# A 3-D constitutive model for shape memory alloys incorporating pseudoelasticity and detwinning of self-accommodated martensite

Peter Popov [b] and Dimitris C. Lagoudas [a,*]

[a] *Department Of Aerospace Engineering, Texas A&M University, TX 77843-3141, USA*

[b] *Institute for Scientific Computation, Texas A&M University, TX 77843-3404, USA*

**Abstract**

A 3-D constitutive model for polycrystalline Shape Memory Alloys (SMAs), based on a modified phase transformation diagram, is presented. The model takes into account both direct conversion of austenite into detwinned martensite as well as the detwinning of self accommodated martensite. This model is suitable for performing numerical simulations on SMA materials undergoing complex thermomechanical loading paths in stress-temperature space. The model is based on thermodynamic potentials and utilizes three internal variables to predict the phase transformation and detwinning of martensite in polycrystalline SMAs. Complementing the theoretical developments, experimental data is presented showing that the phase transformation temperatures for the self accommodated martensite to austenite and detwinned martensite to austenite transformations are different. Determination of some of the SMA material parameters from such experimental data is also discussed. The paper concludes with several numerical examples of boundary value problems with complex thermomechanical loading paths which demonstrate the capabilities of the model.

*Key words:* Shape Memory Alloy, reorientation, detwinning

## 1 Introduction

Shape Memory Alloys (SMAs) are metallic alloys that can undergo martensitic phase transformations as a result of applied thermomechanical loads and are capable of recovering apparently permanent strains when heated above a certain temperature. At high temperatures the crystal lattice is in a high symmetry, parent austenitic phase, while at lower temperatures

* Corresponding author. Tel: 1-979-845-1604; fax: 1-979-845-6051

*Email addresses:* ppopov@tamu.edu (Peter Popov), lagoudas@aero.tamu.edu (Dimitris C. Lagoudas).

a low symmetry martensitic phase appears (Otsuka and Wayman, 1999; Otsuka and Ren, 2005). The key characteristic of all SMAs is the occurrence of a martensitic phase transformation from the austenitic phase to the different variants of the low temperature, low symmetry martensitic phase. The martensitic transformation is a shear-dominant diffusionless solid-state phase transformation occurring by nucleation and growth of the martensitic phase from the parent austenitic phase (Olson and Cohen, 1982). What make SMAs remarkably different from other materials are primarily the *Shape Memory Effect* (SME) and *Pseudoelasticity*, which are associated with the specific way the phase transformation occurs.

When a shape memory alloy undergoes a martensitic phase transformation, it transforms from the parent phase to one or more of the different variants of the martensitic phase (Otsuka and Wayman, 1999). In the absence of applied stresses, the variants of the martensitic phase usually arrange themselves in a self-accommodating manner through twinning, resulting in no observable macroscopic shape change. By applying mechanical loading the martensitic variants are forced to reorient (detwin) into a single variant leading to large macroscopic inelastic strains. After heating above a certain temperature, the martensitic phase returns to the austenitic phase, and the inelastic strains are recovered. This behavior is known as the SME (Otsuka and Wayman, 1999). Pseudoelasticity is observed when the martensitic phase transformation is induced by applied thermomechanical loading of the austenitic phase in which case detwinned martensite is directly produced from austenite. The process is again associated with large inelastic (transformation) strains which are recovered upon unloading due to the reverse phase transformation (Wayman, 1983; Otsuka and Wayman, 1999). The extensive list of alloys exhibiting SME and pseudoelasticity includes the Ni-Ti alloys, and many copper-, iron-, silver- and gold-based alloys (Nishiyama, 1978).

During the last two decades the area of constitutive modeling of *polycrystalline* SMAs has been a topic of many research publications and significant advancements have been reported. One major class of SMA constitutive models is the phenomenological one, which relies on continuum thermodynamics with internal state variables to account for the changes in the microstructure due to phase transformation (Tanaka et al., 1992; Patoor et al., 1988; Ortin and Planes, 1988, 1989; Berveiller et al., 1991; Liang and Rogers, 1992; Sun et al., 1991; Sun and Hwang, 1993a,b; Graesser and Cozzarelli, 1991; Brinson, 1993; Raniecki and Lexcellent, 1994; Lagoudas et al., 1996; Marketz et al., 1995; Leclercq and Lexcellent, 1996; Juhasz et al., 2002; Bo and Lagoudas, 1999a,b,c; Lagoudas and Bo, 1999; Lexcellent et al., 2000; Lagoudas and Entchev, 2004). These type of models usually assume a macroscopic energy function that depends on state and internal variables used to describe the degree of phase transformation. Evolution equations are then postulated for the internal variables. Most phenomenological constitutive models adopt such a thermodynamic structure and select the martensitic volume fraction as an internal state variable to account, on the average, for the influence of the microstructure.

The early constitutive models (Tanaka, 1986; Tanaka et al., 1986; Sato and Tanaka, 1988; Tanaka et al., 1995; Liang and Rogers, 1990, 1992; Brinson, 1993; Boyd and Lagoudas, 1994, 1996a,b) have been used to derive the pseudoelastic response of SMAs and their main focus is the hardening function selected to model the stress-strain response during the stress induced martensitic phase transformation. A unified framework for these early constitutive models has been presented by Lagoudas et al. (1996). Further improvements in the accuracy of SMAs models were achieved by Raniecki and Lexcellent (1998); Qidwai and Lagoudas

(2000b); Lexcellent et al. (2002), who proposed different transformation functions in order to capture the asymmetric response that SMAs exhibit in tension and compression. Qidwai and Lagoudas (2000b) also studied the implications of the principle of maximum dissipation during phase transformation on the transformation surfaces and evolution equations for the martensitic volume fraction. These models are suitable for stress induced martensitic transformations or for SMAs which have already been trained to exhibit the Two-Way Shape Memory Effect (TWSME), (for details on the TWSME, see Otsuka and Wayman, 1999). Modifications, such as making the total amount of transformation strain dependent on the stress, have been introduced by some authors (Bo and Lagoudas, 1999a,b,c; Lagoudas and Bo, 1999; Lagoudas and Entchev, 2004) in order to account for certain aspects of the SME. These changes however do not make an explicit distinction between twinned and detwinned martensite and thus are only suitable for SME related thermomechanical loading paths.

The analysis of the existing models and their comparison to the experimental results has shown that current SMA constitutive models which take into account the development of stress-induced martensite have reached a high level of sophistication. However, such models generally lack the ability to handle other loading paths involving detwinning and reorientation of martensite in conjunction with the pseudoelastic response. Therefore, there is need for a 3D constitutive model that can accurately capture not only the material response during pseudoelastic and SME loading paths, but also loading paths that involve co-existence of all the three material phases - austenite, twinned (self-accommodated) martensite and detwinned martensite. Such a model should also be implemented numerically and tested on a comprehensive set of model problems. This will allow to perform numerical simulations of problems of varying engineering difficulty, such as the actuation of SMA micro-grips (Kohl et al., 2002), the cooling/heating cycles in the manufacturing and deployment of biomedical devices (Jung et al., 2004), temperature actuated flow regulating devices (Popov, 2005) and fuel powered SMA actuators (Jun et al., 2006), to name a few.

The early attempts to combine the pseudoelastic material response with detwinning of martensite were done in one dimension by Brinson (1993); Brinson and Lammering (1993); Boyd and Lagoudas (1993). These models used two internal state variables to model pseudoelasticity and detwinning. In addition, Brinson (1993) used a uniaxial phase diagram in stress-temperature space which conveniently defines the thermodynamically stable domains for the three phases and the possible transformations between them. The work was further refined by Bekker and Brinson (1997, 1998) who incorporated minor loops for the pseudoelastic transformation. However, this basic phase diagram does not account for certain loading paths, especially those that traverse the regions where the three phases can co-exist.

Three dimensional thermodynamics based models of combined detwinning and pseudoelasticity have been proposed by Leclercq and Lexcellent (1996); Lagoudas and Shu (1999); Juhasz et al. (2002). The models of Leclercq and Lexcellent (1996); Lagoudas and Shu (1999) used two scalar volume fractions for twinned and detwinned martensite. While formulated in 3D, they were implemented and tested only on 1D examples. Furthermore, complex loading paths which involve a mixture of the three phases were also not tested. The model of Juhasz et al. (2002) used the entire transformation strain as a tensorial internal variable instead of the volume fraction of detwinned martensite. All three models used phase diagrams which were based on the work of Brinson (1993). While attempts were made to overcome some of its basic limitations, the current work attempts to present an extended phase diagram which

refines existing concepts and also incorporates new experimental results.

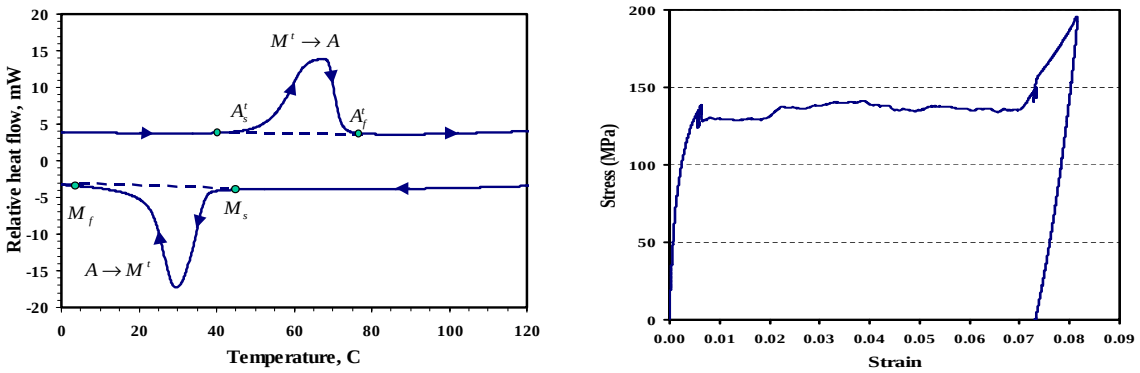
In this paper, a three-dimensional, thermodynamics based model with three internal variables is formulated for the simultaneous modeling of pseudoelasticity and detwinning of self-accommodated martensite in polycrystalline SMAs. The model is consistent with an extended uniaxial phase diagram. The novel characteristics of this model are: *(i)* integration into the phase diagram of new experimental results which demonstrate that twinned and detwinned martensite transform to austenite at different temperatures; *(ii)* refinement of the phase diagram with respect to loading paths that involve a mixture of the three phases; *(iii)* the use of three independent internal variables (in contrast to the usual two, typically used in this class of models) which provides a new approach to modelling the training of SMA materials and the associated evolution of the phase diagram; *(iv)* numerical implementation which tests complex loading paths, including ones that involve a mixture of the three phases.

The paper begins with experimental results which demonstrate that, at zero stress, twinned and detwinned martensite transform to austenite at different temperatures (Section 2). The phase diagram is constructed in Section 3 based on these observations, as well as a careful reexamination of published experimental data on detwinning of twinned martensite and the conversion of twinned martensite to austenite. The 3-D constitutive model is presented in Section 4. A discussion of how to identify the material parameters used in the model from experimentally observable quantities is given in Section 4.5. Finally the numerical implementation of the model into a displacement based Finite Element Method (FEM) code is presented in Section 5 and numerical examples are given in Section 6.

To simplify the presentation, throughout this paper the three phases are denoted by $A$, $M^t$ and $M^d$ for austenite, twinned martensite and detwinned martensite, respectively. The five possible phase transformations are denoted by $A \to M^t$, $A \to M^d$, $M^t \to A$, $M^d \to A$ and $M^t \to M^d$ for austenite to twinned martensite, austenite to detwinned martensite, twinned martensite to austenite, detwinned martensite to austenite and twinned to detwinned martensite, respectively. The detwinning of twinned martensite $M^t \to M^d$ does not involve phase transformation and is, in fact, an inelastic deformation process of reorientation of martensitic variants (c.f. e.g. Otsuka and Wayman, 1999). For the sake of simplicity, the collective term transformations is applied to it whenever the distinction is not important. Note also, that the transformation $M^d \to M^t$ from detwinned to twinned martensite (the so called rubber-like effect, c.f. Otsuka and Wayman (1999)) is not thermodynamically stable and it is not considered. Finally, the critical start and finish transformation temperatures at zero stress level (c.f. Figures 3 and 4) are denoted as follows: $M_s$ and $M_f$ for the $A \to M^t$ transformation, $A_s^t$ and $A_f^t$ for the $M^t \to A$ transformation and $A_s^d$ and $A_f^d$ for the $M^d \to A$ transformation. The clarification that these temperatures are at zero stress level will be omitted, and only the term transformation temperatures will be used.

## 2 Experiments on the transformation temperatures of $M^t \to A$ and $M^d \to A$.

In a recent paper, Sakamoto (2002) questioned the assumption made by many researchers that, at zero stress, the transformation temperatures for $M^t \to A$ and $M^d \to A$ coincide. He introduced the concept of shape change stress which is a local stress field generated at interface between twinned martensitic variants and the surrounding matrix. In stress induced martensite this elastic stress field is absent, and a detailed analysis of the magnitude of




(a) Initial DSC test (b) Mechanical loading

Fig. 1. Results from quasistatic mechanical testing of annealed 2.16mm NiTi wire followed by DSC test. The transformation temperatures for $A \leftrightarrow M^d$, as annealed, are obtained from an initial DSC test (a). A mechanical loading is then performed (b).

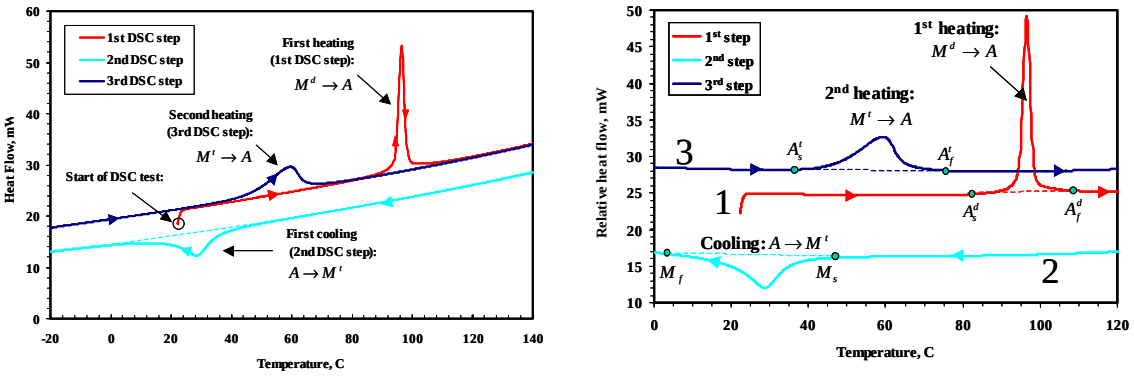
this shape change stress with respect to specimen and martensitic plate sizes leads to the conclusion of different transformation temperatures for twinned and detwinned martensite. In this section, mechanical testing combined with calorimetric measurements are used to confirm this idea.

### *2.1 Setup and experimental procedure*

A 2.16 $mm$ diameter $Ni_{50.3}Ti_{49.7}$ wire was used in the experiment. Two specimens were annealed at 800 $°C$ for 30 $min$, slowly cooled to 0 $°C$, and then brought to room temperature (22 $°C$). Differential scanning calorimetry (DSC) measurements in a Perkin-Elmer Pyrus 1 apparatus were performed in order to establish the transformation temperatures for the $A \rightarrow M^t$ and $M^t \rightarrow A$ transformations and characterize the material state after the annealing. It was found that the transformation temperatures were $M_s = 45\ °C$, $M_f = 3\ °C$, $A_s^t = 40\ °C$ and $A_f^t = 76\ °C$. Since the austenitic start temperature was well above room temperature, it was concluded that, after the heat treatment the wire was entirely in the $M^t$ state. Note, that the transformation temperatures $A_s^d$ and $A_f^d$ (assumed different from $A_s^t$ and $A_f^t$) cannot be determined from a DSC sweep which involves only the $A \rightarrow M^t$ transition. The remainder of this section details the measurement of $A_s^d$ and $A_f^d$ for this SMA material. It will be shown that they are substantially different from the $M^d \rightarrow A$ temperatures.

After establishing the transformation temperatures for the $A \leftrightarrow M^t$ transition, the two specimens were mechanically loaded at room temperature in a MTS 801 loading frame (Figure 1(b)). Due to the initial state of the specimens ($M^t$), the self-accommodated martensite underwent the detwinning ($M^t \rightarrow M^d$) deformation. Upon elastic unloading, large inelastic strain of about than 7.2% was observed, implying a detwinned material state ($M^d$). Note that there was no strain recovery during unloading, indicating the $A_s^d$ (to be determined by subsequent DSC testing) is higher than room temperature. In order to quantify the amount of inelastic strain due to detwinning of $M^t$ and the amount due to plastic deformations the first specimen was heated to about 150 $°C$. During the process about 5% of the inelastic strain was recovered, indicating that it was due to detwinning and the remaining 2% is due to plastic deformations.

The second specimen, immediately after unloading and hence entirely in the $M^d$ state, was




(a) DSC test sequence results

(b) Latent heat of transformation

Fig. 2. Results form DSC testing sequence of an untrained NiTi wire performed immediately after a single mechanical loading. The initial state of the wire is $M^d$. The wire is first heated, revealing the transformation temperatures for the $M^d \rightarrow A$, followed by cooling during which the wire undergoes $A \rightarrow M^t$ transformation, followed by a second heating which shows the transformation temperatures for $M^t \rightarrow A$.

subjected to further DSC testing (Figure 2), described below. Care was taken to prepare the DSC sample so that the material state ($M^d$) achieved at the end of the mechanical unloading step was not altered in the sample preparation process, that is, the specimen was always kept at room temperature, which is below $A_s^d$.

A total of five thermal loading steps were executed in the DSC apparatus. The actual heat flow observed in the specimen during the first three DSC steps is shown in Figure 2(a). The corresponding latent heat is shown in Figure 2(b). The specimen was first heated from room temperature (the temperature at which the mechanical test was performed) to 200 $^\circ C$. The first signs of the forward $M^d \rightarrow A$ transformation were observed at $A_s^d = 82\ ^\circ C$, the peak of the transformation was at approximately 96 $^\circ C$ and the transformation ended at approximately $A_f^d = 108\ ^\circ C$. At this point the sample was in the austenitic phase. The sample was then cooled from 200 $^\circ C$ to $-60\ ^\circ C$. During the cooling a single peak was observed at approximately 28 $^\circ C$, corresponding to the $A \rightarrow M^t$ transformation. Note that, due to the nature of a DSC test, the sample always remains stress free. The beginning of the reverse transformation indicated $M_s = 47\ ^\circ C$ and $M_f = 3\ ^\circ C$, which is consistent with the first DSC test performed before the wire was subjected to mechanical loading. A repeatability in the $A \rightarrow M^t$ temperatures was therefore observed. The third thermal loading step was again heating from $-60\ ^\circ C$ to 200 $^\circ C$. The transformation temperatures were markedly different from the first heating step: $A_s^t = 35\ ^\circ C$, $A_f^t = 76\ ^\circ C$ with the peak at 59 $^\circ C$. At the beginning of this step the sample was entirely in the $M^t$ state (the natural state after the DSC cooling step), therefore the transformation temperatures correspond to the $M^t \rightarrow A$ transformation. Two more loading steps, not shown on Figure 2, were performed. These included an additional cooling and a heating cycle. Due to the stress free state of the SMA, the transformations involved were $A \rightarrow M^t$ and $M^t \rightarrow A$, respectively. Results were close to those from the second (cooling) and third (heating) cycles, respectively, indicating repeatability of the $A \rightarrow M^t$ transformation temperatures. The later are substantially different from the $M^d \rightarrow A$ temperatures.

The same type of mechanical loading followed by the above sequence of DSC tests was

performed for SMA materials with different annealing history and pseudoelastic training. In all cases similar results of markedly different critical temperatures for the $M^t \rightarrow A$ and $M^d \rightarrow A$ were observed (Popov, 2005). The simplest conclusion from these experiments is that the $M^t \rightarrow A$ and $M^d \rightarrow A$ transformation temperatures at zero stress are, generally, different. A qualitative explanation for this results can be done as follows: the twinned martensite, requires some energy input to transform back to austenite. The detwinned martensite, also requires this energy input, but in addition it also needs more energy in order to reverse the inelastic strains which are present (note this always happens in the presence of a local stress field, even when its macroscopic average is zero). Thus, the reverse phase transformation occurs at higher temperatures, compared to twinned martensite. The theoretical study by Sakamoto (2002) arrives at the same conclusion with the help of microstructural arguments and by analyzing the local stresses around the martensite/austenite interfaces which are different for twinned and detwinned martensite. These experimental results motivate a re-examination of the commonly used SMA phase diagram (next section) and the proposed constitutive model takes into account the different temperatures $A_s^t$, $A_s^d$, $A_f^t$ and $A_f^d$.

