One-Dimensional Constitutive Behavior of Shape Memory Alloys: Thermomechanical Derivation with Non-Constant Material Functions and Redefined Martensite Internal Variable

L. C. BRINSON*
Institut für Aeroelastik der DLR
Bunsenstraße 10
3400 Göttingen
Germany

ABSTRACT: A one-dimensional constitutive model for the thermomechanical behavior of shape memory alloys is developed based on previous work by Liang and Tanaka. An internal variable approach is used to derive a comprehensive constitutive law for shape memory alloy materials from first principles without the assumption of constant material functions. This constitutive law is of such a form that it is well suited to further practical engineering applications and calculations.

A separation of the martensite fraction internal variable into temperature-induced and stress-induced parts is presented and justified, which then allows the derived constitutive law to accurately represent both the pseudoelastic and shape memory effects at all temperatures. Several numerical examples are given that illustrate the ability of the constitutive law to capture the unique thermomechanical behavior of shape memory alloys due to their internal phase transformations with stress and temperature.

INTRODUCTION

Shape memory alloy (SMA) materials have been receiving increasingly more attention and study since the discovery and first publication of the shape memory effect by Chang and Read in 1951. These materials have been shown to exhibit extremely large, recoverable strains (on the order of 10%), and it is these properties as functions of temperature and stress which allow SMAs to be utilized in many exciting and innovative engineering applications. At present, SMA materials are being investigated for diverse applications: the driving force in heat cycle engines (Banks and Weres, 1976); the material for implantable Harrington rods for correction of scoliosis; orthodontic wire capable of exerting constant load even after large strains, making frequent readjustments unnecessary (Funakubo, 1987); and integrated actuator/sensor fibers in special composite systems for active control of dynamic and structural behavior (Rogers, Liang et al., 1989).

For all of the various applications, it is necessary to have a precise understanding of the mechanical behavior of SMAs in order to fully develop and exploit their potential. The aforementioned applications of SMAs in intelligent material systems all utilize the SMA material integrated into the structure (or, in some cases, as the entire structure), such that the SMAs provide control for the system and are force bearing members. This high degree of integration of control materials with the structure is a desired goal and approaches the highest level of intelligent systems (Wada, Fanson et al., 1990). However, it simultaneously emphasizes the need for comprehensive material models of shape memory materials that are capable of both accurately representing the thermomechanical behavior and having mathematical expression in a form that is amenable to incorporation into other engineering tools, such as finite element procedures or control analysis programs.

Many scientists and engineers have contributed to the vast available literature on the experimental behavior of shape memory alloys. Likewise, there are several approaches that have been developed to address the constitutive behavior of these materials (Tanaka, Osaka et al., 1982; Falk, 1983; Achenbach and Müller, 1985; McNichols and Cory, 1987; Achenbach, 1989; Liang, 1990). This article will focus, however, on the work presented separately by Tanaka and Liang, which is a unified constitutive law meeting the requirements laid out above (Tanaka, Osaka et al., 1982; Tanaka and Iwasaki, 1985; Tanaka, 1986; Liang, 1990; Liang and Rogers, 1990). In the following sections, background material and a brief summary of the thermomechanical derivation will be presented. A refined definition for the martensite material fraction will be introduced and the consequences discussed. The constitutive law derivation for a special case of non-constant material functions will be presented and finally, several brief numerical examples using the newly derived constitutive law and redefined martensite term will be examined.
MECHANICAL BEHAVIOR OF SHAPE MEMORY ALLOYS

Background

From a macroscopic point of view, one can separate the observable mechanical behavior of SMAs into two major categories: the shape memory effect (SME), in which a specimen exhibits a large residual (apparently plastic) strain after loading and unloading that can be fully recovered upon raising the temperature of the material; and the pseudoelastic effect, in which a specimen achieves a very large (apparently plastic) strain upon loading that is then fully recovered in a hysteresis loop upon unloading.

This surprising ability to fully recover large strains (on the order of 10%) is a result of a martensite phase transformation occurring in the SMA material according to certain environmental and boundary conditions. In a stress-free state, an SMA material at high temperatures exists in the parent phase (usually a body-centered cubic crystal structure, also referred to interchangeably in this paper as the austenite phase) and upon decreasing the material temperature, the crystal structure undergoes a self-accommodating crystal transformation into martensite (usually a face-centered cubic structure). The phase change in the unstressed formation of martensite from austenite is referred to as “self-accommodating” due to the formation of multiple martensitic variants and twins which prohibits the occurrence of a transformation strain. The martensite variants, evenly distributed throughout the material, are all crystallographically equivalent, differing only by habit plane indices, and each variant consists of two twin-related martensites. It is exactly this effect of self-accommodation by twinning which subsequently allows shape memory alloys to exhibit the large reversible strains with stress (Funakubo, 1987; Wayman and Duerig, 1990).

In the stress-free state, an SMA material can be considered to have four transition temperatures, designated as $M_s$, $M_f$, $A_s$, $A_f$: Martensite Finish, Martensite Start, Austenite Start, and Austenite Finish, respectively. Although there are materials in which $A_s < M_f$ and the basic concepts laid out in this treatise generally apply to this case, for clarity of explanation only the so-called “Type I” materials in which $M_f < M_s < A_s < A_f$ will be considered. Note that a change of temperature within the range $M_s < T < A_s$ induces no phase changes and both martensite and austenite can coexist within $M_s < T < A_s$.

