Pergamon

Int. J. Solids Structures Vol. 31, No. 16, pp. 2229–2249, 1994

0020–7683/94 $7.00 + .00

0020–7683(94)E0002–D

A ONE-DIMENSIONAL CONTINUUM MODEL FOR SHAPE-MEMORY ALLOYS

ROHAN ABEYARATNE and SANG-JOO KIM
Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

and

JAMES K. KNOWLES
Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, U.S.A.

(*Received* 17 *February* 1993; *in revised form* 11 *December* 1993)

Abstract—In this paper we construct an explicit one-dimensional constitutive model that is capable of describing some aspects of the thermomechanical response of a shape-memory alloy. The model consists of a Helmholtz free-energy function, a kinetic relation and a nucleation criterion. The free-energy is associated with a three-well potential energy function; the kinetic relation is based on thermal activation theory; nucleation is assumed to occur at a critical value of the appropriate energy barrier. The predictions of the model in various quasi-static thermomechanical loadings are examined and compared with experimental observations.

1. INTRODUCTION

Thermoelasticity theory has been used to study certain general issues pertaining to solids that undergo reversible stress- and temperature-induced phase transitions. In this theory, the potential energy function that characterizes the material, as a function of deformation gradient, has multiple energy-wells for certain ranges of stress and temperature, and each energy-well is identified with a phase (or variant of a phase) of the material. For example, the studies reported by Abeyaratne (1983), Ball and James (1987) and Ericksen (1975) address issues pertaining to stable equilibrium configurations of such materials, while Abeyaratne and Knowles (1988, 1990, 1991, 1992) and Heidug and Lehner (1985) examine questions related to evolution.

A complete constitutive theory that describes the behavior of such materials consists of a Helmholtz free-energy function which describes the response of each individual phase, a nucleation criterion which signals the conditions under which the transition from one phase to another commences, and a kinetic law which characterizes the rate at which this transition progresses. Explicit examples of these ingredients have been constructed by various authors. For example, Ericksen (1986) and Silling (1989) have constructed three-dimensional Helmholtz free-energy functions that model certain crystals. Falk (1980) has studied a one-dimensional polynomial free-energy function and Jiang (1992) has used a similar characterization in anti-plane shear. Models of kinetic relations have been constructed, for example, by Otsuka *et al.* (1976) by relating phase boundary motion to dislocation motion and by Achenbach (1989) and Müller and Wilmansky (1981) by using certain statistical considerations.

Recently, Abeyaratne and Knowles (1993) presented, in one dimension, simple models for each of these ingredients: their Helmholtz free-energy function was associated with a piecewise linear material, while their kinetic relation was based on thermal activation theory; they took nucleation to occur at a critical value of 'driving force'. They also made a qualitative comparison of their theory with certain experiments on some shape memory alloys.

The present study generalizes the model of Abeyaratne and Knowles (1993). The potential energy function in the aforementioned model has at most two energy-wells, and

thus it only accounts for two phases. However, even in the simplest one-dimensional setting, one often encounters three material configurations—a parent phase and two variants of the product phase. The principal generalization of the present paper is to construct a three-well energy function analogous to the two-well model. This allows us to simulate some experiments such as the one carried out by Ehrenstein (1985), in which a bar composed initially of equal amounts of two martensitic variants was subjected to a slowly oscillating stress, and at the same time was heated and then cooled. At various stages during the ensuing process, the bar involved two martensitic variants, two martensitic variants and austenite, one martensitic variant and austenite, and pure austenite. A three-well energy is essential for modeling such a phenomenon.

As in Abeyaratne and Knowles (1993), the kinetic law utilized in the present study is based on thermal activation theory. The nucleation criterion that we adopt here is based on a critical value of the appropriate 'energy barrier' and is motivated, in part, by the models used to describe nucleation in the materials science literature; see e.g. Christian (1975) and Fine (1975).

The number of interfaces that can occur in the bar is controlled by the nucleation criterion. For the particular criterion adopted here, this number is small, usually one, so that if many interfaces are to be allowed, we would have to modify this ingredient of the present model. In some experiments, e.g. those of Burkart and Read (1953), Grujicic *et al.* (1985), Krishnan (1985), Otsuka *et al.* (1976), the transformation has been observed to occur by the propagation of a single interface; in others, however, e.g. those of Müller and Xu (1991), a large number of interfaces have been seen.

In Section 2 we briefly outline the version of thermoelasticity that we intend to use. In Section 3 we construct the three-well energy function and discuss the stability of the various phases. Next, in Section 4 we calculate the energy barriers associated with this energy function and use them to establish the nucleation criterion and kinetic relation. In Section 5 we carry out a number of simulations and compare them with experimental observations.

2. PRELIMINARIES

In this section we briefly review some relevant concepts from the continuum theory of thermomechanical processes within a purely one-dimensional setting that corresponds to uniaxial stress in a bar. A more detailed discussion of the one-dimensional theory may be found in Abeyaratne and Knowles (1993); some general aspects of the three-dimensional theory may be found in Abeyaratne and Knowles (1990).

The bulk behavior of a thermoelastic material may be characterized by its Helmholtz free-energy per unit mass $\psi(\gamma, \theta)$, where γ is strain and θ is temperature; the stress σ and specific entropy η at a particle are then constitutively related to γ, θ through

$$\sigma = \rho\psi_\gamma(\gamma, \theta), \quad \eta = -\psi_\theta(\gamma, \theta), \tag{2.1}$$

where ρ denotes the mass density in the reference configuration. The potential energy per unit reference volume $G(\gamma; \theta, \sigma)$ of the material is defined by

$$G(\gamma; \theta, \sigma) = \rho\psi(\gamma, \theta) - \sigma\gamma, \tag{2.2}$$

and its value at an extremum of $G(\cdot; \theta, \sigma)$ coincides with the Gibbs free energy per unit reference volume $g(\gamma, \theta) = \rho\psi(\gamma, \theta) - \rho\psi_\gamma(\gamma, \theta)\gamma$.

In order to model a material that can undergo a thermoelastic phase transition, the function $G(\cdot; \theta, \sigma)$ should have multiple local minima ('energy wells') when the temperature and stress take on suitable values; the corresponding Helmholtz free-energy potential $\psi(\cdot, \theta)$ will be non-convex, and the stress–strain curve characterized by $\sigma = \rho\psi_\gamma(\gamma, \theta)$ will be non-monotonic. In this theory, each local minimum of the potential energy function G, and therefore each branch with positive slope of the stress–strain curve, is identified with a different (metastable) phase of the material.

Suppose that $G(\cdot\,;\theta,\sigma)$ has at least two local minima corresponding to a given temperature θ and stress σ, and let $\gamma = \bar{\gamma} = \bar{\gamma}(\theta,\sigma)$ and $\gamma = \overset{+}{\gamma} = \overset{+}{\gamma}(\theta,\sigma)$ denote the values of strain at these two energy-wells. Then the strain at a particle that is subjected to this θ and σ could be *either* $\bar{\gamma}$ *or* $\overset{+}{\gamma}$ depending on which energy-well (i.e. phase) the particle is in. Let $x = s(t)$ locate a point in the reference configuration at time t; suppose that the particle immediately to the left of $x = s(t)$ has a strain $\bar{\gamma}$ while the strain on its right is $\overset{+}{\gamma}$; then $x = s(t)$ denotes the location of a phase boundary, i.e. an interface that separates two distinct phases of the material. During a quasi-static process, the rate of entropy production $\Gamma(t)$ in a segment of the bar that contains the interface $x = s(t)$ but no other phase boundaries, and which is at a uniform temperature $\theta(t)$, can be shown to be

$$\Gamma(t) = f(t)\dot{s}(t)/\theta(t), \tag{2.3}$$

where

$$f = G(\overset{+}{\gamma}\,;\theta,\sigma) - G(\bar{\gamma}\,;\theta,\sigma)\,; \tag{2.4}$$

f is known as the 'driving force' acting on the phase boundary. A discussion of the notion of driving force in a general three-dimensional setting which includes the effects of inertia and is not restricted to thermoelastic materials may be found in Abeyaratne and Knowles (1990). The second law of thermodynamics thus requires that the following *entropy inequality* hold:

$$f\dot{s} \geqslant 0. \tag{2.5}$$

If the driving force f happens to vanish, one speaks of the states $(\overset{+}{\gamma},\theta)$ and $(\bar{\gamma},\theta)$ as being in 'phase equilibrium' and of the quasi-static process as taking place 'reversibly'. If $G(\overset{+}{\gamma}\,;\theta,\sigma) > G(\gamma\,;\theta,\sigma)$, then f is positive and so according to (2.5) one has $\dot{s} \geqslant 0$; thus if the phase boundary propagates, it moves into the positive side and thereby transforms particles from $(\overset{+}{\gamma},\theta)$ to $(\bar{\gamma},\theta)$. In this sense, the material prefers the smaller minimum of G. This is also true in the reverse case when $G(\overset{+}{\gamma}\,;\theta,\sigma) < G(\bar{\gamma}\,;\theta,\sigma)$. One therefore speaks of the phase associated with the lowest energy-well as being the *stable* phase.

By using the first law of thermodynamics one can show that the heat generated when a unit mass of material changes phase from $(\overset{+}{\gamma},\theta)$ to $(\bar{\gamma},\theta)$ is $f/\rho+\lambda$ where $\lambda = \theta(\overset{+}{\eta}-\bar{\eta})$ is the *latent heat*; if the phase change occurs under conditions of phase equilibrium, then $f = 0$, and the heat generated is λ.

Let $x = s(t)$ denote the Lagrangian location of a phase boundary at time t. As particles cross this interface, they transform from one phase to another at a rate that is determined by the underlying 'kinetics'. The kinetics of the transformation control the rate of mass flux $\rho\dot{s}$ across the phase boundary. If one assumes that this flux depends only on the states (γ,θ) and $(\overset{+}{\gamma},\theta)$ on either side of the interface, then the propagation of the phase boundary is governed by a relation of the form

$$\dot{s} = v(\bar{\gamma},\overset{+}{\gamma},\theta), \tag{2.6}$$

where the kinetic response function v is determined by the material. Alternatively, since the constitutive relations $\sigma = \rho\psi_\gamma(\overset{+}{\gamma},\theta)$ and $\sigma = \rho\psi_\gamma(\bar{\gamma},\theta)$ can be inverted (separately) for each phase, one can express $\overset{+}{\gamma}$ and $\bar{\gamma}$ in terms of σ and θ, and thus re-write the kinetic law (2.6) in the form

$$\dot{s} = v(\sigma, \theta), \tag{2.7}$$

where the function v depends on the two particular phases involved in the transformation and is different for each pair of phases. Finally, by substituting the inverted stress–strain–temperature relations into (2.4) one can express the driving force acting on an interface between a given pair of phases as $f = \hat{f}(\theta, \sigma)$; this in turn can be inverted at each fixed θ, and so the *kinetic law* can be expressed in the form

$$\dot{s} = V(f, \theta). \tag{2.8}$$

The basic principles of the continuum theory do not provide any further information regarding the kinetic response function V; in particular, explicit examples of V must be supplied by suitable constitutive modeling.