## 3 Modified SMA phase diagram

The phase transformations from austenite to martensite as well as the detwinning of self-accommodated martensite occur due to thermomechanical loading. A convenient way of describing general thermomechanical loading paths leading to the different transformations is to use a phase diagram in stress-temperature space (Figures 3 and 4). Such phase diagrams include the stable domains of $A$, $M^t$ and $M^d$ in stress-temperature space as well as transformation strips in which the various transformations take place. The proposed SMA model is based on the 1-D phase diagram shown in Figure 4. This phase diagram incorporates both the new data presented in Section 2 as well as certain modifications in comparisons with other works. The aim is two-fold: first, to take into account the different critical transformation temperatures for the $M^t \rightarrow A$ and $M^d \rightarrow A$ transformation, which has not been considered previously; second, to define the transformation strips, in agreement with available experimental data, so that non-physical behavior is eliminated for all possible loading paths. In this section the proposed phase diagram is presented and compared with other common choices in the literature (an example is the diagram in Figure 3).

Several SMA models which attempt to take into account both the development of $M^t$ and $M^d$ (c.f. e.g. Brinson, 1993; Leclercq and Lexcellent, 1996; Lagoudas and Shu, 1999; Bekker and Brinson, 1997; Juhasz et al., 2002) use phase diagrams. The phase diagram shown in Figure 3 was used by Brinson (1993) and works well for pure pseudoelastic paths (path 1 on the figure, no $M^t$ is ever produced) and pure SME paths (path 2, complete $M^t \rightarrow M^d$ transformation, stress is zero during heating). However, as more complicated loading paths are considered (3a,b,c, for example) certain non-physical behavior becomes possible, mostly in the intermediate regions where a mixture of the three phases can exist. In particular, there is no agreement how the $M^t \rightarrow M^d$ strip looks in that intermediate region and what is the shape of the $M^t \rightarrow A$ strip. Depending on the thermomechanical loading paths of interest different assumptions and modifications are used by subsequent studies (Leclercq and Lexcellent, 1996; Lagoudas and Shu, 1999; Juhasz et al., 2002). As a result there is no unambiguous understanding of how the phase diagram should look like.

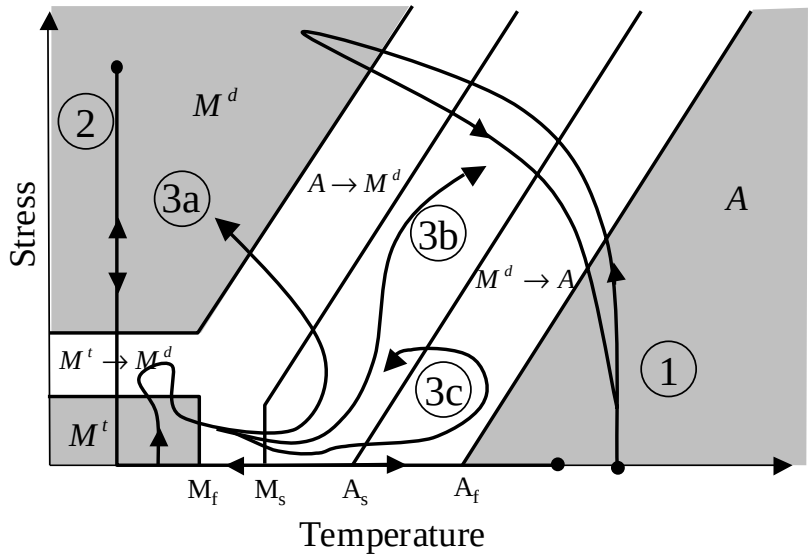



Fig. 3. A typical Phase Diagram for SMA materials

The phase diagram proposed in this work (Figure 4) follows the established literature in assuming three regions where only the pure phases $A$, $M^t$ and $M^d$ can exists (c.f. Figure 3). These regions are shaded and labeled $A$, $M^t$ and $M^d$, respectively. The three regions are separated by transformation strips which are labeled according to the transformations ($A \rightarrow M^t$, $A \rightarrow M^d$, $M^t \rightarrow A$, $M^d \rightarrow A$, $M^t \rightarrow M^d$) which take place. Note that some of these strips overlap and in an overlap region multiple transformations are possible. In the non-shaded region of the phase diagram various mixtures can exist. The critical temperatures for the start and finish of the $A \rightarrow M^t$ transformation are denoted by $M_s$ and $M_f$. Based on the experimental results of Section 2, the critical start and finish temperatures at zero stress for the $M^t \rightarrow A$ transformation are denoted by $A_s^t$ and $A_f^t$. They are assumed different from the corresponding critical temperatures at zero stress for the $M^d \rightarrow A$ transformation which are denoted by $A_s^d$ and $A_f^d$. The start and finish lines for the forward and reverse transformations $A \leftrightarrow M^t$ are vertical and pass through the critical temperatures $M_s$, $M_f$, $A_s^t$, $A_f^t$, respectively. The start and finish lines for the reverse strip $M^d \rightarrow A$ pass through the critical temperatures $A_s^d$ and $A_f^d$ and exhibits a temperature dependence, defined by the positive slope $k$. The critical uniaxial start and finish stresses at $T = M_s$ required for detwinning of twinned martensite ($M^t \rightarrow M^d$) are denoted by $\sigma_s$ and $\sigma_f$, respectively. The transformation strip $M^t \rightarrow M^d$ exhibits a mild temperature dependence characterized by a negative slope $k^d$. The start and finish lines for the forward $A \rightarrow M^d$ transformation exhibits the same temperature dependence as the reverse transformation $M^d \rightarrow A$. The finish line for $A \rightarrow M^d$ passes through or below the point $(M_s, \sigma_f)$.

The modifications of this phase diagram compared to, for example, the one by Brinson (1993), are several. First, and most importantly, based on the experimental results of the previous Section 2, the critical start and finish temperatures at zero stress for the $M^t \rightarrow A$ are assumed different from the corresponding critical temperatures ate zero stress for the $M^d \rightarrow A$ transformation. Secondly, the $M^t \rightarrow M^d$ strip is a single, well-defined strip for the entire temperature range $T < A_f^t$. The original work of Brinson (1993) assumes that the detwinning strip $M^t \rightarrow M^d$ has the same form as the one assumed here for temperatures $T < M_s$ but coincides with the strip for stress induced martensite $A \rightarrow M^d$ at temperatures $T > M_s$. This can lead, for example, to the existence of twinned martensite at high stress

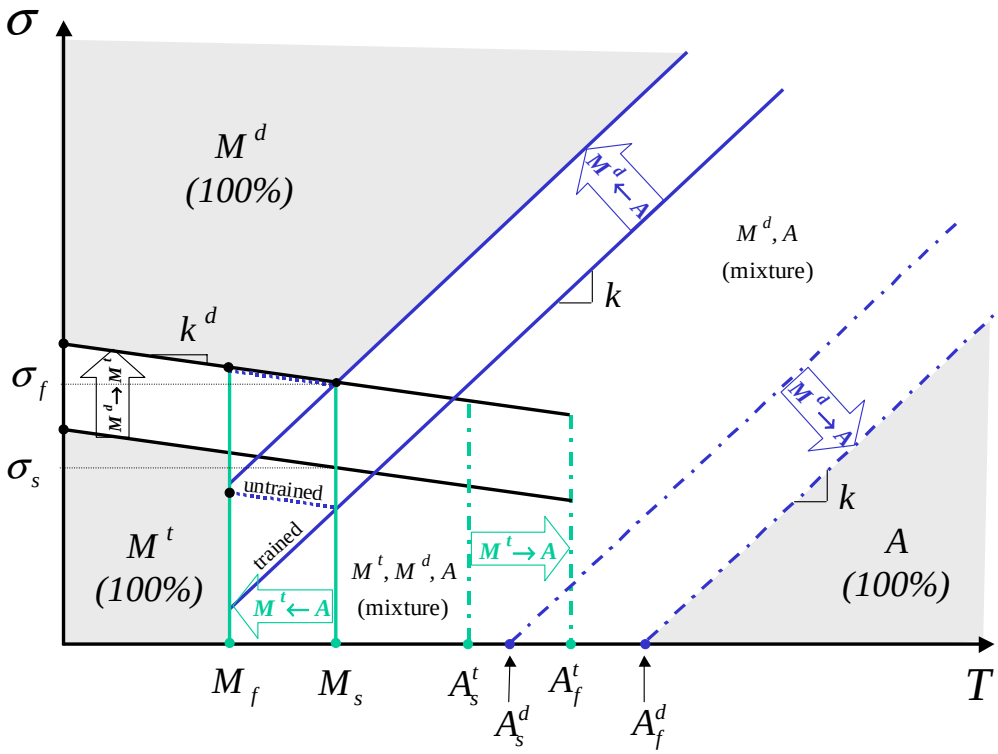



Fig. 4. The SMA phase diagram used in this work. All the three pure phase regions ($A$, $M^d$ and $M^t$) are enclosed by transformation strips. The diagram is completely defined by the respective transformation temperatures $M_s$, $M_f$, $A_s^t$, $A_f^t$, $A_s^d$, $A_f^d$, the critical stresses for detwinning $\sigma_s$ and $\sigma_f$ and the slopes $k$ and $k^d$.

levels (above $\boldsymbol{\sigma}_f$, c.f. path 3b in Figure 3), which is not physically realistic. This problem is critically examined in Section 3.2. With the help of the experimental study of Cross et al. (1969), it is shown that a single transformation strip extending to temperatures as high as $A_s^t$ and possibly to $A_f^t$, as done in this work (Figure 4), is a more natural assumption.

Secondly, there is a disagreement in the literature on the shape of the reverse $M^t \rightarrow A$ strip. In the work of Brinson (1993) and later papers, it is assumed to coincide with the $M^d \rightarrow A$ strip while other authors (Leclercq and Lexcellent, 1996; Lagoudas and Shu, 1999; Juhasz et al., 2002) have used a vertical $M^t \rightarrow A$ strip, which is independent of stress. An argument can be made (Section 3.1) that the latter is a more natural choice. Furthermore, there is an ambiguity in the definition of the $A \rightarrow M^d$ strip at low stresses and temperatures ($T < M_s$ and $\sigma < \sigma_s$). Some authors have extended it to zero-stress level (Bekker and Brinson, 1997), while others (Lagoudas and Shu, 1999) suggest, that in the region $T < M_s$ the dependence on temperature disappears and there is a critical stress below which $A \rightarrow M^d$ does not occur. There are two possible ways of completing it, depending mainly on the training history of the material. In this work, for trained materials, it will be assumed that the $A \rightarrow M^d$ transformation strip extends all the way to zero stress level. For untrained SMA materials a critical stress the respective lines in Figure 4 are labeled accordingly and the start and finish lines (below $M_s$) for the untrained case are also marked with a dashed line.

To fully define the phase diagram one also has to consider the relationship between the $A \rightarrow M^d$, $A \rightarrow M^t$ and $M^t \rightarrow M^d$ strips in the vicinity of $M_s$. The general assumption by most authors is that there exists a triple point $(M_s, \boldsymbol{\sigma}_s)$ where the three onset lines intersect

and another point $(M_s, \boldsymbol{\sigma}_f)$ where the three finish lines intersect (Brinson, 1993; Leclercq and Lexcellent, 1996; Lagoudas and Shu, 1999; Bekker and Brinson, 1997; Juhasz et al., 2002). The experimental evidence cited is usually inconclusive at drawing such a strong relation between the three transformation strips. However, if the finish line for $A \to M^d$ passes above the intersection point of the other two finish lines then one can find a particular isobaric cooling path which leads to jump discontinuities in the strain as the temperature is lowered. This is demonstrated in Section 4.5.2. Therefore one has to assume that the finish line for the $A \to M^d$ transformation passes through or is below that intersection of the other two lines. In the absence of sufficiently clear experimental data we assume that the $A \to M^d$ finish line is below the intersection of the other two (in Figure 4 the extreme case of a triple point is shown). This, along with other restrictions on the relative locations of the transformation strips which arise in the development of the thermodynamically consistent model are discussed in Sections 4.5.2. The remainder of this section presents a detailed description of the proposed extensions and modifications of the phase diagram of Figure 4

### *3.1 Austenite to martensite ($A \leftrightarrow M^t$, $A \leftrightarrow M^d$).*

An early observation in quasi-static isothermal loading tests was that the transformation surfaces for $A \leftrightarrow M^d$ exhibit a strong temperature dependence (Cross et al., 1969; Jackson et al., 1972; Otsuka and Wayman, 1999). These and many other experimental results show that the critical transformation stress required for initiation and completion of both the $A \to M^d$ and $M^d \to A$ forward and reverse transformations increase, more or less linearly, with increase in temperature. The reason for this dependence on temperature is the development of transformation strain during the transformation and the associated work expended by the SMA. The theoretical derivation of the precise functional dependence of the critical transformation stress for detwinning is based on a Clausius-Clapeyron relation (Wollants et al., 1979). After some simplifying assumptions such as equal stiffness and thermal expansion coefficient of austenite and martensite, a linear dependence on temperature is obtained (Wollants et al., 1979; Otsuka and Wayman, 1999). This has been observed consistently by many experimentalists ever since the work of Cross et al. (1969). Virtually any constitutive model for pseudoelastic SMA response, including the current work, takes this into account.

Unlike the $A \leftrightarrow M^d$ transitions, the phase transformation from $A$ to $M^t$ does not involve generation of macroscopic strains. At zero stress level, the $A \to M^t$ phase transformation begins when a critical temperature $M_s$ is reached and is completed when a second, and lower, critical temperature $M_f$ is reached. Due to the lack of transformation strain, a Clausius-Clapeyron argument suggests that there is no dependence of the critical temperatures $M_s$ and $M_f$ on stress. As a consequence one can expect that the transformation strip $A \to M^t$ is nearly vertical when plotted in the stress-temperature space (Figure 4). This fact has been used in most models that take into account the separate development of twinned and detwinned martensite (c.f. e.g. Brinson, 1993; Leclercq and Lexcellent, 1996; Bekker and Brinson, 1997; Lagoudas and Shu, 1999; Juhasz et al., 2002). There is however disagreement on what the shape of the reverse transformation strip $M^t \to A$ should be. Brinson (1993); Bekker and Brinson (1997) assume the same stress-temperature dependence as for the $M^d \to A$ transformation. Others (Leclercq and Lexcellent, 1996; Lagoudas and Shu, 1999; Juhasz et al., 2002) take the $M^t \to A$ strip to be stress independent.

There are not many experiments reported in the literature, which aim at determining the

shape of the $A \leftrightarrow M^t$ strips. Note that, due to lack of macroscopically observable mechanical quantities, such as inelastic strains, it is very difficult to experimentally detect the formation of twinned martensite under applied stress. Differential scanning calorimetry measurements, which are usually employed for revealing the transformation temperatures at zero stress level, cannot be directly used under applied stress. The two direct methods of measuring the progress of martensitic transformation under applied load that have been used by researchers are electrical resistivity measurements (Šittner et al., 2000; Kotil et al., 2003) and in-situ neutron diffraction measurements (Šittner et al., 2003). In both cases, sophisticated testing procedures in a precisely controlled thermal environment in a MTS-type testing frame are required. The focus of these and other direct measurement studies however was not the stress dependence of the critical temperatures for the $A \leftrightarrow M^t$ transformation.

An alternative indirect method, used specifically for determining the $M^t, M^d \rightarrow A$ transformation temperatures at nonzero stress levels during heating and cooling cycles has recently been performed by Tsoi et al. (2003). The experiment is done by first loading an SMA wire and embedding it in a epoxy matrix, as loaded. After the epoxy has cured it keeps the SMA deformed without the need for external apparatus. The composite can further be cut into small enough specimen, suitable for DSC measurements. The tests included pre-strain levels low enough that only $M^t \rightarrow A$ transformation can be expected during heating. While the DSC results are difficult to interpret conclusively, it can be inferred that the $M^t \rightarrow A$ temperatures do not depend on applied stress. Thus, due to the implications of lack of inelastic strains associated with the $M^t \rightarrow A$ transformation, and based on the experimental indications of Tsoi et al. (2003), in this paper, it will be assumed that both $M^t \rightarrow A$ and $A \rightarrow M^t$ are stress independent. In Section 6.1 a different indirect experimental method, based on the different stiffness of the pure martensitic and austenitic phases will be proposed.

### *3.2 Detwinning of self accommodated martensite ($M^t \rightarrow M^d$)*

The three pure phases regions ($A$, $M^t$ and $M^d$) are separated by transformation strips that indicate which transformation occurs ($A \rightarrow M^d$, $A \rightarrow M^t$, etc). In the original phase diagram of Brinson (1993) the transformation strip $M^t \rightarrow M^d$ is not defined at temperatures above $T > M_s$. If the initial conditions are such that $M^t$ is not present and once it is produced, the temperature is never increased beyond $M_s$, this will not cause problems. This is the case with a major class of SME paths where all the $M^t$ is depleted via the $M^t \rightarrow M^d$ deformation before the temperature is increased above $M_s$ (c.f. e.g. path 2 in Figure 3). Since these types of SME loading paths are quite important in characterization and testing of SMAs the possibility that $M^t$ may be present at temperatures in the range $M_s < T < A_f^t$ (for example by detwinning only part of the $M^t$) has generally been overlooked. Brinson (1993) has assumed for simplicity that the transformation strip for $M^t \rightarrow M^d$ coincides with the $A \rightarrow M^d$ strip in this temperature range. This assumption creates the inconvenience of having a concave transformation surface in stress-temperature space. Furthermore, at $T > M_s$ and high stress it is not clear how a single transformation surface can be used to determine the evolution of a two phase mixture which involves two transformation - $M^t \rightarrow A$ and $M^d \rightarrow A$. It can also be argued that the detwinning of martensite is an inelastic deformation process and does not involve change in the crystal lattice. Therefore the temperature dependence of the detwinning surface should not change drastically as suggested, that is, from slightly decreasing yield stress as the temperature is raised in the range $T < M_s$ to rapidly increasing with increase of temperature for $M_s < T < A_f^t$).

More importantly though, it does not seem to be supported by experimental results. A careful review of the pioneering work of Cross et al. (1969) suggests that it extends to temperatures higher than $M_s$. The reader is referred specifically to Figure 16 on page 26 of (Cross et al., 1969), which reports two sets of experiments. In both cases the material is loaded mechanically, under isothermal conditions at several different temperatures. The difference is that prior to the mechanical loading, in the first set, the material is cooled from high temperature and once the prescribed temperature is reached, it is fixed and the SMA is mechanically loaded. In the second set, the material is heated from low temperature, and then loaded. The initial yield stress is recorded in both cases. A look at the transformation temperatures reported by the authors shows that for the first set of experiments the initial material state is $A$, while for the second it is $M^t$. The later implies that the initial yield stress measured in the second set corresponds to the beginning of the $M^t \rightarrow M^d$ deformation over the entire range $T \leq A_s^t$. The results in the range $A_s^t \leq T \leq A_f^t$ cannot be easily interpreted since in this range the material before loading is a mixture of $A$ and $M^t$. The observed values for the critical stress for detwinning exhibit only very slight dependence on temperature, decreasing slowly as temperature is increased. Since for the first experimental set the initial material state is $A$, then a transformation surface for $M^t \rightarrow M^d$ can be inferred from the yield stress results only in the range $T \leq M_f$. Observe that the measured yield stress in the range $T \leq M_f$ for both sets of experiments is the same. This is a consistent experimental result since at these temperatures, the material for both experimental sets is pure $M^t$ when the loading begins.

Based upon this analysis, it is assumed in this work that the shape of the $M^t \rightarrow M^d$ has the same dependence on temperature, both for temperatures below and above $M_s$ (Figure 4). Note that the region of the phase diagram covered both by the $M^t \rightarrow M^d$ and $M^t \rightarrow A$ (to be discussed next) completely surrounds the region where pure $M^t$ can exist. Therefore there is no possibility that a loading path may lead to the existence of $M^t$ at high temperature or high-stress regions of the phase diagram.

### *3.3 Combined austenite to detwinned martensite ($A \rightarrow M^d$) at low stresses*

As was explained in the previous section, it is difficult to determine experimentally when the transformation to twinned martensite is occurring. Therefore another outstanding question, for which there is little experimental information, is what is the shape of the $A \rightarrow M^d$ surface at low stress $\sigma < \sigma_s$ and temperatures. In this region of the phase diagram it can be expected that both $A \rightarrow M^d$ and $A \rightarrow M^t$ occur. Note that the $A \rightarrow M^d$ is measured experimentally by observing the critical transformation stress required for the $A \rightarrow M^d$ transformation.

