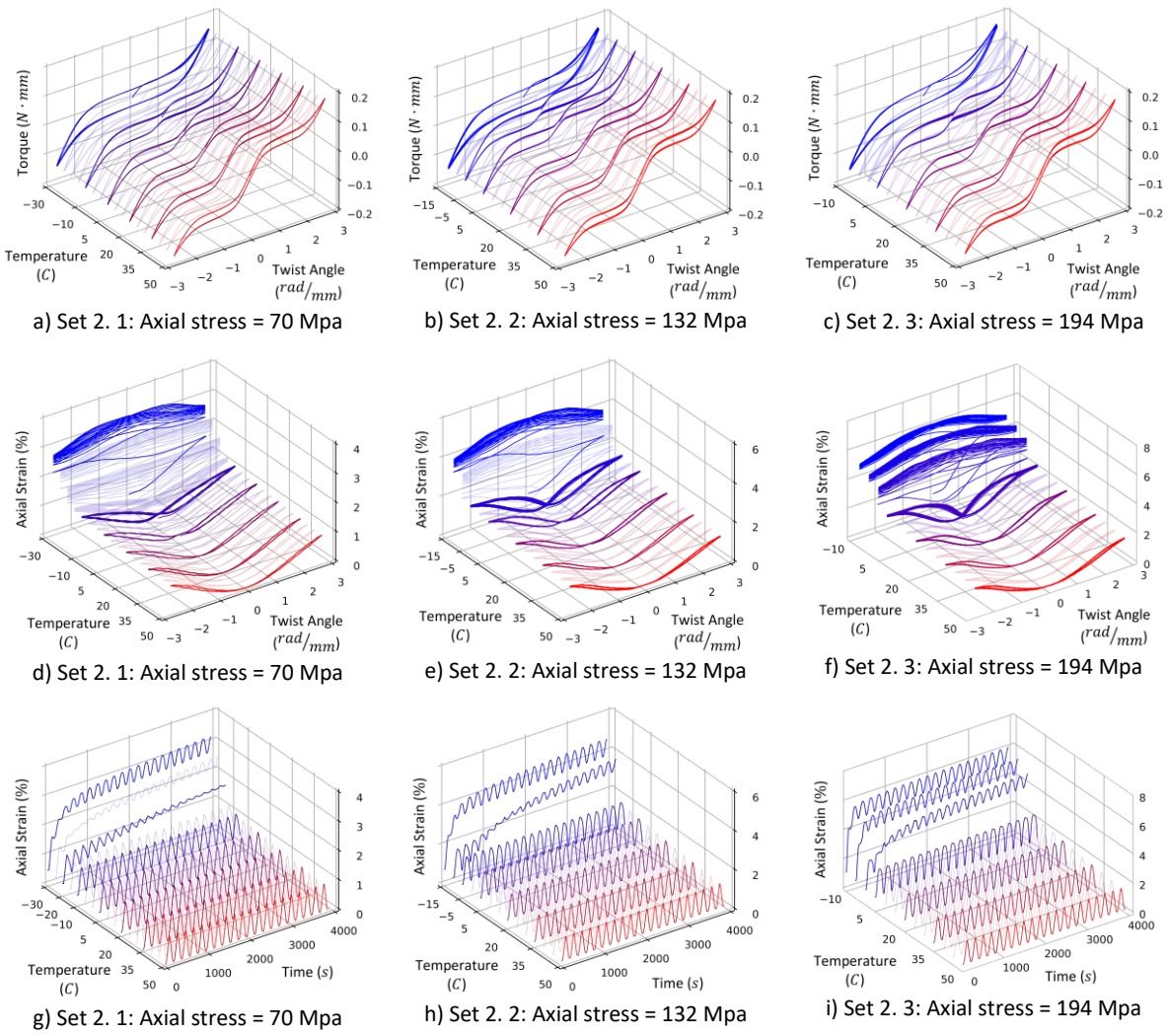
Both sets of plots (a-c and d-f) show that the return to the unloading curves is irregular. This is consistent with the unloading branches from the constant-temperature tension tests (Figures 73 and 74) and is assumed to be caused by localization effects in the wire. This completes the uniaxial tests and the next data set begins the multiaxial loading portion of the *RR* program.

Tension-Torsion Tests

Data set 2 represents a series of experiments where wires were loaded with a constant axial force and then subjected to twisting cycles. Six different axial stresses were examined and, for each one, the wire was twisted at several different temperatures. The test procedure for each set of tension-torsion test was as follows:

1) Apply axial load: Hang a weight from the wire to impose a constant axial stress while leaving the wire free to extend or contract
2) Training: Train the wire for 50 full twist cycles at 50 C
3) Austenite reset: Heat the wire to 100 C then cool it to the specified temperature to remove residual martensite
4) Cycling: Apply full twist cycles at the specified temperature while recording torque and axial strain

The first three data sets, which represent axial stresses of 70, 132, and 194 MPa, are summarized in Figure 77.



a) Set 2. 1: Axial stress = 70 Mpa
b) Set 2. 2: Axial stress = 132 Mpa
c) Set 2. 3: Axial stress = 194 Mpa
d) Set 2. 1: Axial stress = 70 Mpa
e) Set 2. 2: Axial stress = 132 Mpa
f) Set 2. 3: Axial stress = 194 Mpa
g) Set 2. 1: Axial stress = 70 Mpa
h) Set 2. 2: Axial stress = 132 Mpa
i) Set 2. 3: Axial stress = 194 Mpa

Figure 77: Tension-torsion experiments, Axial stresses 70, 132, and 194 Mpa

Due to the number of temperatures examined some of the curves from in Figure 77 have normally colored lines and some are nearly transparent. The goal of displaying the data this way is to convey the general trends while also maintaining clarity. The colors range from red at the higher temperature to blue at lower ones. No color legend is included though because the temperatures for each curve are marked on the lower left axis. The upper row (77a-c) shows torque-twist curves at various temperatures and the curves have the smooth shapes associated with uniaxial polycrystalline or highly trained specimens. The shape of the curves here, however, is mainly due to the radial strain variation in the wire cross section. Each radial segment of the wire experiences a different amount of strain at a given twist angle, therefore the global torque plateaus begin and end gradually. It is also expected that the wire training itself and global lcoalization effects could also be playing a role in the shape of the curves but it is difficult to completely parse the individual effects from the published data. Each set of curves in 77a-c follows the expected progress of an SMA material where the high temperature mode is characterized by two distinct hysteresis loops that are separated by a linear austenite branch. As the temperature is lowered, the hysteresis loops more toward the origin and eventually merge into a single loop that is centered about the origin. This signifies the transition from high-temperature austenite to martensite ($a \leftrightarrow m$) transformations to low-temperature martensite to martensite ($m \leftrightarrow m$) transformations. The effect of the axial stress become apparent when the temperatures where mode transitions occur are examined. At -10 C, the 70 Mpa curve (77a) still has two

distinct hysteresis loops, though they are beginning to merge. In the 194 Mpa set of tests however, -10 C corresponds to a single $m \leftrightarrow m$ hysteresis loop. This implies that the axial stress acts to increase the temperature where the mode shift occurs.

The axial strain-twist-temperature plots in 77d-f also show two distinct modes, one at high temperatures and one at low temperatures. At high temperatures each twist of the wire, whether positive or negative, results in an axial elongation of the wire. The magnitudes of the axial strain remain relatively constant as the temperatures decrease till the $m \leftrightarrow m$ transition occurs. The $m \leftrightarrow m$ transformations cause a large jump in the axial strain and the shape of the loops inverts – the axial strain at low temperatures is now maximal at zero twist. The characteristic shapes of the high and low temperature curves is similar for the three levels of axial stress, but the magnitudes of the axial strains increases with axial stress. At high temperatures the maximum axial strain is approximately 1% at 70 MPa (77a) and 2% at 194 MPa. Both 1% and 2% axial strain are much greater than could be expected from elastic deformation (194 MPa divided by the elastic modulus in the austenite phase yields 0.03% strian). This implies that the effects of the axial strain are transformation-related. The increase in axial strian is more drastic at low temperatures. The maximum low-temperature axial strain at 70 Mpa is about 3% and almost 6% at 194 Mpa.

The high-temperature, low-temperature modes can also be seen in the axial strain-time-temperature plots in 77g-i. These curves more clearly illustrate the racheting behavior that occurs at temperatures in the transition zone. At 194 Mpa (77i), both the -5 C and 0 C curves begin their oscillations at lower axial strians, similar to the high temperature curves. But, after a few cycles, they work their way up to the higher level of the low temperature mode. This implies the presence of microscopic or mesoscopic phase field evolution. The initial, temperature-induced, phase fields and local stress fields can change with cyclice loading as dislocations form, grow, shrink, or move. Global and macroscopic field effects (often bundled with the term localization) tend to appear in the first cycle so they are likely not responsible for the racheting observed here.

The second set of experiments, at 255, 317, and 379 MPa, are shown in Figure 78.

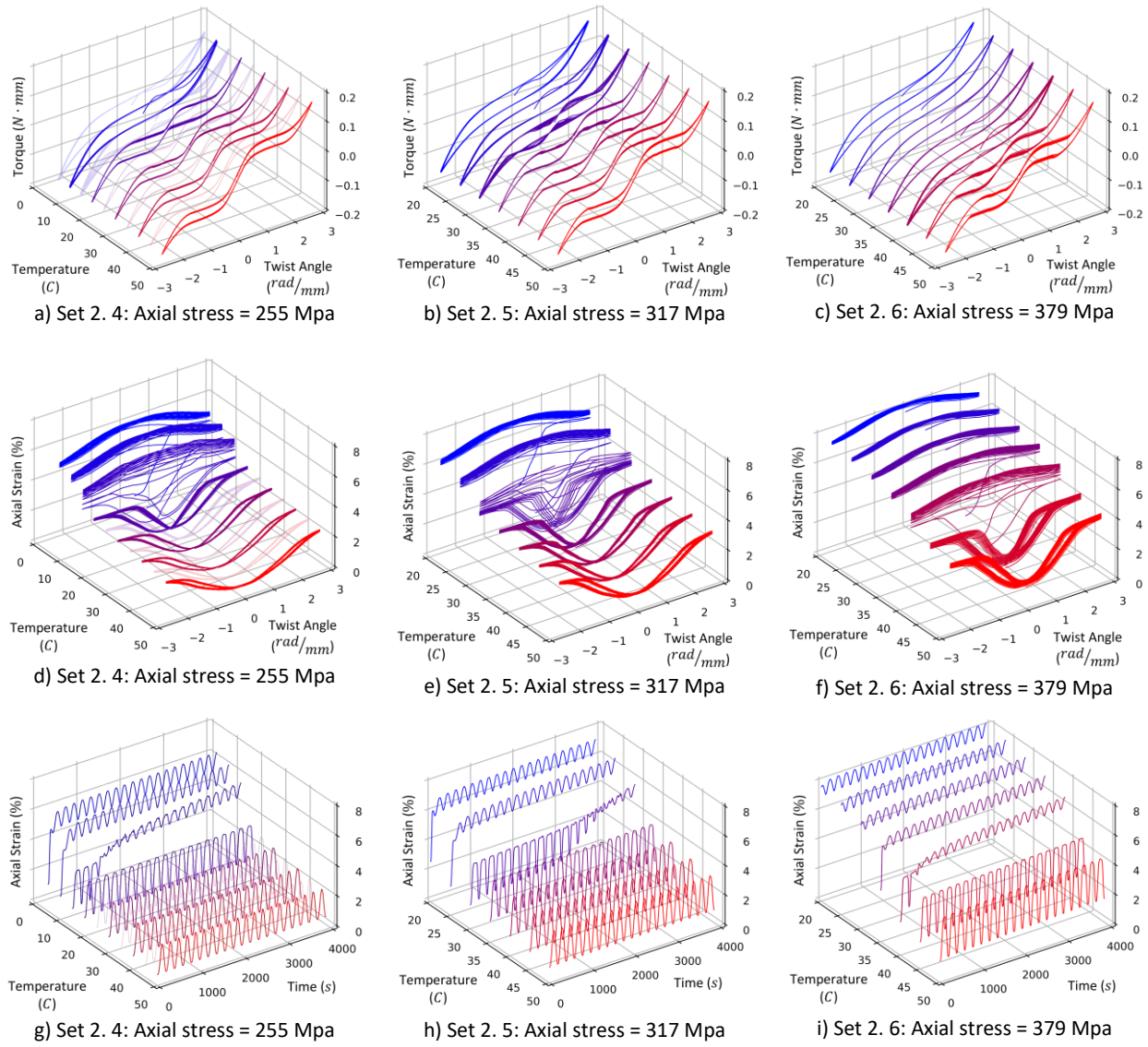


Figure 78: Tension-torsion experiments, Axial stresses 255, 317, and 379 Mpa

The same trends continue at the three higher levels of stress. The transition temperature continues to increase with axial stress; the transitions begin at 40 C in the 379 Mpa data set. This is a 60 C difference from the 70 Mpa experiments. The mode shapes in the torque-twist-temperature curves are similar but the magnitude of the axial strain continues to increase as well. The maximum axial strain in the high temperature, 379 Mpa cycles is over 4% and it reaches as high as 8% at low temperatures. It is also noteworthy that the minimum strain during the high-temperature cycles, increases with axial stress. From the appearance of the torque-twist curves at high temperatures, the material is passing through the austenite phase as the twist passes through the torque-twist origin. The austenite phase, however, is associated with linear elastic behavior and therefore the axial strain at zero twist should reflect the linear axial response of the wire. This is roughly true for the lower levels of axial stress (Figure 77) but the zero-twist axial strain is between 0. 5 and 2% for the higher axial stress data sets. This implies that some additional, residual, strain has been imposed on the austenite configuration.

Summary of Critical Transformation Stresses

The data published from the *RR* tension-torsion experiments was in the form of torque and twist. To develop the theoretical model, which operates in the stress-strain-temperature domain, it was necessary to estimate an

effective shear stress to associate with each torque plateau. Due to the nonlinear stress-strain behavior in SMA materials, the standard torque-shear stress relations were not applicable. The conversion from torque into an *effective* shear stress is discused further with Figure 98. Figure 79 shows five sets of forward transformation stresses, one for each of the tension-torsion data sets (excluding the 379 Mpa set).

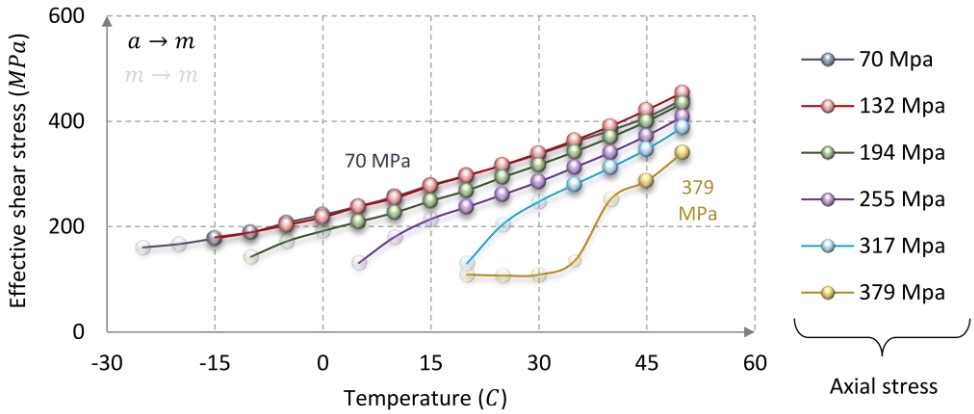


Figure 79: Forward transformation stresses from the tension-torsion experiments

Since focus of this plot is the critical stress for $a \rightarrow m$ transformation, the stress values that correspond to low-temperature ($m \leftrightarrow m$) transformations have been made semi-transparent. Figure 79 shows that the critical stresses are approximately linear with temperature, which is consistent with typical tensile tests on SMA materials. The figure also shows a trend that is not easily observed in the torque-twist plots – the critical shear stresses decrease with increasing axial stress. The physical explanation for this is easily described in terms of energy barriers and wells. Each axial stress represents a force attempting to pull layers out of the austenite well. As this constant force increases, the amount of torque required to pull the layers out of the energy well is lower. It should also be noted, however, that the two lowest values of axial stress (70 and 132 MPA) have almost overlapping transformation stress curves. This implies that the relationship between critical shear stress and axial stress is limited to axial stresses above some threshold between 132 and 194 MPa.

Figure 80 shows the effective shear stresses associated with the reverse ($a \leftarrow m$) transformations; once again, the transformations that are considered to be $m \leftrightarrow m$ are semi-transparent.

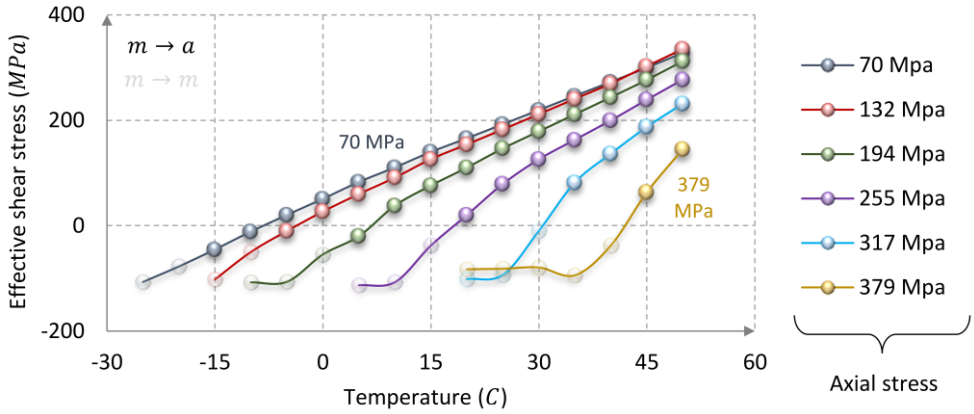


Figure 80: Reverse transformation stresses from the tension-torsion experiments

The set of curves in Figure 80 are roughly linear and the curves get lower as the axial stress increases. The vertical spread between the curves is more drastic than in Figure 79 which means the axial stress has a greater effect on the reverse transformation threshold. The 70 and 132 MPa curves are also similar here, but they begin

to deviate at lower temperatures. It is important to note that the slopes of the curves in Figure 80 are notably different than the slopes of the curves in 79. This indicates temperature-dependent hysteresis behavior, which was added to the 1-D model in Chapter 5. In both the tensile tests and the tension-torison tests, the hysteresis loops get narrower as the temperature increases.

The transformation stresses in the uniaxial tension tests were calculated in a more straightforward manner, as the engineering stresses (which means no consideration of wire diameter changes). These are summarized in Figure 81, where the difference in slopes is also apparent.

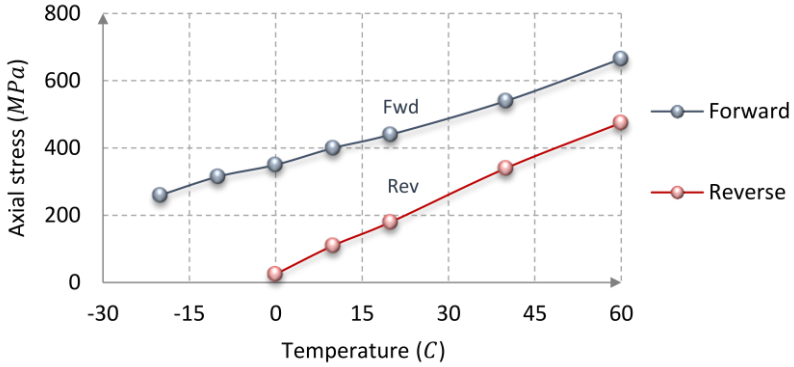
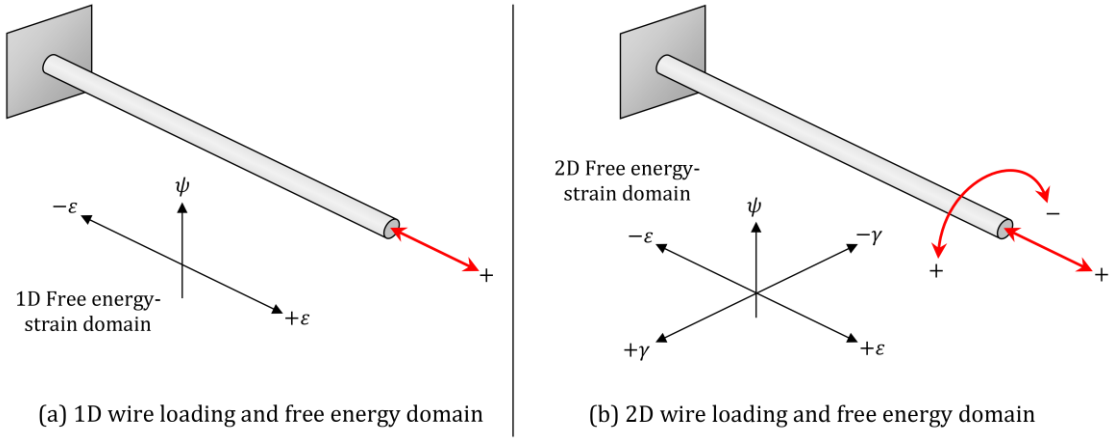


Figure 81: Uniaxial transformation stresses across a range of temperatures

The stress values in Figure 81 are significantly higher than those in Figures 79 and 80. This can mostly be attributed to the difference between normal stiffness (Young's modulus) and shear stiffness (shear modulus). On a free energy landscape, the stiffness equates to the energy well curvature and a lower-curvature well requires less stress to pull layers out. For this reason, it is expected that a lower amount of stress is required to initiate a shear transformation.

6.3 Introduction to the 2-D Model

The 2-D model is aimed at an SMA wire that can experience axial strain or can be twisted (shear strain) and shown in Figure 82b.



(a) 1D wire loading and free energy domain

(b) 2D wire loading and free energy domain

Figure 82: Helmholtz free energy domains for a wire-like specimen (a) 1-D interpretation, (b) 2-D, axial-torsion, interpretation

The main model is a macroscopic constitutive (point-wise) model for the wire material, meaning that it represents the stress-strain-temperature behavior of the macroscopic wire but does not consider field variations along the wire or through the cross section. It is well established that macroscopic field effects (such as transformation localization and front motion) play a significant role in the overall behavior of SMA specimens

and it is likely that they contribute to many of the effects observed in the *RR* experiments. This chapter focuses on the core constitutive model which neglects field effects as well as the influence of neighboring material. Incorporating this model into a full 3-D field solution is left as an opportunity for future work.

In order to account for the linear variation in shear strain through the cross section, we evaluate multiple independent instances of model at multiple "*points*" through the radius and then integrate their individual responses to get the global wire response. This is shown in Figure 112 and discussed further in the *Macroscopic Model* section. The use of multiple instances of the macroscopic model and combining their results at a higher level introduces another scale to the system. We use the term "*Global*" to distinguish the collective response of the integration points from the response of a single *macroscopic* point. The following list summarizes the four different scales that motivate the model and play a role in its application.

Global	→	This scale represents the entire wire and is assumed from the collective response of each integration point. Each integration point is an independent instance of the macroscopic model experiencing a different amount of shear strain and stress.
Macroscopic	→	This is the scale of the main material model which represents the stress-strain-temperature behavior of the SMA material. The scale is macroscopic compared to the underlying SMA effects, but this is the point-wise, continuum material law for the global wire body.
Mesoscopic	→	This represents the scale of the austenite and martensite phase fields (arrangements of layers) and is much smaller than the macroscopic scale. The multiwell free energy that describes the hysteretic and temperature-dependent SMA behavior is formulated at this scale.
Microscopic	→	This is the scale of the atomic lattice and martensite correspondence variants. It takes many alternating layers of correspondence variants to form a single habit plane variant, which is the fundamental material element at the mesoscopic scale. This scale is not directly addressed by the model but motivates the physics and is included for completeness.

Now we focus on the mesoscopic scale to identify the martensitic variants and propose how they can collectively produce the macroscopic effects observed in the *RR* experiments.

6.4 Mesoscopic phase field motivation

We postulate that all of the macroscopic effects observed in the *RR* tension-torsion (as well as the tension) experiments can be explained and modeled using a 2-D, multiwell HFE and four variants of martensite. We begin with the concept of a lattice invariant shear, which is a simple shear on a habit plane that is oriented at some angle as sketched in Figure 83. From the perspective of the X-Y system, the lattice invariant shear results in a combination of axial (X) and shear (XY) strain, where the magnitude of each depends on the habit plane angle.

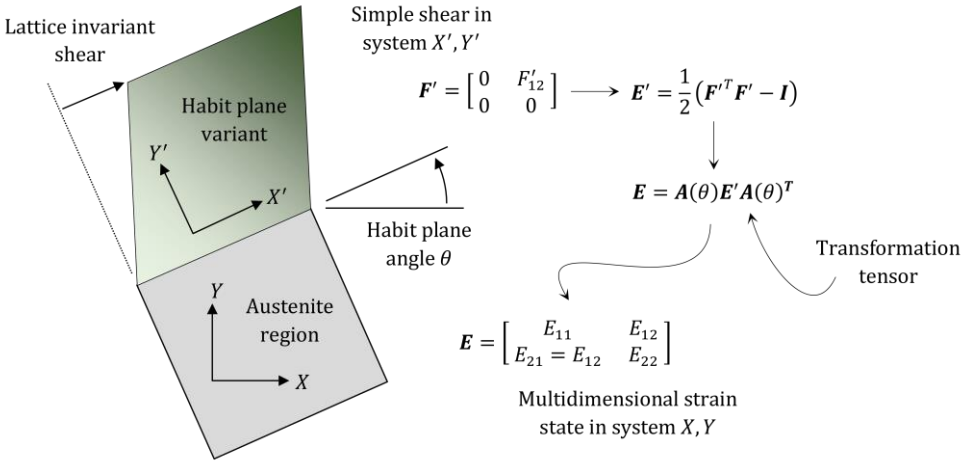


Figure 83: Lattice invariant (simple) shear strain in the habit plane system vs. multidimensional shear strain in the specimen system

In actual SMA materials, the habit plane variants form simultaneously in clusters that have numerous orientations and configurations. These clusters, often called plate groups, usually form in symmetric regions of layered patterns as shown in Figure 84.

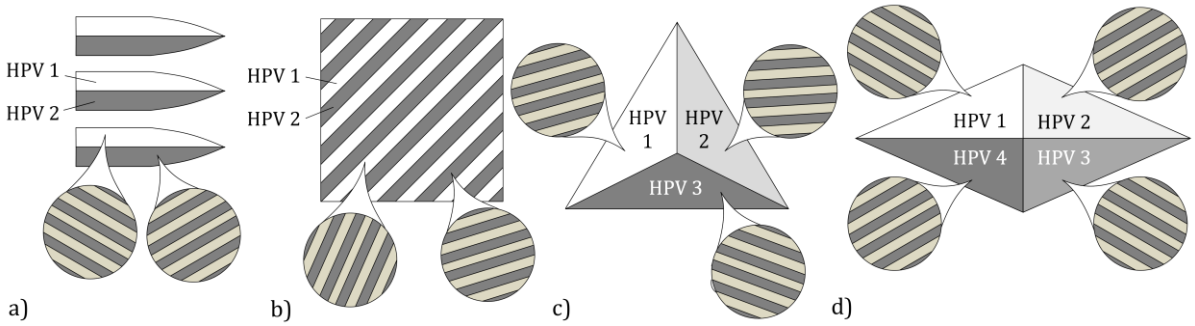


Figure 84: Sketches of observed self-accommodating patches of habit plane variants (a) spear-like plates from Ziolkowski (2015), (b) parallel plates from Ziolkowski (2015), (c) triangle pattern found in Ziolkowski (2015), Otsuka and Ren (2005), and Gao et al. (2000). (d) Diamond pattern from Patoor et al. (2006)

We do not need to resolve the specific shape, number, or orientation of plate groups. The wells in the mesoscopic free energy only need to represent the average states of strain associated with a dominant set of variants. Here, with a set of idealized sketches, we show how the two pairs of habit plane variants assumed by the model can result in any combination of +/- axial deformation and +/- twist in a wire-like body.

Continuing with the idea of habit plane groups forming at various orientations, Figure 85 shows the four variants proposed in the current model arranged in two hypothetical plate groups.

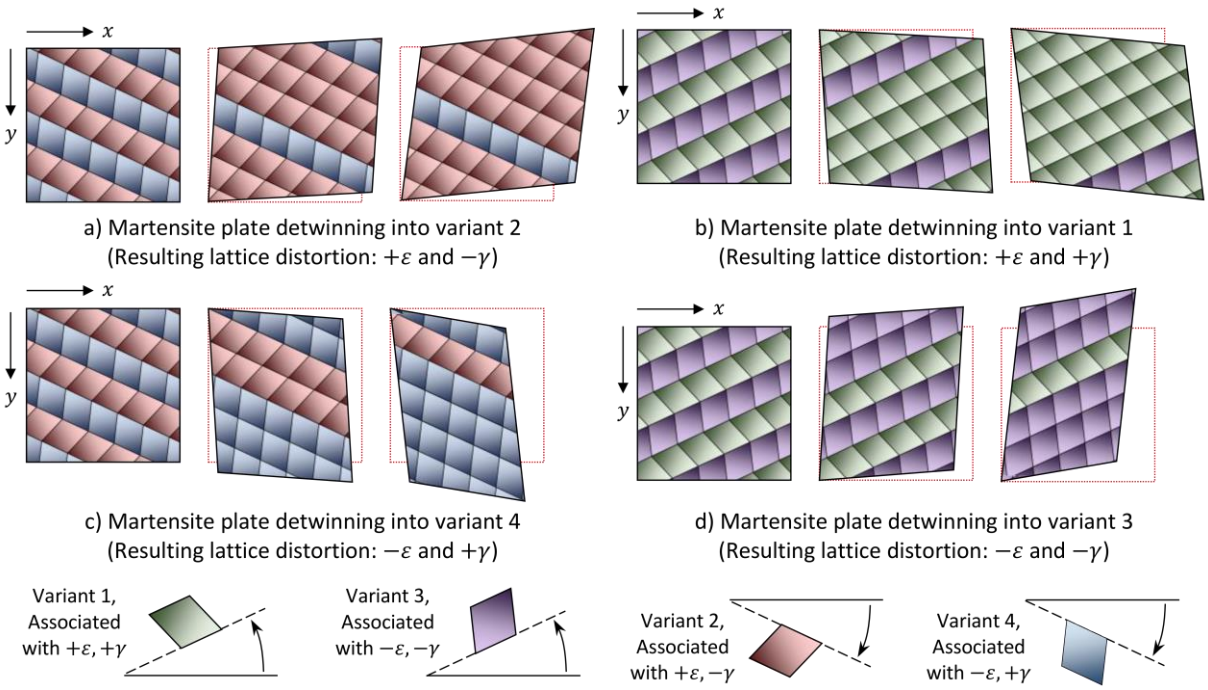


Figure 85: Lattice deformation for each plate group and variant distribution

Each strip of sheared rectangles is referred to as a *layer* can be though of as a representative volume (or material) element (RVE) at the mesoscopic scale, though no individual layer is ever resolved. A layer can assume the parent austenite phase or one of the four martensite variants. Each martensite variant in Figure 85 is signified by its habit plane angle and whether it is shearing to the left or right; they are also colored to help with the distinciton. 85a shows a plate group with a positive habit plane angle and three stages of defromation. At the far

left, the mesoscopic patch is in a self-accommodated configuration where the alternating layers are in roughly equal proportions. The average strain for the entire patch is zero; this effect is also called twinning. The in the center and far right states, the population of layers is becoming increasingly biased in favor of variant 2. As a result, the overall shape of the patch elongates and shears upward. With respect to the corrdinate system shown, this represents a positive axial strain ε and a negative shear γ. 85b shows a similar plate group except the habit plane is now negative. The far left configuration is fully twinned and signifies zero total strain. The two images to the right illustrate the phase field unfolding such that variant 3 diminishes and variant 1 dominates. The resulting deformation here is a positive axial strain and a positive shear. In 85c we return the plate group from 85a, beginning in the same twinned state. Now, the phase fraction of variant 4 grows at the expense of variant 2. The the entire patch now undergoes a negative axial (compressive) strain and a positive shear. Similarly, 85d begins with same plate group as 85b and shows the mesoscopic patch when variant 3 becomes prevalent. Both the axial strain is and shear are now negative. The variant numbers 1-4 are used throughout this chapter and as indices in the model formulation. The details about each variant are summarized here in Table 6.

Variant ID	Habit Plane Angle	Habit Plane Shear Direction	Resulting Normal Strain ε	Resulting Shear Strain γ	Graphical Symbol
1	$-\theta$	Left	$+\varepsilon_R$	$+\gamma_R$	
2	$+\theta$	Left	$+\varepsilon_R$	$-\gamma_R$	
3	$-\theta$	Right	$-\varepsilon_R$	$-\gamma_R$	
4	$+\theta$	Right	$-\varepsilon_R$	$+\gamma_R$	

Table 6: Summary of martensitic variants

Figure 85 shows each variant in free-standing, 2-D patch of mesoscopic material. We extend this to a 3-D wire by breaking the wire into radial slices (tubes) where each tube is dominated by one axial component of normal strain and shear in the axial-tangential plane. Figure 86 shows the mesoscopic material patches from Figure 85 mapped onto a radial slice of a wire-like body.

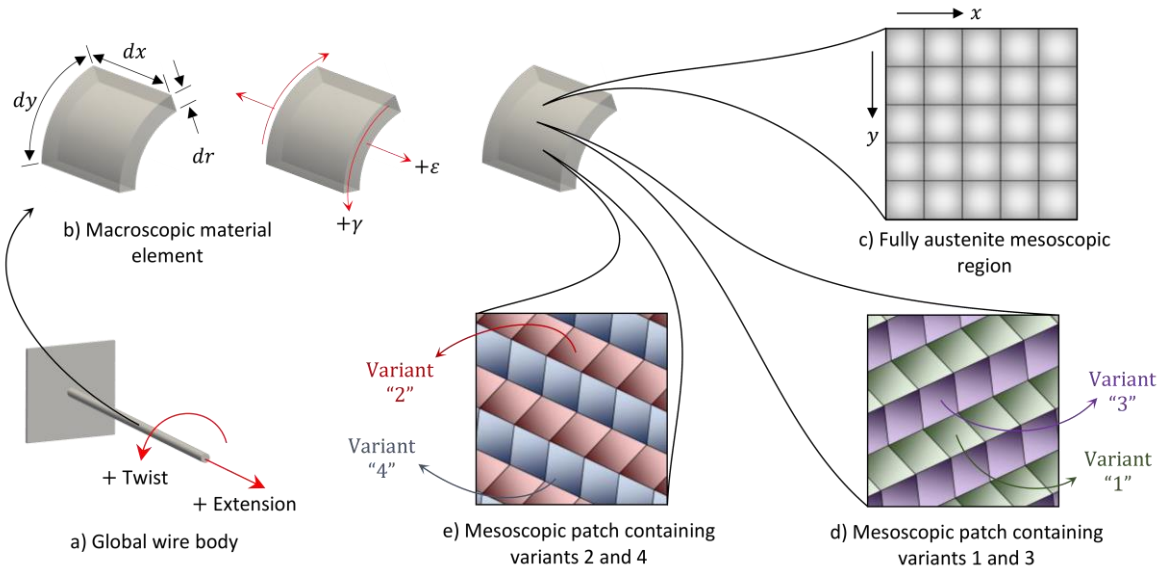


Figure 86: Proposed combinations of habit plane variants occurring in the tension-torsion specimens

The global, wire body that can undergo twist or axial deformation is shown in 86a. A macroscopic material element is illustrated in b where the state of deformation can be described by the strains ε and γ. Any combination of mesoscopic variants can exist with a single macroscopic element and 86c, d, and e show a few different potential mesoscopic arrangements. The macroscopic model samples the mesoscopic free energy to assess the most probable distribution of variants. Any number of arrangements like those shown in Figures 85 and 86 could produce a given set of phase fractions, so the specific mesoscopic arrangement (phase field) in the does not need to be resolved. It is important, however, that each variant is associated with a specific deformation and that the two plate groups yield the same components of strain but with opposite signs. This must be the case for the set of four variants to produce all of the transformation modes observed in the *RR* experiments (uniaxial tension and combined tension-torsion). The symmetry of the two groups also allows pure shear behavior, though this mode was not present in the RR data set. Figure 87 shows the wire in pure tension and how the two plate groups unfold in such a way that uniaxial tension is observed at the macroscopic scale.

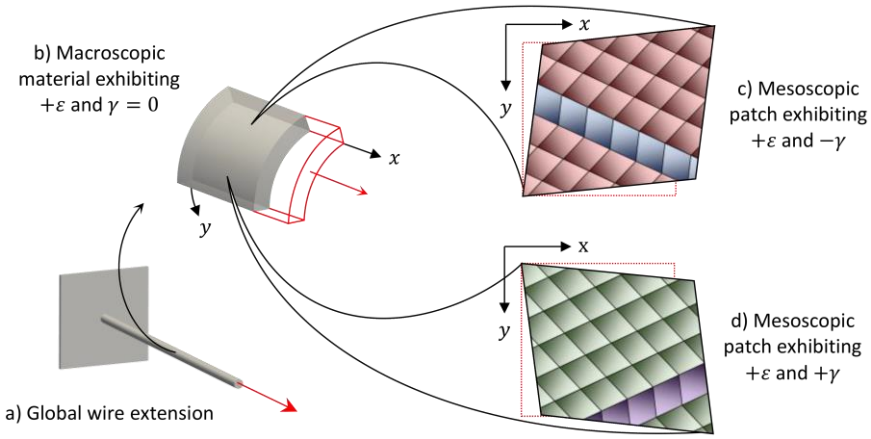


Figure 87: Mesoscopic variants 1 and 2 resulting in axial-only deformation at the macroscopic scale

The presence of both plate groups allows the shear components of the deformation to cancel resulting in uniaxial tension at the macroscopic (and therefore global wire) scale. Figure 88 shows an alternate configuration where the wire is being twisted with no axial deformation.

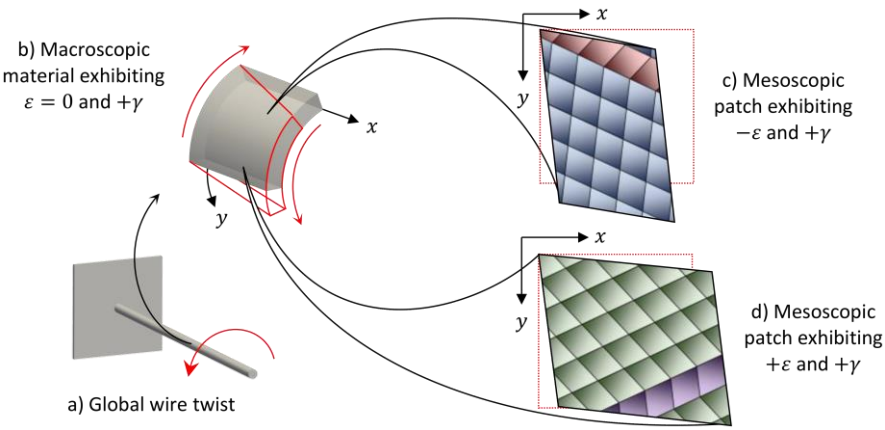


Figure 88: Mesoscopic variants 1 and 4 resulting in shear-only deformation at the macroscopic scale

Once again, the plate groups break their internal symmetry but unfold such that the axial strains negate one another. Figures 87 and 88 represent the two extremes where the plate groups have prefectly opposing responses in one strain axis. Under multiaxial loading, the symmetry is broken the the resulting macroscopic behavior is a

combination of normal and shear strain. Next, we introduce a mesoscopic HFE landscape that reflects the four variants from Figures 85-88.

6.5 Mesoscopic Free Energy

From a thermodynamics perspective, the strain state associated with each variant represents a stable well on a multiwell free energy. We assume that a large population of mesoscopic layers is governed by a single free energy landscape and that different numbers of layers can occupy each phase. The 2-D HFE is shown in Figure 89 at a temperature where all four martensite phases and the parent austenite phase are stable.

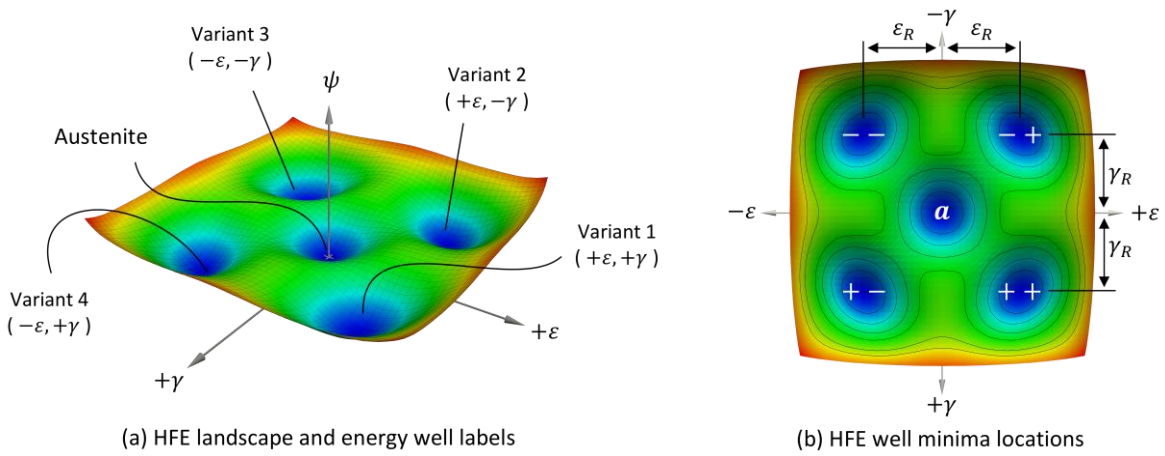


Figure 89: Helmholtz free energy landscape in a normal strain, shear strain domain

Each martensite variant occupies a quadrant in the normal strain – shear strain domain while the parent austenite well rests at the zero-strain origin. Each well is offset from the origin in both strain axes; these offsets represent the stress-free (residual) strains as in the 1-D model. These strains are identified in the same way as the 1-D model, by projecting the superelastic martensite branch to the zero-strain axis (or from actuation experiments). The material parameters are discussed in more detail later in this chapter.