A particle changes its phase either when it crosses a moving phase boundary or by the alternative mechanism of *nucleation*. Consider, for example, a quasi-static process in a bar which involves only a single phase for some initial interval of time, and two-phase states at subsequent times. The kinetic law describes the evolution of *existing* phase boundaries and therefore is only operational once the bar is in a two-phase state. The *transition* of the bar from a single-phase configuration to a two-phase configuration is controlled by a 'nucleation criterion'. If a particle is to change phase from a state $(\bar{\gamma}, \theta)$ to a state $(\overset{+}{\gamma}, \theta)$ through nucleation, some critical condition of the form $\Phi(\theta, \sigma) = 0$ should be reached at that particle, where Φ depends on the pair of phases involved in the transition. Models for Φ are often based on the notion of 'energy barriers', as will be described in Section 4.

We close this section with a brief discussion of some qualitative features of the one-dimensional model that is to be constructed in the sections to follow. Consider momentarily the *three*-dimensional continuum theory, and suppose for purposes of discussion that the material at hand exists in a cubic phase (austenite) and an orthorhombic phase (martensite); an example of this is the class of Cu-Al-Ni shape-memory alloys studied extensively by Otsuka and co-workers (e.g. Otsuka *et al.*, 1976), and more recently by Chu and James (1993). In view of certain invariance requirements, the associated three-dimensional potential energy function G must have seven energy-wells corresponding to the austenite phase and the six 'variants' of martensite. During a uniaxial test of a suitably oriented single crystal specimen, and for suitable values of temperature θ, the material is found to remain in the cubic phase for sufficiently small values of stress, in the orthorhombic phase with the long side of the crystal parallel to the tensile axis for sufficiently large tensile stresses, and in the orthorhombic phase with its long side normal to the tensile axis (i.e. in a different variant of the orthorhombic phase) for sufficiently large compressive stresses. In the *one*-dimensional theory we model this by allowing G to have three energy-wells for suitable θ and σ; the ones at the largest and smallest values of strain correspond to the two variants of martensite just described, while the one at the intermediate value of strain is identified with austenite. Since the variants of martensite are crystallographically identical to each other when $\sigma = 0$, the energy-wells corresponding to them are required to have the same height at all temperatures whenever the stress vanishes. Moreover, all three energy-wells should have the same height if the stress vanishes and the temperature coincides with the transformation temperature θ_T. At higher temperatures, the phase with greater symmetry (austenite) is usually preferred over the low-symmetry phase, and so the model is to be constructed such that the austenite energy-well is lower than the martensite wells when $\theta > \theta_T$; the reverse is true for $\theta < \theta_T$. The crystallographic similarity between the variants also suggests that the specific entropy associated with them should be identical, and therefore that the latent heat associated with the transformation from one variant to another should be zero; this too is a feature of our model.

Finally, a note on terminology: for simplicity of presentation we shall sometimes speak of the 'three phases' rather than the 'one phase and two variants'; similarly we shall often use the term 'phase boundary' generically to refer to both an interface between two phases and to an interface between two variants (which ought to be called a twin boundary).

3. HELMHOLTZ FREE-ENERGY FUNCTION

In this section we construct an explicit three-well Helmholtz free-energy function $\psi(\gamma, \theta)$ that characterizes the response of a multi-phase material; the three energy-wells are viewed as corresponding to an austenitic phase and to two variants of martensite. Recall first that the elastic modulus μ, coefficient of thermal expansion α and specific heat at constant strain c of a thermoelastic material are related to $\psi(\gamma, \theta)$ through

$$\rho\psi_{\gamma\gamma}(\gamma, \theta) = \mu, \quad -\psi_{\gamma\theta}(\gamma, \theta)/\psi_{\gamma\gamma}(\gamma, \theta) = \alpha, \quad -\theta\psi_{\theta\theta}(\gamma, \theta) = c. \tag{3.1}$$

Thus if μ, α and c are constant on some domain of the (γ, θ)-plane, then by integrating (3.1) one finds that

$$\rho\psi(\gamma, \theta) = (\mu/2)(\gamma - g_*)^2 - \mu\alpha\gamma\theta + \rho c(\theta - \theta \log(\theta/\theta_*)) + \rho\psi_* \tag{3.2}$$

on that domain, where g_*, θ_* and ψ_* are constants.

Consider a material which exists in a high temperature phase austenite (A) and as two variants (M^+ and M^-) of a low-temperature phase martensite. Suppose for simplicity that the austenite and both martensitic variants have the same constant elastic modulus $\mu > 0$, the same constant coefficient of thermal expansion $\alpha > 0$ and the same constant specific heat $c > 0$. (The model that follows can be readily generalized to describe the case wherein the different phases have different but constant material properties.) By (3.2), the Helmholtz free-energy function $\psi(\gamma, \theta)$ associated with this material must have the form

$$\rho\psi(\gamma, \theta) = \begin{cases} (\mu/2)(\gamma - g_1)^2 - \mu\alpha\gamma\theta + \rho c(\theta - \theta \log(\theta/\theta_1)) + \rho\psi_1 & \text{on } P_1, \\ (\mu/2)(\gamma - g_2)^2 - \mu\alpha\gamma\theta + \rho c(\theta - \theta \log(\theta/\theta_2)) + \rho\psi_2 & \text{on } P_2, \\ (\mu/2)(\gamma - g_3)^2 - \mu\alpha\gamma\theta + \rho c(\theta - \theta \log(\theta/\theta_3)) + \rho\psi_3 & \text{on } P_3, \end{cases} \tag{3.3}$$

where ρ is the mass density of the material in the reference configuration, and θ_i, g_i, ψ_i, $i = 1, 2, 3$, are nine additional material constants whose physical significance will be made clear in what follows. The regions P_1, P_2 and P_3 of the (γ, θ)-plane on which the three expressions in (3.3) hold are the regions on which the respective phases A, M^+ and M^- exist; they are assumed to have the form shown in Fig. 1, where in particular the boundaries of P_1, P_2 and P_3 have been taken to be straight lines. The temperature levels θ_m and θ_M denote two critical values of temperature: for $\theta > \theta_M$ the material only exists in its austenite

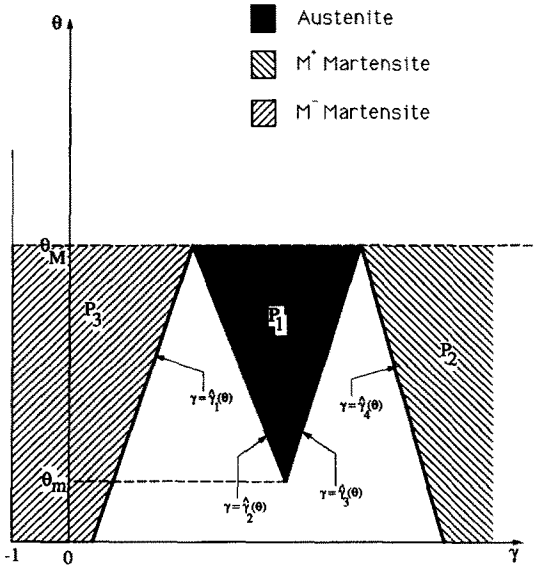


Fig. 1. Regions P_1, P_2 and P_3 in the (γ, θ)-plane.

form, whereas for $\theta < \theta_m$ the material only exists in its martensite forms. Throughout this paper we will restrict attention to temperatures less than θ_M.

We now impose a number of restrictions on the potential function G in order that it properly model the stress-free response of the material we have in mind. Since G and the Helmholtz free-energy function ψ coincide when the stress vanishes, any characteristic to be assigned to G at $\sigma = 0$ could be equivalently imposed on ψ. First, we assume that there is a special temperature θ_T between θ_m and θ_M such that, when $\sigma = 0$, all three phases A, M^+ and M^- are stable. Thus, the function $\psi(\cdot, \theta_T)$ must have three local minima, i.e. 'energy-wells', and the values of ψ at these three minima must be the same. The minima occurring at the smallest, intermediate and largest values of strain identified with M^-, A and M^+ respectively. Next, we require that the values of ψ at the two martensite energy-wells should coincide at every θ at which these energy-wells exist, reflecting the fact that M^+ and M^- are 'variants' of each other. Finally, for $\theta > \theta_T$ the martensite wells must be higher than the austenite well, while for $\theta < \theta_T$ they should be lower, so that austenite is preferred at higher temperatures, martensite at lower temperatures. The special temperature θ_T is called the *transformation temperature*.

On enforcing these requirements on the function ψ defined by (3.3), one finds that

$$\psi_1 - \psi_2 = \psi_1 - \psi_3 \equiv \lambda_T > 0, \tag{3.4}$$

$$\log(\theta_2/\theta_1) = \frac{\mu\alpha}{\rho c}(g_2 - g_1) + \lambda_T/(c\theta_T),$$

$$\log(\theta_3/\theta_1) = \frac{\mu\alpha}{\rho c}(g_3 - g_1) + \lambda_T/(c\theta_T). \tag{3.5}$$

In (3.5) we have let λ_T denote the common value of $\psi_1 - \psi_2$ and $\psi_1 - \psi_3$; one can readily verify that λ_T represents the *latent heat* of the austenite $\rightarrow$ martensite transitions at the transformation temperature θ_T and that the latent heat associated with the M^+–M^- transition is zero.

The stress-response function $\hat{\sigma}(\gamma, \theta) = \rho\psi_\gamma(\gamma, \theta)$ associated with (3.3) is

$$\hat{\sigma}(\gamma, \theta) = \begin{cases} \mu(\gamma - g_1) - \mu\alpha\theta & \text{on } P_1, \\ \mu(\gamma - g_2) - \mu\alpha\theta & \text{on } P_2, \\ \mu(\gamma - g_3) - \mu\alpha\theta & \text{on } P_3. \end{cases} \tag{3.6}$$

Two graphs of $\hat{\sigma}(\gamma, \theta)$ versus γ are shown in Fig. 2: Fig. 2(a) corresponds to a fixed value of temperature in the range $\theta_m < \theta < \theta_M$, and the stress–strain curve shows three rising branches corresponding to the three phases A, M^+ and M^-; Fig. 2(b) is associated with a temperature in the range $0 < \theta < \theta_m$ and the two rising branches of the stress–strain curve correspond to the variants M^+ and M^-.

