INSTITUTE OF PHYSICS PUBLISHING SMART MATERIALS AND STRUCTURES
Smart Mater. Struct. **16** (2007) S51–S62 doi:10.1088/0964-1726/16/1/S06

A micromechanics inspired constitutive model for shape-memory alloys: the one-dimensional case

Amir Sadjadpour and Kaushik Bhattacharya[1]

Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, USA

E-mail: bhatta@caltech.edu

Received 16 July 2005, in final form 4 May 2006
Published 15 January 2007
Online at stacks.iop.org/SMS/16/S51

Abstract
This paper presents a constitutive model for shape-memory alloys that builds on ideas generated from recent micromechanical studies of the underlying microstructure. The presentation here is in one dimension. It is applicable in a wide temperature range that covers both the shape-memory effect and superelasticity, is valid for a wide range of strain rates and incorporates plasticity. The thermodynamic setting of the model is explained and the model is demonstrated through examples.

1. Introduction

Shape-memory alloys are widely used for a variety of applications including micro-actuators, cell phone antennas, energy absorption and biomedical devices. These materials exhibit a strongly nonlinear thermo-mechanical behavior associated with abrupt changes in their lattice structure called martensitic phase transformation. Two common manifestations of this phase transformation are the *shape-memory effect*, wherein an apparently plastic deformation sustained below some critical temperature is recovered on heating above it, and *superelasticity*, wherein significant deformations suffered under loading are recovered on unloading.

The applications of these materials and the need for a design tool have motivated a number of macroscopic constitutive models for these materials (see for example [28, 24, 18, 36, 14, 37, 30, 19, 12, 3, 4] and the references therein). While some of them have been adapted from other phenomena like plasticity, a number of them have tried to build in true micromechanical features. Amongst the latter, there is a whole range that balance simplicity against detail. For a model to be usable in the context of design, it has to be relatively simple, and it should be capable of being implemented in standard stress-analysis software. At the same time, it has to incorporate realistic physics. Indeed, each object and function in the model should in principle be computable from a lower-scale model, but possible to fit to empirical data. This would ensure that the model is widely applicable: it has to be applicable in a wide range of temperatures so that it captures both the shape-memory effect and superelasticity; it has to be adaptable to a wide range of materials and textures; it has to hold for a wide range of loading rates. Finally, since phase transformation often competes with plasticity in shape-memory alloys, the model should incorporate this phenomenon as well.

The last couple of decades has also brought about an increasingly sophisticated understanding of the microstructure of shape-memory alloys and its relation to macroscopic properties in both single and polycrystals (see for example [27, 8, 5, 9, 26, 35, 21, 22, 34, 13, 17, 20, 33, 2] and the references therein). In this paper, we present a constitutive model that builds on the concepts that have emerged from this analysis of microstructure.

The shape-memory effect and superelasticity are consequences of a martensitic phase transformation, which is a diffusionless first-order phase transformation between a high-temperature austenite phase and a low-temperature martensite phase. The crystallographic symmetry of the austenite is higher than that of the martensite, and consequently one can have a number of symmetry-related variants of martensite. The different variants of martensite, along with the austenite, form complex microstructures that can evolve with load and temperature. A key difficulty in the constitutive modeling of these materials is to find an effective means of describing this evolution, especially in polycrystals.

We address this issue in our model by introducing the idea of the *effective transformation strain*. It is the average

[1] Author to whom any correspondence should be addressed.

transformation strain of the different variants averaged over a representative volume containing multiple grains after the material has formed an allowable microstructure. It is allowed to take any value in the set of effective transformation strains, or set of effective recoverable strains. The micromechanical basis for this set can be found in Bhattacharya and Kohn [9] (also see [11]). It is also the convex dual of the transformation yield set introduced by Lexcellent and his coworkers [20]. This set depends on the material and the texture of the specimen, and can be easily fitted to experiment. In one dimension, it is the interval from the recoverable compressive to the recoverable tensile strain. A second important idea in this paper is the use of kinetic relations that cover a wide range of strain rates.

This paper develops the model in one dimension. We have generalized it to three dimensions elsewhere [31]. We have also adapted this model for the study of martensitic phase transformations in iron under dynamic loading [32]. Section 2 introduces the model and describes its thermodynamic setting. Section 3 demonstrates the model by studying its response to applied stress under a variety of conditions. Since load-controlled experiments are difficult, especially at high rates, the model is also studied for its response to Kolsky bar experiments in section 4. We conclude in section 5 with a brief discussion.

2. A thermo-mechanical model

In this section we develop and discuss our phenomenological constitutive model within a continuum thermodynamic framework.

2.1. Kinematics

We are interested in a model that can be used at the macroscopic scale for the design of devices and structures. So we take a multiscale point of view and think of each material point of our continuum as corresponding to a representative volume element (RVE) consisting of a number of grains, each containing a complex microstructure of austenite and variants of martensite. We introduce two kinematic or field or internal variables to represent the consequence of microstructure in an RVE: the *volume fraction of martensite*, $\lambda(x,t)$, and the *effective transformation strain of the martensite*, $\varepsilon_{\mathrm{m}}(x,t)$.

$\lambda(x,t)$ denotes the net or average volume fraction of the martensite; i.e., this would be the value we would obtain if we were to visit each grain in the RVE corresponding to the material point x at time t, add up the volume of all variants of martensite (self-accommodating, internally twinned, detwinned etc) and divide by the total volume of the RVE. To be precise, let χ^{ij} denote the characteristic function of the jth correspondence variant of martensite in the ith grain of the RVE. This function is equal to one at all positions occupied by the jth variant in the i grain and is equal to zero otherwise. Then $\chi^i = \sum_{j=1}^{N} \chi^{ij}$ is the characteristic function of martensite in the ith grain where N, the number of variants, is given by the crystallography of the transformation. We define the volume fraction as

$$\lambda = \langle \chi^i \rangle = \left\langle \sum_{j=1}^{N} \chi^{ij} \right\rangle \tag{1}$$

where $\langle\cdot\rangle$ denotes the mean or expected value over all grains in the RVE, i.e. over all i. λ is constrained to lie between zero and one, with zero signifying that the entire RVE is in the austenite phase and one signifying that the entire element is in the martensite phase.

Since λ cannot differentiate between the different microstructures of martensite like self-accommodating, internally twinned, detwinned etc, we introduce the second internal variable $\varepsilon_{\mathrm{m}}(x,t)$. This is the strain we would obtain if we were to visit every grain in the RVE corresponding to the material point x at time t and average over the transformation or stress-free strain of all the variants of martensite. To be precise, let $\varepsilon_{\mathrm{m}}^{ij}$ denote the transformation or stress-free strain of the jth variant of martensite in the ith grain in the RVE. It is given by

$$\varepsilon_{\mathrm{m}}^{ij} = R_i^T \varepsilon_{\mathrm{m}}^{j} R_i \tag{2}$$

where $\varepsilon_{\mathrm{m}}^{j}$ is the transformation strain of the jth correspondence variant in a reference crystal and is given by crystallography of the transformation, while R_i is the rotation matrix that gives the orientation of the ith grain and is given by the texture of the material. We define the effective transformation strain as

$$\varepsilon_{\mathrm{m}} = \left\langle \sum_{j=1}^{N} \chi^{ij} \varepsilon_{\mathrm{m}}^{ij} \right\rangle. \tag{3}$$

As the material forms different microstructures, the arrangement of variants and thus χ^{ij} changes, and $\varepsilon_{\mathrm{m}}(x,t)$ takes different values. We are only interested in compatible microstructures, and thus we cannot form any arbitrary mixture of variants. Consequently, ε_{m} cannot take any arbitrary average value, but is restricted to those that are obtained from compatible arrangements. We denote the set of all possible values of ε_{m} as the *set of effective transformation strains* or the *set of effective recoverable strains*, P. This set P depends on the crystallography of the material and the texture of the specimen. In a single crystal, P is the set of all possible average transformation strains associated with compatible microstructures of the different variants of martensite. In a polycrystal, the set P is the the macroscopic averages of locally varying strain fields which can be accommodated within each grain by a compatible arrangement of the martensite variants. One can calculate this set in various examples of interest and estimate them in others [9, 8, 11]. Alternately, one can use experimental measurements of recoverable strain to fit this set.