Given these four transformation temperatures and the concepts of self-accommodation, one manifestation of the shape memory effect (SME) can be explained as follows. Consider martensite, formed from the parent phase cooled under stress-free conditions through $M_f$ and $M_s$ at any temperature less than Austenite Start. This material has multiple variants and twins present, all crystallographically equivalent, but with differing orientation (different habit plane indices). When a load applied to this material reaches a certain critical stress, the pairs of martensite twins begin “detwinning” (conversion) to the stress-preferred twins. Subsequently, the multiple martensite variants begin to convert to a single variant, the preferred variant determined by alignment of the habit planes with the axis of loading. The conversion between different martensite variants can also be described to be a twinning deformation process in the materials science definition of the term (Funakubo, 1987; Wayman and Duerig, 1990). During this process of reorientation, the stress raises very slightly in comparison to the strain that is achieved, creating a stress-strain curve as shown schematically in Figure 1(a). As the single variant of martensite is thermodynamically stable at $T < A_s$, upon unloading there is no reconversion to multiple variants and only a small elastic strain is recovered, leaving the material with a large residual strain. The detwinned martensite material can, however, recover the entire residual strain by simply heating above the $A_s$ temperature: the material then transforms to the parent phase (which has no variants) and recovers its original geometric configuration, thus creating the shape memory effect [indicated by the arrow in Figure 1(a)]. See also Figure 2 for an illustrative sketch of the SME.

To explain the pseudoelastic effect, consider the SMA en-

![Figure 1](image.png)

**Figure 1.** Schematic of stress-strain curves of shape memory alloy mechanical behavior. The nonlinear loading portion of the curves represents detwinning of martensite variants or transformation of austenite to martensite. In (c) and (d) the nonlinear portion upon unloading is due to the inverse transformation to austenite. In (e) the maximum stress has exceeded the critical stress for plastic slip and thus there remains irreversible plastic strain upon unloading.
Figure 1 (continued). Schematic of stress-strain curves of shape memory alloy mechanical behavior. The nonlinear loading portion of the curves represents detwinning of martensite variants or transformation of austenite to martensite. In (c) and (d) the nonlinear portion upon unloading is due to the inverse transformation to austenite. In (e) the maximum stress has exceeded the critical stress for plastic slip and thus there remains irreversible plastic strain upon unloading.

tirely in the parent phase (with $T > M_f$). When stress is applied to this material, thermodynamic considerations indicate that there is a critical stress at which the crystal phase transformation from austenite to martensite can be induced. However, due to the presence of stress during the transformation, specific martensite variants will be formed preferentially and at the end of transformation, the stress-induced martensite will consist ideally of a single variant of completely detwinned martensite. If $T \geq A_f$, during unloading of the material a reverse transformation to austenite occurs because of the instability of martensite at these temperatures in the absence of stress. This recovery of high strain values upon unloading yields a characteristic hysteresis loop, which is illustrated in Figure 1(d), and is known as pseudoelasticity (Delaey, Krishnan, et al., 1974; Perkins, Edward, et al., 1976; Funakubo, 1987). If $A_r < T < A_f$, the result upon unloading is a partial pseudoelastic recovery; the remaining residual strain can be fully recovered after heating the material above $A_r$ [see Figure 1(c)]. If $T < A_r$, then there is no pseudoelastic recovery and the result is a different manifestation of the shape memory effect [Figure 1(b)].

There exists at any temperature a critical stress for irreversible plastic slip to occur in the material (this critical stress value decreasing with increasing temperature), and if this stress is exceeded, then the residual strain can no longer, of course, be recovered by unloading or by heating [see Figure 1(e)]. Note that in some SMA materials, it is also possible for the material to undergo transformation from one martensite to another crystallographically different form of martensite with stress. These types of crystal transformations between martensites are not considered here.

Development of the Constitutive Law from Thermodynamics

Here we review Tanaka's approach to the derivation of the constitutive law of shape memory materials. Considering a one-dimensional SMA material undergoing transformation, from principles of thermodynamics the energy balance and Clausius-Duhem inequality can be expressed as

\[
\dot{q}U - \dot{\sigma}L + \frac{\partial q_{\text{int}}}{\partial \dot{\sigma}} - \dot{\sigma}q = 0 \quad (1a)
\]

\[
\dot{q}S - \frac{q}{T} + \frac{\partial q_{\text{int}}}{\partial \dot{\sigma}} \left( \frac{T}{q_{\text{int}}} \right) \geq 0 \quad (1b)
\]

in the current configuration (Tanaka, 1986), where $U$, $\dot{\sigma}$, $q$, and $q_{\text{int}}$ represent the internal energy density, the Cauchy
stress, the heat production term and the heat flux, respectively, and \( S, T, x \) and \( \theta \) represent the entropy density, temperature, the material coordinate and the density in the current configuration, respectively. Following Tanaka’s argument, it is assumed that the thermo-mechanics of an SMA material are fully described by the set of variables \((\epsilon, T, \xi)\), where \( \epsilon \) is the Green strain and \( \xi \) is an internal variable representing the stage of the transformation. The definition of \( \xi \) is the martensite fraction of the material, which varies from zero to one with unity representing 100% martensite, and its value is governed by temperature and stress, this relationship given in the transformation kinetics section.

By introducing the Helmholtz free energy \( \Phi = U - TS \), inequality [Equation (1b)] can be rewritten in the reference configuration as

\[
\rho \frac{\delta \Phi}{\delta \epsilon} \dot{\epsilon} - \left( S + \frac{\partial \Phi}{\partial T} \right) \dot{T} - \frac{\partial \Phi}{\partial \xi} \dot{\xi} - \frac{1}{\rho C_v} \frac{\rho C_v}{\rho} \frac{\partial T}{\partial X} \geq 0
\]  

(2)

where \( \rho \) is the second Piola-Kirchhoff stress, \( F \) is the deformation gradient, and \( \rho \) the density and \( X \) the material coordinate in the reference configuration.

