CrossMark

A Gradient-Based Constitutive Model for Shape Memory Alloys

Majid Tabesh[1] · James Boyd[1] · Dimitris Lagoudas[1]

Abstract Constitutive models are necessary to design shape memory alloy (SMA) components at nano- and micro-scales in NEMS and MEMS. The behavior of small-scale SMA structures deviates from that of the bulk material. Unfortunately, this response cannot be modeled using conventional constitutive models which lack an intrinsic length scale. At small scales, size effects are often observed along with large gradients in the stress or strain. Therefore, a gradient-based thermodynamically consistent constitutive framework is established. Generalized surface and body forces are assumed to contribute to the free energy as work conjugates to the martensite volume fraction, transformation strain tensor, and their spatial gradients. The rates of evolution of these variables are obtained by invoking the principal of maximum dissipation after assuming a transformation surface, which is a differential equation in space. This approach is compared to the theories that use a configurational force (microforce) balance law. The developed constitutive model includes energetic and dissipative length scales that can be calibrated experimentally. Boundary value problems, including pure bending of SMA beams and simple torsion of SMA cylindrical bars, are solved to demonstrate the capabilities of this model. These problems contain the differential equation for the transformation surface as well as the equilibrium equation and are solved analytically and numerically. The simplest version of the model, containing only the additional gradient of martensite volume fraction, predicts a response with greater transformation hardening for smaller structures.

Keywords Shape memory alloys · Size effect · Phase transformation · Analytical modeling · Constitutive modeling · Strain gradient theory

Introduction

Shape memory alloys (SMAs) have recently been used as high-performance actuators in micro-electro-mechanical systems (MEMS) in the form of thin films/beams including micropumps [1], microvalves [2], microgrippers [3], and microactuators [4]. Also, compared to bulk SMAs, thin films can be thermally activated at a higher frequency due to the larger exposed free surfaces. The elastic and inelastic response of materials changes as their size approaches the micron/nanometer region where fluctuations in the microstructural and physical features of the material cannot be resolved by smearing their effect through averaging and homogenization. Such an experimentally observed size-dependent behavior can be modeled using continuum mechanics without resort to cost-prohibitive molecular/atomistic frameworks by using nonlocal or gradient continuum theories. Higher gradients of stress or strain usually occur in smaller sizes. For example, for the same maximum strain on the outer layer, beams with smaller thicknesses undergo higher gradients of strain, motivating the use of higher-order models containing gradient terms for modeling their behavior. Toward this end, the objective for this communication is to develop a thermodynamically consistent gradient-based nonlocal constitutive model for shape memory alloys with the capability of capturing the size effect.

✉ James Boyd
jgboyd@tamu.edu

1   Department of Aerospace Engineering, Texas A&M University, 702 H.R.Bright, 3141 TAMU, College Station, TX 77843-3141, USA

A size effect generally occurs when a characteristic length in the loading, boundary conditions, or geometry of the sample (such as the size of the stress concentration region, curvature of bending, or thickness of the sample) is of the same order of magnitude as a microstructural length scale in the material (such as grain size, size of the dislocation loop in metals, or martensitic twin variants in shape memory alloys). In that case, the heterogeneities in the microstructure of the specimen and hence the local state of the material become significant and cannot be averaged out to produce a larger scale [5]. Size effect has been observed for elastic properties [6–8] as well as inelastic response of materials, such as in dislocation plasticity [9–11].

SMAs are not an exception in this regard because the properties characterizing their unique behavior, i.e., the critical stresses to start and finish the forward and reverse transformations as well as the transformation dissipation and hysteresis, have been shown to depend on the size of the specimen.

The size effect was observed in microcrystalline SMA particles in nontransformable solid matrix, and also in free-standing nano/micro-sized SMA powders. Experimental results demonstrated that by decreasing the size of the SMA particle or powder, the martensitic transformation was fully or partially suppressed [12, 13].

Furthermore, the compressive stress–strain response of SMA micro/nanopillars was shown to depend on the diameter. An increase in the critical stresses to start the austenite to martensite transformation and also in the stress–strain hysteresis was seen when the diameter of the pillars was less than one micron [14–18]. This phenomenon was also observed in SMA wires where the critical stresses for the start of martensite and austenite transformations were reported to increase for diameters below 100 μm [19].

SMA thin films with thicknesses in the micron scale, deposited commonly on glass or silicon substrates, showed reduced transformation temperatures compared to that of the bulk SMA, and in some cases the martensitic transformation was impeded for thicknesses below 100 nm [20–23].

It was also shown through experiments and numerical modeling that martensitic transformation diminishes with decreasing grain size in SMA samples with nano-sized grains, and there exists a critical grain size below which transformation is inhibited [24–29].

It is possible to account for such high-resolution internal microstructural features explicitly by considering the spatial variation of the material properties in a lower scale (for example, using an FEA model including detailed microstructural features or an atomistic/molecular modeling approach). However, this strategy requires much computational resources. The nonlocal models, either differential or integral, introduce one or more length scales in their formulation that can represent the lower scale features in an average sense in the continuum scale.

The standard continuum theories predict a singular state of stress at the tip of a crack or center of a dislocation core. Nonlocal theories of elasticity can eliminate these singularities [30–32]. In addition, standard continuum theories have difficulty predicting size dependence in the elastic torsional or flexural response of certain materials [6, 7, 33, 34, 35, 8]. Furthermore, the elastoplastic response of micron and submicron-sized specimens feature an increased plastic hardening with decreasing size. This effect was modeled using both implicit and explicit strain gradient plasticity models [36, 37]. Plastic size effects include nano-scale indentation depth dependence for hardness of several metals [38], thickness dependence in nondimensional moment-curvature response of Ni thin films [39], and size-dependent yield stress in torsional response of micron-sized Cu wires [9]. The standard plasticity models, due to the lack of intrinsic material length scales, cannot capture such an observed size effect featuring a stronger response as the size of the structure reduces [10].

Nonlocal integral plasticity models are a generalization of the classical plasticity models that include nonlocal measures of the plastic strain, defined via a weighted volume integral over the domain, in their yield function [40, 41]. Strongly nonlocal implicit gradient plasticity models [42] also include an integral nonlocal variable in their constitutive equations. However, the nonlocal variable (such as damage or the accumulated plastic strain) can be obtained via solution of a Helmholtz-type differential equation over the material body with Neumann boundary conditions applied on the external boundary (and not on the evolving elastic-plastic boundary) and solved in a coupled fashion with the equilibrium equation [43]. In contrast, weakly nonlocal constitutive models [42] for inelastic material behavior directly include the spatial gradients of the state variables (hence also called explicit gradient models) that capture the effect of an infinitesimally small neighborhood around a material point. This can be the higher gradients of the displacement field (or the gradients of total strain) [44] or can be exclusive to the gradients of the internal variables such as the accumulated plastic strain [45–47] or the martensite volume fraction (MVF) [48, 49]. Explicit gradient models were shown to be more effective in capturing the strengthening as a result of reducing the specimen size, while nonlocal implicit models proved to be more efficient in regularizing the problem of strain localization in materials with softening behavior [50]. Gudmundson et al. [51] presented a thermodynamic framework that can be used to derive many of the existing strain gradient plasticity models. A generalized principle of virtual power was assumed that considers the contributions

from elastic strains, plastic strains, and the plastic strain gradients along with their respective work conjugates of Cauchy stress, microstress, and moment stress. The principle of virtual power led to the balance equations and corresponding standard and nonstandard boundary conditions with respect to the force and moment force tractions. The balance law regarding microstresses is often denoted as the microforce balance after Gurtin [52]. The microforce balance acts as a generalized flow rule in this context. The rates of the plastic strain and the plastic strain gradient were obtained based on satisfaction of the second law of thermodynamics through derivatives of a yield function. The yield function was defined based on generalized stress and plastic strain measures that accommodated three different intrinsic material length scales. Connections between this theory and strain gradient theories of Aifantis [53–55], and Fleck and Hutchinson [36, 47], were also demonstrated.

Phenomenological modeling of SMAs using internal variables has been under development for the past two decades [56]. The SMA phenomenological models are similar to rate-independent dislocation-based plasticity models. The nonlocal continuum models motivated by the experimentally observed SMA behavior, such as the size effect or transformation front localization, are also similar to the nonlocal plasticity models [57, 58, 48].

Duval et al. [58] provided the nonlocal extension of an existing phenomenological SMA constitutive model [59]. This nonlocal implicit gradient model was inspired by the work of Engelen et al. [43]. In addition to the conventional internal variable of martensite volume fraction (MVF) f, the integral average of it over the domain was also considered as the nonlocal MVF, $\bar{f}$. Similar to the implicit gradient plasticity models, $\bar{f}$ was obtained from an additional PDE with a Neumann-type boundary condition on the domain boundary ensuring the equality of local and nonlocal MVF averaged over the entire domain. The nonlocal model was implemented in the finite element software package ABAQUS through the user element (UEL) feature. The experimentally observed softening behavior (stress peak of nucleation and subsequent stress plateau for propagation) in the SMAs was captured by gradually decreasing the critical force for martensitic transformation through the nonlocal MVF resulting in lower stresses required for propagation of martensite compared to its nucleation. Therefore, the nonlocal model performed as a localization limiter improving the otherwise pathological behaviors of the local models. The nucleation and propagation of martensite transformation front in tensile loading of an SMA plate with a hole were simulated with results showing the dependence of the localization width and the stress peak of nucleation on the intrinsic

length scale introduced in the model through the nonlocal variable.

Another nonlocal implicit gradient SMA model was presented by Badnava et al. [57]. The model was based on a 3D extension of Brinson's SMA model [60] and was intended to capture the unstable softening and localization behavior of SMAs upon nucleation and propagation of martensite transformation. The effectiveness of the model as a localization limiter was demonstrated through FEA simulation of SMA structures after implementation in ABAQUS UEL.

The explicit gradient-based SMA model of Qiao et al. [48] was inspired by the strain gradient plasticity work of Gurtin and Anand [61]. This isothermal one-dimensional model was intended to capture the size-dependent superelastic response of SMA micro/nanopillars under compression tests. The Helmholtz free energy for the material contained the gradient of MVF through which an energetic length scale was introduced. Through the use of a generalized form for the principle of virtual power, stress equilibrium equation and a balance equation regarding the microstresses (conjugates to the martensite volume fraction and its spatial gradient) along with their corresponding boundary conditions were obtained. Constitutive relations were assumed for the microstresses such that the microforce balance played the role of the transformation partial differential equation giving the rate of MVF. The variational form of this equation was implemented in a 1D finite element framework to simulate size effect in the response of Cu-Al-Ni SMA micropillars under compression.

In the work by Sun and He [62], a multiscale continuum phenomenological strain gradient model was developed to study the effect of grain size in the response of polycrystalline SMA specimens. The model was based on a nonlocal nonconvex strain energy function. The characteristic length scales introduced were the specimen size L, grain size l, and the intrinsic material length scale ℓ related to the width of austenite–martensite interface. The results demonstrated that the energy dissipation during phase transformation is governed by the ratios of $\frac{L}{l}$ and $\frac{l}{\ell}$. The martensitic transformation, hence dissipation, was diminished in the case of an SMA with large grains close to the size of the specimen (denoting a single crystal behavior) and also in the case where the grains were of the order of nanometer in size close to the intrinsic length scale (denoting the transformation inhibition observed in ultrafine-grain SMAs).

In the current work, a gradient-based thermodynamically consistent constitutive model is developed for the response of SMAs at small scales. It is assumed that generalized surface and body forces contribute to the free energy as work conjugates to the MVF, transformation

strain tensor, and their spatial gradients. The rates of evolution of these variables are obtained by invoking the maximum dissipation postulate. The generalized transformation surface is a differential equation for the spatial distribution of the thermodynamic forces, whereas in conventional constitutive theories, the transformation surface is an algebraic equation. The connection between this model and the theories that use a configurational force (microforce) balance law is established by showing the latter to be a specific case of a generalized transformation surface.

Using such a framework, gradient-based constitutive models are developed that include energetic and dissipative length scales which can be calibrated experimentally. These models are simplified for 1D. Example boundary value problems are solved analytically and, where impossible, numerically for pure bending of SMA beams and simple torsion of SMA cylindrical bars. The simplest gradient-based SMA constitutive model, containing only the additional gradient of MVF, can capture the size effect in the response of these structures by demonstrating an increased hardening for smaller sizes.

Gradient-Based SMA Constitutive Modeling

In this section, a thermodynamically consistent gradient-based model for SMAs is developed. A treatment based on generalized principles of continuum mechanics is given in the Appendix "Fundamentals" section; however, the internal variable approach is used herein. The development begins by specifying the Gibbs free energy, G. The strain and entropy are shown to be the derivatives of the Gibbs free energy with respect to stress and temperature. A dissipation potential is formulated to obtain (rate-independent) evolution equations for the internal variables.

To begin, the dependent and independent state variables are introduced:

$$G = \hat{G}(\boldsymbol{\sigma}, T, \mathbf{Y}), \quad \boldsymbol{\varepsilon} = \hat{\varepsilon}(\boldsymbol{\sigma}, T, \mathbf{Y}), \quad \mathbf{q} = \hat{\mathbf{q}}(\boldsymbol{\sigma}, T, \mathbf{Y}),$$
$$s = \hat{s}(\boldsymbol{\sigma}, T, \mathbf{Y}),$$

$$\mathbf{Y} \equiv \{\xi, \{\nabla\xi, \nabla\nabla\xi, \ldots, \nabla^n\xi\}, \boldsymbol{\varepsilon}^{\mathrm{tr}}, \{\nabla\boldsymbol{\varepsilon}^{\mathrm{tr}}, \nabla\nabla\boldsymbol{\varepsilon}^{\mathrm{tr}}, \ldots, \nabla^m\boldsymbol{\varepsilon}^{\mathrm{tr}}\}\}. \tag{1}$$

The independent state variables include the Cauchy stress $\boldsymbol{\sigma}$, the absolute temperature T, and a set of internal variables $\mathbf{Y}$ with $\boldsymbol{\varepsilon}^{\mathrm{tr}}$, the tensor for transformation strain, and ξ, the volume fraction of martensite phase in the SMA. Also included are the spatial gradients of the MVF and the transformation strain tensor up to orders n and m. It is assumed, henceforth, that $m = n$. The dependent variables are $\boldsymbol{\varepsilon}$ the infinitesimal strain (we assume infinitesimal strains), $\mathbf{q}$ the heat flux vector, s the specific entropy, and G the specific Gibbs free energy.

Referring to the inequality in Eq. (90) and the development in the Appendix "Fundamentals" section, the Clausius–Plank inequality becomes

$$-\left(\rho\frac{\partial G}{\partial \boldsymbol{\sigma}} + \boldsymbol{\varepsilon}\right) : \dot{\sigma} - \rho\left(\frac{\partial G}{\partial T} + s\right)\dot{T}$$
$$- \rho\left[\frac{\partial G}{\partial \boldsymbol{\varepsilon}^{\mathrm{tr}}} : \dot{\varepsilon}^{\mathrm{tr}} + \frac{\partial G}{\partial \xi}\dot{\xi} + \sum_{k=1}^{n}\left(\frac{\partial G}{\partial \nabla^k\xi}\cdot\nabla^k\dot{\xi} + \frac{\partial G}{\partial \nabla^k\boldsymbol{\varepsilon}^{\mathrm{tr}}}\cdot\nabla^k\dot{\varepsilon}^{\mathrm{tr}}\right)\right] \geq 0. \tag{2}$$

Assuming that all the constitutive variables and their rates are independent, the Clausius–Plank inequality (and hence the second law of thermodynamics as mentioned in the Appendix "Constitutive Equations" subsection) is satisfied given the following constitutive equations:

$$\boldsymbol{\varepsilon} = -\rho\frac{\partial G}{\partial \boldsymbol{\sigma}}, \quad s = -\frac{\partial G}{\partial T}. \tag{3}$$

Therefore, the dissipation is

$$D^{\mathrm{tr}} = \boldsymbol{\Gamma}\cdot\dot{\mathbf{Y}}$$
$$= \varsigma^D : \dot{\varepsilon}^{\mathrm{tr}} + \pi\dot{\xi} + \sum_{k=1}^{n}\bar{\pi}^k\cdot\nabla^k\dot{\xi} + \sum_{k=1}^{n}\boldsymbol{\tau}^k\cdot\nabla^k\dot{\varepsilon}^{\mathrm{tr}} \geq 0, \tag{4}$$

where $\dot{\mathbf{Y}}$ denotes the set of generalized thermodynamic fluxes with corresponding thermodynamic forces, $\boldsymbol{\Gamma}$, given by

$$\boldsymbol{\Gamma} = -\rho\frac{\partial G}{\partial \mathbf{Y}}. \tag{5}$$

That is

$$\varsigma^D = -\rho\frac{\partial G}{\partial \boldsymbol{\varepsilon}^{\mathrm{tr}}}, \quad \pi = -\rho\frac{\partial G}{\partial \xi}, \quad \bar{\pi}^k = -\rho\frac{\partial G}{\partial \nabla^k\xi},$$
$$\boldsymbol{\tau}^k = -\rho\frac{\partial G}{\partial \nabla^k\boldsymbol{\varepsilon}^{\mathrm{tr}}}. \tag{6}$$

It is common to consider the martensitic transformation a volume-preserving process such that $\mathrm{tr}(\boldsymbol{\varepsilon}^{\mathrm{tr}}) = 0$. Therefore, the hydrostatic part of the generalized stress, ς^D, does not contribute to the dissipation. Hence, only its deviatoric part will be considered for determining the evolution of the internal variables:

$$\acute{\varsigma}_{ij} = \varsigma_{ij} - \frac{1}{3}\mathrm{tr}(\varsigma^D)\delta_{ij}, \quad \mathrm{tr}(\varsigma^D) = \acute{\varsigma}_{ii} = 0. \tag{7}$$

It is assumed that a transformation surface defines the region for the thermoelastic state of the material in the space of the generalized forces and that any admissible state of the material must satisfy the condition imposed by

$$\Phi(\boldsymbol{\Gamma}) = \Phi(\varsigma^D, \pi, \bar{\pi}^1, \ldots, \bar{\pi}^n, \boldsymbol{\tau}^1, \ldots, \boldsymbol{\tau}^n) \leq 0. \tag{8}$$

For the rate-independent response considered here, the fluxes are associated with the forces through derivatives of a transformation potential [63]. One approach uses the principle of maximum dissipation (PMD) such that the

fluxes will be normal to the transformation surface. Normality of the fluxes in the space of the generalized forces and also convexity of the transformation surface are the results of the PMD. According to PMD, the transformation state of the SMA material, belonging to the set of admissible states, is the one that maximizes the dissipation D^{tr} or

$$\mathrm{Max}\left\{D^{\mathrm{tr}} = \mathbf{\Gamma} \cdot \dot{\mathbf{Y}} \mid \mathbf{\Gamma}, \ \Phi(\mathbf{\Gamma}) \le 0\right\}. \tag{9}$$

This is a minimization programming which is solved using the method of Lagrange multipliers, further described in the Appendix "Fundamentals" section and Eqs. (98)–(101) including the corresponding Kuhn–Tucker conditions. As a result, the rates of the internal variables can be found from

$$\dot{\varepsilon}^{\mathrm{tr}} = \lambda \frac{\partial \Phi}{\partial \zeta^D}, \quad \dot{\xi} = \lambda \frac{\partial \Phi}{\partial \pi}, \quad \nabla^k \dot{\xi} = \lambda \frac{\partial \Phi}{\partial \bar{\pi}^k},$$
$$\nabla^k \dot{\varepsilon}^{\mathrm{tr}} = \lambda \frac{\partial \Phi}{\partial \boldsymbol{\tau}^k}. \tag{10}$$

Hence, the response is associative in the space of generalized thermodynamic forces.