Note that we do not track the individual volume fractions of the different volumes of martensite. This is too difficult, especially in a polycrystal, where the different grains behave differently depending on orientation, inter-granular constraints and long-range cooperative effects. However, we implicitly account for these effects by tracking the effective transformation strain and confining it to the set P, which depends on material and texture. In other words, the set P incorporates information about the material crystallography, specimen texture and also intergranular constraints.

The set P can be quite complicated in multiple dimensions, but it is relatively simple in one dimension. It is an interval $[\varepsilon_{\mathrm{m}}^{\mathrm{c}}, \varepsilon_{\mathrm{m}}^{\mathrm{t}}]$ where $\varepsilon_{\mathrm{m}}^{\mathrm{c}} < 0$ denotes the largest recoverable compressive strain and $\varepsilon_{\mathrm{m}}^{\mathrm{t}} > 0$ denotes the largest recoverable tensile strain.

In summary, the consequences of the microstructure at a material point are specified by two internal variables λ and ε_{m} which are subject to the following constraints.

$$\lambda \in [0, 1] \qquad \text{and} \qquad \varepsilon_{\mathrm{m}} \in P = [\varepsilon_{\mathrm{m}}^{\mathrm{c}}, \varepsilon_{\mathrm{m}}^{\mathrm{t}}]. \tag{4}$$

Finally, we introduce the plastic strain ε_{p} as an additional field variable. Putting everything together, we say that the total strain can be divided into three parts, elastic, transformation and plastic:

$$\varepsilon(x,t) = \frac{\partial u}{\partial x} = \varepsilon_e(x,t) + \lambda(x,t)\varepsilon_{\mathrm{m}}(x,t) + \varepsilon_{\mathrm{p}}(x,t). \tag{5}$$

It is worth noting that the effective transformation strain of the RVE is $\lambda\varepsilon_{\mathrm{m}}$ since λ is the volume fraction of martensite, ε_{m} is the effective transformation strain of the martensite and the transformation strain of the austenite is zero by choice of reference configuration.

2.2. Balance laws

We assume that the usual balance laws hold. The first is the balance of linear momentum

$$\rho u_{tt} = \sigma_x \tag{6}$$

where ρ is the (referential) mass density and σ is the (Piola–Kirchhoff) stress.

Second, we assume the balance of energy. Writing the balance of energy for any part of the body, localizing and using (6), we obtain

$$\dot{\epsilon} = -q_x + r + \sigma\dot{\varepsilon}. \tag{7}$$

Above, ϵ denotes the internal energy density, q the heat flux and r the radiative heating.

Finally, we write the second law of thermodynamics or the Clausius–Duhem inequality, again in local form:

$$\dot{\eta} \geqslant \left(-\frac{q}{\theta}\right)_x + \frac{r}{\theta} \Rightarrow \theta\dot{\eta} \geqslant -q_x + \frac{q\theta_x}{\theta} + r. \tag{8}$$

Here, η is the entropy density and θ the (absolute) temperature.

Using (7) in the second law (8), we obtain

$$-\dot{\epsilon} + \theta\dot{\eta} + \sigma\dot{\varepsilon} - \frac{q\theta_x}{\theta} \geqslant 0. \tag{9}$$

It is now convenient to introduce Helmholtz free energy density $W = \epsilon - \theta\eta$ and rewrite the second law in the following form:

$$-\dot{W} - \eta\dot{\theta} + \sigma\dot{\varepsilon} - \frac{q\theta_x}{\theta} \geqslant 0. \tag{10}$$

2.3. Constitutive relations, driving forces and kinetic relations

We assume that the Helmholtz free energy density depends on the strain, the temperature and the internal variables:

$$W = W(\varepsilon, \lambda, \varepsilon_{\mathrm{m}}, \varepsilon_{\mathrm{p}}, \theta). \tag{11}$$

We make similar assumptions on the stress. Substituting these in the second law, (10), we obtain

$$-\left(\frac{\partial W}{\partial \varepsilon} - \sigma\right)\dot{\varepsilon} - \frac{\partial W}{\partial \lambda}\dot{\lambda} - \frac{\partial W}{\partial \varepsilon_{\mathrm{m}}}\dot{\varepsilon}_{\mathrm{m}} - \frac{\partial W}{\partial \varepsilon_{\mathrm{p}}}\dot{\varepsilon}_{\mathrm{p}} - \left(\frac{\partial W}{\partial \theta} + \eta\right)\dot{\theta} - \frac{q\theta_x}{\theta} \geqslant 0. \tag{12}$$

Using arguments similar to those of Coleman and Noll [15], we conclude that

$$\sigma = \frac{\partial W}{\partial \varepsilon}, \qquad \eta = -\frac{\partial W}{\partial \theta}. \tag{13}$$

We also assume Fourier's law of heat transfer

$$q = -k\theta_x \tag{14}$$

where $k > 0$ is the conductivity.

We define the *driving forces* associated with the internal variables to be the quantities conjugate to their rates of change in (12):

$$d_\lambda := -\frac{\partial W}{\partial \lambda}, \qquad d_{\varepsilon_{\mathrm{m}}} := -\frac{\partial W}{\partial \varepsilon_{\mathrm{m}}}, \qquad d_{\varepsilon_{\mathrm{p}}} := -\frac{\partial W}{\partial \varepsilon_{\mathrm{p}}}. \tag{15}$$

Substituting these back into (12), and using (13) and (14), we conclude that the second law reduces to the requirement that

$$d_\lambda\dot{\lambda} + d_{\varepsilon_{\mathrm{m}}}\dot{\varepsilon}_{\mathrm{m}} + d_{\varepsilon_{\mathrm{p}}}\dot{\varepsilon}_{\mathrm{p}} \geqslant 0. \tag{16}$$

We have to prescribe the evolution of the internal variables to be consistent with this relation.

We assume that the evolution of the internal variables $\lambda, \varepsilon_{\mathrm{m}}$ depends on the driving forces through the following *kinetic relations*, and subject to the constraints (4):

$$\dot{\lambda} = K_\lambda(d_\lambda, \lambda, \varepsilon_{\mathrm{m}}) \qquad \lambda \in [0, 1], \tag{17}$$

$$\dot{\varepsilon}_{\mathrm{m}} = K_{\varepsilon_{\mathrm{m}}}(d_{\varepsilon_{\mathrm{m}}}, \lambda, \varepsilon_{\mathrm{m}}) \qquad \varepsilon_{\mathrm{m}} \in [\varepsilon_{\mathrm{m}}^{\mathrm{c}}, \varepsilon_{\mathrm{m}}^{\mathrm{t}}]. \tag{18}$$

Finally, we assume that the evolution of the internal variable ε_{p} is prescribed as in the rate-independent theory of plasticity. We postpone its discussion till the next section.

2.4. Specific constitutive assumptions

We specialize to the following constitutive relation for the Helmholtz energy:

$$W = \frac{E}{2}(\varepsilon - \varepsilon_{\mathrm{p}} - \lambda\varepsilon_{\mathrm{m}})^2 + \lambda\omega(\theta) - c_{\mathrm{p}}\theta\ln\left(\frac{\theta}{\theta_0}\right) \tag{19}$$

where E is the elastic modulus (assumed to be equal in both the austenite and the martensite), ω is the difference in chemical energy between the austenite and the martensite, c_{p} is the heat capacity (assumed to be equal in both the austenite and the martensite), and ordinary thermal expansion is neglected. This relation is illustrated in figure 1. We further assume that

$$\omega(\theta) = \frac{\mathcal{L}}{\theta_{\mathrm{cr}}}(\theta - \theta_{\mathrm{cr}}) \tag{20}$$

where $\mathcal{L}$ is the latent heat of transformation and θ_{cr} is the thermodynamic transformation temperature. Substituting (19) and (20) in (13), we obtain

$$\sigma = E(\varepsilon - \varepsilon_{\mathrm{p}} - \lambda\varepsilon_{\mathrm{m}}), \tag{21}$$

$$\eta = \lambda\frac{\mathcal{L}}{\theta_{\mathrm{cr}}} - c_{\mathrm{p}}\left(1 + \ln\left(\frac{\theta}{\theta_0}\right)\right), \tag{22}$$

$$d_\lambda = \sigma\varepsilon_{\mathrm{m}} - \omega, \tag{23}$$

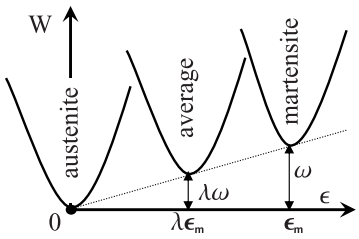

Figure 1. Schematic representation of the Helmholtz energy density.