Of the three parameters g_1, g_2 and g_3, one is fixed by the choice of reference configuration, while the other two are determined through the transformation strains. In particular, if we choose stress-free austenite at the transformation temperature θ_T to be the reference state, then by setting $\hat{\sigma}(0, \theta_T) = 0$ in $(3.6)_1$, one obtains

$$g_1 = -\alpha\theta_T. \tag{3.7}$$

Next, let γ_T (>0) denote the *transformation strain* [Fig. 2(a)] between each martensitic variant and austenite. Then, from (3.6),

$$\gamma_T = g_2 - g_1 = g_1 - g_3 > 0. \tag{3.8}$$

(This can be readily generalized to the case where the transformation strain between phases A and M^+ is, say, γ_T^+, and that between A and M^- is $\gamma_T^- \neq \gamma_T^+$.) Finally, if (3.4), (3.5), (3.7)

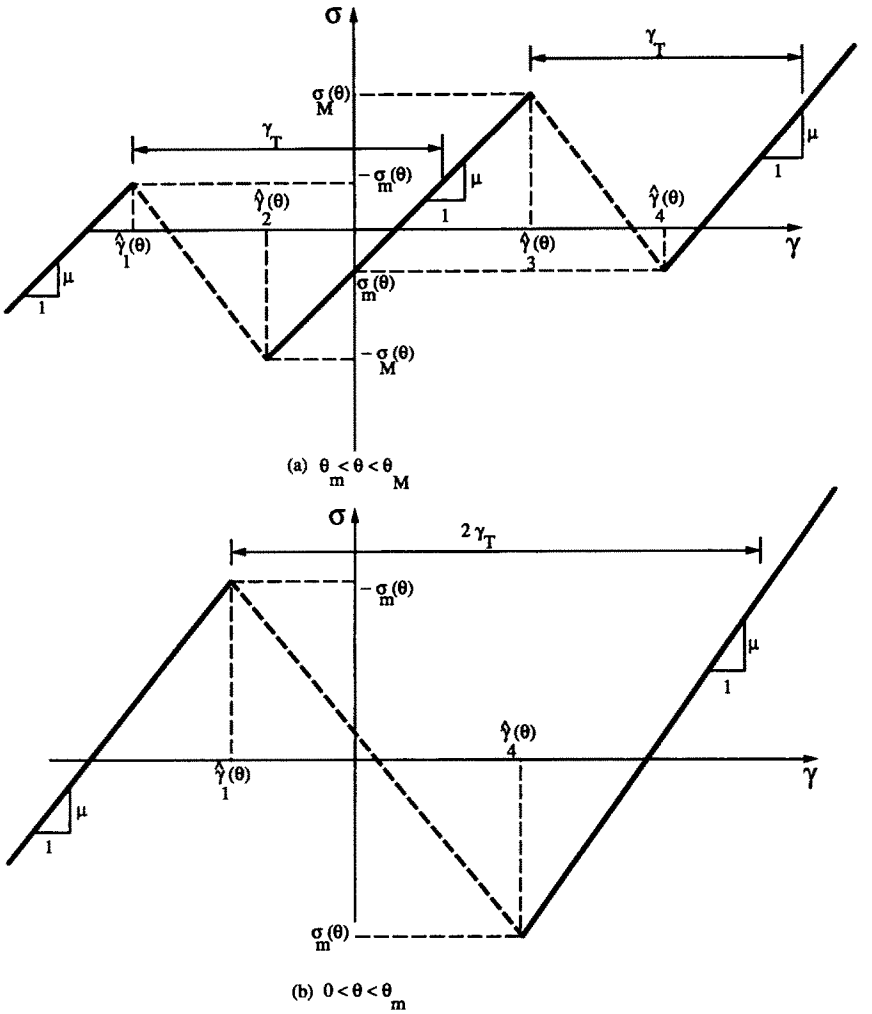


Fig. 2. Stress–strain curves at constant temperatures θ.

and (3.9) are substituted back into (3.3), one finds that $\rho\psi(\gamma,\theta)$ contains the term $\rho\psi_1+(\mu/2)\alpha^2\theta_T^2+\rho c\theta\log(\theta_1/\theta_T)$ as an inessential linear function of temperature which may be eliminated by taking

$$\rho\psi_1 = -(\mu/2)\alpha^2\theta_T^2, \quad \theta_1 = \theta_T. \tag{3.9}$$

In *summary*, the material at hand is characterized by the common elastic modulus μ, specific heat at constant strain c and coefficient of thermal expansion α of the phases; the stress-free transformation temperature θ_T; the mass density ρ in the reference state; the latent heat λ_T at the temperature θ_T and the transformation strain γ_T. The Helmholtz free-energy function is given by combining (3.3) with (3.4), (3.5), (3.7)–(3.9):

$$\rho\psi(\gamma,\theta) = \begin{cases} (\mu/2)\gamma^2-\mu\alpha\gamma(\theta-\theta_T)+\rho c\theta(1-\log(\theta/\theta_T)) & \text{on } P_1,\\ (\mu/2)(\gamma-\gamma_T)^2-\mu\alpha(\gamma-\gamma_T)(\theta-\theta_T)+\rho c\theta(1-\log(\theta/\theta_T))-\rho\lambda_T(1-\theta/\theta_T) & \text{on } P_2,\\ (\mu/2)(\gamma+\gamma_T)^2-\mu\alpha(\gamma+\gamma_T)(\theta-\theta_T)+\rho c\theta(1-\log(\theta/\theta_T))-\rho\lambda_T(1-\theta/\theta_T) & \text{on } P_3. \end{cases} \tag{3.10}$$

The various other thermo-mechanical characteristics of the material can now be derived from (3.10). In particular, the stress-response function $\hat{\sigma}(\gamma,\theta) = \rho\psi_\gamma(\gamma,\theta)$ is given by

$$\hat{\sigma}(\gamma,\theta)=\begin{cases}\mu\gamma-\mu\alpha(\theta-\theta_T) & \text{on } P_1,\\ \mu(\gamma-\gamma_T)-\mu\alpha(\theta-\theta_T) & \text{on } P_2,\\ \mu(\gamma+\gamma_T)-\mu\alpha(\theta-\theta_T) & \text{on } P_3.\end{cases} \tag{3.11}$$

In order to complete the description of the Helmholtz free-energy function we need to specify the regions P_1, P_2 and P_3 of the (γ,θ)-plane shown in Fig. 1, i.e. we need to specify the boundary curves $\gamma=\hat{\gamma}_i(\theta)$ shown in the figure. To this end, we first prescribe the stress-levels at the local maxima and minima of the stress–strain curve. As indicated in Fig. 2, we take, for simplicity, these stress-levels to be given by $\pm\sigma_M(\theta)$ and $\pm\sigma_m(\theta)$. In view of our earlier assumption that the boundaries of the regions P_1, P_2, P_3 are straight, the functions $\sigma_M(\theta)$ and $\sigma_m(\theta)$ must be linear in θ. Moreover, since according to Fig. 1 we must have $\hat{\gamma}_2(\theta_m)=\hat{\gamma}_3(\theta_m)$, $\hat{\gamma}_3(\theta_M)=\hat{\gamma}_4(\theta_M)$ and $\hat{\gamma}_1(\theta_M)=\hat{\gamma}_2(\theta_M)$, it follows that $\sigma_M(\theta_m)=0$ and $\sigma_M(\theta_M)-\sigma_m(\theta_M)=\mu\gamma_T$. Thus

$$\left.\begin{aligned}\sigma_M(\theta)&=\mu M(\theta-\theta_m) && \text{for } \theta_m<\theta<\theta_M,\\ \sigma_m(\theta)&=\mu m(\theta-\theta_M)+\mu M(\theta_M-\theta_m)-\mu\gamma_T && \text{for } 0<\theta<\theta_M,\end{aligned}\right\} \tag{3.12}$$

where m and M are positive material constants. The equations $\gamma=\hat{\gamma}_i(\theta)$, $i=1,2,3,4$, describing the boundaries of P_1, P_2 and P_3 are then given by $\pm\sigma_M(\theta)=\hat{\sigma}(\hat{\gamma}_i(\theta),\theta)$, $i=3,2$, and $\pm\sigma_m(\theta)=\hat{\sigma}(\hat{\gamma}_i(\theta),\theta)$, $i=4,1$.

Thus far, we have only described ψ on the ('metastable') portion $P_1+P_2+P_3$ of the (γ,θ)-plane. It is not necessary, for the purposes of the present section, to specify an explicit form for ψ on the remaining ('unstable') portion of this plane; any function ψ which is once continuously differentiable, is such that ψ_γ is negative on the unshaded portion of Fig. 1, and conforms with (3.3) would be acceptable. An infinite number of such functions exist, provided only that the material parameters satisfy certain inequalities; this is discussed in the Appendix.

Next, it is useful to map the regions P_i, $i=1,2,3$, of the (γ,θ)-plane associated with the respective phases A, M^+ and M^- on to the (θ,σ)-plane using $\sigma=\hat{\sigma}(\gamma,\theta)$. The result of this mapping is displayed in Fig. 3 where P_i' is the image of P_i. Given the stress σ and temperature θ at a particle, Fig. 3 shows all of the phases that are available to that particle.

Finally, we address the issue of the stability of the phases. The potential energy function $G(\gamma;\theta,\sigma)=\rho\psi(\gamma,\theta)-\sigma\theta$ of the material at hand can be calculated using (3.10). At each (θ,σ), G has one or more local minima. Where G has multiple energy-wells, one can use the explicit formula for G to determine the particular minimum that is smallest. This determines the phase that is stable. The result of this calculation is displayed in Fig. 4. The stress-level $\sigma_0(\theta)$ indicated in the figure is given by

$$\sigma_0(\theta)=\frac{\rho\lambda_T}{\gamma_T}\left(\frac{\theta}{\theta_T}-1\right), \tag{3.13}$$

and is known as the Maxwell stress for the $A-M^+$ transition. The Maxwell stress for the $A-M^-$ and M^+-M^- transitions are $-\sigma_0(\theta)$ and 0 respectively. The two states A and M^+ that are both associated with any particular point on the boundary $\sigma=\sigma_0(\theta)$ both have the same value of potential energy G and both are stable; if these states co-exist and are separated by a phase boundary, the driving force on that interface would be zero and the phases would be in phase equilibrium.

Suppose momentarily that a particle always chooses the phase that is stable from among all phases available to it. Then the response of a particle as the stress or temperature is slowly varied is fully determined by Fig. 4. For example, consider a fixed temperature θ greater than the transformation temperature θ_T. As the stress σ is increased monotonically from a sufficiently negative value, the particle is in the martensite variant M^- until the stress reaches the value $-\sigma_0(\theta)$; it then transforms to austenite and remains in this phase

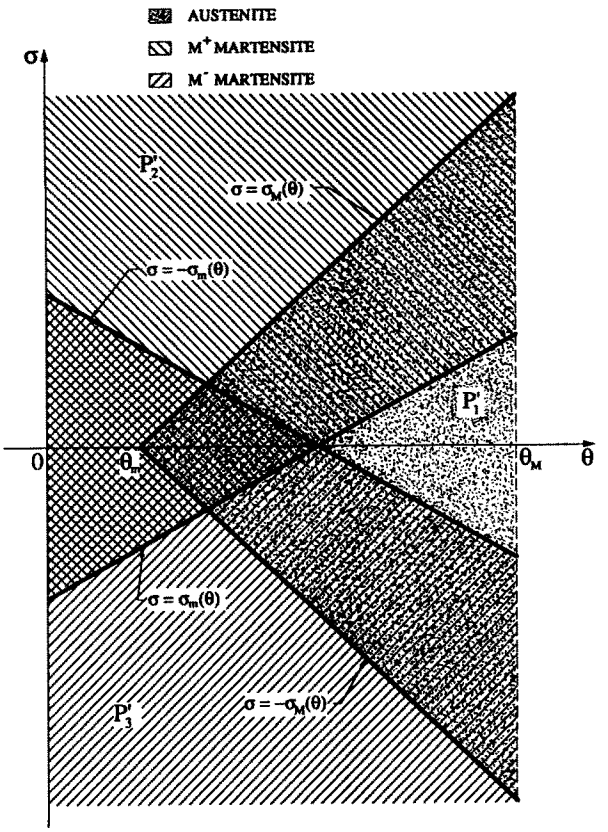


Fig. 3. Available phases at a given (θ, σ).

as the stress increases over the intermediate range $-\sigma_0(\theta) < \sigma < \sigma_0(\theta)$; at $\sigma = \sigma_0(\theta)$ the particle transforms to M^+ and remains there for $\sigma > \sigma_0(\theta)$.