The macroscopic strain is determined from the weighted average of the mesoscopic layers in each phase. As the system stress and temperature change, layers jump from one well to the other which give rise to stress-driven and temperature-driven transformations. Figure 90 shows the eight possible transformation paths.

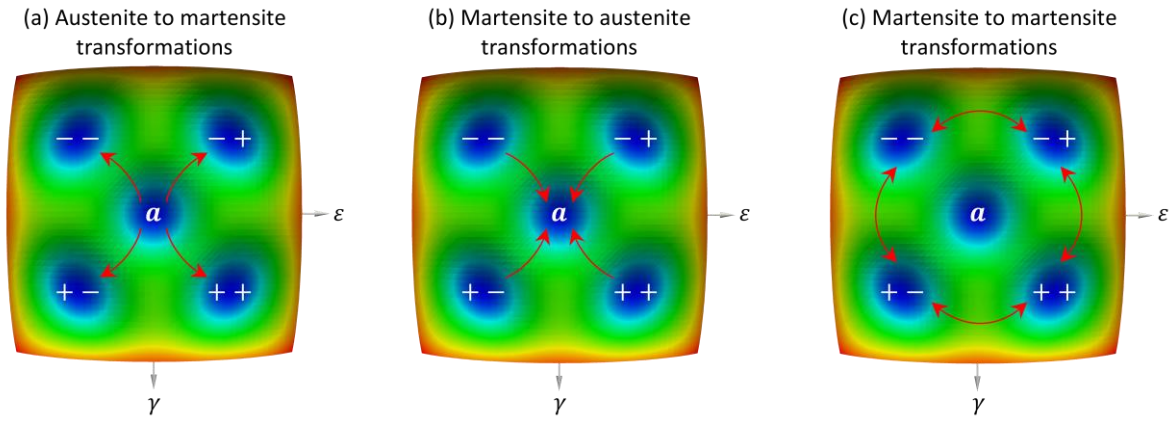
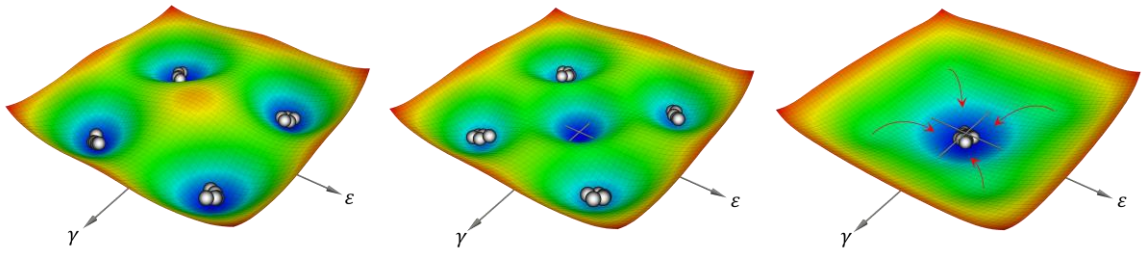


Figure 90: Possible transformation paths (a) Forward → austenite to martensite, (b) Reverse → martensite to austenite, (c) Martensite to martensite

The four paths in 90a are austenite-to-martensite transformation and we call these "*forward*" transformations for brevity. The *reverse* paths are shown in 90b and the four possible martensite-to-martensite paths are shown in c. The martensite-to-martensite transformations represent a behavior often called martensite reorientation. It

should be noted that there is no path across the austenite well. Layers sometimes jump into the austenite well and quickly jump out to the opposite side, but the jump across the austenite well is not necessary for the model to represent the desire behavior.

Figure 91 shows the free energy landscape at three different temperatures and with a series of balls *resting* in the wells. The balls are a graphical representation of how the layers are distributed between the phases. Within each well, the balls are assigned slight random offsets from the well minimum; this allows us to see multiple balls and maintain the idea of a population of layers. In an actual SMA solid, the thermal activation would be very low and the balls in a given well would appear indistinguishably at the minimum.



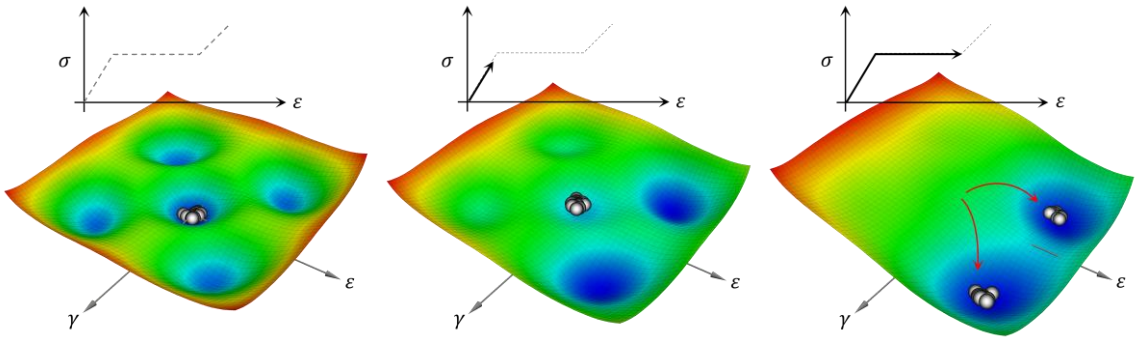
(a) Temperature T_1 (b) Temperature $T_2 > T_1$ (c) Temperature $T_2 > T_3$

(a) Temperature T_1 (b) Temperature $T_2 > T_1$ (c) Temperature $T_3 > T_2$

Figure 91: Helmholtz free energy landscape showing a temperature-induced transformation

Figure 91a shows a temperature below T_a (the austenite start temperature) where the austenite phase is completely unstable. The cooling took place in a stress-free environment and the layers distributed themselves equally between the four martensite wells. In this situation, no macroscopic strain would appear as is commonly observed in stress-free cooling of SMA specimens. In 91b, the temperature has increased to a value above T_a but still below T_m (the martensite start temperature). The austenite well has reformed (become convex) but not layers jump into it because of the energy barriers separating it from the martensite wells.

Figure 92 shows a stress-induced transformation due to uniaxial tensile loading; this free energy configuration corresponds to the case in Figure 87.



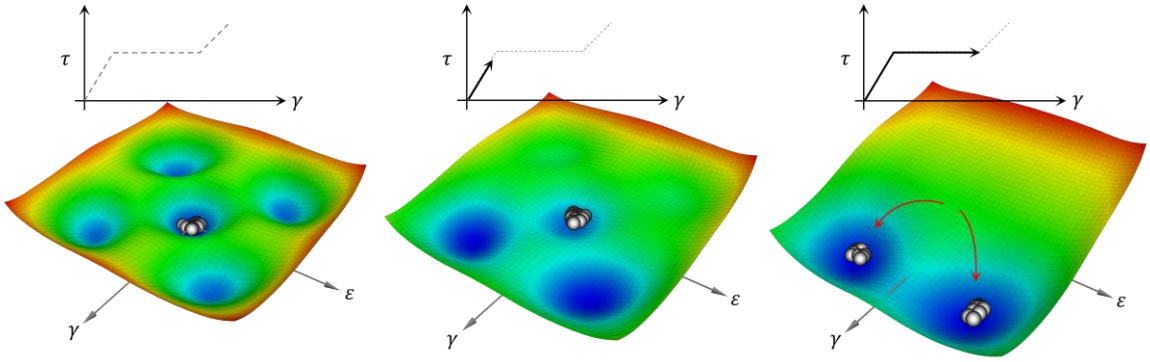
a) Zero stress b) Tensile stress c) Tensile transformation stress

Figure 92: Macroscopic tensile transformation and underlying mesoscopic Gibbs free energy behavior

The free energy shown here is the GFE because it is distorted by the external stress. The system begins (92a) at an intermediate temperature where the material is fully austenitic. In 92b, a tensile stress has been applied and the two wells on the tensile side of the landscape become lower and their convex regions grow – this signifies that tese phases are becoming more favorable. The austenite well, on the other hand, is becoming less favorable as it

is shallower and has a smaller strain footprint. The layers are still trapped there however because a small energy barrier still remains in the path to the two, much lower, martensite wells. The stress-strain curve floating above the free energy shows that the macroscopic response is still on the linear austenite branch but is very close to the transformation plateau. In 92c, the stress (and therefore the GFE) has just reached the critical state where the barriers are eliminated and the layeres are free to jump into the more favorable phases. The results is a large increase in tensile strain with no change in stress which is shown on the stress-strian curve. Since there is no shear stress, the two tensile wells are equally desirable and the layers fill them equally. The mesoscopic shear strain symmetry is maintained and therefore no macroscopic shear strain is observed.

Figure 93 shows the GFE under pure shear loading (which corresponds to Figure 88).



(a) Zero stress (b) Pure shear stress (c) Pure shear transformation stress

Figure 93: Macroscopic shear transformation and underlying mesoscopic Gibbs free energy behavior

At zero stress (93a) the system is identical to the uniaxial tension case (92a); all of the layers are in the central austenite well and average strain is zero. In 93b, a positive shear stress has been appplied and the two energy wells on the $+\gamma$ side of the landscape have become wider and deeper. The stress-strain plot shows that the macroscopic response is still on the linear branch, but that it is close to the transformation threshold. This is indicated on the GFE by the fact that the austenite well has become smaller and the barrier keeping the layers in it is shrinking. The GFE has reached the critical value in 93c, due to shear loading, and the layers jump equally into the two + shear wells. The average axial strain remains zero because of the symmetric transformation but the average shear strain increases suddenly with no change in stress.

Next, in Figure 94, we show an example of combined loading that is similar to the *RR* tension-torsion experiments.

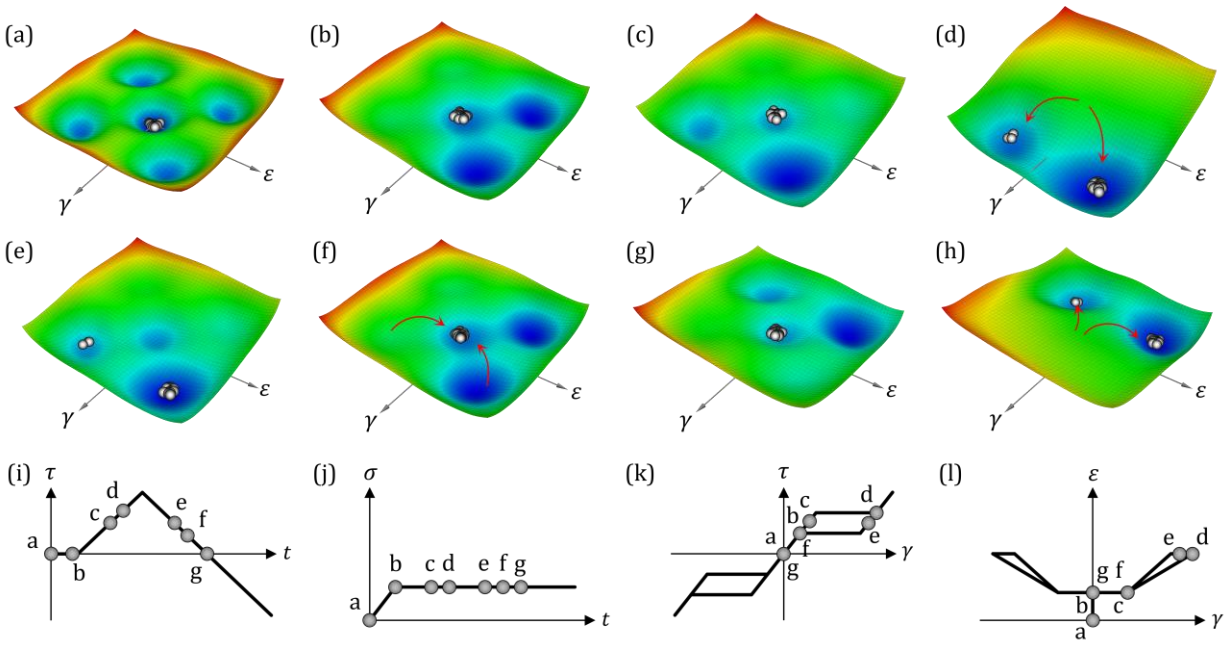


Figure 94: Macroscopic multiaxial loading and underlying mesoscopic Gibbs free energy behavior

The system begins in a fully austenite state (94a) and then a tensile stress is applied in b. The tensile stress is not sufficient to cause transformation but tilts the GFE in the $+\varepsilon$ direction. In 94c, the tensile stress is maintained and an additional shear stress is applied. The shear stress is also not large enough to cause any layers to jump wells but this stage shows an important coupling effect. The two wells on the $+\gamma$ side of the landscape are no longer symmetric about the $+\varepsilon$ axis. The sustained tensile stress causes the $+\varepsilon, +\gamma$ well to be the more favorable of the two $+\gamma$ wells. Now, in 94d, the shear loading continues to the point of transformation and we see that the symmetry is broken and more wells jump into the $+\varepsilon, +\gamma$ well than the $-\varepsilon, +\gamma$ well. The consequence of this is illustrated in 94l, which shows a macroscopic plot of the axial strain vs. shear strain. We see that the path from state c to state d signifies a large jump in both shear and axial strian, eve though the axial stress is constant. The presence of the sustained tensile stress breaks the symmetry between the two $+\gamma$ phases and introduces axial-shear coupling.

In 94e, the shear stress unloading as begun; the two wells are not quite as deep, but no additional transformations initiate. Since no transformations occur, the macroscopic stress-strain response between d and e is linear along the $\tau - \gamma$ unloading branch (k). The $\varepsilon - \gamma$ plot (l) shows a small reduction in shear strian with no change in tensile strain. 94f signifies the reverse transformation where all of the layers return to the parent austenite well. There is a large decrease in shear strain (seen in k and l) and the axial strain associated with the transformation also disappears (l). The axial strain in f is the same as it was in states b and c; the linear-elastic strain associated with the external axial stress. In 94g, the shear unloading continues past zero stress to become a reverse loading. Both of the $-\gamma$ wells have become more convex but the constant axial stress makes the $+\varepsilon, -\gamma$ well the wider and deeper of the two. The critical free energy has been reached in h and the layers transform in the $-\gamma$ direction. Because of the bias toward the tensile side, more of the layers choose the $+\varepsilon, -\gamma$ well and the result is, once again, a transformation that is associated with both axial and shear transformations. The unloading proceeds in a similar manner and the paths are shown in k and l. The resulting $\tau - \gamma$ plot has rotated symmety about the τ axis and the $\varepsilon - \gamma$ plot is mirror symmetric about the ε axis. An important obervation here is that the

characteristic shape of the $\varepsilon - \gamma$ response is similar to the experimentally observed loops in the high-temperature tension-torsion tests shown in Figures 77-78.

6.5.1 Helmholtz Free Energy

In Chapter 5, we presented a version of the 1-D kinetics equations where only the convex free energy wells needed to be defined (Eqn. 55). In this section, we take a similar approach to present a 2-D model where only the convex energy wells are required. If we neglect the barriers connecting the wells in Figure 94a, we can describe the free energy with five individual, parabolic surfaces like those sketched in Figure 95.

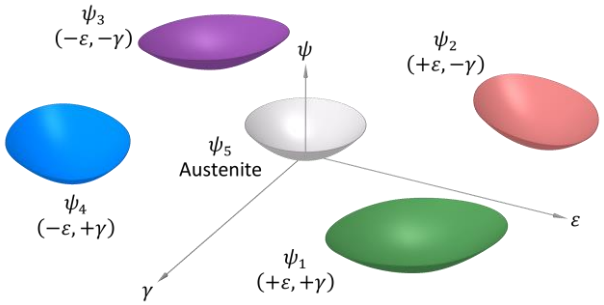


Figure 95: Helmholtz free energy landscape with only the convex wells resolved, At an intermediate temperature where all five phases are stable

It should be noted that the dimensions, locations, and curvatures in Figure 95 have been exaggerated to discuss important features of the HFE functions. Each of the HFE wells is written as a separate function ψ_i but they all have same mathematical form, given by Equation 177.

$$\psi_i(\varepsilon_i, \gamma_i, T) = \frac{E_i}{2}\left(\varepsilon_i - \varepsilon_i^R\right)^2 + \frac{Y_i}{2}\left(\gamma_i - \gamma_i^R\right)^2 + M_i\left(\varepsilon_i - \varepsilon_i^R\right)\left(\gamma_i - \gamma_i^R\right) + \beta_i(T) \tag{177}$$

where

$i \rightarrow$ *Variant index* $(i \in 1-5)$
$\varepsilon_i \rightarrow$ *Average normal strain of phase i*
$Y_i \rightarrow$ *Shear modulus of phase i*
$\gamma_i^R \rightarrow$ *Shear strain offset for well* ψ_i
$\beta_i \rightarrow$ *Chemical free energy of phase i*
$E_i \rightarrow$ *Elastic modulus of phase i*
$\varepsilon_i^R \rightarrow$ *Normal strain offset for well* ψ_i
$\gamma_i \rightarrow$ *Average shear strain of phase i*
$M_i \rightarrow$ *Normal − shear coupling parameter*

For each variant i, ψ expresses the HFE for all strains where that phase is stable. Each HFE function Figure 95 has an oblong shape, due to the difference in stiffness between the tensile and shear deformation. It should also be noted that the oval shapes of the four martensite wells ($\psi_1 - \psi_4$) are rotated, each about an axis parallel to the ψ axis and extending from the well minimum. This is the effect of the coupling parameter M_i. Like the 1-D model from Chapter 4, the austenite well always rests at zero chemical free energy as it does in Figure 95. The marteniste wells, however, translate up or down with temperature according to their chemical free energy $\beta_i(T)$. At lower temperatures, the martensite wells become lower and more stable compred to the central austenite well and the inverse is true at high temperatures.

Figure 95 also illusrates several material property assumptions that are specific to the tension-shear system under examination. The first assumption is the martensite wells are mirror symmetric with respect to each strain axis. This means that the normal and shear stain offsets have equal magnitudes.

$\varepsilon_1^R = +\varepsilon_R \quad \varepsilon_2^R = +\varepsilon_R$
$\varepsilon_2^R = -\varepsilon_R \quad \varepsilon_4^R = -\varepsilon_R$
where $\varepsilon_R \rightarrow$ *Residual normal strain material parameter, determined directly from experimental data*

$$\gamma_1^R = +\gamma_R \quad \gamma_2^R = -\gamma_R \quad \gamma_3^R = -\gamma_R \quad \gamma_4^R = +\gamma_R \qquad where \qquad \gamma_R \rightarrow \begin{matrix} Residual\ shear\ strain\ material\ parameter, \\ determined\ directly\ from\ experimental\ data \end{matrix}$$

Each martensite well is in Figure 95 is rotated about the ψ axis and this represnts the effect of the coupling parameter M. In the figure, we have assumed each well is rotated by the same amount, which means M has the same magniude for each variant.

$$M_1 = M \quad M_2 = -M \quad M_3 = M \quad M_v = -M \qquad where \qquad M \rightarrow \begin{matrix} Normal - shear\ coupling\ material\ parameter, \\ determined\ using\ experimental\ data \end{matrix}$$

This parameter governs axial shear coupling within each energy well, which is the elastic behavior region. In theory, such a parameter could account for the type of behavior seen in bottom of Figure 77, where the torsional cycles cause axial oscillations. Linear axial-shear coupling could also be the cause for the slopes in the axial strain-twist plots in the center row of the same figure (77). These observations are the motivation behind initially including this term. In model operation, however, the system response is only sightly affected by M. Of course, a very large value can cause very different outcomes but these differences to not change the results in a way that makes it more accurate. We find that the axial-shear coupling in the model is dominated the presence of axial stress and shear stress in the Gibbs free energy and that linear coupling in the HFE takes a back seat. In this light, we drop the linear coupling term for the remainder of the main dissertation. Removing M significantly simplifies several model equations. To preserve a record of the full theory, we have included the model equations that maintain the coupling term in Appendix 10.5.

Now we return to Equation 177 final term, which is the chemical free energy. There are four relative chemical free energy parameters, one for each martensite variant, but we assume here that they are all equal. It should be noted that this assumption is an idealization and is only reasonable because the system we address is dominated by two very similar martensite variants.

$$\beta_1(T) = \beta_2(T) = \beta_3(T) = \beta_4(T) = \beta(T) \qquad where \qquad \beta(T) \rightarrow Martensite\ chemical\ free\ energy$$

The CFE $\beta(T)$ is determined from experimental data at two different temperatures but is derived in the context of the GFE, which will be defined in the next section. Now, returning to the HFE equation (177), ψ_i describes the free energy at any location (combination of ε and γ) within the convex energy wells and producing surfaces like those shown in Figures 95. For model operation, however, we assume that the layers in each phase exhbit the same low thermal activation behavior discussed in the *Kinetics* section. This means that the layers all cluster at the well minimum and that the average strains can be determined from the state of stress and the HFE.

$$\sigma(\varepsilon_i) = \frac{\partial \psi_i}{\partial \varepsilon_i} \quad \rightarrow \quad \sigma(\varepsilon_i) = E_i\left(\varepsilon_i - \varepsilon_i^R\right) \quad \rightarrow \quad \varepsilon_i(\sigma) = \frac{\sigma}{E_i} + \varepsilon_i^R \tag{178}$$

$$\tau(\gamma_i) = \frac{\partial \psi_i}{\partial \gamma_i} \quad \rightarrow \quad \tau(\gamma_i) = Y_i\left(\gamma_i - \gamma_i^R\right) \quad \rightarrow \quad \gamma_i(\tau) = \frac{\tau}{Y_i} + \gamma_i^R \tag{179}$$

In Equations 178 and 179, σ and τ are the nominal mesoscopic stresses (normal and shear, respectively), which are assumed to be equal to the macroscopic stresses. This why no distinction is made between the mesoscopic and macroscopic stress parameters. These constitutive equations neglect the spatial variations in stress as well as strain that are known to exist at the mesoscopic scale in actual SMA materials. The justification behind this assumption is that the free energy is used to describe the average (most probable) behavior of the mesoscopic system without actually resolving its configuration. In cases where the mesoscopic texture is too significant to be

neglected, the effects can be introduced via the kinetics equations thus maintaining the simple form of the stress-strain relationships - This was introduced in the context of the 1-D model in section.

6.5.2 Gibbs Free Energy

The HFE (Equation 177) dictates the stress-strain relationships within each phase, but it is the GFE that governs the kinetics behavior. The GFE is a Legendre transformation of the HFE and is given by Equation 180.

$$g_i(\varepsilon,\gamma,T;\sigma,\tau) = \psi_i(\varepsilon_i,\gamma_i,T) - \sigma\varepsilon_i - \tau\gamma_i \tag{180}$$

Inserting Equation 177 for ψ_i (with no coupling term) gives the following expression (181) of Gibbs free energy as a function of strain, stress, and temperature.

$$g_i(\varepsilon,\gamma,T;\sigma,\tau) = \frac{E_i}{2}\left(\varepsilon_i - \varepsilon_i^R\right)^2 + \frac{Y_i}{2}\left(\gamma_i - \gamma_i^R\right)^2 + \beta_i(T) - \sigma\varepsilon_i - \tau\gamma_i \tag{181}$$

As was done in the 1-D model, we use a semicolon between ε,γ,T and σ,τ in the function to denote that the strains and temperature set the HFE and that the stresses transform it into the GFE. Equation 181 can be evaluated at any combination of strains within the stable domain for phase i and thus produces a set of surfaces similar to the convex regions in Figures 89-94. Since we assume that the layers (on average) reside at the local energy minimum, we can use Equation 181 to define a minimum GFE for each phase by substituting Equations 178 into 179 for ε_i and γ_i respectively. The minimum GFE for each phase, denoted with an upper case G_i, is a function of stress and temperature, given by Equation 182.

$$G_i(\sigma,\tau,T) = \frac{E_i}{2}\left(\varepsilon_i(\sigma) - \varepsilon_i^R\right)^2 + \frac{Y_i}{2}\left(\gamma_i(\tau) - \gamma_i^R\right)^2 + +\beta_i(T) - \sigma\varepsilon_i(\sigma) - \tau\gamma_i(\tau) \tag{182}$$

If substitute Equations 178 and 179, the minimum GFE expression can be rearranged and written as shown here.

$$G_i(\sigma,\tau,T) = -\frac{\sigma^2}{2E_i} - \frac{\tau^2}{2Y_i} + \beta(T) - \sigma\varepsilon_i^R - \tau\gamma_i^R \tag{183}$$

Next, we discuss the experimentally measured material parameters (E_a, Y_a, E_m, Y_m, ε_R, γ_R) along with the chemical free energy $\beta(T)$.

6.5.3 Material Parameters

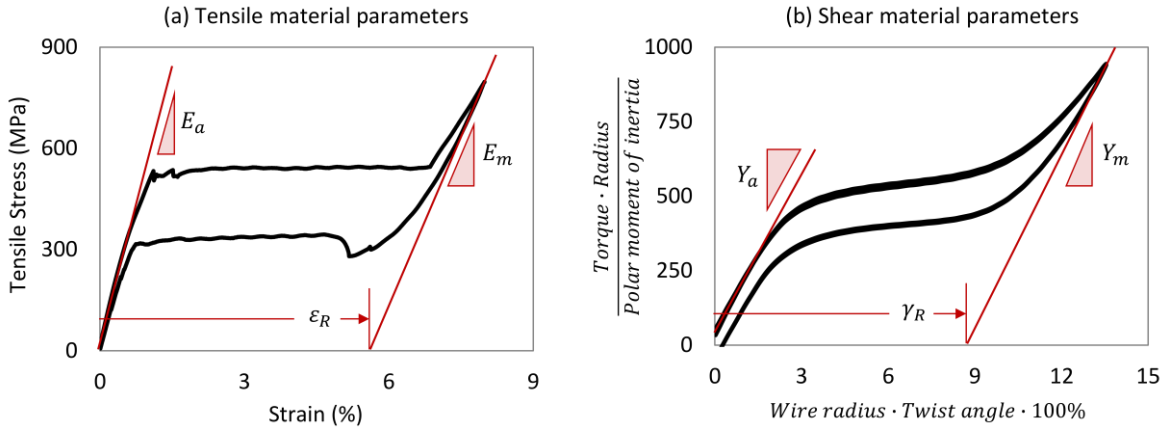
In this section, we discuss the proper sequence and procedure for establishing the material parameters. Because the experimental data has different combinations of tension and shear behavior, a few considerations need to be made. The material parameters are defined in the following order:

1) Define the tensile elastic modulus values and strain offset	→	From uniaxial tension test data and linear stress-strain equations
2) Define the shear modulus values and strain offset	→	From tension-torsion test data and linear torque-twist equations
3) Define the tensile transformation stresses	→	From uniaxial tension test data and linear stress-strain equations
4) Define the shear transformation stresses	→	From tension-torsion test data, linear torque-twist equations, and a nonlinear adjustment factor
5) Define the chemical free energies	→	Calculated from the

		transformation stress parameters
6) Define the latent heats	→	Calculated from the transformation stresses and chemical free energies
7) Define the critical free energies	→	Calculated from the transformation stresses

Steps 1-2: Elastic/Shear Moduli and Strain Offsets

The first set of material parameters measured are the elastic moduli and the strain offsets (residual strains). The tensile properties are measured from one of the tensile test experiments as shown in Figure 96a.



$E_a = 60\ Mpa$	$E_m = 30\ Mpa$	$\varepsilon_R = 5.4\ \%$	$Y_a = 17\ Mpa$	$Y_m = 15\ Mpa$	$\gamma_R = 6.5\ \%$

Figure 96: Material parameter measurements from experimental data

Even though a meaningful reduction in cross section occurs in the transformation zone of SMA tensile specimens, the engineering stress and strain are used. Also, transformations localizations (which are expected to be present because of the appearance of small stress spikes) are ignored in this model and thus the plots and the model represent the average specimen response. The shear modulus parameters (96b) are measured in a similar manner using one of the tension-torsion tests. We used the 70 MPa axial stress case here, but any of the axial stress data sets could be used.

Step 3: Tensile Transformation Stresses

The tensile transformations stresses are obtained the same way as in the 1-D model, by measuring the stress plateaus at a two temperatures.

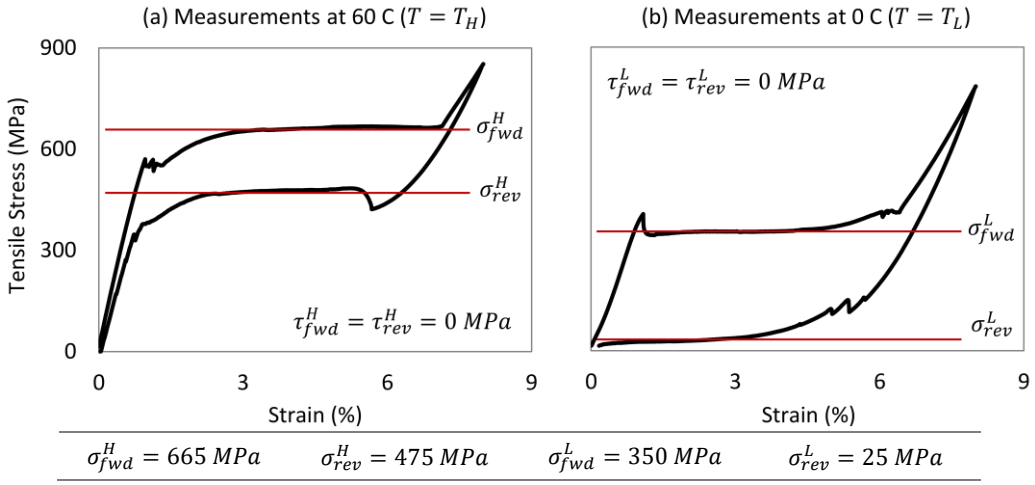


Figure 97: Transformation stress measurements from uniaxial tension experimental data

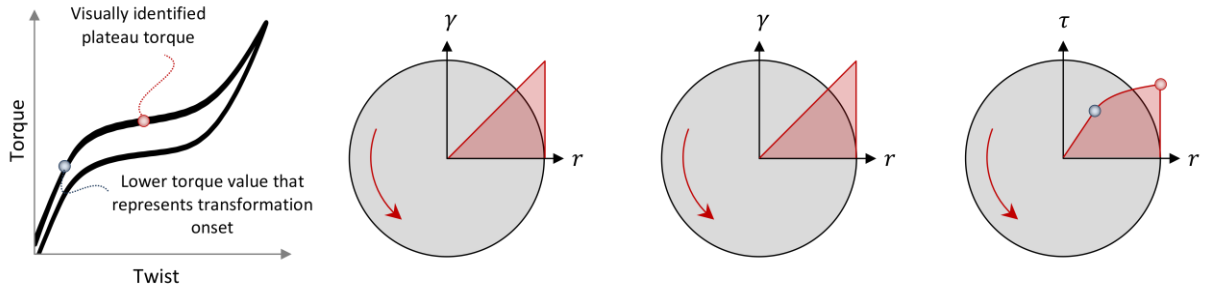
Note here that we ignore the irregularities at the beginnings and ends of the transformation plateaus. These are the results of localization effects, which are significant in these experiment but are not included in the current formulation. The torque was zero during this set of experiments, so the analogous shear parameters are set equal to zero ($\tau^H_{fwd} = \tau^H_{rev} = \tau^L_{fwd} = \tau^L_{rev} = 0$).

Step 4: Shear Transformation Stresses

We take a moment here to discuss an important aspect of the torsional test data, the conversion from torque-twist (which was reported by Sittner et al.) to stress and strain. The traditional linear torque-twist/stress-strain equations for a cylindrical shaft would normally be used for an experiment on a wire, but they are limited to linear materials. The standard linear equation,

$$\tau = \frac{Torque \cdot Radius}{Polar\ moment\ of\ inertia} \quad \rightarrow \quad Only\ correct\ when\ \tau = linear\ f(\gamma),$$

can be used to define the shear modulus values and strain offsets as shown in Figure 96b because the only parameters measured are measured from linear branches. It is important to note that the nonlinear nature of the stress-strain curve must be taken into account to assess the transformation shear stresses.



(a) Plateau torque vs. transformation onset torque

(b) Assumed radial strain profile

(c) Shear stress profile given by linear theory

(d) Stress profile in an SMA wire

Figure 98: Plateau torque vs. torque representing the onset of transformation

The plateau torque, shown in 98a, is easy to identify but the shear stress associated with this torque (98d) are not what is predicted by the standard linear calculation (98d). We overcome this by simulating the experiment and establishing a reduction factor for the shear stress that is obtained with the linear equation. We begin with a single tension-torsion test and measure the forward plateau torque, shown in Figure 98a. Then we convert that torque into shear stress using the standard equation. Next, we simulate that test and use several different

transformation stresses, starting with the one calculated from the lieanr equation and then subsequently lower values. The simulation with the best fit to the data allows us to infer a *stress reduction factor* (F_τ) that can be applied to other curves. An example of this is shown in Figure 99.

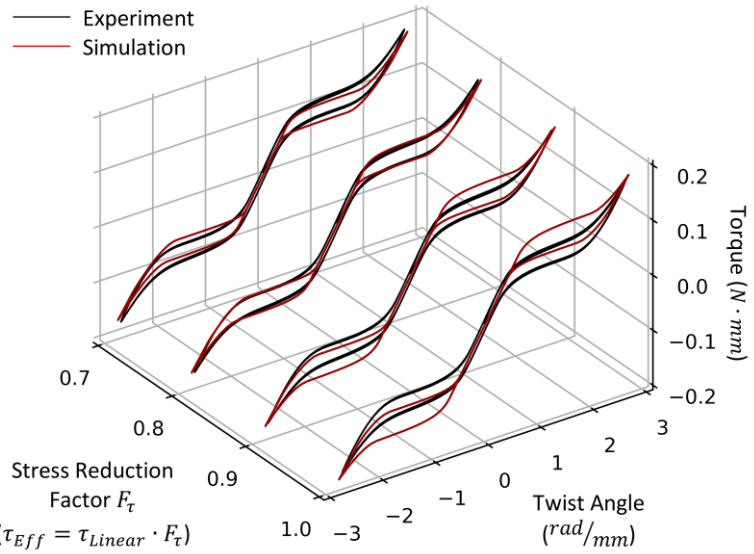


Figure 99: Determination of the stress reduction factor from a set of simulations

τ_{Linear} is the value of shear stress calculated from the mid-plateau torque and τ_{Eff} is the transformation stress used in the model. The plots show that τ_{Linear} must be scaled by 0. 8 to get the proper macroscopic result. This procedure only has to be done once. Now, we can simply measure the plateau torque from any of the tension-torsion curves and estimate the transformation shear stress from the following equation.

$$\tau_{Eff} = F_\tau \cdot \frac{Torque\ at\ the\ center\ of\ the\ plateau \cdot Radius}{Polar\ moment\ of\ inertia} \tag{184}$$

Now, measuring the transformation shear stresses from the torque-twist experiments is straightforward. The plateau torques are measured, the linear equation is used to calculate a shear stress, then that stress is multiplied by F_τ.

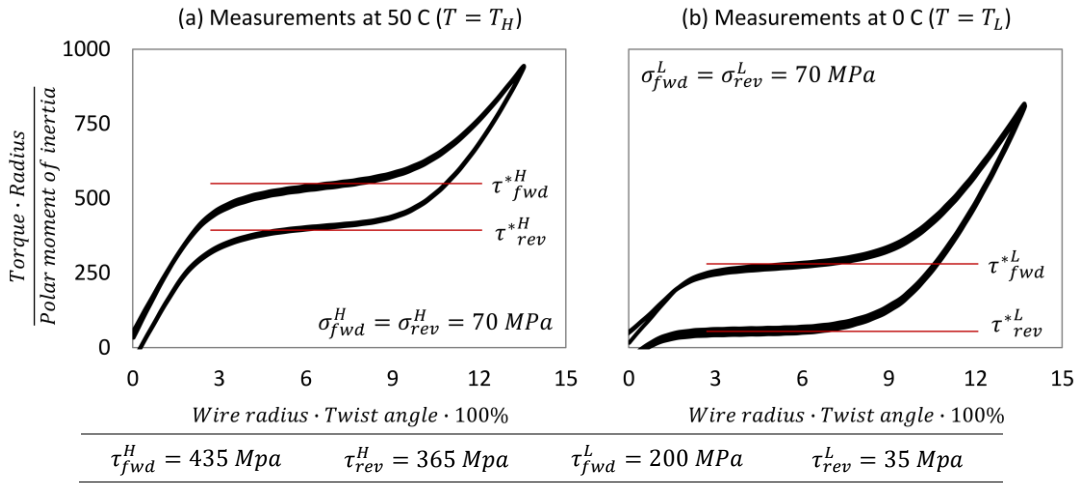


Figure 100: Transformation stress measurement from tension-torsion experimental data

It should be noted that the torque-twist data shown in Figure 100 was from the 70 MPa axial stress data set. The additional axial stress is also included in the material property by setting $\sigma_{fwd}^{H} = \sigma_{fwd}^{L} = \sigma_{rev}^{H} = \sigma_{rev}^{L} = 70\ MPa$.

These assignments conflict with those from Figure 96. This means that we choose either set to populate the model, not both. The normal-shear coupling in the model is generated by the GFE and not from a collection of normal and shear experiments. This means that the model requires only a few material properties but, in the case of the *RR* data, we have to choose one data set from many to use for the material properties. This is discussed further in the simulation section.

Step 5: Chemical Free Energies

In the 1-D model, the chemical free energy was derived from continuity conditions that assumed for the energy wells and the barriers. In the 2-D model, we have taken specific steps to neglect the energy barriers and this means we need a new approach to derive the chemical free energy parameters. Here, in the 2-D model, were estimate the chemical free energy $\beta(T)$ by examining the examining zero-driving force configurations. To begin, we show a set of 1-D GFE wells under zero stress (Figure 101a) and under a tensile stress in 101b.

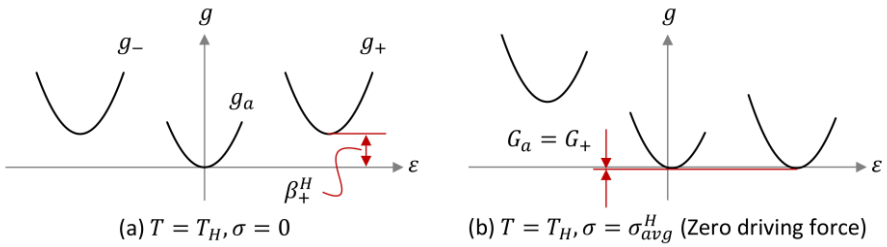


Figure 101: 1-D Gibbs free energy showing (a) martensite chemical free energy and (b) a zero-driving force configuration

The tensile stress Figure 101b is precisely that needed to creaste a zero driving force configuration, i.e. the austenite and tensile martensite wells have the same minimum GFE. To make sure of this, we assume that the zero-driving force configuration occurs at some stress near the average of the experimentally measured forward and reverse transformation stresses. Since we have assumed that the CFE is equal for all four martensite variants, we can focus on one variant and write Equation 185.

$$G_5(\sigma_{avg}, \tau_{avg}, T) \approx G_1(\sigma_{avg}, \tau_{avg}, T) \rightarrow Zero\ driving\ force\ at\ temperature\ T \tag{185}$$

$$where \quad \sigma_{avg} = Average(\sigma_{fwd}, \sigma_{rev}) \qquad \tau_{avg} = Average(\tau_{fwd}, \tau_{rev})$$

$$\sigma_{fwd}, \tau_{fwd} \rightarrow \begin{cases} Combination\ of\ stresses \\ causing\ forward \\ transformation\ at \\ temperature\ T \end{cases} \qquad \sigma_{rev}, \tau_{rev} \rightarrow \begin{cases} Combination\ of\ stresses \\ causing\ reverse \\ transformation\ at \\ temperature\ T \end{cases}$$

Equation 182 is used to define G_5 and G_1 but that expression depends on $\beta(T)$ which is yet to be defined. Recalling that the CFE is a relative quantity and is measured with respect to the austenite phase (i. e $\beta_5 = 0$), we see that β_1, at temperature T, is the only unknown in Equation 185. Rearranging the expression to solve for β at two temperatures is sufficient to derive a linear expression for the CFE vs. temperature.

$$\begin{aligned} G_5(\sigma_{avg}^{L}, \tau_{avg}^{L}, T_L) = G_1(\sigma_{avg}^{L}, \tau_{avg}^{L}, T_L) \quad &\rightarrow \quad Solve\ for\ \beta(T_L) \quad \rightarrow \quad \beta_L = \beta(\sigma_{avg}^{L}, \tau_{avg}^{L}, T_L) \\ G_5(\sigma_{avg}^{H}, \tau_{avg}^{H}, T_H) = G_1(\sigma_{avg}^{H}, \tau_{avg}^{H}, T_H) \quad &\rightarrow \quad Solve\ for\ \beta(T_H) \quad \rightarrow \quad \beta_H = \beta(\sigma_{avg}^{H}, \tau_{avg}^{H}, T_H) \end{aligned} \tag{186}$$

We use the two values of β to derive a linear expression, Equation 187, for $\beta(T)$.

$$\beta(T) = \beta_L + \frac{d\beta}{dT}(T - T_L) \qquad , \qquad \frac{d\beta}{dT} = \frac{\beta_H - \beta_L}{T_H - T_L} \tag{187}$$

The *RR* data set has multiple set of tensile and tension-torsion experiments from which $\beta(T)$ could be derived. The tension tests show forward and reverse transformations at temperatures from -25C to +60C. In this set of experiments the layers are expected to distribute symmetrically in shear. The tension-torsion tests gives us forward and reverse shear transformations at various temperatures as well as varying amounts of constant tensile stress. Because we assume a single energy landscape and set of variants causes both the tensile and shear behavior, we should be able to use either set to calculate the chemical free energy β. Figure 102 shows two sets of $\beta(T)$ values, one calculated using the uniaxial tension test data and one using combined tension-torsion (70 MPa axial stress) data.