A sufficient condition for Equation (2) to hold for every choice of \( \dot{\epsilon}, \dot{T}, \dot{\xi} \), their respective coefficients must vanish, thus yielding

\[
\sigma = \rho \frac{\partial \Phi(\epsilon, \xi, T)}{\partial \epsilon} = \sigma(\epsilon, \xi, T) \tag{3a}
\]

\[
S = -\frac{\partial \Phi}{\partial T} \tag{3b}
\]

Equation (3a) is then the mechanical constitutive equation of the material.

**Constitutive Law with Constant Material Functions**

For later reference, we introduce here Liang’s constitutive model of shape memory alloys derived from Equation (3a).

By differential calculus, one can write Equation (3a) as

\[
d\sigma = \frac{\partial \sigma}{\partial \epsilon} d\epsilon + \frac{\partial \sigma}{\partial \xi} d\xi + \frac{\partial \sigma}{\partial T} dT \tag{4}
\]

leading to the most general equivalent expression

\[
d\sigma = D(\epsilon, \xi, T) d\epsilon + \Omega(\epsilon, \xi, T) d\xi + \Theta(\epsilon, \xi, T) dT \tag{5}
\]

where the material functions are defined by

\[
D(\epsilon, \xi, T) = \frac{\partial^2 \Phi}{\partial \epsilon^2}, \quad \Omega(\epsilon, \xi, T) = \rho \frac{\partial^2 \Phi}{\partial \epsilon \partial \xi}, \quad \Theta(\epsilon, \xi, T) = \rho \frac{\partial^2 \Phi}{\partial \epsilon \partial T} \tag{6}
\]

From the form of the incremental constitutive law [Equation (5)], the function \( D(\epsilon, \xi, T) \) is representative of the modulus of the SMA material, \( \Omega(\epsilon, \xi, T) \) can be considered the “transformation tensor”, and \( \Theta(\epsilon, \xi, T) \) is related to the thermal coefficient of expansion for the SMA material. If these material functions are all assumed to be constants, then the constitutive relation can be easily derived as

\[
\sigma - \sigma_0 = D(\epsilon - \epsilon_0) + \Omega(\xi - \xi_0) + \Theta(T - T_0) \tag{7}
\]

where \((\sigma_0, \epsilon_0, \xi_0, T_0)\) represent the initial state (or initial conditions) of the material.

It is important to note here that the application of a specific material restriction enforces a relationship between the Young’s modulus and the transformation tensor of SMAs. The maximum residual strain of an SMA is a material constant, \( \epsilon_L \). Considering Figure 1(b), \( \epsilon_L \) can be achieved by converting all of the original austenite to completely detwinned martensite, thus going from a state of \( \xi = 0 \) to \( \xi = 1 \). Upon unloading, the maximum residual strain remains until the temperature is increased above \( A_1 \). Using initial conditions of \((\sigma_0 = \epsilon_0 = \xi_0 = 0) \) and final conditions of \((\sigma = 0, \epsilon = \epsilon_L, \xi = 1) \) with \( T = T_0 (M < T < A_1) \) in Equation (7), one obtains the necessary relationship

\[
\Omega = -\epsilon_L D \tag{8}
\]

**Transformation Kinetics**

The phase transformation between austenite and martensite as a function of temperature and stress is governed by chemical free energy as the driving force. A thorough development of the theory of transformation kinetics is contained in work by Funakubo (1987) and Warlimont, Delaey, et al. (1974); a brief description is offered here. A form of the Clausius-Clapeyron equation

\[
\frac{d\sigma}{dT} = -\frac{\Delta H}{T_0 \epsilon} \tag{9}
\]

can be derived from basic thermodynamic principles to describe the relationship between temperature and transformation stress for a single crystal of an SMA material, where \( \epsilon \) is the transformation strain and \( \Delta H \) represents the change in enthalpy between the martensite and austenite phases at the temperature \( T_0 \), at which both phases are in equilibrium under the stress \( \sigma \). Theoretical studies examining phase equilibrium and transformation kinetics have also developed expressions for the martensite fraction as a function of free energy and temperature. Liang and Rogers (1990) have developed an empirically based cosine model to represent the martensite fraction as a function of stress and temperature during transformation, which agrees well with experimental findings. Because this last model contains only measurable engineering variables, it can be practically implemented and is used as a basis in this study.
According to Liang's model, the transformation from the parent phase to martensite can be described by

\[
\xi = \frac{1 - \xi_0}{2} \cos \left[ a_M \left( T - M_f - \frac{\sigma}{C_M} \right) \right] + \frac{1 + \xi_0}{2}
\]

for \( C_M(T - M_s) < \sigma < C_M(T - M_f) \) \( (10) \)

while the reverse transformation from martensite to austenite can be expressed as

\[
\xi = \frac{\xi_0}{2} \cos \left[ a_A \left( T - A_s - \frac{\sigma}{C_A} \right) \right] + \frac{1}{2}
\]

for \( C_A(T - A_f) < \sigma < C_A(T - A_s) \) \( (11) \)

where \( \xi_0 \) is the fraction of the material which is martensite prior to the current transformation, \( \sigma \) is the applied stress, \( T \) the temperature, and \( a_M \) and \( a_A \) are defined by

\[
a_M = \frac{\pi}{M_s - M_f}, \quad a_A = \frac{\pi}{A_s - A_f} \quad (12)
\]

The constants \( C_M \) and \( C_A \) are material properties that describe the relationship of temperature and the critical stress to induce transformation, \( \sigma_{cr} \) (see Figure 3). Although experimental results for the critical stress values at the onset and end of transformations rarely yield precisely linear results, it is nonetheless possible to adequately represent the trend for the transformation stresses as linear functions of temperature. For example, at temperatures above \( M_s \), the stress necessary to induce martensite from austenite increases with temperature and one can define a constant value for \( C_M \). It is often assumed that \( C_M \) and \( C_A \) equal one another and have a continuously constant value over all temperature ranges, as illustrated in Figure 3.