Bo and Lagoudas (1999a); Miller (2000) have measured the development of transformation strain during isobaric heating and cooling of annealed NiTiCu wires at different, constant, stress levels. Such a test can be represented by a horizontal line on the phase diagram and allows to determine the location of the $A \rightarrow M^d$ (during cooling) and $M^d \rightarrow A$ (during heating) transformation surfaces. The results for *untrained* specimen suggest that there $A \rightarrow M^d$ does not take place at stress levels below $40MPa$. They therefore argue that there is a critical stress level, below which detwinned martensite cannot form. This has usually been incorporated into SMA models (Brinson, 1993; Lagoudas and Shu, 1999) by assuming the $M^d \rightarrow A$ surface is independent of temperature below $T < M_s$.

This SMA behavior at low temperatures and stresses however is heavily influenced by the material composition, manufacturing process (e.g. cold work), heat treatments, etc. If a wire is trained for pseudoelastic regime, then development of transformation strain is observed even at zero stress level, which implies that the $M^d \to A$ surface should extend to zero stress. In order to take into account both types of behavior, the model developed here will include both the capability to proceed with the $A \to M^d$ transformation at arbitrary stress level and the possibility of a critical stress below which production of $M^d$ does not happen. In the first case, the $A \to M^d$ transformation strip would reaching zero stress (dotted line in Figure 4), while in the second, it becomes horizontal at $T < M_s$

## 4 Description of the constitutive theory

In order to simplify the presentation, the term "transformation" will be used to denote both the phase transformation from austenite to twinned and detwinned martensite as well as the detwinning deformation of self-accommodated martensite. We start with the volume fractions $c_i$, $i = 1, 2, 3$ of the self accommodated martensite $M^t$, stress-induced martensite $M^d$ and austenite $A$, respectively. The volume fractions are subject to the constraints:

$$c_1 + c_2 + c_3 = 1, \tag{1}$$

$$0 \leq c_i \leq 1, \quad \text{for} \quad i = 1, 2, 3. \tag{2}$$

While the state of the material is represented completely by the three volume fractions $c_i$, it is also useful to know how this state was achieved. To do this, the total amount $\xi_1$ of $M^t$ produced from $A$, the amount $\xi_2$ of $M^d$ produced from $A$, and the amount $\xi_3$ of $M^d$, produced from $M^t$, are introduced. They are connected to the three volume fractions $c_i$ by:

$$c_1 = c_{10} + \xi_1 - \xi_3, \tag{3}$$
$$c_2 = c_{20} + \xi_2 + \xi_3, \tag{4}$$
$$c_3 = c_{30} - \xi_1 - \xi_2, \tag{5}$$

where $c_{i0}$, $i = 1, 2, 3$ are the initial volume fractions of the three phases, subject to the constraint

$$c_{10} + c_{20} + c_{30} = 1.$$

These two representations of the phase state of the material are schematically portrayed in Figure 5. The two *phase* transformation $A \leftrightarrow M^t$ and $A \leftrightarrow M^d$ can proceed both ways, hence, $\dot{\xi}_1$, $\dot{\xi}_2$ can take arbitrary real values. The detwinning deformation $M^t \to M^d$ however is assumed irreversible, e.g. the rubberlike effect (see Otsuka and Wayman (1999) for definition) is not considered. Therefore $\dot{\xi}_3 \geq 0$. Observe that equations (3)-(5) automatically satisfy the constraint (1).

With these preliminary definitions, we now move to the question of selecting the inelastic internal variables. In this work, the internal variables that describe the phase state of the material are selected to be:

$$\xi_i, \quad i = 1, 2, 3. \tag{6}$$

This selection of three independent internal variables requires further discussion. A common choice in the literature is to select two of the three volume fractions $c_i$, say $c_1$ and $c_2$. Such

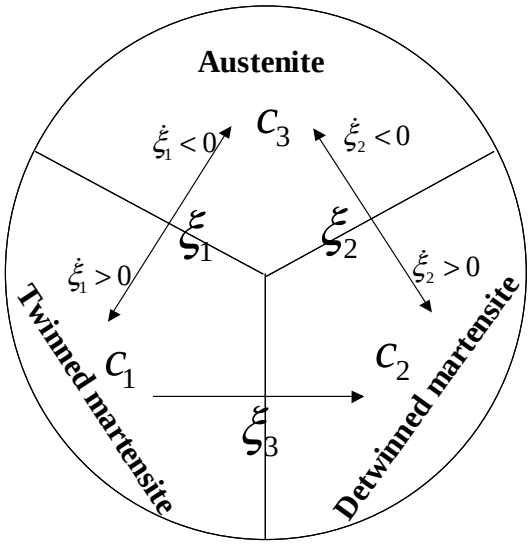



Fig. 5. Schematic of the three phases and the possible transitions between them.

a choice is appropriate when the phase diagram does not change during cyclic loading and when three simultaneous transformation cannot occur (for example $A \rightarrow M^d$, $A \rightarrow M^t$, $M^t \rightarrow M^d$). Observe that, if the three rates $\dot{c}_i$, $i = 1, 2, 3$ are known, equations (3)-(5) cannot be used to determine how exactly the three species transformed, that is, the rates $\dot{\xi}_i$, $i = 1, 2, 3$. Therefore, if three simultaneous transformations occur, one has to use the three rates $\dot{\xi}_i$ to have complete knowledge how the species transformed. This may be necessary, for example, during a simultaneous transformation $A \rightarrow M^d$, $A \rightarrow M^t$ and $M^t \rightarrow M^d$. In such a case the inelastic strains associated with the $A \rightarrow M^d$ and $M^t \rightarrow M^d$ processes may evolve differently from each other and therefor it is needed to keep track of the individual transformations. While some authors (Juhasz et al., 2002) argue that such situations should be forbidden, this in itself is an additional assumption that should not be excluded apriori.

Another reason to use $\xi_i$ is that they provide a complete loading history of the material and as such, can be used to account for various cyclic loading effects. For example, at any instance of time, $\xi_3$ denotes the total amount of detwinned martensite produced from self-accommodated martensite. So if one has a cyclic SME path and wants to account for plastic strains accumulated over all cycles, $\xi_3$ would be a suitable variable. Similarly, $\int_0^T |\dot{\xi}_2| dt$ is a suitable indicator the total amount of $A \leftrightarrow M^d$ transformation. The later variable has been used, for example, by Bo and Lagoudas (1999a,b,c); Lagoudas and Bo (1999) to account for transformation induced plasticity. This issue of modeling cyclic effects is related to the choice of hardening functions and will be revisited in Section 4.3.1.

### *4.1 Kinematic assumptions*

A large class of applications involving SMAs can easily be accommodated within the framework of small deformations. Thus, for the sake of simplicity, the constitutive theory is formulated for linearized strains, that is, the total strain tensor $\boldsymbol{\varepsilon}$ is given by:

$$\boldsymbol{\varepsilon} = \frac{1}{2}\left(\nabla \mathbf{u} + \nabla \mathbf{u}^T\right), \tag{7}$$

where $\mathbf{u}$ is the displacement. Examples of SMA models formulated in terms of finite strains can be found, for example, in Auricchio and Taylor (1997); Qidwai and Lagoudas (2000b);

Anand and Gurtin (2003). The theory is presented from a macroscopic point of view, therefore all quantities involved are macroscopic ones, considered over a suitable representative volume element. Further, it is assumed that the strain can be decomposed additively into elastic $\boldsymbol{\varepsilon}^{el}$, thermal $\boldsymbol{\varepsilon}^{th}$ and inelastic strain $\boldsymbol{\varepsilon}^{in}$ components:

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^{el} + \boldsymbol{\varepsilon}^{th} + \boldsymbol{\varepsilon}^{in}.$$

The inelastic strain $\boldsymbol{\varepsilon}^{in}$ is produced during the forward and reverse stress induced phase transformations ($\dot{\xi}_2 \neq 0$) and during the detwinning deformation ($\dot{\xi}_3 > 0$). Consequently, it can be further decomposed into:

$$\boldsymbol{\varepsilon}^{in} = \boldsymbol{\varepsilon}^{t} + \boldsymbol{\varepsilon}^{d}, \tag{8}$$

where $\boldsymbol{\varepsilon}^{t}$ is the stress induced transformation strain (produced during $A \rightarrow M^d$ transformation) and $\boldsymbol{\varepsilon}^{d}$ is the inelastic strain generated during detwinning ($M^t \rightarrow M^d$). Note that while the formation of martensitic twins in the $A \rightarrow M^t$ transformation does involve local strain fields, the macroscopic strain (averaged over a large representative volume element) is zero. Since this paper deals only with macroscopic description of the SMA, the transformation $A \leftrightarrow M^t$ is associated with zero (macroscopic) strain.

Finally, it is assumed that the transformation and detwinning strains obey the following two transformation/detwinning flow rules:

$$\dot{\boldsymbol{\varepsilon}}^{t} = \boldsymbol{\Lambda}^{t}\dot{\xi}_2, \tag{9}$$

$$\dot{\boldsymbol{\varepsilon}}^{d} = \boldsymbol{\Lambda}^{d}\dot{\xi}_3, \tag{10}$$

where $\boldsymbol{\Lambda}^{t}$ is the transformation tensor for the stress-induced martensitic transformation ($A \leftrightarrow M^d$) and $\boldsymbol{\Lambda}^{d}$ is the inelastic flow tensor for the detwinning of twinned martensite ($M^t \rightarrow M^d$). In general, the transformation tensors $\boldsymbol{\Lambda}^{t}$ and $\boldsymbol{\Lambda}^{d}$ are different (Lagoudas and Shu, 1999). Their specific form is discussed in Section 4.4 in conjunction with the definition of transformation surfaces.

The last assumptions (9) and (10) allow the formulation of the constitutive theory in terms of $\xi_i$, $i = 1, 2, 3$ as the only internal variables. It is convenient to use vector notation $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3)^T$ for the internal variables. The internal variables can be thought of as a time-like parameter because of the following relations:

$$\boldsymbol{\varepsilon}^{in} = \int_0^t \left(\boldsymbol{\Lambda}^{t}\dot{\xi}_2 + \boldsymbol{\Lambda}^{d}\dot{\xi}_3\right) d\tau = \int_0^{\xi_2} \boldsymbol{\Lambda}^{t} d\eta + \int_0^{\xi_3} \boldsymbol{\Lambda}^{d} d\eta, \tag{11}$$

that is, the inelastic strain $\boldsymbol{\varepsilon}^{in}$ can be considered as a path dependent functional of $\boldsymbol{\xi}$.

### *4.2 Free energy for Polycrystalline SMA*

The following form of the Gibbs energy, based on the works of Bo and Lagoudas (1999a); Lagoudas and Shu (1999) is assumed:

$$\begin{aligned} G\left(\boldsymbol{\sigma}, T, \boldsymbol{\xi}, \operatorname{sgn}\left(\dot{\xi}_1\right), \operatorname{sgn}\left(\dot{\xi}_2\right)\right) =& (c_1 + c_2) G^{M}(\boldsymbol{\sigma}, T, \boldsymbol{\xi}) + c_3 G^{A}(\boldsymbol{\sigma}, T, \boldsymbol{\xi}) \\ &+ G^{\text{mix}}\left(\boldsymbol{\sigma}, T, \boldsymbol{\xi}, \operatorname{sgn}\left(\dot{\xi}_1\right), \operatorname{sgn}\left(\dot{\xi}_2\right)\right), \end{aligned} \tag{12}$$

where $G^M$ is the thermoelastic free energy of both martensitic phases (twinned and detwinned), $G^A$ is the thermoelastic component of the free energy of austenite, and $G^{\text{mix}}$ is the free energy of mixing, which is responsible for the transformation behavior of the SMA. The mixing energy and thus the entire free energy depends also on the direction of the $A \leftrightarrow M^t$ and $A \leftrightarrow M^d$ transformations which is expressed through the sign function $\text{sgn}(\cdot)$:

$$\text{sgn}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases} \tag{13}$$

This kind of dependence allows the model to take into account the different material behavior during forward ($\dot{\xi}_i > 0$) and reverse ($\dot{\xi}_i < 0$) phase transformation, $i = 1, 2$.

In the pseudoelastic SMA literature this is often done implicitly, without including $\text{sgn}\left(\dot{\xi}_2\right)$ in the list of parameters of $G$ (c.f. e.g Lagoudas et al., 1996) in order to provide for different hardening during loading and unloading. An argument is then made that such a free energy is allowed to take values form two distinct, possibly discontinuous branches (one for loading, one for unloading) and the laws of thermodynamics should be verified for each branch alone. This argument is made mathematically rigorous in the current work by including $\text{sgn}\left(\dot{\xi}_2\right)$ explicitly in the parameter list (c.f. the discussion after equation (19)). This is also generalized to the $A \leftrightarrow M^t$ transformation by including dependence on $\text{sgn}\left(\dot{\xi}_1\right)$. [1]

Next, the two thermoelastic components are given by:

$$\begin{aligned} G^A(\boldsymbol{\sigma}, T, \boldsymbol{\xi}) = &- \frac{1}{2\rho}\boldsymbol{\sigma} : \boldsymbol{\mathcal{S}}^A : \boldsymbol{\sigma} - \frac{1}{\rho}\boldsymbol{\alpha}^A : \boldsymbol{\sigma}(T - T_0) - \frac{1}{\rho}\boldsymbol{\sigma} : \boldsymbol{\varepsilon}^{in} \\ &+ c^A\left[(T - T_0) - T\ln\left(\frac{T}{T_0}\right)\right] - s_0^A T + u_0^A \end{aligned} \tag{14}$$

$$\begin{aligned} G^M(\boldsymbol{\sigma}, T, \boldsymbol{\xi}) = &- \frac{1}{2\rho}\boldsymbol{\sigma} : \boldsymbol{\mathcal{S}}^M : \boldsymbol{\sigma} - \frac{1}{\rho}\boldsymbol{\alpha}^M : \boldsymbol{\sigma}(T - T_0) - \frac{1}{\rho}\boldsymbol{\sigma} : \boldsymbol{\varepsilon}^{in} \\ &+ c^M\left[(T - T_0) - T\ln\left(\frac{T}{T_0}\right)\right] - s_0^M T + u_0^M \end{aligned} \tag{15}$$

$\boldsymbol{\mathcal{S}}^i$, $\boldsymbol{\alpha}^i$, $c^i$, $s_0^i$ and $u_0^i$ are the compliance tensor, thermal expansion coefficient tensor, specific heat, specific entropy and the specific internal energy at the reference state of the individual phases with the superscript $i = A$ for austenitic and $i = M$ for martensite, respectively. It is assumed that the material properties of the two martensitic phases are the same. Note that this assumption, and correspondingly, the selection of the same energy for $M^t$ and $M^d$ is guided by the fact that from a metallurgical point of view, the two phases are indistinguishable (Leclercq and Lexcellent, 1996). As however was discussed in the beginning of this section, it is the macroscopic mechanical behavior of twinned and detwinned martensite that is different, which is reflected in the kinematic considerations of Section 4.1.

[1] Regardless of whether the dependence is implicit or explicit, continuum thermodynamics of irreversible processes dictates that all fields, including $G$, should be independent of the rates of the internal variables. However, this classical result (c.f. e.g. Coleman and Gurtin, 1967) is applicable only to smooth functions $G$, that is functions that are continuous and all their partial derivatives are also continuous. In the current case $G$ is not smooth since $\text{sgn}(\cdot)$ is a discontinuous, non-differentiable function.

Note that in light of equation (11), the two free energies $G^A$ and $G^M$ and consequently, $G$ are path dependent functionals of $\boldsymbol{\xi}$. Note that this approach of making $G$ a path dependent functional is chosen for convenience (as is done by many other authors c.f. e.g. Bo and Lagoudas, 1999a; Lagoudas and Entchev, 2004), and does not lead to any mathematical inconsistencies in the thermodynamic treatment of the constitutive model.

Further, upon substituting equations (14) and (15) into (12) and by using equations (3)-(5) and the constraint (1), the following expression is obtained for the free energy:

$$\begin{aligned} G = &- \frac{1}{2\rho}\boldsymbol{\sigma} : \boldsymbol{\mathcal{S}}(c_1+c_2) : \boldsymbol{\sigma} - \frac{1}{\rho}\boldsymbol{\sigma} : \left[\boldsymbol{\alpha}(c_1+c_2)(T-T_0) + \boldsymbol{\varepsilon}^{in}\right] \\ &+ c(c_1+c_2)\left[(T-T_0) - T\ln\left(\frac{T}{T_0}\right)\right] - s_0(c_1+c_2)T + u_0(c_1+c_2) + G^{\text{mix}}, \end{aligned} \tag{16}$$

where $\boldsymbol{\mathcal{S}}(c_1+c_2)$, $\boldsymbol{\alpha}(c_1+c_2)$, $c(c_1+c_2)$, $s_0(c_1+c_2)$ and $u_0(c_1+c_2)$ are the effective compliance tensor, thermal expansion coefficient tensor, specific heat, specific entropy and the specific internal energy at the reference state, respectively. These effective material properties are calculated in terms of the total martensitic volume fraction $c_1+c_2$ using the rule of mixtures:

$$\boldsymbol{\mathcal{S}}(\boldsymbol{\xi}) = \boldsymbol{\mathcal{S}}(c_1+c_2) = \boldsymbol{\mathcal{S}}^A + (c_1+c_2)(\boldsymbol{\mathcal{S}}^M - \boldsymbol{\mathcal{S}}^A) = \boldsymbol{\mathcal{S}}^A + (c_1+c_2)\Delta\boldsymbol{\mathcal{S}}, \tag{17a}$$

$$\boldsymbol{\alpha}(\boldsymbol{\xi}) = \boldsymbol{\alpha}(c_1+c_2) = \boldsymbol{\alpha}^A + (c_1+c_2)(\boldsymbol{\alpha}^M - \boldsymbol{\alpha}^A) = \boldsymbol{\alpha}^A + (c_1+c_2)\Delta\boldsymbol{\alpha}, \tag{17b}$$

$$c(\boldsymbol{\xi}) = c(c_1+c_2) = c^A + (c_1+c_2)(c^M - c^A) = c^A + (c_1+c_2)\Delta c, \tag{17c}$$

$$s_0(\boldsymbol{\xi}) = s_0(c_1+c_2) = s_0^A + (c_1+c_2)(s_0^M - s_0^A) = s_0^A + (c_1+c_2)\Delta s_0, \tag{17d}$$

$$u_0(\boldsymbol{\xi}) = u_0(c_1+c_2) = u_0^A + (c_1+c_2)(u_0^M - u_0^A) = u_0^A + (c_1+c_2)\Delta u_0. \tag{17e}$$

A detailed discussion of the functional form (16) for the free energy and the resulting rule of mixtures (17), based on micromechanical averaging over a representative volume element of the polycrystalline SMA can be found in (Bo and Lagoudas, 1999a). Note that, in view of relations (3)-(5), the effective parameters can be viewed either as functions of the total volume fraction of martensite $(c_1+c_2)$ or as functions of the internal variables $\boldsymbol{\xi}$. The later notation is more convenient when performing the algebraic manipulations of this section, while the former gives a better physical understanding of the quantities involved.

The mixing term $G^{\text{mix}}$ in the free energy (c.f. equations (12), (16)) is defined as follows:

$$\begin{aligned} &G^{\text{mix}}\left(\boldsymbol{\xi}, \operatorname{sgn}\left(\dot{\xi}_1\right), \operatorname{sgn}\left(\dot{\xi}_2\right)\right) = \\ &\quad \frac{1}{\rho}\int_0^t \left(f_1(\boldsymbol{\xi}; \operatorname{sgn}\left(\dot{\xi}_1\right))\dot{\xi}_1(\tau) + f_2(\boldsymbol{\xi}; \operatorname{sgn}\left(\dot{\xi}_2\right))\dot{\xi}_2(\tau) + f_3(\boldsymbol{\xi})\dot{\xi}_3(\tau)\right) d\tau, \end{aligned} \tag{18}$$

where the yet to be defined functions $f_i$, $i = 1, 2, 3$ are responsible for the hardening during the $A \leftrightarrow M^t$, $A \leftrightarrow M^d$ and $M^t \rightarrow M^d$ transformations, respectively (see next section). Since many SMAs exhibit different hardening behavior during loading and unloading (Lagoudas et al., 1996) it is necessary to allow $f_1$ and $f_2$ to take different values depending on whether one has forward or reverse transformation, hence the dependence on $\operatorname{sgn}\left(\dot{\xi}_1\right)$, $\operatorname{sgn}\left(\dot{\xi}_2\right)$. It should be noted that the in the case of pseudoelasticity only ($\dot{\xi}_1 = \dot{\xi}_3 = 0$) the above mixing energy leads to free energy which is equivalent to the one used by (Lagoudas et al., 1996).