Next we assume a form for the transformation surface $\Phi(\mathbf{\Gamma})$. For Φ to be a general anisotropic convex function of ζ^D, π, $\bar{\pi}^1, \ldots, \bar{\pi}^n$, and $\boldsymbol{\tau}^1, \ldots, \boldsymbol{\tau}^n$ satisfying the second law of thermodynamics for forward and reverse martensitic transformations, it is sufficient to have

$$\Phi^{\mathrm{fwd}}\left(\zeta^D, \pi, \bar{\pi}^1, \ldots, \bar{\pi}^n, \boldsymbol{\tau}^1, \ldots, \boldsymbol{\tau}^n\right) = \phi - Y,$$
$$\Phi^{\mathrm{rev}}\left(\zeta^D, \pi, \bar{\pi}^1, \ldots, \bar{\pi}^n, \boldsymbol{\tau}^1, \ldots, \boldsymbol{\tau}^n\right) = \phi + Y \tag{11}$$

$$\phi = \tilde{\varphi}(\zeta^D) + \check{\varphi}(\pi) + \sum_{k=1}^{n} \bar{\varphi}^k(\bar{\pi}^k) + \sum_{k=1}^{n} \hat{\varphi}^k(\boldsymbol{\tau}^k) \tag{12}$$

where various φ are homogeneous of degree 1, convex functions of their respective variables. A general quadratic form is commonly assumed such that

$$\tilde{\varphi}(\zeta^D) = \left(\Lambda^I_{ijrs}\, \acute{\zeta}_{ij}\, \acute{\zeta}_{rs}\right)^{\frac{1}{2}},$$
$$\check{\varphi}(\pi) = \pi, \ \bar{\varphi}^k(\bar{\pi}^k) = -\mathbf{\Xi}^{II_k} \cdot \bar{\pi}^k - \left(\bar{\pi}^k \cdot \mathbf{\Lambda}^{II_k} \bar{\pi}^k\right)^{\frac{1}{2}},$$
$$\hat{\varphi}(\boldsymbol{\tau}) = -\mathbf{\Xi}^{III_k} \cdot \boldsymbol{\tau}^k - \left(\boldsymbol{\tau}^k \cdot \mathbf{\Lambda}^{III_k} \boldsymbol{\tau}^k\right)^{\frac{1}{2}} \tag{13}$$

with no summation over k. Here, notice that if $\bar{\pi}^k$ or $\boldsymbol{\tau}^k \in \mathcal{T}^m$ then $\mathbf{\Xi}^{II_k}$ or $\mathbf{\Xi}^{III_k} \in \mathcal{T}^m$ and $\mathbf{\Lambda}^{II_k}$ or $\mathbf{\Lambda}^{III_k} \in \mathcal{T}^{2m}$. The tensors $\mathbf{\Lambda}$ and $\mathbf{\Xi}$ can be assigned the symmetry necessary to achieve the directional dependence of the yield surface and flow observed in the response of the SMA. Y is a constant that is equal to one-half of the amount of energy dissipation in a full transformation path.

The surface ϕ given by Eq. (11) can result in multiple forms of differential equations. For example, consider $\bar{\pi}^1 = -\rho \frac{\partial G}{\partial \nabla \xi} = a\nabla \xi$ and $\bar{\pi}^2 = -\rho \frac{\partial G}{\partial \nabla^2 \xi} = b\nabla^2 \xi$ which can be part of an isotropic SMA constitutive model including ξ,

$\nabla \xi$, and $\nabla^2 \xi$ as internal variables. Hence, one can have first and second-order differential terms, including the Laplacian $\xi_{,ii}$, within the definition of the partial differential equation for the transformation surface:

$$\bar{\pi}_i^1 = a\xi_{,i}, \quad \bar{\varphi}^1(\bar{\pi}^1) = -A\sqrt{\xi_{,i}\xi_{,i}}, \bar{\pi}_{ij}^2 = b\xi_{,ij},$$
$$\bar{\varphi}^2(\bar{\pi}^2) = -B\xi_{,ii} - \sqrt{C\xi_{,ii}^2 + D\xi_{,ij}\xi_{,ij}}. \tag{14}$$

It is possible to eliminate λ from Eq. (10) such that the rate of evolution of the internal variables will be proportional to $\dot{\xi}$ since $\dot{\xi} = \lambda \frac{\partial \Phi}{\partial \pi} = \lambda$. This implies that $\dot{\xi}$ plays the role of the plastic multiplier as in the analysis of dissipation surfaces and flow in plasticity. $\dot{\xi}$ can, furthermore, be determined using the consistency condition.

According to the Kuhn–Tucker condition (iii), while $\dot{\xi} = \lambda \ne 0 \to \Phi = 0$, thus the consistency condition during forward or reverse transformation (fwd/rev) can be obtained considering the transformation surface in Eq. (11):

$$\dot{\Phi}^{\mathrm{fwd/rev}} = 0, \frac{\partial \Phi^{\mathrm{fwd/rev}}}{\partial \boldsymbol{\sigma}} : \dot{\sigma} + \frac{\partial \Phi^{\mathrm{fwd/rev}}}{\partial T} \dot{T} + \frac{\partial \Phi^{\mathrm{fwd/rev}}}{\partial \boldsymbol{\varepsilon}^{tr}} : \dot{\varepsilon}^{tr}$$
$$+ \frac{\partial \Phi^{\mathrm{fwd/rev}}}{\partial \xi} \dot{\xi} + \sum_{k=1}^{n} \frac{\partial \Phi^{\mathrm{fwd/rev}}}{\partial \nabla^k \xi} \cdot \nabla^k \dot{\xi}$$
$$+ \sum_{k=1}^{n} \frac{\partial \Phi^{\mathrm{fwd/rev}}}{\partial \nabla^k \boldsymbol{\varepsilon}^{tr}} \cdot \nabla^k \dot{\varepsilon}^{tr} = 0 \tag{15}$$

which leads to the following for the evolution of the martensite volume fraction:

$$\dot{\xi} = -\frac{\frac{\partial \phi}{\partial \boldsymbol{\sigma}} : \dot{\sigma} + \frac{\partial \phi}{\partial T} \dot{T}}{\frac{\partial \phi}{\partial \xi} + \frac{\partial \phi}{\partial \boldsymbol{\varepsilon}^{tr}} : \frac{\partial \phi}{\partial \zeta^D} + \sum_{k=1}^{n} \frac{\partial \phi}{\partial \nabla^k \xi} \cdot \frac{\partial \phi}{\partial \bar{\pi}^k} + \sum_{k=1}^{n} \frac{\partial \phi}{\partial \nabla^k \boldsymbol{\varepsilon}^{tr}} \cdot \frac{\partial \phi}{\partial \boldsymbol{\tau}^k}}. \tag{16}$$

Once the evolution of internal variables is given by the derivatives of the transformation surface as in Eq. (10), the dissipation due to martensitic transformation can be obtained as follows:

$$D^{\mathrm{tr}} = \lambda \left(\frac{\partial \Phi}{\partial \mathbf{\Gamma}} \cdot \mathbf{\Gamma}\right) = \lambda \left(\frac{\partial \phi}{\partial \mathbf{\Gamma}} \cdot \mathbf{\Gamma}\right)$$
$$= \lambda \, \phi(\mathbf{\Gamma}) = \begin{cases} Y\dot{\xi}, & \dot{\xi} > 0 \\ -Y\dot{\xi}, & \dot{\xi} < 0 \end{cases} \tag{17}$$
$$\ge 0$$

in which Euler's homogeneous function theorem is used as $\phi(\mathbf{\Gamma})$ is continuously differentiable and homogeneous of degree one. The satisfaction of the second law of thermodynamics is guaranteed through Eq. (17), which requires careful attention. First, the rates of internal variables ξ and $\nabla^k \xi$ and also $\boldsymbol{\varepsilon}^{\mathrm{tr}}$ and $\nabla^k \boldsymbol{\varepsilon}^{\mathrm{tr}}$ are not independent because they are related by a spatial gradient. This imposes the following restrictions on the transformation surface:

$$\nabla\left(\lambda\frac{\partial\Phi}{\partial\pi}\right) = \lambda\frac{\partial\Phi}{\partial\bar{\pi}^1}, \quad \nabla\left(\lambda\frac{\partial\Phi}{\partial\varsigma^D}\right) = \lambda\frac{\partial\Phi}{\partial\tau^1},$$
$$\nabla\left(\lambda\frac{\partial\Phi}{\partial\bar{\pi}^k}\right) = \lambda\frac{\partial\Phi}{\partial\bar{\pi}^{k+1}}, \quad \nabla\left(\lambda\frac{\partial\Phi}{\partial\tau^k}\right) = \lambda\frac{\partial\Phi}{\partial\tau^{k+1}}. \tag{18}$$

The first one of these constraints is trivially satisfied in light of $\dot{\xi} = \lambda$ and (10). A discussion about such restrictions on the transformation surface can also be found in [64]. Second, as observed in Eq. (17), the dissipation does not explicitly depend on the rate of the gradient of MVF and transformation strain. This is a consequence of the specific form assumed for the transformation surface, Eq. (11), and also the rates of ξ and $\nabla\xi$ as well as $\boldsymbol{\varepsilon}^{\mathrm{tr}}$ and $\nabla\boldsymbol{\varepsilon}^{\mathrm{tr}}$. However, the inhomogeneous distributions of the martensitic volume fraction and the transformation strain in the body contribute to the dissipation by changing the evolution of ξ as a result of their influence on the transformation surface.

The Most General Anisotropic 3rd Degree SMA Gradient Model

In this section, an anisotropic gradient-based constitutive model for SMAs is presented, with $\mathbf{Y}$ considered as the following list of internal variables:

$$\mathcal{X} \equiv \{\boldsymbol{\sigma}, T, \mathbf{Y}\}, \quad \mathbf{Y} \equiv \{\xi, \nabla\xi, \boldsymbol{\varepsilon}^{\mathrm{tr}}, \nabla\boldsymbol{\varepsilon}^{\mathrm{tr}}\}. \tag{19}$$

The derivation of the constitutive response follows the procedure used from the beginning of this section, so only the key equations will be discussed in this section. The form for the Gibbs free energy

$$G = \hat{G}(\mathcal{X}) = \hat{G}(\boldsymbol{\sigma}, T, \xi, \nabla\xi, \boldsymbol{\varepsilon}^{\mathrm{tr}}, \nabla\boldsymbol{\varepsilon}^{\mathrm{tr}})$$
$$= \hat{G}\left(\sigma_{ij}, T, \xi, \xi_{,i}, \varepsilon_{ij}^{\mathrm{tr}}, \varepsilon_{ij,k}^{\mathrm{tr}}\right), \tag{20}$$

is commonly selected to be a polynomial in terms of the variables considered to determine the state of the material. The polynomial function can be truncated depending on the level of coupling desired between the variables and on the physical phenomena as well as the material symmetry sought to be captured [65, 66]. The reference state, $\mathcal{X}_0$, for the free energy is

$$G_0 = \hat{G}(\mathcal{X}_0) = \hat{G}(\mathbf{0}, T_0, 0, \mathbf{0}, \mathbf{0}, \mathbf{0}). \tag{21}$$

Taylor's expansion with respect to this reference state results in a polynomial form for the free energy function

$$\hat{G}(\mathcal{X}) = \hat{G}(\mathcal{X}_0 + (\mathcal{X} - \mathcal{X}_0)) = G_0 + \mathrm{D}_\mathcal{X}G(\mathcal{X}_0)[\mathcal{X} - \mathcal{X}_0]$$
$$+ \frac{1}{2!}\mathrm{D}_\mathcal{X}^2 G(\mathcal{X}_0)[\mathcal{X} - \mathcal{X}_0, \mathcal{X} - \mathcal{X}_0]$$
$$+ \frac{1}{3!}\mathrm{D}_\mathcal{X}^3 G(\mathcal{X}_0)[\mathcal{X} - \mathcal{X}_0, \mathcal{X} - \mathcal{X}_0, \mathcal{X} - \mathcal{X}_0]$$
$$+ \mathcal{O}(\mathcal{X}^4). \tag{22}$$

The most general anisotropic as well as isotropic forms for Gibbs free energy of Eq. (20) is provided in the Appendix "The Most General Anisotropic/Isotropic Gibbs Free Energy for the Gradient-Based SMA Constitutive Model" section.

The constitutive variables and generalized forces are similarly given by Eqs. (3) and (6). The response of the SMA in this section is assumed to be isotropic so that in Eq. (13), the tensors Ξ, being of odd rank, vanish and the tensors Λ can be rewritten in terms of the fundamental isotropic tensors

$$\tilde{\varphi}(\varsigma^D) = H\left(\frac{3}{2}\varsigma_{ij}\varsigma_{ij}\right)^{\frac{1}{2}}, \quad \bar{\varphi}(\bar{\pi}) = -\left(\frac{\bar{\pi}\cdot\bar{\pi}}{\ell_{d_1}^2}\right)^{\frac{1}{2}}, \tag{23}$$

$$\hat{\varphi}(\boldsymbol{\tau}) = -\left(\frac{1}{\ell_{d_2}^2}\tau_{ijk}\tau_{ijk} + \frac{1}{\ell_{d_3}^2}\tau_{ijk}\tau_{kji} + \frac{1}{\ell_{d_4}^2}\tau_{jji}\tau_{kki}\right.$$
$$\left. + \frac{1}{\ell_{d_5}^2}\tau_{iik}\tau_{kjj} + \frac{1}{\ell_{d_6}^2}\tau_{kii}\tau_{kjj}\right)^{\frac{1}{2}}. \tag{24}$$

The six constants ℓ_d can be regarded as *dissipative* length scales. Hence, the forward and reverse transformation surfaces will be

$$\Phi^{\mathrm{fwd}}\left(\varsigma^D, \pi, \bar{\pi}, \boldsymbol{\tau}\right) = \phi - Y,$$
$$\Phi^{\mathrm{rev}}\left(\varsigma^D, \pi, \bar{\pi}, \boldsymbol{\tau}\right) = -\phi - Y$$
$$\phi = \tilde{\varphi}(\varsigma^D) + \pi + \bar{\varphi}(\bar{\pi}) + \hat{\varphi}(\boldsymbol{\tau}). \tag{25}$$

The Simplified SMA Gradient Model Including $\nabla\xi$ and $\nabla\boldsymbol{\varepsilon}^{\mathrm{tr}}$

In this section, a simplified version of the general isotropic SMA model introduced in the previous section is considered. The free energy is decomposed into local and nonlocal parts with no coupling between the local state variables and the nonlocal internal state variable. For the nonlocal part of the free energy, only three quadratic terms including ξ_i and $\varepsilon_{ij,k}^{\mathrm{tr}}$ are retained. The Gibbs free energy, given to its most general form in Eq. (119), is simplified to

$$G(\boldsymbol{\sigma}, T, \boldsymbol{\varepsilon}^{\mathrm{tr}}, \xi, \nabla\xi, \nabla\boldsymbol{\varepsilon}^{\mathrm{tr}}) = G^{\mathrm{local}}(\boldsymbol{\sigma}, T, \boldsymbol{\varepsilon}^{\mathrm{tr}}, \xi)$$
$$+ G^{\mathrm{nonlocal}}(\nabla\xi, \nabla\boldsymbol{\varepsilon}^{\mathrm{tr}}). \tag{26}$$

The local and nonlocal parts are given as follows:

$$G^{\mathrm{local}}(\boldsymbol{\sigma}, T, \boldsymbol{\varepsilon}^{\mathrm{tr}}, \xi) = -\frac{1}{2\rho}\boldsymbol{\sigma} : \mathbf{S}(\xi)\boldsymbol{\sigma} - \frac{1}{\rho}\boldsymbol{\sigma} : [\boldsymbol{\alpha}(\xi)(T - T_0) + \boldsymbol{\varepsilon}^{\mathrm{tr}}]$$
$$+ c(\xi)\left[(T - T_0) - T\ln\left(\frac{T}{T_0}\right)\right]$$
$$- s_0(\xi)T + u_0(\xi) + \frac{1}{\rho}f(\xi),$$

$$G^{\text{nonlocal}}(\nabla\xi, \nabla\boldsymbol{\varepsilon}^{\text{tr}}) = \frac{1}{2\rho}\left(a_1\ell_1^2\,\xi_{,i}\,\xi_{,i} + a_2\ell_2^2\,\varepsilon^{\text{tr}}_{ij,k}\,\varepsilon^{\text{tr}}_{ij,k}\right.$$
$$\left. + a_3\ell_3^2\,\xi_{,i}\,\varepsilon^{\text{tr}}_{ij,j}\right). \tag{27}$$

The tensor $\mathbf{S}(\xi)$ is the phase-dependent isotropic fourth-order compliance tensor defined linearly (rule of mixtures) with respect to the compliance of the austenite $\mathbf{S}^{\text{A}}$ and martensite $\mathbf{S}^{\text{M}}$ phases. $\boldsymbol{\alpha}(\xi)$ is the phase-dependent effective coefficient of thermal expansion. The material parameters c, s_0, and u_0 are the effective specific heat, effective specific entropy at a reference state, and the effective specific internal energy at the reference state, respectively. They also follow a rule of mixtures with respect to their corresponding values for austenite and martensite phases, i.e.,

$$\mathbf{S} = \mathbf{S}^{\text{A}} + \left(\mathbf{S}^{\text{M}} - \mathbf{S}^{\text{A}}\right)\xi, \quad \boldsymbol{\alpha} = \boldsymbol{\alpha}^{\text{A}} + \left(\boldsymbol{\alpha}^{\text{M}} - \boldsymbol{\alpha}^{\text{A}}\right)\xi,$$
$$s_0 = s_0^{\text{A}} + \left(s_0^{\text{M}} - s_0^{\text{A}}\right)\xi, \quad u_0 = u_0^{\text{A}} + \left(u_0^{\text{M}} - u_0^{\text{A}}\right)\xi,$$
$$c = c^{\text{A}} + \left(c^{\text{M}} - c^{\text{A}}\right)\xi. \tag{28}$$

The function $f(\xi)$ is the hardening function attributable to the obstacles inhibiting the propagation of the transformation phase front. $f(\xi)$ can take various forms, from linear to smooth hardening, as discussed in [67] and [68]. The *energetic* length scales ℓ_1 to ℓ_3 are introduced here with their corresponding constant coefficients a_1 to a_3 (with dimensions of energy per volume).