**A SEPARATION OF THE MARTENSITE FRACTION INTERNAL VARIABLE AND CONSEQUENCES**

Until this point, the internal state variable for the martensite material fraction, \( \xi \), has been considered to simply represent the percentage of the material transformed to martensite. A separation of this variable into two parts is proposed here based on the micromechanics of an SMA material. Let \( \xi \) be further defined by

\[
\xi = \xi_S + \xi_T
\]

where \( \xi_T \) represents the fraction of the material that is purely temperature-induced martensite with multiple variants, and \( \xi_S \) denotes the fraction of the material that has been transformed by stress into a single martensitic variant. With such a definition we will show that it is possible to derive a constitutive relation similar to Equation (7) that is valid for temperatures below \( M_s \) as well as above. That is, the rederived constitutive equation can represent the shape memory effect starting from martensite as well as pseudoelasticity and the SME starting from 100% austenite.

To quickly illustrate the shortcomings of not distinguishing between \( \xi_T \) and \( \xi_S \), consider the application of constitutive Equation (7) in the case of \( T < M_f \). At such temperatures, the material is entirely in a martensitic phase. For simplicity, take the initial conditions to be \( \sigma_0 = \varepsilon_0 = 0 \), \( \xi_0 = 1 \) and hold the temperature constant (\( T = T_0 \)); since \( T < M_f \), \( \xi = 1 \) regardless of the stress and strain values and Equation (7) immediately yields for arbitrary \( \sigma \) and \( \varepsilon \)

\[
\sigma = D \varepsilon + \Omega \cdot (1 - 1) + \Theta(T - T_0)
\]

which is a purely linear elastic stress-strain relation and obviously cannot represent the behavior typical of the shape memory effect shown in Figure 1(a). In effect, Equation (7) cannot capture SMA material behavior at any temperature below Martensite Start, nor at higher temperatures when any temperature-induced martensite is present. In the case of the simple example just given, recall that at such a temperature, the shape memory effect is caused by conversion between martensite variants, not the transformation of austenite to martensite. With the martensite internal variable redefined to separate the converted (or detwinned) martensite fraction of the material from the temperature-induced, fully twinned martensite fraction, the rederivation of the constitutive law will contain these separate contributions and consequently be able to capture the SME of conversion of martensite twins as shown in the next subsection.

It is important to note that this separation of \( \xi \) into two distinct components accurately represents the microscopic behavior of the material. For the case of 100% austenite prior to loading, \( \xi_T = 0 \) and \( \xi = \xi_S \) represents the amount of material transformed by application of stress from austenite to martensite, since the martensite that forms will necessarily be of a single preferred variant. When the material is totally (or partially) martensitic before stress application, then \( \xi_T \) accounts for the purely temperature-induced mar-

![Figure 3. Critical stresses for transformation as functions of temperature. \( C_M \) and \( C_A \) are the slopes of the \( \sigma_{cr}(T) \) curves for the austenite to martensite transformation and the reverse transformation, respectively. Here, \( C_M = C_A \) (after Liang, 1990).](image-url)
tensite consisting of multiple variants and $\xi_\ell$ increases in value only if sufficient stress is applied to convert multiple variants into a single variant (or to transform any austenite present into a single martensite variant).

Another point of note is that the maximum residual strain, $\epsilon_\ell$, is found to be reasonably constant at all temperatures below $A_t$ (Perkins, Edwards, et al., 1976), including $T < M_s$. This will allow the transformation tensor, $\Theta$, to retain its current relationship to the modulus and the maximum residual strain at all temperatures. Note that with this redefinition of $\xi$, as with the previous definition, caution must be exercised in stating the value of $\xi_0$; if the material has been heated, stress-free, above $A_s$ and cooled, then $\xi_0 = 0$ and $\xi_\ell$ depends only on temperature; otherwise these values can be determined from the previous history of the material.

**Modification to the Constitutive Law with Constant Material Functions**

With the introduction of $\xi = \xi_\ell + \xi_\ell$ into the constitutive Equation (3a), it follows immediately from differential calculus that

$$d\sigma = \frac{\partial \sigma}{\partial \epsilon} d\epsilon + \frac{\partial \sigma}{\partial \xi_\ell} d\xi_\ell + \frac{\partial \sigma}{\partial \xi_\ell} d\xi_\ell + \frac{\partial \sigma}{\partial T} dT$$

which can be written

$$d\sigma = Dd\epsilon + \Omega_\ell d\xi_\ell + \Omega_{\ell} d\xi_\ell + \Theta dT$$

once again assuming the material functions, $D$, $\Omega_\ell$, $\Omega_{\ell}$, $\Theta$, to be constants. With the initial conditions of ($\sigma_0$, $\epsilon_0$, $\xi_\ell$, $\xi_\ell$, $T_0$), solving the differential form of the constitutive equation yields

$$\sigma - \sigma_0 = D(\epsilon - \epsilon_0) + \Omega_\ell (\xi_\ell - \xi_\ell) + \Theta (T - T_0)$$

Application of the material restriction of the case of maximum residual strain with the material initially 100% austenite, $\xi_\ell = 0$ and $\xi_\ell = 0$, and the remaining conditions ($\sigma_0 = \epsilon_0 = 0$, $\sigma = 0$, $\epsilon = \epsilon_\ell$, $\xi_\ell = 1$, $\xi_\ell = 0$) and $T = T_0$ ($M_s < T < A_t$) [as for Equation (8)] provides the relationship

$$\Omega_\ell = -\epsilon_\ell D$$

Considering the case of maximum residual strain with the material initially 100% undeformed martensite, $\xi_\ell = 0$ and $\xi_\ell = 1$, and remaining conditions identical to the previous case (except here the temperature constraint can be relaxed to $T < A_t$) results in the restriction that

$$\Omega_\ell = 0$$

Thus, the thermomechanical constitutive law with constant material functions consistent with the separation of the stress-induced and temperature-induced martensite fractions becomes

$$\sigma - \sigma_0 = D(\epsilon - \epsilon_0) + \Omega_\ell (\xi_\ell - \xi_\ell) + \Theta (T - T_0)$$

dropping the subscript on $\Omega$ so that again $\Omega = -\epsilon_\ell D$. Obviously, this equation is capable of capturing the shape memory effect at all temperatures and with any percentage of initial twinned martensite. Applying the same initial conditions as in the example for Equation (14), with the clarification that $\xi_\ell = 1$ but $\xi_\ell = 0$, yields

$$\sigma = D\epsilon + \Omega_\ell \xi_\ell$$

which can accommodate the necessary nonlinear stress-strain behavior.