In order to apply the second law of thermodynamics to the constitutive theory (next section)

it is necessary to derive the rate of change of the free energy (12). It is given by

$$\dot{G} = \dot{\boldsymbol{\sigma}} \frac{\partial G}{\partial \boldsymbol{\sigma}} + \dot{T} \frac{\partial G}{\partial T} + \dot{\boldsymbol{\xi}} \cdot \frac{\partial G}{\partial \boldsymbol{\xi}} \tag{19}$$

The derivation of this relation is not straightforward. The usual way to derive such identities is to consider $G$ as a function of time and apply the chain rule to the definition of $G$, in our case, equations (12) and (18). However in the current case $G$ is not a smooth function of all its internal variables. Indeed, consider a point in the state space, where the $A \leftrightarrow M^t$ or $A \leftrightarrow M^d$ transformation changes sign. At such a point of transformation reversal, $G$ is not differentiable with respect to the two rates $\dot{\xi}_1$ and $\dot{\xi}_2$, (c.f. equation (13)). Moreover, the derivatives $\partial_{\xi_1} G$ and $\partial_{\xi_2} G$ are discontinuous when $\dot{\xi}_1$ or $\dot{\xi}_2$ change sign, respectively (see equation (A-3) in Appendix A). As a result, the chain rule cannot be applied directly to equation (12). To obtain (19), first consider a point in state space where both $\dot{\xi}_1$ and $\dot{\xi}_2$ do not change sign. In the neighborhood of this point $G$ does not depend on $\dot{\boldsymbol{\xi}}$ and it is clearly smooth. Then, (19) is obtained using the chain rule. At points of transformation reversal, where $G$ is not smooth, one has to differentiate directly $G$ with respect to time, the details for which are given in Appendix A. This result demonstrates that the explicit inclusion of $\operatorname{sgn}\left(\dot{\xi}_1\right)$ and $\operatorname{sgn}\left(\dot{\xi}_2\right)$ in the list of parameters of $G$ does not introduce derivatives with respect to $\dot{\xi}_i$ in the expression for $\dot{G}$.

### *4.3 Thermodynamics and constitutive relations*

Every thermomechanical process must satisfy the second law of thermodynamics, which, written in terms of the Gibbs free energy reads (c.f. e.g. Malvern, 1969):

$$\rho \dot{G} + \dot{\boldsymbol{\sigma}} : \boldsymbol{\varepsilon} + \rho s \dot{T} + \frac{\mathbf{q} \cdot \nabla T}{T} \leq 0. \tag{20}$$

Now, by substituting equation (19) into (20), the inequality becomes:

$$\left( \boldsymbol{\varepsilon} + \rho \frac{\partial G}{\partial \boldsymbol{\sigma}} \right) : \dot{\boldsymbol{\sigma}} + \rho \left( s + \frac{\partial G}{\partial T} \right) \dot{T} + \rho \frac{\partial G}{\partial \boldsymbol{\xi}} \cdot \dot{\boldsymbol{\xi}} + \frac{\mathbf{q} \cdot \nabla T}{T} \leq 0$$

Following a standard argument (c.f. e.g. Truesdell and Noll, 1965) the following two constitutive relations are established for the strain and entropy:

$$\boldsymbol{\varepsilon} = -\rho \frac{\partial G}{\partial \boldsymbol{\sigma}}, \tag{21}$$

$$s = -\frac{\partial G}{\partial T}, \tag{22}$$

With the help of equations (16) and (18), the above relations are explicitly written as:

$$\boldsymbol{\sigma} = \boldsymbol{\mathcal{S}}(\boldsymbol{\xi})^{-1} : \left( \boldsymbol{\varepsilon} - \boldsymbol{\alpha}(\boldsymbol{\xi})(T - T_0) - \boldsymbol{\varepsilon}^{in} \right), \tag{23}$$

$$s = \frac{1}{\rho} \boldsymbol{\alpha}(\boldsymbol{\xi}) T + c(\boldsymbol{\xi}) \ln(T/T_0) + s_0(\boldsymbol{\xi}). \tag{24}$$

Further, the entropy inequality (20) also implies:

$$\rho\frac{\partial G}{\partial \boldsymbol{\xi}}\cdot\dot{\boldsymbol{\xi}}+\frac{\mathbf{q}\cdot\nabla T}{T}\leq 0. \tag{25}$$

In this work it will be further assumed (Qidwai and Lagoudas, 2000b) that the SMA material is strongly dissipative, that is, the two terms in the last inequality (25) hold separately [2]:

$$\rho\frac{\partial G}{\partial \boldsymbol{\xi}}\cdot\dot{\boldsymbol{\xi}}\leq 0, \tag{26}$$

$$\frac{\mathbf{q}\cdot\nabla T}{T}\leq 0. \tag{27}$$

Let the thermodynamic forces, conjugate to $\boldsymbol{\xi}$ be denoted by $\boldsymbol{\pi}=(\pi_1,\pi_2,\pi_3)^T$. With the help of (16), (17) and (A-3) they are given by:

$$\pi_1=-\rho\frac{\partial G}{\partial \xi_1}=\tilde{\pi}(\boldsymbol{\sigma},T)-f_1\left(\boldsymbol{\xi},\operatorname{sgn}\left(\dot{\xi}_1\right)\right), \qquad \text{whenever } \dot{\xi}_1\neq 0, \tag{28}$$

$$\pi_2=-\rho\frac{\partial G}{\partial \xi_2}=\boldsymbol{\sigma}:\boldsymbol{\Lambda}^t+\tilde{\pi}(\boldsymbol{\sigma},T)-f_2\left(\boldsymbol{\xi},\operatorname{sgn}\left(\dot{\xi}_2\right)\right), \qquad \text{whenever } \dot{\xi}_2\neq 0, \tag{29}$$

$$\pi_3=-\rho\frac{\partial G}{\partial \xi_3}=\boldsymbol{\sigma}:\boldsymbol{\Lambda}^d-f_3(\boldsymbol{\xi}), \qquad \text{whenever } \dot{\xi}_3>0. \tag{30}$$

where $\tilde{\pi}$ is:

$$\begin{aligned}\tilde{\pi}(\boldsymbol{\sigma},T)=&\frac{1}{2}\boldsymbol{\sigma}:\Delta\boldsymbol{\mathcal{S}}:\boldsymbol{\sigma}+\Delta\boldsymbol{\alpha}:\boldsymbol{\sigma}(T-T_0)\\&-\rho\Delta c\left[(T-T_0)-T\ln\left(\frac{T}{T_0}\right)\right]+\rho\Delta s_0 T-\rho\Delta u_0.\end{aligned} \tag{31}$$

#### *4.3.1 Transformation hardening functions*

The hardening function $f_1$ for the $A\leftrightarrow M^t$ transformation is assumed to depend on $c_1$, and may be different for the forward and reverse transformation:

$$f_1=\begin{cases}\Delta_1^+ f_1^+(c_1) \text{ for } \dot{\xi}_1>0\\ \Delta_1^- f_1^-(c_1) \text{ for } \dot{\xi}_1<0\end{cases}. \tag{32}$$

Here $f_1^{\pm}(c_1)$ are two arbitrary monotonously increasing functions in the interval $[0,1]$ for the forward and reverse transformations $A\rightarrow M^t$ and $M^t\rightarrow A$ respectively which can be determined from experimental measurements. The two material constants $\Delta_1^{\pm}$ serve as a scaling factors for $f_1^{\pm}(c_1)$ respectively, so that

$$f_1^{\pm}(0)=0, \quad f_1^{\pm}(1)=1. \tag{33}$$

The hardening function $f_2^{\pm}$, $f_3$ for the stress induced martensitic transformation $A\leftrightarrow M^d$ and the reorientation of twinned martensite $M^t\rightarrow M^d$ respectively are assumed to depend

[2] Note that the last assumption is equivalent to assuming that $\mathbf{q}$ is independent of $\dot{\boldsymbol{\xi}}$.

on the volume fraction of twinned martensite $c_2$:

$$f_2 = \begin{cases} \Delta_2^+ f_2^+(c_2) \text{ for } \dot{\xi}_2 > 0 \\ \Delta_2^- f_2^-(c_2) \text{ for } \dot{\xi}_2 < 0 \end{cases}, \quad f_3 = \Delta_3 f_3(c_2) \quad \text{for } \dot{\xi}_3 > 0. \tag{34}$$

Similarly to equation (32), the material constants $\Delta_2^{\pm}$ and $\Delta_3$ are scaling factors for the monotonous functions $f_2^{\pm}$ and $f_3$, respectively, and:

$$f_2^{\pm}(0) = 0, \quad f_2^{\pm}(1) = 1, \tag{35}$$
$$f_3(0) = 0, \quad f_3(1) = 1. \tag{36}$$

Several things should be noted about this selection of hardening functions. The choice of $c_2$ as the independent variable for $f_2^{\pm}$ and $f_3$ has generally been accepted in the literature. The choice of $c_1$ as the unknown variable for $f_1^{\pm}$, while often used (Juhasz et al., 2002; Leclercq and Lexcellent, 1996; Brinson, 1993) is not the only possible option. The total amount of austenite $c_3$ may be an equally suitable choice for certain classes of SMA materials.

The specific form of the functions $f_i$ (e.g. polynomials, trigonometric functions, exponents, etc.) is material dependent and should be treated as part of the material specifications. Through the rest of this paper, for the sake of simplicity, it is assumed that the hardening functions are linear:

$$f_1^{\pm}(c_1) = c_1, \quad f_2^{\pm}(c_2) = c_2, \quad f_3(c_2) = c_2. \tag{37}$$

This selection [3] is typical for the pseudoelastic and detwinning response of polycrystalline NiTi SMAs (Lagoudas et al., 1996). In principle however, the model allows for arbitrary monotonous functions that can be curve-fitted from experiments (Section 4.5).

Finally, the hardening functions depend indirectly on $\boldsymbol{\xi}$ through the volume fractions $c_i$ (equations (3)-(5)). The volume fractions $c_i$ have fixed bounds (c.f. equation (2)). Hence, a hardening function which depends explicitly on $c_i$ will have the property that the transformation strips (see next section) will not change with cyclic thermomechanical loading. It should be kept in mind that the position of the transformation strips in the phase diagram do evolve with cyclic repetition of thermomechanical loading paths. Such effects can be accounted for by specifying an explicit dependency of $f_i^{\pm}$ on $\xi_1$, $\xi_2$ and $\xi_3$. For example, if $f_1^+ = \xi_1 - (1+\lambda)\xi_3$ is selected, with $\lambda > 0$ is a small positive parameter, every full SME cycle will increase the $M_s$ and $M_f$ temperatures by fixed amount. This type of modeling the evolution of SMA material response however was outside the scope of this work.

[3] This choice of $f_2$ is consistent with the hardening function of Boyd and Lagoudas (1994); Qidwai and Lagoudas (2000a), who used the derivative of a quadratic polynomial, which is a linear function.

*4.4 Transformation surfaces and flow rules*

It is assumed that a closed elastic domain is associated with each possible transformation, bounded by a transformation surface. The five surfaces are:

$$\Phi_1^+(\boldsymbol{\sigma}, T, \boldsymbol{\xi}) = 0, \quad \text{whenever the } A \to M^t \text{ tranformation takes place,} \tag{38a}$$
$$\Phi_1^-(\boldsymbol{\sigma}, T, \boldsymbol{\xi}) = 0, \quad \text{whenever the } M^t \to A \text{ tranformation takes place,} \tag{38b}$$
$$\Phi_2^+(\boldsymbol{\sigma}, T, \boldsymbol{\xi}) = 0, \quad \text{whenever the } A \to M^d \text{ tranformation takes place,} \tag{38c}$$
$$\Phi_2^-(\boldsymbol{\sigma}, T, \boldsymbol{\xi}) = 0, \quad \text{whenever the } M^d \to A \text{ tranformation takes place,} \tag{38d}$$
$$\Phi_3(\boldsymbol{\sigma}, T, \boldsymbol{\xi}) = 0, \quad \text{whenever the } M^t \to M^d \text{ deformation takes place,} \tag{38e}$$

and the elastic domains in stress-temperature space, for given $\boldsymbol{\xi}$, with respect to $\dot{\xi}_i$ are defined implicitly by the inequalities:

$$\left\{(\boldsymbol{\sigma}, T) | \Phi_i^+(\boldsymbol{\sigma}, T, \boldsymbol{\xi}) \leq 0, \quad \dot{\xi}_i > 0\right\}, \quad \text{for } i = 1, 2, \tag{39}$$
$$\left\{(\boldsymbol{\sigma}, T) | \Phi_i^-(\boldsymbol{\sigma}, T, \boldsymbol{\xi}) \leq 0, \quad \dot{\xi}_i < 0\right\}, \quad \text{for } i = 1, 2, \tag{40}$$
$$\left\{(\boldsymbol{\sigma}, T) | \Phi_3(\boldsymbol{\sigma}, T, \boldsymbol{\xi}) \leq 0, \quad \dot{\xi}_3 > 0\right\}. \tag{41}$$

The first two inequalities describe the elastic domains of the two *forward* transformations, the second two inequalities the elastic domains of the two *reverse* transformations. The last inequality describes the elastic domain for the $M^t \to M^d$ transformation. In contrast to conventional plasticity, the phase transformation terminates, whenever the constraints (2) are violated. Therefore, the elastic domain associated with given phase transformation is assumed to be the entire space, when the transformation is complete or there is no more material to transform.

Following Lagoudas and Shu (1999); Qidwai and Lagoudas (2000b), the following form of the transformation surfaces is suggested:

$$\Phi_1^+(\boldsymbol{\sigma}, T, \boldsymbol{\xi}) = \pi_1 - Y_1^+, \qquad \Phi_1^-(\boldsymbol{\sigma}, T, \boldsymbol{\xi}) = -\pi_1 - Y_1^-, \tag{42}$$
$$\Phi_2^+(\boldsymbol{\sigma}, T, \boldsymbol{\xi}) = \pi_2 - Y_2^+, \qquad \Phi_2^-(\boldsymbol{\sigma}, T, \boldsymbol{\xi}) = -\pi_2 - Y_2^-, \tag{43}$$

$$\Phi_3(\boldsymbol{\sigma}, T, \boldsymbol{\xi}) = \pi_3 - Y_3, \tag{44}$$

where and $Y_1^\pm$, $Y_2^\pm$, $Y_3$ are measures of internal dissipation of the respective transformations. In this work it is assumed that $Y_i^\pm$, $i = 1, 2, 3$ are constants, independent of $\boldsymbol{\sigma}$, $T$ and $\boldsymbol{\xi}$. This, due to the inequalities (39)-(41), implies that the appropriate conjugate forces $\pi_i$ remain constant during the transformation. It also implies that the entropy production due to a phase transformation is proportional to $\dot{\xi}_i$, with $Y_i^\pm$ being the proportionality constant (c.f. equation (26)).

The functions $f_i$ defined by (32), (34) appear in the definition of the transformation function (42)-(44) through the constitutive relations (28)-(30). They are the only terms in the transformation functions dependent on the internal variables $\boldsymbol{\xi}$, hence they are responsible for the transformation hardening.

In order to complete the model, the transformation tensors in the flow rules (9) and (10) should be specified. Let $\mathrm{dev}(\boldsymbol{\sigma})$ be the deviatoric stress and $\|\cdot\|$ the usual tensor norm,

respectively:

$$\mathrm{dev}(\boldsymbol{\sigma}) = \boldsymbol{\sigma} - \frac{1}{3}\mathrm{tr}(\boldsymbol{\sigma})\mathbf{I}, \qquad\qquad \|\mathbf{v}\| = \sqrt{\mathbf{v}\cdot\mathbf{v}}.$$

The detwinning flow tensor is taken to be of the form

$$\boldsymbol{\Lambda}^d = \sqrt{\frac{3}{2}} H^d \frac{\mathrm{dev}(\boldsymbol{\sigma})}{\|\mathrm{dev}(\boldsymbol{\sigma})\|}, \tag{45}$$

where $H^d$ is the maximal uniaxial inelastic strain, assumed to be a material constant.

For the sake of simplicity, the flow rule used for the $A \leftrightarrow M^d$ transformation is, following Lagoudas et al. (1996); Qidwai and Lagoudas (2000a) taken to be

$$\boldsymbol{\Lambda}^t = \begin{cases} \sqrt{\frac{3}{2}} H^t \frac{\mathrm{dev}(\boldsymbol{\sigma})}{\|\mathrm{dev}(\boldsymbol{\sigma})\|} & \text{for } \dot{\xi}_2 > 0 \\ \sqrt{\frac{3}{2}} H^t \frac{\mathrm{dev}(\boldsymbol{\varepsilon}^{in})}{\|\mathrm{dev}(\boldsymbol{\varepsilon}^{in})\|} & \text{for } \dot{\xi}_2 < 0 \end{cases}, \tag{46}$$

where $H^t$ is a material constant having the meaning of maximal uniaxial transformation induced strain. It should be noted that the forward flow rule is a simple $J_2$ based one, which has been used in many of the early works on modeling pseudoelasticity (c.f. e.g. Lagoudas et al., 1996). A number of alternative transformation surfaces have been proposed in the literature (c.f. e.g. Auricchio et al., 1997; Gillet et al., 1998; Huang, 1999; Qidwai and Lagoudas, 2000b; Lexcellent et al., 2002), which account for the observed tension-compression asymmetry of SMA materials as well as the development of a small volumetric strain during the $A \rightarrow M^d$ phase transformations. Due to the large number of different SMA alloys the selection of an appropriate transformation surface can be a difficult task and is specific for each alloy. The simple choice of transformation surface also helps in the next section and in Appendix B, where the necessary relations are found for the material parameters so that the model is consistent with the selected phase diagram. Since the main goal of the current research is the formulation of a consistent model capable of accounting for phase transformation and detwinning of self accommodated martensite over a wide range of stresses and temperatures, the choice of more accurate transformation surfaces/flow rules was not addressed in detail.

The reverse transformation tensor of the last equation (46) also deserves some attention. The reason why two different transformation flow tensors are used for loading and unloading is the need to account for reorientation in multiaxial loading path. In general, if the direction of the stress state is changed, some martensitic variants will reorient in the new direction, thus changing the direction of the inelastic strain. A constitutive model with a single volume fraction for all detwinned variants of martensite cannot account for this process. If the same transformation tensor is used for forward and reverse transformations it may happen that residual inelastic strain is present after unloading to austenite (e.g. $c_3 = 1$ and the stress becomes zero) from a non-proportional loading path. The unloading criterion used above ensures that when $c_3 = 1$, the inelastic strain becomes zero. It reduces to the same transformation tensor used by (Qidwai and Lagoudas, 2000a) when $\dot{\xi}_3 = 0$.

*4.5 Determination of material parameters*

A successful implementation of a material model depends on the ability to express the material parameters from physically observable quantities. The material parameters entering the present model can be divided into two groups. The first group is parameters with direct physical interpretation:

$$\boldsymbol{\mathcal{S}}^i, \boldsymbol{\alpha}^i, c^i, s_0^i, u_0^i, H^t, H^d, \tag{47}$$

and parameters related to the structure of the phase diagram and the transformation hardening of the material:

$$f_1^{\pm}, f_2^{\pm}, f_3, Y_1^{\pm}, Y_2^{\pm}, Y_3, \Delta_1^{\pm}, \Delta_2^{\pm}, \Delta_3. \tag{48}$$

In both groups, the index $i$ takes the values $A, M$ for austenite and martensite, respectively.

The first group of parameters can be measured directly. A polycrystalline SMA, unlike the single crystal SMAs, is an isotropic material. Therefore the compliances $\boldsymbol{\mathcal{S}}^A$, $\boldsymbol{\mathcal{S}}^M$ are determined if the Young's modulus $E^A$, $E^M$ and Poisson's ratio $\nu^A$, $\nu^M$ of the two phases are available. These can be determined from standard mechanical tests. The thermal expansion coefficient $\boldsymbol{\alpha}^A$, $\boldsymbol{\alpha}^M$ for an isotropic material are scalars and are determined from isobaric tests. The specific heats $c^A$, $c^M$, the change in specific entropy $\rho\Delta s_0$ between the two phases and the change of specific internal energy $\Delta u_0$ can be determined from calorimetric measurements (Bo and Lagoudas, 1999a,b). The maximum uniaxial transformation strain $H$ can be obtained from either an isothermal test or from an isobaric test (Bo and Lagoudas, 1999a).