The immediately preceding discussion assumed that a particle is always in the stable phase. In solids however, particles can often remain for long times in states that are merely metastable and the transition from a metastable phase to a stable phase is controlled by additional considerations, viz. nucleation and kinetics. We now turn to these issues.

4. NUCLEATION AND KINETICS

Given the stress σ and temperature θ at a particle, Fig. 3 shows the various phases that that particle can adopt, while Fig. 4 indicates which among them is the stable phase. If a particle happens to be in a phase that is not its stable phase, the questions of whether, and how fast, it will transform into the stable phase are answered by a nucleation criterion and a kinetic law. In this section we will describe simple models for nucleation and kinetics under the assumption that the phase transitions are 'thermally activated'.

A particle can change its phase in one of two ways: a particle whose phase is the same as that of its neighboring particles could change its phase spontaneously through 'nucleation' if the stress and temperature at that particle satisfy an appropriate *nucleation criterion*. Alternatively, a particle in one phase adjacent to a particle in a different phase and separated from it by a phase boundary, will change phase through 'growth' if the phase boundary propagates toward it, the motion of the phase boundary being controlled by a *kinetic law*.

Energy barriers

Figure 5 shows two schematic graphs of the potential energy function $G(\gamma\,;\,\theta, \sigma)$ plotted versus γ for fixed (θ, σ). Figure 5(a) shows three local minima and corresponds to a pair (θ, σ) at which all three phases co-exist [i.e. the point (θ, σ) lies in the region common to

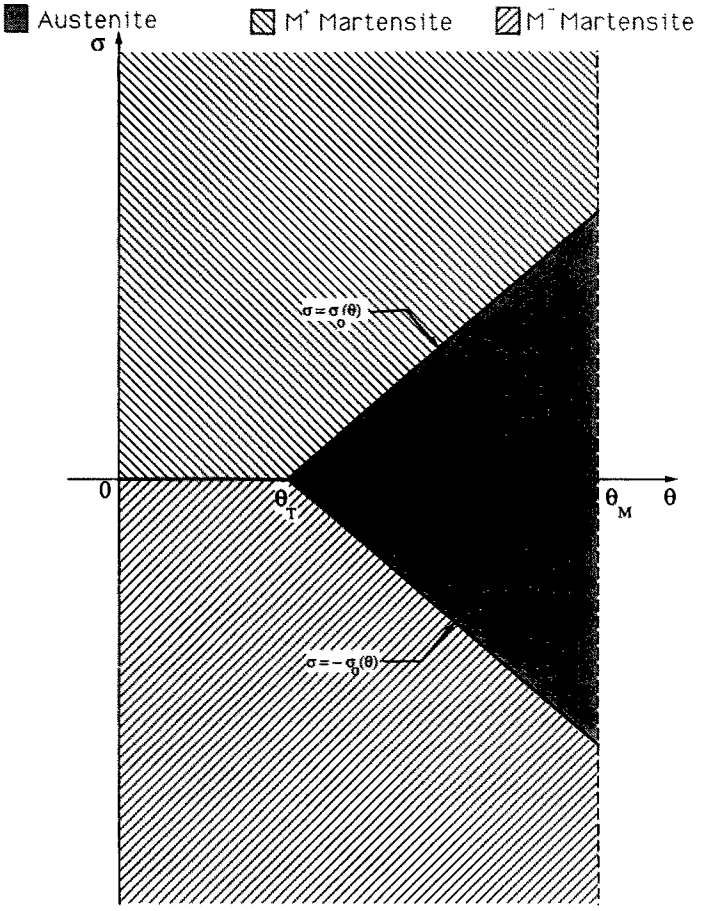


Fig. 4. The stable phases.

P'_1, P'_2 and P'_3 in Fig. 3]; Fig. 5(b) corresponds to a pair (θ, σ) at which only the two martensite variants co-exist. Figures 5(a) and (b) of the potential energy function correspond to the respective Figs 2(a) and (b) of the stress–strain curve. The six quantities $b_{ij}(\theta, \sigma)$ indicated in Fig. 5 are the *energy barriers* to a transformation from phase i to phase j, where it is convenient to use the subscripts 1, 2 and 3 to refer to the phases A, M^+ and M^- respectively. In order to calculate these energy barriers, we need expressions for the values of G at the local maxima, and this in turn requires a knowledge of the Helmholtz free energy function ψ on the unshaded regions of Fig. 1. While we can suitably continue (3.10) into this 'unstable region' in many different ways, our present purpose is merely to construct a simple continuum model that describes the various features of the theory. Consequently, we will simply extend the parabolas which describe G near its local minima in Fig. 5 until they intersect, and use the values of G at these intersection points as estimates for the values of G at the local maxima. Using (3.10) to calculate $G(\gamma\,;\theta,\sigma) = \rho\psi(\gamma,\theta) - \sigma\gamma$ and then carrying out this calculation leads to the following expressions for the six energy-barriers:

$$
\begin{aligned}
b_{13}(\theta,\sigma) &= (2\mu)^{-1}[\sigma+\sigma_0(\theta)+\mu\gamma_T/2]^2,\\
b_{31}(\theta,\sigma) &= (2\mu)^{-1}[\sigma+\sigma_0(\theta)-\mu\gamma_T/2]^2,\\
b_{12}(\theta,\sigma) &= (2\mu)^{-1}[\sigma-\sigma_0(\theta)-\mu\gamma_T/2]^2,\\
b_{21}(\theta,\sigma) &= (2\mu)^{-1}[\sigma-\sigma_0(\theta)+\mu\gamma_T/2]^2,\\
b_{23}(\theta,\sigma) &= (2\mu)^{-1}[\sigma+\mu\gamma_T]^2,\\
b_{32}(\theta,\sigma) &= (2\mu)^{-1}[\sigma-\mu\gamma_T]^2,
\end{aligned}
\tag{4.1}
$$

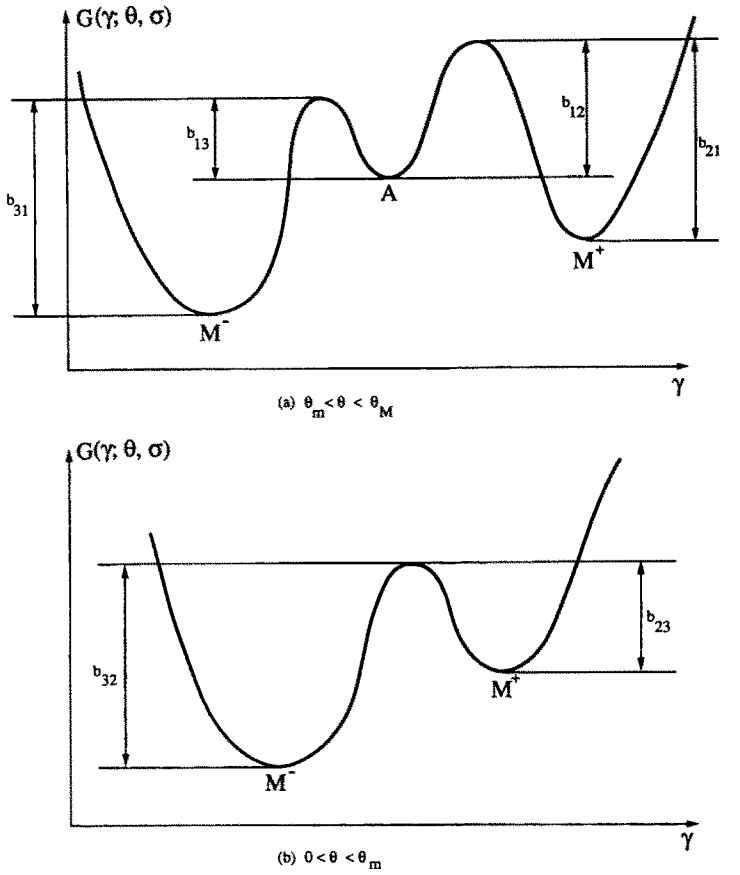


Fig. 5. Potential energy at fixed temperature and stress as a function of strain.

where each $b_{ij}(\theta, \sigma)$ is defined for values of (θ, σ) at which the ith and jth phases co-exist, i.e. the point (θ, σ) lies in the region common to P_i' and P_j' in Fig. 3; $\sigma_o(\theta)$ is the austenite–martensite Maxwell stress introduced previously in (3.13).

Nucleation

Considering first the question of nucleation, we suppose that a particle which is in phase i will transform to phase j by nucleation if the relevant energy barrier $b_{ij}(\theta, \sigma)$ is less than some critical number n_{ij}; the associated nucleation criterion is thus given by setting $b_{ij}(\theta, \sigma) = n_{ij}$ in (4.1). In view of the symmetry of the potential energy function G when $\sigma = 0$, we shall assume that both the $A \to M^+$ and $A \to M^-$ transitions nucleate at the same value of temperature if the stress vanishes; a similar assumption for the reverse $M^+ \to A$ and $M^- \to A$ transitions will also be made. The former value of nucleation temperature is denoted by M_s (for 'martensite start') while the latter is denoted by A_s (for 'austenite start'). Finally, we will also assume that at any given temperature θ, the nucleation stress-level for the $M^+ \to M^-$ transition is the negative of the nucleation stress-level for the reverse $M^- \to M^+$ transition at that same temperature. One can readily enforce these restrictions on the n_{ij}s; when combined with (4.1), this leads to the following *nucleation criteria* for the various transitions:

$$\begin{aligned}
&\sigma + \sigma_o(\theta) \leqslant \sigma_o(M_s) && \text{for } A \to M^-,\\
&\sigma + \sigma_o(\theta) \geqslant \sigma_o(A_s) && \text{for } M^- \to A,\\
&\sigma - \sigma_o(\theta) \geqslant -\sigma_o(M_s) && \text{for } A \to M^+,\\
&\sigma - \sigma_o(\theta) \leqslant -\sigma_o(A_s) && \text{for } M^+ \to A,\\
&\sigma \geqslant \Sigma && \text{for } M^- \to M^+,\\
&\sigma \leqslant -\Sigma && \text{for } M^+ \to M^-,
\end{aligned} \tag{4.2}$$

where $\sigma_0(\theta)$ is the austenite–martensite Maxwell stress given by (3.13) and the constants M_s, A_s and Σ are characteristic of the material; necessarily, $M_s < \theta_T < A_s$ and $\Sigma > 0$.