The constitutive variables and thermodynamic forces can be determined using Eqs. (3), (6), and the free energy given in Eq. (27):

$$\boldsymbol{\varepsilon} = -\rho\frac{\partial G}{\partial\boldsymbol{\sigma}} = \mathbf{S}(\xi)\boldsymbol{\sigma} + \boldsymbol{\alpha}(\xi)(T - T_0) + \boldsymbol{\varepsilon}^{\text{tr}},$$
$$s = -\frac{\partial G}{\partial T} = \frac{1}{\rho}\boldsymbol{\sigma}:\boldsymbol{\alpha} + c\ln\left(\frac{T}{T_0}\right) + s_0,$$
$$\varsigma^{\text{D}} = -\rho\frac{\partial G}{\partial\boldsymbol{\varepsilon}^{\text{tr}}} = \boldsymbol{\sigma},$$
$$\pi = -\rho\frac{\partial G}{\partial\xi} = \frac{1}{2}\boldsymbol{\sigma}:\left(\mathbf{S}^{\text{M}} - \mathbf{S}^{\text{A}}\right)\boldsymbol{\sigma} + \boldsymbol{\sigma}:\left(\boldsymbol{\alpha}^{\text{M}} - \boldsymbol{\alpha}^{\text{A}}\right)(T - T_0)$$
$$\qquad - \rho(c^{\text{M}} - c^{\text{A}})\left[(T - T_0) - T\ln\left(\frac{T}{T_0}\right)\right]$$
$$\qquad + \rho\Delta s_0 T - \rho\Delta u_0 - \frac{\partial f}{\partial\xi}, \tag{29}$$
$$\bar{\pi}_i = -\rho\frac{\partial G}{\partial\xi_{,i}} = -a_1\ell_1^2\,\xi_{,i} - \frac{1}{2}a_3\ell_3^2\,\varepsilon^{\text{tr}}_{ij,j},$$
$$\tau_{ijk} = -\rho\frac{\partial G}{\partial\varepsilon^{\text{tr}}_{ij,k}} = -a_2\ell_2^2\,\varepsilon^{\text{tr}}_{ij,k} - \frac{1}{4}a_3\ell_3^2\left(\delta_{ij}\,\xi_{,k} + \delta_{ik}\,\xi_{,j} + \delta_{jk}\,\xi_{,i}\right).$$

Also, the transformation surfaces in Eq. (23) are simplified to

$$\Phi^{\text{fwd}} = \phi - Y, \quad \Phi^{\text{rev}} = -\phi - Y,$$
$$\phi = \tilde{\varphi}(\varsigma^{\text{D}}) + \check{\varphi}(\pi) + \bar{\varphi}(\bar{\pi}) + \hat{\varphi}(\tau),$$

$$\tilde{\varphi}(\varsigma^{\text{D}}) = H\left(\frac{3}{2}\dot{\varsigma}_{ij}\dot{\varsigma}_{ij}\right)^{\frac{1}{2}}, \quad \bar{\varphi}(\bar{\pi}) = -\left(\frac{1}{\ell_{\text{d}_1}^2}\bar{\pi}_i\bar{\pi}_i\right)^{\frac{1}{2}},$$
$$\hat{\varphi}(\tau) = -\left(\frac{1}{\ell_{\text{d}_2}^2}\tau_{ijk}\tau_{ijk}\right)^{\frac{1}{2}}, \tag{30}$$

which can be rewritten by substituting for the generalized forces determined in (29):

$$\tilde{\varphi}(\varsigma^{\text{D}}) = H\left(\frac{3}{2}\acute{\sigma}_{ij}\acute{\sigma}_{ij}\right)^{\frac{1}{2}},$$
$$\bar{\varphi}(\bar{\pi}) = -\frac{1}{\ell_{\text{d}_1}}\left[\left(a_1\ell_1^2\right)^2\,\xi_{,i}\xi_{,i} + \frac{1}{4}\left(a_3\ell_3^2\right)^2\,\varepsilon^{\text{tr}}_{ij,j}\varepsilon^{\text{tr}}_{ik,k}\right.$$
$$\left. + a_1\ell_1^2 a_3\ell_3^2\,\xi_{,i}\varepsilon^{\text{tr}}_{ij,j}\right]^{\frac{1}{2}}, \tag{31}$$
$$\hat{\varphi}(\tau) = -\frac{1}{\ell_{\text{d}_2}}\left[\left(a_2\ell_2^2\right)^2\,\varepsilon^{\text{tr}}_{ij,k}\varepsilon^{\text{tr}}_{ij,k} + \frac{15}{16}\left(a_3\ell_3^2\right)^2\,\xi_{,i}\xi_{,i}\right.$$
$$\left. + a_2\ell_2^2 a_3\ell_3^2\,\xi_{,i}\varepsilon^{\text{tr}}_{ij,j}\right]^{\frac{1}{2}}.$$

In (31), $\acute{\sigma}_{ij}$ are the components of the deviatoric stress and the newly introduced dissipative length scales, ℓ_{d_1} and ℓ_{d_2}, are assumed to be positive. Thus, this model has three nonlocal constants in the Gibbs free energy and two in the transformation surface.

The rates of the internal variables are given by

$$\dot{\varepsilon}^{\text{tr}}_{ij} = \frac{3}{2}H\frac{\acute{\sigma}_{ij}}{\sqrt{\frac{3}{2}\acute{\sigma}_{ij}\acute{\sigma}_{ij}}}\,\dot{\xi},$$
$$\dot{\xi}_{,i} = \frac{1}{\ell_{\text{d}_1}}\frac{a_1\ell_1^2\,\xi_{,i} + \frac{1}{2}a_3\ell_3^2\,\varepsilon^{\text{tr}}_{ij,j}}{\sqrt{\bar{\pi}_i\bar{\pi}_i}}\,\dot{\xi},$$
$$\dot{\varepsilon}^{\text{tr}}_{ij,k} = \frac{1}{\ell_{\text{d}_2}}\frac{a_2\ell_2^2\,\varepsilon^{\text{tr}}_{ij,k} + \frac{1}{4}a_3\ell_3^2\left(\delta_{ij}\,\xi_{,k} + \delta_{ik}\,\xi_{,j} + \delta_{jk}\,\xi_{,i}\right)}{\sqrt{\tau_{ijk}\tau_{ijk}}}\,\dot{\xi}. \tag{32}$$

In order to further investigate the contribution from the gradient of transformation, the model is reduced for a one-dimensional problem where only a single component of stress (σ_{11}) is nonzero. In this case, it is possible to analytically integrate the equation for evolution of transformation strain, (32), such that $\varepsilon^{\text{tr}}_{11} = H\xi$, $\varepsilon^{\text{tr}}_{22} = -\frac{1}{2}H\xi$, $\varepsilon^{\text{tr}}_{33} = -\frac{1}{2}H\xi$. Hence, the surfaces for forward and reverse transformations become

$$H|\sigma_{11}(x)| + \pi - \mathcal{M}\left|\frac{\text{d}\xi(x)}{\text{d}x}\right| = \pm Y, \tag{33}$$

where

$$\mathcal{M} = \frac{a_1\ell_1^2 + \frac{1}{2}Ha_3\ell_3^2}{\ell_{\text{d}_1}}$$
$$+ \frac{1}{\ell_{\text{d}_2}}\sqrt{\frac{3}{2}\left(a_2\ell_2^2\right)^2 H^2 + \frac{15}{16}\left(a_3\ell_3^2\right)^2 + a_2\ell_2^2 a_3\ell_3^2 H^2}. \tag{34}$$

This shows that the 1D transformation surface, as a result of the nonlocal terms, is a first-order differential equation and the nonlocal material constants can be grouped in a *nonlocal parameter*, $\mathcal{M}$. Specifically, because the transformation strain evolves proportional to the MVF, adding the variable $\nabla \boldsymbol{\varepsilon}^{\mathrm{tr}}$ results in additional restrictions and constants without changing the form of the transformation surface that must be finally solved in conjunction with the equilibrium equation. Therefore, $\nabla \boldsymbol{\varepsilon}^{\mathrm{tr}}$ will be excluded from the next simplified SMA gradient model and the focus will be on the coupling of $\nabla \xi$ with other state variables.

The Simplified SMA Gradient Model Including Only Terms with $\nabla \xi$

The general isotropic SMA model introduced in "The Most General Anisotropic 3rd-Degree SMA Gradient Model" section is simplified in this section to include the terms with $\nabla \xi$ but not $\nabla \boldsymbol{\varepsilon}^{\mathrm{tr}}$. The resulting Gibbs free energy cannot be decomposed into two additive parts as before because the cubic terms provide coupling between the local and nonlocal state variables:

$$
\begin{aligned}
G(\boldsymbol{\sigma}, T, \boldsymbol{\varepsilon}^{\mathrm{tr}}, \xi, \nabla \xi) = & -\frac{1}{2\rho} \boldsymbol{\sigma} : \mathbf{S}(\xi) \boldsymbol{\sigma} - \frac{1}{\rho} \boldsymbol{\sigma} : [\boldsymbol{\alpha}(\xi)(T - T_0) + \boldsymbol{\varepsilon}^{\mathrm{tr}}] \\
& + c(\xi) \left[(T - T_0) - T \ln\left(\frac{T}{T_0}\right) \right] \\
& - s_0(\xi) T + u_0(\xi) + \frac{1}{\rho} f(\xi) \\
& + \frac{1}{2\rho} [a_1 \ell_1^2 \, \nabla \xi \cdot \nabla \xi + a_2 \ell_2^2 \, \nabla \xi \cdot \boldsymbol{\sigma} \nabla \xi \\
& + a_3 \ell_3^2 \, T \, \nabla \xi \cdot \nabla \xi + a_4 \ell_4^2 \, \xi \, \nabla \xi \cdot \nabla \xi \\
& + a_5 \ell_5^2 \, \nabla \xi \cdot \boldsymbol{\varepsilon}^{\mathrm{tr}} \nabla \xi + a_6 \ell_6^2 \, \mathrm{tr}(\boldsymbol{\sigma}) \, \nabla \xi \cdot \nabla \xi].
\end{aligned}
$$

(35)

Six energetic length scales are introduced in this model that attribute to the coupling between $\nabla \xi$ and the other state variables.[1] The material constants a_1, a_4, and a_5 have dimensions of energy per volume, while a_3 has dimension of energy per volume per temperature and a_2 and a_6 are dimensionless. Because they originate from the same tensorial constant, there exists a relation between the constants coupling $\boldsymbol{\sigma}$ and $\nabla \xi$, i.e., $a_2 \ell_2^2 = 2 a_6 \ell_6^2$. The local material constants of Eq. (35) follow a similar linear rule of mixtures as listed in Eq. (28).

The constitutive variables and thermodynamic forces can be determined via (3) and (6) using the free energy given in Eq. (35):

$$
\begin{aligned}
\boldsymbol{\varepsilon} = & -\rho \frac{\partial G}{\partial \boldsymbol{\sigma}} \\
= & \mathbf{S}(\xi) \boldsymbol{\sigma} + \boldsymbol{\alpha}(\xi)(T - T_0) + \boldsymbol{\varepsilon}^{\mathrm{tr}} - \frac{1}{2} [a_2 \ell_2^2 \, (\nabla \xi \otimes \nabla \xi) \\
& + a_6 \ell_6^2 \, (\nabla \xi \cdot \nabla \xi) \, \mathbf{I}],
\end{aligned}
$$

$$
s = -\frac{\partial G}{\partial T} = \frac{1}{\rho} \boldsymbol{\sigma} : \boldsymbol{\alpha} + c \ln\left(\frac{T}{T_0}\right) - \frac{1}{2\rho} a_3 \ell_3^2 \, (\nabla \xi \cdot \nabla \xi) + s_0,
$$

$$
\varsigma^{\mathrm{D}} = -\rho \frac{\partial G}{\partial \boldsymbol{\varepsilon}^{\mathrm{tr}}} = \boldsymbol{\sigma} - \frac{1}{2} a_5 \ell_5^2 \, (\nabla \xi \otimes \nabla \xi),
$$

$$
\begin{aligned}
\pi = & -\rho \frac{\partial G}{\partial \xi} = \frac{1}{2} \boldsymbol{\sigma} : (\mathbf{S}^{\mathrm{M}} - \mathbf{S}^{\mathrm{A}}) \boldsymbol{\sigma} + \boldsymbol{\sigma} : (\boldsymbol{\alpha}^{\mathrm{M}} - \boldsymbol{\alpha}^{\mathrm{A}})(T - T_0) \\
& - \rho(c^{\mathrm{M}} - c^{\mathrm{A}}) \left[(T - T_0) - T \ln\left(\frac{T}{T_0}\right) \right] + \rho \Delta s_0 T \\
& - \rho \Delta u_0 - \frac{\partial f}{\partial \xi} - \frac{1}{2} a_4 \ell_4^2 \, (\nabla \xi \cdot \nabla \xi),
\end{aligned}
$$

(36)

$$
\begin{aligned}
\bar{\pi}_i = & -\rho \frac{\partial G}{\partial \xi_i} \\
= & -a_1 \ell_1^2 \, \nabla \xi - a_2 \ell_2^2 \, \boldsymbol{\sigma} \nabla \xi - a_3 \ell_3^2 \, T \, \nabla \xi - a_4 \ell_4^2 \, \xi \, \nabla \xi \\
& - a_5 \ell_5^2 \, \boldsymbol{\varepsilon}^{\mathrm{tr}} \nabla \xi - a_6 \ell_6^2 \, \mathrm{tr}(\boldsymbol{\sigma}) \, \nabla \xi.
\end{aligned}
$$

The transformation surfaces, $\Phi(\boldsymbol{\Gamma})$, are slightly modified for the forward and reverse processes:

$$
\Phi^{\mathrm{fwd}} = \phi^{\mathrm{fwd}} - Y, \quad \Phi^{\mathrm{rev}} = -\phi^{\mathrm{rev}} - Y,
$$

$$
\phi^{\mathrm{fwd/rev}} = \tilde{\varphi}(\varsigma^{\mathrm{D}}) + \check{\varphi}(\pi) + \bar{\varphi}^{\mathrm{fwd/rev}}(\bar{\pi}),
$$

$$
\tilde{\varphi}(\varsigma^{\mathrm{D}}) = H \left(\frac{3}{2} \acute{\varsigma}_{ij} \acute{\varsigma}_{ij} \right)^{\frac{1}{2}}, \quad \bar{\varphi}^{\mathrm{fwd}}(\bar{\pi}) = -\frac{1}{\ell_{\mathrm{d}}^{\mathrm{fwd}}} \sqrt{\bar{\pi} \cdot \bar{\pi}},
$$

$$
\bar{\varphi}^{\mathrm{rev}}(\bar{\pi}) = -\frac{1}{\ell_{\mathrm{d}}^{\mathrm{rev}}} \sqrt{\bar{\pi} \cdot \bar{\pi}},
$$

(37)

which can be rewritten by substituting for the generalized forces determined in Eq. (36). Notice that the deviatoric part of the generalized stress tensor conjugate to transformation strain is $\varsigma^{\mathrm{D}} = \varsigma^{\mathrm{D}} - \frac{1}{3} \mathrm{tr}(\varsigma^{\mathrm{D}})$ and the constant H is the maximum achievable transformation strain. Therefore, $\tilde{\varphi}$ is a Mises-type transformation surface; ℓ_d is the *dissipative* length scale that has different values during the forward or reverse transformations, hence ℓ_d^{fwd} and ℓ_d^{rev}.

The energetic and dissipative intrinsic length scales, ℓ_1 to ℓ_6 plus ℓ_{d}, are combined in the *nonlocal parameter* $\mathcal{M}_i$. $\mathcal{M}_i$ has dimensions of length$\times$ energy/volume. Appearing in the transformation surfaces, $\mathcal{M}_i$ are given by (no summation on i)

$$
\mathcal{M}_i^{\mathrm{fwd/rev}} \equiv \frac{a_i \ell_i^2}{\ell_{\mathrm{d}}^{\mathrm{fwd/rev}}}.
$$

(38)

It will be shown later that such nonlocal parameters appear in the differential equation for the transformation surface as part of the analysis of SMA structures.

---

[1] Notice that the nonlocal length scales here are independent of the ones in Eq. (27).

The rates of the internal variables can thus be obtained as

$$\dot{\varepsilon}^{\mathrm{tr}} = \frac{3}{2} H \frac{\varsigma^{\mathrm{D}}}{\sqrt{\frac{3}{2}\varsigma^{\mathrm{D}}:\varsigma^{\mathrm{D}}}} \dot{\xi},$$ (39a)

$$\nabla\dot{\xi} = -\frac{1}{\ell_{\mathrm{d}}} \frac{\bar{\pi}}{\sqrt{\bar{\pi}\cdot\bar{\pi}}} \dot{\xi},$$ (39b)

in which

$$\varsigma^{\mathrm{D}} = \acute{\sigma} - \frac{1}{6} a_5 \ell_5^2 [3(\nabla\xi \otimes \nabla\xi) - (\nabla\xi\cdot\nabla\xi)\mathbf{I}],$$

$$\varsigma^{\mathrm{D}}:\varsigma^{\mathrm{D}} = \acute{\sigma}:\acute{\sigma} - a_5\ell_5^2\,\nabla\xi\cdot\acute{\sigma}\nabla\xi + \frac{1}{6}[a_5\ell_5^2\,(\nabla\xi\cdot\nabla\xi)]^2.$$ (40)

The rate of MVF is determined using the consistency condition, $\dot{\Phi} = 0$ or (16), repeated here:

$$\dot{\xi} = -\frac{\frac{\partial\phi}{\partial\sigma}:\dot{\sigma} + \frac{\partial\phi}{\partial T}\dot{T}}{\frac{\partial\phi}{\partial\xi} + \frac{\partial\phi}{\partial\varepsilon^{\mathrm{tr}}}:\frac{\partial\phi}{\partial\varsigma^{\mathrm{D}}} + \frac{\partial\phi}{\partial\nabla\xi}\cdot\frac{\partial\phi}{\partial\bar{\pi}} + \frac{\partial\phi}{\partial\nabla\varepsilon^{\mathrm{tr}}}\cdot\frac{\partial\phi}{\partial\tau}} = -\frac{\frac{\partial\phi}{\partial\sigma}:\dot{\sigma} + \frac{\partial\phi}{\partial T}\dot{T}}{\mathcal{A}},$$ (41)

in which we have

$$\frac{\partial\phi}{\partial\sigma} = \frac{3}{2} H \frac{\varsigma^{\mathrm{D}}}{\sqrt{\frac{3}{2}\varsigma^{\mathrm{D}}:\varsigma^{\mathrm{D}}}} + \frac{a_2\ell_2^2}{\ell_{\mathrm{d}}}\left[\frac{\bar{\pi}\otimes\nabla\xi + \frac{1}{2}(\bar{\pi}\cdot\nabla\xi)\mathbf{I}}{\sqrt{\bar{\pi}\cdot\bar{\pi}}}\right]$$
$$+ (\mathbf{S}^{\mathrm{M}} - \mathbf{S}^{\mathrm{A}})\sigma + (\alpha^{\mathrm{M}} - \alpha^{\mathrm{A}})(T - T_0),$$

$$\frac{\partial\phi}{\partial T} = \sigma$$
$$: (\alpha^{\mathrm{M}} - \alpha^{\mathrm{A}}) + \rho(c^{\mathrm{M}} - c^{\mathrm{A}})\ln\left(\frac{T}{T_0}\right) + \rho\Delta s_0$$
$$+ \frac{a_3\ell_3^2}{\ell_{\mathrm{d}_1}}\frac{\bar{\pi}\cdot\nabla\xi}{\sqrt{\bar{\pi}\cdot\bar{\pi}}}$$ (42)

and

$$\mathcal{A} = 3H\frac{a_5\ell_5^2}{\ell_{\mathrm{d}}}\frac{1}{\sqrt{\bar{\pi}\cdot\bar{\pi}}}\frac{1}{\sqrt{\frac{3}{2}\varsigma^{\mathrm{D}}:\varsigma^{\mathrm{D}}}}\bar{\pi}\cdot$$
$$\left[\acute{\sigma}\nabla\xi - \frac{1}{3}a_5\ell_5^2(\nabla\xi\cdot\nabla\xi)\nabla\xi\right] + 2\frac{a_4\ell_4^2}{\ell_{\mathrm{d}}}\frac{1}{\sqrt{\bar{\pi}\cdot\bar{\pi}}}(\bar{\pi}\cdot\nabla\xi)$$
$$-\frac{1}{\ell_{\mathrm{d}}}\Big[(a_1\ell_1^2 + a_3\ell_3^2\,T + a_4\ell_4^2\,\xi + a_6\ell_6^2\,\mathrm{tr}(\sigma))$$
$$+\frac{a_2\ell_2^2\,\sigma + a_5\ell_5^2\,\varepsilon^{\mathrm{tr}}}{\bar{\pi}\cdot\bar{\pi}}:(\bar{\pi}\otimes\bar{\pi})\Big] - \frac{\partial^2 f}{\partial\xi^2}.$$ (43)

Thus, the evolution of MVF, Eq. (41), is affected by the heterogeneity in its distribution coupled with the temperature as well as the stress tensor.