The remaining parameters (48) are related to the position of the transformation strips in the uniaxial phase diagram (Figure 4). The key to determining them is to measure the critical temperatures $M_s$, $M_f$, $A_s^t$, $A_f^t$, $A_s^d$, $A_f^d$ as well as the critical stresses $\sigma_s$ and $\sigma_f$. Then $f_1^{\pm}$, $f_2^{\pm}$, $f_3$, $Y_1^{\pm}$, $Y_2^{\pm}$, $Y_3$, $\Delta_1^{\pm}$, $\Delta_2^{\pm}$ and $\Delta_3$ are obtained by simulating several simple thermomechanical paths on the phase diagram for a uniaxial stress state. The calculations are straightforward, and for the sake of brevity are given in Appendix B. The critical temperatures can be measured in the following way: The transformation temperatures $M_s$, $M_f$, $A_s^t$, $A_f^t$ are easily determined from a DSC test such as the one shown in Figure 2. Knowing the critical temperatures at zero stress is sufficient to determine the $A \to M^t$ and $M^t \to A$ transformation strips. The $A_s^d$, $A_f^d$ temperatures on the other hand can be found by first loading a specimen in detwinning conditions until the specimen has entirely detwinned. It is then mechanically unloaded in a way which preserves the material state and then a DSC test is performed as described in Section 2. To do the DSC test it is necessary to perform the mechanical loading and the subsequent preparation of a DSC sample from the loaded specimen at temperatures below $A_s^d$, which may not always be possible. A more direct approach relies on several isothermal test above $A_f^t$, which will allow to construct both the $A \to M^d$ and $M^d \to A$ strips and therefore, also the $A_s^d$, $A_f^d$ temperatures. Isothermal tests at temperatures below $M_f$ can be used to determine $\sigma_s$ and $\sigma_f$ and hence the $M^t \to M^d$ strip.

The presentation of the current model is concluded in the next two sections by demonstrating that the it reproduces the phase diagram of Figure 4 and by discussing certain restrictions on the relative position of the phase transformation strips.

*4.5.1 The uniaxial transformation strips and the phase diagram*

The one dimensional reduction of the model (Appendix B) resulted in the inequalities (B-3)-(B-7) for the elastic domains of the respective transformations. It is clear from equation (B-9)

that the transformation strip in stress-temperature space for the $M^t \rightarrow M^d$ deformation is the horizontal strip

$$\sigma_s \leq \sigma \leq \sigma_f,$$

which is consistent with the assumptions of Section 3.

Next, assume for a moment that the elastic moduli of the two phases, the thermal moduli and the specific heats of the two phases are equal:

$$\boldsymbol{\mathcal{S}}^A = \boldsymbol{\mathcal{S}}^M, \quad \boldsymbol{\alpha}^A = \boldsymbol{\alpha}^M, \quad c^a = c^M.$$

In this case, equation (B-8) reduces to

$$\hat{\pi} = \rho \Delta s_0 T - \rho \Delta u_0.$$

Then, equation (B-10) implies that the transformation strip for the $A \rightarrow M^t$ is defined by

$$M_f \leq T \leq M_s$$

and from equation (B-13), the transformation strip for the $M^t \rightarrow A$ is the vertical region

$$A_s^t \leq T \leq A_f^t.$$

It can also be seen form equations (B-19) and (B-16) that, for any given $c_2$, the transformation line for both $A \rightarrow M^d$ and $M^d \rightarrow A$ transformation is linear and has slope

$$k = -\frac{\rho \Delta s_0}{H}.$$

Therefore, the $A \rightarrow M^d$ and $M^d \rightarrow A$ strips have the shape shown in Figure 4, and the slope $k$ is given by the above formula [4].

When the moduli for the two phases are different, the transformation lines for the $A \leftrightarrow M^t$ and $A \leftrightarrow M^d$ depart from the above linear relationships. However, the terms $\Delta S \sigma^2$, $\Delta \alpha \sigma$ and $\rho \Delta c \Big[ (T - T_0) - T \ln \left( \frac{T}{T_0} \right) \Big]$ which will now appear in (B-8) are all an order of magnitude smaller than the leading term $\rho \Delta s_0 T$. The departure from a linear shape is therefore visible for high stress (several hundred $MPa$) for $A \leftrightarrow M^t$ transformation and for both higher stresses and away from the equilibrium temperature $T_0$ for the $A \leftrightarrow M^d$ transformation. It is easy to show, that in the general case of different elastic and thermal moduli, the meaning of the slope $k$ becomes now the tangent to the transformation surface at zero stress.

### *4.5.2 Relative position of the transformation surfaces*

As mentioned in Section 3 the current model does not assume any triple point as often done in the literature Brinson (1993). As a result, the $A \rightarrow M^d$ strip can be translated according tho experimental measurements. However, certain restrictions, which result from the assumed functional dependence of $f_1$, $f_2$ and $f_3$ are still valid. To the best of the authors knowledge, two of these exist and will be mentioned briefly here. Both of them occur for

[4] This last formula is frequently used (for example, by Qidwai and Lagoudas, 2000a) as an alternative method to determine the difference in specific entropies $\Delta s_0$.

certain specific material parameters and it may happen that such classes of SMA materials do not exist in practice.

First, the transformation strips $M^t \rightarrow A$, $M^d \rightarrow A$ and $M^t \rightarrow M^d$ must have a zero intersection. It is easy to show, that if they do, the three inequalities (B-4), (B-6) and (B-7) cannot be satisfied simultaneously. In other words, a simultaneous transformation $M^t \rightarrow A$, $M^d \rightarrow A$ and $M^t \rightarrow M^d$ is not possible. In light of the experimental results of Section 2, it seems unlikely that such a situation can occur. Also, it is physically difficult to explain why some twinned martensite will transform to austenite through an intermediate detwinned phase, while the rest of the twinned martensite will transform directly to austenite. This limitation can be removed by assuming a different functional dependence of $f_1$, for example on $c_3$. Note that if such a transformation is allowed, it is necessary to use all the tree internal variables $\boldsymbol{\xi}$, as just knowing the rate of change of $c_i$ is not sufficient to determine exactly how the species transformed (c.f. equations (3)-(5)).

The second limitation of the theory is associated with a bifurcation in some loading paths (in $\boldsymbol{\sigma} - \boldsymbol{\varepsilon} - T$ space) and certain material parameters. Suppose that a material is cooled from pure austenite at constant stress, which is higher by a finite amount than the critical stress for the $M^t \rightarrow M^d$ deformation. As there is no available $M^t$, no $M^d$ can be produced so that the inequality (B-7) is turned to equality. Suppose further, that as the cooling proceeds, the $A \rightarrow M^t$ surface is reached and there is still some available $A$. That is, $\Phi_1^+ = 0$, $\Phi_2^+ = 0$ and $\Phi_3 > 0$. As the $\Phi_1$ surface is first activated, some (small) amount of $M^t$ is produced so that (B-3) becomes equality. This $M^d$ must all be detwinned via $M^t \rightarrow M^d$ ($\dot{\xi}_3 > 0$) in order to relax the violation of (B-7). As this happens, $c_1$ again becomes 0, thus (B-3) is again violated, hence more $M^t$ must be produced, and so forth, until all the austenite is exhausted via this transformation $A \rightarrow M^t \rightarrow M^d$. Thus an infinitesimal drop in the temperature, which activates $\Phi_1^+$ (c.f. equations (42) and (B-3)) and produces the first $M^t$, will result in a finite amount of $A \rightarrow M^d$. This implies a finite production of transformation strain and thus will results in strain discontinuity, which to the best of the our knowledge has not been observed in practice. Clearly, this situation is possible if the finish line for the $A \rightarrow M^d$ transformation pass above the point $(M_s, \sigma_f)$ in stress temperature space. Then, as long as $f_1$ is a function of $c_1$ and $f_3$ a function of $c_2$, and regardless of the functional form of $f_i$ (c.f. equation (37)), a simple isobaric path at stress equal to $\boldsymbol{\sigma}_f$ will result in the above situation.

Hence, to prevent such behavior, it is necessary (but not sufficient) to require that the finish $A \rightarrow M^d$ line passes at or below the point $(M_s, \sigma_f)$. It is easy to see that for the selected linear form of the hardening function (37) the above loading path discontinuity is not possible if the finish $A \rightarrow M^d$ line passes at or below the point $(M_s, \sigma_f)$ and $\Delta_3 > \Delta_2^+$. For nonlinear functions $f_i$ it is more difficult to derive sufficient conditions for which there is no discontinuity. Whether such materials, for which the finish $A \rightarrow M^d$ line passes above the point $(M_s, \sigma_f)$, exist is an open question. Related to this is if loading path discontinuity (in $\boldsymbol{\sigma} - \boldsymbol{\varepsilon} - T$ space) should be prohibited. Note that, for such class of materials a different functional dependence of $f_1$ may provide a solution to the discontinuity problem.

## 5 Numerical Implementation of the model

When solving numerically boundary value problems (for example by finite elements) one is presented with the following problem: at a given material point, the history as well as the

current material state (strain, temperature, internal variables) is known and new values are given for the strain and temperature. One has to find numerically the corresponding stress and internal variables at this particular material point. To do this, an extension of the closest point projection algorithm of Qidwai and Lagoudas (2000a) is used. Qidwai and Lagoudas (2000a) implemented the class of SMA models of Boyd and Lagoudas (1996a) which have only one active transformation surface (either $A \to M^d$ or $M^d \to A$). The numerical scheme described in this section is adapted to the multiple transformation surfaces present in the current model. The scheme belongs to the general family of return mapping algorithms of (Ortiz and Popov, 1985; Simo and Hughes, 1987, 1998) which couple in a natural with a displacement based finite element methods. In this section, the major steps of the numerical implementation of the SMA constitutive model are presented, while the details can be found in Popov (2005).

Consider a single material point. First, rewrite equation (23) as:

$$\boldsymbol{\varepsilon} = \boldsymbol{\mathcal{S}}(\boldsymbol{\xi}) : \boldsymbol{\sigma} + \boldsymbol{\alpha}(\boldsymbol{\xi})(T - T_0) + \boldsymbol{\varepsilon}^{in}. \tag{49}$$

The evolution equations (9) and (10) and the decomposition (8) imply that the total inelastic strain $\boldsymbol{\varepsilon}^{in}$ can be written as

$$\dot{\boldsymbol{\varepsilon}}^{in} = \boldsymbol{\Lambda}^t(\boldsymbol{\sigma})\dot{\xi}_2 + \boldsymbol{\Lambda}^d(\boldsymbol{\sigma})\dot{\xi}_3, \tag{50}$$

where $\boldsymbol{\Lambda}^t(\boldsymbol{\sigma})$ and $\boldsymbol{\Lambda}^d(\boldsymbol{\sigma})$ are defined by equations (46) and (45) respectively. It is also convenient to write the consistency conditions imposed by the transformation surfaces (39)–(44) for the evolution of the internal variables $\boldsymbol{\xi}$ in the following compact form:

$$\dot{\xi}_1 \geq 0, \quad \Phi_1^+ \leq 0, \quad \Phi_1^+\dot{\xi} = 0, \tag{51a}$$
$$\dot{\xi}_1 \leq 0, \quad \Phi_1^- \leq 0, \quad \Phi_1^-\dot{\xi} = 0, \tag{51b}$$
$$\dot{\xi}_2 \geq 0, \quad \Phi_2^+ \leq 0, \quad \Phi_2^+\dot{\xi} = 0, \tag{51c}$$
$$\dot{\xi}_2 \leq 0, \quad \Phi_2^- \leq 0, \quad \Phi_2^-\dot{\xi} = 0, \tag{51d}$$
$$\dot{\xi}_3 \geq 0, \quad \Phi_3 \leq 0, \quad \Phi_3\dot{\xi} = 0. \tag{51e}$$

Thus, at each material point, the state variables satisfy the nonlinear system of differential-algebraic equations (49), (50) along with the constraints (51).

Assume now, that the history of all field and internal variables at the material point is known. In particular, the values of $\boldsymbol{\varepsilon}_n$, $T_n$, $\boldsymbol{\sigma}_n$, $\boldsymbol{\varepsilon}_n^{in}$, and $\boldsymbol{\xi}_n$ are known. The subscript $n$ is used to denote a history/time parameter [5]. The new values of $\boldsymbol{\varepsilon}_{n+1}$ and $T_{n+1}$ for the strain and temperature respectively are also given [6]. The Closest Point Projection Return Mapping Algorithm is a numerical method which computes the values for $\boldsymbol{\sigma}_{n+1}$, $\boldsymbol{\varepsilon}_{n+1}^{in}$ and $\boldsymbol{\xi}_{n+1}$ by solving equations (49)–(50) along with the constraints (51). This is done by first discretizing the evolution equation (50):

$$\boldsymbol{\varepsilon}_{n+1}^{in} = \boldsymbol{\varepsilon}_n^{in} + (\xi_{2_{n+1}} - \xi_{2_n})\boldsymbol{\Lambda}^t(\boldsymbol{\sigma}_{n+1}) + (\xi_{3_{n+1}} - \xi_{3_n})\boldsymbol{\Lambda}^d(\boldsymbol{\sigma}_{n+1}), \tag{52}$$

[5] For a quasi-static problem, this would be the values at the $n$-th loading step, while in a dynamic problem this would be the values of the field and internal variables at some discrete instance of time $t_n$.

[6] Alternatively, the increments $\Delta\boldsymbol{\varepsilon}_{n+1} = \boldsymbol{\varepsilon}_{n+1} - \boldsymbol{\varepsilon}_n$ and $\Delta T_{n+1} = T_{n+1} - T_n$ may be given, which, of course, is equivalent to knowing $\boldsymbol{\varepsilon}_{n+1}$ and $T_{n+1}$.

For a geometric interpretation of this backward Euler discretization, see Qidwai and Lagoudas (2000a). The stress-strain relation (49) is then written as:

$$\boldsymbol{\sigma}_{n+1} = \boldsymbol{\mathcal{S}}(\boldsymbol{\xi}_{n+1})^{-1} : \left(\boldsymbol{\varepsilon}_{n+1} - \boldsymbol{\varepsilon}^{in}_{n+1} - \boldsymbol{\alpha}(\boldsymbol{\xi}_{n+1})(T_{n+1} - T_0)\right). \tag{53}$$

In order to solve the discrete system (52), (53) subject to the constraints (51), first substitute $\boldsymbol{\varepsilon}^{in}_{n+1}$ from equation (52) into (53) and then rearrange the terms to obtain:

$$\begin{aligned} &\boldsymbol{\mathcal{S}}(\boldsymbol{\xi}_{n+1})\boldsymbol{\sigma}_{n+1} - \boldsymbol{\varepsilon}_{n+1} + \alpha(\boldsymbol{\xi}_{n+1})\,(T_{n+1} - T_0) \\ &+\boldsymbol{\varepsilon}^{in}_{n} + (\xi_{2n+1} - \xi_{2n})\boldsymbol{\Lambda}^{t}(\boldsymbol{\sigma}_{n+1}) + (\xi_{3n+1} - \xi_{3n})\boldsymbol{\Lambda}^{d}(\boldsymbol{\sigma}_{n+1}) = \mathbf{0}. \end{aligned} \tag{54}$$

Note that in the above equation, all members with subscript $n$ as well as $\boldsymbol{\varepsilon}_{n+1}$ and $T_{n+1}$ have known values. It is convenient to introduce the residual $\mathbf{F}$:

$$\begin{aligned} \mathbf{F}(\boldsymbol{\sigma}, \boldsymbol{\xi}) =&\boldsymbol{\mathcal{S}}(\boldsymbol{\xi})\boldsymbol{\sigma} - \boldsymbol{\varepsilon}_{n+1} + \boldsymbol{\alpha}(\boldsymbol{\xi})\,(T_{n+1} - T_0) \\ &+ \boldsymbol{\varepsilon}^{in}_{n} + (\xi_2 - \xi_{2n})\boldsymbol{\Lambda}^{t}(\boldsymbol{\sigma}) + (\xi_3 - \xi_{3n})\boldsymbol{\Lambda}^{d}(\boldsymbol{\sigma}). \end{aligned} \tag{55}$$

Observe, that the system (52), (53) is now equivalent to

$$\mathbf{F}(\boldsymbol{\sigma}_{n+1}, \boldsymbol{\xi}_{n+1}) = \mathbf{0}. \tag{56}$$

The Closest Point Projection method, like most return mapping algorithms, first performs a linear thermoelastic loading using equation (53), called **thermoelastic prediction**. It then determines, using (51), if phase transformation occurs or not. If it does not, then the solution is accepted. If it does, it determines which one and performs a **transformation correction**. Without loss of generality, suppose that during the loading step the forward stress-induced phase transformation occurs and the rest of the phase transitions do not happen. This implies $\xi_{2n+1} - \xi_{2n} > 0$ and (51) reduces to

$$\Phi^{+}_{2}(\boldsymbol{\sigma}_{n+1}, T_{n+1}, \boldsymbol{\xi}_{n+1}) = 0. \tag{57}$$

The Closest Point Projection method then does nothing else but to solve (56) and (57) by Newton's method in order to obtain a consistent material state. These two steps are explained in details below. Both the predictor and corrector steps can be viewed as part of an iterative process which solves the nonlinear algebraic system of equations (56), subject to the constraints (51), by constructing a converging sequence

$$\boldsymbol{\sigma}^{(k)}_{n+1} \underset{k\to\infty}{\longrightarrow} \boldsymbol{\sigma}_{n+1}, \quad \boldsymbol{\varepsilon}^{in(k)}_{n+1} \underset{k\to\infty}{\longrightarrow} \boldsymbol{\varepsilon}^{in}_{n+1}, \quad \boldsymbol{\xi}^{(k)}_{n+1} \underset{k\to\infty}{\longrightarrow} \boldsymbol{\xi}_{n+1}. \tag{58}$$

### *5.1 Thermoelastic prediction*

As the first step, a **thermoelastic prediction** is performed during which, the internal variables do not change:

$$\boldsymbol{\varepsilon}^{in(0)}_{n+1} = \boldsymbol{\varepsilon}^{in}_{n}, \tag{59}$$

$$\boldsymbol{\xi}^{(0)}_{n+1} = \boldsymbol{\xi}_{n}, \tag{60}$$

$$\boldsymbol{\sigma}^{(0)}_{n+1} = \boldsymbol{\mathcal{S}}(\boldsymbol{\xi}_{n})^{-1} : \left[\boldsymbol{\varepsilon}_{n+1} - \boldsymbol{\varepsilon}^{in}_{n} - \boldsymbol{\alpha}(\boldsymbol{\xi}_{n})(T_{n+1} - T_0)\right]. \tag{61}$$

It should be noted, that this first step corresponds to purely thermoelastic loading without any transformation ($\dot{\boldsymbol{\xi}} = 0$), hence its name thermoelastic prediction. The corresponding values of the five transformation functions are then evaluated:

$$\Phi_\alpha^{(0)} = \Phi_\alpha(\boldsymbol{\sigma}_{n+1}^{(0)}, T_{n+1}, \boldsymbol{\xi}_{n+1}^{(0)}). \tag{62}$$

The subscript $\alpha$ is understood in the sense $\Phi_\alpha \in \left\{\Phi_1^+, \Phi_1^-, \Phi_2^+, \Phi_2^-, \Phi_3\right\}$. If the value of all transformation functions satisfy $\Phi_\alpha^{(0)} \leq 0$ then all equations and constraints are satisfied and the iteration is terminated for $k = 0$.

### *5.2 Transformation correction*

The predictor step assumed that $\boldsymbol{\xi}_n = \boldsymbol{\xi}_{n+1}$, hence (51) are satisfied if and only if all $\Phi_\alpha \leq 0$. Therefore, if at least one of the transformation functions $\Phi_\alpha > 0$ then the corresponding consistency condition is violated. Such surfaces will be referred to as *inconsistent*. The existence of inconsistent surfaces implies that during the loading step, phase transformation takes place and a **transformation correction** is needed (Qidwai and Lagoudas, 2000a). During this step, the stress and the internal variables are modified in accordance with the transformation flow rules so that the consistency conditions are satisfied.

The consistency condition(s) which correspond to the phase transformation(s) taking place will be called *active*. The same term will be used for the respective transformation surfaces. The consistency conditions that the elastic predictor violates are not necessarily the active ones, nor are they necessarily the only ones active. For examples of such cases, as well as an approach how to find the active surfaces, the reader is referred to Popov (2005).