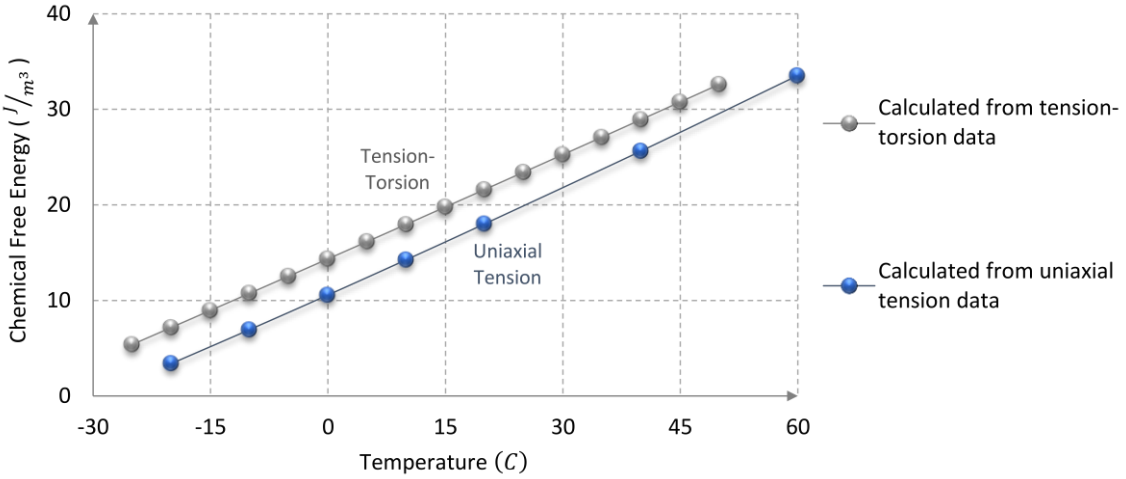


Figure 102: Comparison of chemical free energies calculated from tension-torsion and uniaxial tension data

The two curves do have almost identical slopes, which implies that the entropy differences are equal for the two types of transformation. The vertical difference between the curves could be attributed to several factors. It is only speculation that all of the martensite variants have the same chemical free energy functions. It is possible that, between the pure tension and tension-torsion test, the sets of variants activated have different chemical free energies. This would yield different curves as shown in Figure 102. It should also be pointed out that different inhomoegeities could be present in the tension and tension-torsion specimens. This could also manifest in different apparent transformation stresses, which would alter the data used to drive the β function. For the purposes of opperating the model, we assume that these two curves are similar enough. The choice of which data is used to define the chemical free energy is dictated by which data is used to populate the eight transformation stress parameters (σ^L_{fwd}, σ^L_{rev},…, τ^H_{fwd}, τ^H_{rev}).

Step 6: Latent heats

The identification of an expression for the CFE vs. temperature allows us to define several additional caloric properties. We begin by parsing the CFE into an internal energy constant terms and an entropy constant-temperature term.

$$\beta_i(T) = u_i^R - s_i^R T \qquad where \qquad \begin{matrix} u_i^R & \rightarrow & Relative\ martensite\ internal\ energy\ constant, \\ & & u_i^R = \hat{u}_i^R - \hat{u}_a^R \\ s_i^R & \rightarrow & Relative\ martensite\ entropy\ constant, \\ & & s_i^R = \hat{s}_i^R - \hat{s}_a^R \end{matrix} \tag{188}$$

Constants u_i^R and s_i^R in Equation 188 are relative quantities, measured with respect to the austenite phase. The ^'s in the right-most set of equations indicated absolute values and the constants are defined this way to be consistent with the CFE, which is a relative energy. Since Equation 187 also expresses the CFE as a function of temperature, we can use its slope term define the entropy costant.

$$s_i^R = -\frac{d\beta_i}{dT} \qquad where \qquad \frac{d\beta_i}{dT} \rightarrow Determined\ from\ experimental\ stress-strain\ data \tag{189}$$

Now, the only unknown in Equation 188 is the internal energy constant u_i^R and we can find this be evaluating the expression at $T = 0$.

$$u_i^R = \beta_i(T = 0) \tag{190}$$

Next, we can use the relationship $g = h - Ts$ (another way to express Gibbs free energy) to extract enthalpy h from Equations 182 for G and Equation 188 (for β). We begin by writing an expression for the driving force from austenite to martensite variant i.

$$F_{a\rightarrow i} = G_a(\sigma,\tau,T) - G_i(\sigma,\tau,T) = \hat{G}_a(\sigma,\tau,T) - \hat{G}_i(\sigma,\tau,T) \tag{191}$$

Equation 191 is written in two equivalent statements, one with the relative GFE's and once with absolute GFE's. Now, we express the absolute GFE's in terms of enthalpy, entropy constant, and temperature.

$$F_{a\rightarrow i} = \hat{G}_a(\sigma,\tau,T) - \hat{G}_i(\sigma,\tau,T) = \left(\hat{h}_a - T\hat{s}_a^R\right) - \left(\hat{h}_i - T\hat{s}_i^R\right) \tag{192}$$

We rearrange the right-hand side of Equation 192 to have a relative enthalpy that is constistent with other free energy quantities ($h_i = \hat{h}_i - \hat{h}_a$).

$$F_{a\rightarrow i} = \left(\hat{h}_a - T\hat{s}_a^R\right) - \left(\hat{h}_i - T\hat{s}_i^R\right) = \left(\hat{h}_a - \hat{h}_i\right) - T\hat{s}_a^R + T\hat{s}_i^R = -h_i + \left(\hat{s}_i^R - \hat{s}_a^R\right)T = -h_i + s_i^R T \tag{193}$$

Now we can write an expression for enthalpy in terms of the driving force or directly in terms of the GFE's.

$$h_i = s_i^R T - F_{a\rightarrow i} = s_i^R T + G_i(\sigma,\tau,T) - G_a(\sigma,\tau,T) \tag{194}$$

Equation 194 independently gives the enthalpy for each martensite variant i (recall that $a = N + 1$ and N is the number of martensite variants). For the system examined here (the *RR* data set), we have assumed four martensite variants so $i = 1 - 4$ and $a = 5$. We have also assumed that the caloric properties are equal for all of the martensites. In this situation, we can drop the indices on h, u_R, and s_R as was done for β.

Step 7: Critical Free Energies

The transformation criteria for this model is based on a critical Gibbs free energy, which is analogous to the 1-D inflection point-based model in Chapter 3. The critical free energies are determined from a single set of experimentally measured stresses and temperatures but can then be used to predict transformation in a more general manner.

$$\begin{matrix} Combned\ loading \\ experiment\ at \\ two\ different \\ temperatures \end{matrix} \rightarrow \begin{matrix} T^H, \sigma_{fwd}^H, \sigma_{rev}^H, \tau_{fwd}^H, \tau_{rev}^H \\ T^H, \sigma_{fwd}^H, \sigma_{rev}^H, \tau_{fwd}^H, \tau_{rev}^H \end{matrix} \rightarrow \begin{matrix} Convert\ to \\ critical\ GFE\ vs. \\ temperature \\ G_{fwd}^c(T), G_{rev}^c(T) \end{matrix} \rightarrow \begin{matrix} Use\ to\ predict \\ transformations \\ for\ arbitrry\ stress \\ and\ temperature \\ combinations \end{matrix}$$

To capture hysteretic effects, we distinguish between forward and reverse transformations by defining one critical energy for each. G_{fwd}^c and G_{rev}^c are temperature dependent parameters representing forward and reverse transformation conditions respectively. It is important to note that G_{fwd}^c and G_{rev}^c are not specific to any particular variant, component of stress, or whether the transformation is stress-induced vs. temperature-induced. Once

established, the two temperature-dependent parameters are used to predict all of the behaviors shown in Figures 73-77. For the set of experimental values chosen, we begin by defining four independent transformation stress vs. temperature functions.

$$Experimenal\ data \rightarrow \begin{matrix} T^H, \tau_{fwd}^H, \\ T^L, \tau_{fwd}^L \end{matrix} \rightarrow \frac{d\tau_{fwd}}{dT} = \frac{\tau_{fwd}^H - \tau_{fwd}^L}{T^H - T^L} \rightarrow \tau_{fwd}(T) = \tau_{fwd}^L + \frac{d\tau_{fwd}}{dT} \cdot (T - T_L) \quad (195)$$

$$Experimenal\ data \rightarrow \begin{matrix} T^H, \sigma_{fwd}^H, \\ T^L, \sigma_{fwd}^L \end{matrix} \rightarrow \frac{d\sigma_{fwd}}{dT} = \frac{\sigma_{fwd}^H - \sigma_{fwd}^L}{T^H - T^L} \rightarrow \sigma_{fwd}(T) = \sigma_{fwd}^L + \frac{d\sigma_{fwd}}{dT} \cdot (T - T_L) \quad (196)$$

$$Experimenal\ data \rightarrow \begin{matrix} T^H, \sigma_{rev}^H \\ T^L, \sigma_{rev}^L \end{matrix} \rightarrow \frac{d\sigma_{rev}}{dT} = \frac{\sigma_{rev}^H - \sigma_{rev}^L}{T^H - T^L} \rightarrow \sigma_{rev}(T) = \sigma_{rev}^L + \frac{d\sigma_{rev}}{dT} \cdot (T - T_L) \quad (197)$$

$$Experimenal\ data \rightarrow \begin{matrix} T^H, \tau_{rev}^H, \\ T^L, \tau_{rev}^L \end{matrix} \rightarrow \frac{d\tau_{rev}}{dT} = \frac{\tau_{rev}^H - \tau_{rev}^L}{T^H - T^L} \rightarrow \tau_{rev}(T) = \tau_{rev}^L + \frac{d\tau_{rev}}{dT} \cdot (T - T_L) \quad (198)$$

It is important to note that Equations 195-198 are not directly used as the transformation criteria. The normal stress equations do not include the influence of shear stress and the shear stress equations do not account for the influence of normal stress. To account for the coupling expeerienced during combined loading, we use the expression to derive the critical Gibbs free energies in Equations 199 and 200.

$$G_{fwd}^c(T) = G_5\big(\sigma_{fwd}(T), \tau_{fwd}(T), T\big) \quad (199)$$

$$G_{rev}^c(T) = G_1(\sigma_{rev}(T), \tau_{rev}(T), T) \quad (200)$$

Equation 199 assumes that the layers in the austenite phase will transform at the same, critical, value of Gibbes free energy regardless of the temperature or normal/shear stress ratio. Equation is 200 is similar, except it ddresses the layers in the martensite $++$ well and uses the reverse transformation stress. We assume that it apples to all of the martensite variants and that, if the layers in any martensite energy well exceed the critical Gibbs free energy, they will exit that phase.

Now, we have a transformation criteria that is independent of stress and therefore is applicable to any type of loading (i.e. uniaxial, proportional, nonproportional). It is derived from a specific set of transformation stresses, but the use of critical free energy generalizes it to any situation. Figure 103 shows a set of critical free energy values calculated from the 70 MPa axial stress, combined tension-torsion test data.

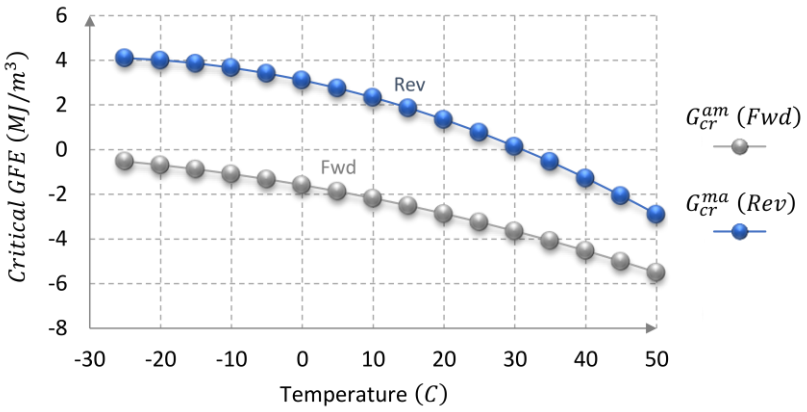


Figure 103: Critical Gibbs free energies

We can see that both critical free energy functions are nonlinear due to the quadratic nature of the GFE function (Equation 182). This nonlinearity is important because it is common in 1-D SMA applications to assume a linear transformation stress function. We make a similar assumption with Equations 195-198, where we use stress values at two temperatures to define a set of linear relationships (one for normal stress and one for shear). We then convert the linear transformation stress into critical free energies. The four stress values used for input reflect one specific loading, but the critical free energies serve as transformation thresholds that apply to other combination of stress and temperature. In this way, the critical free energy criterion more general than a stress-based one.

Figure 103 is useful to gain a sense of the magnitude and relative values of the critical energies, but not for understanding their physical meaning. A better way to visualize the critical free energies is to plot them on the GFE landscape. At a given temperature, G^c_{fwd} and G^c_{rev} are constants and there exists a set of stresses that yield G^c_{fwd} as well as a set of stresses that yield G^c_{rev}. These sets of stresses can be converted to strains (by solving Equation 182 via the quadratic equation) and visualized as paths on the free energy landscape. These paths designate the stability boundaries for each phase as shown in Figure 104a. The redlines mark the critical free energies at the current temperature (which happens to be superelastic here) and the hash marks indicate which side of the boundary is the stable side. Note that the hash marks are only shown for the austenite and martensite variant 1; this is only for visual clarity. The stable/unstable domains for the other martensite variants are symmetric about the two strain axes.

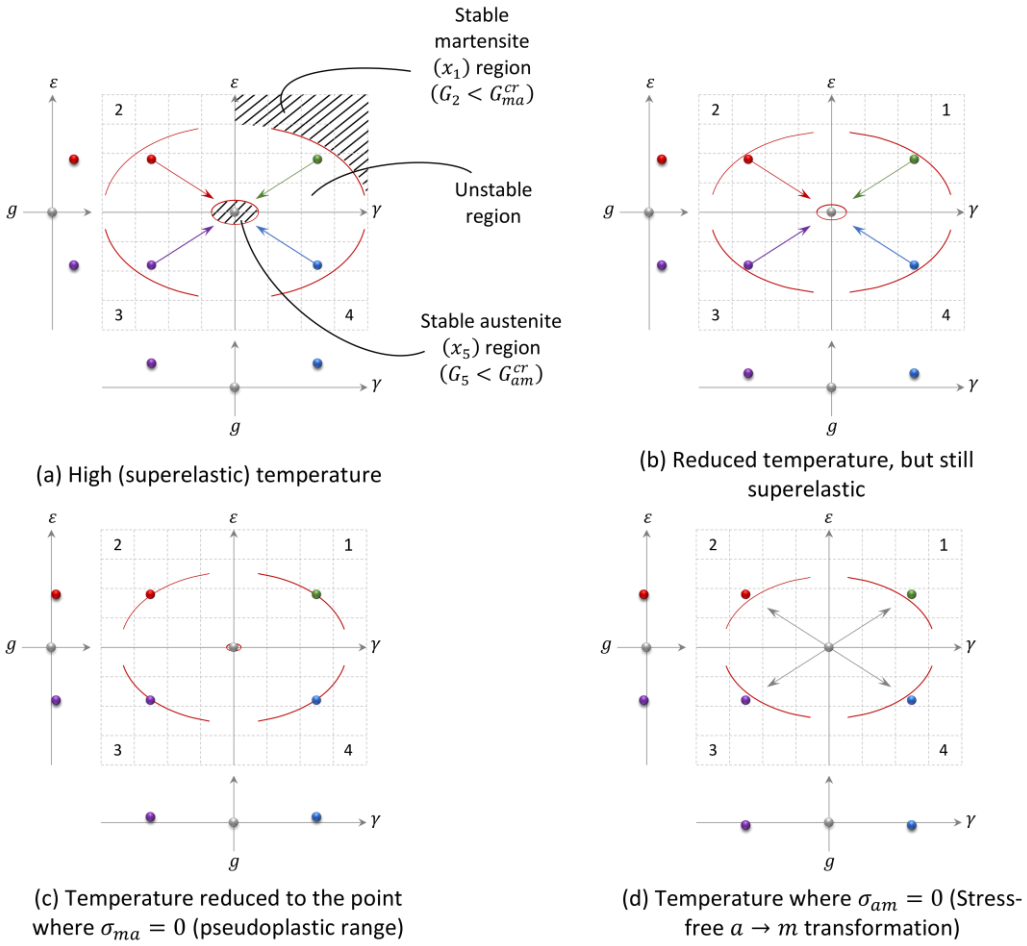


Figure 104: Critical free energies as stability boundaries in a temperature-driven transformation

There are a few additional features in the figure that illustrate the effects of the critical energy boundaries and their influence on the system. Each of the four plots contains five points; one in the center and one in each strain quadrant. These represent the location of the GFE well minimum for each phase, which also designates the average strain for each phase. The point in the center represents austenite and other four points represent the martensite variants; the variant numbers are shown for each strain quadrant. In Figures 104a and b, arrows extend from the martensite variants toward the austenite well and, in Figure 104d, they go the opposite way. The arrows indicate the transformation probabilities for each configuration. The probabilites are numerically calculated in the following section because they are a feature of the macroscopic kinetics model, but they are motivated by the mesoscopic states shown here. In Figure 104 (as well as Figures 105, 106, 107, 108, and 109), they are to be interpretted as markers of high vs. low probability for a given transformation. The presence of an arrow indicates a high probability for that transformation; the absense of an arrow signifies that a transformation is unlikely. The arrows do not necessarily mean transformation is happening, only that the probability for that transformation to occur is high. For example: the likelihood may be high that transformation $x \rightarrow y$ happens but, if well x is empty, then nothing happens. Returning to Figure 104a, we see that all the four martensite variants have a high probability of jumping to the austenite well. This is because, at the temperature shown, the equilibrium locations of the martenste phase are outside of their stable GFE domains. When a given phase leaves its stable zone, the probability that layers will exit that phase becomes high.

We have established that there are stable and unstable GFE zones for each phase and that layers exit any phase when it crosses into an unstable zone. But, this begs the question "*Which well(s) do the layers jump into?*". Notice that, in Figure 104a (and b), there are no arrows from the martensite wells to the other martensites. When a phase becomes unstable, the subset of possible transformation paths that become active is dictated by the driving forces. We assume that exiting layers will travel simultaneously to all of the other phases that have a lower GFE (a positive driving force). Furthermore, we assume that the fraction of layers that travel to each phase is proportional to the relative driving forces. For example: Phase w becomes unstable and there are three other phases, x, y, and z. The driving force to phase x is negative so the $w \rightarrow x$ transformation does not activate. The driving force to phases y and z are positive but phase z has a much lower GFE. Both transformations $w \rightarrow y$ and $w \rightarrow z$ will activate, but the number of layers jumping into to the z well will be much higher than the number of layers jumping into the y well.

The driving forces are shown in Figure 104 by the two secondary plots to the left and the bottom of the main plot. The g, ε plot shows the GFE for the layers in each phase as if you were standing to the left of the main plot and looking down the γ axis. Martensites 1 and 4 are hidden by 2 and 3 because they have the same GFE values and 3-D perspective is not used in these plots. The g, γ plot shows a view as if you were standing at the negatve end of the ε axis looking in the $+\varepsilon$ direction. In this view, the martensite 3 and 4 phases are in the forefront. The GFE plots show that, in Figures 104a-c, the GFE's of the martensites are higher than the austenite (i.e. the 1→5, 2→5, 3→5, and 4→5 driving forces are positive). Figure 104d shows that the arrows have switched directions because the austenite→martensite driving forces have become positive and the martensite→ austenite driving forces negative.

The collective behavior of the critical free energies, the well locations, and the driving forces in Figure 104 shows how the model produces a stress-free, temperature-driven transformation. Beginning in a superelastic state (104a), the martensite phases are unstable and any layers finding themselves there will instantly jump into the austenite phase. In 104b, the temperature has been decreased and the martensite stability boundary has moved inward and is not closer to the well mimina locations. Also, the stable domain of the austenite phase has become smaller which indicates that a lower amount of stress would be required to cause an austenite → martensite transformation. The temperature has been reduced further in 104c and the arrows has disappeared. The martensite wells are now just inside their stable zones so the probability of martensite → austenite transformation is low. The stable zone for the austenite phase has become very small, but still finite. This represents a temperature in the hysteresis zone, between the austenite and martensite start temperatures. Finally, in Figure 104d, the temperature has been lowered below the martensite start temperature, where the austenite phase becomes unstable. The stable austenite zone disappears and, because the $a \rightarrow m$ driving forces are positive, the $5 \rightarrow i \in 1,2,3,4$ transformation arrows appear.

Figure 105 illustrates a tensile, forward, transformation at a superelastic temperature. In 105a, the system is under uniaxial tensile stress which causes all of the phases to shift position updward (in the $+\varepsilon$ direction). The stress is not sifficient to push the austenite out of its stable zone nor is it sufficient to push the tensile martensite variants (1 and 2) into their respective stable zones. The transformation probabilites reflect this state because no arrows extend from the austenite phase and all four martensite → austenite arrows are still present. Comparing 105a to 104a, which shows a similar temperature, we see another important feature. There are no $m \rightarrow m$ transformation arrows in Figure 104a and this is because, under zero stress, all of the $m \rightarrow m$ driving forces were

equal to zero. Now, in 105a, the tensile stress has cause a positive driving force to appear between the 3,4 wells and the 1,2 wells. Because of this, two transformation arrows extend from each of the two compressive m variants. This implies that any layers that happened to be in the 3 well would jump proportionally into the a and 2 wells and that any layers that happened to be in the 4 well would jump proportionally into the a and 1 wells. The two $m \rightarrow a$ transformation arrows are longer than the two $m \rightarrow m$ arrows because the driving force, at this temperature, into the central well is higher than between the m wells. This indicates that, if layers were to transform out of the 3 and 4 wells, a greater fraction of them would find the a well. In this case the 3 and 4 wells are empty so the arrows only indicate local transformation probabilities, not active transformations.

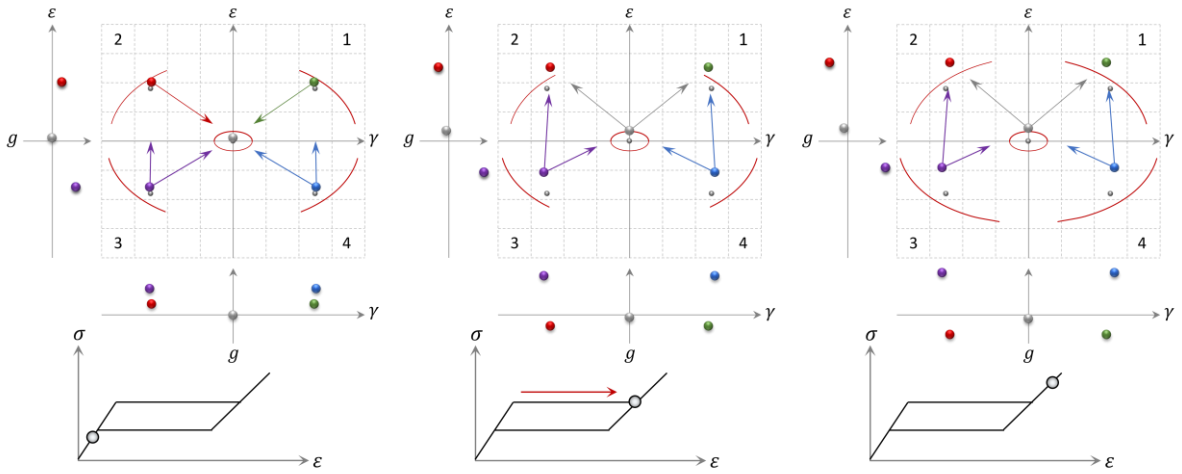
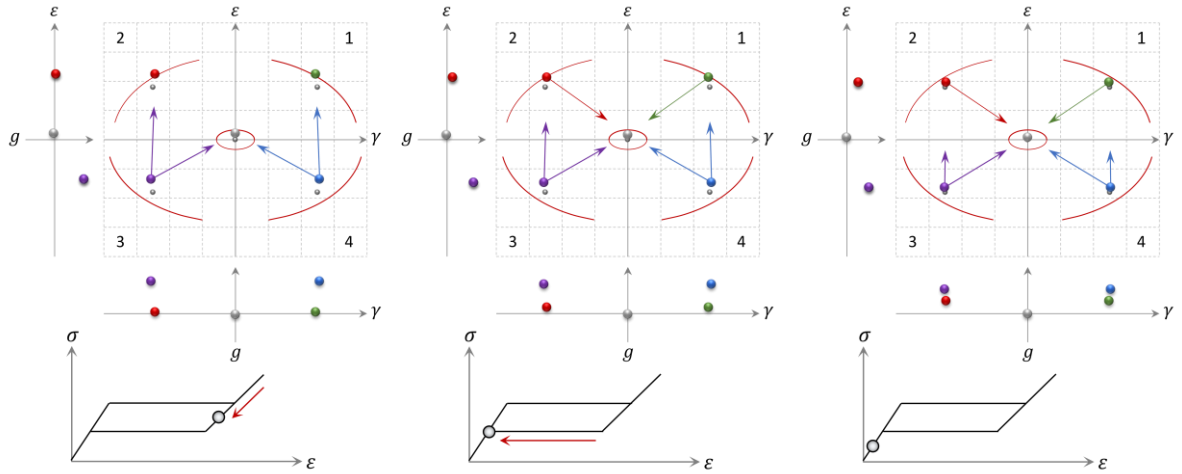


Figure 105: Mesoscopic system and resulting macroscopic stress-strain response,
Forward transformation under uniaxial tension

In Figure 105b, the tensile stress has been increased to the point that the a phase has been pushed beyond the stability boundary created by its critical free energy. Once it crosses the boundary, transformation arrows appear leading to the 1 and 2 wells because they are lower (i.e. the $5 \rightarrow 1$ and $5 \rightarrow 2$ driving forces are positive). The two tensile m phases have also been pushed into their stable regions which means that the layers coming from the a well will have a stable destination. The combination of unstable austenite and layers jumping into stable m wells results in a tensile-forward transformation. Since there is no shear stress, the two $a \rightarrow m$ transformation arrows have exactly the same length and this is to illustrate that the two transformations ($5 \rightarrow 1$ and $5 \rightarrow 2$) occur symmetrcally. One half of the layers end up in each of the two tensile m phases. This type of symmetry is how the model can represent uniaxial tension-compression as well as pure shear behavior. Normal-shear coupling does not appear unless multiaxial loading breaks the symmetry of the phase fractions. 105c shows further tensile loading after the transformations in b have been completed. Each phase is pushed further in the $+\varepsilon$ direction but the transformation arrows do not change. Further increases in stress will not change the stability status for any of the phases and this yields the linear martensite branch on the $\sigma - \varepsilon$ curve.

Figure 106 illustrates three stages during the unloading of the same system from 105. In 106a, the tensile stress has been decreased to a value that is lower than the critical forward threshold. The a phase has crossed the critical free energy boundary and returned to its stable zone. The $5 \rightarrow 1$ and $5 \rightarrow 2$ arrows have disappeared because the probability of these two transformations has dropped to ≈zero. The difference is subtle, but the GFE of the a phase is also lower than that of the two tensile m phases. This means that a positive driving force exists

that pulls layers from the 1 and 2 wells into the 5 well. Despite the stable a phase and the positive driving forces, there are no $1 \rightarrow 5$ or $2 \rightarrow 5$ transformation arrows in 106a. This is explained by the fact that the two tensile m phases are still in their respective stable zones (they have not yet been pulled back across their critical free energy boundaries). Each phase has its own stability boundary that changes with stress as well as temperature and this allows hysteresis to be captured. The stress that pushed the a phase over its boundary is higher than the stress that would allow the two m phases to slide back over their boundaries. The result is a hysteresis loop in the $\sigma - \varepsilon$ response, which is not completed till Figure 106b.



(a) Tensile unloading to a stress below the forward transformation stress

(b) Further unloading causes the tensile m variants to become unstable

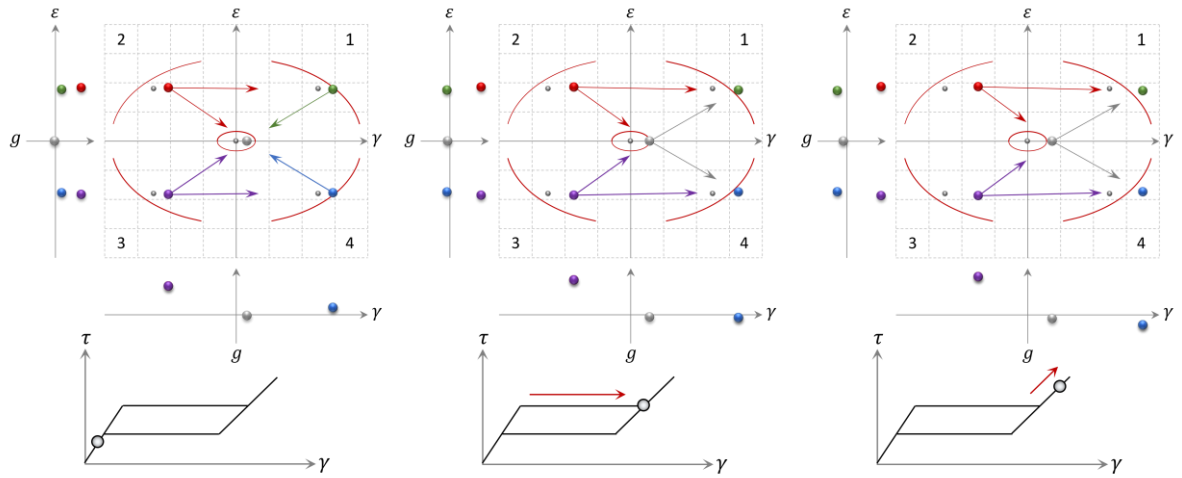
(c) Further unloading follows the linear austenite branch back to the origin

Figure 106: Mesoscopic system and resulting macroscopic stress-strain response, Reverse transformation under uniaxial tension

The stress in Figure 106b is precisely the value where the m phases hit the critical free energy boundary. Now, the two requirements for $1 \rightarrow 5$ and $2 \rightarrow 5$ transformation have been met: (1) the source phase is unstable and (2) the driving force to the destination phase is positive. All of the layers that were equally distributed between the tensile m wells instantly jump into the a well and the $\sigma - \varepsilon$ curve returns to the linear austenite branch. 106c shows continued unloading toward zero stress. The unchanged transformation arrows indicate that the stability status does not change for any of the phases and no new transformations activate; the $\sigma - \varepsilon$ behavior is linear at any stress below the reverse threshold (from 106b).

It is also worth noting that, in both Figures 105 and 106, the arrows extending from the 3 and 4 phases persist relatively unchanged. In all six configurations, these phases remain in their unstable zones and the driving forces toward the other three phases remain positive. The arrows indicate that the probability of layers transforming out of these wells is very high, but this ony applies to layers actually in phases 3 or 4. Layers never enter the compressive m wells therefore the high probabilities never activate any transformations.

Figure 107 shows a forward ($a \rightarrow m$) transformation in torsion (+ shear). The a phase in 107a is once again beging pushed toward its critical free energy boundry, but now to the right. Since the shear stress is below the that required for forward transformation, there are no arrows extending from the a phase. The stress is also below the reverse transformation stress, which is the stress where the two +shear m phases become stable. Because of this, the 1 and 4 phases are unstable and transformation arrows extend toward the a energy well.



(a) Pure shear stress at a superelastic temperature

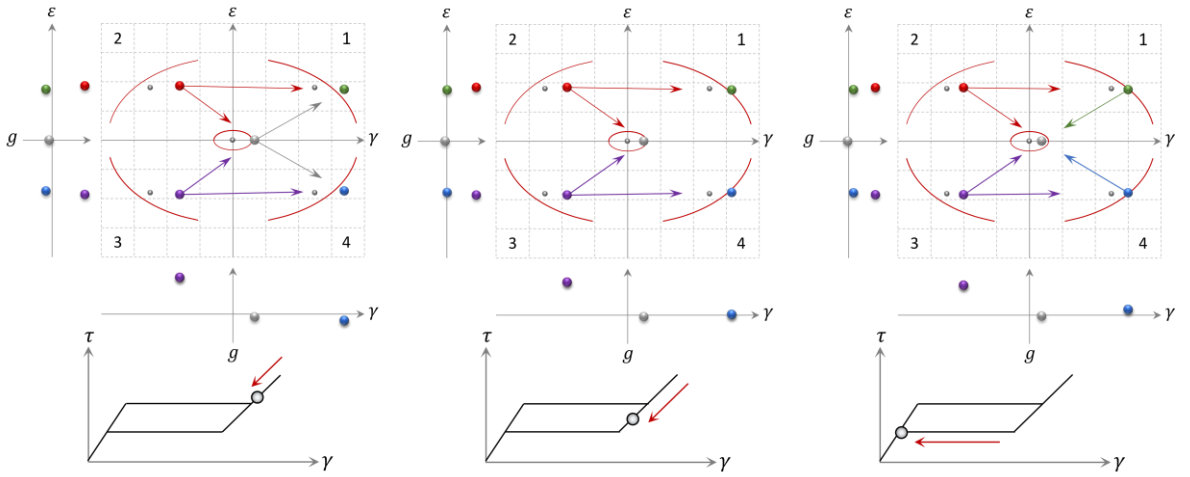
(b) Stress increased to the point where the austenite phase become unstable

(c) Additional stress causes linear behavior in the two + shear variants

Figure 107: Mesoscopic system and resulting macroscopic stress-strain response, Forward transformation in torsion

The critical shear stress is reached in Figure 107b which is the point where the a phase goes unstable. All of the layers immediately jump out of the a well and distribute themselves among all of the other phases where positive driving forces ($a \rightarrow \#$) exist. In this case, only the $a \rightarrow 1$ and $a \rightarrow 4$ driving forces are positive so only these two transformations activate. Since there is no tensile or compressive stress, the symmetry about the ε axis is maintained. The driving forces into the two +shear phases are exactly equal, the layers jump into them symmetrically, and no macroscopic normal strain is observable while the system undergoes a large change in shear strain. In 107c, the transformation is complete the system moves along the linear martensite branch of the $\tau - \gamma$ curve.

Figure 108 represents the unloading process for system in 107. In 108a the shear stress is being lowered and the macroscopic $\tau - \gamma$ curve follows the martensite branch. The transformation probabilities look essentially the same as in 107c because the shear stress is still above both the reverse and forward thresholds.



(a) Unloading on the + shear, fully martensite branch

(b) Continued unloading, to a stress lower than the forward threshold

(c) Continued unloading, to the reverse transformation threshold

Figure 108: Mesoscopic system and resulting macroscopic stress-strain response, Reverse transformation in torsion

The $a \rightarrow 1$ and $a \rightarrow 4$ transformation arrows have disappeared in 108b because the a well has reentered its stable region. This coincides with the macroscopic τ becoming lower than the forward plateau. The system is in

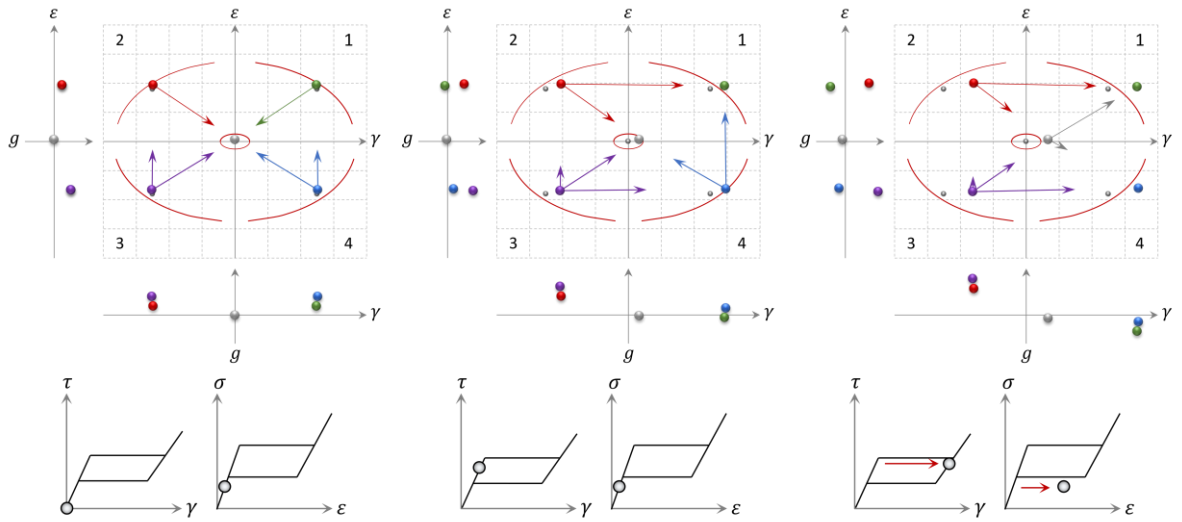
the hysteretic range, between the forward and reverse stresses, which is why no transformation arrows extend from the two +shear m phases either. Both the 1 and 4 phases are still in their respective stable strain domains and layers will remain trapped there till the reverse transformation stress is reached in 108c. Now the two +shear phases simultaneously cross their critical free energy boundaries and the transformation arrows leading into the a well appear. The signifies the reverse transformation seen in the macroscopic $\tau \rightarrow \gamma$ curve.

Figures 105-108 illustrate examples of pure tensile and pure shear transformations, where the variants remain symmetric. Next, we examine an example of non-proportional loading where both shear and axial responses activate and interact with one another. Figure 109a shows a system, at a superelastic temperature, after a tensile stress has been applied. The tensile stress shifts all of the phases in the $+\varepsilon$ direction, but is not sufficient to activate any transformations (all of the layers remain in the a phase). The two macroscopic stress-strain plots show that a linear normal strain is present but there is no shear strain. No layers are affected by it, but we also see that the $m \leftrightarrow m$ transformation arrows have become asymmetric about the γ axis. The tensile stress has created a driving force between the $-\varepsilon$ and $+\varepsilon$ sets of variants causing the $3 \rightarrow 2$ and $4 \rightarrow 1$ arrows to appear. In 109b, the tensile stress is held fixed and a small shear stress has also been introduced. The shear stress is not quite enough to activate transformation, but close. The a phase has moved diagonally toward its elliptical stability boundary and is near the edge; this is also shown by the $\tau - \gamma$ plot where the shear stress is near the forward plateau. We see that the 1 phase has entered stable territory and, therefore, transformation arrows out of that well are inactive. Viewing the $\varepsilon - \sigma$ plot alone would imply that tensile martensite is still unstable, because the stress is below the reverse plateau. But it is now the combination of normal and shear stress that dictate the stability. We see also that a $2 \rightarrow 1$ transformation arrow has appears; this is due to the shear stress creating a driving force between the two wells. The fact that non of the phases are hidden in the $\varepsilon - g$ and $\gamma - g$ lots indicates that, under this multiaxial loading, none of the m phases have the sam GFE. This means that transformations will now occur asymmetrically when activated.

109c shows the instant where transformation occurs. The a phase has been pushed out of the stable zone and its transformation arrows indicate that layers flow into both of the +shear m phases. The relative lengths of the two arrows ($5 \rightarrow 1$ and $5 \rightarrow 4$) idicate that one of the two available wells is much lower than the other. In model operation, the rate of layers moving to the 1 well will far exceed the rate going into the 4 well due to driving force disparity. The two stress-strain plots show several important multiaxial effects. Now, the actual path of the system (denoted by the movement of the point) does not follow the black stress-strain paths. The black lines, which mapped the path of the point in Figures 105-108, represent the expected uniaxial responses. In the $\tau - \gamma$ plot, we see that the stress plateau forms at a lower stress value than in the pure shear. This is an effect of the tensile stress, which lowers the shear stress required to activate transformation. Figure 79 shows that this effect is very distinct in the *RR* experimental data; at each increasing level of tensile stress, the forward shear stresses become lower. In the model, this effect is captured by the elliptical stability boundary around the a well. The vector sum of the normal and shear strains dictate the status of the transformations (on or off). In the $\sigma - \varepsilon$ plot, we see a normal strain plateau but at a stress far below even the reverse transformation stress. Physically, this reflects that fact that each m variant is offset in both the ε and γ axes. Even if a transformation is activated by shear stress, any asymmetry in the tensile stress will cause a normal transformation strain appear also. This can be seen in the *RR* experimental data (Figure 77 for example) where each torsional transformation loop also causes an axial transformation loop. The axial strains experinced in the tension-torsion experiments had smaller magnitudes than

a uniaxial tensile transformation. This is implied in 109c because the axial strain, in the $\sigma - \varepsilon$ plot, stops short of the martensite branch. The explanation for this lies in the imbalanced transformation probabilities coming out of the a well. An axial transformation strain appears because the layers do not transform symmetrically but, because some of the layers go to the 1 well and some go to the 4 well, the average macroscopic transformation strain is reduced.