To complete the modeling section, it is necessary to introduce the hardening function, $f(\xi)$. A linear form is chosen for $f(\xi)$ with the relevant material constants b^{M}, b^{A}, μ_1, and μ_2:

$$f(\xi) = \begin{cases} \frac{1}{2}\rho b^{\mathrm{M}}\xi^2 & + (\mu_1 + \mu_2)\xi, & \dot{\xi} > 0 \\ \frac{1}{2}\rho b^{\mathrm{A}}\xi^2 & + (\mu_1 - \mu_2)\xi, & \dot{\xi} < 0 \end{cases}.$$ (44)

In addition, a smooth hardening function can be introduced that results in a continuous transition between the linear elastic and transforming regimes, for example in a uniaxial stress–strain loading plot. Its differentiated form is given by [68]

$$\frac{\partial f(\xi)}{\partial\xi} = \begin{cases} \frac{1}{2}b_1[1 + \xi^{n_1} - (1-\xi)^{n_2}] + b_3, & \dot{\xi} > 0 \\ \frac{1}{2}b_2[1 + \xi^{n_3} - (1-\xi)^{n_4}] - b_3, & \dot{\xi} < 0 \end{cases},$$ (45)

where n_1 to n_4 and b_1 to b_3 are constants.

One-Dimensional Analysis of SMA Structures

The gradient-based SMA constitutive model introduced in "The Simplified SMA Gradient Model Including Only Terms with $\nabla\xi$" section is reduced to consider the case when only one component of stress is nonzero. The proposed model is then used to investigate the response of SMA wires and thin films which are commonly encountered in microactuators. Extension of an SMA homogenous prismatic bar is analytically investigated to demonstrate that the gradient-based model recovers the results of a conventional SMA model in the case of a uniform deformation. In addition, the torsion of SMA wires is studied in which the variation of stress, and hence the martensitic volume fraction, over the diameter of the wire demonstrates the size-dependent features of the gradient-based model. The response of an SMA thin film is also analytically modeled for forward transformation as a beam under pure bending. The loading–unloading response of the SMA beam under pure bending, including the forward and reverse transformations, is numerically modeled.

Extension of SMA Prismatic Bars

The purpose of this section is to show that the proposed nonlocal model yields a local response if the material undergoes a homogeneous deformation. A prismatic bar is assumed to undergo a uniaxial isothermal loading at a temperature above the austenite finish temperature, A_f. The bar is isothermally loaded from austenite so that the transformation to martensite occurs. The schematic for the 1D boundary value problem, considering the symmetry (at the left end) and boundary conditions, is shown in Fig. 1. It
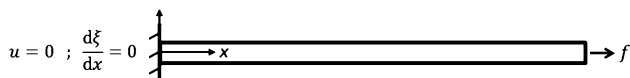
Fig. 1 Schematic including boundary conditions for uniaxial loading of an SMA bar

is assumed that the state of the bar, including the MVF, is symmetric about the left end.

The kinematic and equilibrium equations, along with the equation for decomposition of strain, (36a), result in the following:

$$\varepsilon_x = \frac{\partial u}{\partial x}, \quad \frac{\partial \sigma_x}{\partial x} = 0,$$

$$\varepsilon_x = \frac{\sigma_x}{E(\xi(x))} - \frac{3}{2} a_6 \ell_6^2 \left(\frac{d\xi(x)}{dx}\right)^2 + H\xi(x). \tag{46}$$

The stress does not vary through the length of the bar. The equation for transformation surface, after simplifying Eq. (37) and considering linear hardening for the transformation, can be rewritten as follows:

$$H \left| \sigma_x - \frac{1}{2} a_5 \ell_5^2 \left(\frac{d\xi(x)}{dx}\right)^2 \right| - \frac{1}{2} a_4 \ell_4^2 \left(\frac{d\xi(x)}{dx}\right)^2$$
$$+ \left[\frac{1}{2} \left(\frac{1}{E^M} - \frac{1}{E^A}\right) \sigma_x^2 + \rho \Delta s_0 T - \rho \Delta u_0 - \left(\rho b^M \xi + \mu_1 + \mu_2\right) \right]$$
$$- \frac{1}{\ell_d^{\text{fwd}}} \left| a_1 \ell_1^2 + a_3 \ell_3^2 T + a_4 \ell_4^2 \xi + 3 a_6 \ell_6^2 \sigma_x \right.$$
$$+ a_5 \ell_5^2 H\xi(x) \left| \left| \frac{d\xi(x)}{dx} \right| - Y = 0. \tag{47}$$

Upon satisfying the symmetry boundary condition, $\frac{d\xi}{dx}\big|_{x=0}$ at the left end, the nonlocal contribution due to the derivative of ξ vanishes. Therefore, MVF depends only on the stress with a constant homogeneous distribution along the bar:

$$\xi|_{\text{fwd}} = \frac{H\sigma_x + \frac{1}{2}\left(\frac{1}{E^M} - \frac{1}{E^A}\right)\sigma_x^2 + \rho\Delta s_0 T - \rho\Delta u_0 - (\mu_1 + \mu_2) - Y}{\rho b^M}. \tag{48}$$

A similar analysis for the reverse transformation results in

$$\xi|_{\text{rev}} = \frac{H\sigma_x + \frac{1}{2}\left(\frac{1}{E^M} - \frac{1}{E^A}\right)\sigma_x^2 + \rho\Delta s_0 T - \rho\Delta u_0 - (\mu_1 - \mu_2) + Y}{\rho b^A}. \tag{49}$$

This solution will be used in the Appendix "Experimental Measurement of the SMA Material Properties" section in order to relate the typical experimentally observed SMA properties to the material constants of the developed constitutive model.

Simple Torsion of an SMA Bar

The solution for the torsion of an SMA bar using the nonlocal model of "The Simplified SMA Gradient Model Including Only Terms with $\nabla\xi$" section is obtained in this section. It is assumed that the bar has a circular cross section representing a wire with a diameter of D. An opposite and equal torque T is applied to the ends of the wire (Fig. 2). Also, it is assumed that the wire is initially in a fully austenitic state at a temperature above A_f.

The twist per unit length of the circular bar is denoted by Θ, a constant. This problem is axisymmetric, and it is assumed that the radii on the cross section remain straight. The shear strain in the plane perpendicular to the axis is, therefore, given by

$$u_\theta = r\Theta z, \quad \varepsilon_{\theta z} = \frac{1}{2}\left(\frac{\partial u_\theta}{\partial z} + \frac{1}{r}\frac{\partial u_z}{\partial \theta}\right) = \frac{1}{2}r\Theta. \tag{50}$$

The state of stress in the SMA bar, satisfying the boundary conditions for end moments and traction-free side, is considered to be

$$\sigma_{ij} = 0, \quad ij \neq \theta z. \tag{51}$$

Therefore, the equilibrium equations reduce to

$$\frac{\partial \sigma_{\theta z}}{\partial z} = 0, \quad \frac{1}{r}\frac{\partial \sigma_{\theta z}}{\partial \theta} = 0, \tag{52}$$

which are all satisfied in the case of $\sigma_{\theta z} = \hat{\sigma}_{\theta z}(r)$. This implies that $\xi = \hat{\xi}(r)$. Also, for the regions in the material undergoing elastic loading $\sigma_{\theta z} = Gr\Theta$ where $G = \frac{E}{2(1+\nu)}$.

The relation for total strain, according to the model of "The Simplified SMA Gradient Model Including Only Terms with $\nabla\xi$" section, is given in Eq. (36a). For this problem, after considering that the MVF changes only with the radial coordinate, the θz component of strain tensor can be written as
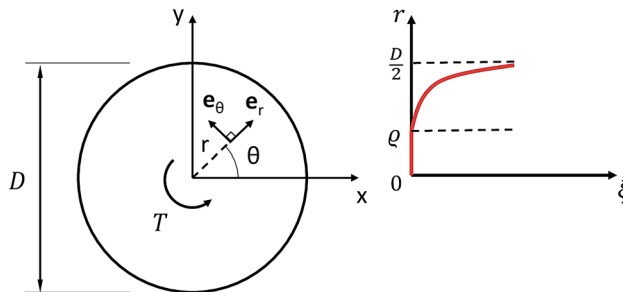
Fig. 2 Simple torsion of an SMA bar with a circular cross section showing, to the right side, a typical distribution for the MVF during forward transformation

$$\varepsilon_{\theta z} = \frac{\sigma_{\theta z}}{2G(\xi)} + \varepsilon_{\theta z}^{\mathrm{tr}} - \frac{1}{2} a_6 \ell_6^2 \left(\frac{\partial \xi}{\partial r} \right)^2. \tag{53}$$

The generalized stress tensor, ς^{D}, conjugate to transformation strain can be determined by simplifying Eq. (36) which finally leads to

$$\varsigma^{\mathrm{D}} : \varsigma^{\mathrm{D}} = 2\sigma_{\theta z}^2 + \frac{2}{3} \left[\frac{1}{2} a_5 \ell_5^2 \left(\frac{\partial \xi}{\partial r} \right)^2 \right]^2. \tag{54}$$

In this sense then

$$\tilde{\varphi}(\varsigma^{\mathrm{D}}) = H \sqrt{3\sigma_{\theta z}^2 + \left[\frac{1}{2} a_5 \ell_5^2 \left(\frac{\partial \xi}{\partial r} \right)^2 \right]^2}, \tag{55}$$

which means, after Eq. (39a), the rate of transformation strain can be given by

$$\dot{\varepsilon}^{\mathrm{tr}} = \frac{3}{2} H \frac{\dot{\xi}}{\sqrt{3\sigma_{\theta z}^2 + \left[\frac{1}{2} a_5 \ell_5^2 \left(\frac{\partial \xi}{\partial r} \right)^2 \right]^2}} \varsigma^{\mathrm{D}}, \tag{56}$$

or in terms of components,

$$\dot{\varepsilon}_{rr}^{\mathrm{tr}} = -2\dot{\varepsilon}_{\theta\theta}^{\mathrm{tr}} = -2\dot{\varepsilon}_{zz}^{\mathrm{tr}} = H\dot{\xi} \frac{-\frac{1}{2} a_5 \ell_5^2}{\sqrt{3\sigma_{\theta z}^2 + \left[\frac{1}{2} a_5 \ell_5^2 \left(\frac{\partial \xi}{\partial r} \right)^2 \right]^2}} \left(\frac{\partial \xi}{\partial r} \right)^2,$$

$$\dot{\varepsilon}_{\theta z}^{\mathrm{tr}} = \frac{3}{2} H \frac{\dot{\xi}}{\sqrt{3\sigma_{\theta z}^2 + \left[\frac{1}{2} a_5 \ell_5^2 \left(\frac{\partial \xi}{\partial r} \right)^2 \right]^2}} \sigma_{\theta z}, \quad \dot{\varepsilon}_{r\theta}^{\mathrm{tr}} = \dot{\varepsilon}_{rz}^{\mathrm{tr}} = 0. \tag{57}$$

For this simple torsion problem, in addition to transformation shear strains, the SMA gradient model results in transformation strain in normal directions as well. Therefore, by the initiation and propagation of transformation in the bar, the cross section may undergo swelling or warping. This point can be used for identification and calibration of the nonlocal parameter $a_5 \ell_5^2$ as discussed in the results section. Equation (36d) reduces to

$$\pi = \frac{1}{2} \left(\frac{1}{G^{\mathrm{M}}} - \frac{1}{G^{\mathrm{A}}} \right) \sigma_{\theta z}^2 + \rho \Delta s_0 T - \rho \Delta u_0 - \frac{\partial f^{\mathrm{fwd/rev}}}{\partial \xi}$$
$$- \frac{1}{2} a_4 \ell_4^2 \left(\frac{\partial \xi}{\partial r} \right)^2. \tag{58}$$

The generalized force $\bar{\pi}$ in component form can be given by

$$\bar{\pi}_r = -\left[a_1 \ell_1^2 + a_3 \ell_3^2 T + a_4 \ell_4^2 \xi + a_5 \ell_5^2 \varepsilon_{rr}^{\mathrm{tr}} + a_6 \ell_6^2 \operatorname{tr}(\boldsymbol{\sigma}) \right] \frac{\partial \xi}{\partial r},$$
$$\bar{\pi}_\theta = -a_5 \ell_5^2 \frac{\partial \xi}{\partial r} \varepsilon_{r\theta}^{\mathrm{tr}}, \bar{\pi}_z = -a_5 \ell_5^2 \frac{\partial \xi}{\partial r} \varepsilon_{rz}^{\mathrm{tr}}. \tag{59}$$

The last two components vanish in light of the relations given for the rate of transformation strain components, Eq. (57c). Hence after Eq. (37) and with $\ell_d = \ell_d^{\mathrm{fwd}} = \ell_d^{\mathrm{rev}}$,

$$\bar{\varphi}(\bar{\pi}) = -\left(\frac{1}{\ell_d^2} \bar{\pi}_i \bar{\pi}_i \right)^{\frac{1}{2}}$$
$$= -\frac{1}{\ell_d} \left| a_1 \ell_1^2 + a_3 \ell_3^2 T + a_4 \ell_4^2 \xi + a_5 \ell_5^2 \varepsilon_{rr}^{\mathrm{tr}} \right| \left| \frac{\partial \xi}{\partial r} \right|. \tag{60}$$

It is reasonable to assume that $\sigma_{\theta z} \geq 0$ and $\frac{\partial \xi}{\partial r} \geq 0$, after which the forward transformation surface becomes

$$\Phi^{\mathrm{fwd}} = H \sqrt{3\sigma_{\theta z}^2 + \left[\frac{1}{2} a_5 \ell_5^2 \left(\frac{\partial \xi}{\partial r} \right)^2 \right]^2} + \frac{1}{2} \left(\frac{1}{G^{\mathrm{M}}} - \frac{1}{G^{\mathrm{A}}} \right) \sigma_{\theta z}^2$$
$$+ \rho \Delta s_0 T - \rho \Delta u_0 - \frac{\partial f^{\mathrm{fwd}}}{\partial \xi}$$
$$- \frac{1}{2} a_4 \ell_4^2 \left(\frac{\partial \xi}{\partial r} \right)^2 - \frac{1}{\ell_d} \left| a_1 \ell_1^2 + a_3 \ell_3^2 T + a_4 \ell_4^2 \xi \right.$$
$$+ a_5 \ell_5^2 \varepsilon_{rr}^{\mathrm{tr}} \left| \frac{\partial \xi}{\partial r} - Y = 0. \tag{61}$$

The boundary conditions required to solve this differential equation are determined by considering the interface between the elastic and transforming regions in the circular bar. The forward transformation begins at the outer surface, while the inner core is still loading elastically. The schematic for the distribution of MVF in this case is shown in Fig. 2 where ϱ denotes the radial position of the interface. Hence,

$$\xi(\varrho) = 0, \quad \frac{\mathrm{d}\xi}{\mathrm{d}r}|_\varrho = 0, \tag{62}$$

due to the zeroth- and first-order continuity of $\xi(r)$. Equations (61) and (62) enable one to find ϱ.

For the reverse transformation,

$$\Phi^{\mathrm{rev}} = -H \sqrt{3\sigma_{\theta z}^2 + \left[\frac{1}{2} a_5 \ell_5^2 \left(\frac{\partial \xi}{\partial r} \right)^2 \right]^2} - \frac{1}{2} \left(\frac{1}{G^{\mathrm{M}}} - \frac{1}{G^{\mathrm{A}}} \right) \sigma_{\theta z}^2$$
$$- \rho \Delta s_0 T + \rho \Delta u_0 + \frac{\partial f^{\mathrm{rev}}}{\partial \xi}$$
$$+ \frac{1}{2} a_4 \ell_4^2 \left(\frac{\partial \xi}{\partial r} \right)^2 + \frac{1}{\ell_d} \left| a_1 \ell_1^2 \right| + a_3 \ell_3^2 T + a_4 \ell_4^2 \xi$$
$$+ a_5 \ell_5^2 \varepsilon_{rr}^{\mathrm{tr}} \frac{\partial \xi}{\partial r} - Y = 0. \tag{63}$$

Similarly, the boundary conditions for (63) are the continuity of $\xi(x)$ and $\frac{\mathrm{d}\xi(r)}{\mathrm{d}r}$ across the interface, $x = \acute{\varrho}$, where the transition form the reverse transformation to elastic unloading takes place.

Fig. 3 Pure bending of an SMA beam showing, to the right side, a typical distribution for the MVF through the (half) thickness during forward transformation
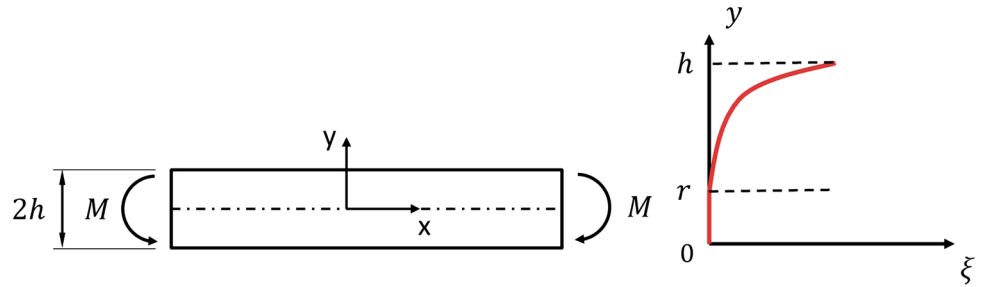
For this problem, the differential equations for the transformation surfaces, Eqs. (61) or (63), must be solved in a coupled fashion with the equations for the rate of transformation strain, (57), and the equation for the decomposition of strain, (53), in order to find the distribution of $\xi(r)$ and $\sigma_{\theta z}(r)$.