$$d_{\varepsilon_m} = \lambda\sigma, \tag{24}$$

$$d_{\varepsilon_p} = \sigma. \tag{25}$$

The kinetic relation describing the evolution of the martensite volume fraction λ is taken to be the following:

$$\dot\lambda = \begin{cases} \dot\lambda^+(1+(d_\lambda - d_\lambda^+)^{-1})^{-\frac{1}{p}} & d_\lambda > d_\lambda^+ \text{ and } \lambda < 1, \\ \dot\lambda^-(1+(d_\lambda^- - d_\lambda)^{-1})^{-\frac{1}{p}} & d_\lambda < d_\lambda^- \text{ and } \lambda > 0, \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

where $\dot\lambda^\pm$, $d_\lambda^\pm$, p are material parameters. This relation is shown in figure 2. Note that the kinetic relation is characterized by a 'stick–slip' character at small driving forces. The phase transformation requires a critical driving force before it can proceed; i.e., the rate of change of volume fraction is zero for driving forces below a critical driving force. As one exceeds the critical driving force, note that the curve is vertical, meaning that the rate of phase transformation is indeterminate or equivalently the driving force is independent of rate of phase transformation. This is rate-independent behavior. The reason for this is a combination of metastability [6] and pinning by defects [1, 7, 10]. However, at large driving forces, it becomes rate dependent and in fact asymptotes to a limiting rate. The reason for this is that phase boundaries require an unboundedly increasing driving force for the propagation speeds to reach towards some sound speed [10, 29]. Note that this kinetic relation is consistent with the experimental observations, which say that for low driving forces, around $d_\lambda^\pm$, phase transformation is rate independent [27], and at large driving forces the material shows rate dependency [25].

The evolution of the effective transformation strain ε_m describes the twinning, detwinning and other such processes that convert one martensitic variant to another. We assume a rather simple law for its evolution:

$$\dot\varepsilon_m = K_{\varepsilon_m}(d_{\varepsilon_m}, \lambda, \varepsilon_m) = \frac{\alpha}{\lambda} d_{\varepsilon_m} = \begin{cases} \alpha\sigma & \varepsilon_m \in [\varepsilon_m^c, \varepsilon_m^t] \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

where α is a material parameter and is chosen large enough to guarantee a very quick process, so that ε_m is essentially equal to ε_m^t and ε_m^c under tension and compression respectively.

Finally, we assume a rate-independent plasticity relation that neglects the Bauschinger effect:

$$\dot\varepsilon_p = K_{\varepsilon_p}(d_{\varepsilon_p}, \sigma_y) = \frac{\dot d_{\varepsilon_p}}{H} = \begin{cases} \dfrac{\dot\sigma}{H} & \sigma \geqslant \sigma_y \text{ or } \sigma \leqslant -\sigma_y, \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

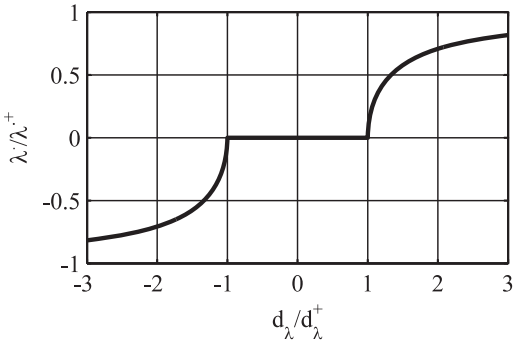

Figure 2. The kinetic relation between $\dot\lambda$ and the driving force d_λ.

where H is the hardening parameter and σ_y is the plastic yield stress.

This completes the specification of the model.

2.5. Temperature evolution

The energy balance along with the constitutive relations describes the evolution of the temperature. However, this is rather complicated, and therefore it is useful to make some simplifying assumptions.

We begin by substituting for the internal energy in terms of the Helmholtz free energy and entropy in the energy balance (7) to rewrite it as

$$\dot W + \theta\dot\eta + \dot\theta\eta = -q_x + r + \sigma\dot\varepsilon. \tag{29}$$

Using the constitutive assumption (19) for W to expand $\dot W$, using the various definitions of driving force and simplifying, we obtain

$$\theta\dot\eta = -q_x + r + d_\lambda\dot\lambda + d_{\varepsilon_m}\dot\varepsilon_m + d_{\varepsilon_p}\dot\varepsilon_p. \tag{30}$$

Specializing now to the specific constitutive relation and in particular (22), we obtain

$$c_p\dot\theta = \dot\lambda\theta\frac{\mathcal{L}}{\theta_{cr}} - q_x + r + d_\lambda\dot\lambda + d_{\varepsilon_m}\dot\varepsilon_m + d_{\varepsilon_p}\dot\varepsilon_p. \tag{31}$$

In the following, we shall be interested in processes where we can assume adiabatic conditions and neglect heat transfer ($q = r = 0$). Further, it turns out that the latent heat of transformation is large compared to the energy dissipated during transformation, martensitic variant reorientation and plasticity during typical processes of interest. Therefore, we assume

$$c_p\dot\theta = \theta\dot\lambda\frac{\mathcal{L}}{\theta_{cr}}. \tag{32}$$

Integrating this, we obtain a relation between temperature, volume fraction of martensite, latent heat and specific heat:

$$\theta(t) = \theta_0 \exp\left(\frac{(\lambda(t)-\lambda_0)\mathcal{L}}{c_p\theta_{cr}}\right). \tag{33}$$

3. Demonstration

We now demonstrate the model by calculating the response of a material point to a given applied stress history, and then conduct a parameter study.

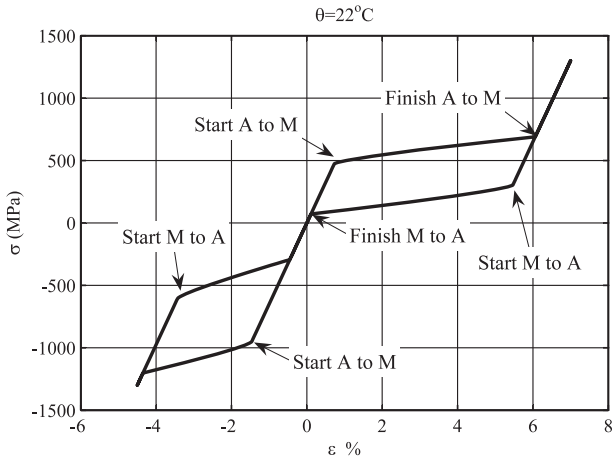

Figure 3. A typical stress–strain curve generated by the model.

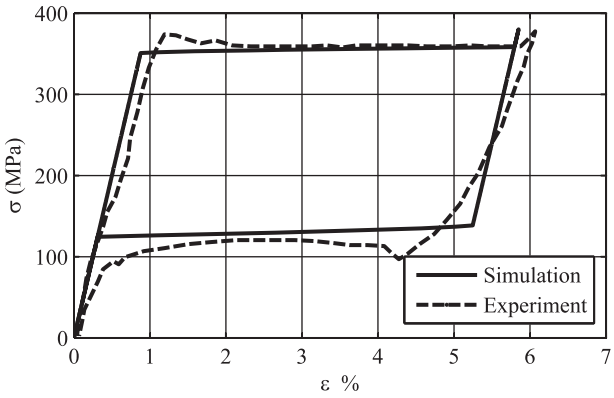

Figure 4. Comparison between the experiment [23] and the fit to the model.