At the end of the loading in 109c, most of the layers are in the 1 phase with just a few in the 4 phase. The axial stress is maintained, but the shear unloading begins in 109d. As the shear stress is reduced, the layers move to the left toward their equilibrium locations. Because of the tensile stress-induced upward shift, the 4 phase crosses its stability boundary before the 1 well as shown by the transformation probability arrows extending from it. In this configuration, the driving force into the a phase is greater, so most of the layers will undergo a $4 \rightarrow a$ jump. This results in a small reduction in γ as well as ε, but not enough to signify a stress plateau due to the small fraction of layers. The shear stress if lowered to the point where the 1 phases loses stability in 109e. Here, the layers in the 1 phase only have one choice for their transformation path ($1 \rightarrow a$) because the a well is the only lower well. The a phase now contains all of the layers and the stress-strain responses return to the linear austenite brach with almost no change in stress (reverse plateaus).



(a) Pure tensile loading, well below the transformation threshold

(b) Torsional stress added, but not enough to cause transformation

(c) Imbalanced transformation into the + twist variants causes axial strain

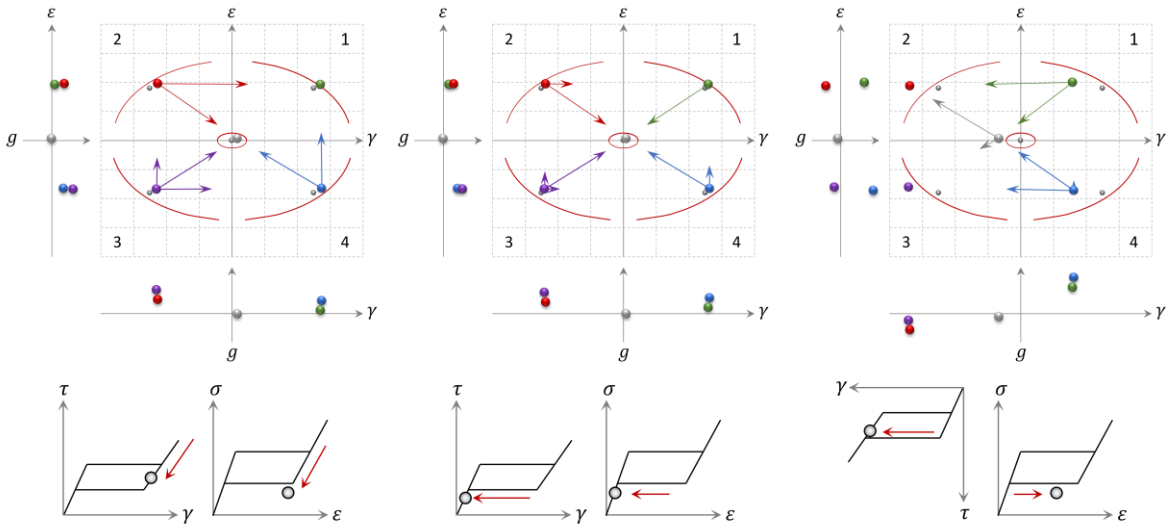


Figure 109: Mesoscopic system and resulting macroscopic stress-strain response, Combined tension and torsional transformations

In Figure 109f the shear unloading has continued past zero and the system is now under -shear stress. The GFE plot is mirror symmetric about the γ axis when compared to 109c because the magnitudes of shear stress are the same. The $\tau - \gamma$ plot is now shows in the thrid quadrant but it is important to note that the $\sigma - \varepsilon$ plot is the same as in 109c. Desipte the reversal in shear loading, the positive tensile stress has remained fixed. This causes the layers to prefer the $a \rightarrow 2$ path over $a \rightarrow 3$ and results in a net positive tensile strain.

Figures 104-109 illustrate the concept of how the critical free energy boundaries dictate transformations, whether induced by temperature, normal stress, or shear stress. We can evaluate the theory using the transformation stresses and temperatures from the *RR* experimental data (Figures 79-81) and the critical free energies shown in Figure 103. For a given experiment, we know the temperature and the combination of stresses that caused transformation. We also have calculated the GFE's for each experimental temperature. For the tension-torsion tests, we consider the applied tensile stress (either 70, 132, 255, 317, or 379 MPa) and the critical GFE to be known constants, then solve Equation 182 for the shear stress. Since the GFE is equal to the critical value, then the shear stress should be the transformation stress and we can compare these values to the experimental values. Figure 110 shows the experimental transformation shear stresses (a) along with the calculated values (b).

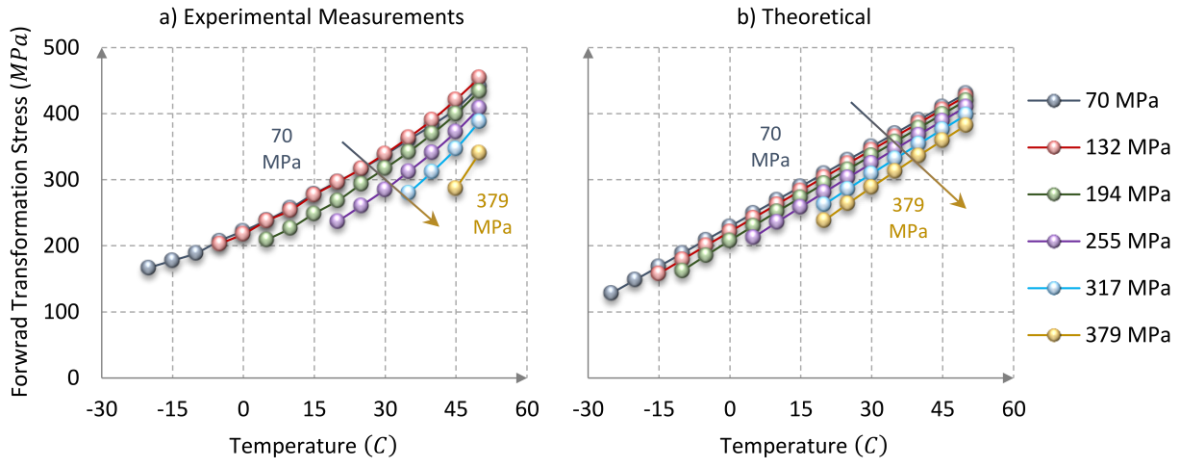


Figure 110: Shear stresses causing forward transformations stresses under combined tension-torsion loading (a) experimental (b) theoretical

The first observation that should be made in Figure 110 is that the theory captures both of the salient trends in the experimental data. 1) At each level of axial stress, the theoretical transformation decreases linearly with temperature and all of the lines have the same slope. The experimental measurements have a modest curvature but can reasonably be described as linear. The slopes are similar with the most notable exception being the 379 MPa curve, which deviates from the other curves in both slope and separation. At this level of axial stress, the notion of a *forward* transformation ($a \rightarrow m$) is beginning to lose meaning because the axial stress alone is approaching a level of stress that makes the austenite phase unstable. 2) The second trend that is captured by 110b is that, for each increasing value of axial stress, the entire transformation stress-temperature curve shifts downward. Physically, this indicates that the shear stress required to pull layers out of the a energy well gets lower as more axial stress is present. The downward shift is slightly more drastic in the experimental data and, once again, the theory does not capture the more pronounced deviation in 379 MPa curve.

It should also be pointed out that data points representing $m \rightarrow m$ transformations were excluded from Figure 110. The occurrence of $a \rightarrow m$ vs. $m \rightarrow m$ behavior is addressed at the kinetics model level by including the effects of the driving forces. The comparison here is restricted to experimental data points where only one type of transformation ($a \rightarrow m$) is clearly active. Figure 111 is similar to 110 except it shows the reverse transformation shear stresses.

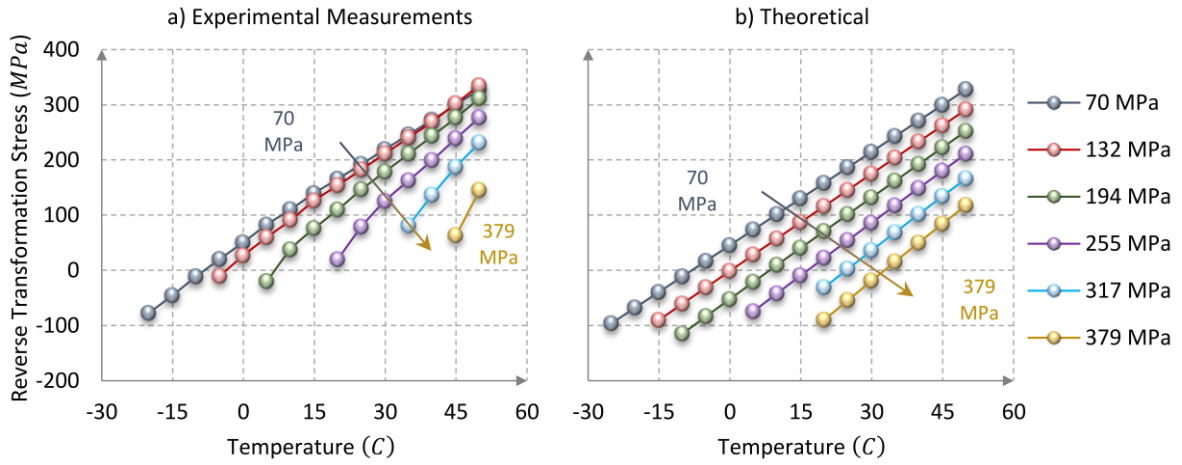


Figure 111: Shear stresses causing reverse transformations stresses under combined tension-torsion loading (a) experimental (b) theoretical

The reverse transformation stresses exhibit the same trends as their forward counterparts, but the influence of axial stress much more pronounced. At 50 C, for example, the maximum deviation in forward transformation stress (Figure 110a) is $\approx$100 MPa whereas the deviation in Figure 111a (at 50 C) is $\approx$180 MPa. The experimentally measured data sets have a slight curvature, but the sign of the curvature is reversed compared to the forward measurements. In both cases, however, the curvature is small and neglected by the model. We maintain the assumption that, if one component of stress is held constant, the other component that is required for transformation will vary linearly with temperature. The curves in Figure 110 have the same general characteristic shape, with the exception of the 379 MPa curve. In Figure 111a, by contrast, the curves show a prominent change in shape as the axial stress is increased. The 70 and 312 MPa curves are essentially linear, then the curves become more curved and downward sloping as the axial stress continues to increase. The theoretical predictions (111b), at a qualititative level, capture most of the trends in the experimental measurements. Higher values of axial stress mean that lower values of shear stress are required to cause transformation. There is a very slight nonlinearity in the separation between the theoretical curves, but it fails to capture the clustering of the experimental curves at low axial stress nor the large deviation between the high axial stress curves. The theoretical curves all have the same slope, which means they also do not capture the increase in curvature exhibited by the experimental measurements.

It is important to note that each theoretical data point in Figures 110b and 111b relflects a single shear stress value calculated from a critical GFE and axial stress. The experimental data, on the other hand, represents a wire with nonuniform stress, strian, and phase fraction fields. We speculate that the curvature in the theoretical curves is due to macroscopic inhomogeneity in the specimens. For example, if the local transformation regions in the wire had different sizes from one test to the next, this would add nonlinearity to the measurements because they only reflect the overall wire response. Some field effects (like radial strain variation) are captured by the employing multiple instances of the macroscopic model, but others also neglected (like axial localization) and left for future work.

6.6 Macroscopic Model

The macroscopic model represents the stress-strain-temperature-phase behavior of a single material point in a tensile-test specimen-sized body. The distinction of a *single point* implies that no spatial variations are considered for any of the fields. This means that the influence of surrounding material, field gradients, field motion, field shape, etc. are neglected in this formulation. These effects play a significant role the macroscopic behavior of SMA specimens, as was discussed in *Section 5*, but the goal of this section is to lay the foundation for the core constitutive model.

Point-wise material models are not limited to describing uniform specimens, however. Describing field effects using a point-wise material model simply requires multiple instances of the model to be associated with different spatial regions in the specimen. We employ this approach to account for the shear strain gradient that is known to exist in a wire under torsion. Each simulation used to test and evaluate the model, consisted of a series of material points that were spread radially through the cross-section as shown in Figure 112b.

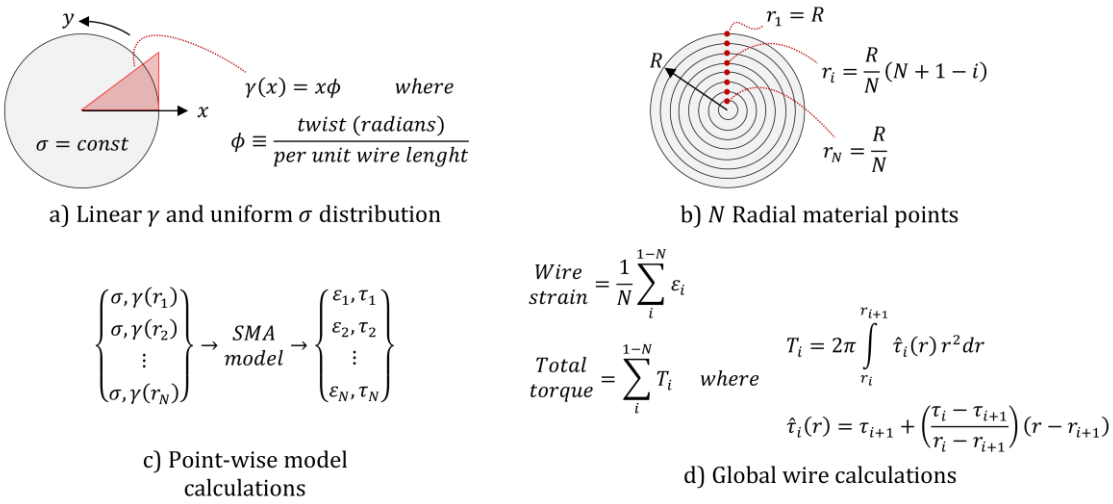


Figure 112: Wire cross-section evaluation points and calculation of total wire response

The goal of this set up is to capture the smooth transitions into and out of transformation observed in the experiments. These curves have shapes like those typically associated with polycrystalline SMAs and polycrystallinity may be a contributing factor, but it is not the only one. In a torsion experiment, the strain experienced is nonuniform through the cross-section. This means that different radial locations will begin and complete transformations at different levels of stress, which will also give the stress-strain curves smooth, rounded shapes. For example, the simulations that generated the curves in Figure 99 were executed with 20 material points and no polycrystalline effects. The shape of the simulated stress-strain curves agrees well with the experimental data. This implies that it is possible for a single crystal model to capture the basic response from the *RR* tension-torsion tests. On the other hand, the torsion specimens were trained prior to testing. The typically causes specimens, even ones that begin yielding single crystal-type curves, to assume polycrystalline characteristics. It is likely that a combination of polycrystalline effects and macroscopic field variations were active the tension-torsion experiments and this is discussed further as the model is compared to the experiments.

Figure 112a illustrates a linear distribution of shear strain with radius and, which is the traditional assumption in linear torque-twist equations. We make the same assumption here, but also acknowledge that the strain field in an SMA wire cross-section may be more complex. It is conceivable that, as the outer potion of the wire begins to transform, the inner portion does not and a nonlinear shear strian distribution forms. This could be investigated by implementing the model is a more general simulation (such as a 3-D finite element code), but this is outside the scope of the current study. Figure 112c and d show the input and output data from the material points in b. We assign a constant axial stress and a radially-varying shear strain (as well as a constant temperature) to each material point. The model is executed independently for each point and the output is a series of axial strian and shear stress pairs. The wire axial strian is assumed to be the average of the each point-wise axial strain. The shear stress is assumed to vary linearly between the points and then integrated to obtain the total torque (112d). Now, having introduced how the point-wise model was used to simulate the experiments, we focus on the evaluation of a single material point, i.e. the main macroscopic model.

6.6.1 Stress-Strain Relationship

At the macroscopic scale, each material point has a normal strain ε and a shear strain γ and these are computed as the sum of the mesoscopic strains, weighted by the phase fractions.

$$\varepsilon = \sum_i x_i \cdot \varepsilon_i(\sigma) \quad , \quad \gamma = \sum_i x_i \cdot \gamma_i(\tau) \tag{201, 202}$$

The normal strain represents the length-wise deformation in a thin, wire-like body and the shear strain is associated with wire twisting (see Figure 82). The normal and shear stresses (σ and τ respectively) are the macroscopic work conjugates to the strian measures; engineering strains and stresses are assumed here. ε_i and γ_i are the mesoscopic strains for each variant i and these are calculated using Equations 178 into 179. These quantities are mesoscopic but can be used in the macroscopic equations here because strain is a per-unit length quantitiy. As discussed in the *Mesoscopic Free Energy* section, when the macroscopic stresses σ and τ are fed to the mesoscopic strain equations they take the meaning of the nominal stresses in the mesoscopic system. The phase fractions, x_i, dictate the amount of material in each phase and are detailed later in this section.

Equations 201 and 202 represent the main stress-strain relationships for the model. If we expand the mesoscopic strains and rearrange terms, we can express each component of stress as a function of the strains and phase fractions. Using Equation 178, we can rewrite Equation 201 and then solve for normal stress.

$$\varepsilon = \sum_i x_i \cdot \left(\frac{\sigma}{E_i} + \varepsilon_i^R\right) \quad \rightarrow \begin{matrix} Solve \\ for\ \sigma \end{matrix} \rightarrow \quad \sigma(\varepsilon, x_i) = \frac{\varepsilon - \sum_i x_i \varepsilon_i^R}{\sum_i \frac{x_i}{E_i}} \quad i \in i-5 \tag{203}$$

Similarly, we use Equation 179 and Equation 202 to solve for shear stress.

$$\gamma = \sum_i x_i \cdot \left(\frac{\tau}{Y_i} + \gamma_i^R\right) \quad \rightarrow \begin{matrix} Solve \\ for\ \tau \end{matrix} \rightarrow \quad \tau(\gamma, x_i) = \frac{\gamma - \sum_i x_i \gamma_i^R}{\sum_i \frac{x_i}{Y_i}} \quad i \in i-5 \tag{204}$$

All of the parameters required to evaluate Equations 201 and 202 (or 203 and 204) have been defined except for the phase fractions x_i. The phase fractions are macroscopic quantities and are governed by the model's kinetics law.

6.6.2 Kinetics Equations

The phase fractions are macroscopic quantities that represent the amount of material in each phase. Their values are governed by a set of kinetic ODE's given by Equations 205 and 206.

$$\dot{x_i} = \sum_j x_j p_{ji} - x_i \sum_j p_{ij} \quad where \quad i \in 1-4 \quad and \quad j \in 1-5 \tag{205}$$

$$x_5 = 1 - \sum_j x_j \quad where \quad j \in 1-4 \tag{206}$$

Each martensite variant i has a governing equation where the first term represents layers transforming into that phase ($j \rightarrow i$ transformations) and the second terms represents layers exiting it ($i \rightarrow j$ transformations). Since the phase fractions always sum to one, we do not need a differential equation for the austenite phase fraction which is given by Equation 206. The parameters p_{ij} represent the rate of transformation from phase i to phase j and are derived from the GFE. These are often referred to as transformation *probabilities* because of their geneses in statistical mechanics but, in the 2-D formulation presented here, they are relative free energy-based quantities. The term transformation probability is still an apt description, however, because of the multiscale nature of the model. The macroscopic equations bypass many of the complexities in the mesoscopic system by sampling the free energy, then operating only on what the mesoscopic system is most likely to do. The transformation probabilities are given by Equation 207.

$$p_{ij} = p_{max} \cdot F_{ij} \cdot \Delta G_j \cdot e^{-\alpha \Delta G_i} \quad where \quad \begin{matrix} for\ i = 1, j \in 2,4,5 \\ for\ i = 2, j \in 1,3,5 \\ for\ i = 3, j \in 2,4,5 \\ for\ i = 4, j \in 1,3,5 \\ for\ i = 5, j \in 1-4 \end{matrix} \quad \rightarrow \quad p = \begin{bmatrix} - & p_{12} & - & p_{14} & p_{15} \\ p_{21} & - & p_{13} & - & p_{25} \\ - & p_{32} & - & p_{34} & p_{35} \\ p_{41} & - & p_{43} & - & p_{45} \\ p_{51} & p_{52} & p_{53} & p_{54} & - \end{bmatrix} \tag{207}$$

The transformation probabilities p_{ij} act like the start/stop conditions for the various transformations $i \rightarrow j$. Each term contributes to activating or deactivating the various transformations as shown here in Figure 113.

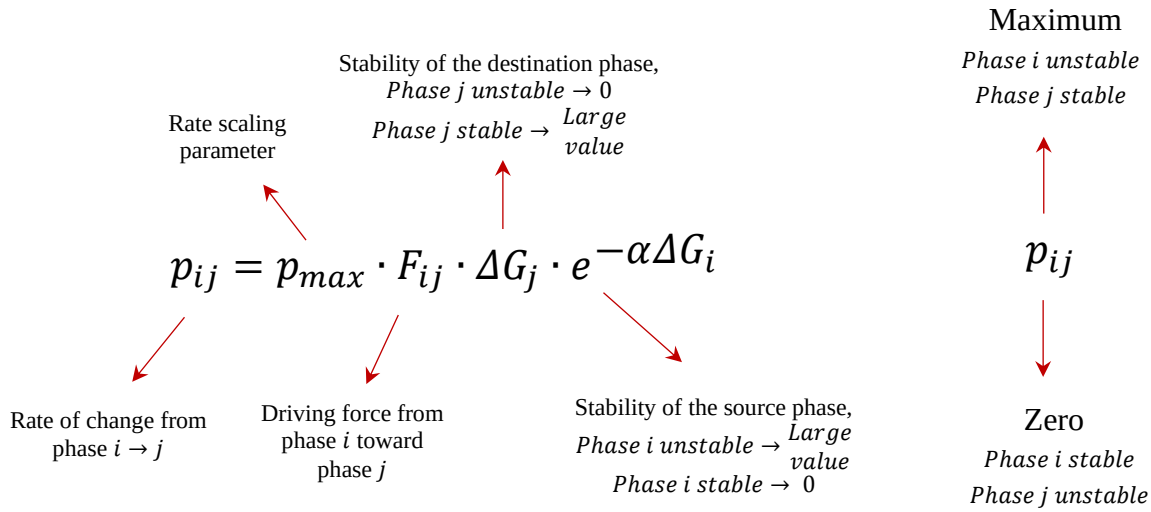


Figure 113: Diagram of terms in the transformation probability expression

The right-hand expression included with Equation 207 shows a single matrix quantity p. The matrix form is mainly used as a concise way to show the sequence of indices for transformation-related quantities like p. The i index represents the beginning phase as well as the row number, the j index represents the destination phase along with the column, and the "−"'s indicate transformations that are not allowed. The parameter F_{ij} is the driving force between energy wells i and j which is determined from the GFE's as shown here by Equation 208.

$$F_{ij} = max(G_i - G_j, 0) \qquad where \qquad i \in 1-5 \quad , \quad j \in 1-5 \tag{208}$$

The driving force parameter is only used for the active terms in p, but all 25 values can be easily calculated and the notation is simpler without the selective sequence of indices. The values for $G_{i\ or\ j}$ are given by Equation 182 from the *Mesoscopic Free Energy* section. The parameter p_{max} is the maximum possible rate of change in phase fraction. In the 1-D model this term was written as $1/\tau$, where τ was related to the attempt frequency of the layers in a given phase. This term has its origin in statistical mechanics but, for a solid material, acts simply as a transformation rate tuning parameter. The rates of transformation must be finite for the ODE's to be numerically integrated, but they must also be very high compared to the macroscopic strain rates in order to prevent artificial damping and artificial widening of the hysteresis loops. These effects were examined in detail in the context of the 1-D model. For the 2-D model we adopt the parameter p_{max} because τ is reserved for shear stress. The parameter α is also derived from statistical considerations and took the form of $1/\gamma$ or $K \cdot T/v$ in various versions of the 1-D model. Once again, for a solid material, the need to individually include the Boltzmann constant (K), temperature (T), and activation volume (v) is eliminated and the entire term acts like a ramp rate for the transformations. When the GFE conditions dictate that a given transformation should become active, α controls how rapidly the rate of transformation jumps from zero to p_{max}. For the 2-D model, we have switched to the

parameter α because γ is used for shear strain and there is no need to individually resolve K, T, and v in this context. ΔG, like F, is calculated from the GFE but it reflects the relative energy between the layers in a given energy well and the critical GFE value. Because the austenite and martensite phases have in inverted stability vs. temperature relationships (i.e. a is stable at high temperatures while m is stable at low temperatures), their ΔG expressions are different. For the austenite phase, ΔG is given by Equation 209.

$$\begin{aligned} \Delta G_{i\in 1-4} = max(G_{rev}^{c} - G_i, 0) \quad &\rightarrow \quad \begin{cases} > 0 & \rightarrow \quad Layers\ will\ remain\ in\ well\ i \\ \cong 0 & \rightarrow \quad Layers\ will\ exit\ well\ i \end{cases} \\ \Delta G_5 = max\left(G_5 - G_{fwd}^{c}, 0\right) \quad &\rightarrow \quad \begin{cases} > 0 & \rightarrow \quad Layers\ will\ remain\ austenite \\ \cong 0 & \rightarrow \quad Layers\ will\ exit\ the\ austenite\ well \end{cases} \end{aligned} \tag{209}$$

The G values are already available because they were calculated during the calculation of F. The critical free energies, G_{fwd}^{c} and G_{rev}^{c}, were given by Equations 199 and 200. Equation 209 complets the calculations required to build the system of kinetics equations, but a handful of additional relationships must be defined to operate the macroscopic model. Next, we show how to use the mesoscopic stress-strain equations to relate macroscopic stress and strain quantities under several different scenarios.

This completes the presentation of the model equations and the following section is devoted to simulations of the *RR* experiments and discussion of the model's performance.

6.7 Simulations

In this section we demonstrate the model's performance by the *RR* tension tests and the combined tension-torsion tests.

6.7.1 Uniaxial Tension Cases

The tests shown in Figure 96 were used to populate the material parameters and the values are summarized here in Table 7.

Parameter	Value	Parameter	Value
E_a	60 (GPa)	G_a	17 (GPa)
E_m	30 (GPa)	G_m	15 (GPa)
ε_R	5. 4 (%)	γ_R	6. 5 (%)
T_H	60 (C)	T_L	0 (C)
σ_{fwd}^{H}	665 (MPa)	σ_{fwd}^{L}	350 (MPa)
σ_{rev}^{H}	475 (MPa)	σ_{rev}^{L}	25 (MPa)
τ_{fwd}^{H}	0 (MPa)	τ_{fwd}^{L}	0 (MPa)
τ_{rev}^{H}	0 (MPa)	τ_{rev}^{L}	0 (MPa)
M	0 (GPa)		

Table 7: Material parameters used for the uniaxial tension and thermal cycling simulations

Figure 114 shows the full-cycle tension experiments (—) plotted with the simulation results (—).

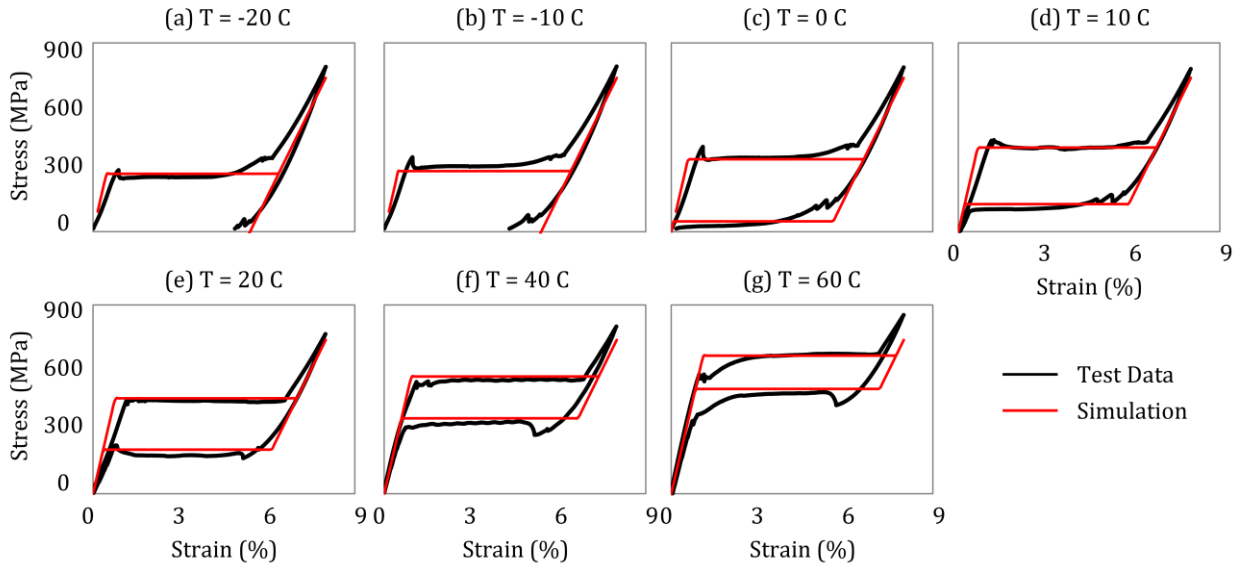


Figure 114: Uniaxial tension tests with a full cycle

The overall agreement is good and the effects of interest, the pseudoplastic → superelastic transition and shrinking hysteresis, are well-captured. The test data curves contain some irregularities due to global localization effects and these effects are not included in the current model. These are most prominent in the 60 C case (g). The experimental transformation strain also begins to decrease at the highest two temperatures (f and g). This is an unexpected result and may be related to the narrower hysteresis loop, but this is speculation only. This effect is not expected and is not part of the intended model behavior.

The next set of tests was similar to those shown in Figure 114, except there was a partial unloading near the mid-transformation point. The simulation results are plotted along with the test data in Figure 115.

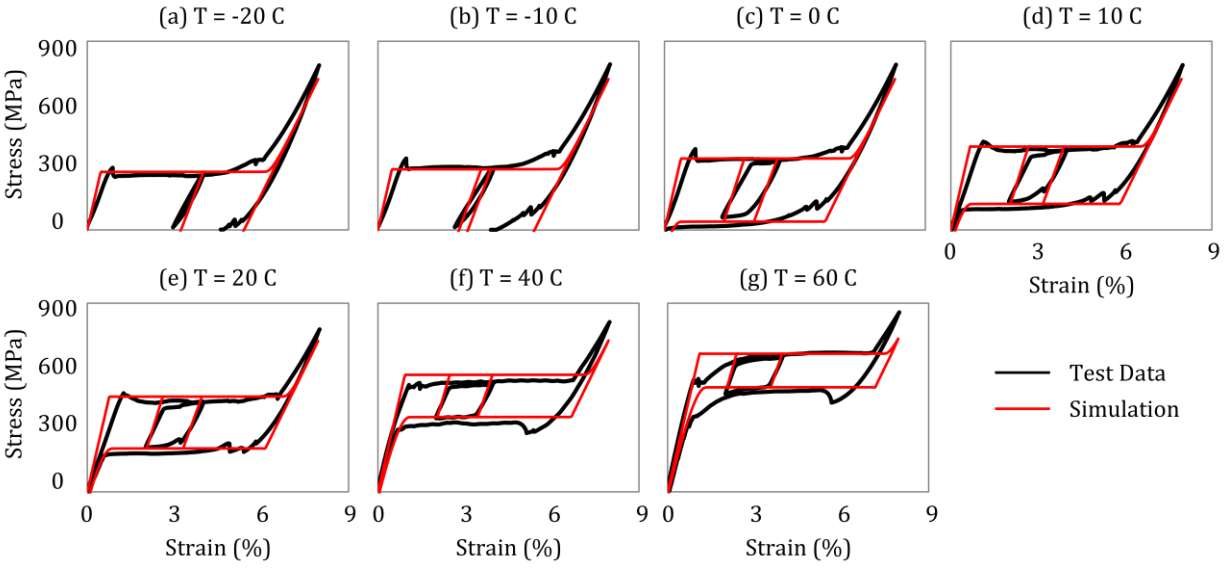


Figure 115: Axial tension tests with a partial unload cycle

The first observation to be made is that the model captures the partial unloading behavior accurately. The slopes of the partial unload and reload loops in the simulations are slightly steeper than the test data. We assume that this is related to the presence of an R-phase, which is not included in the model. We also speculate that the R-phase is why the initial loading branches in the experiments below 40 C (a-e) are slightly nonlinear and have slopes that deviate from the simulations. The second interesting observation is that the curve irregularities (i.e. the

stress spikes at the beginning of the transformation plateaus) are almost identical to those in Figure 114. The macroscopic localization events that cause these irregularities often initiate at the grip-ends of the specimen and, as indicated by the tests here, are repeatable events over at least a few cycles.

The next set of simulations where of the actuation tests, where the specimens were temperature cycled under different amount of static, axial load. The results are shown in Figure 116.

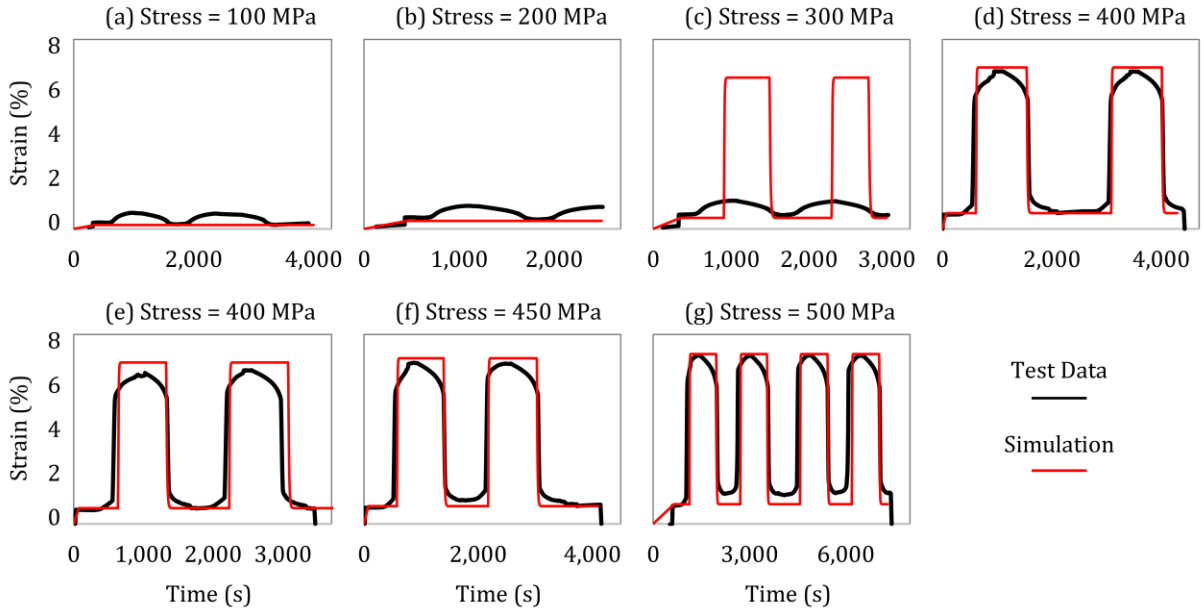


Figure 116: Temperature cycles under constant axial load

The experimental data in Figure 116a, b, and c show almost no actuation; this happens for two reasons: 1) the minimum temperature in the cycles was approximately -15 C. This is, by our calculations, well above the martensite start temperature (almost -70 C) where the transformation would occur at essentially zero stress. 2) Local stress fields can also be present which can trap austenite layers, thus requiring higher stresses or lower temperatures to spur actuation. We see that the actuation begins at 400 MPa (d) because temperature cycles are sufficient to push the forward transformation stress below this value. The simulations are off by one temperature increment and show actuation beginning at 300 MPa. Once the actuation starts, the model agrees well with the test data. The magnitudes of the maximum strain are roughly equal for all stress values and this is because the layers are transforming to the same set of variants (energy well) each time. The experimental curves are rounded as the strain approaches the maximum value whereas the simulations have sharp corners. This is due to the single crystal nature of the model, which is a limiting case idealization. Even though the specimens are single crystal-type specimens, mesoscopic field effects are still present to some degree which make the responses smoother. Global localization effects could also be playing a role in the shapes of the curves.

Figure 117 show simulations of the fixed-strain temperature cycling tests. In 117a-c, the loading was paused during the forward transformation then the temperature cycle was applied where as in d-f the loading was pasued during the reverse stress plateau. This can be thought of as a type of actuation test where the specimens are not allowed to contract. Instead of producing changes in strain during thermal cycles, changes in stress are observed.

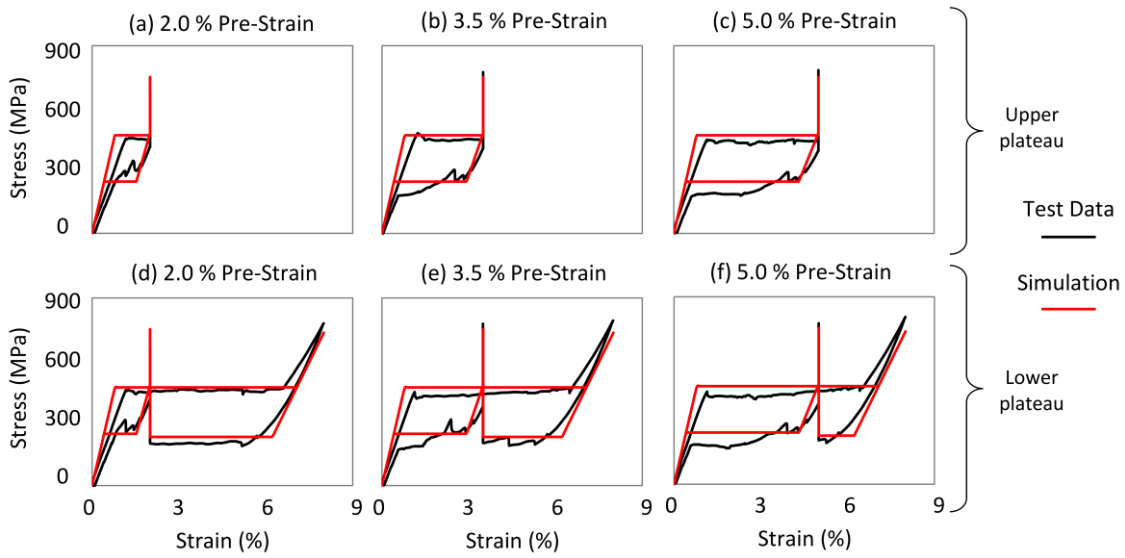


Figure 117: One temperature cycle at a fixed, applied displacement

The simulations here agree with the experiments for all cases. The experimental curves have irregular curves during unloading, which happened after the temperature cycling. The simulation curves are smooth with sharp corners. When the specimen is heated, the stress-strain curve is perfectly vertical. The simulations capture this and they also, for the most part, capture the proper height. When the specimen is cooled, the stress-strain response travels down the vertical segment till it reaches the top hysteresis loop, then it assumes a slope somewhere between the austenite and martensite slopes. The simulations also capture this effect.

Figures 114-117 show that the model is capable of uniaxial tension behavior including a number of effects, pseudoplasticity, superelasticity, temperature-dependent hysteresis, and the shape memory effect (actuation). Next we proceed to the combined tension-torsion experiments where we demomnstrate the model's multaxial capabilities and limitations.

6.7.2 Combined Tension-Torsion Cases

The same formulation of the model and the same set of parameters were used to simulation the tension-torsion and the uniaxial tension experiments. The model converts a set of transformation stress input into a critical free energy criterion that can then be applied to any combination of stresses and temperature. It should be noted however, that different values were assigned to the transformation stress parameters for the tension-torsion simulations. Figure 102 shows the chemical free energy calculated once using the tension data for the inputs and once using the shear data as inputs. The two curves are similar but there is also a nontrivial separation between them. This means there is some type of imperfection in the process of measuring the shear stresses or some limitation in the theory. Ideally, either type of transformation stress could be used to derive the same free energy. The same situation is true for the critical free energy, which governs transformation in the model. The values are again similar enough that either could be used, but have enough of a difference that the model's agreement would be affected in one set of simulations. We choose to focus on the overall behavior of the model and it's capture of the various experimental effects. Using a second set of input parameters here allows us to accentuate where the model agrees with the data and also where it falls short. It should also be noted that this set of parameter values was used across all six of the different tension-torsion data sets. The 70-MPa-axial stress set of tension-torsion tests was used to populate the transformation stress parameters, as shown in Figure ##. The values are summarized here in Table 8.

Parameter	Value	Parameter	Value
E_a	60 (GPa)	G_a	17 (GPa)
E_m	30 (GPa)	G_m	15 (GPa)
ε_R	5. 4 (%)	γ_R	6. 5 (%)
T_H	60 (C)	T_L	0 (C)
σ_{fwd}^H	70 (MPa)	σ_{fwd}^L	70 (MPa)
σ_{rev}^H	70 (MPa)	σ_{rev}^L	70 (MPa)
τ_{fwd}^H	435 (MPa)	τ_{fwd}^L	200 (MPa)
τ_{rev}^H	365 (MPa)	τ_{rev}^L	35 (MPa)
M	2 (GPa)		

Table 8: Material parameters used for the uniaxial tension and thermal cycling simulations

Figure 118 shows three temperatures from the 70 MPa axial stress set of tension-torsion simulations overlayed with the experimental data. We see on the left (a) that the model matches the torque-twist behavior well across the temperature range, which includes a transition from superelastic to pseudoplastic.