So, assume that it is known which transformation(s) are active during the load step. We will consider the case of a single or two simultaneous active surfaces. Suppose first, that only one transformation is active, say $\Phi_\alpha$. This implies that the corresponding volume fraction, denoted also by $\xi_\alpha$, has nonzero rate[7]. That is, $\dot{\xi}_\alpha \neq 0$, and the corresponding consistency conditions (51) is satisfied, if and only if,

$$\Phi_\alpha(\boldsymbol{\sigma}_{n+1}, T_{n+1}, \boldsymbol{\xi}_{n+1}) = 0. \tag{63}$$

Therefore, during the transformation correction, one has to solve (56) along with the last equation. This is done by Newton's method: *For the given k-th iterate of* $\boldsymbol{\sigma}_{n+1}^{(k)}$, $\boldsymbol{\varepsilon}_{n+1}^{in(k)}$ *and* $\boldsymbol{\xi}_{n+1}^{(k)}$, *find the* $k+1$ *iterates by linearizing* $\mathbf{F}$ *and* $\Phi_\alpha$ *around* $\left(\boldsymbol{\sigma}_{n+1}^{(k)}, \boldsymbol{\xi}_{n+1}^{(k)}\right)$ *and requiring that:*

$$\mathbf{F}^{(k)} + \frac{\partial \mathbf{F}^{(k)}}{\partial \boldsymbol{\sigma}} : \Delta\boldsymbol{\sigma}^{(k)} + \frac{\partial \mathbf{F}^{(k)}}{\partial \xi_\alpha} \Delta\xi_\alpha^{(k)} = 0, \tag{64}$$

$$\Phi_\alpha^{(k)} + \frac{\partial \Phi_\alpha^{(k)}}{\partial \boldsymbol{\sigma}} : \Delta\boldsymbol{\sigma}^{(k)} + \frac{\partial \Phi_\alpha^{(k)}}{\partial \xi_\alpha} \cdot \Delta\xi_\alpha^{(k)} = 0. \tag{65}$$

*The shortcut notation for* $\mathbf{F}^{(k)} = \mathbf{F}\left(\boldsymbol{\sigma}_{n+1}^{(k)}, \boldsymbol{\xi}_{n+1}^{(k)}\right)$, $\Phi_\alpha^{(k)} = \Phi_\alpha\left(\boldsymbol{\sigma}_{n+1}^{(k)}, T_{n+1}, \xi_{n+1}^{(k)}\right)$ *and all their derivatives is used. When the increments* $\Delta\boldsymbol{\sigma}^{(k)}$ *and* $\Delta\xi_\alpha^{(k)}$ *are determined from the above*

[7] If $\Phi_\alpha \in \left\{\Phi_1^+, \Phi_1^-\right\}$ then the internal variable responsible is $\xi_\alpha = \xi_1$, if $\Phi_\alpha \in \left\{\Phi_2^+, \Phi_2^-\right\}$ then $\xi_\alpha = \xi_2$ and if $\Phi_\alpha = \Phi_3$ then $\xi_\alpha = \xi_3$

*system of linear equations, the stress and the internal variable are updated according to*

$$\boldsymbol{\sigma}_{n+1}^{(k+1)} = \boldsymbol{\sigma}_{n+1}^{(k)} + \Delta\boldsymbol{\sigma}^{(k)}, \quad \xi_{\alpha}{}_{n+1}^{(k+1)} = \xi_{\alpha}{}_{n+1}^{(k)} + \Delta\xi_{\alpha}^{(k)},$$

*and* $\boldsymbol{\varepsilon}_{n+1}^{in(k)}$ *is updated according to equation (52).*

If two of the transformations are active, say $\Phi_\alpha$ and $\Phi_\beta$, then during the correction, equation (56) along with

$$\Phi_\alpha(\boldsymbol{\sigma}_{n+1}, T_{n+1}, \boldsymbol{\xi}_{n+1}) = 0, \tag{66}$$

$$\Phi_\beta(\boldsymbol{\sigma}_{n+1}, T_{n+1}, \boldsymbol{\xi}_{n+1}) = 0, \tag{67}$$

is being solved, again by a Newton's method: *For the given k-th iterate of* $\boldsymbol{\sigma}_{n+1}^{(k)}$, $\boldsymbol{\varepsilon}_{n+1}^{in(k)}$ *and* $\boldsymbol{\xi}_{n+1}^{(k)}$ *find the* $k+1$ *iterates by linearizing* $\mathbf{F}$, $\Phi_\alpha$ *and* $\Phi_\beta$ *around* $\left(\boldsymbol{\sigma}_{n+1}^{(k)}, \boldsymbol{\xi}_{n+1}^{(k)}\right)$ *and requiring that:*

$$\mathbf{F}^{(k)} + \frac{\partial \mathbf{F}^{(k)}}{\partial \boldsymbol{\sigma}} : \Delta\boldsymbol{\sigma}^{(k)} + \frac{\partial \mathbf{F}^{(k)}}{\partial \xi_\alpha}\Delta\xi_\alpha^{(k)} + \frac{\partial \mathbf{F}^{(k)}}{\partial \xi_\beta}\Delta\xi_\beta^{(k)} = 0, \tag{68}$$

$$\Phi_\alpha^{(k)} + \frac{\partial \Phi_\alpha^{(k)}}{\partial \boldsymbol{\sigma}} : \Delta\boldsymbol{\sigma}^{(k)} + \frac{\partial \Phi_\alpha^{(k)}}{\partial \xi_\alpha} \cdot \Delta\xi_\alpha^{(k)} + \frac{\partial \Phi_\alpha^{(k)}}{\partial \xi_\beta} \cdot \Delta\xi_\beta^{(k)} = 0, \tag{69}$$

$$\Phi_\beta^{(k)} + \frac{\partial \Phi_\beta^{(k)}}{\partial \boldsymbol{\sigma}} : \Delta\boldsymbol{\sigma}^{(k)} + \frac{\partial \Phi_\beta^{(k)}}{\partial \xi_\alpha} \cdot \Delta\xi_\alpha^{(k)} + \frac{\partial \Phi_\beta^{(k)}}{\partial \xi_\beta} \cdot \Delta\xi_\beta^{(k)} = 0. \tag{70}$$

*When the increments* $\Delta\boldsymbol{\sigma}^{(k)}$, $\Delta\xi_\alpha^{(k)}$ *and* $\Delta\xi_\beta^{(k)}$ *are determined from the above system of linear equations, the stress and the internal variable are updated according to*

$$\boldsymbol{\sigma}_{n+1}^{(k+1)} = \boldsymbol{\sigma}_{n+1}^{(k)} + \Delta\boldsymbol{\sigma}^{(k)}, \quad \xi_{\alpha}{}_{n+1}^{(k+1)} = \xi_{\alpha}{}_{n+1}^{(k)} + \Delta\xi_{\alpha}^{(k)}, \quad \xi_{\beta}{}_{n+1}^{(k+1)} = \xi_{\beta}{}_{n+1}^{(k)} + \Delta\xi_{\beta}^{(k)},$$

*and* $\boldsymbol{\varepsilon}_{n+1}^{in(k)}$ *is updated according to equation (52).*

This completes the outline the numerical implementation of the model. Several important aspects of the numerical implementation will not be given here. These include the details of solving the linear system (64)-(65) or (68)-(70), including the functional form of the derivatives involved; an approach to finding the active surfaces; the important aspect of completing the transformations (the internal variables $\xi$ are bounded by the constraints (2), so they cannot evolve indefinitely); algorithmic tangent moduli that are useful when coupling this numerical scheme with a finite element method (c.f. e.g. Simo and Hughes, 1998); integration of the closest point projection algorithm into a displacement based FEM. For details, the reader is referred to Popov (2005).

It is important to note that when $\Phi_2^{\pm}$ is the only active surface, the iteration (64),(65) reduces to the Closest Point Projection method of Qidwai and Lagoudas (2000a). In the later work, the algorithm is formulated by defining a residual for the flow rule (52), instead of (54). It is easy to show that the two lead to the same algorithm. The current approach has the advantage that by taking the residual of Hooke's law the algorithm generalizes for the twinning transformation $A \leftrightarrow M^t$ in which no transformation strain is generated.

Table 1
Material parameters used in the SMA model

| Material constant | Value | Material constant | Value | Material constant | Value |
|---|---|---|---|---|---|
| $E^A$ | $70 \times 10^9$ Pa | H | 0.05 | $A_f^t$, $A_f^d$ | 315 $^\circ K$ |
| $E^M$ | $30 \times 10^9$ Pa | $k$ | $4.5 \times 10^6$ $Pa/(m^3K)$ | $A_s^t$, $A_s^d$ | 295 $^\circ K$ |
| $\alpha^A$ | $22 \times 10^{-6}/K$ | $M_f$ | 275 $^\circ K$ | $\sigma_s$ | 100 $MPa$ |
| $\alpha^M$ | $10 \times 10^{-6}/K$ | $M_s$ | 291 $^\circ K$ | $\sigma_f$ | 200 $MPa$ |

## 6 Numerical examples

The numerical examples in this section were selected so that complex loading paths in stress-temperature phase space could be tested. First, a uniaxial example (Section 6.1) of a constrained SMA rod is considered. One-dimensional setting allows to easily determine the loading path in stress-temperature space, and solve the relevant equations by symbolic software. In the example, an SMA rod is cooled from the austenitic phase to low temperature while the strain is kept constrained. This problem allows to demonstrate the cut-off of the $A \rightarrow M^d$ transformation in untrained SMA materials and the predominant development of $M^t$ from $A$ at low stress levels. Secondly, the constitutive model was then numerically implemented in the full 3-D setting and was tested on a plane strain problem. An SMA block with a square hole was loaded at low temperature and then heated while constrained, so that multiple transformations could take place. The results are described in Section 6.2. The basic material parameters used in both examples are given in Table 1 and represent a generic SMA properties (Qidwai and Lagoudas, 2000a). The selection of same critical temperatures for $M^d \rightarrow A$ and $M^t \rightarrow A$ was done in order to maximize the intersection region of the two transformation strips (c.f. the phase diagram of Figure 4) and thus to test the model for multiple transformation in the most severe case from numerical point of view.

### *6.1 Constrained cooling of an SMA rod*

In order to get a feeling of the thermomechanical response predicted by this model first a very simple example is presented. Consider a rod in uniaxial stress state (B-1), (B-2). The rod is first loaded mechanically to $200MPa$ from a stress free, fully austenitic configuration, at a constant temperature of 320 $^\circ K$. Then the two ends of the rod are fixed and it is cooled to a temperature of 260 $^\circ K$ which is well below the $M_f$ temperature. The loading process is plotted in the stress-temperature space (Figure 6).

Due to the uniform stress state this problem is simple enough and can be solved semi-analytically. The stress in the rod is related to the strain by (c.f. equations (17), (23) and (B-1)):

$$\sigma = E(c_1 + c_2)\left(\varepsilon - \alpha(c_1 + c_2)(T - T_0) - \varepsilon^{in}\right) \tag{71}$$

The maximal detwinning and transformation strains are the same, e.g. $H^t = H^d = H$ (see Table 1). In the uniaxial case, the inelastic strain is proportional to the volume fraction of detwinned martensite $c_2$ (c.f. equations (8)-(10) and (B-2)):

$$\varepsilon^{in} = Hc_2. \tag{72}$$

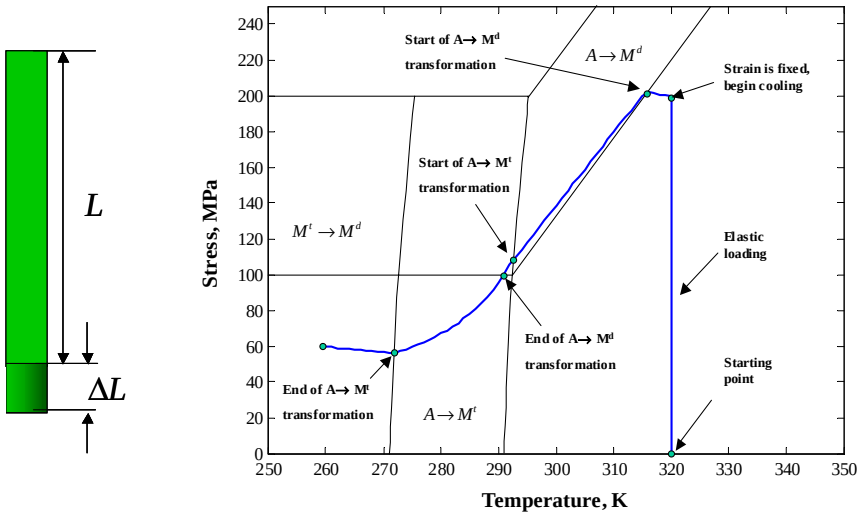



Fig. 6. A constrained cooling path in stress-temperature space. The rod is loaded in tension at the austenitic phase to a stress lower than required for phase transformation. The strain is then fixed and the rod is cooled. The rapid drop of the stress during the phase transformation is caused by the development of inelastic strains. Since the total achievable inelastic strain is an order of magnitude larger than the initial elastic strain, very little $A \to M^d$ transformation occurs. For clarity, only the $A \to M^d$, $A \to M^t$, $M^t \to M^d$ and transformation strips are shown.

For the particular example under consideration, the rod is initially in the austenitic phase up to a stress $\sigma_0 = 200MPa$, which is below the critical stress required to initiate the forward, $A \to M^d$ phase transformation. Without loss of generality, let this be a tensile stress. Then the inelastic strain is identically zero:

$$\varepsilon^{in} = 0.$$

and from equation (71), the rod has developed uniform elastic strain

$$\varepsilon_0 = \sigma_0 / E^A.$$

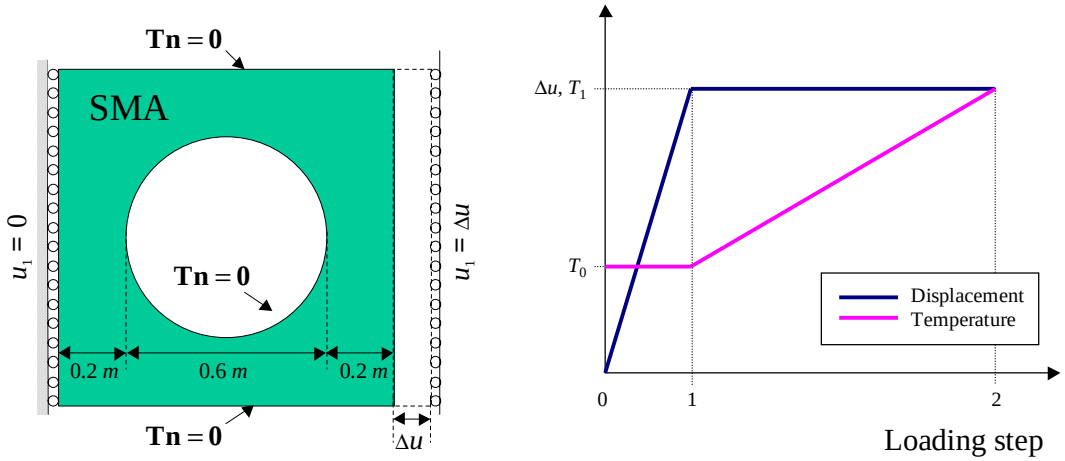
At this point of the loading path, the strain is then fixed and the SMA is gradually cooled. At first, a thermoelastic contraction of the rod slightly increases the stress. When the $A \to M^d$ transformation surface is reached, transformation strains begin to develop. At this point one has to solve equations (71) and (72), along with the rule of mixtures (17) and the relevant transformation surfaces (51). This is done using symbolic software (Mathematica). The material state during the entire loading path is plotted in stress-temperature space in Figure 6 and the relevant transformation strips are also shown.

Now, as the transformation strains becomes nonnegative it will relax the stress state. Observe that the maximum possible value of the transformation strain $H$ is an order of magnitude larger than the elastic strain $\varepsilon_0$ (which is kept fixed during the cooling). Therefore, very little phase transformation is required to produce transformation strains comparable with $\varepsilon_0$ and, thus to drastically reduce the stress. In this example the $A \to M^d$ surface terminates at some finite value of stress $\sigma_s$ (which, as discussed before is material dependent). Slightly before this point the $A \to M^t$ transformation surface is also reached and the material undergoes combined transformation.

As the stress decreases below the critical stress $\sigma_s$, only the $A \rightarrow M^t$ transformation proceeds. In the process, no further transformation strain is produced, however the stiffness changes. The stiffness of the martensite $E^M$ is less than the stiffness of austenite $E^A$, so the effective stiffness decreases (c.f. equation (17)). On the other hand, the total strain is fixed. Therefore, neglecting the thermal strains, and noting that very small amount of $A \rightarrow M^d$ has occurred, from equation (71) it follows that the stress in the rod will decrease by a factor of $E^A/E^M$. This is clearly visible in Figure 6. Upon completion of the transformation, the a material again exhibits thermoelastic contraction, which causes small increase in stress. For additional uniaxial examples which illustrate multiple transformations while heating a constrained rod, the reader is referred to Popov (2005).

### *6.2 Complex thermomechanical loading path*

The numerical example in this section involves a complex thermal and mechanical loading path applied to an SMA block with a cylindrical hole. A 2D cross-section of the geometry is shown in Figure 7(a). It is assumed that in the off-plane direction the body is constrained, so that plane strain conditions are achieved in the cross-section. As is usual for plane strain, the stress is still three-dimensional, and in light of equations (46), (45) and (11) so are the inelastic strains (8). This, combined with both thermal and mechanical loading applied to the body, allows to test the SMA model during a complex loading path. In addition, it will be shown in this section, the thermal loading in conjunction with the plane strain conditions also leads to evolution of non-proportional stress state in the body.



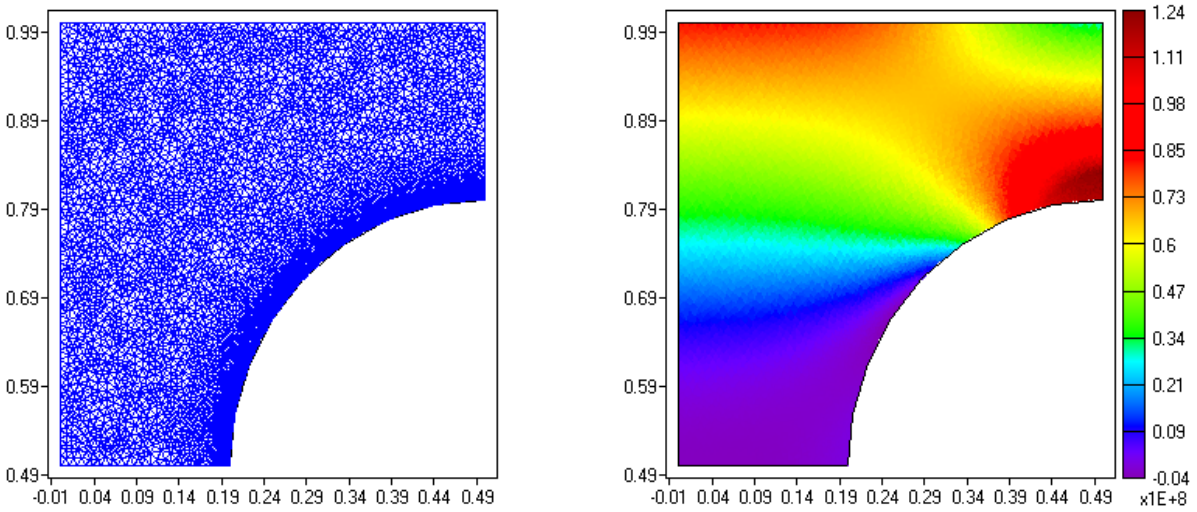

(a) Problem geometry and initial boundary conditions

(b) Loading path followed for the boundary conditions and temperature

Fig. 7. A plane strain perforated square model problem

The numerical example is solved using a displacement based Finite Element Method (FEM). The constitutive model was implemented numerically using return mapping algorithms (Section 5). All discretizations use triangular meshes and standard linear Lagrangian finite element spaces. The SMA material properties used represent generic NiTi SMA material used previously in the literature (Qidwai and Lagoudas, 2000a) and are given in Table 1.

Initially, the SMA block is stress free and in the self-accommodated phase, i.e. $c_1 = 1$ everywhere in the domain. The thermomechanical loading that the square is subjected to




(a) Mesh, $h \sim 0.01$, 9274 DOF (b) $\sigma_{11}$

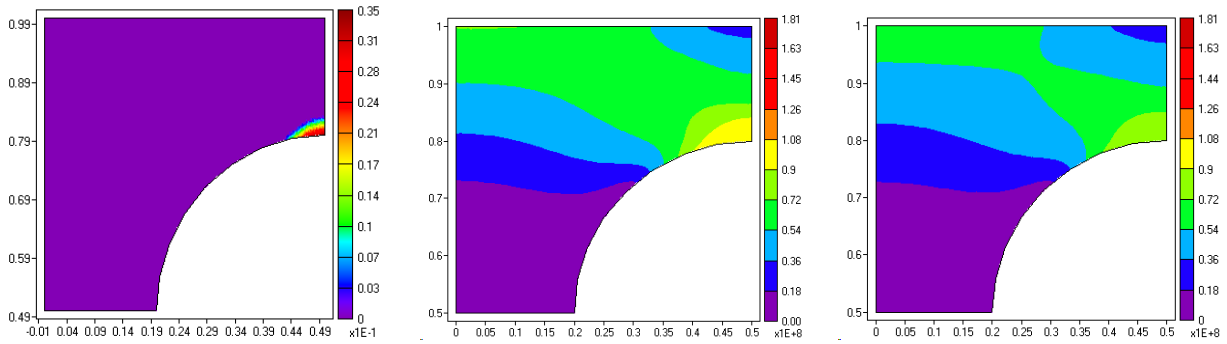
Fig. 8. The triangular FEM mesh selected for the entire computation and associated stresses and internal variable at the end of first loading step ($T = 260\ ^\circ K$).

is shown on Figure 7(b). The block is first loaded mechanically at constant temperature $T = 260\ ^\circ K$ as follows: the left side of the square is fixed against horizontal displacement, but is allowed to roll in the vertical direction; the right side is pulled uniformly by the amount of $0.002m$ in the horizontal direction and the side is again allowed to move in the vertical direction; the remaining part of the boundary (including the hole) is stress free. The second loading step consists of keeping the horizontal component of the displacement fixed and uniformly raising the temperature to $T = 350\ ^\circ K$.