Ultimately, it is possible to find the relation between the twist per length in the bar Θ, as the loading parameter, and the applied torque:

$$T(\Theta) = 2\pi \int_0^{\frac{D}{2}} \sigma_{\theta z}(r) r \, dr. \qquad (64)$$

The complexity of the problem here prohibits one from obtaining a closed-form solution. Therefore, such system of differential equations was solved numerically using the finite difference method. The following nondimensional parameters are, notably, introduced for the purpose of results.

$$T^* = \frac{T}{T_s^{\text{fwd}}} = \frac{T}{\frac{\pi}{2}\left(\frac{D}{2}\right)^3 \tau_s^{\text{fwd}}}, \qquad \Theta^* = \frac{\Theta}{\Theta_s^{\text{fwd}}} = \frac{G\frac{D}{2}}{\tau_s^{\text{fwd}}}\Theta. \qquad (65)$$

The nondimensional torque, T^*, and nondimensional angle of twist per length, Θ^*, are obtained by considering the critical torque to start the transformation at the surface of the circular SMA bar, T_s^{fwd}, and the corresponding critical angle of twist per length, Θ_s^{fwd}. The critical shear stress to start forward transformation is given via $\tau_s^{\text{fwd}} = \frac{\sigma_s^{\text{fwd}}}{\sqrt{3}}$.

In the next section, pure bending of an SMA beam is considered with its solution for when only one of the length scales, $a_1\ell_1^2$, is nonzero.

Pure Bending of an SMA Beam

The deflection of SMA thin film actuators is investigated in this section. For simplicity, all of the nonlocal length scales in the nonlocal SMA constitutive model except $a_1\ell_1^2$ are considered to vanish. It will be shown that even in this simplified version, the SMA gradient model is capable of capturing the size effect in the response.

It is considered that a slender prismatic SMA beam (with a width of W and thickness of $2h$) undergoes a state of pure bending (Fig. 3) such that the Euler–Bernoulli assumptions hold, i.e., the planar sections remain planar and perpendicular to the neutral axis and rotate with reasonably small slopes. Hence,

$$u(y) = -y\tan(\theta) = -y\frac{\partial v(x)}{\partial x}, \quad \varepsilon_x = -y\frac{\partial^2 v}{\partial x^2} = \kappa y, \quad (66)$$

where κ is the curvature of bending deformation. We assume that the only nonzero component of stress is the normal stress on the cross section, σ_x. In addition to that, the decomposition of strain in the axial x direction using the SMA gradient-based model results in

$$\frac{\partial \sigma_x(y)}{\partial x} = 0, \quad \varepsilon_x = \frac{\sigma_x}{E(\xi(x))} + H\xi(x). \qquad (67)$$

First, the start of forward transformation on the outermost layers of the beam under loading is investigated to find an analytical nondimensional response. The state of the beam including stress, strain, and martensitic volume fraction reduces to a one-dimensional problem with variation along the thickness of the beam or y direction; $\sigma_x(y)$, $\varepsilon_x(y)$, $\xi(y)$. The normal stress in the x direction is the only nonzero component of the stress tensor. The deformation is isothermal at a temperature above A_f, thus the loading begins, while the beam is entirely in an austenite state. The loading causes the transformation to martensite to begin at the outer layers and propagate towards the neutral axis. The interface between transforming outer layers and the elastic austenitic inner layers of the beam is denoted by r on the y-axis, $\xi(y = r) = 0$. Because the stress state is 1D, the forward transformation equation for the SMA gradient-based model becomes

$$H|\sigma_x(y)| + \left[\frac{1}{2}\left(\frac{1}{E^M} - \frac{1}{E^A}\right)\sigma_x(y)^2 + \rho\Delta s_0^{\text{fwd}}T - \rho\Delta u_0 \right.$$
$$\left. - \left(\rho b^M \xi(y) + \mu_1 + \mu_2\right)\right]$$
$$- \mathcal{M}_1^{\text{fwd}}\left|\frac{d\xi(y)}{dy}\right| - Y = 0, \qquad (68)$$

in which a linear hardening function is used. Also, $\mathcal{M}_1^{\text{fwd/rev}} \equiv \frac{a_1\ell_1^2}{\ell_d}$, after the definition given in Eq. (38). Considering the symmetry about the neural axis, only the top portion of the beam is studied hence $\sigma_x \geq 0$ and,

therefore, $\frac{d\xi}{dy} \geq 0$. The material constants in the above equation can be collected together to reach

$$H\sigma_x(y) + A\sigma_x(y)^2 + B\xi(y) - \mathcal{M}_1^{\text{fwd}}\frac{d\xi(y)}{dy} + D = 0, \quad (69)$$

with the newly introduced constants of forward transformation with linear hardening (Table 2),

$$A = \frac{1}{2}\left(\frac{1}{E^M} - \frac{1}{E^A}\right), \quad B = -\rho b^M = -H\left(\sigma_f^{\text{fwd}} - \sigma_s^{\text{fwd}}\right),$$
$$D = \rho\Delta s_0^{\text{fwd}}T - \rho\Delta u_0 - (\mu_1 + \mu_2) - Y = -H\sigma_s^{\text{fwd}}. \quad (70)$$

Equation (69) is a differential equation that requires a boundary condition for its solution. Obviously, the condition at the boundary of the elastic-transforming regions or, $\xi(r) = 0$, must be satisfied. But this relation gives the constant of integration for (69) as a function of r which is still an unknown parameter of loading. Implicit in the analysis of the nonlocal constitutive model developed here is the fact that the forward or reverse transformation front entails a smooth transition from the elastic to the transforming region, i.e., all the state variables, including $\nabla\xi(\mathbf{X})$, must be continuous across the interface. In a polycrystalline SMA sample, the transformation front progresses by the martensitic transformation first occurring in the grains with favorable crystalline direction to the stress direction and then gradually transitioning to the elastic behavior in the transformed regions. This can be simulated by having a smooth scalar field variable, ξ. Thus, at the elastic-transforming boundary interface,

$$\xi|_{y=r} = 0, \quad \frac{d\xi}{dy}\Big|_{y=r} = 0. \quad (71)$$

Satisfaction of the boundary conditions in (71) gives a relation for both the constant of integration in the ODE of (69) and the location of the elastic-transforming boundary r with respect to the current curvature, κ, as the loading parameter. The transition to the fully martensitic region is not smooth in terms of ξ, i.e., $\nabla\xi$ will not be continuous across the transformation finish front.

Interestingly, using (71) in the equation for forward transformation surface (69) results in

$$[H\sigma_x(y) + D]_{y=r} = 0 \quad \Rightarrow \quad \sigma_x(r) = \sigma_s^{\text{fwd}} \quad \Rightarrow$$
$$r = \frac{\sigma_s^{\text{fwd}}}{E^A\kappa}. \quad (72)$$

The loading here is proportional, so it is possible to directly integrate the rate of evolution of transformation strains

$$\dot{\varepsilon}_x^{\text{tr}} = H\dot{\xi}\,\text{sgn}(\sigma_x), \quad \dot{\varepsilon}_y^{\text{tr}} = \dot{\varepsilon}_z^{\text{tr}} = -\frac{1}{2}\dot{\varepsilon}_x^{\text{tr}}, \quad (73)$$

to obtain

$$\varepsilon_x^{tr}(y) = H\xi(y), \quad \sigma_x(y) = E(\xi)(\kappa y - H\xi(y)). \quad (74)$$

In order to attain a closed-form solution, the difference in the elastic modulus of austenite and martensite is ignored, i.e., $E = E^M = E^A$ which means $A = 0$.

The differential equation in Eq. (69) (with $A = 0$) is linear first order with a general solution given by

$$\xi(y) = -\frac{\frac{\mathcal{M}_1^{\text{fwd}}}{EH}}{\left(\frac{B}{EH} - H\right)^2}\kappa - \frac{\frac{D}{EH} + \kappa y}{\frac{B}{EH} - H} + \zeta e^{\left(\frac{\frac{B}{EH} - H}{\mathcal{M}_1^{\text{fwd}}}y\right)}, \quad (75)$$

which after applying the boundary condition in Eq. (71) and replacing ζ yields

$$\xi(y) = -\frac{D + HE\kappa y}{B - EH^2} - \frac{HE\mathcal{M}_1^{\text{fwd}}}{(B - EH^2)^2}\kappa\left(1 - e^{\frac{B - EH^2}{\mathcal{M}_1^{\text{fwd}}}(y-r)}\right). \quad (76)$$

The first term on the R.H.S. of the above equation pertains to the local part of the model and gives a bilinear distribution in ξ, while the second term is the exponential contribution from the nonlocal part. The moment, M, applied to the beam section corresponding to the current curvature, κ, can be readily calculated from

$$M(\kappa) = \int_{-h}^{h} W\sigma_x(y)y\,dy = 2W\int_0^h \sigma_x(y)y\,dy. \quad (77)$$

Using Eqs. (74) and (76) in the integration of (77) results in

$$M^* = \kappa^* + \frac{3}{\beta - 1}\left\{\frac{1}{2}\frac{\mathcal{M}_1^{\text{fwd}*}}{\beta - 1}\left(\kappa^* - \frac{1}{\kappa^*}\right) + \frac{1}{3}\kappa^* + \frac{1}{6}\left(\frac{1}{\kappa^*}\right)^2 - \frac{1}{2}\right.$$
$$- \left(\frac{\mathcal{M}_1^{\text{fwd}*}}{\beta - 1}\right)^2$$
$$\left.\kappa^*\left[\left(1 - \frac{\mathcal{M}_1^{\text{fwd}*}}{\beta - 1}\right)e^{\frac{\beta-1}{\mathcal{M}_1^{\text{fwd}*}}\left(1 - \frac{1}{\kappa^*}\right)} - \left(\frac{1}{\kappa^*} + \frac{\mathcal{M}_1^{\text{fwd}*}}{\beta - 1}\right)\right]\right\}, \quad (78)$$

which is the equation of moment-curvature in a nondimensional form for pure bending of an SMA beam.

The nondimensional parameters introduced in (78) are now defined. The nondimensional moment, M^*, and nondimensional curvature are obtained based on a normalization using the critical moment and curvature to start the forward transformation, M_s^{fwd} and κ_s^{fwd}:

$$M^* = \frac{M}{M_s^{\text{fwd}}} = \frac{M}{\frac{2}{3}Wh^2\sigma_s^{\text{fwd}}}, \quad \kappa^* = \frac{\kappa}{\kappa_s^{\text{fwd}}} = \frac{Eh}{\sigma_s^{\text{fwd}}}\kappa. \quad (79)$$

In addition, $\mathcal{M}^{\text{fwd}*}$ can be defined as the nondimensional form of the nonlocal parameter as well as the nondimensional hardening, β

$$\beta = \frac{B}{EH^2}, \quad \mathcal{M}_1^{\text{fwd}*} = \frac{1}{EH^2}\frac{\mathcal{M}_1^{\text{fwd}}}{h} \quad \text{or}$$
$$\frac{\mathcal{M}_1^{\text{fwd}*}}{\beta - 1} = \frac{1}{h}\frac{\mathcal{M}_1^{\text{fwd}}}{B - EH^2}. \tag{80}$$

It is also helpful to write down the local nondimensional response of the SMA beam

$$M^* = \kappa^* + \frac{3}{\beta - 1}\left\{\frac{1}{3}\kappa^* + \frac{1}{6}\left(\frac{1}{\kappa^*}\right)^2 - \frac{1}{2}\right\}. \tag{81}$$

Comparing Eqs. (78) and (81) reveals the size effect in the response for pure bending of SMA beams. The first term on the R.H.S of both equations pertains to the elastic portion of the SMA behavior, i.e., $M^* = \kappa^*$. This portion is size-independent because the nonlocality is only considered for the *inelastic* part of the response. Nonetheless, the remainder of Eq. (81) is also size-independent, because there is no effect from the thickness of the beam in the hardening parameter β. Therefore, the moment-curvature response using the conventional local model cannot capture the size effect in pure bending of SMAs because it can be fully normalized with respect to the thickness h. The response according to the nonlocal constitutive model, on the other hand, is able to capture the size effect, due to the nondimensional nonlocal parameter $\mathcal{M}_1^{\text{fwd}*}$ which *depends* on the thickness; $\mathcal{M}_1^{\text{fwd}*} = \frac{1}{EH^2}\frac{\mathcal{M}^{\text{fwd}}}{h}$. Hence by changing $\mathcal{M}_1^{\text{fwd}*}$ either through $\mathcal{M}_1^{\text{fwd}}$ or h, the moment-curvature response of the beam changes.

The analysis of the reverse transformation can be performed by following similar steps. Upon unloading, the reverse transformation first begins where its corresponding differential equation is first satisfied:

$$-H\sigma_x(y) - A\sigma_x(y)^2 + B'\xi(y) + \mathcal{M}_1^{\text{rev}}\frac{d\xi(y)}{dy} + D' = 0, \tag{82}$$

$$\begin{aligned} B' &= \rho b^{\text{A}} = H(\sigma_s^{\text{rev}} - \sigma_f^{\text{rev}}), \\ D' &= -\rho\Delta s_0^{\text{rev}}T + \rho\Delta u_0 + (\mu_1 - \mu_2) - Y = H\sigma_f^{\text{rev}}. \end{aligned} \tag{83}$$

It can be shown that the reverse transformation begins at the outer layer and propagates towards the neutral axis with progression of unloading. The interface between the region of reverse transformation and the region where unloading takes place elastically is denoted by r'. The reverse transformation finishes when r' reaches r. Although extremely

lengthy, it is possible to develop a closed-form analytical response for unloading step as well. Equation (82) can be integrated considering the continuity of ξ and $\frac{d\xi(y)}{dy}$ at $y = r'$.

However, a numerical algorithm was developed to solve both of the ODEs in (69) and (82) for forward and reverse transformations using the finite difference method. The thickness of the beam was discretized through which the differential equations were integrated using a forward Euler method. The imposed curvature was considered as the loading parameter, and having solved for the MVF at each discretization point, the stress was found using the decomposition of strain in Eq. (67) and integrated over the thickness, Eq. (77), to obtain the corresponding moment. The results of this analysis are presented in the next section.

Results and Discussions

The modeling results developed in the previous section for pure bending and simple torsion of the SMA structures are now presented. Where nondimensional numbers could not be realized, the material constants listed in Table 1 are used for the local part of the SMA constitutive model. They are selected from the properties of a typical shape memory alloy.

It is noted that the nonlocal model in "The Simplified SMA Gradient Model Including Only Terms with $\nabla\xi$" section contains six energetic length scales. The analysis for simple torsion included all, whereas only one was considered for the pure bending analysis.

The effect of the nonlocal parameter $\mathcal{M}_1^{\text{fwd/rev}} = \frac{a_1\ell_1^2}{\ell_d^{\text{fwd/rev}}}$ [with the units of (mm MPa)] on the pure bending of the SMA beam is investigated. It is assumed that $\mathcal{M}_1 = \mathcal{M}_1^{\text{fwd}} = \mathcal{M}_1^{\text{rev}}$. For an SMA beam with $W = 1$ (mm), the plot of the nondimensional moment vs. curvature is illustrated in Fig. 4.

The case of $\mathcal{M}_1^* = 0$ pertains to the local conventional model. Inclusion of the nonlocal effect in the model results in a macroscopic hardening effect. More energy is dissipated due to the existence of the nonlocal term and therefore more flexural work is required to achieve a given curvature. This plot can act as the characterization point for the nonlocal parameter. Given a fixed thickness and known

Table 1 SMA material properties used in the presentation of the modeling results

| Thermoelastic constants	Transformation properties			
$E^{\text{A}} = 85{,}000$ MPa	$A_s = 27\ °C$	$A_f = 47\ °C$	$C_A = 10\left(\frac{\text{MPa}}{°C}\right)$	$H = 5.5\%$
$E^{\text{M}} = 75{,}000$ MPa	$M_s = -53\ °C$	$M_f = -73\ °C$	$C_A = 10\left(\frac{\text{MPa}}{°C}\right)$	$T = 67\ °C$
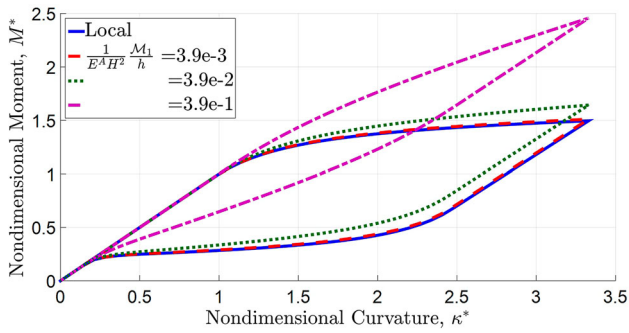
Fig. 4 Effect of the nonlocal parameter $\mathcal{M}_1^* = \frac{1}{E^A H^2} \frac{\mathcal{M}_1}{D}$ on the nondimensional moment–curvature response of an SMA beam with $W = 1$
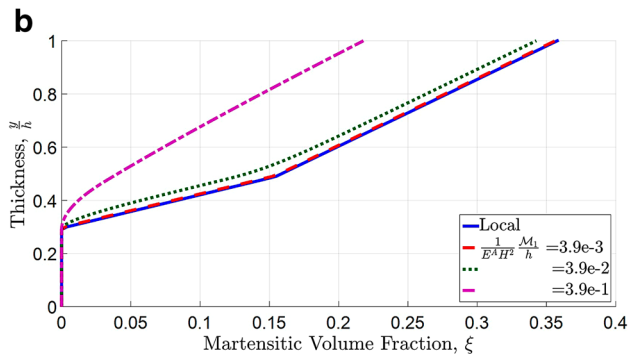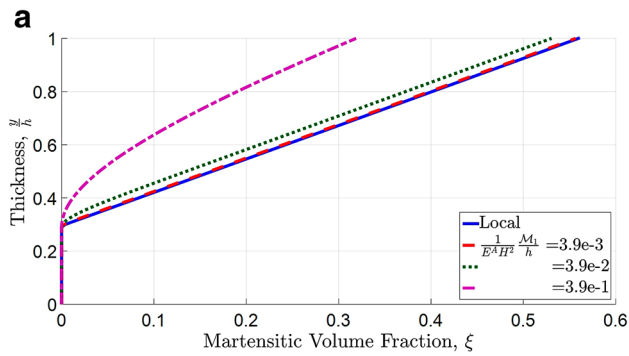
Fig. 5 Effect of the nonlocal parameter $\mathcal{M}_1^* = \frac{1}{E^A H^2} \frac{\mathcal{M}_1}{D}$ on the distribution of MVF in the SMA beam of Fig. 4. **a** During loading at $\kappa^* = 3.3$, **b** during unloading at $\kappa^* = 1.6$

tensile SMA properties, it is possible to determine the value of the nonlocal parameter by calibrating the nondimensional flexural response of an SMA sample.