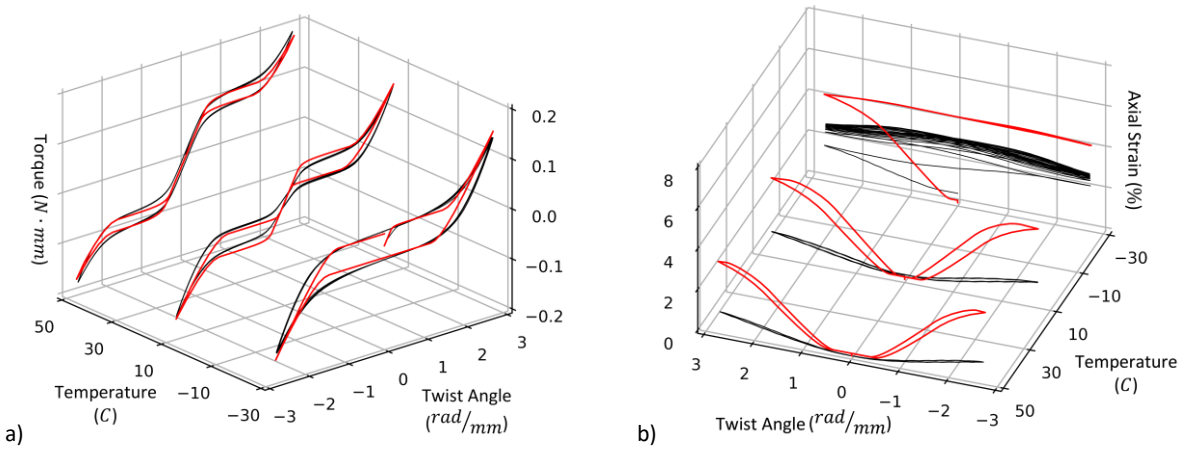


Figure 118: Selected tension-torsion tests, Axial stress = 70 MPa, Experiment (black) vs. simulation (red)

The plot on the right (b) shows the axial strain vs. twist angle at the same temperatures and we see that the model agreement here is poor. The model drastically overpredicts the axial strain and does not capture the large jump in strain between the two lower temperatures. Apart from the magnitude, the simulations do capture the both modes; the high temperature V-shape curves and the low-temperature mode where the axial strain is roughly constant. Even at only 70 MPa of axial stress, the two tensile wells dominate as the GFE landscape tilts back and forth in the shear direction. For the model to predict a small amount of axial strain, each shear transformation would need to cause a more balanced distribution between the tensile and compressive wells. Instead, all of the layers that become free from the austenite well choose the well on the tension side. This imbalance causes the macroscopic axial strain to be larger than that measured in the experiment.

The next set of tension-compression tests was under a 132 MPa axial stress. This set of simulations is shown in Figure 119. The torque-twist plots once again agree well with the experimental data (a) and the axial strain does not. Figures 110 and 111 show that the transformation shear stresses are very similar in the 70 MPa and 132 MPa data sets, so similar agreement to the previous plot is expected. The axial strain-twist plots in (b), however, are different from the 70 MPa set. The experimental axial strains have increased slightly at the higher temperatures

and increased by almost 2% at the lowest temperature (-25 C). Here, in the low-temperature case, the simulation matches the experimental curve reasonably well.

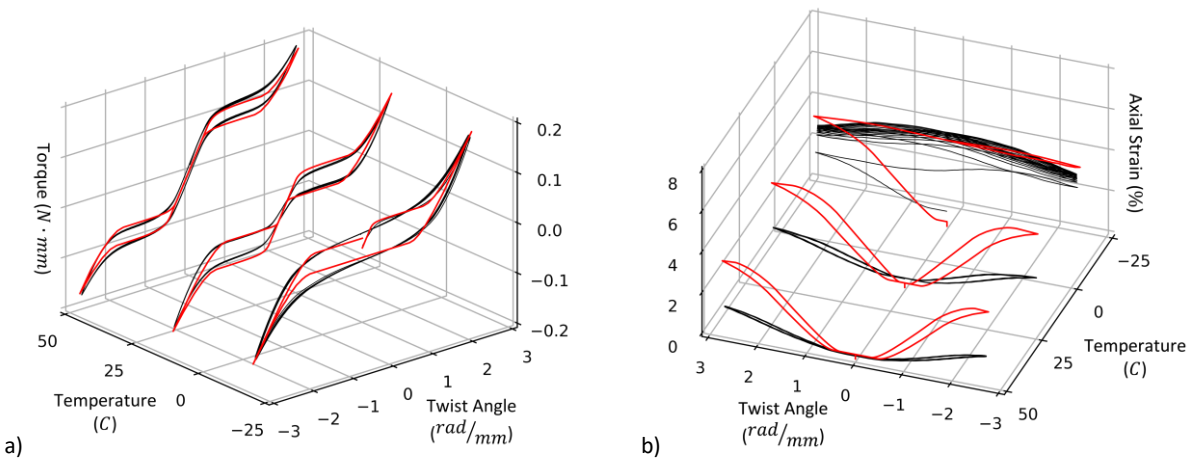


Figure 119: Selected tension-torsion tests, Axial stress = 132 MPa, Experiment (black) vs. simulation (red)

The simulation starts in the austenite phase where the axial strain is very low, then transitions to the martensite-to-martensite mode after the first twist cycle. The test specimen takes approximately two twist cycles to transition to the low-tempertaure mode. The model can switch modes as shown in the Figure but the type of ratcheting effect seen here, where it takes a few cycles to gradually transition is not included in the current theory. Another relevant observation is in the axial strain at zero-twist. In the high temperature cases, the zero-twist axial strain should be equal to the axial stress divided by the elastic modulus of the austenite phase. The same is true for the low temperature cases as well, for the axial strain before the twisting starts (because residual martensite was removed via heating before each test). For the high-temperature curve in Figure 119b, the experimental zero-twist axial strain agrees with the model which predicts the linear elastic value but this is not the case for the other two sets of curves. In the intermdiate and low-temperature curves, the zero-twist axial strain increases as the temperature gets lower. At lower levels of axial stress (70, 132, and 194 Mpa), this is an artifact of the polycrystalline-type of behavior under thermal cycling that was observed in Figure 116. Despite the fact that the specimens were heated between tests to remove residual martensite, the applied axial stress of 132 MPa in 119 results in a small but noticable amount of axial strain before the twist cycles ever begin. This effect is outside the scope of the current model, but it causes only a minor disagreement between the curves and is not one of the main mechanisms of interest in the tension-torsion tests.

Figure 120 shows the 194 MPa axial stress set of simulations where we see the same trends continue, but we can also see the axial stress begin to play more of a role.

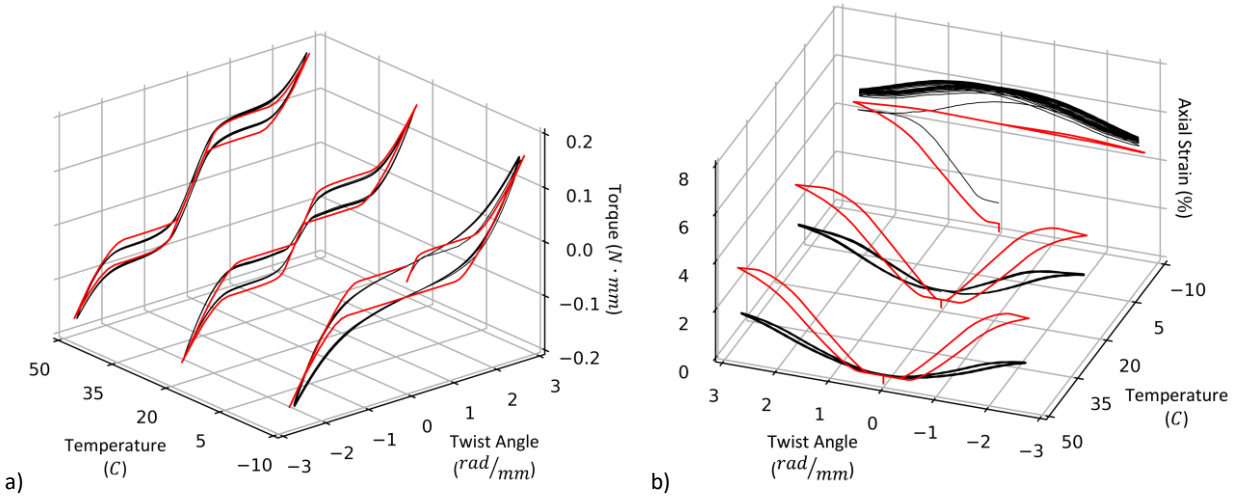


Figure 120: Selected tension-torsion tests, Axial stress = 194 MPa, Experiment (black) vs. simulation (red)

The first observation to be made is the temperature associated with the single hysteresis-loop curve in 120a, which is at -10 C. Recalling Figure 77a, the -10 C curve in the 70 MPa set of tension-torsion tests still has two somewhat distinct hysteresis loops. This means the the axial stress-alone has caused the -10 C response to shift from the double hysteresis-loop type to to the single-loop type. The simulation captures this effect naturally. The transformation stress and temperature inputs were based on the 70 MPa axial stress data set and did not have any information from the other axial stress responses. The 70 MPa parameters were used to generate the critical GFE functions and those functions sucessfully account for the coupling between axial stress, transformation torque, and temperature.

The axial strain, showin in 120b, shows the same relationship between the simulation and experiment data as the previous cases. The model continues to predict simular values for the axial strain while the test specimens show axial strains that are higher for each axial stress. The model does, however, capture some of the axial strain effects. Comparing the high-tempeature curves in 120b to the intermediate-temperature curves, we see that the high-temperature curves have thinner legs in their “V” shapes whereas the intermediate-temperature legs are thicker. The simulation V’s have thicker legs in general, but the capture the effect of widening with lower temperature. Also, like the previous case, the model agrees better with the test data at the lowest temperature. This is especially true at the maximum and minimum twist points. The experimental data has a slight inverted “V” shape which the model does not have. The right side of the simulation does, however, show a small hysteresis loop like the experiment does. There is no loop on the negative twist side only because the simulation was stopped after a single twist cycle.

We see also that the experimental zero-twist axial strain is much higher than predicted by the simulation. This is due to an unintended training effect that occurred over the course of the experiments. Each set of tests began with the highest temperature specimen and worked its way toward the lowest temperature. The specimens were heated between tests which should have removed any residual martensite and this means that the zero-twist axial strain strain should always be equal to the linear elastic prediction. This is what the model shows but not the experiments. The deviation in the zero-twist axial strain due to multiple phenomena: (1) polycrystalline effects which were discussed with Figure 119 and (2) ratcheting effects. In SMA materials, *ratcheting* refers to the cyclical accumulation of residual strain and stems from microscopic effects like dislocation slipping in the high strain-energy regions near phase boundaries (Kang (2013), Yu et al. (2014)). In the *RR* tension-torsion

experiments, especially at higher axial stress values, all three active phases exhibit ratcheting. Over the course of the experiments, the stress-strain origin of the austenite phase drifts due to cyclic loading and this causes an axial offset even under zero-twist and high temperature. The effect is not present, or noticable, at the two lower values of axial stress but it begins to play a role at 194 MPa. We also speculate that the deviation between the simualtion and the experiments at the minimum and maximum twist points is due to the two martensite wells accummulating additional residual strain.

Ratching effects are not currently included in the model. The offsets in the experimental curves are easy to see so it is not difficult to ignore them when comparing to the model's behavior, but this also presents an interesting opportunity for future work. The residual strains are currently material constants in the model and it would be a simple procedure to formualte them as exponential functions operating on some cycle-like measure. This type of technique is widely used in rubber modeling (the *Mullins* effect) and material damage modeling.

Figure 121 shows the next value of axial stress, 255 MPa. We see the the torque-twist curves (a) continue to have good agreement, that the single hysteresis loop now appears at +5 C, and that the experimental axial strains continue to increase. In 121a we see that the simulations continue, in general, to match the experiments in transformation torques across the temperature ranges tested. The single hysteresis loop is now occurring at a temperature that is 35 C higher than the initial data set from Figure 118 and the model captures it. In both the experimental and simulation curves at 5 C, we see an austenite-to-martensite plateau on the first cycle then a transition into the single loop configuration. The simulation over-predicts the torque for this initial transformation but matches the average transformation torque for the subsequent cycles.

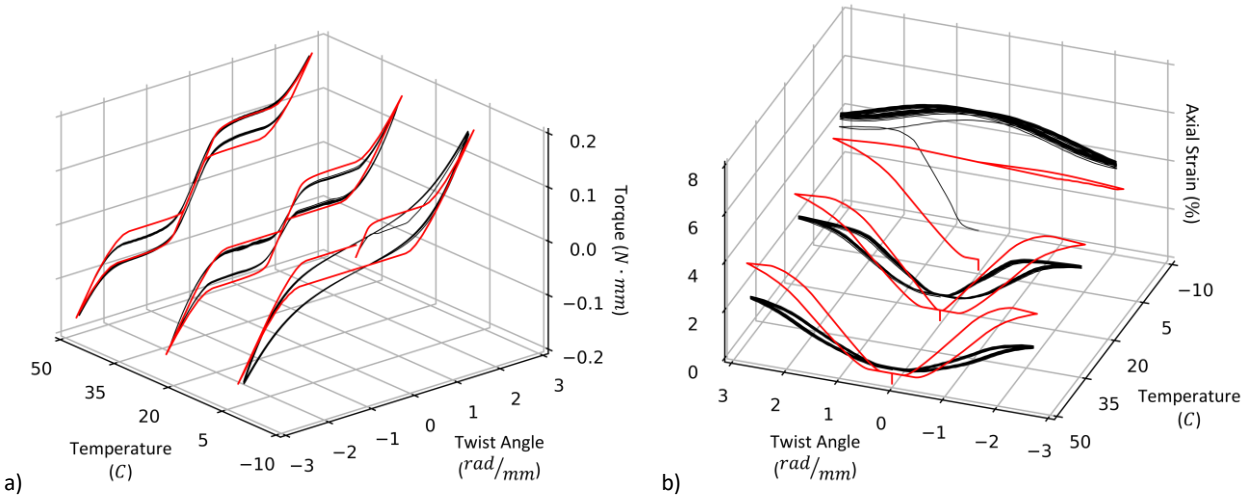


Figure 121: Selected tension-torsion tests, Axial stress = 255 MPa, Experiment (black) vs. simulation (red)

The low-temperature experimental curve in 121a also shows another trend that has been developing as the axial stress increases – the slanted, smooth shape of the hysteresis loop. The simulations exhibit a flat box-like hysteresis loop with rounded transitions into and out of the plateaus. This is also what the experimental response looked like in the 70 MPa axial stress case (Figure 118), but not here. The experimental, low-temperature hysteresis loop now has the shape of a very highly textured (polycrystalline) SMA material. The model does not capture this mode and we can only speculate about its physical explanation. We cannot assume that the material has become more polycrystalline because of observations in later tests. After austenite resetting, the higher-temperature response continues to have the same flat torque plateaus as the previous tests (see Figure 122). If the

specimen were becoming more polycrystalline as the testing progressed, the shape of both the low and high-temperature curves would change. In the experiments, only the low-temperature curves do. This change in low-temperature curve shape could be due to macroscopic field effects, but this would have to be a special case that only appeared at low temperatures and high levels of axial stress. It is possible that the elastic coupling term (M from the HFE function) becomes more pronounced at low temperatures. This would explain why the high-temperature axial strain-twist plots (121b) are flat in the martensite branches, yet the low temperature curves show almost a 1% deviation. A higher amount of coupling could also conceivably result in a more gradual transformation between the two martensite wells as is observed in the experiments. Currently, the model assumes the coupling term M is a constant but this would be a simple modification to the theory and is recommended for future examination.

The agreement between the torque-twist curves (121a), while still reasonable, is not as good as in the lower axial stress cases. Comparison of the previous simulations (as well as examination of Figures 77 and 78) shows that the hysteresis width at a given temperature also increases with axial stress. The model captures this effect nicely, especially considering that no axial stress-dependence was programmed a priori into the transformation shear stress functions. The model however, as shown here, slightly exaggerates the widening of the hysteresis loop at the highest temperatures. This could be improved with minor modifications to the transformation stress material parameters, but the values used were chosen to give the best overall agreement to the large set of curves.

Figure 121b shows the axial strain-twist plots where the experimental curves continue to drift upward. The disagreement in zero-twist axial strain is now very noticeable at both the intermediate and low temperatures. If we neglect the zero-twist strain offset and visually shift the experimental curves downward, we see that the model still over-predicts the higher-temperature axial strains and reasonably matches the low-temperature value. Figure 122 shows the final set of tension-torsion simulations which represent an axial stress of 317 MPa. The 379 MPa set of tests was excluded because this level of axial stress is close to the thermal actuation limit stress observed in Figure 75 and this introduces ambiguity between thermal and mechanically-motivated transformations.

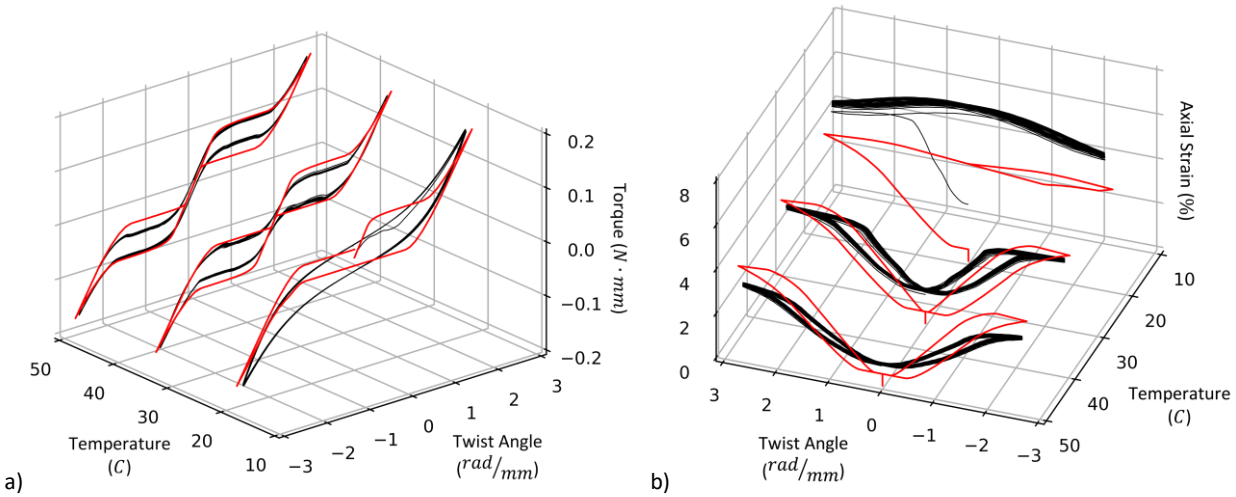


Figure 122: Selected tension-torsion tests, Axial stress = 317 MPa, Experiment (black) vs. simulation (red)

The axial stress-related trends continue here, and this set of simulations has similar agreement to the previous two data sets. The single-hysteresis loop now appears at 20 C and the hysteresis loop is very slanted and smooth. The width of the hysteresis loop predicted by the simulation is also narrow compared to the experiment. This is

due to an effect that occurs in the model when the austenite and martensite CFE's are similar. Through a certain range of temperatures, the relative CFE is near zero and the transformation is dominated by the stress. At high levels of axial stress, the hysteresis between the martensite wells is diminished because the wells are being pulled upward (as in Figure 105) where the critical GFE boundaries are closer together. The low-temperature hysteresis loops in the experiments are narrow, so it seems the model is predicting the correct effect; just doing so a little too aggressively. It should also be noted that this effect could also be related to the effects that cause the slanted and smooth hysteresis loops in the test data. In 122b, we remain in the same situation. If we shift the curves downward to "eliminate" the ratcheting effect, the model over-predicts the high and intermediate -temperature axial strains and matches the low-temperature curve. This concludes the simulations and the presentation of the 2-D model. We end the chapter with a recap of the model's successes, shortcommings, and opportunities for future development.

6.8 Chapter Review and Discussion

In this chapter we have presented an extension of the 1-D M-A-S model into multiple dimensions. The formulation was specified to represent an axial strain-shear strain (2-D) case, based on the *RR* experimental data set. The *RR* tests included uniaxial tension tests, uniaxial actuation tests, combined tension-torsion tests, and thermal cycles under combined loading. The tension tests showed typical single-crystal SMA behavior with the exception of having a hysteresis loop that became narrower with increasing temperature. The tension-torsion tests examined the torsional behavior wires across a wide range of temperatures and six different values of axial stress. The torque-twist responses exhibited high (double hysteresis loop) and low-temperature (single hysteresis loop) modes that were similar in characteristic to uniaxial tests. The torque-twist curves, however, had smooth transitions into and out of the transformation plateaus due to cross-sectional stress variation. The axial stress played a similar role to temperature in the torque-twist behavior in that it affected the amount of torque required for transformation which in-turn dictated whether the response was of the double-hysteresis loop type or the single-loop type. There were also two distinct modes in the axial strain-twist plots. At high temperatures the curves had a "V" shape where the axial strain was minimal at zero-twist. At low temperatures the curves assumed a much flatter shape (less oscillation in axial strain) and the V was inverted, where the maximum axial strain occurred at zero-twist. This behavior was consistent across all of the axial stresses, but the magnitudes of the axial strain increased as the axial stress did.

The 2-D model developed in this chapter is formulated by extending concepts from the 1-D model onto a 2-D, mesoscopic, multiwell free energy domain. We assume two sets of two martensite variants that are offset from one another on the normal strain and shear strain axes. Physically, this implies two sets of habit plane variant groups forming on opposing habit planes. Any transformation results in a combination of normal and shear strain for each variant, but the symmetric nature of the free energy means that either normal strain or shear strain can be canceled out if the right combination of variants form. Another benefit of using a single, multiwell, free energy to describe all of the variants is that the natural distortion of the free energy (from stress and temperature) provides the interaction effects between the normal and shear components. The use of a multiwell free energy and variant symmetry gives us a mechanism to produce the complex tension-torsion-temperature coupling that is:

- Simple (only four variants)
- Visualizable (balls jumping between wells)
- Can be traced directly back to physical effects (lattice invariant shears)

The kinetics model, which relates the mesoscopic free energy to macroscopic stress-strain-temperature behavior, was also reformulated to accommodate the 2-D model. The new kinetics equations were built such that the energy barriers between the wells, which are prohibitively complex in multiple dimensions, did not have to be defined or even approximated. This was based on a the 1-D *Inflection Point* model but extended to a 2-D energy landscape where the inflection points become curves. Each energy well has a critical boundary where it shifts from convex to concave (i.e. where it becomes unstable and layers will exit). We constructed these lines by assuming that each signifies a constant value Gibbs free energy, which we call the critical free energy. The critical free energy is temperature dependent parameter, which is calculated from a set of experimentally measured transformation stresses and temperatures. There is one critical energy to exit the austenite well and another to exit any of the martensite wells. In the 1-D model, each ΔG value was associated with a specific transformation, but this is not the case in the 2-D model. Each phase now has a single ΔG value that simply signifies whether that phase is stable (layers will remain) or unstable (layers will exit). Two additional terms had to be added to the kinetics equations to dictate where the layers go when they exit a well. 1) A driving force term was added which encourages layers to favor lower energy wells. 2) We also included a second ΔG term to represent he destination well. This term ensures that layers will not flow to an unstable well, even if the driving force (i.e. stress) pushes them in that direction.

To test the model in uniaxial tension, we evaluated a single instance of the model and assumed zero shear strain at all times. For the combined loading tests, we assumed a linear strain profile through the cross section of the wire and evaluated the stress at 12 radially distributed points. The stress from each point was used to calculate the total torque and the average was used for the total wire axial strain. The model showed excellent agreement in the uniaxial cases and mixed, but promising, results for the tension-torsion simulations. The main effects from the experiments and that we were attempting to capture are listed here along with an evaluation of the success.

Success?	Effect	Comment
✓	Superelastic behavior	Uniaxial tension, shear, and combined loading
✓	Pseudoplastic behavior	Uniaxial tension, shear, and combined loading
✓	Temperature-dependent hysteresis	Uniaxial tension, shear, and combined loading
✓	Thermal actuation	Uniaxial tension
✓	Axial stress lowers the transformation torque	Both forward and reverse transformations
✓	Axial stress widens the hysteresis width in torsion	Effect is slightly exaggerated but reasonable
✓	Distinct high and low-temperature modes of axial strain during torsion cycles	Model shows both modes and transitions at the proper the stress-temperature combinations
✗	Smooth/slanted shape of the low-temperature torsional hysteresis loops	The cause of this is not clear and model does not reflect it
✗	Magnitude of axial strain during torsion cycles	The model over-predicts the high-temperature axial strains for all axial stress cases
✗	Axial strain during torsion cycles increases with axial stress	The model has this capability, but the parameters assigned and conditions imposed do not produce it

The good agreement in the uniaxial cases is to be expected because the 1-D model showed similar accuracy. This repetition of results is significant because it confirms that the 2-D model can function in a uniaxial mode without any loss in performance. The first three items in the list, it should be noted, apply to uniaxial tension as well as shear and combined loading. Both the tension test stress-strain curves and the torque-twist curves from the combined loading tests showed all three effects (superelasticity, pseudoplasticity, and temperature-dependent hysteresis). The model captured these effects accurately in both scenarios. Similar curves are also generated for a pure shear (zero axial stress) case, but this case was not presented because it was not part of the experimental data

set. The next effect in the list, thermal actuation, was also only simulated in uniaxial tension. The model agreed vert well with the experiment except that the threshold stress, where actuation begins, was slightly over-predicted. The constant torque tests would have been a platform to test the model in thermal actuation under combined loading, but this was not necessary. The critical free energy is the modeling feature that causes both temperature-dependent transformation stress and thermal actuation. The fact that the transformation torques decreased with temperature to eventually form a single hysteresis loop means that the thermal actuation mechanism is functioning in shear and combined loading.

The next two items in the list involve the effects of axial stress on the transformation shear stress; these are the main demonstrations of coupling between the stress components. As the axial stress increases, the forward and reverse transformation torque decrease. The torque plateaus in the simulations match the experimental curves reasonably well for all values of axial stress. This is a natural consequence of using a free energy-based transformation criterion; the axial stress pulls the phases closer to the stability boundaries which means that less shear stress (torque) is required for transformation to occur. The axial stress also causes the torsional hysteresis width to increase in both the tests and the simulations. The free energy distortion affects the forward and reverse stability boundaries differently and therefore the difference in transformation stress changes with the axial-shear stress ratio. The next item refers to the fact that the low-temperature hysteresis loop exhibited a notable change in shape as the axial stress increased. This effect was not captured by the model and, without the addition of some type of inhomogeneity or global field effects, the model has no mechanism to represent this behavior. The physical cause for this is not clear from the test data, but the next logical step would be to implement the current model in a full 3-D field solution and examine if stress, strain, or phase fields in the wire could be the culprit.

The final two effects involve the axial strain in the wire as the torsion cycles are imposed. The axial strain vs. twist plots showed an intriguing set of behaviors that, to the author's knowledge, had not previously been demonstrated in ferroelastic SMA experiments. It should be noted, however, that some aspects of the axial strain-twist responses were similar to those observed in ferroelectric materials (Kim and Seelecke (2007)) and this fact motivated the initial concept for the model developed in this chapter. In the end, the model was able to reproduce the two major modes that were observed in the experiments but failed to capture the magnitudes of the axial strain. The model is capable of producing the two axial strain-twist modes over the full range of strains observed in the tests, but the this would require a high level of symmetry to be maintained in the combined loading cases. Instead, the axial stresses cause the layer population to be highly biased toward the tensile side of the free energy. This means that the model predicts roughly the same values of axial transformation strain for all axial stress cases whereas the experimental values increase almost linearly. Inhomogeneity effects, which were not included in the 2-D model, could possibly improve the behavior in this case but it is not likely that they would completely resolve the disparity. It more likely that global field effects are again at play and this also promotes the notion of 3-D field implementation of the current model for future work.

We conclude this dissertation with a set of appendices including assorted material from the previous chapters. This is material that was not required to accomplish the fundamental goals of each chapter but provides important background information or convenient references.

7 References

[1] Müller, I. , Wilmanski, K. "Memory alloys—phenomenology and Ersatzmodel. " Volume 4. 1981, p 495.
[2] Achenbach, M. , Müller, I. "Simulation of Material Behaviour of Alloys with Shape Memory. " Volume 37. 6, p 573-585.
[3] Achenbach, M. "A Model For An Alloy With Shape Memory. " Volume 5. 4, p 371-395.
[4] Müller, I. , Seelecke, S. "Thermodynamic Aspects of Shape Memory Alloys. " Volume 34. 12, p 1307-1355.
[5] Sittner, P. "Roundrobin SMA modeling. " ESOMAT 2009-8th European Symposium on Martensitic Transformations, 2009.
[6] Ziolkowski, A. "Pseudoelasticity of shape memory alloys: theory and experimental studies. " Butterworth-Heinemann, 2015.
[7] Barbarino, S. "A review on shape memory alloys with applications to morphing aircraft. " Volume 23. 6, p 63001.
[8] Lagoudas, D. C. "Shape memory alloys, Part II: Modeling of polycrystals. " Volume 38. 5, p 430-462.
[9] Patoor, E. "Shape memory alloys, Part I: General properties and modeling of single crystals. " Volume 38. 5, p 391-429.
[10] Paiva, A. , Savi, M. A. "An Overview of Constitutive Models for Shape Memory Alloys. " Volume 2006. 1, p 1-30.
[11] Boyd, J. G. , Lagoudas, D. C. "A Thermodynamical Constitutive Model for Shape Memory Materials. Part I. the Monolithic Shape Memory Alloy. " Volume 12. 6, p 805-842.
[12] Bo, Z. , Lagoudas, D. C. "Thermomechanical modeling of polycrystalline SMAs under cyclic loading, Part I: theoretical derivations. " Volume 37. 9, p 1089-1140.
[13] Bo, Z. , Lagoudas, D. C. "Thermomechanical modeling of polycrystalline SMAs under cyclic loading, Part III: evolution of plastic strains and twoway shape memory effect. " Volume 37. 9, p 1175-1203.
[14] Lagoudas, D. C. , Bo, Z. "Thermomechanical modeling of polycrystalline SMAs under cyclic loading, Part II: material characterization and experimental results for a stable transformation cycle. " Volume 37. 9, p 1141-1173.
[15] Bernardini, D. , Pence, T. J. "Models for one-variant shape memory materials based on dissipation functions. " Volume 37. 8, p 1299-1317.
[16] Müller, C. , Bruhns, O. T. "A thermodynamic finite-strain model for pseudoelastic shape memory alloys. " Volume 22. 9, p 1658-1682.
[17] Zaki, W. "An efficient implementation for a model of martensite reorientation in martensitic shape memory alloys under multiaxial nonproportional loading. " Volume 37. 2012, p 72-94.
[18] Morin, C. , Moumni, Z. , Zaki, W. "Thermomechanical coupling in shape memory alloys under cyclic loadings: Experimental analysis and constitutive modeling. " Volume 27. 12, p 1959-1980.
[19] Zaki, W. , Moumni, Z. "A three-dimensional model of the thermomechanical behavior of shape memory alloys. " Volume 55. 11, p 2455-2490.
[20] Thamburaja, P. "A finite-deformation-based phenomenological theory for shape-memory alloys. " Volume 26. 8, p 1195-1219.
[21] Armattoe, K. M. , Haboussi, M. , Zineb, T. B. "A 2D finite element based on a nonlocal constitutive model describing localization and propagation of phase transformation in shape memory alloy thin structures. " Volume 51. 6, p 1208-1220.
[22] Chemisky, Y. "Constitutive model for shape memory alloys including phase transformation, martensitic reorientation and twins accommodation. " Volume 43. 7, p 361-376.
[23] Saleeb, A. F. , Padula, S. A. , Kumar, A. "A multi-axial, multimechanism based constitutive model for the comprehensive representation of the evolutionary response of SMAs under general thermomechanical loading conditions. " Volume 27. 5, p 655-687.
[24] Xiao, H. "An explicit, straightforward approach to modeling SMA pseudoelastic hysteresis. " Volume 53. 2014, p 228-240.
[25] Peultier, B. , Zineb, T. B. , Patoor, E. "Macroscopic constitutive law of shape memory alloy thermomechanical behaviour. Application to structure computation by FEM. " Volume 38. 5, p 510-524.
[26] Zhou, B. "A macroscopic constitutive model of shape memory alloy considering plasticity. " Volume 48. 2012, p 71-81.
[27] Auricchio, F. , Sacco, E. "A temperature-dependent beam for shape-memory alloys: constitutive modeling, finite-element implementation and numerical simulations. " Volume 174. 1, p 171-190.
[28] Masud, A. , Panahandeh, M. , Aurrichio, F. "A finite-strain finite element model for pseudoelastic behavior of shape memory alloys. " Volume 148. 1, p 23-37.
[29] Auricchio, F. , Taylor, R. L. , Lubliner, J. "Shape-memory alloys: macromodelling and numerical simulation of the superelastic behavior. " Volume 146. 3, p 281-312.

[30] Auricchio, F. , Lubliner, J. "A Uniaxial Model for Shape Memory Alloys. " Volume 34. 27, p 3601-3618.
[31] Auricchio, F. , Sacco, E. "A Superelastic Shape-Memory-Alloy Beam Model. " Volume 8. 6, p 489-501.
[32] Lubliner, J. , Auricchio, F. "Generalized plasticity and shape memory alloys. " Volume 33. 7, p 991-1003.
[33] Arghavani, J. "A 3-D phenomenological constitutive model for shape memory alloys under multiaxial loadings. " Volume 26. 7, p 976-991.
[34] Auricchio, F. , Marfia, S. , Sacco, E. "Modelling of SMA materials: Training and two way memory effects. " Volume 81. 24, p 2301-2317.
[35] Auricchio, F. , Sacco, E. "Thermo-mechanical modeling of a superelastic shape memory-alloy wire under cyclic stretching-bending loadings. " Volume 38. 34, p 6123-6145.
[36] Bodaghi, M. "A phenomenological SMA model for combined axial–torsional proportional/non-proportional loading conditions. " Volume 587. -, p 42730.
[37] Stebner, A. P. , Brinson, L. C. "Explicit finite element implementation of an improved three dimensional constitutive model for shape memory alloys. " Volume 257. 2013, p 17-35.
[38] Christ, D. , Reese, S. "A finite element model for shape memory alloys considering thermomechanical couplings at large strains. " Volume 46. 20, p 3694-3709.
[39] Reese, S. , Christ, D. "Finite deformation pseudo-elasticity of shape memory alloys – Constitutive modelling and finite element implementation. " Volume 24. 3, p 455-482.
[40] Coleman, B. D. , Noll, W. "The thermodynamics of elastic materials with heat conduction and viscosity. " Volume 13. 1, p 167-178.
[41] Coleman, B. D. , Gurtin, M. E. "Thermodynamics with Internal State Variables. " Volume 47. 2, p 597-613.
[42] Patoor, E. , Eberhardt, A. , Berveiller, M. "Thermomechanical Behavior of Shape Memory Alloys. " European Symposium on Martensitic Transformations, 1989.
[43] Patoor, E. , Eberhardt, A. , Berveiller, M. "Micromechanical Modelling of Superelasticity in Shape Memory Alloys. " Volume IV. 6, p C1-277.
[44] Entemeyer, D. "Strain rate sensitivity in superelasticity. " Volume 16. 10, p 1269-1288.
[45] Huang, M. , Brinson, L. C. "A Multivariant model for single crystal shape memory alloy behavior. " Volume 46. 8, p 1379-1409.
[46] Yu, C. "Micromechanical constitutive model considering plasticity for super-elastic NiTi shape memory alloy. " Volume 56. 2012, p 1-5.
[47] Wang, Y. F. , Yue, Z. F. , Wang, J. "Experimental and numerical study of the superelastic behaviour on NiTi thin-walled tube under biaxial loading. " Volume 40. 2, p 246-254.
[48] Wang, X. M. , Xu, B. X. , Yue, Z. F. "Micromechanical modelling of the effect of plastic deformation on the mechanical behaviour in pseudoelastic shape memory alloys. " Volume 24. 8, p 1307-1332.
[49] Richards, A. W. , Lebensohn, R. A. , Bhattacharya, K. "Interplay of martensitic phase transformation and plastic slip in polycrystals. " Volume 61. 12, p 4384-4397.
[50] James, R. , Kinderlehrer, D. "Theory of diffusionless phase transitions. " Lecture Notes in Physics, 1989.
[51] Ball, J. , Richard, M. , James, D. "Proposed Experimental Tests of a Theory of Fine Microstructure and the Two-Well Problem. " Volume 338. 1650, p 389-450.
[52] Abeyaratne, R. , Knowles, J. K. "On the dissipative response due to discontinuous strains in bars of unstable elastic material. " Volume 24. 10, p 1021-1044.
[53] Abeyaratne, R. , Knowles, J. K. "Kinetic Relations And The Propagation Of Phase Boundaries In Solids. " Volume 114. 2, p 119-154.
[54] Abeyaratne, R. , Knowles, J. K. "On the driving traction acting on a surface of strain discontinuity in a continuum. " Volume 38. 3, p 345-360.
[55] Abeyaratne, R. , Knowles, J. K. "Implications of Viscosity and Strain-Gradient Effects for the Kinetics of Propagating Phase Boundaries in Solids. " Volume 51. 5, p 1205-1221.
[56] Leo, P. H. , Shield, T. W. , Bruno, O. P. "Transient heat transfer effects on the pseudoelastic behavior of shape-memory wires. " Volume 41. 8, p 2477-2485.
[57] Shaw, J. A. , Kyriakides, S. "Thermomechanical Aspects of NiTi. " Volume 43. 8, p 1243-1281.
[58] Shaw, J. A. "A thermomechanical model for a 1-D shape memory alloy wire with propagating instabilities. " Volume 39. 5, p 1275-1305.
[59] Shaw, J. A. , Churchill, C. B. , Iadicola, M. A. "Tips And Tricks For Characterizing Shape Memory Alloy Wire: Part 1 - Scanning Calorimetry And Basic Phenomena. " Volume 32. 5, p 55-62.
[60] Churchill, C. B. , Shaw, J. A. , Iadicola, M. A. "Tips And Tricks For Characterizing Shape Memory Alloy Wire: Part 2 — Fundamental Isothermal Responses. " Volume 33. 1, p 51-62.
[61] Churchill, C. B. , Shaw, J. A. , Iadicola, M. A. "Tips And Tricks For Characterizing Shape Memory Alloy Wire: Part 3 — Part 3-Localization And Propagation Phenomena. " Volume 33. 5, p 70-78.
[62] Churchill, C. B. , Shaw, J. A. , Iadicola, M. A. "Tips And Tricks For Characterizing Shape Memory Alloy Wire: Part 4 — Thermo-Mechanical Coupling. " Volume 34. 2, p 63-80.