3.1. Parameters

Consistent with typical experiments on NiTi (see for example [23]), we consider the following parameters:

$$\begin{gathered} M_s = -51.55\,^\circ\mathrm{C} \qquad \text{and} \qquad A_s = -6.36\,^\circ\mathrm{C} \\ \mathcal{L} = 79\ \mathrm{MJ\ m^{-3}} \qquad \text{and} \qquad c_p = 5.4\ \mathrm{MJ\ m^{-3}\ K^{-1}} \\ \varepsilon_m^c = -2.5\% \qquad \text{and} \qquad \varepsilon_m^t = 5\% \\ E = 65\ \mathrm{GPa} \qquad \text{and} \qquad \sigma_y = 1500\ \mathrm{MPa} \end{gathered} \tag{34}$$

where M_s and A_s are the martensite start and austenite start temperatures respectively[2]. Recalling (26) and (23), we obtain

$$d_\lambda^+ = -\omega(M_s), \qquad d_\lambda^- = -\omega(A_s). \tag{35}$$

Assuming further that

$$d_\lambda^+ = -d_\lambda^-, \tag{36}$$

we conclude that

$$d_\lambda^+ = -d_\lambda^- = \mathcal{L}\left(\frac{A_s - M_s}{A_s + M_s}\right), \qquad \theta_{cr} = \frac{A_s + M_s}{2}. \tag{37}$$

Note that one has to be careful to specify the absolute temperature in kelvin to use these formulae.

We also assume the following kinetic coefficients:

$$\dot\lambda^+ = -\dot\lambda^- = 10^4, \qquad \alpha = 1, \qquad p = 2$$

$$\text{and} \qquad H = \frac{E}{50}.$$

The parameters $\dot\lambda^\pm$ that control the evolution rate of the volume fraction of martensite are kept fixed in the entire paper and chosen so that $\dot\lambda\,\varepsilon_m$ is smaller than $\dot\varepsilon$.

Parameter α is chosen to guarantee the fast speed of the evolution of phase transformation strain ε_m both under tension and compression as described earlier. While larger α values do not change the results, smaller α values may lead to slow evolution of ε_m, which is not acceptable in this setting.

The power of the kinetic law p controls the shape of its function; for the given value of $p = 2$ its form is shown in figure 2. For higher values the kinetic relation would get closer to a Heaviside function in the first and third quadrants.

3.2. Demonstration

We consider an applied stress of the form

$$\sigma(t) = A \sin \omega t \tag{38}$$

with $A = 1300$ MPa and $\omega = 2\pi/T$ with $T = 5 \times 10^{-3}$ s unless otherwise mentioned. In particular, this value of loading is below the chosen yield strength, and thus plasticity is suppressed. We examine this aspect in section 3.5. We integrate (26)–(28), (33) as well as (21) simultaneously subject to the initial conditions

$$\begin{gathered} \varepsilon(0) = 0.0\%, \qquad \varepsilon_p(0) = 0.0\%, \qquad \varepsilon_m(0) = 0.0\%, \\ \lambda(0) = 0 \qquad \text{and} \qquad \theta(0) = 22\,^\circ\mathrm{C} \end{gathered} \tag{39}$$

to obtain the time trajectory of ε, ε_m, ε_p, λ and θ.

The result is shown in figure 3. As we start loading, the material deforms elastically till it reaches the point at the top left corner of the upper *flag*. At this point the austenite to martensite phase transformation starts and this results in a change of slope. It continues till the material fully transforms to martensite and we reach the top right corner of the upper flag. The material deforms elastically beyond this. As material is unloaded, it deforms back elastically till the driving force is equal to d_λ^- at the bottom right corner of the upper flag. The reverse phase transformation starts as we traverse the lower side of the upper flag. The reverse phase transformation continues till material transforms back to austenite completely at the bottom left corner of the upper flag. Material undergoes the same process under compression; however, the bottom flag has a different shape, consistent with the well known asymmetry of shape-memory alloys.

Figure 4 compares the stress versus strain relation obtained from experiments by McNaney *et al* [23] with that

[2] M_s is the temperature at which the specimen completely in the austenite will begin to transform to martensite during cooling and A_s is the temperature at which the specimen completely in the martensite will begin to transform to austenite during heating. Though stated in celsius, they have to be converted to kelvin for use in the model.

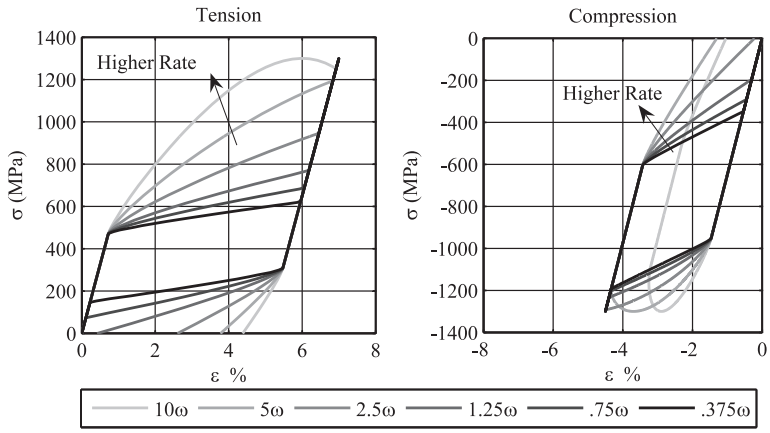

Figure 5. Tensile and compressive half-cycles at different loading rates, $\omega \mapsto 0.375\omega, 0.75\omega, 1.25\omega, 2.5\omega, 5\omega, 10\omega$.

generated by our model[3]. It is clear that the model is able to reproduce the overall features of the experiment. However, we see some differences. First, notice that the unloading begins with a greater slope in our model compared to the experiment. This is because our model assigns the same elastic moduli for both the austenite and the martensite while the experimental values are different. This may be easily changed. Second, the experiment shows a slight overshoot in both loading and unloading, but the model does not. This is related to localization and beyond the scope of the model.

3.3. Loading rate

Figure 5 shows the results of both the tensile and compressive half cycles of loading at different loading rates. Each curve was generated by starting with the initial conditions (39), and carrying through the calculation for a tensile or compressive half-cycle of loading. As stress rate increases, the transformation begins at the same level of stress, but the stress increases faster than the strain can evolve, giving rise to an apparent hardening and increase in the size of the hysteresis loop. This hardening is a result of both the particular kinetic relation (specifically the rate dependence at high rates) and also an increase of temperature of the material. We shall examine the two effects separately below. At higher rates we also notice two other aspects. First, there is an apparent residual strain on full unloading. The reason is that the stress unloads too fast to allow for the completion of the reverse transformation. This is a consequence of the rate dependence of the chosen kinetic relation at high rates. Second, at the highest rate (10ω), the material appears to soften as the loading proceeds. The reason is that the loading rate is so high that the transformation is not complete as the load increases to its peak and thus the transformation continues even as the load begins to decrease. All of these are consistent with observations in the literature [25].

Figure 6 shows the results of a full tension–compression cycle starting with tension.

We conclude this section by examining the relative role of the two factors that lead to the apparent hardening at high rate. Figure 7 shows the stress–strain response of the rate-independent (gray) and rate-dependent (black) kinetic relations. The kinetic relations are shown on the right. The rate-dependent kinetic relation is the one stated earlier in (26). The rate-independent kinetic relation may be described as follows:

$$|d_\lambda| \leqslant d_\lambda^+, \qquad \dot{\lambda} = 0 \text{ if } |d_\lambda| < d_\lambda^+, \tag{40}$$

and is shown as a gray line on the right of figure 7.

Let us begin with the rate-independent response. If the loading is quasistatic and in fact slow enough to dissipate the latent heat so that the response is isothermal, one obtains the stress–strain response shown as a gray solid line and marked 'Isothermal'. As the loading rate increases, there is less time to dissipate the latent heat and the temperature rises as the transformation progresses. Consequently, the stress required to sustain the transformation increases with increasing volume fraction, causing an apparent hardening. We eventually reach the response shown in a gray dashed line and marked 'Adiabatic'. Further increase in loading rate does not change the response.

We can estimate the amount of hardening for this rate-independent situation. By combining the definition of driving force (23) with the constitutive assumption (20) for ω, we

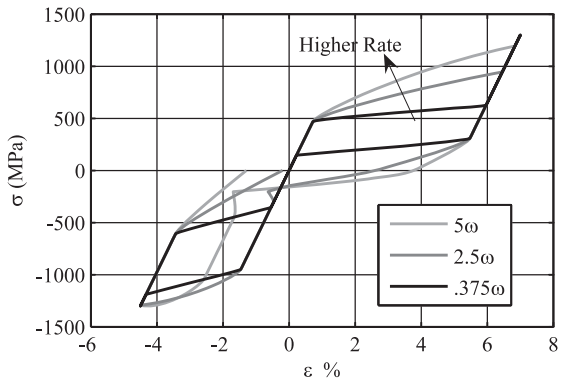

Figure 6. Complete tension–compression cycles at different loading rates, $\omega \mapsto 0.375\omega, 2.5\omega, 5\omega$.