**Transformation Equations and Transformation Stresses**

The transformation equations of Liang [Equations (10) and (11)], with critical stresses as defined in Figure 3, must now be modified to accommodate the definition of $\xi_\ell$ and $\xi_\ell$ and to allow for the shape memory effect at temperatures below $M_s$. According to Delaey, Krishnan, et al. (1974), the critical stress for conversion of martensite variants below $M_s$ is a constant and thus can be considered to be a material property along with the $C_m$ and $C_m$ parameters, which remain valid for $T > M_s$. The variation with temperature of the critical stresses for transformation consistent with the separation of $\xi$ into two components is shown schematically in Figure 4. Experimental results by Dyc (1990) and others show a slight increase in $\sigma_{crit}$ with decreasing temperature below $M_s$ (indicated by dotted lines). Such a variation could be easily incorporated into this model, however, for simplicity we currently assume the critical stress values below $M_s$ to be constant and denoted by $\sigma_{crit}^\ell$ and $\sigma_{crit}^\ell$ for the critical stresses at the start and finish of the conversion of the martensitic variants. It might also be possible to predict the value of $\sigma_{crit}$ for reorientation of martensite twins from theoretical work by Achenbach and Muller (1985), who have developed a model based on the potential energy necessary

![Figure 4](imageмещение.png)

*Figure 4. Critical stresses for transformation or martensite twin conversion as functions of temperature. Dotted lines indicate possibility for non-constant values of $\sigma_{crit}^\ell$ and $\sigma_{crit}^\ell$. If the elastic modulus of the martensite phase is taken to be a constant, which is the case throughout this paper.
to overcome the chemical energy barrier for conversion of twins.

The evolution equations for calculation of the martensite fractions according to temperature and stress, recalling that $\xi = \xi_s + \xi_r$, can now be represented in conjunction with Figure 4 as†:

**CONVERSION TO DETWINNED MARTENSITE**

for $T > M_i$ and $\sigma'_u + C_m(T - M_i) < \sigma < \sigma''_u + C_m(T - M_i)$‡‡:

$$\xi_s = \frac{1 - \xi_{so}}{2} \cos \left( \frac{\pi}{\sigma'_u - \sigma''_u} \right) \times \left[ \sigma - \sigma''_u - C_m(T - M_i) \right] + \frac{1 + \xi_{so}}{2} \quad (22a)$$

$$\xi_r = \frac{\xi_{ro} - \xi_{ro}}{1 - \xi_{ro}} (\xi_s - \xi_{so}) \quad (22b)$$

for $T < M_i$ and $\sigma'_u < \sigma < \sigma''_u$:

$$\xi_s = \frac{1 - \xi_{so}}{2} \cos \left( \frac{\pi}{\sigma'_u - \sigma''_u} (\sigma - \sigma''_u) \right) + \frac{1 + \xi_{so}}{2} \quad (22c)$$

$$\xi_r = \xi_{ro} - \frac{\xi_{ro}}{1 - \xi_{so}} (\xi_s - \xi_{so}) + \Delta r \quad (22d)$$

where, if $M_i < T < M_j$, and $T < T_0$,

$$\Delta r = \frac{1 - \xi_{ro}}{2} \left[ \cos \left( \frac{\sigma}{C_m(T - M_j)} \right) + 1 \right] \quad (22e)$$

else, $\Delta r = 0$

**CONVERSION TO AUSTENITE**

for $T > A_i$ and $C_a(T - A_i) < \sigma < C_a(T - A_i)$:

$$\xi = \frac{\xi_0}{2} \cos \left( \frac{\sigma}{C_a(T - A_i)} \right) + 1 \quad (23a)$$

$$\xi_s = \frac{\xi_{so} - \xi_{so}}{\xi_0} (\xi_s - \xi) \quad (23b)$$

†Note that in the cases of $a_s < C_m(T_0 - A_i)$ for transformation to austenite or $a_s > \sigma'_u$ or $a_s > \sigma''_u + C_m(T_0 - M_i)$ for transformation to martensite, the values of $\xi_s$ and $\xi_{so}$ used in Equations (22) and (23) must be the values $\xi_s$ and $\xi_{so}$ would have had at the beginning of the transformation region for consistency.

‡‡Note that if $\sigma'_u = 0$ and $\sigma''_u = C_m(M_i - M_r)$, then the stress range for conversion to detwinned martensite is $C_m(T - M_i) < \sigma < C_m(T - M_j)$ as before and the expression for $\xi_e$ reduces to

$$\xi_e = \frac{1 - \xi_{ro}}{2} \cos \left( \frac{\sigma}{C_m(T - M_j)} \right) + \frac{1 + \xi_{ro}}{2}$$

as expected. Such would be the case for materials in which the curve for the critical stress for the start of the transformation to martensite passes through $M_i$ at zero stress.