[63] Dai, H. H. , Wang, J. "Instabilities induced by phase transformation fronts coalescence during the phase transitions in a thin SMA layer: Mechanism and analytical descriptions. " Volume 48. 11, p 1146-1163.
[64] Dhote, R. P. , Melnik, V. N. , Zu, J. "Dynamic thermo-mechanical coupling and size effects in finite shape memory alloy nanostructures. " Volume 63. -, p 105-117.
[65] Dhote, R. P. "Hysteresis phenomena in shape memory alloys by non-isothermal Ginzburg–Landau models. " Volume 18. 9, p 2549-2561.
[66] Song, Z. , Dai, H. H. , Sun, Q. P. "Propagation stresses in phase transitions of an SMA wire: New analytical formulas based on an internal-variable model. " Volume 42. 2013, p 101-119.
[67] Rogers, C. "Balloon-Artery Interactions During Stent Placement : A Finite Element Analysis Approach to Pressure, Compliance, and Stent Design as Contributors to Vascular Injury. " Volume 84. 4, p 378-383.
[68] Duerig, T. W., Tolomeo, D. E., Wholey, W. "An overview of superelastic stent design." Volume 9.42433, p 235-246.
[69] Migliavacca, F. "Mechanical behavior of coronary stents investigated through the finite element method. " Volume 35. 6, p 803-811.
[70] Gong, X. Y. "Finite Element Analysis and Experimental Evaluation of Superelastic Nitinol Stent. " SMST–2003 Proc. International Conference on Shape Memory and Superelastic Technologies, 2003.
[71] Stoeckel, D. , Pelton, A. , Duerig, T. "Self-expanding nitinol stents: material and design considerations. " Volume 14. 2, p 292-301.
[72] Kleinstreuer, C. "Computational mechanics of Nitinol stent grafts. " Volume 41. 11, p 2370-2378.
[73] Ballew, W. , Seelecke, S. "A Comparison of FE Beam and Continuum Elements for Typical Nitinol Stent Geometries. " Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring, 2009.
[74] Zaki, W. "An efficient implementation for a model of martensite reorientation in martensitic shape memory alloys under multiaxial nonproportional loading. " Volume 37. 2012, p 72-94.
[75] Hartl, D. J. , Lagoudas, D. C. "Aerospace Applications of Shape Memory Alloys. " Volume 221. 4, p 535-552.
[76] Barbarino, S. "A review on shape memory alloys with applications to morphing aircraft. " Volume 23. 6, p 63001.
[77] Bouvet, C. , Calloch, S. , Lexcellent, C. "A phenomenological model for pseudoelasticity of shape memory alloys under multiaxial proportional and nonproportional loadings. " Volume 23. 1, p 37-61.
[78] Arghavani, J. "A 3-D phenomenological constitutive model for shape memory alloys under multiaxial loadings. " Volume 26. 7, p 976-991.
[79] Saleeb, A. F. , Padula, S. A. , Kumar, A. "A multi-axial, multimechanism based constitutive model for the comprehensive representation of the evolutionary response of SMAs under general thermomechanical loading conditions. " Volume 27. 5, p 655-687.
[80] Sedlak, P. "Thermomechanical model for NiTi-based shape memory alloys including R-phase and material anisotropy under multi-axial loadings. " Volume 39. 2012, p 132-151.
[81] Gibeau, E. , Lexcellent, C. , Boubakar, L. "Application with the experimental results of the Roundrobin of a 3D modelling SMA based on the based on the phase transformation surface identification under proportional loading and anisothermal conditions. " ESOMAT 2009-8th European Symposium on Martensitic Transformations, 2009.
[82] Rio, G. , Favier, D. , Liu, Y. "Elastohysteresis model implemented in the finite element sofware HEREZH++. " ESOMAT 2009-8th European Symposium on Martensitic Transformations, 2009.
[83] Chemisky, Y. "Numerical tool for Shape Memory Alloys structures simulations including twinning effects. " ESOMAT 2009-8th European Symposium on Martensitic Transformations, 2009.
[84] Hartl, D. , Oehler, S. , Lagoudas, D. C. "Constitutive Modeling of Phase Transformation and Plastic Yield in SMAs: Application to the S3T-RoundRobin. " ESOMAT 2009-8th European Symposium on Martensitic Transformations, 2009.
[85] Frost, M. , Sedlak, P. "Modeling of Two-dimensional Thermomechanical Loading of NiTi Wires. " ESOMAT 2009-8th European Symposium on Martensitic Transformations, 2009.
[86] Sedlak, P. "Thermomechanical model for NiTi-based shape memory alloys including R-phase and material anisotropy under multi-axial loadings. " Volume 39. 2012, p 132-151.
[87] Bodaghi, M. "A phenomenological SMA model for combined axial–torsional proportional/non-proportional loading conditions. " Volume 587. -, p 42730.
[88] Auricchio, F. , Morganti, S. , Reali, A. "SMA numerical modeling versus experimental results. " ESOMAT 2009-8th European Symposium on Martensitic Transformations, 2009.
[89] Auricchio, F. "Theoretical and numerical modeling of shape memory alloys accounting for multiple phase transformations and martensite reorientation. " Volume 59. -, p 30-54.
[90] Chatziathanasiou, D. "Modeling of coupled phase transformation and reorientation in shape memory alloys under non-proportional thermomechanical loading. " Volume 82. -, p 192-224.
[91] Müller, I. "A Model for a body with shape memory. " Volume 70. 1, p 61-77.

[92] Falk, F. "Model Free Energy, Mechanics, and Thermodynamics of Shape Memory Alloys. " Volume 28. 12, p 1773-1780.
[93] Falk, F. "Ginzburg-Landau Theory of Static Domain Walls in Shape-Memory Alloys. " Volume 51. 2, p 177-185.
[94] Seelecke, S. "Modeling the dynamic behavior of shape memory alloys. " Volume 37. 8, p 1363-1374.
[95] Seelecke, S. , Müller, I. "Shape memory alloy actuators in smart structures-modeling and simulation. " Volume 57. 1, p 23-46.
[96] Heintze, O. , Seelecke, S. "A coupled thermomechanical model for shape memory alloys—From single crystal to polycrystal. " Volume 481. -, p 389-394.
[97] Rizzello, G. , Mandolino, M. A. , Schmidt, M. , Naso, D. , Seelecke, S. "An accurate dynamic model for polycrystalline shape memory alloy wire actuators and sensors. " Volume 28. 2, p 025020.
[98] Furst, S. J. , Crews, J. H. , Seelecke, S. "Numerical and experimental analysis of inhomogeneities in SMA wires induced by thermal boundary conditions. " Volume 24. 4-6, p 485-504.
[99] Schmidt, M. "Experimental Methods for Investigation of Shape Memory Based Elastocaloric Cooling Processes and Model Validation. " Volume 111. 2016, p e53626-e53626.
[100] Ullrich, J. "Experimental Investigation and Numerical Simulation of the Mechanical and Thermal Behavior of a Superelastic Shape Memory Alloy Beam During Bending. " ASME 2014 Conference on Smart Materials, Adaptive Structures and Intelligent Systems, 2014.
[101] Welsch, F. "Numerical simulation and experimental investigation of the elastocaloric cooling effect in sputter-deposited TiNiCuCo thin films. " Volume 30. 1, p 53-68.
[102] Kim, S. J. , Seelecke, S. "A rate-dependent three-dimensional free energy model for ferroelectric single crystals. " Volume 44. 3, p 1196-1209.
[103] Krevet, B. "Magnetization-and strain-dependent free energy model for FEM simulation of magnetic shape memory alloys. " Volume 158. 1, p 205-211.
[104] Morrison, P. "A free energy model for magneto-mechanically coupled NiMnGa single crystals. " Volume 158. 1, p 213-220.
[105] Morrison, P. "Development of a meso-scale thermo-magneto-mechanical free energy model for NiMnGa. " The 15th International Symposium on: Smart Structures and Materials & Nondestructive Evaluation and Health Monitoring. International Society for Optics and Photonics, 2008.
[106] Seelecke, S. "A rate-dependent two-dimensional free energy model for ferroelectric single crystals. " Volume 17. 4, p 337-350.
[107] Smith, R. C. "A Free Energy Model for Hysteresis in Ferroelectric Materials. " Volume 14. 11, p 719-739.
[108] Smith, R. C. "A unified framework for modeling hysteresis in ferroic materials. " Volume 54. 1, p 46-85.
[109] Ballew, W. , Seelecke, S. "Mesoscopic free energy as a framework for modeling shape memory alloys. " Volume 30. 13, p 1969-2012.
[110] Chang, B. C. , Shaw, J. A. , Iadicola, M. A. "Thermodynamics of shape memory alloy wire: modeling, experiments, and application. " Volume 18. 42371, p 83-118.
[111] Falk, F. "Model Free Energy, Mechanics, and Thermodynamics of Shape Memory Alloys. " Volume 28. 12, p 1773-1780.
[112] Falk, F. "Ginzburg-Landau Theory of Static Domain Walls in Shape-Memory Alloys. " Volume 51. 2, p 177-185.
[113] Ortin, J. "Calorimetric measurements on the β⇌ γ'and β⇌ β'martensitic transformations in a cu-al-ni single crystal subjected to uniaxial tensile stress. " Volume 65. 2, p 461-475.
[114] Wollants, P. , Roos, J. R. , Delaey, L. "Thermally-and stress-induced thermoelastic martensitic transformations in the reference frame of equilibrium thermodynamics. " Volume 37. 3, p 227-288.
[115] Iadicola, M. A. , Shaw, J. A. "Rate and thermal sensitivities of unstable transformation behavior in a shape memory alloy. " Volume 20. 4, p 577-605.
[116] Schmidt, M. , Schütze, A. , Seelecke, S. "Scientific test setup for investigation of shape memory alloy based elastocaloric cooling processes. " Volume 54. 2015, p 88-97.
[117] Schmidt, M. "Thermal stabilization of NiTiCuV shape memory alloys: observations during elastocaloric training. " Volume 1. 2, p 132-141.
[118] Schmidt, M. "Elastocaloric cooling: from fundamental thermodynamics to solid state air conditioning. " Volume 22. 5, p 475-488.
[119] Schmidt, M. , Schütze, A. , Seelecke, S. "Elastocaloric cooling processes: The influence of material strain and strain rate on efficiency and temperature span. " Volume 4. 6, p 064107.
[120] Holzapfel, G. , A. "Nonlinear solid mechanics. " Chichester: Wiley, 2000.
[121] Liu, Y. "Asymmetry Of Stress-Strain Curves Under Tension And Compression For NiTi Shape Memory Alloys. " Volume 46. 12, p 4325-4338.
[122] Sittner, P. "In situ neutron diffraction studies of martensitic transformations in NiTi polycrystals under tension and compression stress. " Volume 378. 1, p 97-104.

[123] Reedlunn, B. "Tension, compression, and bending of superelastic shape memory alloy tubes. " Volume 63. 2014, p 506-537.
[124] Xiao, H. "An explicit, straightforward approach to modeling SMA pseudoelastic hysteresis. " Volume 53. 2014, p 228-240.
[125] Nascimento, M. M. S. F. "A mathematical model for the strain–temperature hysteresis of shape memory alloy actuators. " Volume 30. 3, p 551-556.
[126] Hamilton, R. F. "Stress dependence of the hysteresis in single crystal NiTi alloys. " Volume 52. 11, p 3383-3402.
[127] Ortin, J. "Calorimetric measurements on the β⇌ γ'and β⇌ β'martensitic transformations in a cu-al-ni single crystal subjected to uniaxial tensile stress. " Volume 65. 2, p 461-475.
[128] Fu, S. , Huo, Y. , Müller, I. "Thermodynamics of pseudoelasticity—an analytical approach. " Volume 99. 1-4, p 1-19.
[129] Otsuka, K. , Ren, X. "Physical metallurgy of Ti–Ni-based shape memory alloys. " Volume 50. 5, p 511-678.
[130] Gao, X. , Huang, M. , Brinson, L. C. "A multivariant micromechanical model for SMAs Part 1. Crystallographic issues for single crystal model. " Volume 16. 10, p 1345-1369.
[131] Kang, G. "Advances in transformation ratcheting and ratcheting-fatigue interaction of NiTi shape memory alloy. " Volume 26. 3, p 221-236.
[132] Yu, C. , Kang, G. , Kan, Q. "Crystal plasticity based constitutive model of NiTi shape memory alloy considering different mechanisms of inelastic deformation. " Volume 54. 2014, p 132-162.

8 List of Tables

9 List of Figures

10 Appendices

10.1 Evaluation of the General Kinetics Model Probability Expressions

This section is devoted to the evaluation of the transformation probability expressions given by Equations 48-51. Analytically and graphically the expressions are straightforward but for numerical evaluation, such as in Matlab or Fortran, several considerations must be made and we discuss those here.

For a given stress and temperature, the probability densities f_-, f_a, and f_+ are given by Equations 45-47. To evaluate these quantities we begin by substituting the Gibbs free energy expressions g_-, g_a, and g_+ from Equation 14. After some algebraic manipulation, the exponents can be separated into three terms, where only the first depends on strain ε.

$$e^{-\frac{1}{\gamma}g_-(\varepsilon,\sigma,T)} = e^{-\frac{E_m}{2\gamma}\left(\varepsilon+\varepsilon_R-\frac{\sigma}{E_m}\right)^2} e^{\frac{\sigma^2}{2\gamma E_m}} e^{\frac{\sigma\varepsilon_R}{\gamma}} \tag{210}$$

$$e^{-\frac{1}{\gamma}g_a(\varepsilon,\sigma,T)} = e^{-\frac{E_a}{2\gamma}\left(\varepsilon-\frac{\sigma}{E_a}\right)^2} e^{\frac{\sigma^2}{2\gamma E_a}} e^{-\frac{\beta}{\gamma}} \tag{211}$$

$$e^{-\frac{1}{\gamma}g_+(\varepsilon,\sigma,T)} = e^{-\frac{E_m}{2\gamma}\left(\varepsilon-\varepsilon_R-\frac{\sigma}{E_m}\right)^2} e^{\frac{\sigma^2}{2\gamma E_m}} e^{-\frac{\sigma\varepsilon_R}{\gamma}} \tag{212}$$

Recall here that, at a given instant in time, the entire range of strain within g_i is possible for the current stress state there for σ and ε are independent variables. When expressions 210-212 are inserted in to Equations -47, the last two terms can be removed from the integrals and canceled with the like terms in the numerator.

$$f_- = \frac{e^{-\frac{E_m}{2\gamma}\left(\varepsilon+\varepsilon_R-\frac{\sigma}{E_m}\right)^2}}{\int_{-\infty}^{\varepsilon_{-a}} e^{-\frac{E_m}{2\gamma}\left(\varepsilon+\varepsilon_R-\frac{\sigma}{E_m}\right)^2} d\varepsilon} \tag{213}$$

$$f_a = \frac{e^{-\frac{E_a}{2\gamma}\left(\varepsilon-\frac{\sigma}{E_a}\right)^2}}{\int_{\varepsilon_{-a}}^{\varepsilon_{a+}} e^{-\frac{E_a}{2\gamma}\left(\varepsilon-\frac{\sigma}{E_a}\right)^2} d\varepsilon} \tag{214}$$

$$f_+ = \frac{e^{-\frac{E_m}{2\gamma}\left(\varepsilon-\varepsilon_R-\frac{\sigma}{E_m}\right)^2}}{\int_{\varepsilon_{+a}}^{\infty} e^{-\frac{E_m}{2\gamma}\left(\varepsilon-\varepsilon_R-\frac{\sigma}{E_m}\right)^2} d\varepsilon} \tag{215}$$

For conciseness we collapse the exponential terms into $Z_i(\varepsilon,\sigma)$ expressions and rewrite the probability equations.

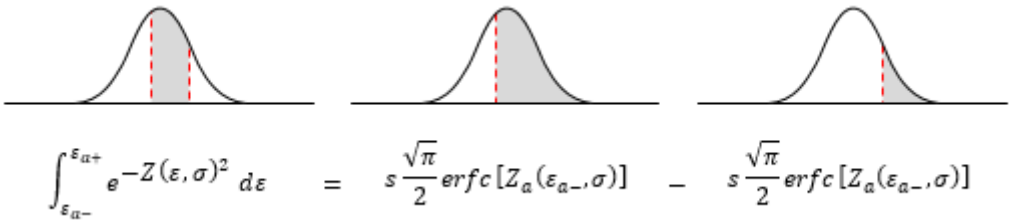
$$f_- = \frac{e^{-Z_-(\varepsilon,\sigma)^2}}{\int_{-\infty}^{\varepsilon_{-a}} e^{-Z_-(\varepsilon,\sigma)^2} d\varepsilon} \quad where \quad Z_-(\varepsilon,\sigma) = \sqrt{\frac{E_m}{2\gamma}}\left(\varepsilon+\varepsilon_R-\frac{\sigma}{E_m}\right) \tag{216}$$

$$f_a = \frac{e^{-Z_a(\varepsilon,\sigma)^2}}{\int_{\varepsilon_{a-}}^{\varepsilon_{a+}} e^{-Z_a(\varepsilon,\sigma)^2} d\varepsilon} \quad where \quad Z_a(\varepsilon,\sigma) = \sqrt{\frac{E_a}{2\gamma}}\left(\varepsilon-\frac{\sigma}{E_a}\right) \tag{217}$$

$$f_+ = \frac{e^{-Z_+(\varepsilon,\sigma)}}{\int_{\varepsilon_{+a}}^{\infty} e^{-Z_+(\varepsilon,\sigma)^2} d\varepsilon} \quad where \quad Z_+(\varepsilon,\sigma) = \sqrt{\frac{E_m}{2\gamma}}\left(\varepsilon-\varepsilon_R-\frac{\sigma}{E_m}\right) \tag{218}$$

At a given stress and temperature, Equations 216-218 can be integrated to calculate the probability of finding a layer over a range of strain. In model operation, we evaluate f_i at finite values of ε and consider it to represent the probability over a very small range of strain $d\varepsilon$.

Integrals such as those in Equations 216-218 are typically replaced with error (erf) and complimentary error ($erfc$) function expressions, which are acceptably accurate approximations available in most computer codes.



$$\int_{\varepsilon_{a-}}^{\varepsilon_{a+}} e^{-Z(\varepsilon,\sigma)^2}\, d\varepsilon \quad = \quad s\frac{\sqrt{\pi}}{2} erfc\left[Z_a(\varepsilon_{a-},\sigma)\right] \quad - \quad s\frac{\sqrt{\pi}}{2} erfc\left[Z_a(\varepsilon_{a-},\sigma)\right]$$

Making use of these expressions we rewrite Equations 216-218 to take the form of Equations 219-221.

$$f_-(\varepsilon,\sigma,T) = \frac{e^{-Z_-(\varepsilon,\sigma)^2}}{\sqrt{2\gamma/E_m}\frac{\sqrt{\pi}}{2}\left(2 - erfc\left(Z_-(\varepsilon_{-a}(T),\sigma)\right)\right)} \tag{219}$$

$$f_a(\varepsilon,\sigma,T) = \frac{e^{-Z_a(\varepsilon,\sigma)^2}}{\sqrt{2\gamma/E_a}\frac{\sqrt{\pi}}{2}\left(erfc\left(Z_a(\varepsilon_{a-}(T),\sigma)\right) - erfc\left(Z_a(\varepsilon_{a+}(T),\sigma)\right)\right)} \tag{220}$$

$$f_+(\varepsilon,\sigma,T) = \frac{e^{-Z_+(\varepsilon,\sigma)^2}}{\sqrt{2\gamma/E_m}\frac{\sqrt{\pi}}{2} erfc\left(Z_+(\varepsilon_{+a}(T),\sigma)\right)} \tag{221}$$

The transformation probabilities are based on evaluation of the probability expressions at the inflection points. Substituting Equations 219-221 into Equations 48-51 we have the following.

$$p_{-a}(\sigma,T) = \frac{1}{\tau}\frac{e^{-Z_-(\varepsilon_{-a}(T),\sigma)^2}}{\sqrt{2\gamma/E_-}\frac{\sqrt{\pi}}{2}\left(2 - erfc\left(Z_-(\varepsilon_{-a}(T),\sigma)\right)\right)} \tag{222}$$

$$p_{a-}(\sigma,T) = \frac{1}{\tau}\frac{e^{-Z_a(\varepsilon_{a-}(T),\sigma)^2}}{\sqrt{2\gamma/E_a}\frac{\sqrt{\pi}}{2}\left(erfc\left(Z_a(\varepsilon_{a-}(T),\sigma)\right) - erfc\left(Z_a(\varepsilon_{a+}(T),\sigma)\right)\right)} \tag{223}$$

$$p_{a-}(\sigma,T) = \frac{1}{\tau}\frac{e^{-Z_a(\varepsilon_{a-}(T),\sigma)^2}}{\sqrt{2\gamma/E_a}\frac{\sqrt{\pi}}{2}\left(erfc\left(Z_a(\varepsilon_{a-}(T),\sigma)\right) - erfc\left(Z_a(\varepsilon_{a+}(T),\sigma)\right)\right)} \tag{224}$$

$$p_{+a}(\sigma,T) = \frac{1}{\tau}\frac{e^{-Z_+(\varepsilon_{+a}(T),\sigma)^2}}{\sqrt{2\gamma/E_m}\frac{\sqrt{\pi}}{2} erfc\left(Z_+(\varepsilon_{+a}(T),\sigma)\right)} \tag{225}$$

At times, the probability densities f_i used to construct the p_{ij} expressions are evaluated far away from the mean strain which results in an $\approx \frac{0}{0}$ calculation. Furthermore, at low temperatures, the austenite well becomes completely unstable which causes the probability of layers exiting the well to become infinite ($\propto \frac{1}{0}$). Both of these effects are representative of the underlying physics and they encourage rate-independence of the kinetic equations (in the absence of self-heating) however, they must be addressed for the purposes of numerical implementation. We begin with an illustration in Figure 123 and discussion of the case where the austenite well becomes unstable.

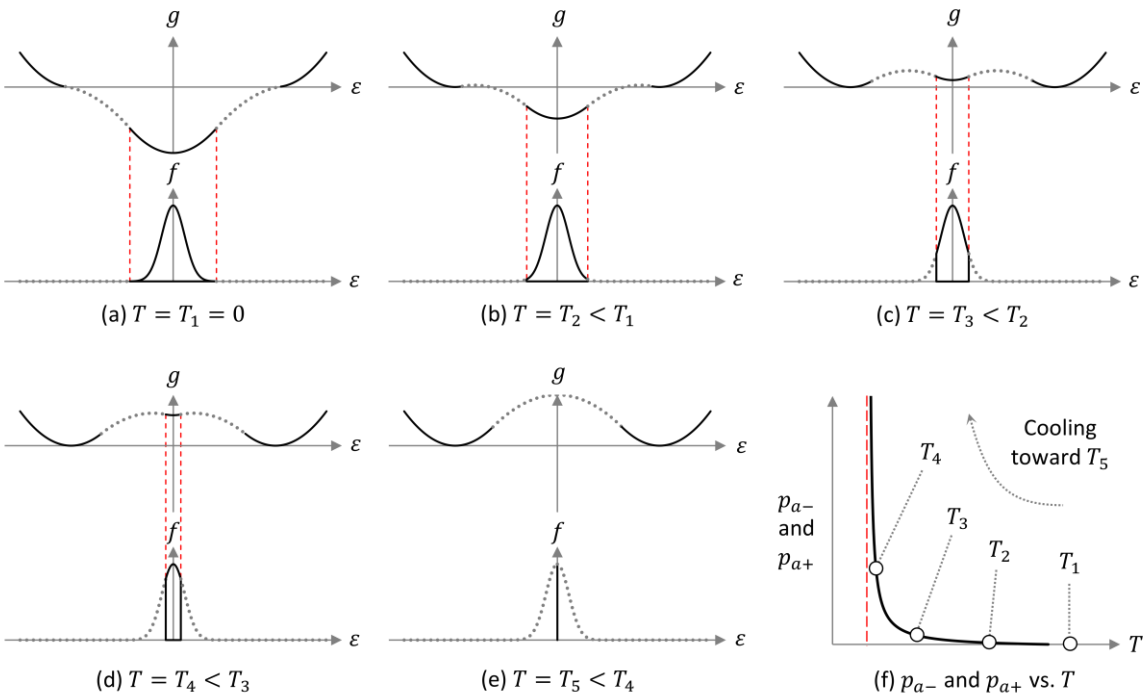


Figure 123: GFE, probability density, and transformation rate under decreasing temperature

Figure 123 shows that, as the convex austenite well shrinks, the left and right boundaries of the probability density function f_a simultaneously move toward $\varepsilon = 0$. The black-outlined f regions in Figure 123(a-e) represent the domain of integration for the denominators in both Equation 223 and 224. Mathematically, the result is that transformation rates p_{a-} and p_{a+} approach infinity as the width of the austenite well (and thus the value denominator integral) approaches zero. Physically this represents temperature-induced martensite formation and the temperature where it occurs, shown here as T_5, is often called the martensite start temperature.

The infinite-rate effect shown in Figure 123 is consistent with the physics of the system because the austenite well is no longer stable, therefore any layers will exit instantaneously. For numerical implementation, however, the $p_{a-} = p_{a+} = {}^{1}/_{\tau}\,({}^{1}/_{0})$ result shown in Figure 123e is not allowed. We bypass this by imposing a lower limit on the strains ε_{a-} and ε_{a+} during the evaluation of Equations 223 and 224. The limiting strain value ($1\mathrm{E}-10$ used here) is chosen to achieve transformation rates p_{a-} and p_{a+} that are finite but sufficiently high to yield transformation behavior that is much faster than any strain rates encountered in the simulation. The final, numerical evaluation-ready, set of transformation rate equations is given at the end of this section after discussion of the p $\approx {}^{0}/_{0}$ cases.

We now turn our attention to stress conditions where the second numerical limitation must be addressed. In cases of extremely large positive or negative stress magnitudes, each of the four transformation rate equations (222-225) encounters a situation where the probability term approaches ${}^{0}/_{0}$. A graphical example is shown here in Figure 124 for the austenite to martensite + transformation rate p_{a+}.

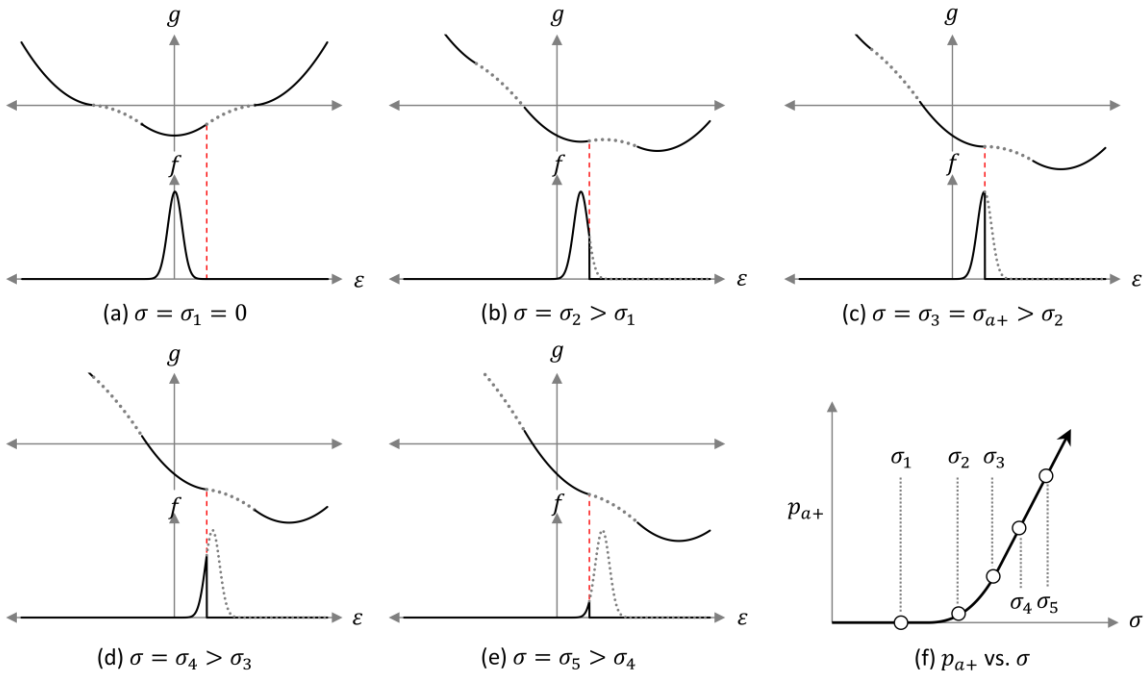


Figure 124: GFE, probability density, and transformation rate under increasing tensile stress

Again, the solid black portion of the f vs. ε curves represents the area included in the integral which is in the denominator of Equation 223. The numerator can be visualized in the Figure as the value of f at the point where it intersects the red dashed line; note that the probability density is truncated at this location. In Figure 124(a), the value of the numerator is approximately zero and the denominator integral includes most of the non-zero f curve and is approximately equal to 1. This results in a transformation rate p_{a+} that is nearly zero as shown in 124(f). In Figure 124(b and c), the stress level increases and the f curve moves to the right following the austenite well minimum. The critical strain (red dashed line) remains fixed which causes the value of the numerator to increase as the truncation point climbs the f curve. The integrated portion of the f curve (the denominator) simultaneously diminishes and the cumulative effect is a transformation rate p_{a+} that increases. The stress continues to increase in Figure 124(d and e) which causes the f curve to move past the critical strain and thus the numerator falls back toward zero. The transformation rate, however, continues to rise (124(f)) because the integral that constitutes the denominator also races toward zero and does so at a greater pace. A Bernoulli-L'Hopital substitution in Equation 223 reveals a finite limit as the fraction approaches $^0/_0$ and this is shown in Figure 124(f) where the rate reaches a linear ramp trajectory beyond 5th stress state. This mechanism encourages rapid stress-induced transformations in model operation which it is consistent with observed physical behavior – i.e. rate independence in the absence of self-heating or cooling. Below, we discuss the stress conditions and give the final expressions for each of the 4 transformation rate equations including limit substitution expressions and logic.

For the rate of transformation p_{-a}, large positive values of stress result in $p_{-a} \to {}^0/_0$. A limit expression derived from the derivatives of the numerator and denominator is substituted at $Z^* = -5.5$ yielding the final expression for p_{-a} that was used for the simulations in this paper.

$$p_{-a}=\frac{1}{\tau}\begin{cases} if \quad Z<Z^* & -\dfrac{2Z}{s} \\ if \quad Z^*\le Z & \dfrac{e^{-Z^2}}{s\frac{\sqrt{\pi}}{2}\left(2-erfc(Z)\right)} \end{cases} \qquad where \qquad \begin{cases} Z=\dfrac{1}{s}\left(\varepsilon_{-a}(T)+\varepsilon_R-\dfrac{\sigma}{E_m}\right) \\ s=\sqrt{2\gamma/E_m} \end{cases} \tag{226}$$

The limit substitution point Z^* was chosen to reflect a large stress value in order to allow a smooth transition from the main expression to the limit expression but prior to encountering numerical overflow errors. A 64-bit Dell Precision M6600 with an Intel® Core ™ i7-2720QM CPU and Intel Fortran compiler edition 11. 1 was used for these simulations and the proceeding limit substitution points (denoted by Z^*, Z^{**}, or ε^*) were chosen in a similar manner.

The rate of transformation p_{+a} has a similar structure except a limit expression was substituted at $Z^*=26.3$ which corresponds to the case of large negative values of stress.

$$p_{+a}(\sigma,T)=\frac{1}{\tau}\begin{cases} if \quad Z<Z^* & \dfrac{e^{-Z^2}}{s\frac{\sqrt{\pi}}{2}erfc(Z)} \\ if \quad Z^*\le Z & \dfrac{2Z}{s} \end{cases} \qquad where \qquad \begin{cases} Z=\dfrac{1}{s}\left(\varepsilon_{+a}(T)-\varepsilon_R-\dfrac{\sigma}{E_m}\right) \\ s=\sqrt{2\gamma/E_m} \end{cases} \tag{227}$$

The rates of transformation out of the central austenite well (p_{a-} and p_{a+}) each have three overflow substitutions representing the cases of large $+/-$ stress values and low temperatures. The numerical overflow switch points for both expressions are $Z^*=-80.0$, $Z^{**}=26.0$, and $\varepsilon^*=1\cdot10^{-10}$.

$$p_{a-}=\frac{1}{\tau}\begin{cases} if \quad Z_1<Z^* & 0 \\ if \quad Z^*\le Z_1\le Z^{**} & \left|\dfrac{e^{-{Z_1}^2}}{s\frac{\sqrt{\pi}}{2}\left(erfc(Z_1)-erfc(Z_2)\right)}\right| \\ if \quad Z^{**}<Z_1 & \dfrac{2Z_1}{s} \end{cases} \qquad where \qquad \begin{cases} Z_1=\dfrac{1}{s}\left(-max(\varepsilon_{a+}(T),\varepsilon^*)-\dfrac{\sigma}{E_a}\right) \\ Z_2=\dfrac{1}{s}\left(max(\varepsilon_{a+}(T),\varepsilon^*)-\dfrac{\sigma}{E_a}\right) \\ s=\sqrt{2\gamma/E_a} \end{cases} \tag{228}$$

$$p_{a+}=\frac{1}{\tau}\begin{cases} if \quad Z_2<Z^* & -\dfrac{2Z_2}{s_a} \\ if \quad Z^*\le Z_2\le Z^{**} & \left|\dfrac{e^{-{Z_2}^2}}{s\frac{\sqrt{\pi}}{2}\left(erfc(Z_1)-erfc(Z_2)\right)}\right| \\ if \quad Z^{**}<Z_2 & 0 \end{cases} \qquad where \qquad \begin{cases} Z_1=\dfrac{1}{s}\left(-max(\varepsilon_{a+}(T),\varepsilon^*)-\dfrac{\sigma}{E_a}\right) \\ Z_2=\dfrac{1}{s}\left(max(\varepsilon_{a+}(T),\varepsilon^*)-\dfrac{\sigma}{E_a}\right) \\ s=\sqrt{2\gamma/E_a} \end{cases} \tag{229}$$

Expressions 226, 227, 228, and 229 allow the evaluation of the transformation rate Equations and therefore the kinetics equations for all values of stress and temperature.

10.2 Energy Barriers for the Asymmetric Free Energy

In Chapter 5, only the essential components of the free energy function were included. The energy barriers in the new 1-D model are not necessary to operate the model and were therefore not fully defined. Several free energy plots, however, do show the full landscape. For completeness and to show how to define the continuous free energy plots, we have included the expressions to define the energy barrier coefficients here. The two energy

barriers were given in Equation 139 and are repeated here, where six coefficients $A^{\pm}$, $B^{\pm}$, and $C^{\pm}$ were left undefined.

$$\psi_{\mp}^{B}(\varepsilon,T) = A^{\mp}(T)\varepsilon^2 + B^{\mp}(T)\varepsilon + C^{\mp}(T) \tag{230}$$

The coefficients are determined in the same manner as in Chapter 2, except it is done twice – once for the ψ_{+}^{B} coefficients and once for the ψ_{-}^{B} set. The continuity conditions imposed on each set of parabolas gives us the following two sets of four equations.

Compression		Tension	
$\psi_a\vert_{\varepsilon_{a-}(T)} = \psi_{-}^{B}\vert_{\varepsilon_{a-}(T)}$	$\psi_{-}^{B}\vert_{\varepsilon_{-a}(T)} = \psi_{-}\vert_{\varepsilon_{-a}(T)}$	$\psi_a\vert_{\varepsilon_{a+}(T)} = \psi_{+}^{B}\vert_{\varepsilon_{a+}(T)}$	$\psi_{+}^{B}\vert_{\varepsilon_{+a}(T)} = \psi_{+}\vert_{\varepsilon_{+a}(T)}$
$\left.\frac{\partial\psi_a}{\partial\varepsilon}\right\vert_{\varepsilon_{a-}(T)} = \left.\frac{\partial\psi_{-}^{B}}{\partial\varepsilon}\right\vert_{\varepsilon_{a-}(T)}$	$\left.\frac{\partial\psi_{-}^{B}}{\partial\varepsilon}\right\vert_{\varepsilon_{-a}(T)} = \left.\frac{\partial\psi_{-}}{\partial\varepsilon}\right\vert_{\varepsilon_{-a}(T)}$	$\left.\frac{\partial\psi_a}{\partial\varepsilon}\right\vert_{\varepsilon_{a+}(T)} = \left.\frac{\partial\psi_{+}^{B}}{\partial\varepsilon}\right\vert_{\varepsilon_{a+}(T)}$	$\left.\frac{\partial\psi_{+}^{B}}{\partial\varepsilon}\right\vert_{\varepsilon_{+a}(T)} = \left.\frac{\partial\psi_{+}}{\partial\varepsilon}\right\vert_{\varepsilon_{+a}(T)}$
(231,232,233,234)		(235,236,237,238)	

Simultaneous solution of Equations 231-234 allows us to solve for A^{-}, B^{-}, C^{-} and β_{-} while simultaneous solution of 235-238 yields A^{+}, B^{+}, C^{+} and β_{+}. The solutions for the β's were given by Equation 156; the solutions for the other three pairs of parameters are given here.

Compression	Tension
$A^{-}(T) = \frac{\sigma_{\Delta}^{-}(T)}{2\left(\varepsilon_{a-}(T) - \varepsilon_{-a}(T)\right)}$	$A^{+}(T) = \frac{\sigma_{\Delta}^{+}(T)}{2\left(\varepsilon_{a+}(T) - \varepsilon_{+a}(T)\right)}$
$B^{-}(T) = \frac{\sigma_{-a}(T)\varepsilon_{a-}(T) - \sigma_{a-}(T)\varepsilon_{-a}(T)}{\varepsilon_{a-}(T) - \varepsilon_{-a}(T)}$	$B^{+}(T) = \frac{\sigma_{+a}(T)\varepsilon_{a+}(T) - \sigma_{a+}(T)\varepsilon_{+a}(T)}{\varepsilon_{a+}(T) - \varepsilon_{+a}(T)}$
$C^{-}(T) = \left(\frac{E_a}{2} - A^{-}(T)\right)\varepsilon_{a-}(T)^2 - B^{-}(T)\varepsilon_{a-}(T)$	$C^{+}(T) = \left(\frac{E_a}{2} - A^{+}(T)\right)\varepsilon_{a+}(T)^2 - B^{+}(T)\varepsilon_{a+}(T)$
(239,240,241)	(242,243,244)

Some of the expressions have different forms than their analogous coefficients from the symmetric version (Equations 10-13). This is due to the β terms being in different locations prior to the algebraic manipulation. It should also be noted that the expressions above assume that the linear transformation stress expression is being used, so Equations 141, 142, 145, and 146 should be used for the $\sigma_{ij}(T)$ parameters and Equations 152 and 153 should be used for the $\varepsilon_{ij}(T)$'s.

10.3 Free Energy Landscape Plots

Six different experimental data sets were used in the *Simulations* section to demonstrate the model's behavior and their material parameters were given in Table 3. In the *Mesoscopic Free Energy* section, free energy plots were only shown as needed to visualize the features being discussed – in most cases the tension-compression asymmetric data from Reedlunn et al. (2014) was used. Here, in Figure 125, we show the HFE landscape using the material parameters from each of the six experimental data sets for visual comparison. Each plot shows the same five temperatures and has axes scaled to the same ranges (both energy and strain).

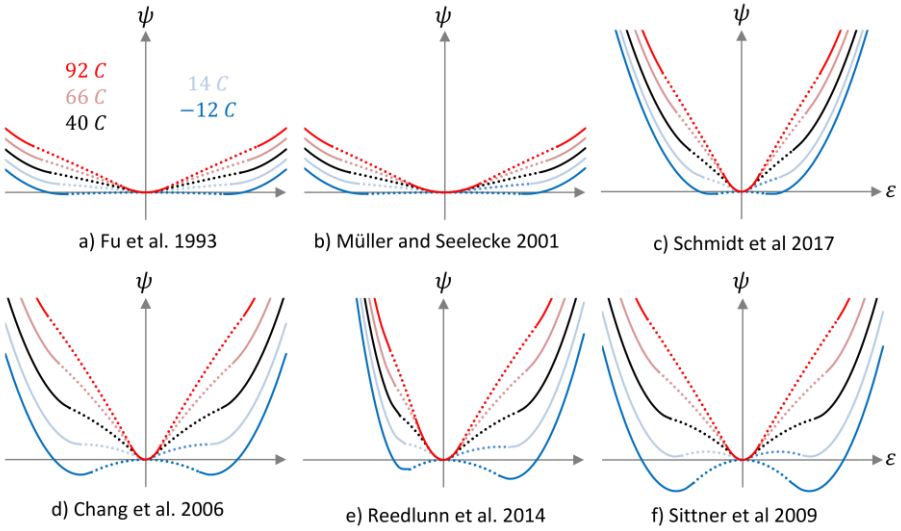


Figure 125: Helmholtz free energy plots for SMA materials from six different experimental data sets

10.4 Additional Calculations for the 2-D Model

10.4.1 Calculating Stress from a Critical Free Energy

Each of the critical free energy curves shown in Figures 104-109 represents a set of strain combinations $\{\varepsilon\},\{\gamma\}$. Since each strain is a function of stress (178 and 179) we are really finding a series of normal stress, shear stress pairs that result in the critical free energy and we write this condition using Equation 183.

$$g_c = -\frac{\sigma^2}{2E_i} - \frac{\tau^2}{2Y_i} + \beta(T) - \sigma\varepsilon_i^R - \tau\gamma_i^R \tag{245}$$

The easiest way to generate the critical free energy curves is to treat either normal stress or shear stress as an independent parameter and then solve for the other. It should be noted that the choice for the independent stress value has to be within the operating range of the model – i.e. the stress and temperature ranges where material properties are scoped. In the case where the normal stress is an input parameter, we can write a quadratic expression for the shear stress.

$$\tau(\sigma, g_c) = \frac{-B \pm \sqrt{B^2 - 4\cdot A\cdot C}}{2A} \quad \rightarrow \quad A = -\frac{1}{2Y_i} \quad B = -\gamma_i^R \quad C = -\frac{\sigma^2}{2E_i} + \beta(T) - \sigma\varepsilon_i^R - g_c \tag{246}$$

Or, in the case where the shear stress is an input parameter, we can write a quadratic expression for the normal stress.

$$\sigma(\tau, g_c) = \frac{-B \pm \sqrt{B^2 - 4\cdot A\cdot C}}{2A} \quad \rightarrow \quad A = -\frac{1}{2E_i} \quad B = -\varepsilon_i^R \quad C = -\frac{\tau^2}{2Y_i} + \beta(T) - \tau\gamma_i^R - g_c \tag{247}$$

Equations 246 and 247 are not necessary to operate the model because but they are useful for visualizing the free energy boundaries and they were used to generate the $\sigma = 132 - 379\ MPa$ curves in Figures 103 and 110.

10.5 2-D Model Equations with the Coupling Parameter

The normal-shear coupling parameter M, that was briefly discussed in Chapter 6, was neglected in the main dissertation simply because it proved not to play a significant role. The model developed, however, was presented as a first step in extending the 1-D M-A-S model to multiple dimensions and was limited in several ways, such as neglecting macroscopic field effects as well as the simplified manner of discretizing the radial wire domain. We acknowledge that a more complete material-level theory may be called for in a future work where these macro-

scale inhomogeneities are better addressed. In this light, we have included the free energy-related model equations in this section with the M term included to provide a complete record of the theory.