[3] The parameters used were as described earlier in (34) except $\mathcal{L} = 8.8$ J g^{-1} and $E = 40$ GPa.

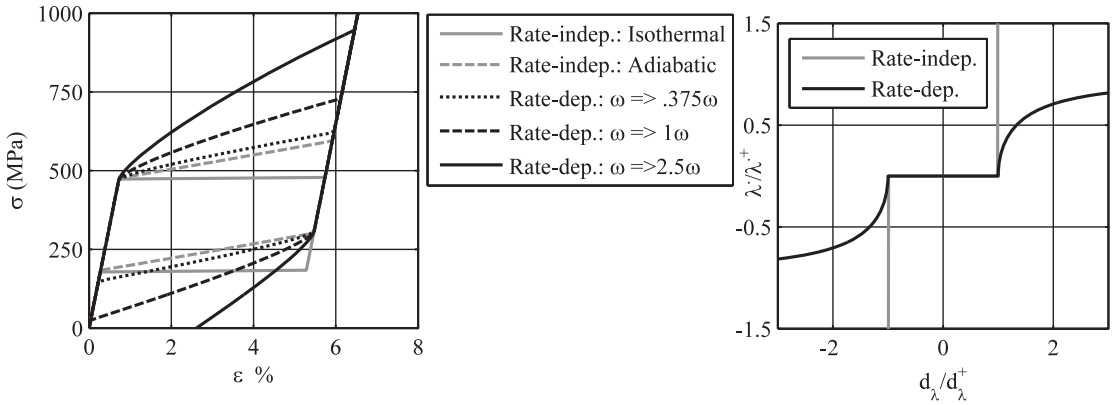

Figure 7. Comparison of the stress–strain relations for rate-independent (gray) and rate-dependent transformation (black) kinetics.

obtain

$$d_\lambda = \sigma\varepsilon_{\rm m} - \frac{\mathcal{L}}{\theta_{\rm cr}}(\theta - \theta_{\rm cr}). \tag{41}$$

For rate-independent kinetics, (40), the driving force $d_\lambda = d_\lambda^+$ for transformation to proceed. Thus, the stress required for the transformation to proceed is given by

$$\sigma = \frac{d_\lambda^+}{\varepsilon_{\rm m}} + \frac{\mathcal{L}}{\varepsilon_{\rm m}\theta_{\rm cr}}(\theta - \theta_{\rm cr}). \tag{42}$$

At the start of the transformation and for isothermal conditions, $\theta = \theta_0$. However, under adiabatic conditions, the temperature rises with transformation according to (33), so that final temperature at the completion of transformation is given by

$$\theta_f = \theta_0 \exp\left(\frac{\mathcal{L}}{c_{\rm p}\theta_{\rm cr}}\right). \tag{43}$$

Thus the difference in stress between the start and finish of transformation during rate-independent adiabatic conditions[4] is given by

$$\Delta\sigma = \frac{\theta_0\mathcal{L}}{\varepsilon_{\rm m}\theta_{\rm cr}}\left(\exp\left(\frac{\mathcal{L}}{c_{\rm p}\theta_{\rm cr}}\right) - 1\right). \tag{44}$$

We now turn to the rate-dependent kinetic relation (26). The response is black in figure 7. Under quasistatic, isothermal loading the response is identical to the rate-independent case and thus not shown. Since this kinetic relation is rate independent for small rates, we first see a transition from isothermal to adiabatic, as before. This is also not shown. On increasing the rate even further, we access the regime at which the kinetic relation becomes rate dependent and thus observe further hardening. We first see this at 0.375ω, as shown in figure 7. Further increases in rate cause further hardening.

In conclusion, both the transition from isothermal to adiabatic conditions as well as inherent rate-dependence can give rise to hardening of the response with increasing rate. However, the former is limited to an increase of stress as described in (44). Any further hardening is necessarily a manifestation of inherent rate-dependence of the transformation.

[4] This is also equal to the difference in stress at the finish between adiabatic and isothermal conditions.

3.4. Ambient temperature, superelasticity and shape-memory effect

Figure 8 shows the effect of change of the ambient temperature in the stress–strain hysteresis. The initial temperature is taken to be the ambient temperature (instead of that in (39)), and subsequently allowed to evolve according to (33).

At 200 K (−73 °C), the lowest ambient temperature shown in figure 8, λ quickly increases to unity as the austenite transforms to martensite and remains there independent of loading. So the stress–strain curve reflects the evolution of $\varepsilon_{\rm m}$, or reorientation of martensitic variants. This changes as the ambient temperature increases through $M_{\rm s}$ and $A_{\rm s}$, till at 300 K (27 °C) one is completely in the austenite at zero stress independent of the loading history. So this stress–strain curve reflects the stress-induced transformation. As the ambient temperature increases, stress required to induce the transformation increases (faster in compression than in tension since the transformation strains are different), till no transformation is observed at 440 K (167 °C). All of this is consistent with observations and the well known Clausius–Clapeyron relation [27].

These plots show that this model captures both the shape-memory effect and the superelasticity. The latter is observed at 300 K and beyond. To observe the former, suppose we cool the specimen with no stress. As remarked above, the material transforms to martensite as λ increases to unity as we pass below $M_{\rm s}$ to say 200 K. However, the transformation strain $\varepsilon_{\rm m}$ remains at zero, and it follows from (21) that the strain ε remains at zero. Thus, transformation induced by cooling produces no strain, something called self-accommodation, and this is captured by the model. Now, deform the specimen. It is clear from the stress–strain curve at 200 K that unloading will cause a residual strain or permanent deformation. Finally heat the specimen to above A_f to say 300 K. Note from the stress–strain curve that the only strain consistent with zero stress is zero, and thus the specimen recovers its permanent deformation. This is the shape-memory effect.

3.5. Yield strength

The results discussed so far do not consider any plasticity, since the yield strength was chosen to be higher than the maximum applied load. We now change this by taking the

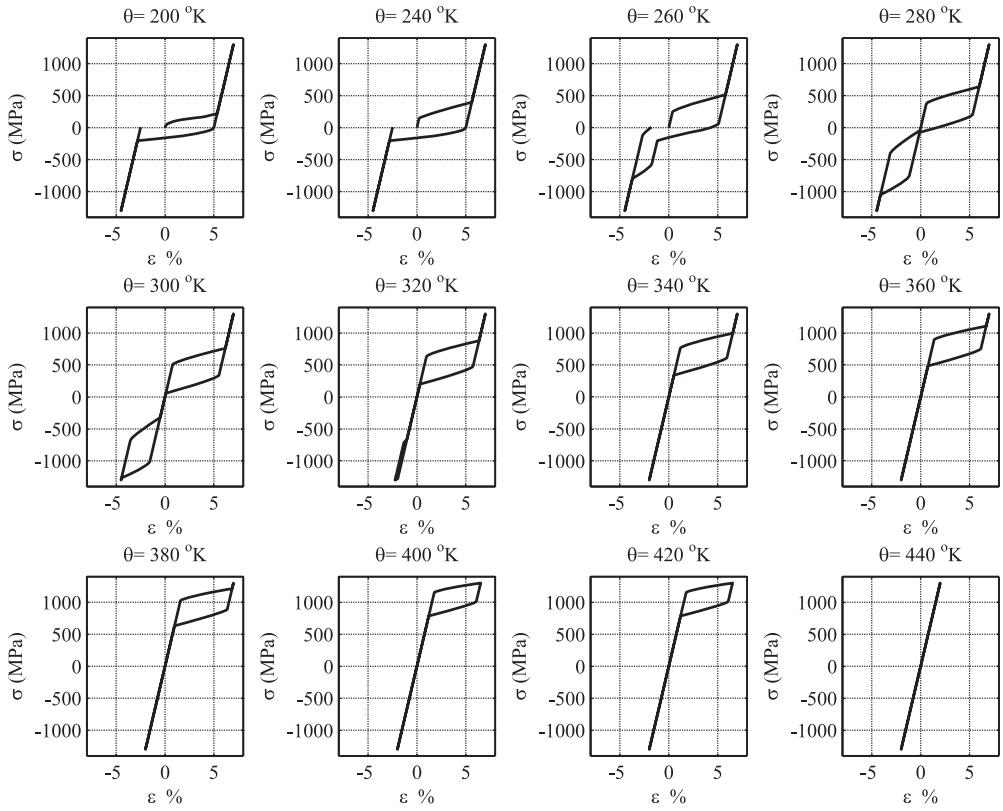

Figure 8. The strain–strain behavior at different ambient temperatures.