Figure 6 illustrates these nucleation criteria on the (θ, σ)-plane. If the inequalities in (4.2) hold with equality, the resulting equations describe a set of straight lines in the (θ, σ)-plane; the nucleation criterion states that, as indicated in the figure, crossing one of these lines nucleates an associated transition. Figure 6, as shown, corresponds to a material for which the critical nucleation stress-level given by (4.2) for the $M^- \to A$ transition exceeds the corresponding stress-level for the $A \to M^+$ transition for some range of temperature, i.e.

$$\Sigma > (1/2)(\rho\lambda_T/\gamma_T\theta_T)(A_s - M_s); \tag{4.3}$$

this need not necessarily be the case.

If the current state of the bar were to involve either a temperature or stress gradient, one could determine the *location* in the bar at which nucleation occurs. In this paper we will consider a uniform bar that is always subjected to uniform stress and temperature fields; the location of the nucleation site in this bar is therefore rather arbitrary. If the bar had been rendered inhomogeneous by a slight uniform taper with the small end at $x = 0$ and the large end at $x = L$, then the transition from a low-strain phase to a high-strain phase would necessarily commence at $x = 0$ and the reverse transition would occur at $x = L$. We shall arbitrarily assume that this is the case in our uniform bar as well.

Kinetics

Let $x = s(t)$ denote the location at time t of a phase boundary in a bar, and let $(\bar{\gamma}, \theta, \sigma)$ and $(\dot{\bar{\gamma}}, \theta, \sigma)$ denote the strain, temperature and stress at the two particles adjacent to the

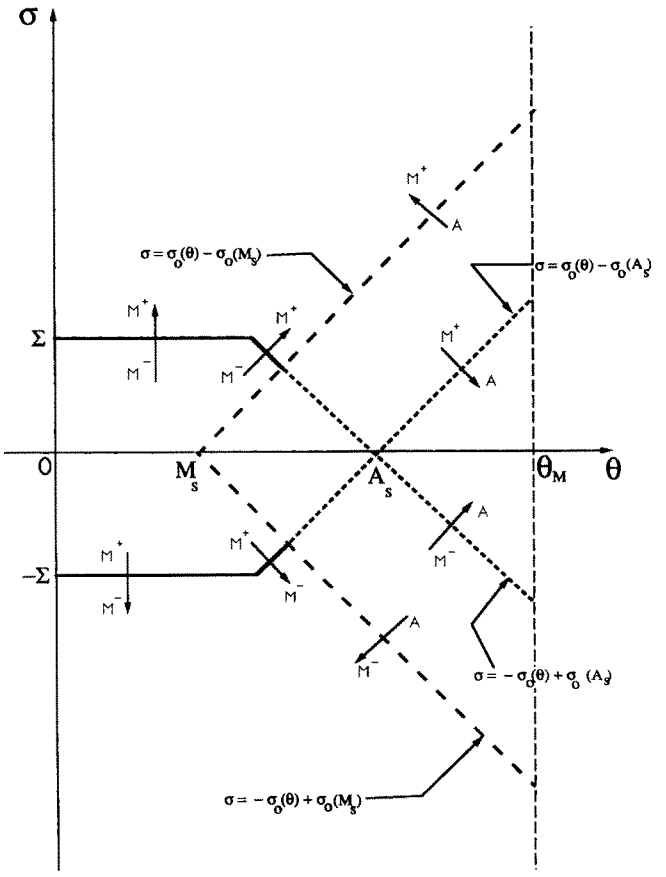


Fig. 6. Nucleation criteria.

phase boundary on its left and right, respectively. Suppose that the particle on the left is in phase i while the particle on the right is in phase j (recall that $i = 1, 2, 3$ corresponds to phases A, M^+, M^- respectively). The driving force $f = G(\overset{+}{\gamma}, \theta, \sigma) - G(\bar{\gamma}, \theta, \sigma)$ on an i/j phase boundary can be calculated from (4.1):

$$\begin{aligned}
f &= \gamma_T(\sigma - \sigma_0(\theta)) && \text{for an } M^+/A \text{ interface,}\\
f &= -\gamma_T(\sigma - \sigma_0(\theta)) && \text{for an } A/M^+ \text{ interface,}\\
f &= \gamma_T(\sigma + \sigma_0(\theta)) && \text{for an } A/M^- \text{ interface,}\\
f &= -\gamma_T(\sigma + \sigma_0(\theta)) && \text{for an } M^-/A \text{ interface,}\\
f &= 2\sigma\gamma_T && \text{for an } M^+/M^- \text{ interface,}\\
f &= -2\sigma\gamma_T && \text{for an } M^-/M^+ \text{ interface.}
\end{aligned} \tag{4.4}$$

As the phase boundary propagates through the bar, the particle immediately in front of it 'jumps' from one local minimum of G to another, and an explicit model of the kinetic relation may be constructed by viewing this jumping process on an atomic scale. In order to jump from one minimum of G to the other, an atom must acquire an energy at least as great as that represented by the relevant energy barrier: for an atom undergoing a phase $i \to$ phase j transition this barrier is $b_{ij}(\sigma, \theta)$; for the reverse phase $j \to$ phase i transition it is $b_{ji}(\sigma, \theta)$. Under suitable assumptions about the statistics of this process, the probability that the energy of an atom is at least as great as B is $\exp(-B/K\theta)$ where K is Boltzmann's constant. The average rate at which atoms jump from one minimum to the other is taken to be proportional to the probability of exceeding the corresponding energy barrier; we assume for simplicity that the proportionality factor is the same for the phase $i \to$ phase j and phase $j \to$ phase i transitions. The velocity $\dot{s}$ of the phase boundary, being the macroscopic measure of the net rate at which atoms change from phase i to phase j, is then taken to be the difference in the average rates associated with the $i \to j$ and the $j \to i$ transitions:

$$\dot{s} = R_{ij}\left\{\exp\left(-\frac{b_{ij}(\sigma,\theta)}{rK\theta}\right) - \exp\left(-\frac{b_{ji}(\sigma,\theta)}{rK\theta}\right)\right\} \tag{4.5}$$

where r denotes the number of atoms per unit *reference* volume and R_{ij} is a positive proportionality factor, related in part to the frequency with which atoms attempt to cross over to the new phase.

Substituting (4.1) into (4.5) now leads to an explicit representation for the kinetic relations of the various transitions in the form $\dot{s} = v_{ij}(\sigma, \theta)$. By using (4.4), they can be expressed in the alternative form $\dot{s} = V_{ij}(f, \theta)$:

$$\dot{s} = 2R_{ij}\exp\left\{-\frac{f^2 + \mu^2 g^4/4}{2\mu r g^2 K\theta}\right\}\sinh\left\{\frac{f}{2rK\theta}\right\}, \tag{4.6}$$

where $g = \gamma_T$ for both M^+/A and A/M^- interfaces, and $g = 2\gamma_T$ for an M^+/M^- interface. These kinetic relations automatically satisfy the condition $fV_{ij}(f, \theta) \geqslant 0$, so that any motion consistent with them will conform with the entropy inequality (2.5). Moreover, at each fixed θ, $V_{ij}(f, \theta)$ increases monotonically with f, so that the greater the driving force, the faster the speed of propagation. If the driving force f is small, so that quasi-static processes take place close to phase equilibrium, one can approximate (4.6) to obtain the following kinetic relation which is linear in f:

$$\dot{s} \approx \frac{R_{ij}}{rK\theta} \exp\left\{-\frac{\mu g^2}{8rK\theta}\right\} f. \tag{4.7}$$

5. THERMO-MECHANICAL RESPONSE OF THE MODEL

In this section we will utilize the constitutive model constructed in Sections 3 and 4 to determine the uniaxial response of a bar when it is subjected to various thermo-mechanical loadings. We describe the analysis associated with one of these loadings in some detail; the analysis corresponding to the remaining cases is conceptually similar.

Consider the isothermal mechanical loading of a uniform bar at a temperature $\theta > A_s$. The bar is initially unstressed and is composed of austenite. As the stress $\sigma(t)$ is monotonically increased, the bar remains in this phase for some time $0 < t < t_1$. By (3.11), the elongation δ of the bar during this stage of loading (measured from the reference state) is given by

$$\delta(t)/L = \sigma(t)/\mu + \alpha(\theta - \theta_T) \quad \text{for } 0 < t < t_1. \tag{5.1}$$

From Fig. 6 and the paragraph below (4.3) we conclude that M^+ martensite is nucleated at the left end of the bar at the instant $t = t_1$, where t_1 is given by

$$\sigma(t_1) = \sigma_0(\theta) - \sigma_0(M_s). \tag{5.2}$$

During the next stage $t_1 < t < t_2$, the bar is composed of M^+ martensite on the interval $0 < x < s(t)$ and austenite on $s(t) < x < L$. By (3.11),

$$\delta(t) = s(t)[\sigma(t)/\mu + \gamma_T + \alpha(\theta - \theta_T)] + [L - s(t)][\sigma(t)/\mu + \alpha(\theta - \theta_T)] \quad \text{for } t_1 < t < t_2, \tag{5.3}$$

where the phase boundary location $s(t)$ is found by integrating the appropriate kinetic relation in (4.6), (4.4), i.e.

$$\dot{s}(t) = 2R \exp\left\{-\frac{f^2(t) + \mu^2\gamma_T^4/4}{2\mu r\gamma_T^2 K\theta}\right\} \sinh\left\{\frac{f(t)}{2rK\theta}\right\}, \quad s(t_1) = 0, \tag{5.4}$$

with the driving force $f(t) = [\sigma(t) - \sigma_0(\theta)]\gamma_T$. At the instant $t = t_2$, the phase boundary reaches the right end of the bar, $s(t_2) = L$. For $t \geqslant t_2$ the bar consists entirely of M^+ martensite and its response, according to (3.11), is given by

$$\delta(t)/L = \sigma(t)/\mu + \gamma_T + \alpha(\theta - \theta_T) \quad \text{for } t > t_2. \tag{5.5}$$

A similar analysis can be used to describe the response corresponding to subsequent unloading, as well as to loading by compressive stress in which case the M^- variant is involved instead of M^+.