Figure 5 shows the variation in the distribution of MVF for this SMA beam under a curvature of $\kappa^* = 3.3$ during loading and that of $\kappa^* = 1.6$ during unloading. Through the nonlocal model, the inelastic energy generated contributes both to the martensitic transformation and also its gradient. The higher the nonlocal effect, the lower the total amount of transformation. Note that the conventional local model ($\mathcal{M}_1^* = 0$) results in a linear distribution of ξ, as
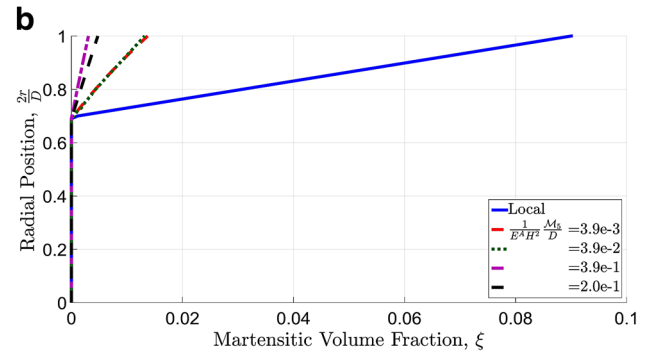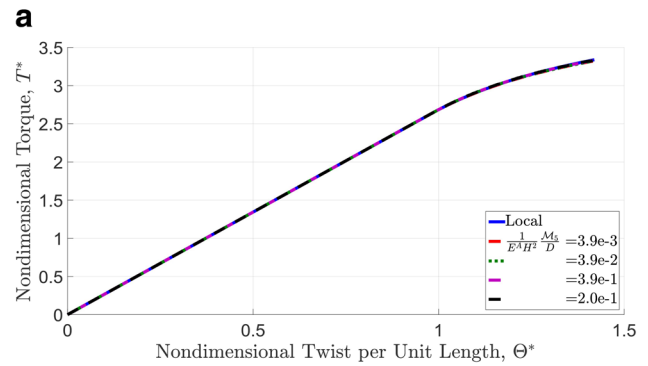
Fig. 6 Effect of the nonlocal parameter $\mathcal{M}_5^* = \frac{1}{E^A H^2} \frac{\mathcal{M}_5}{D}$ on the nondimensional torque–twist response of an SMA bar under simple torsion

expected from the linear choice for the hardening function. Also, note that the SMA gradient model does not capture the size effect in the elastic regime.

The model used for the pure bending analysis included only one nonlocal parameter, $\mathcal{M}_1$. That corresponds to considering only the quadratic terms in the free energy for the gradient of MVF and showed the capability in capturing the size effect.

The SMA nonlocal model, in general, has 6 parameters coupling the gradient of MVF with the rest of the state variables, Eq. (35). In what follows, the effect of coupling between the two internal variables is investigated. As shown in Eq. (35), the nonlocal energetic length scale, $a_5 \ell_5^2$ with dimensions of [mm^2 MPa], acts as the coupling term between the gradient of MVF and the transformation strain tensor. The significance of this coupling appears when the solution for simple torsion of an SMA bar is considered.

According to (35), axial, circumferential, and radial transformation strains are generated under simple torsion in addition to the common shear term for the transformation strain. Here then we assume $\mathcal{M}_5 = \mathcal{M}_5^{\text{fwd}} = \mathcal{M}_5^{\text{rev}}$ where $\mathcal{M}_5^{\text{fwd/rev}} = \frac{a_5 \ell_5^2}{\ell_d^{\text{fwd/rev}}}$.

Figures 6 and 7 show the response of an SMA bar with a diameter $D = 1$ mm under pure torsion. The only nonzero nonlocal parameter here is $\mathcal{M}_5$, and according to Fig. 6a,
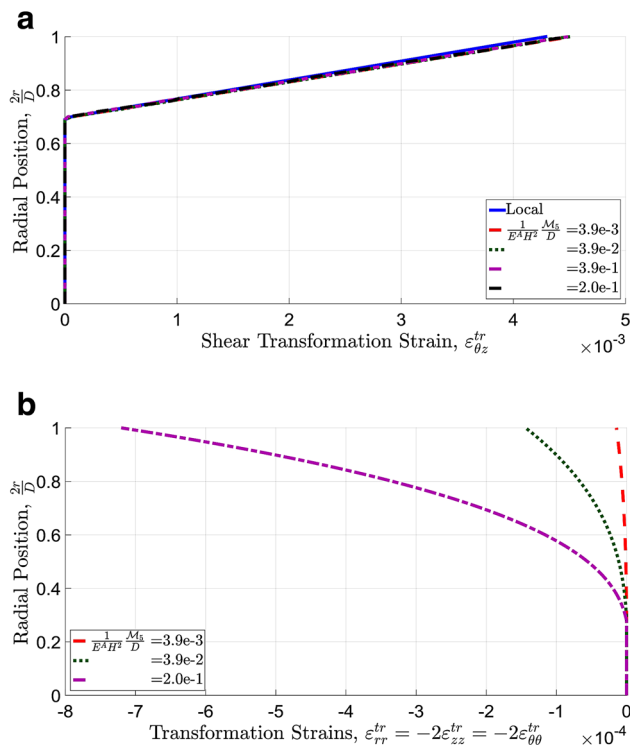
Fig. 7 Effect of the nonlocal parameter $\mathcal{M}_5^* = \frac{1}{E^A H^2} \frac{\mathcal{M}_5}{D}$ on the distribution of shear and normal transformation strains in simple torsion of an SMA bar

the nonlocal torque–twist response is close to the local one. The distribution of the martensite volume fraction across the diameter, however, varies by changing the nonlocal parameter. This can be clarified by referring to Fig. 7a, where the shear transformation strain is plotted, in which the local and nonlocal responses also differ insignificantly. According to the equation for the transformation surface, Eq. (61), and what has been observed so far, the nonlocal effect tends to inhibit the transformation as plotted in Fig. 6b. But this will not affect the evolution of the shear transformation strain, as shown in Eq. (57), where a decrease in the MVF is counteracted by a larger nonlocal parameter. On the other hand, the nonlocal parameter gives rise to transformation strains in the radial, hoop, and axial directions, Eq. (57) and Fig. 7b. Therefore, a pure torsion experiment cannot be used to calibrate such a nonlocal parameter unless the experimentalist obtains the distribution of MVF or measures the axial/radial transformation strain, if any.

The ability of the proposed nonlocal model to capture the size effect in the bending response of SMA thin films as well as simple torsion of SMA bars was demonstrated. Beams of various thicknesses from the same SMA show different responses. As the thickness of the SMA beam increases, its response becomes closer to the results of the local model. This is supported by the fact that by moving away from the small scale, the transformation behavior becomes independent of the size and can be nondimensionalized based on the mode and complexity of the deformation. SMA beams of smaller thickness, according to Fig. 4, show a hardening effect demonstrating the experimentally established "smaller is stronger."

Conclusions

The response of SMA structures with submicron dimensions deviate from that of the bulk material. Conventional constitutive models, due to lack of intrinsic length scales, cannot capture such an experimentally observed size effect.

In the current work, a thermodynamically consistent constitutive model for the behavior of SMAs is developed that includes additional internal variables for the spatial gradients of martensite volume fraction and the transformation strain. The gradient-based constitutive model includes various energetic and dissipative length scales that can be calibrated experimentally. The length scales contribute to additional hardening in the structural SMA response.

A boundary value problem, defined by such model, contains the differential equation for the transformation surface including the gradients of internal variables as well as the equilibrium equation. The boundary value problems are solved analytically and, where impossible, numerically for simple extension of an SMA bar, pure bending of an SMA beam, and simple torsion of an SMA cylindrical bar. The gradient model attains to the results of a conventional SMA model in the case of a homogeneous problem. Unlike pure bending of a beam, the solution to simple torsion of an SMA bar featured a coupling between the gradient of MVF and transformation strain that resulted in transformation strains in normal directions in addition to transformation shear strains. Moreover, for structures that include gradients of stress, the most simplified version of the nonlocal SMA constitutive model, containing only one nonlocal parameter, can capture the size effect demonstrating a stronger response for smaller sizes.

Appendices

Fundamentals

The independent state variables, assumed in the development of the strain gradient constitutive model, include the stress $\boldsymbol{\sigma}$, the absolute temperature T, and a set of generalized state variables $\mathbf{Y}$ related to the microstructure of the material. The definition of $\mathbf{Y}$ as the set of internal variables or internal degrees of freedom is discussed later in this

section. The dependent state variables that change in response to the changes in the independent state variables are assumed to be

$$\boldsymbol{\varepsilon} = \hat{\varepsilon}(\boldsymbol{\sigma}, T, \mathbf{Y}), \quad \mathbf{q} = \hat{q}(\boldsymbol{\sigma}, T, \mathbf{Y}), \quad s = \hat{s}(\boldsymbol{\sigma}, T, \mathbf{Y})$$

$$\boldsymbol{\mu}^s = \hat{\mu}^s(\boldsymbol{\sigma}, T, \mathbf{Y}), \quad \boldsymbol{\mu}^b = \hat{\mu}^b(\boldsymbol{\sigma}, T, \mathbf{Y}), \quad G = \hat{G}(\boldsymbol{\sigma}, T, \mathbf{Y}).$$
(84)

The response function $\boldsymbol{\varepsilon}$ is the infinitesimal strain, $\mathbf{q}$ is the heat flux vector, s is the specific entropy, G is the specific Gibbs free energy, and $\boldsymbol{\mu}^s$ and $\boldsymbol{\mu}^b$ are the generalized forces conjugate to $\mathbf{Y}$ through surface and body action, respectively. The specific Gibbs free energy can be related to the specific internal energy, u, through the Legendre transformation.

Tensors of any arbitrary rank can constitute the set of generalized state variables $\mathbf{Y} = \{\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, \ldots\}$ and their corresponding generalized forces $\boldsymbol{\mu}^s = \{\boldsymbol{\mu}_1^s, \boldsymbol{\mu}_2^s, \boldsymbol{\mu}_3^s, \ldots\}$ and $\boldsymbol{\mu}^b = \{\boldsymbol{\mu}_1^b, \boldsymbol{\mu}_2^b, \boldsymbol{\mu}_3^b, \ldots\}$.[2] The variables $\mathbf{Y}_i$ are related to the mechanisms of microstructural rearrangement; for example, plastic deformation due to the movement of dislocations, twinning of crystals, or slip of grain boundaries or transformation strain due to the stress/temperature-induced martensitic transformation. For the current study of an SMA gradient theory, the martensitic transformation is considered as the mechanism of microstructural changes. Therefore, $\mathbf{Y}$ includes the MVF, ξ, and its spatial gradients, $\nabla\xi, \ldots, \nabla^p\xi$, as well as the martensitic transformation strain tensor, $\boldsymbol{\varepsilon}^{\text{tr}}$, and its spatial gradients, $\nabla\boldsymbol{\varepsilon}^{\text{tr}}, \ldots, \nabla^q\boldsymbol{\varepsilon}^{\text{tr}}$. We assume here that the $\mathbf{Y}_i$ are independent, but later we will see that some $\mathbf{Y}_i$ are related by the evolution equations.

The conventional conservation laws of continuum mechanics as for the principle of conservation of mass, linear momentum, and angular momentum as well as the second law of thermodynamics in the form of the Clausius–Duhem inequality are considered [69]. The localization procedure, in going from the integral to the PDE form, assumes a continuous and smooth function for all of the fields integrated.

Generalized Principle of Conservation of Energy

The first law of thermodynamics states that the rate of change of kinetic and internal energy for a continuum body is equal to the rate at which external mechanical work is exerted on that body as well as external heat added to it. The external work is performed partly due to the contact and body forces. In its generalized form herein, it is assumed that the generalized state variables are capable of performing work through generalized surface and body forces. No physical justification is given here. However,

one can argue that since the generalized state variables contribute to the internal energy, they must do external work in the balance of energy:

$$\frac{D}{Dt}\left(\int_\Omega \frac{1}{2}\rho\mathbf{v}\cdot\mathbf{v}\,\mathrm{d}V + \int_\Omega \rho\mathbf{u}\,\mathrm{d}V\right)$$
$$= \int_{\partial\Omega}\mathbf{t}\cdot\mathbf{v}\,\mathrm{d}S + \int_\Omega \mathbf{b}\cdot\mathbf{v}\,\mathrm{d}V + \int_{\partial\Omega}-\mathbf{q}\cdot\mathbf{n}\,\mathrm{d}S + \int_\Omega \rho r\,\mathrm{d}V$$
$$+ \int_{\partial\Omega}(\boldsymbol{\mu}_i^s\cdot\mathbf{n})\cdot\dot{\mathbf{Y}}_i\,\mathrm{d}S + \int_\Omega \boldsymbol{\mu}_i^b\cdot\dot{\mathbf{Y}}_i\,\mathrm{d}V,$$
(85)

where $\frac{D}{Dt}$ and $(\dot{\cdot})$ denote the material (or Lagrangian) time derivative, V is the volume of the body with mass density ρ, and $\mathbf{v} = \dot{\mathbf{u}}$ being the velocity of the material point. Also, $\mathbf{b}$ is the body force per unit volume, $\mathbf{t}$ is the surface traction, and $\boldsymbol{\sigma}$ is the symmetric Cauchy stress tensor which is related to the surface traction operating on a surface with a unit normal $\mathbf{n}$ through $\boldsymbol{\sigma}\mathbf{n} = \mathbf{t}$. In addition, r is the body heat source, and $\boldsymbol{\mu}_i^s$ and $\boldsymbol{\mu}_i^b$ are the generalized surface and body forces, respectively. Equation (85) can be localized using the equilibrium equation, $\nabla\cdot\boldsymbol{\sigma} + \mathbf{b} = \rho\dot{\mathbf{v}}$, and the divergence theorem to get the local form of the generalized principle of conservation of energy,

$$\rho\dot{\mathbf{u}} = \boldsymbol{\sigma} : \dot{\varepsilon} - \nabla\cdot\mathbf{q} + \rho r + \boldsymbol{\mu}_i^s\cdot\nabla\dot{\mathbf{Y}}_i + \left[\mathrm{Div}(\boldsymbol{\mu}_i^s) + \boldsymbol{\mu}_i^b\right]\cdot\dot{\mathbf{Y}}_i.$$
(86)

For the special case of $\mathbf{Y} \equiv \{\xi, \boldsymbol{\varepsilon}^{\text{tr}}, \nabla\xi, \nabla\boldsymbol{\varepsilon}^{\text{tr}}\}$, $\boldsymbol{\mu}^s \equiv \{\mu_1^s, \boldsymbol{\mu}_2^s, \boldsymbol{\mu}_3^s, \boldsymbol{\mu}_4^s\}$, and $\boldsymbol{\mu}^b \equiv \{\mu_1^b, \boldsymbol{\mu}_2^b, \boldsymbol{\mu}_3^b, \boldsymbol{\mu}_4^b\}$, the local form of the generalized principle of conservation of energy will be

$$\rho\dot{\mathbf{u}} = \boldsymbol{\sigma} : \dot{\varepsilon} - \nabla\cdot\mathbf{q} + \rho r + \left[\mathrm{Div}(\boldsymbol{\mu}_1^s) + \mu_1^b\right]\dot{\xi}$$
$$+ \left[\mathrm{Div}(\boldsymbol{\mu}_2^s) + \boldsymbol{\mu}_2^b\right] : \dot{\varepsilon}^{\text{tr}} + \left[\mathrm{Div}(\boldsymbol{\mu}_3^s) + \boldsymbol{\mu}_3^b + \boldsymbol{\mu}_1^s\right]\cdot\nabla\dot{\xi}$$
$$+ \left[\mathrm{Div}(\boldsymbol{\mu}_4^s) + \boldsymbol{\mu}_4^b + \boldsymbol{\mu}_2^s\right]\cdot\nabla\dot{\varepsilon}^{\text{tr}} + \boldsymbol{\mu}_3^s : \nabla\nabla\dot{\xi} + \boldsymbol{\mu}_4^s\cdot\nabla\nabla\dot{\varepsilon}^{\text{tr}}.$$
(87)

Constitutive Equations

External thermodynamic state variables perform work on the surface of the body. Internal thermodynamic state variables (ISV), on the other hand, are identifiable and measurable, but cannot be linked to any external force variable. Therefore, ISVs do not perform work, i.e., they do not appear in the mechanical work in the first law of thermodynamics [70]. The constitutive equations give relationships between the dependent state variables at time t and the independent external state variables and the ISVs. These state equations must be accompanied by evolution laws for the internal variables.

The generalized state variables, $\mathbf{Y}$, contribute to the internal energy through their surface and body work

---

[2] It is assumed here that if $\mathbf{Y}_i \in \mathcal{T}^m$, the space of mth-rank tensors, then $\boldsymbol{\mu}_i^s \in \mathcal{T}^{m+1}$ and $\boldsymbol{\mu}_i^b \in \mathcal{T}^m$.

conjugates. According to the aforementioned definition, therefore, $\mathbf{Y}_i$ are not internal state variables. They are rather the internal degrees of freedom related to the microstructure. In order to derive the constitutive equations, it is assumed that all the material points in the continuum body at any time, t, go through thermodynamic processes that obey the 2nd law of thermodynamics. Using the Legendre transform for G, it is possible to combine the first and second laws of thermodynamics in their local form to obtain

$$-\rho\dot{G} - \rho s\dot{T} - \boldsymbol{\varepsilon} : \dot{\sigma} + \boldsymbol{\mu}_i^s \cdot \nabla\dot{\mathbf{Y}}_i + \left[\mathrm{Div}(\boldsymbol{\mu}_i^s) + \boldsymbol{\mu}_i^b\right] \cdot \dot{\mathbf{Y}}_i \geq$$
$$\nabla \cdot \mathbf{q} - T\nabla \cdot \left(\frac{\mathbf{q}}{T}\right) = \frac{1}{T}\,\mathbf{q} \cdot \nabla T. \qquad (88)$$

This can be reduced to the Clausius–Plank inequality assuming the Fourier's law of conduction.

$$-\left(\rho\frac{\partial G}{\partial \boldsymbol{\sigma}} + \boldsymbol{\varepsilon}\right) : \dot{\sigma} - \rho\left(\frac{\partial G}{\partial T} + s\right)\dot{T} + \boldsymbol{\mu}_i^s \cdot \nabla\dot{\mathbf{Y}}_i$$
$$+ \left[\mathrm{Div}(\boldsymbol{\mu}_i^s) + \boldsymbol{\mu}_i^b - \rho\frac{\partial G}{\partial \mathbf{Y}_i}\right] \cdot \dot{\mathbf{Y}}_i \geq 0. \qquad (89)$$

Thus, by assigning surface work to generalized state variables, $\mathbf{Y}$, the gradients $\nabla\mathbf{Y}$ appear in the rate of energy dissipation.