amplitude of loading to 800 MPa and then taking a range of yield strengths from 350 to 850 MPa. The results are shown in figure 9. There is no plasticity as before when the yield strength is above stress amplitude ($\sigma_y = 850$ MPa). One begins to observe plasticity as the yield strength decreases. At $\sigma_y = 750, 650$ MPa, note that the yielding does not begin till the transformation is complete. The unloading is similar as before, though offset by the residual plastic strain, and this remains as a permanent strain even after full unloading. At $\sigma_y = 550$ MPa, the transformation begins at 500 MPa; the plasticity begins at 550 MPa (indicated by an arrow) even before the transformation is complete, and proceeds even after the transformation is complete (again indicated by an arrow). At $\sigma_y = 350$ MPa, the yield begins at 350 MPa, and continues as the transformation begins at 500 MPa (indicated by an arrow) and is complete (again indicated by an arrow). The unloading is as before.

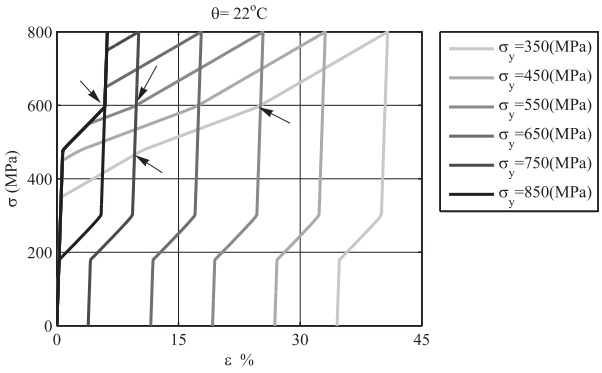

Figure 9. The stress–strain relations at different yield strengths.

3.6. Internal loops

Inspired by some experiments, we discuss internal loops of the stress–strain hysteresis for a simple triangular applied stress function. In figure 10 the applied stress function σ is shown at the top left corner, the strain ε evolution at the top right, the evolution of the martensitic volume fraction λ at the bottom left, and the stress strain curves at the bottom right. The stress–strain curves consist of two parallel, linearly elastic branches and two almost horizontal lines where forward and reverse phase transformation happens. In particular, for the cases in which loading is interrupted before material is fully transformed to martensite, material starts unloading along a path parallel to the elastic loading–unloading branches and then starts the reverse phase transformation and goes back to the initial austenite phase at an almost constant stress level. This is consistent with the observations and arguments of Abeyaratne *et al* [1] but not with those of Huo and Müller [16].

We also note that some of the curves appear to show a softening. This is a consequence of the triangular applied load. The ramp-down begins before the transformation is complete, so during the initial ramp-down the load is decreasing but still high enough for the transformation to continue, giving rise to an apparent softening.

4. Kolsky bar

The previous section demonstrated the model under stress control. However, this is difficult to attain experimentally. In

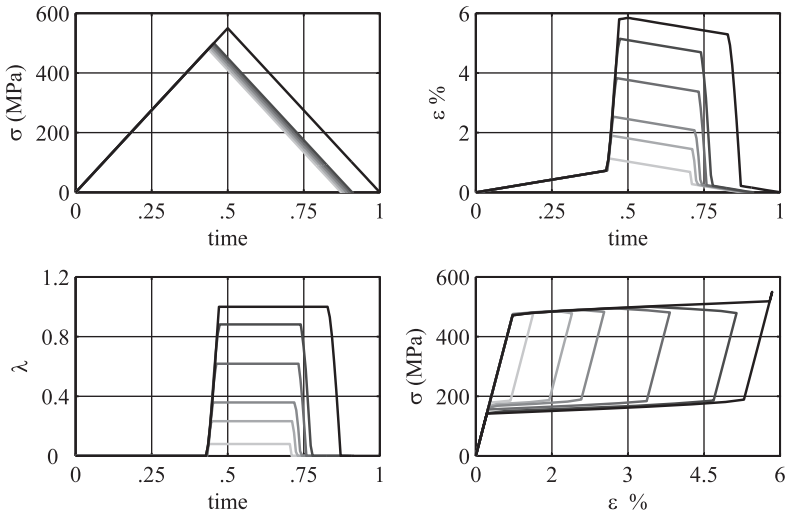

Figure 10. The internal hysteresis loops.

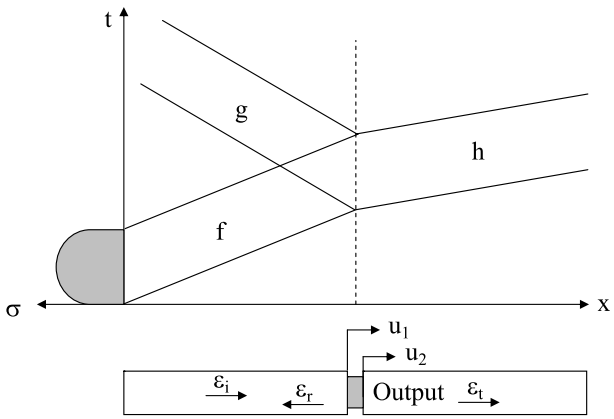

Figure 11. Kolsky bar and time–space diagram.

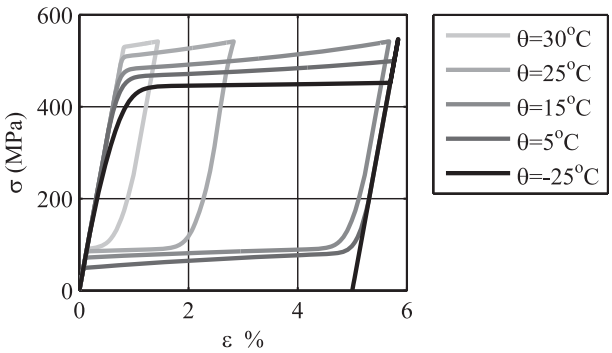

Figure 12. The stress–strain curve in a Kolsky bar for different ambient temperatures.

materials like shape-memory alloys that involve the evolution of internal variables, the stress–strain curve varies with the methodology of the experiment. A particularly popular means of measuring material properties at high deformation rates is the Kolsky or split-Hopkinson bar. So we consider it in this section. It also serves to emphasize how sensitively the stress–strain curve can depend on the experimental methodology.

4.1. Kolsky bar

In the Kolsky or split-Hopkinson bar experiment shown in figure 11, a thin specimen is placed between an incident bar (left) and output bar (right), both made of a linear elastic material and designed to have very little dispersion. A compression stress wave of known amplitude, duration and shape is generated in the incident bar through a striker bar (not shown). As this wave reaches the specimen, a portion is reflected while another portion is transmitted into the output bar. The length of the specimen is very small compared to the ratio of the wave speed to the duration of the pulse so that one may assume that the specimen is in equilibrium at any given time.