The first loading step was used to determine a suitable mesh size for the entire simulation. This was done by starting with a very coarse mesh and consecutively refining it. Due to obvious symmetry consideration, only one quarter of the domain was used in the calculations. The solution was judged to be accurate enough when the non-smoothed, piece-wise constant stress components in each element showed little variation over element boundaries. Based on this, the mesh shown in Figure 8 (8964 elements, 9274 DOF) was selected for the rest of the computation.

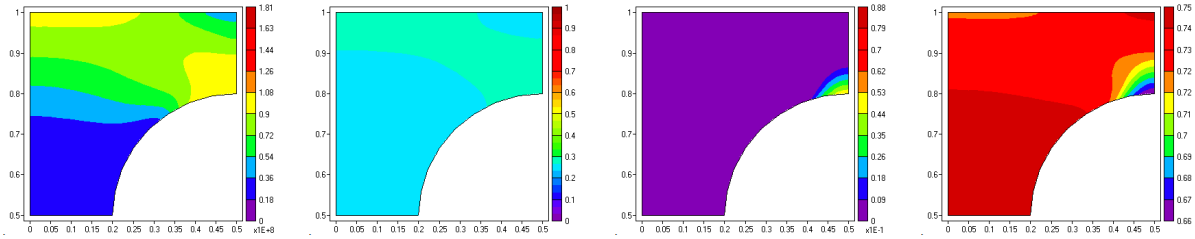
It can be seen from this first part of the simulation that stress concentrations develop near the top and bottom edge of hole. The effective stress in these locations become sufficiently high for the detwinning of small amounts of self-accommodated martensite as shown on Figure 9. If the stress concentrations are compared to a pure elastic solution, the development of inelastic strains in the detwinning process tends to reduce this stress concentration.

During the heating process, the material undergoes two distinct processes - first simultaneous $M^t \rightarrow A, M^d$ the transformations then followed by a $M^d \rightarrow A$ transformation. When the second loading step begins, the material first experiences initial linear thermoelastic expansion. The conditions are of plane strain, the horizontal component of the displacements is fixed and the vertical displacements are not constrained during the heating. Hence any thermal expansion will results in a nonhomogeneous change in the stress state. During the linear thermoelastic expansion ($T < A_s^t$) this results in small relaxation of the effective stress (Figure 9(c)). As the critical temperature for the $M^t \rightarrow A$ transformation is reached. The



(a) $c_2$ at $T = 260, 280\ ^\circ K$ (b) Von-Mises Stress, $T = 260\ ^\circ K$ (c) Von-Mises Stress, $T = 280\ ^\circ K$

Fig. 9. The martensitic volume fraction (a) and the Von-Mises stress (b) at the completion of the initial loading. As the heating begins the material undergoes only thermal expansion and no transformation occurs. As a result the Von-Mises stress decreases slightly by $T = 280\ ^\circ K$ and the internal variables do not change.

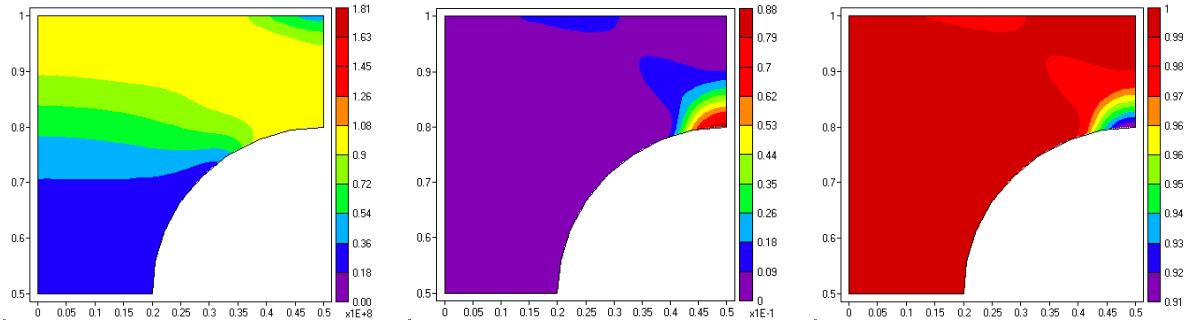


(a) Von-Mises stress (b) $c_1$ $(M^t)$ (c) $c_2$ $(M^d)$ (d) $c_3$ $(A)$

Fig. 10. Solution for square hole problem at $T = 310\ ^\circ K$. As the material is slowly heated, the reverse transformation $M^t \to A$ occurs (c). Since the austenite is considerably stiffer than martensite, and the material is constrained, the stresses also increase. This results in a simultaneous $M^t \to A, M^d$ transformation, which is manifested in increase in the volume fraction of $M^d$ (a,b).

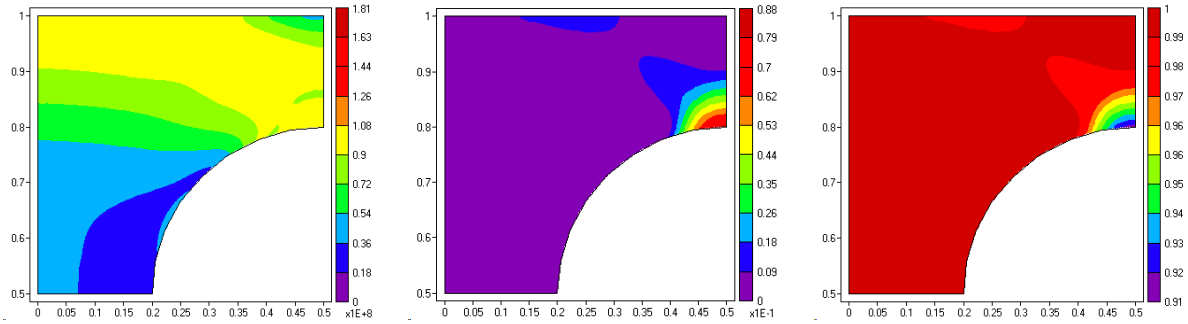
self-accommodated martensite begins to transform to austenite. The stiffness of austenite is approximately 2.3 times that of martensite (see Table 1) and due to the fixed horizontal displacement the stresses increase throughout the block. The effective stress increases correspondingly and this causes further detwinning of martensite in some areas of the block, resulting in a simultaneous $M^t \to A, M^d$ transformations, mostly near the top and bottom parts of the hole (Figure 10). This is manifested in further increase in $c_2$ in comparisons to the amount that was produced during the first loading step. The effective stress and $c_2$ after the completion of the $M^t \to A$ transformation are shown on Figure 11.

The areas, where detwinned martensite is present have generally higher effective stress (above $\sigma_s$), compared to the rest of the domain. Hence, the reverse transformation of detwinned martensite ($M^d \to A$) does not happen until much higher temperature, due to the fact that the corresponding transformation surfaces exhibit stress dependence (see Figure 4). Around $T = 335\ ^\circ K$, the $M^d \to A$ transformation begins in areas with lowest effective stress. During this phase, the inelastic strains decrease according to the transformation rule (9),(46). Again, due to the constraint on the displacements, imposed by the boundary conditions, the elastic portion of the stress generally increases which leads to a corresponding (non-uniform) increase in the stress during the reverse transformation (Figures 12-14).



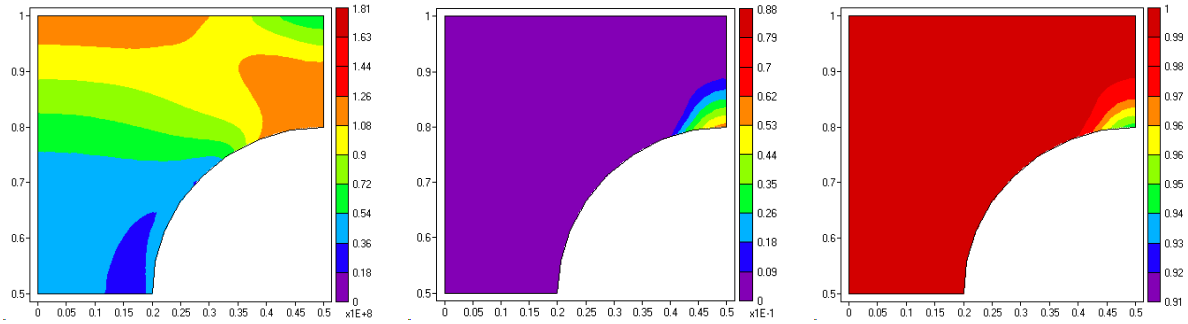
(a) Von-Mises Stress (b) $c_2$ ($M^d$) (c) $c_3$ ($A$)

Fig. 11. By $T = 320\ ^\circ K$ the twinned martensite is depleted. The volume fraction of $M^d$ (b) has reached approximately 11% and the rest is in the austenitic phase (c).



(a) Von-Mises Stress (b) $c_2$ ($M^d$) (c) $c_3$ ($A$)

Fig. 12. Around $T = 335\ ^\circ K$ the temperature is sufficient for the reverse, $M^d \rightarrow A$ transformation to begin.

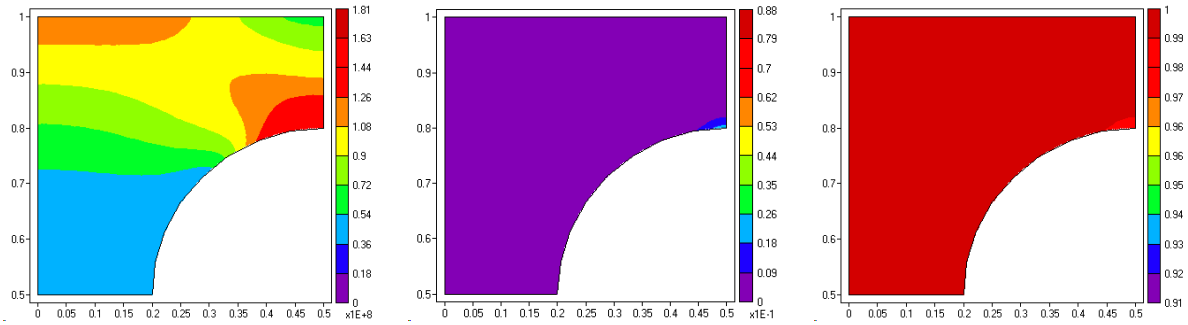


(a) Von-Mises Stress (b) $c_2$ ($M^d$) (c) $c_3$ ($A$)

Fig. 13. Further heating ($T = 340\ ^\circ K$ causes further $M^d \rightarrow A$ transformation and subsequent decrease in $c_2$ (b) and the inelastic strain $\boldsymbol{\varepsilon}^{in}$, and increase of $c_3$ (c). Due to the constrained displacement and the decrease of the inelastic strain, the stresses begin to increase (a)

## 7 Conclusions

A new 3-D constitutive model for polycrystalline SMAs based on thermodynamic potentials is presented. The model can account for both development of stress induced martensite directly from austenite (pseudoelasticity) as well as detwinning of twinned martensite. This is accomplished by describing the material state as a mixture of three phases - twinned martensite, detwinned martensite and austenite and by using the three possible "reactions" between these phases as internal variables.



(a) Von-Mises Stress (b) $c_2$ $(M^d)$ (c) $c_3$ $(A)$

Fig. 14. The $M^d \rightarrow A$ is complete almost everywhere in the domain (c) by the time the temperature reaches $T = 350\ ^\circ K$. Note that the maximum value for the effective stress has reached increased to approximately $166 MPa$ (a).

The model is made consistent with a modified phase diagram in stress-temperature space. A key new experimental finding is the existence of separate reverse transformation temperatures for detwinned and twinned martensite. This is obtained through a series of calorimetric measurements and is incorporated in the model. The phase diagram incorporates a single transformation strip for the $M^t \rightarrow M^d$ inelastic deformation over the temperature range $T \leq A_f^t$. It also assumed the same temperature independent transformation regions for the $A \rightarrow M^t$ and $M^t \rightarrow A$ phase transformations.

The constitutive model was numerically implemented using return mapping algorithms. The implementation was integrated into an numerical implementation and tested for several model problems. Presented in this work two cooling/heating loading paths of a rod in uniaxial stress state. In order to demonstrate the 3-D capabilities of the model a complex loading path for a perforated square under conditions of plane strain was also presented.

There are several areas for future work on this types of models. On the experimental side it is necessary to further characterize the differences between $M^d \rightarrow A$ and $M^t \rightarrow A$ transformations for different material compositions and different material training, and the evolution of this difference with cycling loading. Also Sakamoto (2002) suggests a size effect, so it will therefore be of particular interest to correlate the observed difference of the $M^t \rightarrow A$ and $M^d \rightarrow A$ transformation temperatures with the diameter of the SMA wires.

On the modeling side, the selection made for the variables the hardening functions depend on are not the only possible ones. The current selection imposes certain constraints on the arrangement of the transformation regions in the phase diagram. Other choices, for example a dependence of $f_1$ (the $A \leftrightarrow M^t$ hardening function) on $c_3$, rather than $c_1$ is possible and should be explored. Further, the model should be tested for the classes of SMA materials for which the critical temperatures $M_s$, $M_f$, $A_s^t$, $A_f^t$ define overlapping regions.

Also, the phase diagram itself evolves as the material is cycled through a certain thermomechanical loading path. The evolution is fairly well understood in the special case of pseudoelastic loading paths. More general cases however are not explored either from a modeling or an experimental point of view. It should me noted that translational movements of the transformation strips of the phase diagram can be accomplished by using $\xi_i$ instead of $c_i$ as the independent variables for the hardening functions.

**Acknowledgements**

The authors would like to acknowledge the financial support of the Air Force Office of Scientific Research, grant F49620-01-0196, P00005, as well as NASA URETI grant NCC-1-02038. We would also specifically like to acknowledge the contributions of P. Kumar (Department of Aerospace Engineering, Texas A&M University, College Station, TX 77843, *email: parik@tamu.edu*) for carrying out the experiments of Section 2.

## Appendix A: The time derivative (rate) of the free energy

We consider the set of thermodynamical processes for which the functions $f_1$, $f_2$ and $f_3$ are smooth in their first argument ($\boldsymbol{\xi}$). Further more we assume that either of the phase transformation from $A$ to $M^t$ and $M^d$ can be reversed a countable number of times. That is, there exist (possibly infinitely many) occasions in time $T_0 < T_1 \leq T_3 \cdots \leq T_{2n} < T_{2n+1} \ldots$ where $\dot{\xi}_i = 0$, for $i = 1$ and/or $i = 2$ and is strictly positive or strictly negative on each interval $(T_{2n}, T_{2n+1})$. Observe that only the term $G^{\text{mix}}$ in the definition of the free energy (16) is not a smooth function of its variables. Therefore, to show the identity (19), it is sufficient to show that:

$$\dot{G}^{\text{mix}} = \sum_{i=1}^{3} \frac{\partial G^{\text{mix}}}{\partial \xi_i} \dot{\xi}_i(t). \tag{A-1}$$

First, using the definition of $G^{\text{mix}}$ given by equation (18), its rate is directly computed as:

$$\dot{G}^{\text{mix}}\left(\boldsymbol{\xi}, \operatorname{sgn}\left(\dot{\xi}_1\right), \operatorname{sgn}\left(\dot{\xi}_2\right)\right) = \frac{1}{\rho} \sum_{i=1}^{3} \left[ f_i\left(\boldsymbol{\xi}; \operatorname{sgn}\left(\dot{\xi}_i\right)\right) \dot{\xi}_i(t) \right]. \tag{A-2}$$

Assume now that $t \in [T_{2n}, T_{2n+1}]$ and let the values of $\xi_i$ at the beginning of this interval be denoted by $\Xi_i$, $i = 1, 2$, that is, $\Xi_1 = \xi_1(T_{2n})$ and $\Xi_2 = \xi_2(T_{2n})$. Without loss of generality, suppose that $\dot{\xi}_i > 0$ in the interval $[T_{2n}, T_{2n+1}]$. Then one has:

$$\int_{T_{2n}}^{t} f_i\left(\boldsymbol{\xi}, \operatorname{sgn}\left(\dot{\xi}_i\right)\right) \dot{\xi}_i(\tau) d\tau = \int_{\Xi_i}^{\xi_i} f_i\left(..., \eta, ..., 1\right) d\eta,$$

On the other hand,

$$\frac{\partial}{\partial \xi_i} \int_{\Xi_i}^{\xi_i} f_i\left(..., \eta, ..., 1\right) d\eta = f_i\left(\xi_1, \xi_2, \xi_3, 1\right), \text{ for all } \xi_i \in [\Xi_1, \xi_i(T_{2n+1})].$$

Now, $G^{\text{mix}}$ is a path-dependent functional of $\boldsymbol{\xi}$. Therefore, the derivative $\partial_{\xi_i} G^{\text{mix}}$ with respect to the current state $\boldsymbol{\xi}$ (and the path which lead to it) can only be meaningfully defined by continuing the current path through $\boldsymbol{\xi}$ and allowing only $\xi_i$ to change. That is, $\dot{\xi}_i \neq 0$ and $\dot{\xi}_j = 0$, $j \neq i$ for times greater then $t$. For such paths only the term containing $\xi_i$ in the integral expression in equation (18) will change past the point $\boldsymbol{\xi}$ and therefore:

$$\begin{aligned} \rho \frac{\partial G^{\text{mix}}}{\partial \xi_i} = \frac{\partial}{\partial \xi_i} \left[ \int_0^t f_i\left(\boldsymbol{\xi}; \operatorname{sgn}\left(\dot{\xi}_i\right)\right) \dot{\xi}_i(\tau) d\tau \right] &= \frac{\partial}{\partial \xi_i} \left[ \int_{\Xi_i}^t f_i\left(\boldsymbol{\xi}; \operatorname{sgn}\left(\dot{\xi}_i\right)\right) \dot{\xi}_i(\tau) d\tau \right] \\ &= f_i\left(\boldsymbol{\xi}; \operatorname{sgn}\left(\dot{\xi}_i\right)\right). \end{aligned} \tag{A-3}$$

The last equation, combined with equation (A-2), leads to the identity (A-1). Note that one may want to avoid defining the derivatives $\partial_{\xi_i} G^{\text{mix}}$ along specific paths altogether and instead

split the free energy $G$ into a potential part (in our case, $(c_1+c_2)G^M(\boldsymbol{\sigma},T,\boldsymbol{\xi})+c_3G^A(\boldsymbol{\sigma},T,\boldsymbol{\xi})$) and a path dependent functional ($G^{\text{mix}}$). The path dependent part is then defined as a function of time only, i.e. $G^{\text{mix}}(t) := G^{\text{mix}}(\boldsymbol{\xi}(t))$, whose rate can be explicitly calculated. Thus the rate $\dot{G}$ is known and can be used in the entropy inequality (20).

## Appendix B: Uniaxial reduction of the model

In order to express the model parameters in terms of the physical quantities defining the phase diagram, it is necessary to write the transformation surfaces explicitly in the uniaxial stress case. Assume a stress state

$$\sigma_{11} = \sigma, \quad \sigma_{12} = ... = \sigma_{33} = 0. \tag{B-1}$$

Since uniaxial loading is always proportional, any combination of detwinning $M^t \to M^d$, forward $A \to M^d$ or reverse $M^d \to A$ by virtue of (45) and/or (46) will result in a transformation direction tensor:

$$\boldsymbol{\Lambda}^{t,d}_{11} = H, \quad \boldsymbol{\Lambda}^{t,d}_{22} = \boldsymbol{\Lambda}^{t,d}_{33} = -\frac{1}{2}H, \quad \boldsymbol{\Lambda}^{t,d}_{12} = \boldsymbol{\Lambda}^{t,d}_{13} = \boldsymbol{\Lambda}^{t,d}_{23} = 0. \tag{B-2}$$

With this in mind, and in light of equations (28)-(30) and (42)-(44), the inequalities (39)-(41) take the form:

$$\hat{\pi}(\sigma,T) - \Delta_1^+ f_1^+(c_1) \le Y_1^+ \tag{B-3}$$
$$-\hat{\pi}(\sigma,T) + \Delta_1^- f_1^-(c_1) \le Y_1^- \tag{B-4}$$
$$\sigma H + \hat{\pi}(\sigma,T) - \Delta_2^+ f_2^+(c_2) \le Y_2^+ \tag{B-5}$$
$$-\sigma H - \hat{\pi}(\sigma,T) + \Delta_2^- f_2^-(c_2) \le Y_2^- \tag{B-6}$$
$$\sigma H - \Delta_3 f_3(c_2) \le Y_3 \tag{B-7}$$

where,

$$\begin{aligned}\hat{\pi}(\sigma,T) =& \Delta S\sigma^2 + \Delta\alpha\sigma(T-T_0) \\ &- \rho\Delta c\left[(T-T_0) - T\ln\left(\frac{T}{T_0}\right)\right] + \rho\Delta s_0 T - \rho\Delta u_0.\end{aligned} \tag{B-8}$$

The material constants $\Delta S$, $\Delta\alpha$, $\Delta s_0$ and $\Delta u_0$, entering the expression for $\hat{\pi}$, are known from either mechanical or calorimetric measurements (See Section 4.5), hence $\hat{\pi}$ is a well defined function of $\sigma$ and $T$. Whenever one or more transformations are taking place (that is, $\dot{\xi}_i \neq 0$, $i = 1,2,3$) the respective inequalities (B-3)-(B-7) turn into equalities (c.f. conditions (38)).