In order to study the quantitative, as well as qualitative, suitability of our model, we chose values for the material parameters that are of the correct order of magnitude for a Cu-Al-Ni shape-memory alloy:

$$\mu = 3 \times 10^{10}\ \text{N/m}^2, \quad \rho = 8000\ \text{kg/m}^3, \quad \alpha = 16 \times 10^{-6}/^\circ\text{K}, \quad c = 400\ \text{J/kg}\,^\circ\text{K},$$
$$\gamma_T = 0.05, \quad \lambda_T = 5.7 \times 10^3\ \text{J/kg}, \quad \theta_T = 307\ ^\circ\text{K},$$
$$A_s = 308\ ^\circ\text{K}, \quad M_s = 306\ ^\circ\text{K}, \quad \Sigma = 2.5 \times 10^7\ \text{N/m}^2,$$
$$K = 1.381 \times 10^{-23}\ \text{J/}^\circ\text{K}, \quad R_{ij} = 0.0448\ \text{m/s}, \quad r = 9.046 \times 10^{28}\ \text{atoms/m}^3. \tag{5.6}$$

The values of the transformation and nucleation temperatures θ_T, A_s and M_s were taken

from Otsuka *et al.* (1979) and correspond to an alloy whose composition is Cu–14.0 Al–4.2 Ni (wt%). By comparison, the remaining material parameters are less sensitive to the alloy composition and the values chosen for them do not correspond to an alloy of this precise composition. The modulus μ, the transformation strain γ_T and the latent heat λ_T at the transformation temperature were estimated using data from Otsuka and Wayman (1976). The values of the mass density ρ, coefficient of thermal expansion α and specific heat c were estimated using information in the *Metals Handbook* (1979). The value of the M^+/M^- nucleation stress Σ was obtained from Sakamoto *et al.* (1979). The value of R_{12} was estimated by using our kinetic relation (4.6) in conjunction with the austenite–martensite interface velocities measured by Grujicic *et al.* (1985) and reported in their Fig. 5; the remaining R_{ij}s were arbitrarily assumed to have this same value. The number of atoms per unit reference volume, r, was calculated by using the mass density, alloy composition and the atomic masses of Cu, Al and Ni.

There are four other material parameters, viz. m, M, θ_m and θ_M, that are involved in the description of our model. Even though they do *not* affect the response of the material, the validity of the model requires that such numbers exist in a manner that is consistent with the various constitutive inequalities given in the appendix. One can show that there is a range of acceptable values for these parameters.

6. RESULTS

(i) Figure 7 shows two force–elongation curves corresponding to isothermal mechanical load cycling by the application of a prescribed stress. Figure 7(a) corresponds to a temperature below M_s, with the bar transforming between the M^- and M^+ variants without

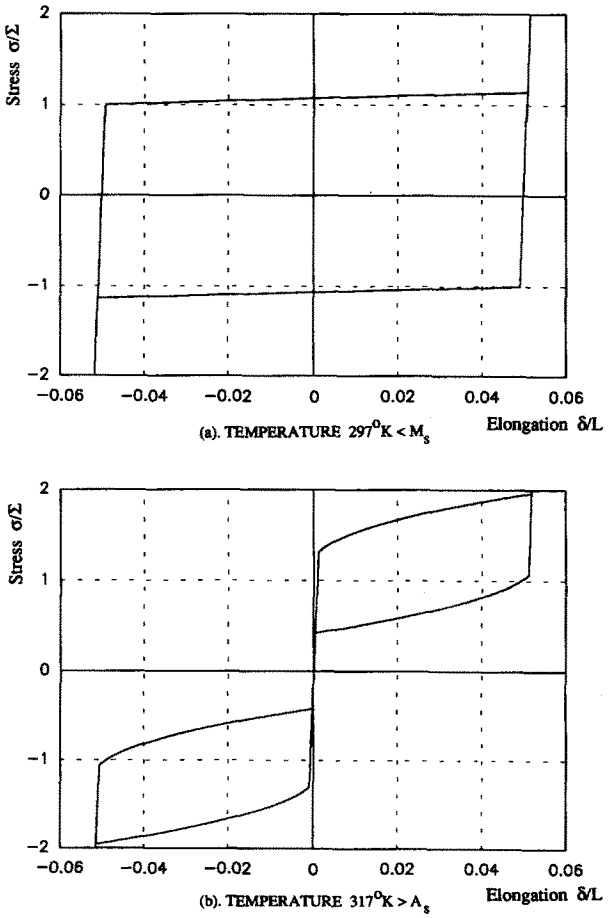


Fig. 7. Isothermal mechanical loading.

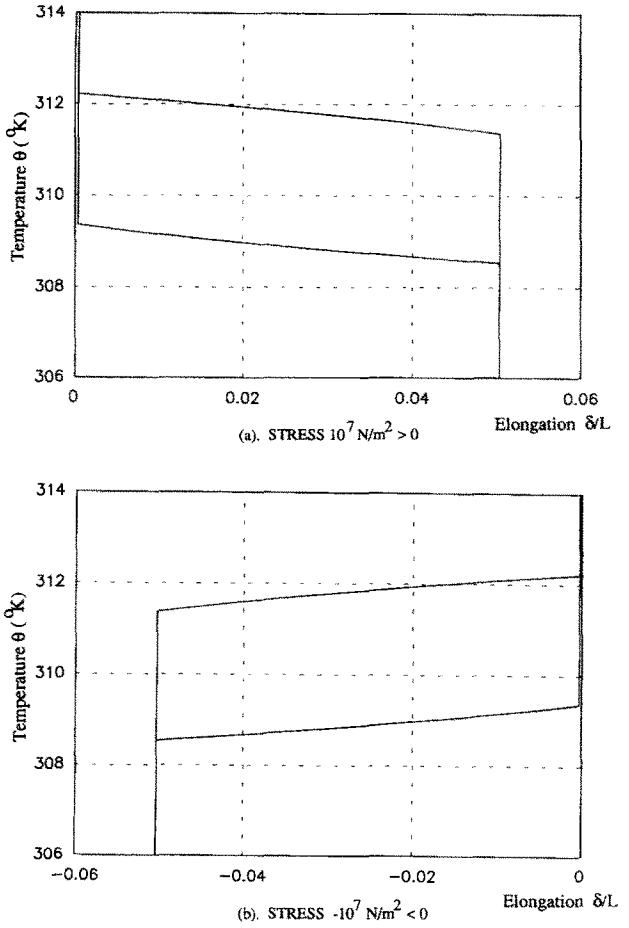


Fig. 8. Thermal loading at constant stress.

involving austenite. Figure 7(b) corresponds to a temperature greater than A_s; as the stress increases from a sufficiently negative value, the bar transforms from M^- to A and then from A to M^+. The loading and unloading rate underlying both of these figures is $|\dot{\sigma}(t)| = 5 \times 10^4$ N/m^2 s. The response depicted in these figures is similar to that observed by Nakanishi (1983) for Au-Cd; see Figs 10, 13 in their paper.

(ii) Figure 8 shows two elongation–temperature curves which result from cycling the temperature with the stress held fixed; Figs 8(a) and (b) correspond respectively to tensile and compressive values of the applied stress. The bar transforms between the phases A and M^+ in the former case, between the phases A and M^- in the latter. Observe that the transformation from the high temperature phase (austenite) to the low temperature phase involves an elongation in Fig. 8(a) and a contraction in Fig. 8(b). The heating and cooling rate underlying both of these figures is $|\dot{\theta}(t)| = 0.001$ °K/s. The response depicted in these figures is similar to the response described by Krishnan *et al.* (1974), Fig. 13 for Cu-Zn at a constant tensile stress and that observed by Burkart and Read (1953), Fig. 6 for In-Tl under compressive stress.

If the initially austenitic bar remains *stress-free* as it is cooled from a high temperature, the phases M^+ and M^- are both nucleated simultaneously at $\theta = M_s$, the former at the left end of the bar, the latter at the right end; they then grow at the same rate according to their kinetic relations, and once the transformation is complete, the bar consists of an equal mixture of M^+ and M^-. Since the transformation strains involved in the $A \to M^+$ and $A \to M^-$ transformations have been taken to be γ_T and $-\gamma_T$ respectively, the length of the bar does *not* change due to transformation. The elongation–temperature response in this case is therefore a straight line. This is a trivial, one-dimensional manifestation of what is sometimes referred to as 'self accommodation'.

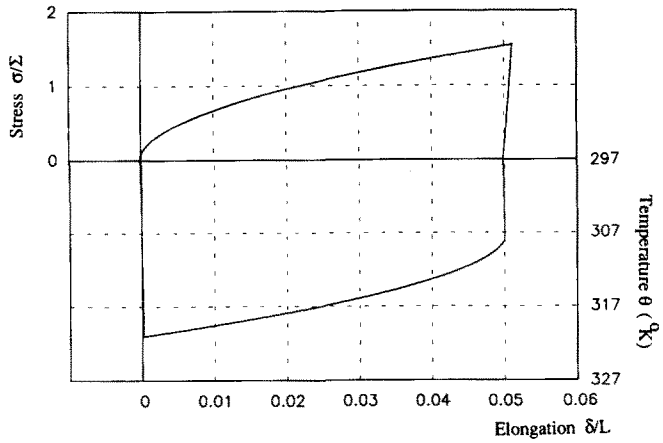
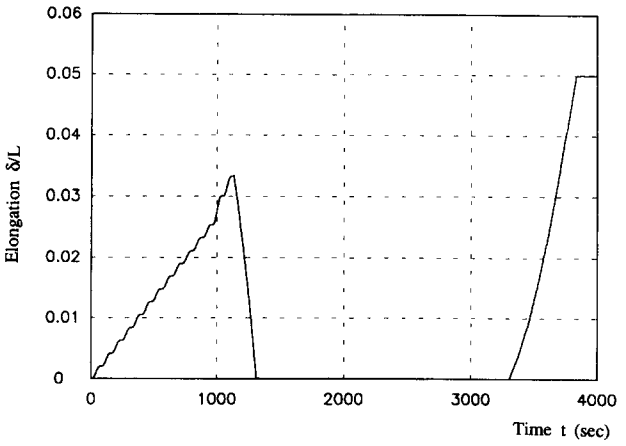


Fig. 9. The shape-memory effect.

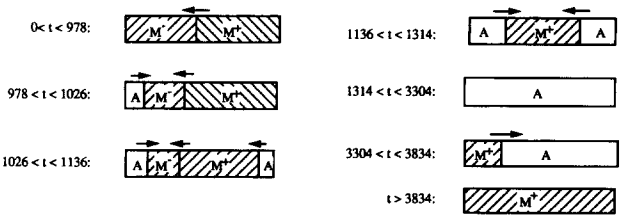
(iii) Figure 9 displays the result of a calculation which attempts to model the 'shape memory effect'. We begin with a martensitic bar which is composed of M^+ for $0 < x < L/2$ and M^- for $L/2 < x < L$ and whose initial temperature is less than M_s. The bar is first subjected to a program of isothermal mechanical loading during which time the stress is first increased and then decreased back to zero. At the end of this stage of loading, the bar is composed of M^+ only (as can be deduced from Fig. 6), the stress has returned to the value zero, and the bar has suffered a permanent elongation. During the next stage, the bar remains unstressed while it is first heated to a temperature greater than A_s (which, according to Fig. 6, transforms it to phase A) and is then cooled back to its original temperature (which, by Fig. 6, leads to a configuration involving equal amounts of M^+ and M^-); at the end of this thermal loading, the state of the bar is identical to its original state. In the calculations underlying Fig. 9 we took the mechanical loading and unloading rate to be $|\dot{\sigma}(t)| = 8 \times 10^5$ N/m^2 s and the heating and cooling rate to be $|\dot{\theta}(t)| = 0.08$ °K/s. Schematic figures similar to Fig. 9 may be found, for example, in Krishnan *et al.* (1974).