Since the input bar is linear elastic, the displacement in the bar is of the form

$$u_L(x,t) = f(x-ct) + g(x+ct) \tag{45}$$

so that the strain, the stress and the particle velocity are given by

$$\varepsilon_L(x,t) = (u_L)_{,x}(x,t) = f'(x-ct) + g'(x+ct) \tag{46}$$

$$\sigma_L(x,t) = E(f'(x-ct) + g'(x+ct)) \tag{47}$$

$$v_L(x,t) = (u_L)_{,t}(x,t) = c(-f'(x-ct) + g'(x+ct)) \tag{48}$$

that includes both the incident and reflected waves. Above, the prime denotes the differential with respect to its native variable. Similarly, the output bar is linear elastic and

$$u_R(x,t)_{\text{left}} = h(x-ct) \tag{49}$$

$$\varepsilon_R(x,t) = h'(x-ct) \tag{50}$$

$$\sigma_R(x,t) = Eh'(x-ct) \tag{51}$$

$$v_R(x,t) = -ch'(x-ct) \tag{52}$$

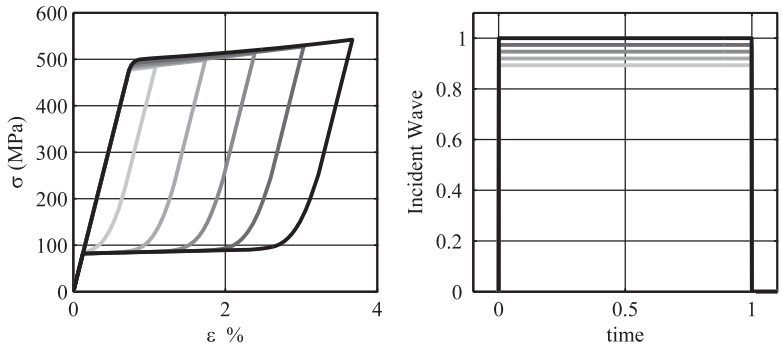

Figure 13. The stress–strain response of a specimen in a Kolsky bar (left) subjected to pulses of varying amplitude (right). The amplitude and pulse duration are normalized with those described in section 4.1.

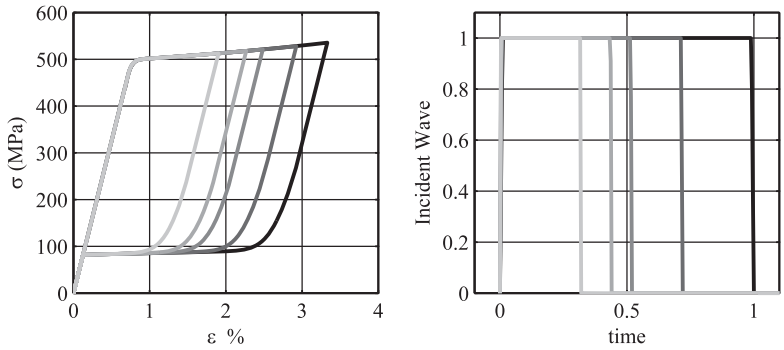

Figure 14. The stress–strain response of a specimen in a Kolsky bar (left) subjected to pulses of varying duration (right). The amplitude and pulse duration are normalized with those described in section 4.1.

since there is only a transmitted wave. Assuming the equilibrium of the specimen, the stress in the specimen is uniform and one has continuity of forces so that

$$\sigma(t)A_s = \sigma_L(0,t)A_b = \sigma_R(0,t)A_b, \tag{53}$$

where A_s and A_b are the cross-sectional areas of the specimen and the input/output bars respectively. Further, the overall strain in the specimen is given as

$$\varepsilon = \frac{u_R(0,t) - u_L(0,t)}{l_s} \tag{54}$$

where l_s is the length of the specimen. It follows that

$$\sigma(t) = \frac{EA_b}{A_s}(f'(-ct) + g'(+ct)) = \frac{EA_b}{A_s}h'(-ct) \tag{55}$$

$$\dot{\varepsilon}(t) = \frac{c(-h'(-ct) + f'(-ct) - g'(ct))}{l_s}. \tag{56}$$

In a typical Kolsky bar experiment, f is applied, g and h are measured, and (55) and (56) are used to obtain the stress and strain rate.

Our purpose here is to see how a material described by our model would behave when subjected to a Kolsky bar experiment. So we apply a given input pulse $f(t)$ and integrate the three equations in (55) and (56) as well as (26)–(28), (33) and (21) to obtain g, h, σ, ε, ε_m, ε_p, λ and θ. The initial conditions are the same as before.

We choose the same parameters as described in section 3.1. Further,

$$E_{\text{specimen}} = 65 \text{ GPa} \qquad \text{and} \qquad E_{\text{Kolsky bars}} = 200 \text{ GPa}$$

$$\text{Diameter}_{\text{specimen}} = 1.5 \text{ cm} \qquad \text{and}$$

$$\text{Diameter}_{\text{Kolsky bars}} = 2 \text{ cm}$$

$$\text{Length}_{\text{specimen}} = 0.5 \text{ cm} \qquad \text{and}$$

$$\text{Wave speed}_{\text{specimen}} = 3500 \text{ m s}^{-1}.$$

The yield strength is chosen high enough so that plasticity plays no role. We assume that the incident wave f' is a square pulse with amplitude 1.5×10^{-3}, duration 50×10^{-4} s and rise time 50×10^{-6} s unless otherwise stated.

4.2. Ambient temperature

Figure 12 shows the result for different ambient (initial) temperatures. The critical stress at which phase transformation starts rises as temperature increases. Therefore, at the highest temperature displayed, 30 °C, the transformation begins very late, just before the unloading. Therefore, the material does not have enough time to transform fully, and consequently the amount of strain is small. It fully transforms back on unloading. As the ambient temperature drops, the transformation begins earlier and thus proceeds further, and the

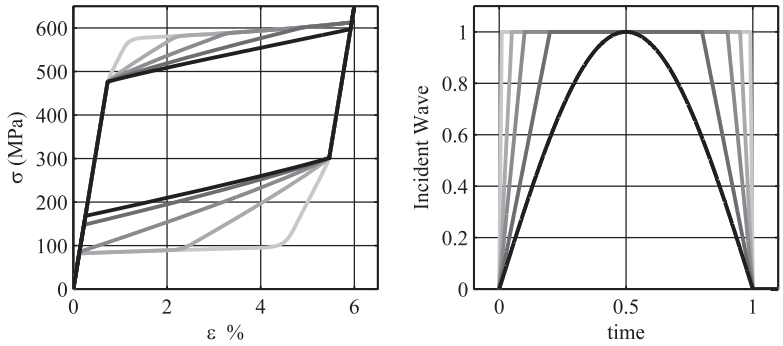

Figure 15. The stress–strain response of a specimen in a Kolsky bar (left) subjected to pulses of varying shape (right). The amplitude and pulse duration are normalized with those described in section 4.1.

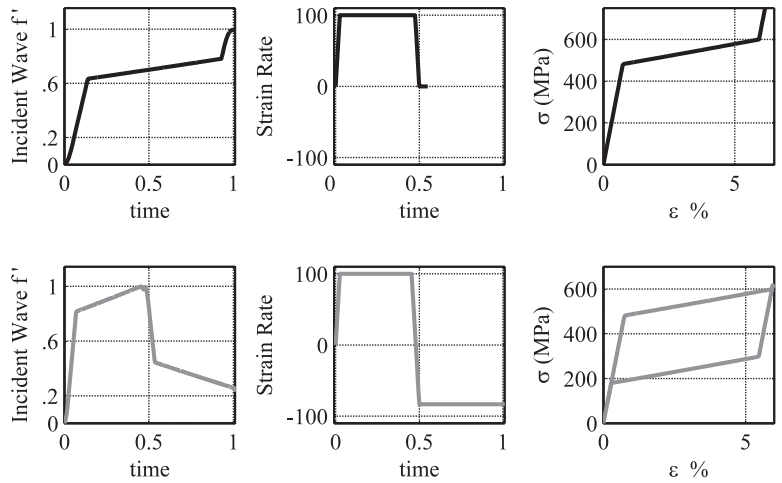

Figure 16. Design of the input pulse for obtaining a constant strain-rate in the specimen.

amount of strain increases. It continues to fully transform back except at the lowest displayed temperature of −25 °C.

Notice that the shapes of the stress–strain curves are less boxy than those obtained with the given stress history in section 3. This is our first indication that the stress–strain curve can depend on experimental methodology.

4.3. Pulse amplitude, size and shape

Figure 13 shows the effect of changing the pulse amplitude while figure 14 shows the effect of changing pulse duration. As expected, the transformation is incomplete with smaller or shorter pulses. Figure 15 shows the stress–strain curve for various pulse shapes. The shape of the stress–strain curve can vary widely with the input pulse.