If the set of generalized variables, $\mathbf{Y}$, used in Eq. (87) are considered, we will have

$$-\left(\rho\frac{\partial G}{\partial \boldsymbol{\sigma}} + \boldsymbol{\varepsilon}\right) : \dot{\sigma} - \rho\left(\frac{\partial G}{\partial T} + s\right)\dot{T} + D \geq 0. \qquad (90)$$

The inequality (90) is satisfied by considering the following constitutive equations:

$$\boldsymbol{\varepsilon} = -\rho\frac{\partial G}{\partial \boldsymbol{\sigma}}, \quad s = -\frac{\partial G}{\partial T}, \quad D \geq 0, \qquad (91)$$

where D, the dissipation rate, is given by

$$D = \left[\mathrm{Div}(\boldsymbol{\mu}_1^s) + \mu_1^b - \rho\frac{\partial G}{\partial \xi}\right]\dot{\xi} + \left[\mathrm{Div}(\boldsymbol{\mu}_2^s) + \boldsymbol{\mu}_2^b - \rho\frac{\partial G}{\partial \boldsymbol{\varepsilon}^{tr}}\right] : \dot{\varepsilon}^{\mathrm{tr}}$$
$$+ \left[\mathrm{Div}(\boldsymbol{\mu}_3^s) + \boldsymbol{\mu}_3^b + \boldsymbol{\mu}_1^s - \rho\frac{\partial G}{\partial \nabla\xi}\right] \cdot \nabla\dot{\xi}$$
$$+ \left[\mathrm{Div}(\boldsymbol{\mu}_4^s) + \boldsymbol{\mu}_4^b + \boldsymbol{\mu}_2^s - \rho\frac{\partial G}{\partial \nabla\boldsymbol{\varepsilon}^{tr}}\right] \cdot \nabla\dot{\varepsilon}^{\mathrm{tr}}$$
$$+ \boldsymbol{\mu}_3^s : \nabla\nabla\dot{\xi} + \boldsymbol{\mu}_4^s \cdot \nabla\nabla\dot{\varepsilon}^{\mathrm{tr}}. \qquad (92)$$

Whether internal degrees of freedom or internal state variables, the above equations must be augmented by evolution equations for the state variables $\dot{\mathbf{Y}} = \{\dot{\mathbf{Y}}_i\}$, as

well as constitutive equations for $\boldsymbol{\mu}^s = \{\boldsymbol{\mu}_i^s\}$ and $\boldsymbol{\mu}^b = \{\boldsymbol{\mu}_i^b\}$. In any case, it is possible to use the concept of generalized thermodynamic forces, $\boldsymbol{\Gamma}$, and fluxes, $\dot{\Pi}$. In relation to Eq. (89), after including (91)a and b, one can write

$$\boldsymbol{\Gamma} \equiv \left\{\left[\mathrm{Div}(\boldsymbol{\mu}_i^s) + \boldsymbol{\mu}_i^b - \rho\frac{\partial G}{\partial \mathbf{Y}_i}\right], \boldsymbol{\mu}_i^s\right\}, \quad \boldsymbol{\Pi} \equiv \{\mathbf{Y}_i, \nabla\mathbf{Y}_i\},$$
$$D = \boldsymbol{\Gamma}_1 \cdot \dot{\Pi}_1 + \boldsymbol{\Gamma}_2 \cdot \dot{\Pi}_2 = \boldsymbol{\Gamma} \cdot \dot{\Pi} \geq 0. \qquad (93)$$

For the case of SMA response, by redefining the variables in Eq. (92), we will have

$$\boldsymbol{\Gamma}_1 \equiv \mathrm{Div}(\boldsymbol{\mu}_1^s) + \mu_1^b - \rho\frac{\partial G}{\partial \xi}, \quad \boldsymbol{\Pi}_1 \equiv \xi;$$

$$\boldsymbol{\Gamma}_2 \equiv \mathrm{Div}(\boldsymbol{\mu}_2^s) + \boldsymbol{\mu}_2^b - \rho\frac{\partial G}{\partial \boldsymbol{\varepsilon}^{tr}}, \quad \boldsymbol{\Pi}_2 \equiv \boldsymbol{\varepsilon}^{tr};$$

$$\boldsymbol{\Gamma}_3 \equiv \mathrm{Div}(\boldsymbol{\mu}_3^s) + \boldsymbol{\mu}_3^b + \boldsymbol{\mu}_1^s - \rho\frac{\partial G}{\partial \nabla\xi}, \quad \boldsymbol{\Pi}_3 \equiv \nabla\xi;$$

$$\boldsymbol{\Gamma}_4 \equiv \mathrm{Div}(\boldsymbol{\mu}_4^s) + \boldsymbol{\mu}_4^b + \boldsymbol{\mu}_2^s - \rho\frac{\partial G}{\partial \nabla\boldsymbol{\varepsilon}^{tr}}, \quad \boldsymbol{\Pi}_4 \equiv \nabla\boldsymbol{\varepsilon}^{\mathrm{tr}}; \qquad (94)$$

$$\boldsymbol{\Gamma}_5 \equiv \boldsymbol{\mu}_3^s, \quad \boldsymbol{\Pi}_5 \equiv \nabla\nabla\xi;$$

$$\boldsymbol{\Gamma}_6 \equiv \boldsymbol{\mu}_4^s, \quad \boldsymbol{\Pi}_6 \equiv \nabla\nabla\boldsymbol{\varepsilon}^{tr};$$

$$\boldsymbol{\Gamma} \equiv \{\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \boldsymbol{\Gamma}_3, \boldsymbol{\Gamma}_4, \boldsymbol{\Gamma}_5, \boldsymbol{\Gamma}_6\},$$
$$\boldsymbol{\Pi} \equiv \{\boldsymbol{\Pi}_1, \boldsymbol{\Pi}_2, \boldsymbol{\Pi}_3, \boldsymbol{\Pi}_4, \boldsymbol{\Pi}_5, \boldsymbol{\Pi}_6\};$$

$$D = \boldsymbol{\Gamma}_i \cdot \dot{\Pi}_i \geq 0.$$

A common method to achieve a rate-independent response is to define a threshold for $\boldsymbol{\Gamma}$ at which the dissipative mechanism activates. This can be done by assuming the existence of a convex set K such that

$$\forall \boldsymbol{\Gamma}^* \in K : \quad (\boldsymbol{\Gamma}_i - \boldsymbol{\Gamma}_i^*) \cdot \dot{\Pi}_i \geq 0. \qquad (95)$$

No dissipation occurs if $\boldsymbol{\Gamma}$ is inside K. Equation (95) is known as Hill–Mandel's principle of maximum dissipation. A hypersurface

$$\Phi(\boldsymbol{\Gamma}, \boldsymbol{\Pi}) = \phi(\boldsymbol{\Gamma}) - g(\boldsymbol{\Pi}), \qquad (96)$$

can be defined in the space of generalized forces that define the convex boundary of the convex set K. In the case of dislocation plasticity, $\Phi = 0$ is the yield surface. Φ can also be associated with the dissipation potential the result of which is the normality in the space of generalized forces. To that end, the state of generalized variables is considered to be the one that maximizes the dissipation, D, or

$$\mathrm{Max}\left\{D = \boldsymbol{\Gamma}_i \cdot \dot{\Pi}_i \mid \boldsymbol{\Gamma}, \ \Phi(\boldsymbol{\Gamma}, \boldsymbol{\Pi}) \leq 0\right\}. \qquad (97)$$

It is shown that the problem in (97) is equivalent to finding the minimum of a Lagrangian, L,

$$L = -\mathbf{\Gamma}_i \cdot \dot{\Pi}_i + \dot{\lambda}\Phi(\mathbf{\Gamma}, \mathbf{\Pi}), \tag{98}$$

so that

$$\frac{\partial L}{\partial \mathbf{\Gamma}_i} = -\dot{\Pi}_i + \dot{\lambda}\frac{\partial \Phi}{\partial \mathbf{\Gamma}_i} = 0, \tag{99}$$

on the condition that the following Kuhn–Tucker conditions are satisfied:

1. $\Phi(\mathbf{\Gamma}, \mathbf{\Pi})$ is convex in $\mathbf{\Gamma}$ and $\Phi \leq 0$,

2. $\dot{\lambda} \geq 0$, $\tag{100}$

3. $\dot{\lambda}\Phi(\mathbf{\Gamma}) = 0.$

Therefore, the rates of the generalized fluxes can be found via

$$\dot{\Pi}_i = \dot{\lambda}\frac{\partial \Phi}{\partial \mathbf{\Gamma}_i} = \dot{\lambda}\frac{\partial \phi}{\partial \mathbf{\Gamma}_i}, \tag{101}$$

demonstrating the associativity of the response in the space of generalized forces. The existence of the time derivative on both sides does not imply a rate-dependent response because it can be viewed as differentiation with respect to a loading parameter.

The rate of the Lagrange multiplier can be determined from the consistency condition:

$$\dot{\Phi} = \dot{\phi} - \dot{g} = 0, \quad \frac{\partial \phi}{\partial \mathbf{\Gamma}_i} \cdot \dot{\Gamma}_i + \dot{\lambda}\left(\frac{\partial g}{\partial \mathbf{\Pi}_j} \cdot \frac{\partial \phi}{\partial \mathbf{\Gamma}_j}\right) = 0. \tag{102}$$

It is customary for $\phi(\mathbf{\Gamma})$ to be assumed a homogeneous function of degree k. In that case the dissipation, (94) satisfying the second law of thermodynamics, can be obtained from

$$D = \mathbf{\Gamma} \cdot \dot{\Pi} = \dot{\lambda}\left(\mathbf{\Gamma}_i \cdot \frac{\partial \phi}{\partial \mathbf{\Gamma}_i}\right) = k\dot{\lambda}\phi = k\dot{\lambda}g \geq 0, \tag{103}$$

in which Euler's homogeneous function theorem is used. The connection between the general constitutive model developed in this section and the higher-order theories that use a *microscopic force balance* or *microforce balance* law is shown in "Connection with the Gradient Theories Based on the Microforce Balance Law" section.

Reformulation of the Gradient Theory with Internal Variables

The approach of thermodynamics with internal variables is chosen to capture the microstructural changes in an SMA due to the martensitic phase transformation. According to the discussion in "Constitutive Equations" section, the internal variables, by definition, cannot perform work through surface or body forces. The internal variable theory is recovered by setting the generalized forces $\mathbf{\mu}^s$ and $\mathbf{\mu}^b$ to zero such as in Eq. (89), resulting in

$$\mathbf{\varepsilon} = -\rho\frac{\partial G}{\partial \mathbf{\sigma}}, \quad s = -\frac{\partial G}{\partial T}, \quad \mathbf{\mu}^s \equiv \mathbf{0}, \quad \mathbf{\mu}^b \equiv \mathbf{0},$$
$$D = -\rho\frac{\partial G}{\partial \mathbf{Y}} \cdot \dot{\mathbf{Y}} \geq 0. \tag{104}$$

Thus with $\mathbf{Y} \equiv \{\xi, \mathbf{\varepsilon}^{\mathrm{tr}}, \nabla\xi, \nabla\mathbf{\varepsilon}^{\mathrm{tr}}\}$, Eq. (92) reduces to

$$D = \left[-\rho\frac{\partial G}{\partial \xi}\right]\dot{\xi} + \left[-\rho\frac{\partial G}{\partial \mathbf{\varepsilon}^{\mathrm{tr}}}\right]$$
$$: \dot{\varepsilon}^{\mathrm{tr}} + \left[-\rho\frac{\partial G}{\partial \nabla\xi}\right] \cdot \nabla\dot{\xi} + \left[-\rho\frac{\partial G}{\partial \nabla\mathbf{\varepsilon}^{\mathrm{tr}}}\right] \cdot \nabla\dot{\varepsilon}^{\mathrm{tr}} \geq 0. \tag{105}$$

It is acknowledged that the equivalence of "nonexistence" versus "being present with zero value" is a point of controversy [71]. Nonetheless, the final outcome is equivalent for the purpose of the current constitutive formulation.

Connection with the Gradient Theories Based on the Microforce Balance Law

Several gradient-based theories of plasticity postulate an additional balance law that incorporates the *configurational* or *microscopic* forces associated with the effective or accumulated plastic strain [72, 73]. This *microscopic force balance* or *microforce balance* law [52] can be obtained from a generalized form of the principle of virtual power, in which it is assumed that *microtractions* act as work conjugates to the generalized variable of the accumulated plastic strain and contribute to the external power. Also, it is assumed that the accumulated plastic strain and its spatial gradient contribute to the internal power.

As an example, Eq. (3.9) from Gurtin and Anand [72] is rewritten here

$$\mathcal{W}_{ext} = \int_{\partial P} (\mathbf{t}(\mathbf{n}) \cdot \dot{\mathbf{u}} + \chi(\mathbf{n})\dot{\gamma}^p) \, \mathrm{d}A + \int_P \mathbf{b} \cdot \dot{\mathbf{u}} \, \mathrm{d}V,$$
$$\mathcal{W}_{int} = \int_P \left(\mathbf{T} : \dot{\mathbf{E}}^{\mathrm{el}} + \pi\dot{\gamma}^p + \mathbf{\xi} \cdot \nabla\dot{\gamma}^p\right) \mathrm{d}V. \tag{106}$$

Following Gurtin and Anand's notation, the scalar microstress π is the power conjugate to the accumulated plastic strain $\dot{\gamma}^{\mathrm{p}} = \left\|\dot{\mathbf{E}}^{\mathrm{p}}\right\|$, $\mathbf{\xi}$ a vector microstress conjugate to $\nabla\dot{\gamma}^{\mathrm{p}}$, and $\chi(\mathbf{n})$ is the surface microtraction. It is then assumed that $\mathcal{W}_{\mathrm{int}} = \mathcal{W}_{\mathrm{ext}}$ for any arbitrary subregion P of the body and for any consistent virtual velocities $\mathcal{V} = \left(\dot{\mathbf{u}}, \dot{\mathbf{E}}^{\mathrm{el}}, \dot{\gamma}^{\mathrm{p}}\right)$.

The generalized principle of virtual power results in the conventional and microforce balance laws

$$\text{Div}(\mathbf{T}) + \mathbf{b} = \mathbf{0}, \quad \tau - \pi + \text{Div}(\boldsymbol{\xi}) = 0, \tag{107}$$

as well as standard traction and microtraction boundary conditions

$$\mathbf{t}(\mathbf{n}) = \mathbf{T}\mathbf{n}, \quad \chi(\mathbf{n}) = \boldsymbol{\xi} \cdot \mathbf{n}, \tag{108}$$

where τ is the (Von Mises) equivalent stress. Furthermore, the dissipation inequality as given in Eq. (6.13) of [72] is

$$D = \pi_{\text{dis}} \dot{\gamma}^{\text{p}} + \boldsymbol{\xi}_{\text{dis}} \cdot \nabla \dot{\gamma}^{\text{p}} \geq 0, \quad \pi_{\text{dis}} = \pi - \frac{\partial \psi}{\partial \gamma^{\text{p}}},$$

$$\boldsymbol{\xi}_{\text{dis}} = \boldsymbol{\xi} - \frac{\partial \psi}{\partial \nabla \gamma^{\text{p}}},$$

where $\psi = \hat{\psi}\left(\mathbf{E}^{\text{el}}, \gamma^{\text{p}}, \nabla \gamma^{\text{p}}\right)$ is the Helmholtz free energy. Replacing π and $\boldsymbol{\xi}$ in the microforce balance (107) yields

$$\tau - \pi_{\text{dis}} - \frac{\partial \psi}{\partial \gamma^{\text{p}}} + \text{Div}\left(\boldsymbol{\xi}_{\text{dis}} + \frac{\partial \psi}{\partial \nabla \gamma^{\text{p}}}\right) = 0, \tag{109}$$

which is Eq. (6.15) in [72]. Equation (109), together with constitutive equations for π_{dis} and $\boldsymbol{\xi}_{\text{dis}}$, acts as the flow rule for the gradient plasticity theory. It will now be shown that (109) can be obtained following the general model established in "Constitutive Equations" section. This, however, will be performed for the shape memory response by replacing the accumulated plastic strain with the MVF. To that end, it is assumed that the generalized variables are

$$\mathbf{Y} \equiv \{\xi, \boldsymbol{\varepsilon}^{\text{tr}}, \nabla\xi\}, \quad \boldsymbol{\mu}^s \equiv \{\boldsymbol{\mu}_1^s\}, \quad \boldsymbol{\mu}^b \equiv \{\mu_1^b, \boldsymbol{\mu}_3^b\}, \tag{110}$$

where $\boldsymbol{\mu}_1^s$ is the surface power conjugate to ξ and μ_1^b and $\boldsymbol{\mu}_3^b$ are, respectively, the body power conjugates to ξ and $\nabla\xi$ in the generalized first law of thermodynamics (86). The dissipation in Eq. (92) reduces to

$$D = \left[\text{Div}(\boldsymbol{\mu}_1^s) + \mu_1^b - \rho\frac{\partial G}{\partial \xi}\right]\dot{\xi} + \left[-\rho\frac{\partial G}{\partial \boldsymbol{\varepsilon}^{\text{tr}}}\right] : \dot{\boldsymbol{\varepsilon}}^{\text{tr}}$$
$$+ \left[\boldsymbol{\mu}_3^b + \boldsymbol{\mu}_1^s - \rho\frac{\partial G}{\partial \nabla\xi}\right] \cdot \nabla\dot{\xi} \geq 0. \tag{111}$$

Hence, the list of generalized forces and fluxes is as follows:

$$\boldsymbol{\Gamma}_1 \equiv \text{Div}(\boldsymbol{\mu}_1^s) + \mu_1^b - \rho\frac{\partial G}{\partial \xi}, \quad \boldsymbol{\Pi}_1 \equiv \xi,$$

$$\boldsymbol{\Gamma}_2 \equiv -\rho\frac{\partial G}{\partial \boldsymbol{\varepsilon}^{\text{tr}}}, \quad \boldsymbol{\Pi}_2 \equiv \boldsymbol{\varepsilon}^{\text{tr}},$$

$$\boldsymbol{\Gamma}_3 \equiv \boldsymbol{\mu}_3^b + \boldsymbol{\mu}_1^s - \rho\frac{\partial G}{\partial \nabla\xi}, \quad \boldsymbol{\Pi}_3 \equiv \nabla\xi,$$

$$\boldsymbol{\Gamma} \equiv \{\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \boldsymbol{\Gamma}_3\}, \quad \boldsymbol{\Pi} \equiv \{\boldsymbol{\Pi}_1, \boldsymbol{\Pi}_2, \boldsymbol{\Pi}_3\},$$
$$D = \boldsymbol{\Gamma}_i \cdot \dot{\Pi}_i \geq 0. \tag{112}$$

Constitutive assumptions are now considered for the generalized forces and fluxes. First, it is assumed that $\boldsymbol{\Gamma}_3 \equiv \mathbf{0}$ such that

$$\boldsymbol{\mu}_1^s = \rho\frac{\partial G}{\partial \nabla\xi} - \boldsymbol{\mu}_3^b. \tag{113}$$

In addition, the transformation surface in Eq. (96) is considered to have the form

$$\Phi(\boldsymbol{\Gamma}, \boldsymbol{\Pi}) = \phi(\boldsymbol{\Gamma}) - g(\boldsymbol{\Pi}) = \sqrt{\boldsymbol{\Gamma}_2 : \boldsymbol{\Gamma}_2} + \boldsymbol{\Gamma}_1 - Y = 0 \tag{114}$$

or

$$\Phi = \sqrt{\boldsymbol{\Gamma}_2 : \boldsymbol{\Gamma}_2} + \text{Div}\left(\rho\frac{\partial G}{\partial \nabla\xi} - \boldsymbol{\mu}_3^b\right) + \mu_1^b - \rho\frac{\partial G}{\partial \xi} - Y$$
$$= 0. \tag{115}$$

Now, if the form of the Gibbs free energy is such that

$$\boldsymbol{\Gamma}_2 \equiv -\rho\frac{\partial G}{\partial \boldsymbol{\varepsilon}^{\text{tr}}} = \boldsymbol{\sigma}, \tag{116}$$

then

$$\Phi = \tau + \text{Div}\left(\rho\frac{\partial G}{\partial \nabla\xi} - \boldsymbol{\mu}_3^b\right) - \rho\frac{\partial G}{\partial \xi} + \mu_1^b - Y = 0, \tag{117}$$

equivalent to the microforce balance in Eq. (109) augmented with the constitutive relations $\pi_{\text{dis}} \equiv Y - \mu_1^b$ and $\boldsymbol{\xi}_{\text{dis}} \equiv -\boldsymbol{\mu}_3^b$.