(iv) Next we simulate one of the experiments carried out by Ehrenstein (1985). Consider a martensitic bar which is initially at zero stress and at a temperature $< M_s$; the segment $0 < x < L/2$ of the bar consists of M^- while $L/2 < x < L$ consists of M^+. We consider a time interval $0 \leqslant t \leqslant t_F$ and apply a stress $\sigma(t) = \varepsilon_1(1-\cos(2\pi t/80))$ while simultaneously varying the temperature according to $\theta(t) = \theta(0)+\varepsilon_2(1-\cos(2\pi t/t_F))$. We take $0 < \varepsilon_1 < \Sigma/2$ and $\sigma(t_F/2) < \sigma_0(\theta(t_F/2))-\sigma_0(A_s)$ which ensures that the hottest temperature $\theta(t_F/2)$ is large enough to nucleate austenite. The loading parameters underlying our calculation are $\varepsilon_1 = 5 \times 10^5$ N/m^2, $\varepsilon_2 = 4$ °K, $\theta(0) = 304$ °K and $t_F = 4000$ s. The resulting elongation history is shown in Fig. 10(a) and may be compared with Ehrenstein's observations [see Fig. 2 in Achenbach (1989)]; the qualitative behavior is seen to be the same; since the material used in the experiment was Nitinol whereas the material parameters employed here correspond to Cu-Al-Ni, a quantitative comparison cannot be made. A simulation of this experiment has also been carried out by Achenbach (1989) using a different model.

The calculations show that the macroscopic response of the bar plotted in Fig. 10(a) is associated with the local transitions shown in Fig. 10(b): during an initial stage, roughly $0 < t < 978$ s, the bar consists of only phases M^- and M^+; the driving force on the M^-/M^+ interface is negative [by (4.4)] and this causes the interface to move leftward in accordance with the appropriate kinetic law (4.6); thus during this stage the amount of M^+ increases at the expense of M^-. For the heating rate used in our calculation, this leftward moving interface has not yet reached the end $x = 0$ of the bar when $t = 978$ s; at roughly this instant, the nucleation criterion for the $M^- \to A$ transition is satisfied and phase A is nucleated at the left end of the bar. As t continues to increase, the newly emerged A/M^- interface propagates to the right (since its driving force is positive) in accordance with its kinetic law, while the M^-/M^+ interface continues to move leftward; the amount of phase



(a) RESPONSE δ(t) VERSUS t.



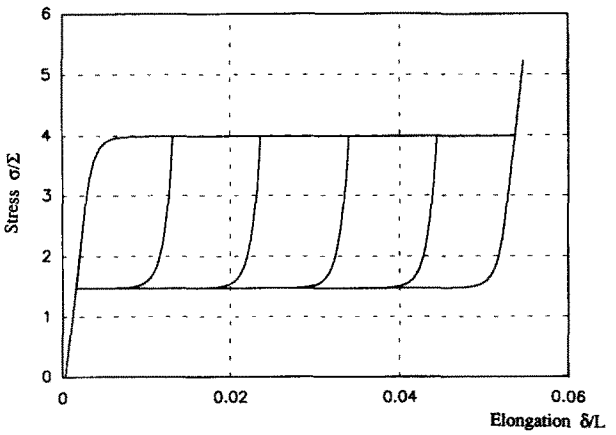
(b) PHASES IN BAR AT VARIOUS STAGES.

Fig. 10. Response to thermo-mechanical loading.

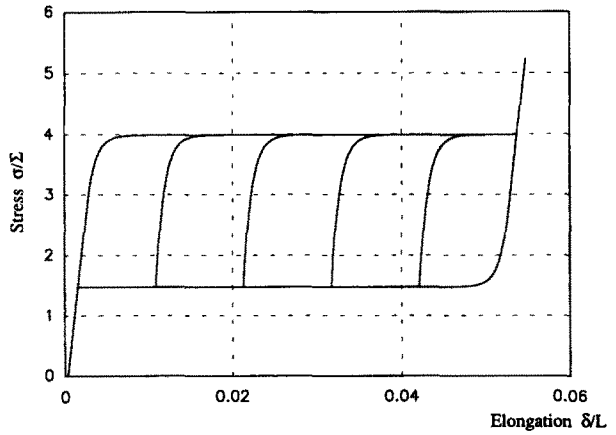
M^- thus continues to decrease while the amounts of phases A and M^+ increase. A short while later ($t \approx 1026$ s) the nucleation criterion for the $M^+ \to A$ transition is satisfied and phase A is nucleated at the right end of the bar. There are three interfaces in the bar at this time, viz. a rightward moving A/M^- interface and two leftward moving interfaces, one M^-/M^+ and the other M^+/A. At $t \approx 1136$ s the first two of these interfaces meet so that during the next stage, the bar transforms from $M^+ \to A$ as the A/M^+ interface advances towards the M^+/A interface. Eventually these two phase boundaries meet at $t \approx 1314$ s and the entire bar consists of phase A. During the next stage $1314 < t < 3304$, the bar continues to remain in phase A. The temperature which was increasing for $0 < t < 2000$ begins to decrease at $t = 2000$ s; at $t \approx 3304$ s the bar is sufficiently cool for phase M^+ to nucleate and begin to grow, until eventually at $t \approx 3834$ s the entire bar consists of M^+.

The qualitative features of the elongation history shown in Fig. 10(a) can be understood from the preceding discussion [Fig. 10(b)] by keeping in mind that M^- is the low-strain phase, A is the intermediate-strain phase and M^+ is the high-strain phase. During the initial stages $0 < t < 978$ and $978 < t < 1026$, when M^- is disappearing, first due to the growth of M^+ and then due to the growth of both M^+ and A, the bar gets longer. During the stage $1136 < t < 1314$ the bar is transforming from M^+ to A and so the bar gets shorter. Next, for $1314 < t < 3304$, the bar remains in phase A and so its length does not change appreciably. Finally, for $3304 < t < 3834$, the bar transforms from A to M^+ and so it gets longer again.

(v) Finally we calculate the response associated with repeated unloading and reloading. Consider a bar of austenite at an initial temperature $> A_s$. In the first calculation, the bar is subjected to an isothermal mechanical loading during which the elongation is increased monotonically until M^+ martensite has nucleated and begun to grow; then, *before* the bar has transformed *completely* to martensite, it is unloaded by decreasing the elongation back to zero. The calculation is now repeated, with unloading commencing at different instants. Figure 11(a) shows the result of this calculation (carried out at $\theta = 330$ °K and $|\dot{\delta}| = 8.333 \times 10^{-6}$ m/s).



(a) PARTIAL LOADING FOLLOWED BY UNLOADING.



(b) PARTIAL UNLOADING FOLLOWED BY RELOADING.

Fig. 11. Isothermal mechanical loading.

In the second calculation, the bar is subjected to an isothermal mechanical loading during which the elongation is increased until the bar has transformed completely into M^+ martensite. Next, the elongation is decreased monotonically until austenite has been nucleated and begun to grow; then, *before* the bar has transformed *completely* back to austenite, it is reloaded by increasing the elongation. The calculation is now repeated, with reloading commencing at different instants. Figure 11(b) shows the result of this calculation which was also carried out at $\theta = 330\ ^\circ\text{K}$ and $|\dot{\delta}| = 8.333 \times 10^{-6}$ m/s.

Acknowledgements—We gratefully acknowledge the support provided by ONR N00014-90-J-1871, A.R.O. DAAL 03-91-G-0011 and N.S.F. MSS 9121257.

REFERENCES

Abeyaratne, R. (1983). An admissibility condition for equilibrium shocks in finite elasticity. *J. Elasticity* **13,** 175–184.

Abeyaratne, R. and Knowles, J. K. (1988). On the dissipative response due to discontinuous strains in bars of unstable elastic material. *Int. J. Solids Structures* **24,** 1021–1044.

Abeyaratne, R. and Knowles, J. K. (1990). On the driving traction acting on a surface of strain discontinuity in a continuum. *J. Mech. Phys. Solids* **38,** 345–360.

Abeyaratne, R. and Knowles, J. K. (1991). Kinetic relations and the propagation of phase boundaries in solids. *Arch. Ration. Mech. Anal.* **114,** 119–154.

Abeyaratne, R. and Knowles, J. K. (1992). Nucleation, kinetics and admissibilitiy criteria for propagating phase boundaries. In *Shock Induced Transitions and Phase Structures* (edited by E. Dunn, R. Fosdick and M. Slemrod), IMA Volumes in Mathematics and its Applications, Vol. 52. Springer, Berlin.

Abeyaratne, R. and Knowles, J. K. (1993). A continuum model of a thermoelastic solid capable of undergoing phase transitions. *J. Mech. Phys. Solids* **41,** 541–571.

Achenbach, M. (1989). A model for an alloy with shape memory. *Int. J. Plasticity* **5,** 371–395.