This points to the importance of designing an appropriate pulse shape in experiments. We address this by using the model to design a pulse that will yield a desired strain rate history. We set our given strain rate $\dot{\varepsilon}(t)$ and integrate the three equations in (55) and (56) as well as (26)–(28), (33) and (21) to obtain f, g, h, σ, ε_{m}, ε_{p}, λ and θ. The initial conditions are the same as before. The results are shown in figure 16 for two desired strain rates. The first is a constant strain rate to investigate the loading. Note that the pulse cannot have constant amplitude but has to gradually increase in amplitude. The second is a jump test to study both the loading and the unloading.

5. Conclusion

A one-dimensional thermodynamically consistent constitutive framework for the dynamic behavior of polycrystalline shape-memory alloys is presented. The model is demonstrated using both stress-controlled and Kolsky bar experiments. This model is consistent with the observed asymmetric response of shape memory alloys under tension and compression. It is able to capture the behavior under a wide range of temperature spanning both the shape-memory effect and superelasticity, and a wide range of loading rates. The model is relatively easy to use, and this is demonstrated by designing the necessary pulse for a constant strain rate in a Kolsky bar experiment. A generalization of the model in multiple dimensions with both proportional and non-proportional loading conditions will be presented elsewhere [31].

Acknowledgment

We gratefully appreciate the partial financial support of the Office of Naval Research through the MURI grant N00014-02-1-0666.

References

[1] Abeyaratne R, Chu C and James R D 1996 Kinetics of materials with wiggly energies: theory and application to the evolution of twinning microstructure in a Cu–Al–Ni alloy *Phil. Mag.* A **73** 457–97

[2] Abeyaratne R and Knowles J K 1993 A continuum model of a thermoelastic solid capable of undergoing phase-transitions *J. Mech. Phys. Solids* **41** 541–71

[3] Auricchio F, Taylor R L and Lubliner J 1997 Shape-memory alloys: macromodelling and numerical simulations of the superelastic behavior *Comput. Meth. Appl. Mech. Eng.* **146** 281–312

[4] Auricchio F and Petrini L 2004 A three-dimensional model describing stress-temperature induced solid phase transformations: solution algorithm and boundary value problems *Int. J. Numer. Methods Eng.* **61** 807–36

[5] Ball J M and James R D 1987 Fine phase mixtures as minimizers of energy *Arch. Ration. Mech. Anal.* **100** 13–52

[6] Ball J M, Chu C and James R D 1995 Hysteresis during stress-induced variant rearrangement *J. Physique Coll.* **5** 245–51

[7] Bhattacharya K 1999 Phase boundary propagation in heterogeneous bodies *Proc. R. Soc.* A **455** 757–66

[8] Bhattacharya K 2003 *Microstructure of Martensite* (Oxford: Oxford University Press)

[9] Bhattacharya K and Kohn R V 1997 Energy minimization and the recoverable strains of polycrystalline shape-memory alloys *Arch. Ration. Mech. Anal.* **139** 99–180

[10] Bhattacharya K, Purohit P and Craciun B 2003 The mobility of twin and phase boundaries *J. Physique Coll.* **112** 163–6

[11] Bhattacharya K and Schlömerkemper A 2004 Transformation yield surface of shape-memory alloys *J. Physique Coll.* **115** 155–62

[12] Bhattacharyya A, Lagoudas D C, Wang Y and Kinra V K 1995 On the role of thermoelectric heat transfer in the design of SMA actuators: theoretical modelling and experiments *Smart Mater. Struct.* **4** 252–63

[13] Boyd J G and Lagoudas D C 1996 A thermodynamic constitutive model for shape memory materials: 1. The monolithic shape memory alloy *Int. J. Plast.* **12** 843–73

[14] Brocca M, Brinson L C and Bazant Z 2002 Three-dimensional constitutive model for shape memory alloys based on microplane *J. Mech. Phys. Solids* **50** 1051–77

[15] Coleman B D and Noll W 1963 The thermodynamics of elastic materials with heat conduction and viscosity *Arch. Ration. Mech. Anal.* **13** 167–83

[16] Huo Y and Müller I 1993 Nonequilibrium thermodynamics of pseudoelasticity *Cont. Mech. Thermodyn.* **5** 163–204

[17] Huang M and Brinson L C 1998 Multivariant model for single crystal shape memory alloy behavior *J. Mech. Phys. Solids* **46** 1379–409

[18] Jung Y J, Papadopoulos P and Ritchie R O 2004 Constitutive modelling and numerical simulation of multivariant phase transformation in superelastic shape-memory alloys *Int. J. Numer. Methods Eng.* **60** 429–60

[19] Lexcellent C, Goo B C, Sun Q P and Bernardini J 1996 Characterization, thermomechanical behaviour and micromechanical-based constitutive model of shape-memory Cu–Zn–Al single crystals *Acta Mater.* **44** 3773–80

[20] Lexcellent C, Vivet A, Bouvet C, Calloch S and Blanc P 2002 Experimental and numerical determinations of the initial surface of phase transformation under biaxial polycrystalline shape-memory alloys *J. Mech. Phys. Solids* **50** 2717–35

[21] Lu Z K and Weng G J 1997 Martensitic transformation and stress–strain relations of shape-memory alloys *J. Mech. Phys. Solids* **45** 1905

[22] Marketz F and Fischer F D 1996 Modelling the mechanical behavior of shape memory alloys under variant coalescence *Comput. Mater. Sci.* **5** 210–26

[23] McNaney J M, Imbeni V, Jung Y, Papadopoulos P and Ritchie R O 2003 An experimental study of the superelastic effect in a shape-memory Nitinol alloy under biaxial loading *Mech. Mater.* **35** 969–86

[24] Naito H, Matsuzaki Y and Ikeda T 2004 A unified constitutive model of phase transformations and rearrangements of shape memory alloy wires subjected to quasistatic load *Smart Mater. Struct.* **13** 535–43

[25] Nemat-Nasser S, Choi J Y, Guo W G, Isaacs J B and Taya M 2005 High strain-rate, small strain response of a NiTi shape-memory alloy *J. Eng. Mater. Technol.* **127** 83–9

[26] Niclaeys C, Ben Zineb T, Arbab-Chirani S and Patoor E 2002 Determination of the interaction energy in the martensitic state *Int. J. Plast.* **18** 1619–47

[27] Otsuka K and Wayman C M 1998 *Shape Memory Materials* (Cambridge: Cambridge University Press)

[28] Paiva A, Savi M A, Braga A M B and Pacheco P M C L 2005 A constitutive model for shape memory alloys considering tensile-compressive asymmetry and plasticity *Int. J. Solid Struct.* **42** 3439–57

[29] Purohit P K 2002 Dynamics of phase transformations in strings, beams and atomic chains *PhD Thesis* California Institute of Technology

[30] Qidwai M A and Lagoudas D C 2000 Numerical implementation of a shape memory alloy thermomechanical constitutive model using return mapping algorithms *Int. J. Numer. Methods Eng.* **47** 1123–68

[31] Sadjadpour A and Bhattacharya K 2006 A micromechanics inspired constitutive model for shape-memory alloys, in preparation

[32] Sadjadpour A, Rittel D, Ravichandran G and Bhattacharya K 2006 Dynamic deformation of iron under shear, in preparation

[33] Stoijov V and Bhattacharyya A 2002 A theoretical framework of one-dimensional sharp phase fronts in shape memory alloys *Acta Mater.* **50** 4939–52

[34] Sun Q P and Hwang K C 1993 Micromechanics modeling for the constitutive behavior of polycrystalline shape-memory alloys: 1. Derivation of general relations *J. Mech. Phys. Solids* **41** 1–17

Sun Q P and Hwang K C 1993 Micromechanics modeling for the constitutive behavior of polycrystalline shape-memory alloys: 2. Study of individual phenomena *J. Mech. Phys. Solids* **41** 19–33

[35] Thamburaja P and Anand L 2001 Polycrystalline shape-memory materials: effect of crystallographic texture *J. Mech. Phys. Solids* **49** 709–37

[36] Wang Y H and Fang D N 2003 A three-dimensional constitutive model for shape memory alloys *Int. J. Nonlinear Sci. Numer. Sim.* **4** 81–7

[37] Zhu J J, Huang N G, Liew K M and Zhu Z H 2002 A thermodynamic constitutive model for stress induced phase transformation in shape memory alloys *Int. J. Solid Struct.* **39** 741–63