10.5.1 Stress and Strain Equations

The expressions for average strain within a given phase, originally Equations 178 and 179, are now written as follows.

$$\varepsilon_i(\sigma,\tau) = \frac{\sigma Y_i - \tau M_i}{E_i Y_i - {M_i}^2} + \varepsilon_i^R = \sigma\left(\frac{Y_i}{E_i Y_i - {M_i}^2}\right) - \tau\left(\frac{M_i}{E_i Y_i - {M_i}^2}\right) + \varepsilon_i^R$$

$$\varepsilon_i(\sigma,\tau) = \sigma\frac{Y_i}{X} - \tau\frac{M_i}{X} + \varepsilon_i^R \qquad where \qquad X = E_i Y_i - {M_i}^2 \tag{248}$$

$$\gamma_i(\sigma,\tau) = \frac{\tau}{Y_i}\left(1 + \frac{{M_i}^2}{E_i Y_i - {M_i}^2}\right) - \left(\frac{\sigma M_i}{E_i Y_i - {M_i}^2}\right) + \gamma_i^R = \tau\left(\frac{1}{Y_i} + \frac{{M_i}^2}{Y_i X}\right) - \sigma\frac{M_i}{X} + \gamma_i^R \tag{249}$$

To write stress as a function of strian and phase fractions, originally Equations 203 and 204, we now write the following expressions.

$$\varepsilon = \sum_i x_i \cdot \varepsilon_i(\sigma,\tau) = \sum_i x_i \cdot \sigma\frac{Y_i}{X} - \sum_i x_i \cdot \tau\frac{M_i}{X} + \sum_i x_i \cdot \varepsilon_i^R$$

$$\sigma(\varepsilon,\tau,x_i) = \frac{\varepsilon - \sum_i x_i\varepsilon_i^R + \sum_i \frac{x_i\tau M_i}{X}}{\sum_i \frac{x_i Y_i}{X}} \qquad , \qquad i \in i-5 \tag{250}$$

$$\gamma = \sum_i x_i \cdot \gamma_i(\sigma,\tau) = \sum_i x_i\tau\left(\frac{1}{Y_i} + \frac{{M_i}^2}{Y_i X}\right) - \sum_i x_i\sigma\frac{M_i}{X} + \sum_i x_i\gamma_i^R$$

$$\tau(\gamma,\sigma,x_i) = \frac{\gamma - \sum_i x_i\gamma_i^R + \sum_i \frac{x_i\sigma M_i}{X}}{\sum_i x_i\left(\frac{1}{Y_i} + \frac{{M_i}^2}{Y_i X}\right)} \qquad , \qquad i \in i-5 \tag{251}$$

10.5.2 Calculating Stress from Total Strain and Phase Fractions

During model operation, the order of evaluation of the macroscopic stress and strain equations depends on which parameters are known and which need to be calculated. For example, to simulate the *RR* tension-torsion tests, the axial stress σ and the shear strain γ are known leaving the axial strain ε and shear stress τ to be calculated. For the *RR* tension tests, on the other hand, both of the strains are known inputs and the stresses need to be found. In the latter case, we must solve Equations 203 and 204 to obtain expressions for stress as a function of strains and phase fractions. We begin with Equations 203 and 204 by collecting terms and introducing temporary place holder parameters B and C.

$$\sigma = \frac{\varepsilon - \sum_i x_i\varepsilon_i^R + \sum_i \frac{x_i\tau M_i}{X}}{\sum_i \frac{x_i Y_i}{X}} = \frac{\varepsilon - \sum_i x_i\varepsilon_i^R + \sum_i \frac{x_i\tau M_i}{X}}{B} = \frac{\varepsilon}{B} - \frac{\sum_i x_i\varepsilon_i^R}{B} + \frac{\tau A}{B} \tag{252}$$

$$\tau = \frac{\gamma - \sum_i x_i\gamma_i^R + \sum_i \frac{x_i\sigma M_i}{X}}{\sum_i x_i\left(\frac{1}{Y_i} + \frac{{M_i}^2}{Y_i X}\right)} = \frac{\gamma - \sum_i x_i\gamma_i^R + \sum_i \frac{x_i\sigma M_i}{X}}{C} = \frac{\gamma}{C} - \frac{\sum_i x_i\gamma_i^R}{C} + \frac{\sigma A}{C} \tag{253}$$

Now, substituting Equation 253 into Equation 252 we add one more place holder A and write an expression for normal stress as a function of strain and phase fractions.

$$\sigma = \frac{\varepsilon}{B} - \frac{\sum_i x_i \varepsilon_i^R}{B} + \left(\frac{\gamma}{C} - \frac{\sum_i x_i \gamma_i^R}{C} + \frac{\sigma A}{C}\right)\frac{A}{B} = \frac{\varepsilon}{B} - \frac{\sum_i x_i \varepsilon_i^R}{B} + \frac{\gamma A}{BC} - \frac{\sum_i x_i \gamma_i^R A}{BC} + \frac{\sigma A^2}{BC}$$

$$\sigma - \frac{\sigma A^2}{BC} = \frac{1}{BC}\left(C\varepsilon - C\sum_i x_i \varepsilon_i^R + A\gamma - A\sum_i x_i \gamma_i^R\right) \rightarrow \sigma BC - \sigma A^2 = C\varepsilon - C\sum_i x_i \varepsilon_i^R + A\gamma - A\sum_i x_i \gamma_i^R$$

$$\sigma = \frac{C\left(\varepsilon - \sum_i x_i \varepsilon_i^R\right) + A\left(\gamma - \sum_i x_i \gamma_i^R\right)}{BC - A^2} \qquad where \qquad \begin{aligned} A &= \sum_i \frac{x_i M_i}{X} \qquad C = \sum_i x_i \left(\frac{1}{Y_i} + \frac{{M_i}^2}{Y_i X}\right) \\ B &= \sum_i \frac{x_i Y_i}{X} \qquad X = E_i Y_i - {M_i}^2 \end{aligned} \tag{254}$$

After solving for normal stress σ via Equation 254, Equation 253 can then be evaluated for shear stress τ.

10.5.3 Calculating Stress from a Critical Free Energy

When the M parameter is kept, the process of find stress pairs that follow the constant critical free energy lines is more complex. We begin with the GFE expression that maintains the coupling term.

$$\begin{aligned} g_c = \; & \frac{E_i}{2}\left(\varepsilon_i(\sigma,\tau) - \varepsilon_i^R\right)^2 + \frac{Y_i}{2}\left(\gamma_i(\sigma,\tau) - \gamma_i^R\right)^2 \\ & + M_i\left(\varepsilon_i(\sigma,\tau) - \varepsilon_i^R\right)\left(\gamma_i(\sigma,\tau) - \gamma_i^R\right) + \beta_i(T) - \sigma\varepsilon_i(\sigma,\tau) - \tau\gamma_i(\sigma,\tau) \end{aligned} \tag{255}$$

Now, we substitute Equations 248 and 249 and then rearrange into a quadratic function of σ.

$$\begin{aligned} g_c = \; & \frac{E_i}{2}\left(\sigma\frac{Y_i}{X} - \tau\frac{M_i}{X}\right)^2 + \frac{Y_i}{2}\left(\tau\left(\frac{1}{Y_i} + \frac{{M_i}^2}{Y_i X}\right) - \sigma\frac{M_i}{X}\right)^2 + M_i\left(\sigma\frac{Y_i}{X} - \tau\frac{M_i}{X}\right)\left(\tau\left(\frac{1}{Y_i} + \frac{{M_i}^2}{Y_i X}\right) - \sigma\frac{M_i}{X}\right) + \beta_i(T) \\ & -\sigma\left(\sigma\frac{Y_i}{X} - \tau\frac{M_i}{X} + \varepsilon_i^R\right) - \tau\left(\tau\left(\frac{1}{Y_i} + \frac{{M_i}^2}{Y_i X}\right) - \sigma\frac{M_i}{X} + \gamma_i^R\right) \\ g_c = \; & \tau^2\left(\frac{E_i {M_i}^2}{2X^2}\right) + \tau^2\frac{Y_i}{2}\left(\frac{1}{Y_i} + \frac{{M_i}^2}{Y_i X}\right)^2 - \tau^2\left(\frac{{M_i}^2}{Y_i X}\right) - \tau^2\left(\frac{{M_i}^4}{Y_i X^2}\right) - \tau^2\left(\frac{1}{Y_i} + \frac{{M_i}^2}{Y_i X}\right) \\ & -\tau\left(\frac{E_i \sigma Y_i M_i}{X^2}\right) - \tau\left(\frac{\sigma M_i}{X}\right) - \tau\left(\frac{\sigma {M_i}^3}{X^2}\right) + \tau\left(\frac{\sigma M_i}{X}\right) + \tau\left(\frac{\sigma {M_i}^3}{X^2}\right) + \tau\left(\frac{\sigma {M_i}^3}{X^2}\right) + \tau\left(\frac{2\sigma M_i}{X}\right) \\ & -\tau\gamma_i^R + \left(\frac{E_i \sigma^2 {Y_i}^2}{2X^2}\right) + \left(\frac{\sigma^2 Y_i {M_i}^2}{2X^2}\right) - \left(\frac{\sigma^2 Y_i {M_i}^2}{X^2}\right) + \beta_i(T) - \left(\frac{\sigma^2 Y_i}{X}\right) - \sigma\varepsilon_i^R \end{aligned} \tag{256}$$

Now, Equation 256 can be separated into the proper terms to use the quadratic equation to solve for τ in Equation 257.

$$\tau(g_c,\sigma)_i = \frac{-B \pm \sqrt{B^2 - 4\cdot A\cdot C}}{2A} \qquad where \qquad \begin{cases} A = \left(\frac{E_i {M_i}^2}{2X^2}\right) + \frac{Y_i}{2}\left(\frac{1}{Y_i} + \frac{{M_i}^2}{Y_i X}\right)^2 - \left(\frac{2{M_i}^2}{Y_i X}\right) - \left(\frac{{M_i}^4}{Y_i X^2}\right) - \left(\frac{1}{Y_i}\right) \\ B = -\left(\frac{E_i \sigma Y_i M_i}{X^2}\right) + \left(\frac{\sigma {M_i}^3}{X^2}\right) + \left(\frac{2\sigma M_i}{X}\right) - \gamma_i^R \\ C = \left(\frac{E_i \sigma^2 {Y_i}^2}{2X^2}\right) - \left(\frac{\sigma^2 Y_i {M_i}^2}{2X^2}\right) + \beta(T) - \left(\frac{\sigma^2 Y_i}{X}\right) - \sigma\varepsilon_i^R - g_c \end{cases} \tag{257}$$

If a shear stress τ is assumed, the corresponding normal stress σ can be obtained in a similar manner.

$$\sigma(g_c,\tau)_i = \frac{-B \pm \sqrt{B^2 - 4\cdot A\cdot C}}{2A} \qquad where \qquad \begin{cases} A = \left(\frac{E_i {Y_i}^2}{2X^2}\right) - \left(\frac{Y_i {M_i}^2}{2X^2}\right) - \left(\frac{Y_i}{X}\right) \\ B = \left(\frac{2\tau M_i}{X}\right) - \frac{\tau E_i Y_i M_i}{X^2} + \left(\frac{\tau {M_i}^3}{X^2}\right) - \varepsilon_i^R \\ C = \beta_i(T) + \frac{\tau^2 E_i {M_i}^2}{2X^2} - \left(\frac{\tau^2 {M_i}^2}{Y_i X}\right) - \left(\frac{\tau^2 {M_i}^4}{2Y_i X^2}\right) - \left(\frac{\tau^2}{2Y_i}\right) - \tau\gamma_i^R - g_c \end{cases} \tag{258}$$

10.6 Summary of SMA Roundrobin Model Results

The 2-D models from several groups or researchers were discussed in the *Introduction* section. In most cases these models were tested using the *RR* experimental data and we include a brief summary and discussion of their results here. It should be pointed out that the plots shown here were taken directly from the published papers, not from reproducing their models or simulations.

The model described in Gibeau et al. (2009) was also included in the *RR* program and uses stress-induced martensite as well as self-accommodated martensite internal variables. A macroscopic Helmholtz free energy contains internal energy, entropy, elastic energy, specific heat, and interaction terms. A Gibbs free energy is derived from the Helmholtz energy and it also has a phase dependent elastic modulus. The free energies are used, via partial derivatives, to define driving forces for elastic strain, thermal strain, transformation strain, entropy, as well as proportions of stress-induced and self-accommodated martensite. A yield function is used to define the threshold for transformation while a Claussius-Duhem inequality along with several logical conditions (such as loading vs. unloading and forward vs. reverse) are used to supply flow rules. A specific choice of stress-invariant parameters is used to also enable tension-compression asymmetry, similarly to Drucker-Prager methods which are common concrete and soil modeling.

This model represented the uniaxial RR tests reasonably (Figures 126 and 127), though the martensite branch under-predicts the strain at some temperatures (20 and 40 C).

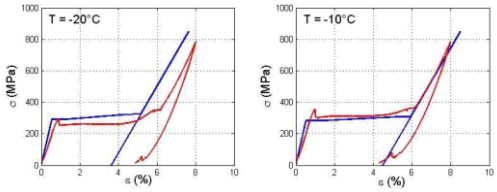


Figure 126: Tension test simulations from Gibeau et al. (2009) compared to experiments from the *RR* program, Pseudoplastic temperatures

The progression of the transformation stresses and the shape of the hysteresis loop also seem irregular between 10 C and 60 C (Figure 127).

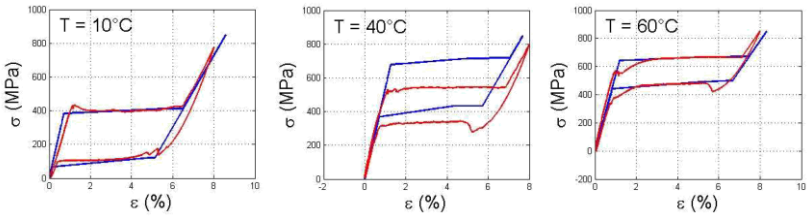


Figure 127: Tension test simulations from Gibeau et al. (2009) compared to *RR* experiments, Superelastic temperatures

The actuation behavior is well captured (Figure 128) although the lower-stress experiments, where a no-actuation threshold was observed, were not included in the simulations.

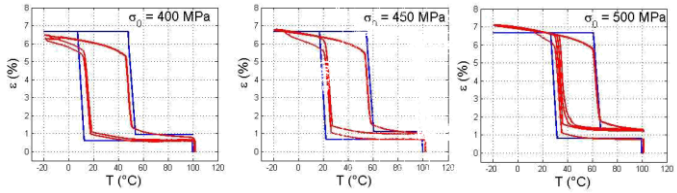


Figure 128: Tensile actuation test simulations from Gibeau et al. (2009) compared to *RR* experiments, Superelastic temperatures

For the tension-torsion tests, the model failed to capture the smooth shape of the curves and underpredicted the transformation strains in the low-temperature cases (Figure 129).

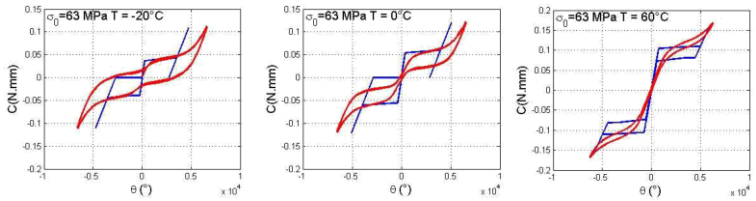


Figure 129: Tension-torsion test simulations from Gibeau et al. (2009) compared to *RR* experiments, Several temperatures and a low level of axial stress

Only a single axial strain was presented for the tension-torsion simulations so no assessment could be made about the model's handling of the coupling between the axial stress and the tension-torsion response.

Rio et al. (2009) presented an elasto-hysteresis model where the total stress was composed of a hyperelastic component and an elastoplastic hysteresis component. The hyperelastic potential is based on strain invariants, relative volume variation, deviatoric deformation intensity, and an angle giving the direction of the deformation. The hysteresis is governed by an incremental model with discrete memory points. This model was able to accurately match the RR experiments in most cases and, in the cases where it did not match well, the overall characteristics of the responses were still correct. In the uniaxial tension tests the model shows a transition from pseudoplastic to superelastic, accurate partial unloading loops, as well as shrinking hysteresis (Figure 130).

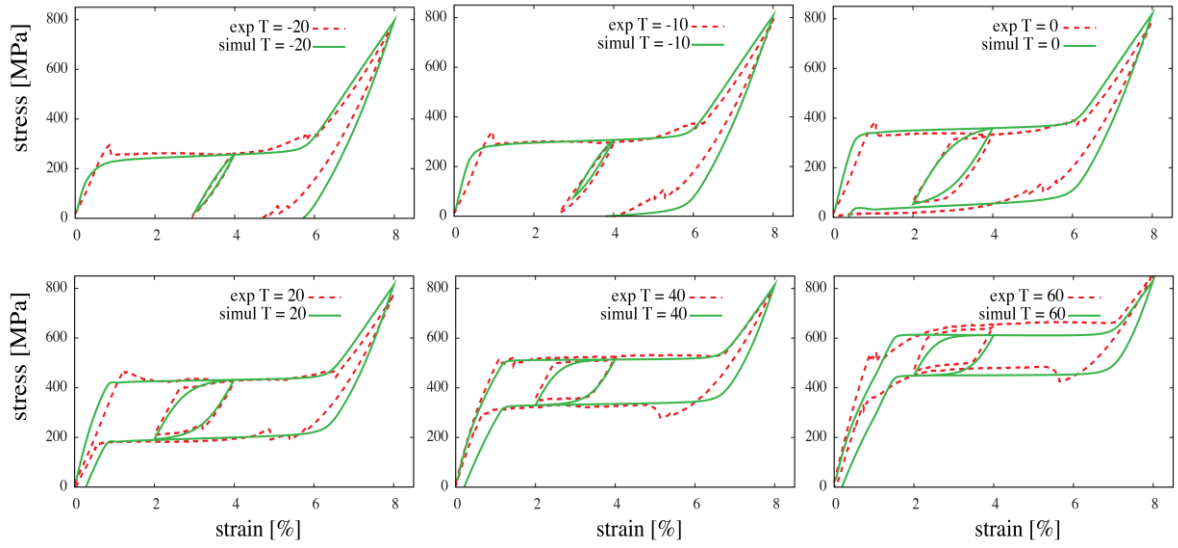


Figure 130: Tensile test simulations from Rio et al. (2009) compared to *RR* experiments, Superelastic temperatures with a partial unloading loop

In the actuation tests, the minimum axial stress required to produce actuation was slightly off but only by one 100 MPa increment (Figure 131).

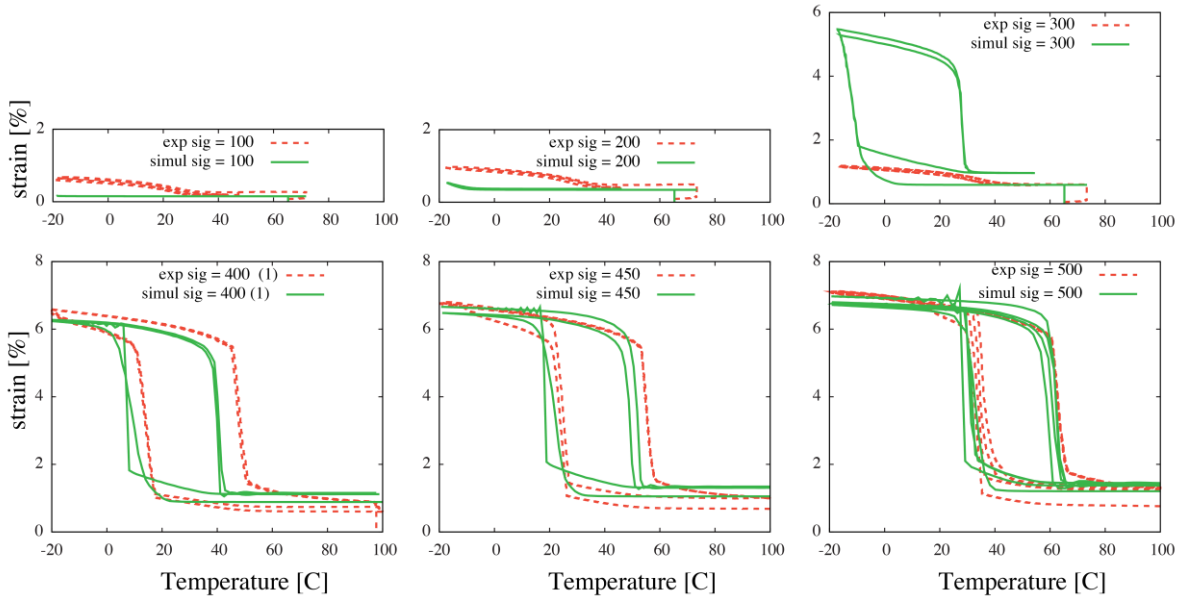


Figure 131: Tensile actuation test simulations from Rio et al. (2009) compared to *RR* experiments, At several values of axial stress

In the fixed-strain thermal cycle simulations, the stress-strain response is accurately predicted but the model produces stress vs. temperature curves that under-predict the stress (Figure 132).

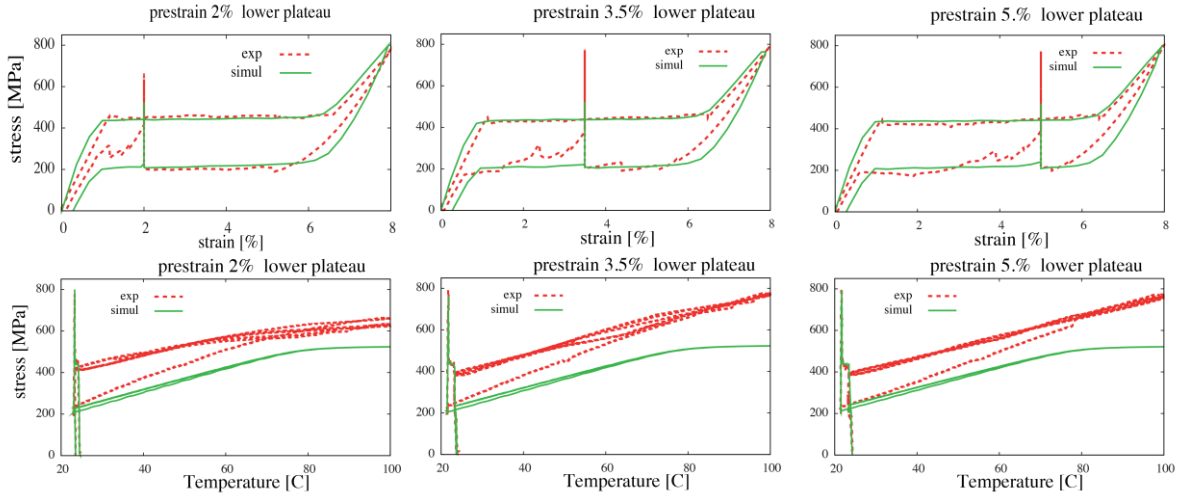


Figure 132: Fixed strain temperature cycle simulations from Rio et al. (2009) compared to *RR* experiments

The simulations of the tension-torsion tests (Figures 133, 134, and 135)demonstrated that the model captured both the high-temperature and low-temperature modes, as well as the transition between the two.

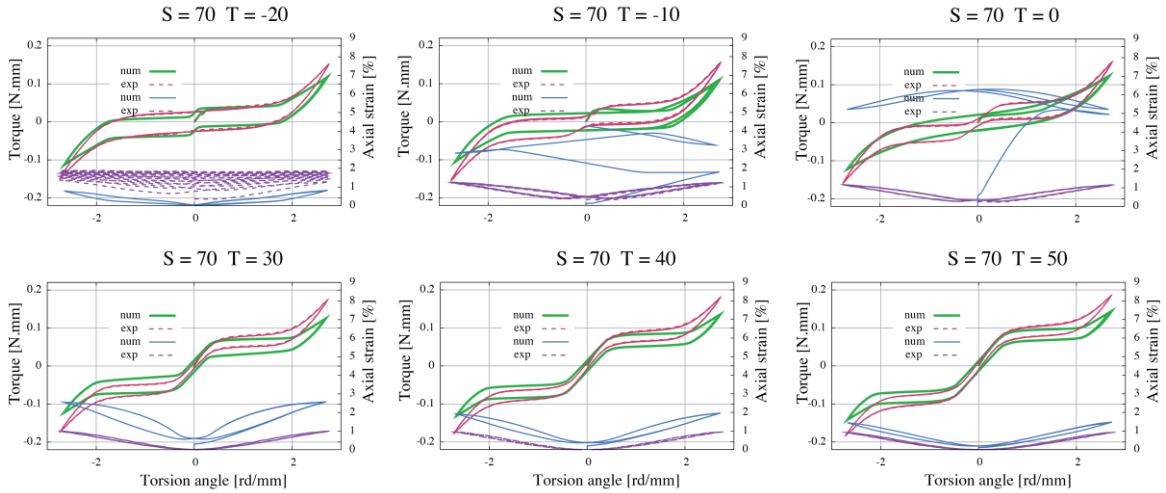


Figure 133: Tension-torsion simulations from Rio et al. (2009) compared to *RR* experiments, 70 MPa axial stress data set

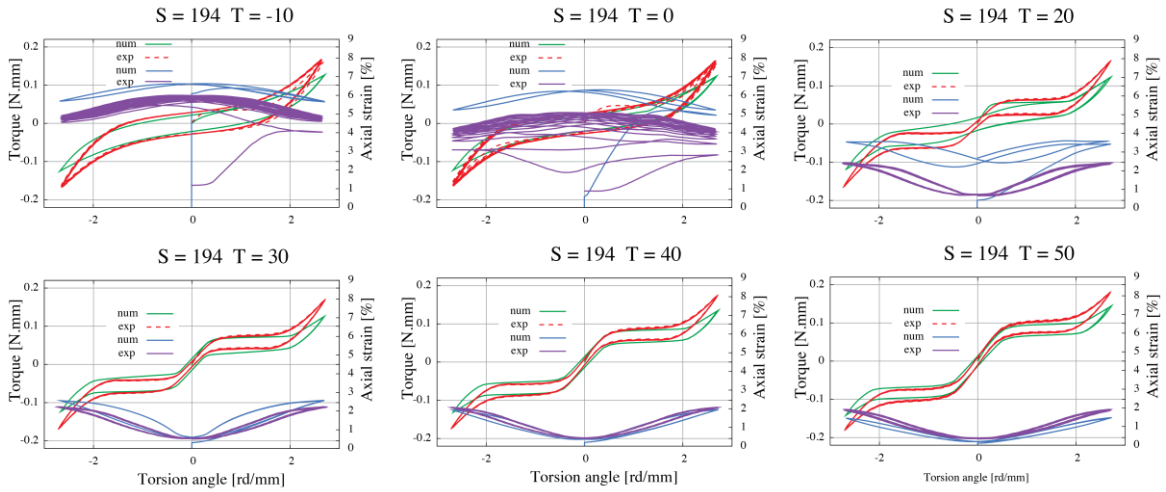


Figure 134: Tension-torsion simulations from Rio et al. (2009) compared to *RR* experiments, 194 MPa axial stress data set

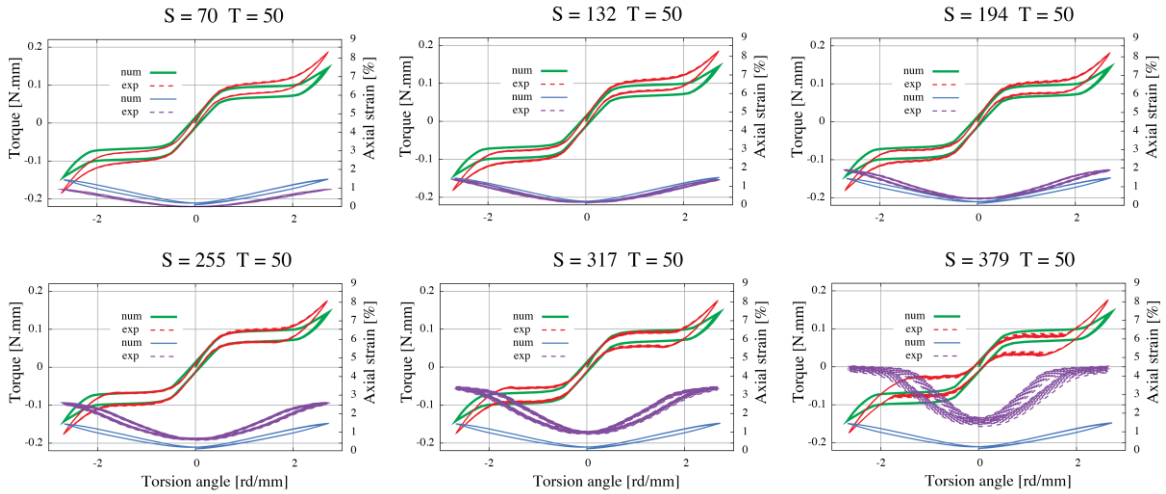


Figure 135: Tension-torsion simulations from Rio et al. (2009) compared to *RR* experiments, 317 MPa axial stress data set

The axial responses also exhibited low and high-temperature modes, but the magnitudes were off. The experimental data shows a gradual increase in axial strain as the tensile stress increases. The axial strain in the simulations is relatively insensitive to the external axial load.

Chemisky and colleagues (Chemisky et al., 2009) present results from a model where the inelastic strain is an average of that from martensite transformation and twinned martensite. A Gibbs free energy is written containing chemical, elastic, thermal strain, inelastic, and mixing energy terms. A set of dissipation potentials and evolution rules govern the progression of the internal variables and phase diagram-like method is used to identify the material parameters. The model is implemented in a 3-D FEA simulation (Abaqus via UMAT) and it captures the uniaxial tension tests as well as the thermal cycle tests reasonably well (Figures 136, 137, and 138).

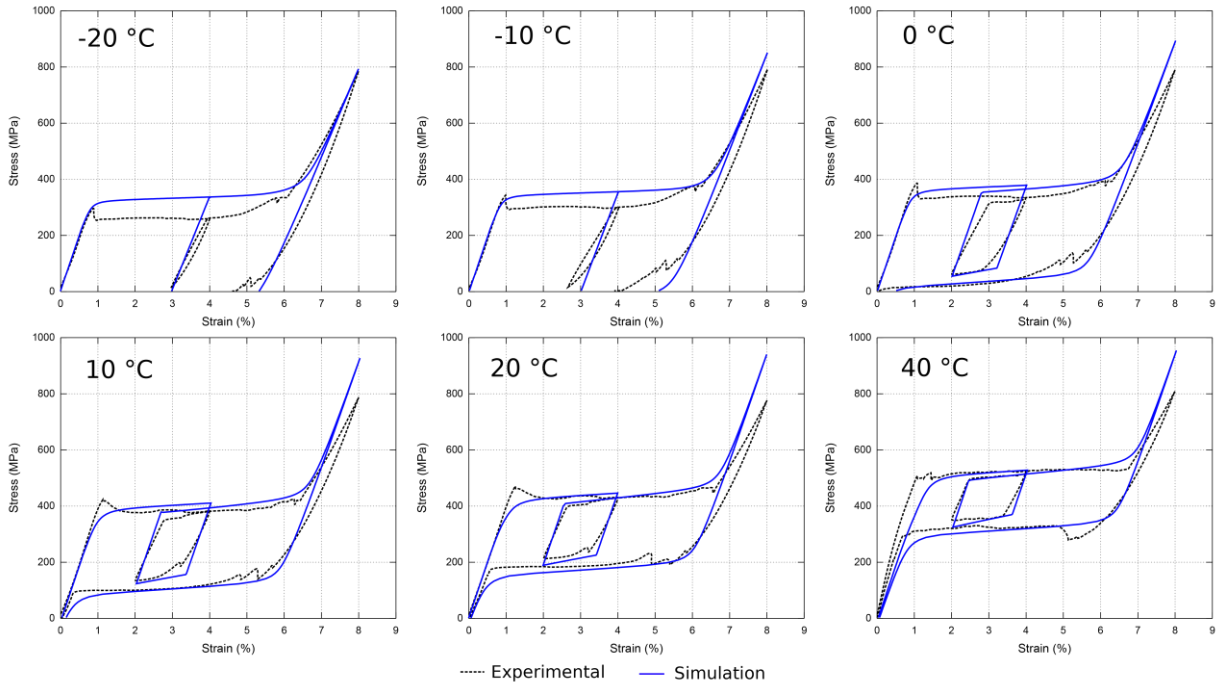


Figure 136: Tension test simulations with partial unloading loops from Chemisky et al. (2009) compared to *RR* experiments

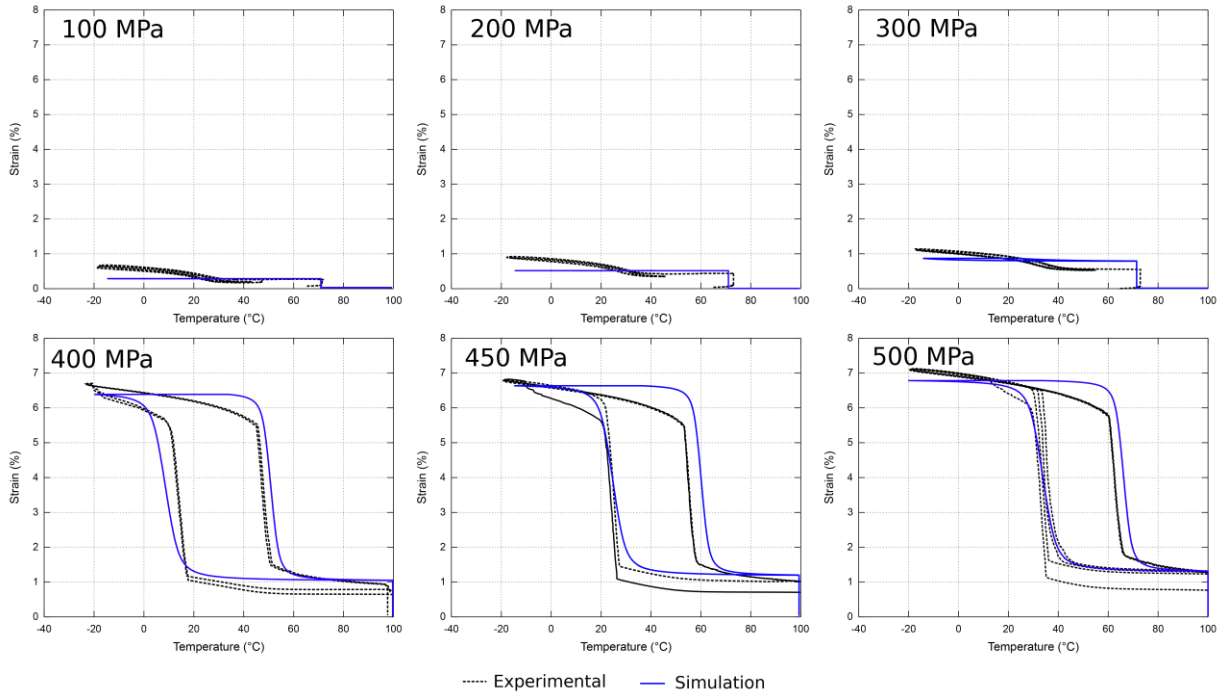


Figure 137: Actuation test simulations from Chemisky et al. (2009) compared to *RR* experiments

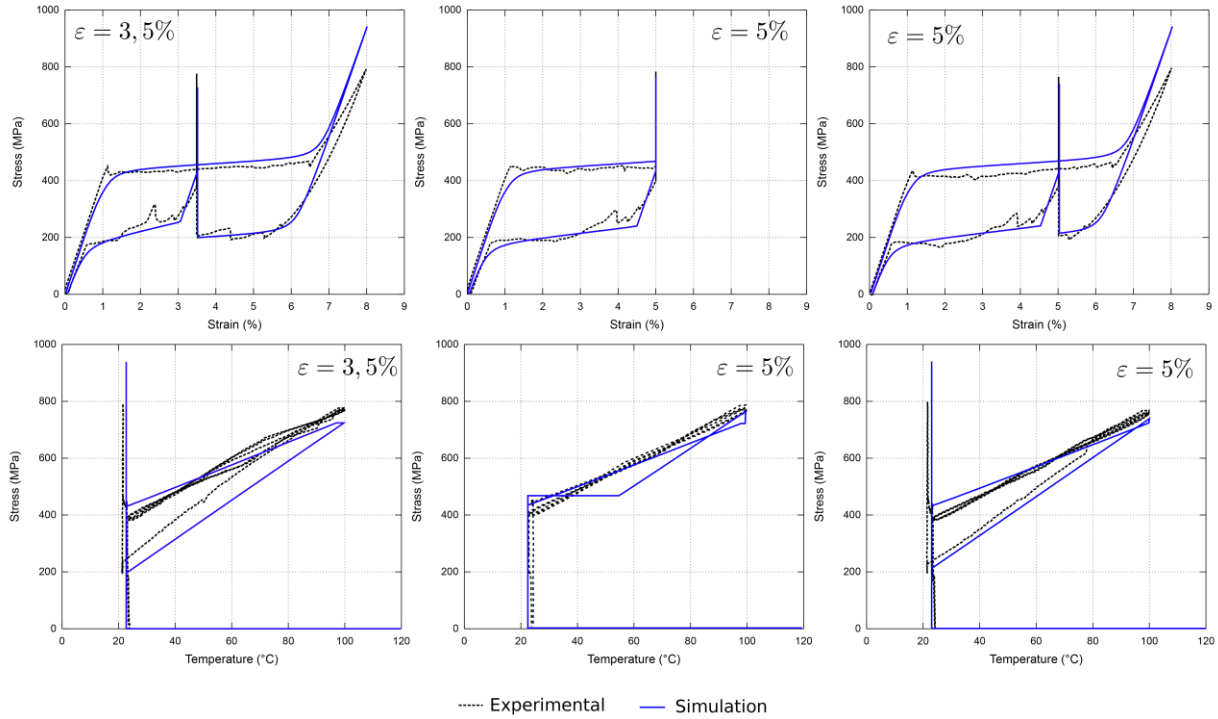


Figure 138: Fixed-strain temperature cycle simulations from Chemisky et al. (2009) compared to *RR* experiments

In combined tension and torsion, the model represents the two modes of deformation (high and low temperature), but the magnitudes are inaccurate in some of the cases and the characteristic shapes of the curves do not match (Figures 139, 140, and 141).