Ball, J. M. and James, R. D. (1987). Fine phase mixtures as minimizers of energy. *Arch. Ration. Mech. Anal.* **100,** 15–52.
Burkart, M. W. and Read, T. A. (1953). Diffusionless phase change in the indium–thallium system. *Trans. AIME—J. Metals* **197,** 1516–1524.
Christian, J. W. (1975). *The Theory of Transformations in Metals and Alloys*, Part I. Pergamon, Oxford.
Chu, C. and James, R. D. (1993). Work in progress. Department of Aerospace Engineering and Mechanics, University of Minnesota, Minneapolis.
Ehrenstein, H. (1985). *Die Herstellung und das Formerinnerungsvermogen von Nitinol.* Dissertation, Tech. Univ. Berlin.
Ericksen, J. L. (1975). Equilibrium of bars. *J. Elasticity* **5,** 191–201.
Ericksen, J. L. (1986). Constitutive theory for some constrained elastic crystals. *Int. J. Solids Structures* **22,** 951–964.
Falk, F. (1980). Model free energy, mechanics and thermodynamics of shape memory alloys. *Acta Metall.* **28,** 1773–1780.
Fine, M. E. (1975). *Introduction to Phase Transformations in Condensed Systems.* Macmillan, New York.
Grujicic, M., Olson, G. B. and Owen, W. S. (1985). Mobility of the β_1–γ'_1 martensitic interface in Cu-Al-Ni: Part 1. Experimental measurements. *Metall. Trans.* **16A,** 1723–1734.
Heidug, W. and Lehner, F. K. (1985). Thermodynamics of coherent phase transformations in nonhydrostatically stressed solids. *Pure Appl. Geophys.* **123,** 91–98.
Jiang, Q. (1993). On the modeling of thermo-elastic phase transformations in solids. *J. Elasticity* **32,** 61–91.
Krishnan, R. V. (1985). Stress induced martensitic transformations. *Mater. Sci. Forum* **3,** 387–398.
Krishnan, R. V., Delaey, L., Tas, H. and Warlimont, H. (1974). Thermoplasticity, pseudoelasticity and the memory effects associated with martensitic transformations Part 2. The macroscopic mechanical behavior. *J. Mater. Sci.* **9,** 1536–1544.
American Society of Metals (1979). Properties and selection: Nonferrous alloys and pure metals. *Metals Handbook*, Vol. 2, 9th edn, pp. 434–435.
Müller, I. and Wilmansky, K. (1981). Memory alloys—phenomenology and Ersatzmodel. In *Continuum Models of Discrete Systems* (edited by O. Brulin and R. K. T. Hsieh). North Holland, Amsterdam, pp. 495–509.
Müller, I. and Xu, H. (1991). On the pseudo-elastic hysteresis. *Acta Metall. Materialia*, **39,** 263–271.
Nakanishi, N. (1983). Characteristics of stress–strain behavior associated with thermoelastic martensitic transformation. *Arch. Mech.* **35,** 37–62.
Otsuka, K., Sakamoto, H. and Shimizu, K. (1979). Successive stress-induced martensitic transformations and associated transformation pseudoelasticity in Cu-Al-Ni alloys. *Acta Metall.* **27,** 585–601.
Otsuka, K. and Wayman, C. M. (1976). Superelasticity effects and stress-induced martensitic transformations in Cu-Al-Ni alloys. *Acta Metall.* **24,** 207–226.
Otsuka, K., Wayman, C. M., Nakai, K., Sakamoto, H. and Shimizu, K. (1976). Superelasticity effects and stress-induced martensitic transformations in Cu-Al-Ni alloys. *Acta Metall.* **24,** 207–226.
Sakamoto, H., Tanigawa, M., Otsuka, K. and Shimizu, K. (1979). Effect of tensile and compressive stress on martensitic transformations and deformation behavior of Cu-Al-Ni alloys. In *Proc. Int. Conf. Martensitic Transformations*, ICOMAT-79 (edited by C. M. Wayman *et al.*) Cambridge, MA, pp. 633–638.
Silling, S. A. (1989). Phase change induced by deformation in isothermal elastic crystals. *J. Mech. Phys. Solids* **37,** 293–316.

APPENDIX: RESTRICTIONS ON THE MATERIAL PARAMETERS

Here we shall list all of the inequalities not displayed previously which the material parameters must satisfy. According to the statement below (3.12) the equations of the boundaries of the regions P_i in the (γ, θ)-plane are given by

$$\left.\begin{aligned} \hat{\gamma}_1(\theta) &= -\sigma_m(\theta)/\mu - \gamma_T + \alpha(\theta-\theta_T) && \text{for } 0 < \theta < \theta_M, \\ \hat{\gamma}_2(\theta) &= -\sigma_M(\theta)/\mu + \alpha(\theta-\theta_T) && \text{for } \theta_m < \theta < \theta_M, \\ \hat{\gamma}_3(\theta) &= \sigma_M(\theta)/\mu + \alpha(\theta-\theta_T) && \text{for } \theta_m < \theta < \theta_M, \\ \hat{\gamma}_4(\theta) &= \sigma_m(\theta)/\mu + \gamma_T + \alpha(\theta-\theta_T) && \text{for } 0 < \theta < \theta_M, \end{aligned}\right\} \tag{A.1}$$

where the stress-levels $\sigma_m(\theta)$ and $\sigma_M(\theta)$ are given by (3.12). In order that the corresponding straight lines in the γ, θ-plane be arranged as shown in Fig. 1, it is necessary that $\hat{\gamma}_4(\theta) > \hat{\gamma}_3(\theta) > \hat{\gamma}_2(\theta) > \hat{\gamma}_1(\theta) > -1$ for $\theta_m < \theta < \theta_M$ and that $\hat{\gamma}_4(\theta) > \hat{\gamma}_1(\theta) > -1$ for $0 < \theta < \theta_m$. These inequalities can be expressed, upon using (A.1), as

$$\left.\begin{aligned} &0 < \sigma_M(\theta) < \sigma_m(\theta) + \mu\gamma_T < \mu + \mu\alpha(\theta-\theta_T) && \text{for } \theta_m < \theta < \theta_M, \\ &0 < \sigma_m(\theta) + \mu\gamma_T < \mu + \mu\alpha(\theta-\theta_T) && \text{for } 0 < \theta < \theta_m. \end{aligned}\right\} \tag{A.2}$$

Next, since we assumed in Section 3 that all three phases M^-, A and M^+ exist when $\sigma = 0$ and $\theta = \theta_T$, it is necessary that the corresponding strains $\gamma = -\gamma_T$, 0 and γ_T lie in the appropriate strain ranges as defined by Fig. 1. In view of (A.1) and (A.2), one finds that this holds if and only if

$$\gamma_T < 1, \quad \sigma_m(\theta_T) < 0. \tag{A.3}$$

We turn finally to the issue of extending the Helmholtz free-energy function (3.10) to the unshaded ('unstable') region of the (γ, θ)-plane shown in Fig. 1. Even though we do not need an expression for ψ on this region, it is still necessary to know that (3.10) can be extended to that region in the manner previously assumed [see paragraph below (3.12)]. The ability to do this is equivalent to the ability to connect each adjacent pair of rising branches

of the stress–strain curve in Fig. 2 by a declining branch *with prescribed area under it*. Since the stress–strain curve is to be declining for strains in the intervals $(\gamma_1(\theta), \gamma_2(\theta))$ and $(\gamma_3(\theta), \gamma_4(\theta))$ when $\theta_m < \theta < \theta_M$, and on $(\gamma_1(\theta), \gamma_4(\theta))$ when $0 < \theta < \theta_m$, it is necessary that

$$\sigma_M(\theta) > \sigma_m(\theta) \quad \text{for } \theta_m < \theta < \theta_M, \qquad -\sigma_m(\theta) > \sigma_m(\theta) \quad \text{for } 0 < \theta < \theta_m. \tag{A.4}$$

Next, as is readily seen from Fig. 2(a), for $\theta_m < \theta < \theta_M$, the area under the graph of $\hat{\sigma}(\cdot, \theta)$ between $\gamma = \hat{\gamma}_3(\theta)$ and $\gamma = \hat{\gamma}_4(\theta)$, must necessarily lie between the areas of the two rectangles with the same base $(\hat{\gamma}_3(\theta), \hat{\gamma}_4(\theta))$ and with heights $\sigma_M(\theta)$ and $\sigma_m(\theta)$. A similar restriction applies to the area between $\gamma = \hat{\gamma}_1(\theta)$ and $\gamma = \hat{\gamma}_2(\theta)$, and for $0 < \theta < \theta_m$ to the area between $\gamma = \hat{\gamma}_1(\theta)$ and $\gamma = \hat{\gamma}_4(\theta)$. Thus it is necessary that

$$\left.\begin{aligned} -\sigma_M(\theta)(\hat{\gamma}_2(\theta) - \hat{\gamma}_1(\theta)) < \rho\psi(\hat{\gamma}_2(\theta), \theta) - \rho\psi(\hat{\gamma}_1(\theta), \theta) < -\sigma_m(\theta)(\hat{\gamma}_2(\theta) - \hat{\gamma}_1(\theta)), \\ \sigma_m(\theta)(\hat{\gamma}_4(\theta) - \hat{\gamma}_3(\theta)) < \rho\psi(\hat{\gamma}_4(\theta), \theta) - \rho\psi(\hat{\gamma}_3(\theta), \theta) < \sigma_M(\theta)(\hat{\gamma}_4(\theta) - \hat{\gamma}_3(\theta)), \\ \sigma_m(\theta)(\hat{\gamma}_4(\theta) - \hat{\gamma}_1(\theta)) < \rho\psi(\hat{\gamma}_4(\theta), \theta) - \rho\psi(\hat{\gamma}_1(\theta), \theta) < -\sigma_m(\theta)(\hat{\gamma}_4(\theta) - \hat{\gamma}_1(\theta)), \end{aligned}\right\} \tag{A.5}$$

where the first two sets of inequalities in (A.5) hold for $\theta_m < \theta < \theta_M$, while the last set holds for $0 < \theta < \theta_m$. Conversely, given two points $(\hat{\gamma}_3(\theta), \sigma_M(\theta))$ and $(\hat{\gamma}_4(\theta), \sigma_m(\theta))$ in the (γ, σ)-plane, with $\hat{\gamma}_4(\theta) > \hat{\gamma}_3(\theta)$, a *sufficient* condition for the existence of a continuous decreasing function $\hat{\sigma}(\cdot, \theta)$ connecting these two points, which is such that the area under it is $\rho\psi(\hat{\gamma}_4(\theta), \theta) - \rho\psi(\hat{\gamma}_3(\theta), \theta)$, is that $(A.4)_1$ and $(A.5)_2$ hold. The requirements (A.4), (A.5) are therefore necessary and sufficient for the extendability of the Helmholtz free-energy function (3.10) to the unstable region.

The inequalities (A.5) can be expressed equivalently in terms of the stresses $\sigma_m(\theta)$, $\sigma_M(\theta)$ and $\sigma_o(\theta)$ as

$$\left.\begin{aligned} &[\sigma_M(\theta) - \sigma_m(\theta)]^2 < 2\mu\gamma_T[\sigma_o(\theta) - \sigma_m(\theta)] && \text{for } \theta_m < \theta < \theta_M, \\ &[\sigma_M(\theta) - \sigma_m(\theta)]^2 < 2\mu\gamma_T[\sigma_M(\theta) - \sigma_o(\theta)] && \text{for } \theta_m < \theta < \theta_M, \\ &-\mu\gamma_T < \sigma_m(\theta) < 0 && \text{for } 0 < \theta < \theta_m. \end{aligned}\right\} \tag{A.6}$$

The inequalities (A.2)–(A.4) and (A.6) must be enforced on the material model. They can be reduced by using (3.12), (3.13) into temperature independent inequalities that involve only the material parameters. We shall not display the resulting inequalities here. These inequalities, as well as (4.3), are to be imposed on the material constants entering into our model. One can verify that the particular values (5.6) of the material constants, together with a range of values of the four remaining parameters m, M, θ_m and θ_M, do satisfy these inequalities. For example, one possible set of values of the latter four parameters is $m = 9.7253 \times 10^{-5}/^{\circ}\text{K}$, $M = 10.1371 \times 10^{-5}/^{\circ}\text{K}$, $\theta_m = 285\ ^{\circ}\text{K}$, $\theta_M = 10\,000\ ^{\circ}\text{K}$; as mentioned previously, the particular values of these four material constants do *not* affect the response of the material.