Note that these transformation equations [as with Equations (10) and (11)] are based upon an array of experimental results performed on a particular SMA material. Although these curves agree well with the behavior of a wide range of SMAs, it is possible that a different set of equations would be better suited to a different particular alloy. In addition, there may be certain physical processes in an SMA for which Equations (22) and (23) would need modification. For example, there is evidence that reloading an SMA after partial unloading leads to a slightly different path in the internal loop and the assumed hysteretic behavior than predicted by these equations. Thus, although the predictions of Equations (22) and (23) would be qualitatively correct in such partial hysteresis loops, if the precise path followed in such excursions is of importance to a particular application, then Equations (22) and (23) would need to be expanded by a model which accounts for memory of return points (points at which transformation is reversed). Recent work by Ortiz (1991) investigates the use of a Preisach model on stress induced SMA transformations which could be utilized in conjunction with the results in this paper.

Sufficient experimental evidence is not available to distinguish between the critical stresses for transformation of austenite to martensite and the critical stresses for conversion of martensite twins at a given temperature. At temperatures in the range $M_i < T < A_i$, it is always possible to have some austenite present, as well as some unconverted martensite twins. If a distinction exists between the transformation and conversion stresses, the transformation Equations (22) would have to be altered accordingly. Since the stress-strain curves of shape memory alloys do not generally exhibit two distinct transformation regions (excluding crystallographic martensite-martensite transformations), it is reasonable to assume that the transformation and conversion stresses are identical for the purposes of this paper.

**CONSTITUTIVE LAW WITH NON-CONSTANT MATERIAL FUNCTIONS**

The constitutive Equation (20) with constant material functions is quite trivial to derive from the incremental constitutive Equation (16). However, experimental evidence on the modulus of SMA materials indicates clearly that the Young's modulus, $D_s$, has a strong dependence on the martensite fraction of the material, $\xi$. A reasonable assumption for the modulus function of an SMA material, as suggested by Liang (1990) and Sato and Tanaka (1988) is

$$D(\varepsilon, \xi, T) = D(\xi) + \xi(D_m - D_0) \quad (24)$$

where $D_m$ is the modulus value for the SMA as 100% martensite and $D_0$ is the modulus value for the SMA as 100% austenite. The ratio of the magnitudes of $D_m$ to $D_0$ usually have a value of 3 or greater. Firm experimental evidence of the variation of the remaining material properties with the state variables is lacking. Thus, here the derivation of the
one obtains

\[ \sigma + K = D_{s} \epsilon + (D_{m} - D_{s}) \xi \epsilon + C(\xi) - \epsilon_{t} D_{s} \xi \]

\[ - \epsilon_{t} (D_{m} - D_{s}) \left[ \frac{\xi}{2} + \xi \epsilon \right] + \Theta T \] (29b)

where \( K \) is an arbitrary constant and \( C(\xi) \) is an arbitrary function of \( \xi \). Rearranging terms and recognizing the expansion

\[ D_{\xi_{o}} \xi = D_{s} \xi + (D_{m} - D_{s}) \xi \epsilon + (D_{m} - D_{s}) \xi \epsilon \]

(29c)

Equation (29b) simplifies to

\[ \sigma + K = D(\xi) \epsilon + \Omega(\xi) \xi \epsilon + \Theta T \]

\[ + \epsilon_{t} (D_{m} - D_{s}) \frac{\xi}{2} + C(\xi) \] (29d)

which is the general solution to the governing differential Equation (28). To obtain a particular solution, first apply initial conditions: Equation (29d) must hold at the initial state \((\sigma_{0}, e_{0}, \xi_{0}, T_{0})\). Thus, the unknown constant, \( K \), is determined

\[ K = D(\xi_{0}) e_{0} + \Omega(\xi_{0}) \xi_{0} + \Theta T_{0} \]

\[ + \epsilon_{t} (D_{m} - D_{s}) \frac{\xi_{0}}{2} + C(\xi_{0}) - \sigma_{0} \] (30)

and Equation (29d) then becomes

\[ \sigma - \sigma_{0} = D(\xi) \epsilon - D(\xi_{0}) e_{0} + \Omega(\xi) \xi \epsilon - \Omega(\xi_{0}) \xi_{0} \]

\[ + \Theta (T - T_{0}) + \epsilon_{t} (D_{m} - D_{s}) \left[ \frac{\xi}{2} - \frac{\xi_{0}}{2} \right] \]

\[ + C(\xi) - C(\xi_{0}) \] (31)

To determine the unknown function \( C(\xi) \), consider again the specific material restriction of residual strain in shape memory alloys. Here we consider the general case of residual strain such that \( \epsilon_{res} < \epsilon_{t} \) upon unloading to zero stress as in Figure 1(c). (Note that \( \epsilon_{res} < \epsilon_{t} \) can also be achieved by unloading the material before \( \xi \) achieves a value of 1.) Using the definition of maximum residual strain, and the micromechanics concepts governing the value of the martensite fraction, it follows that

\[ \epsilon_{res} = \epsilon_{t} \xi_{s} \] (32)

where the subscript “res” indicates the residual value of strain after unloading to \( \sigma = 0 \). Applying the simplest case of residual strain to Equation (31), take an initial state of \((\sigma_{0} = \epsilon_{0} = \xi_{0} = 0)\) and a final state of \((\sigma = 0, \epsilon = \cdots)\).
\( \varepsilon_m = \varepsilon_L \tilde{\varepsilon}_L \) with \( T = T_0 \) (and consequently \( \xi_T = \xi_{T0} \)). This yields

\[
0 = D(\xi)\varepsilon_L \tilde{\varepsilon}_L + \Omega(\xi)\tilde{\varepsilon}_L \\
+ \varepsilon_L(D_m - D_s) \frac{\xi_T^3}{2} + C(\xi) - C(0)
\]

which, upon recalling the definition [Equation (27c)], implies

\[
C(\xi) = -\varepsilon_L(D_m - D_s) \frac{\xi_T^3}{2}
\]