The unknown quantities to be determined are $Y_1^+$, $Y_1^-$, $Y_2^+$, $Y_2^-$, $Y_3$, $\Delta_1^+$, $\Delta_1^-$, $\Delta_2^+$, $\Delta_2^-$, $\Delta_3$ which are responsible for the relative position and width of the transformation strips in the phase diagram. The transformation strips on the other hand are completely determined by the start and finish detwinning stresses $\sigma_s$ and $\sigma_f$, the transformation temperatures $M_s$, $M_f$, $A_s^t$, $A_f^t$, $A_s^d$, $A_f^d$, as well as the start and finish temperatures $T_s(\sigma_f)$ and $T_f(\sigma_f)$ for the $A \to M^d$ transformation at constant (uniaxial) stress $\sigma = \sigma_f$ [8]. Finally, the functional form of $f_i$ has to be determined as well.

[8] In this model there is no assumption of a triple point, so the $A \to M^d$ strip can be located at or to the right of the intersection point of the finish lines for the $A \to M^t$ and $M^t \to M^d$ transformations

The model parameters are then established as follows. Consider a loading path in which a purely twinned SMA ($c_1 = 1$, $c_2 = c_3 = 0$ is loaded at temperature below $M_f$. As the detwinning deformation progresses, $\dot{\xi}_3 > 0$, and the inequality (B-7) becomes an equality:

$$\sigma H - \Delta_3 f_3(c_2) = Y_3, \tag{B-9}$$

Therefore, at the beginning of the detwinning deformation one has $\sigma = \sigma_s$, $c_2 = 0$, and the last equation, together with (36) implies:

$$Y_3 = \sigma_s H.$$

Similarly, upon completion of the deformation, one has $\sigma = \sigma_f$, $f_3(1) = 1$ and $\Phi_3 = 0$, hence:

$$\Delta_3 = H(\sigma_f - \sigma_s).$$

The function $f_3$ itself is curve-fitted from a stress-strain relationship obtained in a standard isothermal loading test at some fixed temperature below $M_f$.

The material parameters $Y_1^{\pm}$, $\Delta_1^{\pm}$ for the $A \leftrightarrow M^t$ are determined with the help of a zero stress cooling/heating cycle. During cooling, the forward transformation surface (B-3) turns into equality:

$$\hat{\pi}(\sigma, T) - \Delta_1^+ f_1^+(c_1) = Y_1^+, \tag{B-10}$$

which, in conjunction with (33) yields:

$$\begin{aligned} Y_1^+ &= \hat{\pi}(0, M_s) = \\ &\quad - \rho\Delta u_0 + \rho\left(M_s\Delta s_0 - \Delta c\left[(M_s - T_0) - M_s \ln\left(\frac{M_s}{T_0}\right)\right]\right), \qquad \text{(B-11)} \\ \Delta_1^+ &= \hat{\pi}(0, M_f) - Y_1^+ = \hat{\pi}(0, M_f) - \hat{\pi}(0, M_s) \\ &= \rho\left((M_f - M_s)\Delta s_0 - \Delta c\left[M_f - M_s + M_s \ln\left(\frac{M_s}{T_0}\right) - M_f \ln\left(\frac{M_f}{T_0}\right)\right]\right). \qquad \text{(B-12)} \end{aligned}$$

Similarly, during the heating (B-4) becomes:

$$-\hat{\pi}(\sigma, T) + \Delta_1^- f_1^-(c_1) = Y_1^-, \tag{B-13}$$

hence $Y_1^-$ and $\Delta_1^-$ can be determined:

$$\begin{aligned} Y_1^- &= -\hat{\pi}(0, A_f^t) \\ &= \rho\Delta u_0 - \rho\left(A_f^t\Delta s_0 - \Delta c\left[(A_f^t - T_0) - A_f^t \ln\left(\frac{A_f^t}{T_0}\right)\right]\right), \qquad \text{(B-14)} \\ \Delta_1^- &= \hat{\pi}(0, A_s^t) + Y_1^- = \hat{\pi}(0, A_s^t) - \hat{\pi}(0, A_f^t) \\ &= \rho\left((A_s^t - A_f^t)\Delta s_0 - \Delta c\left[A_s^t - A_f^t + A_f^t \ln\left(\frac{A_f^t}{T_0}\right) - A_s^t \ln\left(\frac{A_s^t}{T_0}\right)\right]\right). \qquad \text{(B-15)} \end{aligned}$$

Determining the parameters for the stress-induced martensitic transformation is done by considering two loading paths. First, assume a fully detwinned state at some temperature below $A_s^d$ and at zero stress (this can be obtained by loading isothermally at $T \leq M_f$ until

all the material has detwinned and then unloading until zero stress us reached) and heat, while maintaining the material stress free. Then $\dot{\xi}_2 < 0$ and (B-6) becomes an equality:

$$-\sigma H - \hat{\pi}(\sigma, T) + \Delta_2^- f_2^-(c_2) = Y_2^-. \tag{B-16}$$

Noting that $\sigma = 0$ throughout the loading path, and with the help of (35), $Y_2^-$ and $\Delta_2^-$ are found to be:

$$\begin{aligned} Y_2^- &= -\hat{\pi}(0, A_f^d) \\ &= \rho\Delta u_0 - \rho\left(A_f^d \Delta s_0 - \Delta c\left[(A_f^d - T_0) - A_f^d \ln\left(\frac{A_f^d}{T_0}\right)\right]\right), \qquad \text{(B-17)} \\ \Delta_2^- &= \hat{\pi}(0, A_s^d) + Y_2^- = \hat{\pi}(0, A_s^d) - \hat{\pi}(0, A_f^d) \\ &= \rho\left((A_s^d - A_f^d)\Delta s_0 - \Delta c\left[A_s^d - A_f^d + A_f^d \ln\left(\frac{A_f^d}{T_0}\right) - A_s^d \ln\left(\frac{A_s^d}{T_0}\right)\right]\right). \qquad \text{(B-18)} \end{aligned}$$

Finally, in order to determine $Y_2^+$ and $\Delta_2^+$, load the material in austenite to some stress level, for example, $\sigma_f$ and then cool the material. Let the critical temperatures for the $A \to M^d$ transformation at this stress level be $T_s(\sigma_f)$ for the start and $T_f(\sigma_f)$ for the finish. Then the constraint (B-5) becomes:

$$\sigma H + \hat{\pi}(\sigma, T) - \Delta_2^+ f_2^+(c_2) = Y_2^+ \tag{B-19}$$

which results in

$$Y_2^+ = \sigma_f H + \hat{\pi}(\sigma_f, T_s(\sigma_f)), \tag{B-20}$$

$$\Delta_2^+ = \sigma_f H + \hat{\pi}(\sigma_f, T_f(\sigma_f)) - Y_2^+. \tag{B-21}$$

Note that it is necessary to load to a stress equal or higher then $\sigma_f$, in order to avoid development of twinned martensite.

With this last equation, all the material parameters, except for the functional form of $f_i^\pm$ are expressed from physically observable quantities. While in this model only linear functions are considered, the model allows for arbitrary monotonous functions which can be curve-fitted from experiments. The curve-fit for $f_2^\pm$ can be done from a single uniaxial loading (Lagoudas et al., 1996, c.f., e.g.) and a curve-fit for $f_1^\pm(c_1)$ can be obtained by using a DSC measurement in conjunction with the balance of energy. These issues, however, will not be discussed further.

## References


Anand, L., Gurtin, M., 2003. Thermal effects in the superelasticity of crystalline shape-memory materials. Journal of the Mechanics and Physics of Solids 51 (6), 1015–1058.

Auricchio, F., Taylor, R. L., 1997. Shape-memory alloys: Modelling and numerical simulations of the finite-strain superelastic behavior. Computer Methods in Applied Mechanics and Engineering 143, 175–194.

Auricchio, F., Taylor, R. L., Lubliner, J., 1997. Shape-memory alloys: Macromodelling and numerical simulations of the superelastic behavior. Computer Methods in Applied Mechanics and Engineering 146, 281–312.

Bekker, A., Brinson, L. C., 1997. Temperature-induced phase transformation in a shape memory alloy: Phase diagram based kinetics approach. J. Mech. Phys. Solids 45 (6), 949–988.

Bekker, A., Brinson, L. C., 1998. Phase diagram based description of the hysteresis behavior of shape memory alloys. Acta Materialia 46 (10), 3649–3665.

Berveiller, M., Patoor, E., Buisson, M., 1991. Thermomechanical constitutive equations for shape memory alloys. Journal de Physique IV 1, C.4,387, European Symposium on Martensitic Transformation and Shape Memory Properties.

Bo, Z., Lagoudas, D. C., 1999a. Thermomechanical modeling of polycrystalline SMAs under cyclic loading, Part I: Theoretical derivations. Int. J. Eng. Sci. 37, 1089–1140.

Bo, Z., Lagoudas, D. C., 1999b. Thermomechanical modeling of polycrystalline SMAs under cyclic loading, Part III: Evolution of plastic strains and two-way memory effect. Int. J. Eng. Sci. 37, 1175–1204.

Bo, Z., Lagoudas, D. C., 1999c. Thermomechanical modeling of polycrystalline SMAs under cyclic loading, Part IV: Modeling of minor hysteresis loops. Int. J. Eng. Sci. 37, 1205–1249.

Boyd, J. G., Lagoudas, D. C., 1993. A thermodynamically based costitutive model for the SME due to transformation and reorientation. In: Proceedings of the 4th International Symposium on Plasticity and its Current Applications.

Boyd, J. G., Lagoudas, D. C., 1994. Thermomechanical response of shape memory composites. J. Intell. Mater. Systems Struct. 5, 333–346.

Boyd, J. G., Lagoudas, D. C., 1996a. A thermodynamic constitutive model for the shape memory materials. Part I. The monolithic shape memory alloys. Int. J. Plasticity 12 (6), 805–842.

Boyd, J. G., Lagoudas, D. C., 1996b. A thermodynamical constitutive model for shape memory materials. Part II. The SMA composite material. Int. J. Plasticity 12 (7), 843–873.

Brinson, L. C., 1993. One-dimensional constitutive behavior of shape memory alloys: Thermomechanical derivation with non-constant material functions and redefined martensite internal variable. J. of Intell. Mater. Syst. and Struct. 4, 229–242.

Brinson, L. C., Lammering, R., 1993. Finite element analysis of the behavior of shape memory alloys and their applications. Int. J. Solids Struct. 30 (23), 3261–3280.

Coleman, B., Gurtin, M., 1967. Thermodynamics with internal state variables. The Journal of Chemical Physics 47, 597–613.

Cross, W. B., Kariotis, A. H., Stimler, F. J., 1969. Nitinol characterization study. Tech. Rep. CR-1433, NASA.

Gillet, Y., Patoor, E., Berveiller, M., 1998. Calculation of pseudoelastic elements using a non symmetrical thermomechanical transformation criterion and associated rule. Journal of Intelligent Materials and Technology 9, 366–378.

Graesser, E. J., Cozzarelli, F. A., 1991. Shape memory alloys as new materials for aseismic isolation. Journal of Engineering Mechanics 117 (11), 2590–2608.

Huang, W., 1999. Yield surfaces of shape memory alloys and their applications. Acta Materialia 47 (9), 27692776.

Jackson, C. M., Wagner, H. J., Wasilewski, R. J., 1972. 55-nitinol—The alloy with a memory: Its physical metallurgy, properties and applications. Tech. Rep. NASA SP-5110, NASA, Technology Utilization Office, Washington, D.C.

Juhasz, L., Schnack, E., Hesebeck, O., Andra, H., 2002. Macroscopic modeling of shape memory alloys under non-proportional thermo-mechanical loadings. J. Intell. Mater. Systems

Struct. 13, 825–836.
Jun, H. Y., Rediniotis, O. K., Lagoudas, D. C., 2006. Development of a fuel-powered shape memory alloy actuator system Submitted.
Jung, Y., Papadopoulos, P., Ritchie, R., 2004. Constitutive modelling and numerical simulation of multivariant phase transformation in superelastic shape-memory alloys. International Journal for Numerical Methods in Engineering 60, 429–460.
Kohl, M., Krevet, B., Just, E., 2002. Sma microgripper system. Sensors and Actuators A97-98, 646–652.
Kotil, T., Sehitoglu, H., Maier, H., Chumlyakov, Y., 2003. Transformation and detwinning induced electrical resistance variations in NiTiCu. Materials and Engineering A359, 280–289.
Lagoudas, D., Shu, S., 1999. Residual deformations of active structures with SMA actuators. International Journal of Mechanical Sciences 41, 595619.
Lagoudas, D. C., Bo, Z., 1999. Thermomechanical modeling of polycrystalline SMAs under cyclic loading, Part II: Material characterization and experimental results for a stable transformation cycle. Int. J. Eng. Sci. 37, 1205–1249.
Lagoudas, D. C., Bo, Z., Qidwai, M. A., 1996. A unified thermodynamic constitutive model for SMA and finite element analysis of active metal matrix composites. Mech. Composite Mater. Struct. 3, 153–179.
Lagoudas, D. C., Entchev, P. B., 2004. Modeling of transformation-induced plasticity and its effect on the behavior of porous shape memory alloys. Part I: Constitutive model for fully dense SMAs. Mech. Mater. 36 (9), 865–892.
Leclercq, S., Lexcellent, C., 1996. A general macroscopic description of the thermomechanical behavior of shape memory alloys. J. Mech. Phys. Solids 44 (6), 953–980.
Lexcellent, C., Leclerq, S., Gabry, B., Bourbon, G., 2000. The two way shape memory effect of shape memory alloys: an experimental study and a phenomenological model. Int. J. Plasticity 16, 1155–1168.
Lexcellent, C., Vivet, A., Bouvet, C., Calloch, S., Blanc, P., 2002. Experimental and numerical determinations of the initial surface of phase transformations under biaxial loading in some polycrystalline shape-memory alloys. J. Mech. Phys. Solids (to appear).
Liang, C., Rogers, C. A., 1990. One-dimensional thermomechanical constitutive relations for shape memory materials. Journal of Intelligent Material Systems and Structures 1, 207–234.
Liang, C., Rogers, C. A., 1992. A multi-dimensional constitutive model for shape memory alloys. Journal of Engineering Mathematics 26, 429–443.
Malvern, L. E., 1969. Introduction to the Mechanics of a Continuous Medium. Prentice-Hall, Inc., Englewood Cliffs, NJ.
Marketz, F., Fischer, F., Tanaka, K., 1995. A computational micromechanics study on variant-coalescence in a Cu-Al-Ni shape memory alloy. Journal de Physique IV, Colloque C2 5, 537–542.
Miller, D., 2000. Thermomechanical characterization of plastic deformation and transformation fatigue in shape memory alloys. Ph.D. thesis, Texas A&M University.
Nishiyama, Z., 1978. Martensitic Transformations. Academic Press, San Diego.
Olson, G. B., Cohen, M., 1982. Stress assisted isothermal martensitic transformation: Application to TRIP steels. Metall. Trans. A 13A, 1907–1914.
Ortin, J., Planes, A., 1988. Thermodynamic analysis of thermal measurements in thermoelastic martensitic transformations. Acta Metallurgica 36 (8), 1873–1889.

Ortin, J., Planes, A., 1989. Thermodynamics of thermoelastic martensitic transformations. Acta Metall. 37 (5), 1433–1441.

Ortiz, M., Popov, E. P., 1985. Accuracy and stability of integration algorithms for elastoplastic constitutive relations. International Journal for Numerical Methods in Engineering 21, 1561–1576.

Otsuka, K., Ren, X., 2005. Physical metallurgy of Ti-Ni-based shape memory alloys. Progress in Materials Science 50, 511–678.

Otsuka, K., Wayman, C. M., 1999. Shape Memory Materials. Cambridge University Press, Cambridge, Ch. Introduction, pp. 1–26.

Patoor, E., Eberhard, A., Berveiller, M., 1988. Thermomechanical behaviour of shape memory alloys. Arch. of Mech. 40 (5-6), 775–794.

Popov, P., 2005. Constitutive modelling of shape memory alloys and upscaling of deformable porous media. Ph.D. thesis, Texas A&M University.

Qidwai, M. A., Lagoudas, D. C., 2000a. Numerical implementation of a shape memory alloy thermomechanical constitutive model using return mapping algorithms. Int. J. Numer. Meth. Eng. 47, 1123–1168.

Qidwai, M. A., Lagoudas, D. C., 2000b. On thermomechanics and transformation surfaces of polycrystalline NiTi shape memory alloy material. Int. J. Plasticity 16, 1309–1343.

Raniecki, B., Lexcellent, C., 1994. RL-models of pseudoelasticity and their specification for some shape memory solids. Eur. J. Mech. A/Solids 13 (1), 21–50.

Raniecki, B., Lexcellent, C., 1998. Thermodynamics of isotropic pseudoelasticity in shape memory alloys. Eur. J. Mech. A/Solids 17 (2), 185–205.

Sakamoto, H., 2002. Distinction between thermal and stress-induced martensitic transformations and inhomogeneity in internal stress. Materials Transactions 43 (9), 2249–2255.

Sato, Y., Tanaka, K., 1988. Estimation of energy dissipation in alloys due to stress-induced martensitic transformation. Res Mechanica 23, 381–392.

Simo, J. C., Hughes, T. J. R., 1987. General return mapping algorithms for rate-independent plasticity. In: Desai, C. S. (Ed.), Constitutive Laws for Engineering Materials: Theory and Applications. Elsevier Science Publishing Co., Inc., pp. 221–231.

Simo, J. C., Hughes, T. J. R., 1998. Computational Inelasticity. Springer-Verlag New York Inc.

Sun, Q. P., Hwang, K. C., 1993a. Micromechanics modeling for the constitutive behavior of polycrystalline shape memory alloys — I. Derivation of general relations. J. Mech. Phys. Solids 41 (1), 1–17.

Sun, Q. P., Hwang, K. C., 1993b. Micromechanics modelling for the constitutive behavior of polycrystalline shape memory alloys — II. Study of the individual phenomena. J. Mech. Phys. Solids 41 (1), 19–33.

Sun, Q. P., Hwang, K. C., Yu, S. W., 1991. A micromechanics constitutive model of transformation plasticity with shear and dilatation effect. J. Mech. Phys. Solids 9 (4), 507–524.

Tanaka, K., 1986. A thermomechanical sketch of shape memory effect: One-dimensional tensile behavior. Res Mechanica 18, 251–263.

Tanaka, K., Hayashi, T., Itoh, Y., Tobushi, H., 1992. Analysis of thermomechanical behavior of shape memory alloys. Mechanics of Materials 13, 207–215.

Tanaka, K., Kobayashi, S., Sato, Y., 1986. Thermomechanics of transformation pseudoelasticity and shape memory effect in alloys. Int. J. Plasticity 2, 59–72.

Tanaka, K., Nishimura, F., Hayashi, T., Tobushi, H., Lexcellent, C., 1995. Phenomenological analysis on subloops and cyclic behavior in shape memory alloys under mechanical and/or

thermal loads. Mechanics of Materials 19, 281–292.
Truesdell, C., Noll, W., 1965. The Non-Linear Field Theories of Mechanics. Springer-Verlag, Berlin.
Tsoi, K., Stlmans, R., Schrooten, J., 2003. Transformational behavior of constrained shape memory alloys. Acta Materialia 50, 3535–3544.
Šittner, P., Lukáš, P., Neov, D., Lugovyy, D., 2003. Martensitic tranformations in NiTi polycryctals investigated by in-situ neutron diffraction. Materials Science Forum 426-432, 2315–2320.
Šittner, P., Vokoun, D., Dayananda, G., Stalmans, R., 2000. Recovery stress generation in shape memory Ti50Ni45Cu5 thin wires. Material Science and Engineering A 286, 298–311.
Wayman, C. M., 1983. Phase transformations, nondiffusive. In: Cahn, R. W., Haasen, P. (Eds.), Physical Metallurgy. North-Holland Physics Publishing, New York, pp. 1031–1075.
Wollants, P., De Bonte, M., Roos, J., 1979. A thermodynamic analysis of the stress-induced martensitic transformation in a single crystal. Z. Metallkd. 70, 113–117.