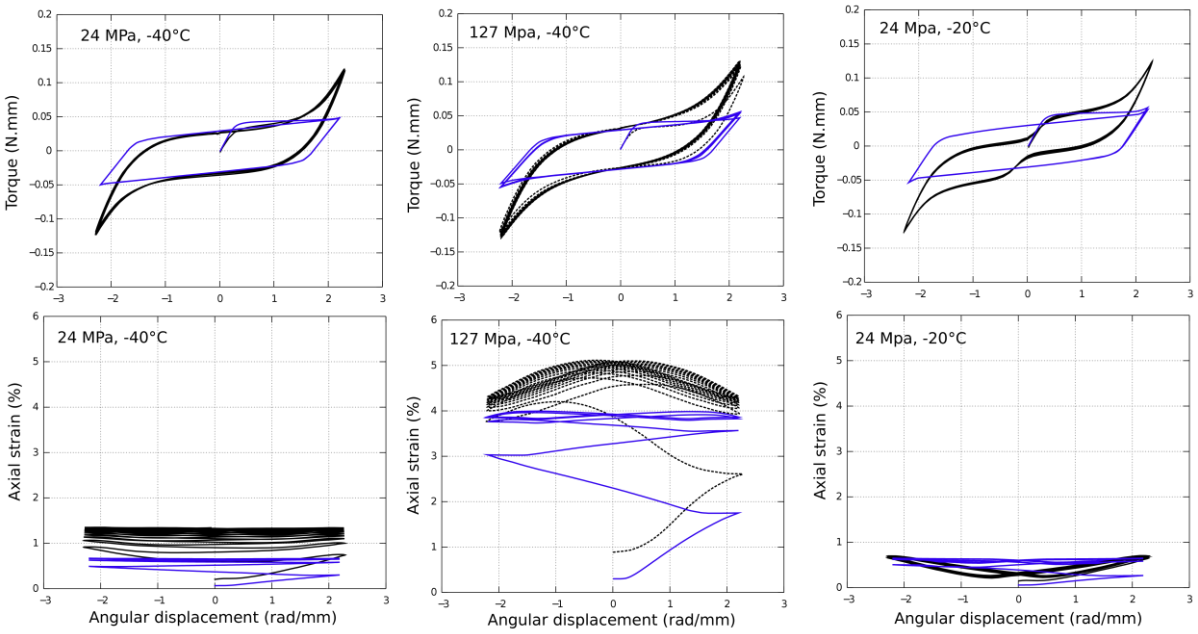


Figure 139: Tension-torsion test simulations from Chemisky et al. (2009) compared to *RR* experiments, Sampling of temperature and axial stress cases

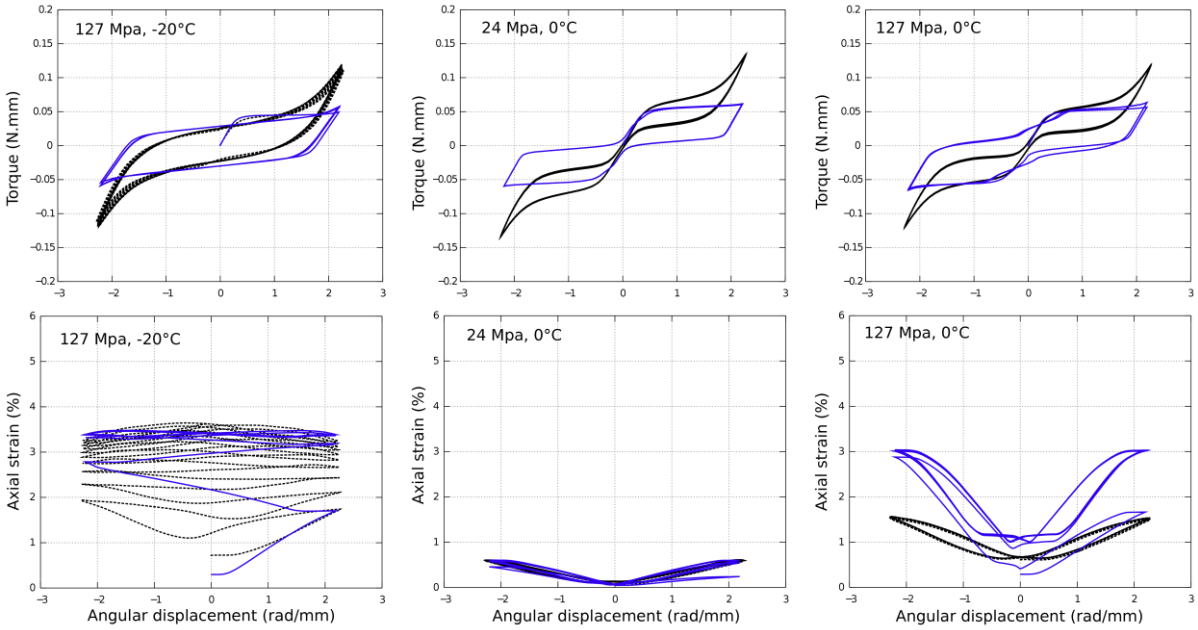


Figure 140: Tension-torsion test simulations from Chemisky et al. (2009) compared to *RR* experiments, Sampling of temperature and axial stress cases

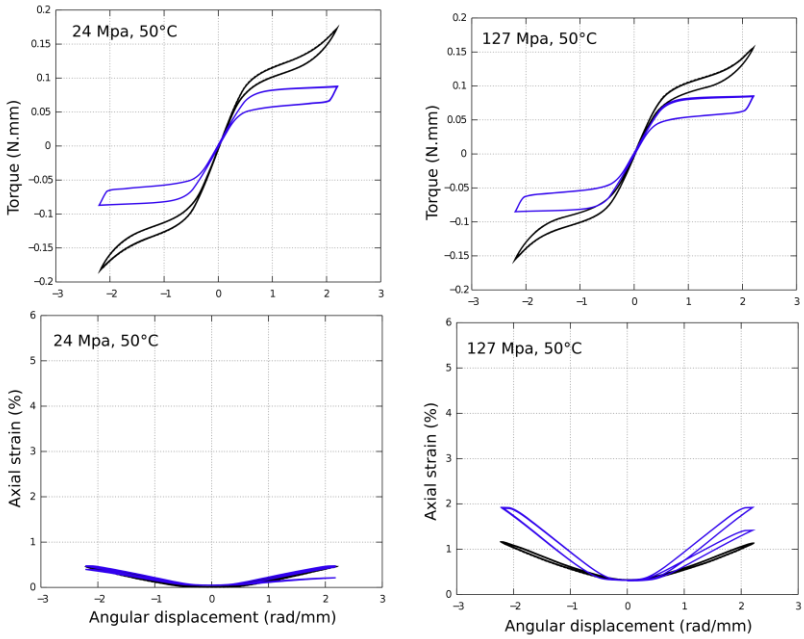


Figure 141: Tension-torsion test simulations from Chemisky et al. (2009) compared to *RR* experiments, Sampling of temperature and axial stress cases

This model assumes a total strain composed of an elastic component (with a stiffness that is phase fraction-dependent), a thermal component, a transformation component, and a plastic component. The formulation is of the plasticity-type, where the inelastic strains accumulate in the direction of a deviator stress invariant. The evolution equations use loading vs. unloading logic, a maximum achievable strain parameter that is stress dependent, as well as back stress and transformation hardening terms. The start-stop criteria are based on driving forces which are also related to the various terms from the evolution equations.

The model's agreement with the uniaxial tension tests (Figure 142) is qualitatively good but the transformation hardening is exaggerated in the tension tests; this is especially true in the low-temperature (pseudo-plastic) case.

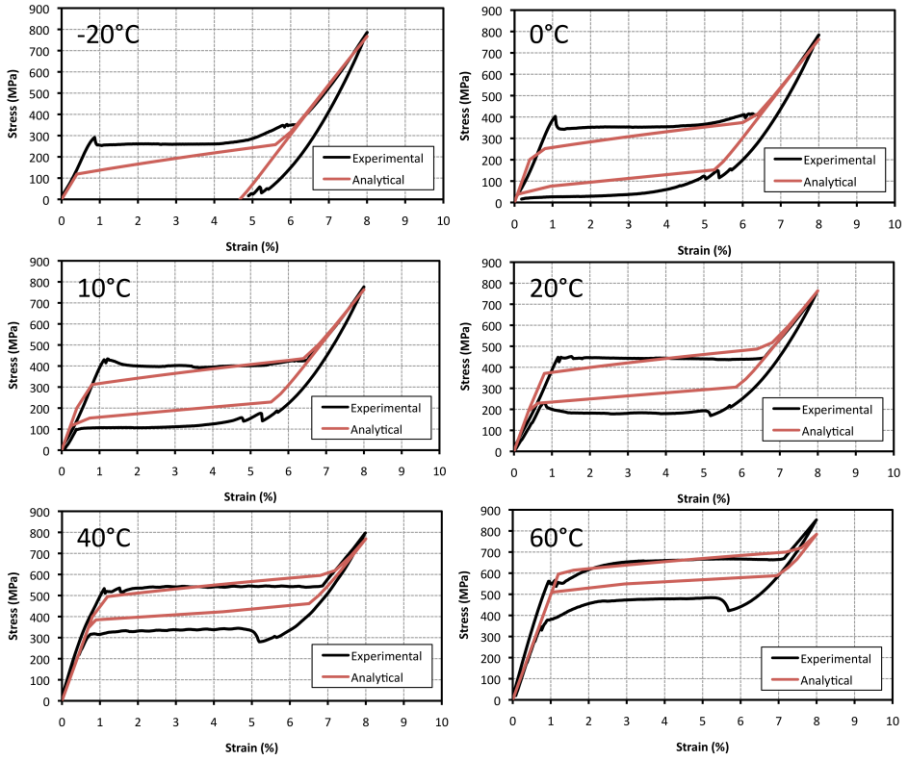


Figure 142: Tension test simulations from Hartl et al. (2009) compared to *RR* experiments

The actuation tests, Figure 143, show better agreement in the characteristic shape of the curves, but the forward actuation temperatures are under-predicted and the authors do not show if the model predicts the low-stress no-actuation cases.

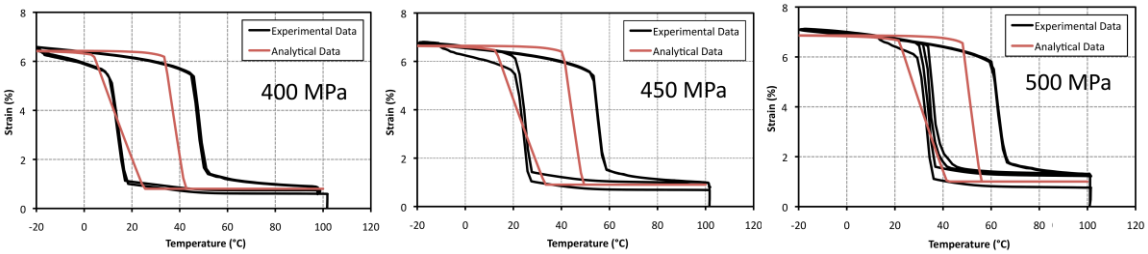


Figure 143: Actuation test simulations from Hartl et al. (2009) compared to *RR* experiments

The constant-strain thermal cycle simulations show good agreement with the experiments within the limitations of the model's representation of the hysteresis loop (Figure 144).

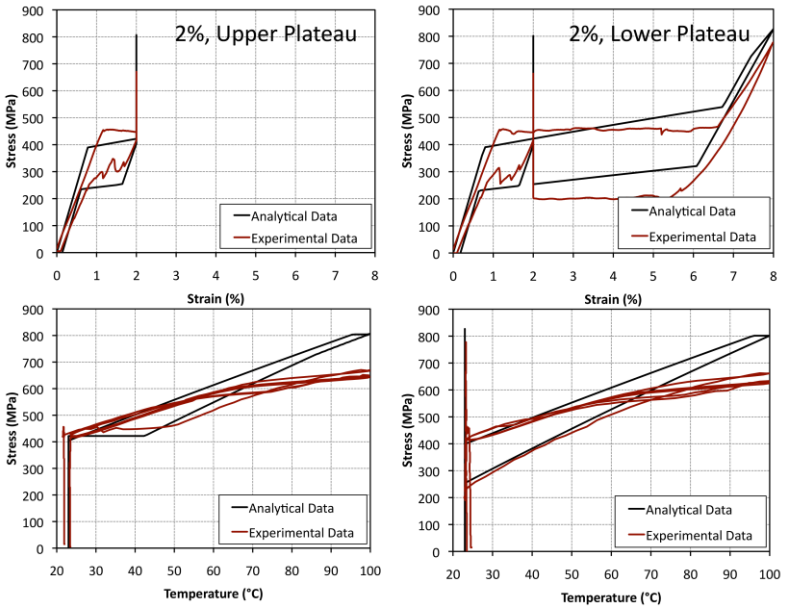


Figure 144: Fixed-strain temperature cycle simulations from Hartl et al. (2009) compared to *RR* experiments

The authors show results from a few tension-torsion test simulations (Figure 145), but they only include the torque vs. twist plots. The agreement is reasonable, but the axial response is not included, therefore no evaluation can be made about the axial-twist coupling in the model.

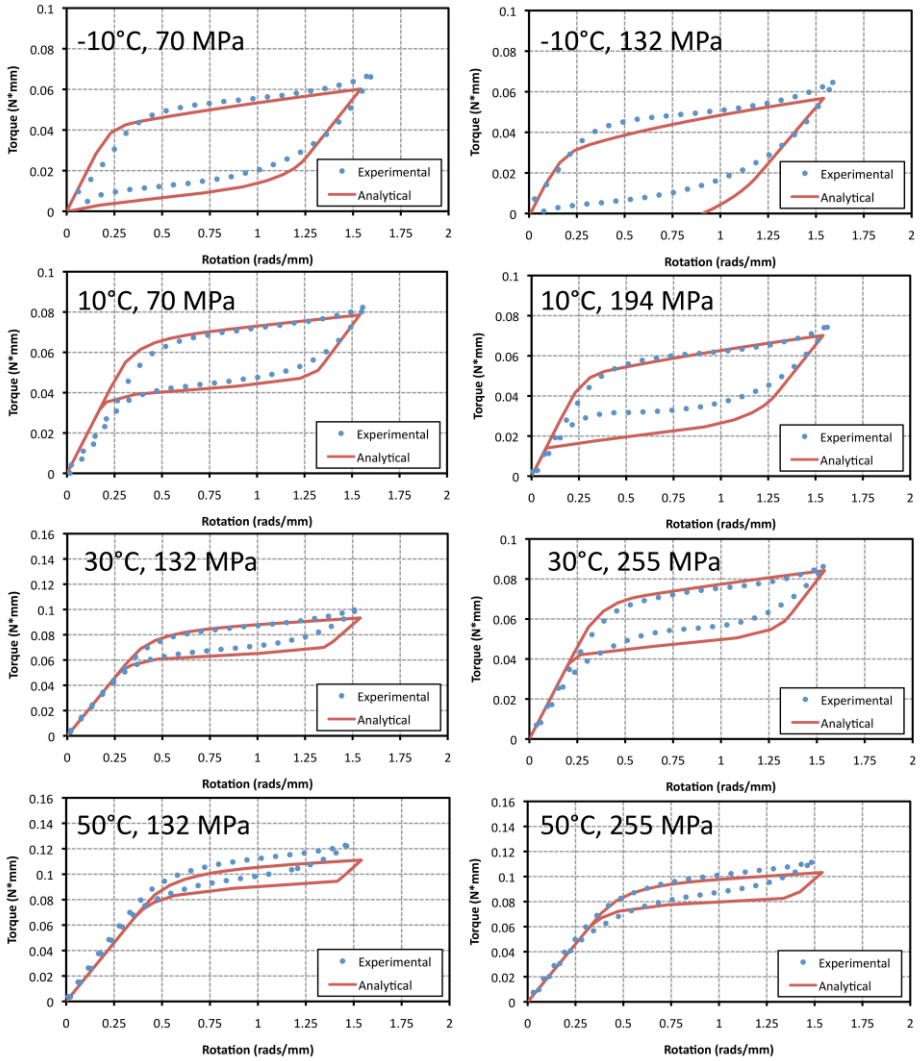


Figure 145: Tension-torsion test simulations from Hartl et al. (2009) compared to *RR* experiments

An earlier model was described in Frost and Sedlak (2009), which was published as part of the *RR* program, but no simulations were included in that work. The 2012 paper presents a more mature model with material anisotropy, an R phase, martensite reorientation, along with multi-axial loading. The total strain is segregated into elastic and inelastic parts, where the inelastic strain is based on a transformation strain tensor and a martensite phase fraction. A HFE is introduced with a mechanical and chemical component; penalty-type energy terms are added to enforce strain constraints, such as limiting the inelastic strain to deviatoric components and keep the phase fractions between zero and one. Dissipation functions are introduced as alternatives to yield functions to govern the onset of stress and temperature-induced transformations as well as martensite reorientation. The reorientation term has a temperature-dependent component which allows temperature dependent hysteresis. A set of thermodynamic forces along with minimization routines for the dissipation functions play the role of flow rules and dictate the evolution of the inelastic strains.

Simulations are presented for a small sampling of the *RR* experimental data. The model shows good agreement for the uniaxial tension (Figure 146) and actuation (Figure 147) experiments; especially in capturing the non-actuation response that was observed at stresses below 400 MPa.

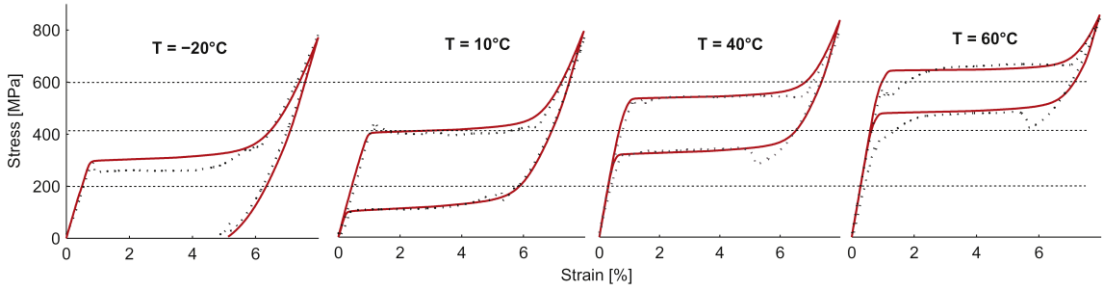


Figure 146: Tension test simulations from Sedlak et al. (2012) compared to *RR* experiments

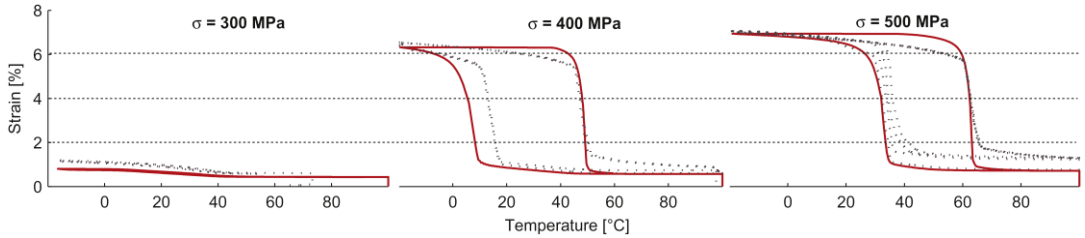


Figure 147: Actuation test simulations from Sedlak et al. (2012) compared to *RR* experiments

For the tension-torsion tests (Figures 148-151), the agreement between the model and the experiments appears to be good, though only two of six axial stresses are shown. In Figure 148 we see that the model captures both modes of torque-twist response (single-loop and double loop), the transition between them.

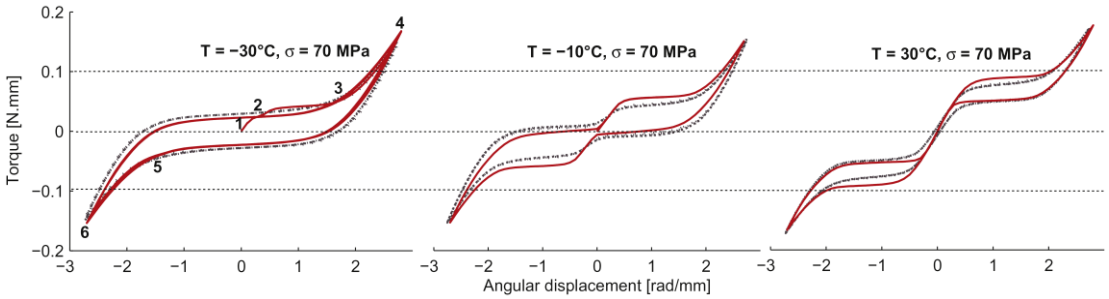


Figure 148: Tension-torsion test simulations from Sedlak et al. (2012) compared to *RR* experiments, Torque-twist plots

The two modes of axial strain vs. twist are also captured, as is the ratcheting into the low temperature mode (Figure 149).

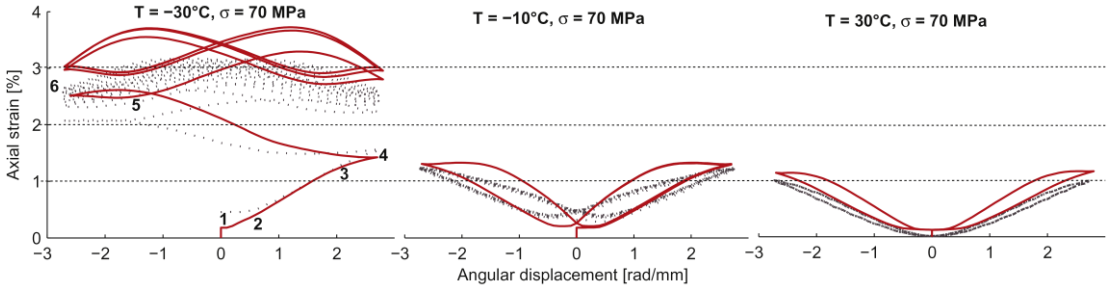


Figure 149: Tension-torsion test simulations from Sedlak et al. (2012) compared to *RR* experiments, Axial strain-twist plots

At a higher value of axial stress, the shape of the low-temperature hysteresis loop changes to a thin, sloped configuration. The model captures this effect accurately (Figure 150).

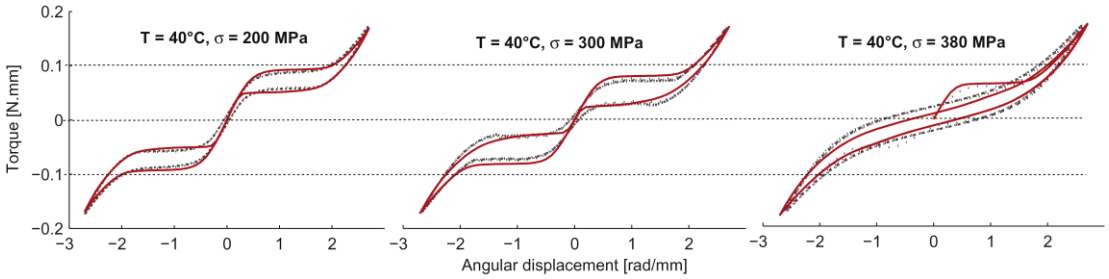


Figure 150: Tension-torsion test simulations from Sedlak et al. (2012) compared to *RR* experiments, Torque-twist plots at various axial stress values

Also, the increase in axial strain that occurs with increasing the axial stress is reflected remarkably well by the model (Figure 151).

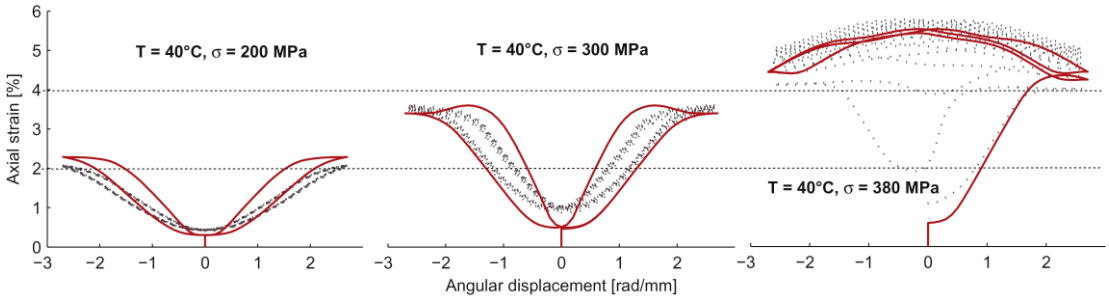


Figure 151: Tension-torsion simulations from Sedlak et al. (2012) compared to experiments from the *RR* program, Axial strain-twist plots at various axial stress values

This model is proposed in the context of the RR program's combined loading and uses self-accommodated martensite, oriented martensite, as well as preferred-direction martensite internal variables to describe the phase composition of the material. The strain is decomposed into elastic and inelastic components and the strain domain is made to be two dimensional, only considering a normal and a shear (torsional) component. A HFE function is written with elastic energy, internal energy, entropy, and transformation hardening terms followed by a total dissipation inequality of the Claussius-Duhem type. Evolution equations are derived from the free energy and dissipation expressions (acting similar to flow rules) and then optimization conditions (Kuhn-Tucker) as well as limit functions are added (behaving like yield surfaces).

Simulations of the tension and actuation tests show good correlation (Figure 152).

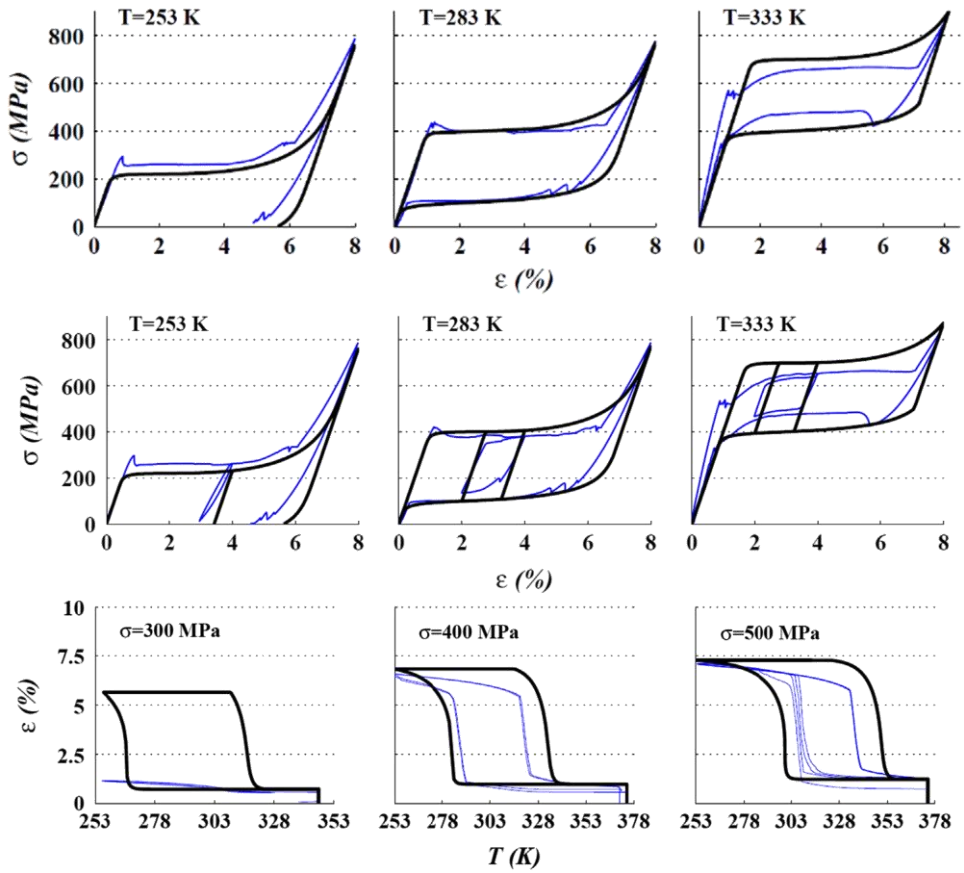


Figure 152: Uniaxial tension and actuation test simulations from Bodaghi et al. (2013) compared to experiments from the *RR* program

For the tension-torsion cases (Figure 153), the model …

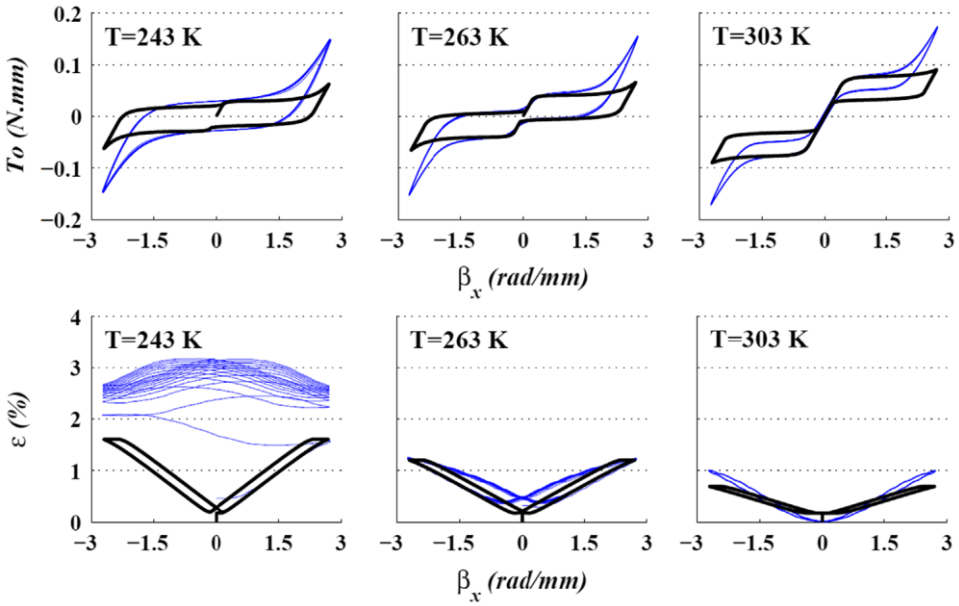


Figure 153: Tension-torsion test simulations from Bodaghi et al. (2013) compared to experiments from the *RR* program

A version of this model was described in Auricchio et al. (2009), which was as a conference paper for the *RR* program, but no simulation results were reported in the initial paper. In 2014, the work is continued on a model suitable for the type of behavior observed in the *RR* experimental program. In this model, three internal variables are used to represent SMA phases: austenite, single-variant martensite, and multivariant martensite. A HFE is

written as a function of total strain, temperature, phase fractions, and a transformation direction tensor. A chemical free energy is also written for each phase along with a set of interaction energies, which influence the behavior when multiple phases are present. Partial derivatives of the HFE are used to derive constitutive equations and thermodynamic forces. The driving force also have logical terms based on Kuhn-Tucker conditions and the transformation direction is assumed to be the direction of the deviatoric stress vector. Each driving force, along a consistency parameter and limit function, is used to derive an evolution equation which also include logical and optimization (Kuhn-Tucker) conditions. Next, some physical assumptions are discussed which simplify the model and some solution techniques are discussed.

As set of simulations shows good agreement with the *RR* uniaxial tension tests; Figure 154 shows several cases with a partial unloading loop.

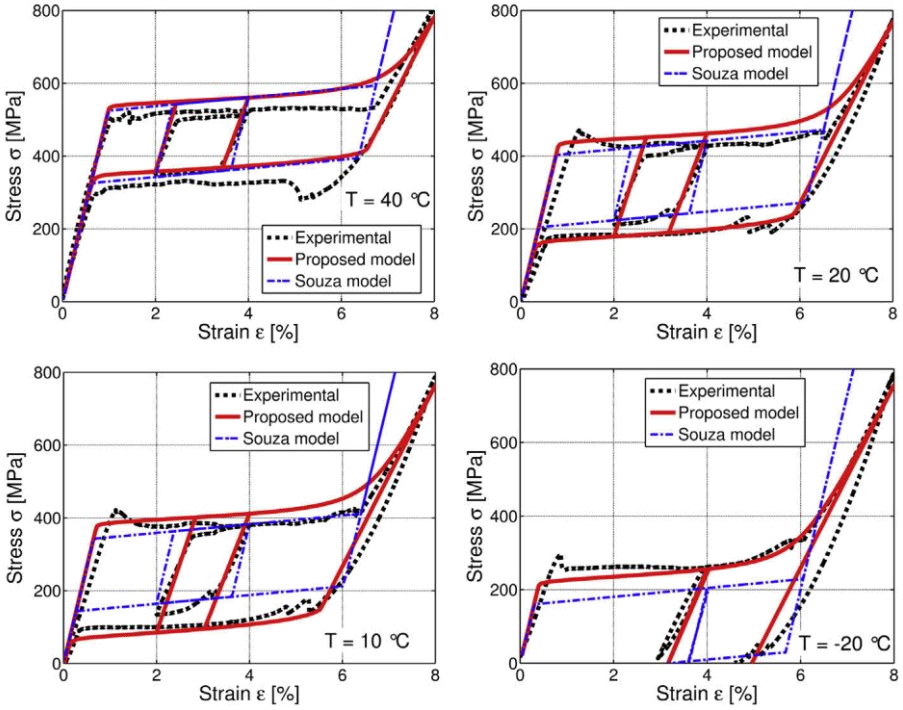


Figure 154: Uniaxial tension test simulations from Auricchio et al. (2014) compared to experiments from the *RR* program

The model captures the actuation response at the higher levels of axial stress (Figure 155, right) but appears to not capture the stress threshold where actuation does not occur (Figure 155, left).

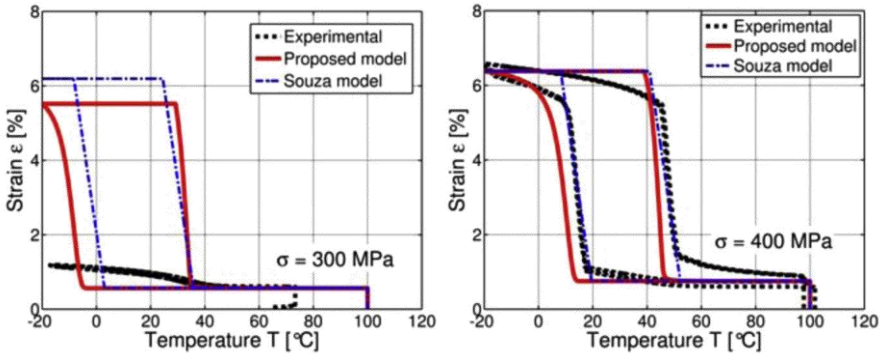


Figure 155: Actuation tension test simulations from Auricchio et al. (2014) compared to experiments from the *RR* program

The model also reasonably matches the stress-strain and stress-temperature response for the fixed-strain thermal cycle experiments (Figure 156).

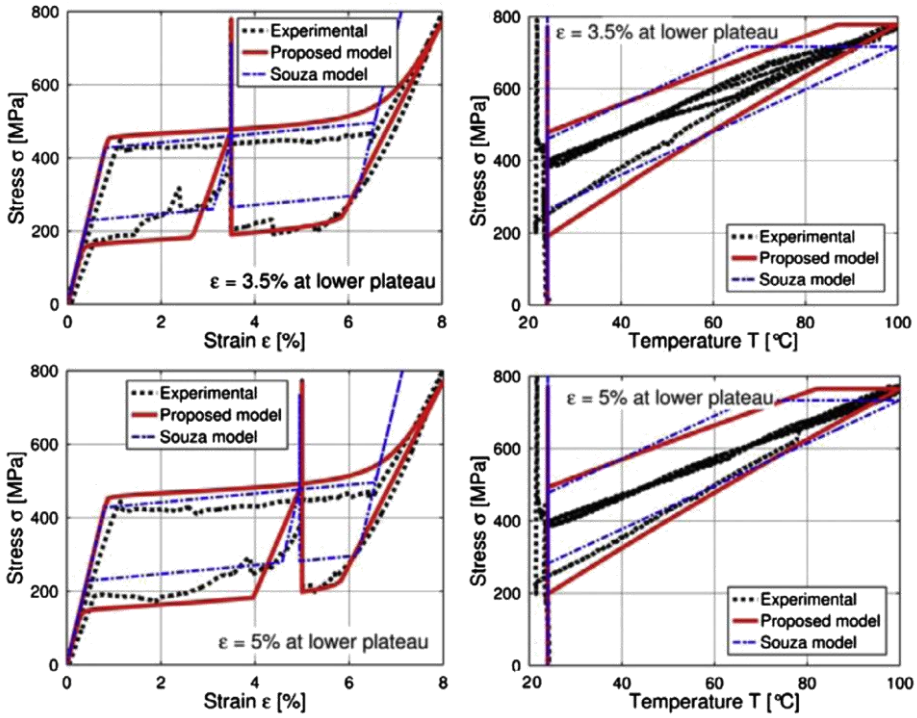


Figure 156: Fixed-strain thermal cycle test simulations from Auricchio et al. (2014) compared to experiments from the *RR* program

Only a small sampling of the tension-torsion experiments was included in the paper. For the cases shown, the agreement between the model and the test data is promising but imperfect (Figure 157).

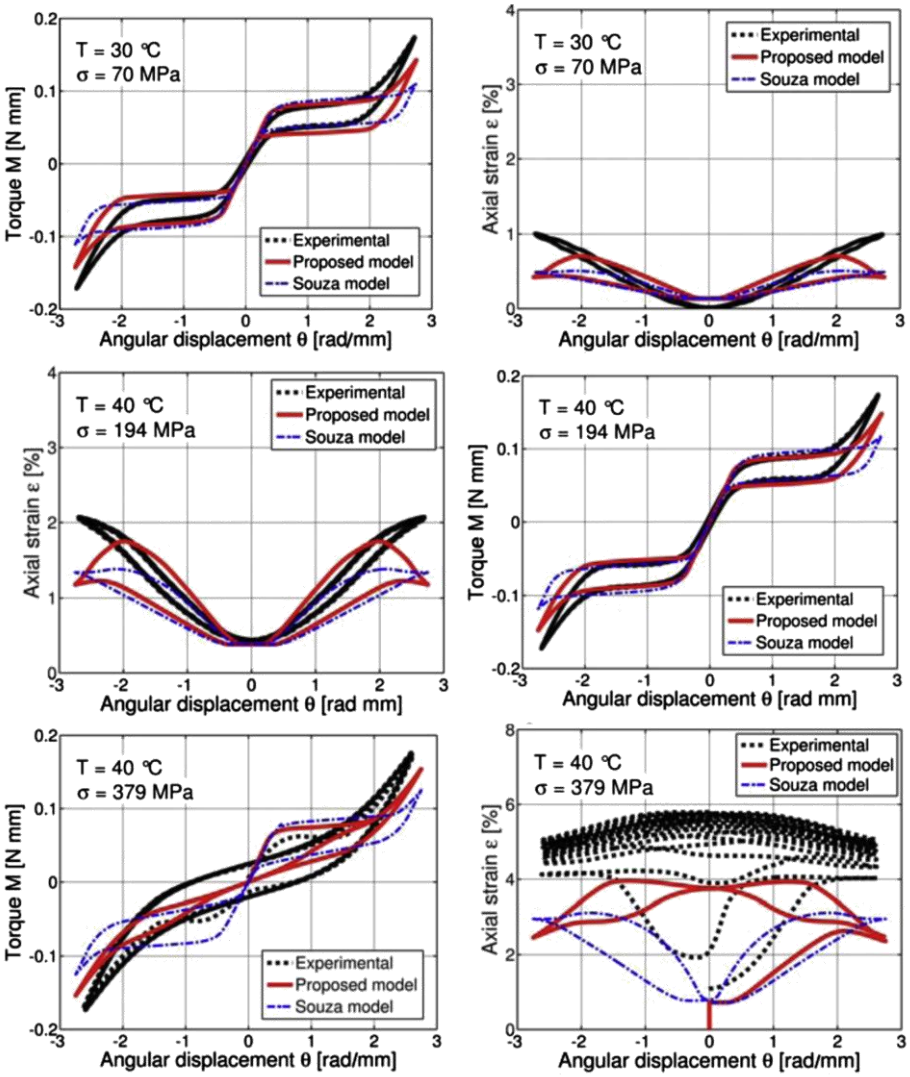


Figure 157: Tension-torsion test simulations from Auricchio et al. (2014) compared to experiments from the *RR* program

The torque-twist plots at lower levels of axial stress (70 MPa and 194 MPa) show good agreement, though the twist is slightly over-predicted. In the 379 MPa torque-twist plot, the model follows the experimental data in a general way, but the curve never forms the single hysteresis loop. The axial strain-twist plots show a similar characteristic shape to the test data and show that the axial strain increases with axial stress (though it is under-predicted for the 379 MPa case).

In Chatziathanasiou et al. (2016), the authors present a plasticity type of model that considers forward transformation, reverse transformation, and reorientation. The transformation strain parameter represents the average strain in a martensite RVE, where the strain direction can change (reorient) and the phase fraction can grow or shrink. A rate of change for each internal variable is derived from a GFE and a set of yield surfaces is used to define the activation criteria. A Claussius-Duhem expression is written to add some additional constraints to the system and maximum transformation strain parameters are used to allow transformations to complete.

Simulations of a small sample of the RR tension experiments are presented showing that the model captures superelasticity, pseudoplasticity, as well as a transition between the two in Figure 158.

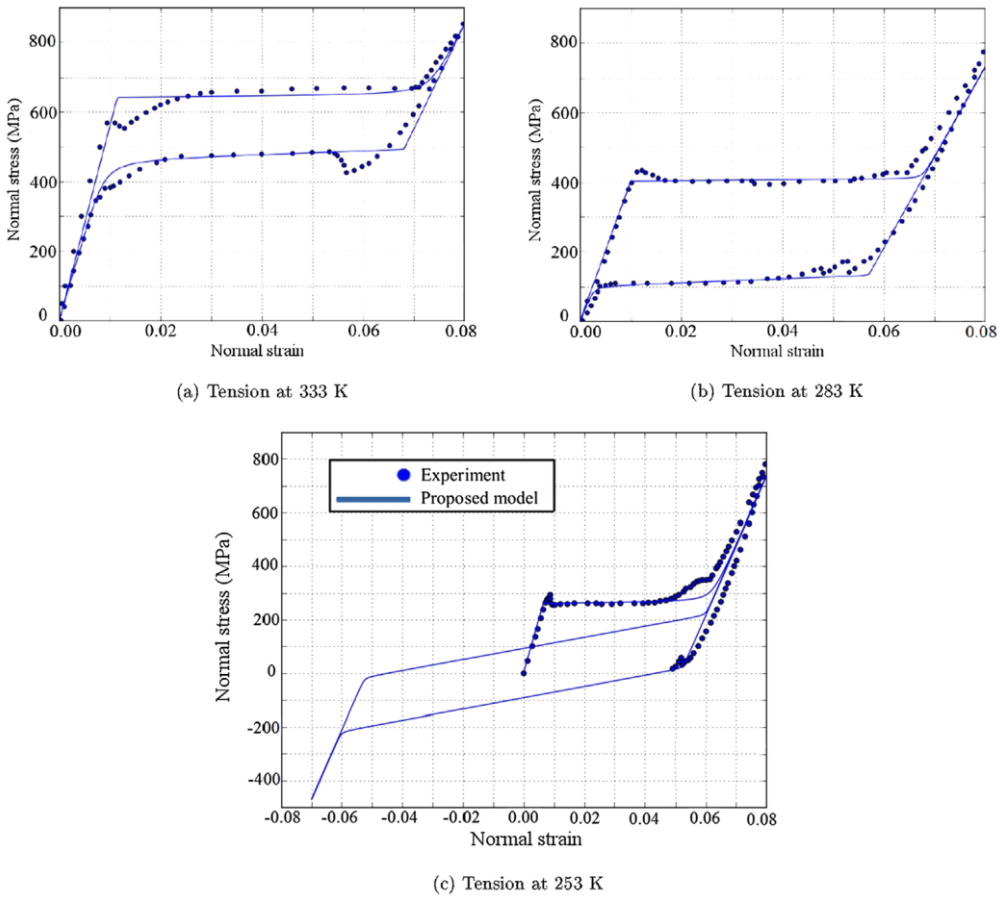


Figure 158: Uniaxial tension test simulations from Chatziathanasiou et al. (2016) compared to experiments from the *RR* program

The second cycle in 158c has a different characteristic shape than the first cycle. The experimental data does not include a second cycle so no evaluation can be made on the model's accuracy in this context. It is worth noting, however, that the characteristic shape of the second cycle is qualitatively similar to the second cycles in the low temperature tension-torsion experiments from the RR program (Figures 77 and 78).

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form in einem Verfahren zur Erlangung eines akademischen Grades vorgelegt.

Ort, Datum

Cary, NC, USA, 2020-11-01

Unterschrift