(33)

And consequently the final constitutive equation for shape memory alloy behavior with material functions that are linear in \( \xi \) is

\[
\sigma - \sigma_0 = D(\xi)\varepsilon_L - D(\xi_0)\varepsilon_L + \Omega(\xi)\tilde{\varepsilon}_L \\
- \Omega(\xi_0)\tilde{\varepsilon}_0 + \Theta(T - T_0)
\]

(34)

Following the derivation technique outlined here, it would be possible to obtain the constitutive relation for SMAs with material properties that are more general functions of the martensite fraction or functions of the other independent variables. It is important to note that the constitutive law [Equation (7) or (20)], derived assuming that the material functions \( D, \Omega, \Theta \) are all constant, can then not subsequently be utilized with non-constant material functions. With any change in the functionality of the material properties with the state variables, the constitutive law must be rederived from the basic differential form.

**NUMERICAL EXAMPLES**

In this section, the constitutive Equation (34) coupled with transformation Equations (22) and (23) is utilized to calculate the thermomechanical response of shape memory alloys in several cases. Stress-strain curves representative of the shape memory effect and the pseudoelastic effect are given first. Since this model is most closely based on previous work by Liang and Rogers, two effects clearly illustrated in their work, typical of shape memory alloys in general, will also be presented here: free strain recovery and restrained recovery. In all cases, the numerical results agree well with experimental observations.

The material properties for the shape memory alloy in the following examples are taken from data given by Dye (1990) and Liang (1990) on a nitinol alloy (NiTi). The values for the necessary material properties are listed in Table 1. Note that the value for \( C_m \) is taken from the experimental curves for critical transformation stress given by Dye, which correspond well with Figure 4, i.e., the curve does not pass through the Martensite Start temperature at zero stress. The experimental data indicate a slight increase in the values of critical transformation stress at temperatures below \( M_s \), but \( \sigma_m^c \) and \( \sigma_m^a \) are taken to be constants here. Additionally, although the experiments show a decrease in the maximum residual strain at temperatures above \( A_s \), this decrease is not considered in these examples.

Constitutive Equation (34), transformation Equations (22) and (23), and the data from Table 1 were utilized to calculate the stress-strain curves of the shape memory alloy at various temperatures. The results for a wide range of temperatures are shown in Figure 5 and the results for temperatures less than \( M_s \) are grouped together in Figure 6. For all of these curves, the initial value of the stress-induced martensite variable is clearly zero, indicated by lack of residual initial strain. For temperatures above \( M_s \), the initial value of the temperature-induced martensite variable was taken to be zero and for temperatures less than \( M_s \), the initial values of \( \xi_T \) were proportional to temperature as indicated by Equation (22e). With these initial conditions, only the curve for \( T = 5^\circ C \) is representative of a fully martensitic specimen before loading. At \( T = 12^\circ C \) and \( T = 15^\circ C \), the material is partially martensitic and partially austenitic prior to application of stress, and at all higher temperatures the material is fully austenitic. The subsequent different initial values of the modulus functions are manifested in the difference in slope of the linear loading portion of the stress-strain curves. The slight variation of slope of the unloading portion of the curves in Figure 6 arises from small austenite contributions to the modulus at \( T = 12^\circ C \) and \( T = 15^\circ C \), since the material has not completely converted to detwinned martensite at the final strain shown here. See Figure 7 for comparison, where stress-strain curves extending to 100% transformation for two temperatures are given. In Figure 7, the material transforms completely to detwinned martensite at both temperatures, after which the stress-strain curve again becomes linear with a slope of \( D_s \). Upon unloading the maximum residual strain \( \varepsilon_L \) is achieved. Due to the assumed constant values of \( \sigma_m^c \) and \( \sigma_m^a \) in this example, the stress-strain curves for materials with \( \xi_T = 1 \) prior to loading will be identical regardless of temperature and will coincide with the curve for \( T = 5^\circ C \).

The curves at temperatures less than Austenite Start are all indicative of the shape memory effect: the material loads

| Table 1. Material properties for the nitinol alloy used in the following examples [Dye, 1990; Liang, 1990]. |
|---|---|---|---|
| Moduli | Transformation Temperatures | Transformation Constants | Maximum Residual Strain |
| \( D_s = 67 \times 10^3 \) MPa | \( M_s = 9^\circ C \) | \( C_m = 8 \) MPa/\( ^\circ C \) | \( \varepsilon_L = 0.067 \) |
| \( D_s = 26.3 \times 10^3 \) MPa | \( M_s = 18.4^\circ C \) | \( C_s = 13.8 \) MPa/\( ^\circ C \) | |
| \( \Theta = 0.55 \) MPa/\( ^\circ C \) | \( A_s = 34.5^\circ C \) | \( \sigma_m^c = 100 \) MPa | |
| | \( A_s = 49^\circ C \) | \( \sigma_m^a = 170 \) MPa | |
elastically, then undergoes conversion of martensite variants and/or transformation of austenite to deformed martensite during the nonlinear portion of the stress-strain curve, and finally unloads elastically (with no pseudoelastic recovery), incurring a residual strain. To complete the shape memory effect for the curves at temperatures below $A_r$, the material temperature must be raised above the Austenite Finish temperature at zero stress for the material to recover all of the residual strain. Equations (23a) and (34) together provide for this condition (Figure 9). The pseudoelastic effect is demonstrated in Figure 5 by the curves for $T = 40^\circ C$ and $T = 60^\circ C$. Since the lower of these two temperatures is less than $A_r$, there is only a partial pseudoelastic strain recovery on unloading and the material consists of both detwinned martensite and austenite after unloading. At $T = 60^\circ C$, above $A_r$, the material exhibits a complete hysteresis loop during the procedure: the material is austenite prior to loading, transforms to detwinned martensite during loading and completes the inverse transformation to austenite upon unloading.

Figure 8 illustrates the change in the two portions of the martensite fraction while loading the material above the
critical transformation stress. This figure shows the case of \( T = 5^\circ\text{C} \) where the material is initially 100\% temperature-induced martensite and then begins conversion to stress-induced martensite after the critical stress is exceeded. Note that here the sum \( \xi_t + \xi_s \) is always identically 1 since the test temperature is below \( M_f \).

Figures 9 and 10 demonstrate free strain recovery of the material, in which the material recovers an initial residual strain at zero stress by raising the temperature above \( A_f \). This case is then illustrative of the completion of the SME for curves in Figures 5 and 6. The example shown here is of a specimen at \( T_a = 20^\circ\text{C} \) with \( \xi_{T0} = 0.5 \) and \( \varepsilon_0 = 0.02 \); \( \xi_{SO} \) is then defined by \( \varepsilon_0 / \xi_T \). As the temperature is raised from \( T_a \), the inverse transformation to austenite begins and the material starts to recover the residual strain at \( A_s \). At \( A_f \) this transformation is complete and the strain and both martensite fraction variables are zero. Note that in Figure 10 \( \xi = \xi_T + \xi_s \) at all times as required.

**Figure 7.** Stress-strain curves to maximum residual strain, \( \varepsilon_L \).

**Figure 8.** Martensite fractions vs. stress: \( \xi_{T0} = 1, \xi_{SO} = 0, T = 5^\circ\text{C} \) from stress-strain analysis with \( \sigma_0 = \varepsilon_0 = 0 \), loading portion only.
The final example presented here is the case of restrained recovery in which a material with residual strain is constrained to maintain that deformation as the temperature is raised through $A_s$ and $A_f$. Figure II shows a material with 0.5% residual strain subjected to these conditions ($\varepsilon = \varepsilon_0 = 0.005$). Because the material is restrained as the inverse transformation to austenite occurs, and the material would recover the residual strain as in the last example if unrestrained, extremely large internal stresses are incurred. Figure II also shows the critical transformation stress curves for ease of comparison. It is clear then that upon heating, the internal stress increases rapidly during transformation to austenite, after transformation is complete the stress remains essentially constant, and as the specimen is cooled the internal stress decreases rapidly in the region of austenite to martensite transformation. Note that the hysteresis loop here starts and ends with a definite value of stress; this value is the amount of stress required at $T = 20^\circ C$ to achieve a 0.005 strain. The hysteresis loop can also be calculated starting from zero stress with the same $\varepsilon_0$, as shown in Figure 12; after the austenite to martensite transformation reaches the original temperature $T_0$ upon cooling, the stress is then decreased to zero to complete the cycle.

It is important to note that all the results presented agree
extremely well both quantitatively and qualitatively with experimental data on SMA materials. The one exception to this statement is the case of restrained recovery in which, with the current material parameters, the stress required for 100% transformation to austenite upon heating increasingly exceeds experimental values with increasing initial residual strain. The general characteristics of the model’s prediction and the experimental results agree quite well, only the magnitude of the maximum stress achieved differs significantly. This discrepancy is most likely due to the experimental material exceeding the critical stress for true plasticity before achieving complete conversion to austenite. Experimental measurements for restrained recovery are generally done with 0.01 or 0.02 residual strain (detwinned martensite), which is indeed far too large to expect that the material could attain 100% austenite conversion without encountering plasticity. Given experimental data on the critical plastic limits for SMAs, these could be simply incorporated into the model so that the maximum stress for restrained recovery would reflect their influence.
CONCLUSION

A comprehensive one-dimensional constitutive law for shape memory alloys has been derived from first principles based on previous work by Tanaka and Liang. A separation of the martensite fraction internal variable into stress-induced and temperature-induced components was proposed and introduced into the model. This subdivision is justified by the micromechanical behavior of SMAs and is effected here such that the stress-induced martensite fraction represents the extent of transformation of the material into a single martensite-variant oriented with the direction of loading. With this distinction it was shown that the model can capture both stress-induced martensitic transformation at temperatures above Austenite Start with accompanying pseudoelastic behavior and the shape memory effect associated with conversion of martensite variants (or transformation of austenite to martensite) with applied stress at all temperatures.

In this paper, the constitutive law was derived for the case of specific material functions suggested by current experimental evidence. The method, however, is extensible to material properties which are general functions of the state variables. For example, if the material properties for certain SMA materials are found to differ not only for the austenite and martensite phases, but also for twinned and detwinned phases, the separation of the martensite variable fraction presented in this paper will allow such properties to be included in the constitutive description of the material.

In general, the derivation method presented here can be followed to incorporate any new experimental evidence into the constitutive law as such evidence becomes available. The current model is also easily expanded according to new experimental findings for critical conversion and transformation stresses. It is most important to realize that the basic constitutive law must be rederived from the differential form with each change in the expression of material functions or state variables.

The current model exhibits the desired characteristics stated in the Introduction for material characterization: this model accurately represents shape memory material behavior, uses common engineering variables and measurable material properties, and the mathematical formulation is easily incorporated into further practical applications. In addition, the formulation is easily adaptable as new information becomes available on SMA materials in general or certain subclasses. Ongoing investigations will extend the model to two dimensions and pursue a finite element implementation, with the goal of using this theory as an active part of engineering design with shape memory alloys.

ACKNOWLEDGEMENTS

The author wishes to gratefully acknowledge the Deutsche Forschungsanstalt für Luft- und Raumfahrt and the American Association for University Women Educational Foun-

dation—Postdoctoral Fellowship Fund for supplying the resources and the funding with which this work was completed.

REFERENCES