The Most General Anisotropic/Isotropic Gibbs Free Energy for the Gradient-Based SMA Constitutive Model

The form for the Gibbs free energy of Eq. (20) is given in this appendix. The most general anisotropic Gibbs free energy with a polynomial form including terms up to 3rd degree is

$$\rho\, G(\boldsymbol{\sigma}, T, \xi, \boldsymbol{\varepsilon}^{\mathrm{tr}}, \nabla\xi, \nabla\boldsymbol{\varepsilon}^{\mathrm{tr}}) = \rho\, G_0$$
$$+ (\xi P_{ijkl} + TQ_{ijkl} + F_{ijklmn}\sigma_{mn} + N_{ijklmn}\varepsilon_{mn}^{\mathrm{tr}} + O_{ijklm}\xi_{,m}$$
$$+ \mathbb{A}_{ijklmno}\varepsilon_{mn,o}^{\mathrm{tr}} + A_{ijkl})\sigma_{ij}\sigma_{kl}$$
$$+ (\xi\bar{A}_{ijkl} + T\bar{B}_{ijkl} + S_{ijklmn}\varepsilon_{mn}^{\mathrm{tr}} + Z_{ijklm}\xi_{,m}$$
$$+ \mathbb{F}_{ijklmno}\varepsilon_{mn,o}^{\mathrm{tr}} + C_{ijkl})\sigma_{ij}\varepsilon_{kl}^{\mathrm{tr}}$$
$$+ (\mathbb{B}_{ijklmno}\varepsilon_{mn,o}^{\mathrm{tr}} + \xi V_{ijkl} + T U_{ijkl} + K_{ijklmn}\varepsilon_{mn}^{\mathrm{tr}}$$
$$+ \tau_{ijklm}\xi_{,m} + B_{ijkl})\varepsilon_{ij}^{\mathrm{tr}}\varepsilon_{kl}^{\mathrm{tr}}$$
$$+ (\mathbb{C}_{ijlmn}\varepsilon_{lm,n}^{\mathrm{tr}} + L_{ij}\xi + M_{ij}T + W_{klij}\sigma_{kl} + Y_{klij}\varepsilon_{kl}^{\mathrm{tr}}$$
$$+ L_{ijk}\xi_{,k} + C_{ij})\xi_{,i}\xi_{,j}$$
$$+ (\mathbb{H}_{ijklmn}\sigma_{ij} + \mathbb{K}_{ijklmn}\varepsilon_{ij}^{\mathrm{tr}} + \mathbb{U}_{klmn}\xi + \mathbb{V}_{klmn}T)\xi_{,k}\varepsilon_{lm,n}^{\mathrm{tr}}$$
$$+ (\tilde{N}_{ijklmnopq}\varepsilon_{op,q}^{\mathrm{tr}} + \tilde{H}_{ijklmnpq}\sigma_{pq} + \tilde{J}_{ijklmnpq}\varepsilon_{pq}^{\mathrm{tr}}$$
$$+ \tilde{K}_{ijklmnp}\xi_{,p} + \tilde{L}_{ijklmn}\xi + \tilde{M}_{ijklmn}T)\varepsilon_{ij,k}^{\mathrm{tr}}\varepsilon_{lm,n}^{\mathrm{tr}}$$
$$+ (\mathbb{I}_{ijlmn}\xi\varepsilon_{lm,n}^{\mathrm{tr}} + \mathbb{J}_{ijlmn}T\varepsilon_{lm,n}^{\mathrm{tr}} + D_{ij}\xi + \bar{C}_{ijk}\xi\xi_{,k}$$
$$+ \bar{D}_{ijk}T\xi_{,k} + N_{ij}\xi^2 + E_{ij}T + P_{ij}\xi T + Q_{ij}T^2 + \alpha_{ij})\sigma_{ij}$$
$$+ (\mathbb{L}_{ijlmn}\xi\varepsilon_{lm,n}^{\mathrm{tr}} + \mathbb{M}_{ijlmn}T\varepsilon_{lm,n}^{\mathrm{tr}} + F_{ij}\xi + O_{ij}\xi^2 + K_{ij}T$$
$$+ \bar{K}_{ij}T\xi + S_{ij}T^2 + \bar{E}_{ijk}\xi\xi_{,k} + \bar{F}_{ijk}T\xi_{,k} + \beta_{ij})\varepsilon_{ij}^{\mathrm{tr}}$$
$$+ (D_{mni}\sigma_{mn} + E_{mni}\varepsilon_{mn}^{\mathrm{tr}} + g_iT + f_i\xi + r_iT\xi$$
$$+ m_i\xi^2 + o_iT^2 + a_i)\xi_{,i}$$
$$+ (\tilde{A}_{ijklmn}\varepsilon_{lm,n}^{tr} + \tilde{B}_{ijklm}\sigma_{lm} + \tilde{C}_{ijklm}\varepsilon_{lm}^{\mathrm{tr}} + \tilde{D}_{ijkl}\xi_{,l}$$
$$+ \tilde{E}_{ijk}\xi + \tilde{F}_{ijk}T + \mathbb{O}_{ijk}T\xi + \mathbb{D}_{ijk}\xi^2 + \mathbb{E}_{ijk}T^2 + \tilde{\alpha}_{ijk})\varepsilon_{ij,k}^{\mathrm{tr}}$$
$$+ (lT^2 + dT + b)T + (k\xi^2 + c\xi + a)\xi$$
$$+ (pT + n\xi + f)T\xi$$

$$(118)$$

The free energy includes all possible couplings between the state variables with their corresponding tensorial material constants. The tensorial material constants are the derivatives of the free energy calculated at the reference state leading to the major and minor symmetries.

The Gibbs free energy for the case of isotropic materials can be obtained by rewriting the constant tensors included in Eq. (118) in terms of their isotropic invariants. That is,

$$\rho\, G(\boldsymbol{\sigma}, T, \xi, \boldsymbol{\varepsilon}^{\mathrm{tr}}, \nabla\xi, \nabla\boldsymbol{\varepsilon}^{\mathrm{tr}}) = \rho\, G_0$$
$$+ (\alpha + \xi D_1 + \xi^2 N_6 + TE_1 + T\xi P_3 + T^2Q_3)\mathrm{tr}(\boldsymbol{\sigma})$$
$$+ (A_1 + \xi P_1 + TQ_1 + \mathrm{tr}(\boldsymbol{\sigma})F_1)\mathrm{tr}(\boldsymbol{\sigma}^2)$$
$$+ 2(A_2 + \xi P_2 + TQ_2 + 3\mathrm{tr}(\boldsymbol{\sigma})F_2)\mathrm{tr}(\boldsymbol{\sigma}^2)$$
$$+ 8F_5\,\mathrm{tr}(\boldsymbol{\sigma}^3) + 2(C_2 + \xi\bar{A}_2 + T\bar{B}_2 + 2\mathrm{tr}(\boldsymbol{\sigma})N_2)\mathrm{tr}(\boldsymbol{\sigma}\boldsymbol{\varepsilon}^{\mathrm{tr}})$$
$$+ 8N_5\,\mathrm{tr}(\boldsymbol{\sigma}^2\boldsymbol{\varepsilon}^{\mathrm{tr}}) + 8S_5\,\mathrm{tr}(\boldsymbol{\sigma}\boldsymbol{\varepsilon}^{\mathrm{tr2}})$$
$$+ 2(B_2 + \xi V_2 + TU_2 + \mathrm{tr}(\boldsymbol{\sigma})S_2)\mathrm{tr}(\boldsymbol{\varepsilon}^{\mathrm{tr2}})$$
$$+ 8K_5\,\mathrm{tr}(\boldsymbol{\varepsilon}^{\mathrm{tr3}}) + \Big[(c_1 + \xi L_1 + TM_1 + \mathrm{tr}(\boldsymbol{\sigma})W_1)\delta_{ij}$$
$$+ 2W_1\sigma_{ij} + 2Y_1\varepsilon_{ij}^{\mathrm{tr}}\Big]\xi_{,i}\xi_{,j}$$

$$+ (\tilde{H}_{ijklmnpq}\sigma_{pq} + \tilde{J}_{ijklmnpq}\boldsymbol{\varepsilon}_{pq}^{\mathrm{tr}})\varepsilon_{ij,k}^{\mathrm{tr}}\varepsilon_{lm,n}^{\mathrm{tr}}$$
$$+ 4(\xi\tilde{L}_5 + T\tilde{M}_5 + \tilde{A}_5)\varepsilon_{ij,j}^{\mathrm{tr}}\varepsilon_{ik,k}^{\mathrm{tr}}$$
$$+ 2(\xi\tilde{L}_{10} + T\tilde{M}_{10} + \tilde{A}_{10})\varepsilon_{ij,k}^{\mathrm{tr}}\varepsilon_{ij,k}^{\mathrm{tr}}$$
$$+ 4(\xi\tilde{L}_{11} + T\tilde{M}_{11} + \tilde{A}_{11})\varepsilon_{ij,j}^{\mathrm{tr}}\varepsilon_{ik,j}^{\mathrm{tr}}$$
$$+ 2(\tilde{D}_1 + \xi U_1 + T\tilde{V}_1 + \mathrm{tr}(\boldsymbol{\sigma})\tilde{H}_1)\xi_i\varepsilon_{ij,j}^{\mathrm{tr}}$$
$$+ 2\Big[(\xi_j\varepsilon_{ik,k}^{\mathrm{tr}} + \xi_i\varepsilon_{jk,k}^{\mathrm{tr}}) + (\varepsilon_{ij,n}^{\mathrm{tr}} + \varepsilon_{ni,j}^{\mathrm{tr}} + \varepsilon_{nj,i}^{\mathrm{tr}})\xi_n\Big]$$
$$(\tilde{H}_1\sigma_{ij} + \tilde{K}_1\varepsilon_{ij}^{\mathrm{tr}}) + (lT^2 + dT + b)T$$
$$+ (k\xi^2 + c\xi + a)\xi + (pT + n\xi + f)T\xi.$$

$$(119)$$

In this equation, the notation for the scalar constants is chosen based on their corresponding tensor. The rank-8 tensors, as listed, are not expanded due to the prohibitively large number, 91, of scalar constants involved. Four conditions are taken into account for the coefficient tensorial material constants in order to derive Eq. (119) from Eq. (118) as follows: (1) isotropy, (2) major symmetry as a result of interchangeability in differentiation of G with respect to state variables, (3) minor symmetry as a result of symmetry in σ, $\boldsymbol{\varepsilon}^{\mathrm{tr}}$, and $\nabla\boldsymbol{\varepsilon}^{\mathrm{tr}}$, and (4) $\mathrm{tr}(\boldsymbol{\varepsilon}^{\mathrm{tr}}) = 0$ due to the volume-preserving assumption for the transformation strain.

Experimental Measurement of the SMA Material Properties

An attempt is made herein to relate the material constants of the constitutive model in "The Simplified SMA Gradient Model Including Only Terms with $\nabla\xi$" section to the typical experimentally observed properties of shape memory alloys. For that, the solution to uniaxial loading of an SMA bar is considered ("Extension of SMA Prismatic Bars" section). The typical pseudoelastic response of SMAs, based on Eqs. (48) and (49), is shown in Fig. 8. σ_s^{fwd} and σ_f^{fwd} are the stresses to start and finish the forward transformation to martensite at the current temperature and σ_s^{rev} and σ_f^{rev} are the stresses to start and finish the reverse transformation to austenite, respectively.

The material constants used in the model can be found based on the following critical transformation start and finish conditions for a stress-free heating and cooling path:

1. At the start of forward transformation, $T = M_s$, $\xi = 0$,
2. At the end of forward transformation, $T = M_f$, $\xi = 1$,
3. At the start of reverse transformation, $T = A_s$, $\xi = 1$,
4. At the end of reverse transformation, $T = A_f$, $\xi = 0$.
5. The Gibbs free energy and thus the hardening function, $f(\xi)$, must be continuous, i.e., $f(\xi = 1)|_{\mathrm{fwd}} = f(\xi = 1)|_{\mathrm{rev}}$. The form of f is given in Eq. (44) for the linear hardening

Fig. 8 A typical stress–temperature phase diagram and pseudoelastic loading path for SMAs
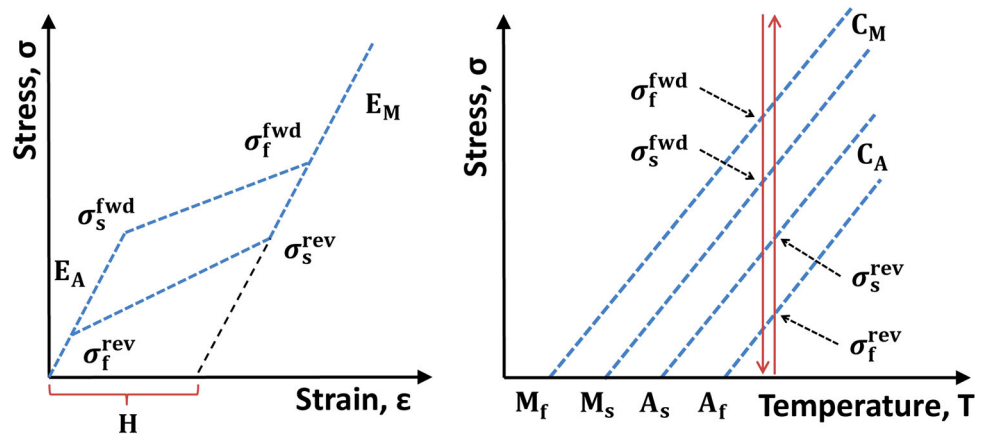
Table 2 Material constants used in the local part of the SMA constitutive model using the linear hardening function

$$Y = \tfrac{1}{4}\Delta s_0 (M_s + M_f - A_s - A_f)$$

$$\rho b^M = -\rho \Delta s_0 (M_s - M_f)$$

$$\rho b^A = -\rho \Delta s_0 (A_s - A_f)$$

$$\mu_1 + \rho \Delta u_0 = \tfrac{1}{2}\rho \Delta s_0 (M_s + A_f)$$

$$\mu_2 = -\tfrac{1}{4}\rho \Delta s_0 (M_f - M_s + A_f - A_s)$$

$$\rho \Delta s_0 = -\left[H + \left(\tfrac{1}{E^M} - \tfrac{1}{E^A}\right)\sigma^*\right]C_M, \quad \rho \Delta s_0 = -\left[H + \left(\tfrac{1}{E^M} - \tfrac{1}{E^A}\right)\sigma^*\right]C_A$$

Table 3 Material constants used in the local part of the SMA constitutive model using the smooth hardening function

$$Y = -\tfrac{1}{2}\rho \Delta s_0 (M_s - A_f) - b_3$$

$$b_1 = \rho \Delta s_0 (M_f - M_s)$$

$$b_2 = \rho \Delta s_0 (A_s - A_f)$$

$$b_3 = -\tfrac{b_1}{4}\left(1 + \tfrac{1}{n_1+1} - \tfrac{1}{n_2+1}\right) + \tfrac{b_2}{4}\left(1 + \tfrac{1}{n_3+1} - \tfrac{1}{n_4+1}\right)$$

$$\rho \Delta u_0 = \tfrac{1}{2}\rho \Delta s_0 (A_f + M_s)$$

$$\rho \Delta s_0 = -\left[H + \left(\tfrac{1}{E^M} - \tfrac{1}{E^A}\right)\sigma^*\right]C_M, \quad \rho \Delta s_0 = -\left[H + \left(\tfrac{1}{E^M} - \tfrac{1}{E^A}\right)\sigma^*\right]C_A$$

Table 4 SMA material properties used in the developed nonlocal model

	Thermoelastic constants			Transformation properties				Nonlocality
v^A	E^A (MPa)	$\alpha^A \left(\tfrac{1}{°C}\right)$		A_s (°C)	A_f (°C)	$C_A \left(\tfrac{MPa}{°C}\right)$	H (%)	$\mathcal{M}^{rev}$ (MPa mm)
v^M	E^M (MPa)	$\alpha^M \left(\tfrac{1}{°C}\right)$		M_s (°C)	M_f (°C)	$C_M \left(\tfrac{MPa}{°C}\right)$		$\mathcal{M}^{fwd}$ (MPa mm)

rule. Notice that this condition for $\xi = 0$ is trivially satisfied. For the smooth hardening function, $\frac{\partial f(\xi)}{\partial \xi}$ is explicitly given in Eq. (45). Hence, the continuity condition can be written in the integral form as $\int_0^1 \frac{\partial f^{fwd}(\xi)}{\partial \xi}d\xi = \int_0^1 \frac{\partial f^{rev}(\xi)}{\partial \xi}d\xi$.

6. By taking the differential of (48) and (49) and letting $d\xi$ to vanish, it is possible to recover the slopes of the transformation regions on the stress–temperature phase diagram at a specific stress (for here the calibration stress $\sigma = \sigma^*$). This leads to a relation for $\rho \Delta s_0$ based on the forward and reverse transformation slopes, $C_M \equiv \frac{d\sigma}{dT}\big|^{fwd}_{\sigma=\sigma^*}$ and $C_A \equiv \frac{d\sigma}{dT}\big|^{rev}_{\sigma=\sigma^*}$.

The results of the conditions listed above are summarized in Table 2 for the linear hardening and in Table 3 for the smooth hardening rules.

The SMA model in this form cannot capture a response with different Clausius–Clapeyron slopes, as per the last equation shown in Tables 2 or 3, hence $C_A = C_M$.

In addition,

$$\sigma_s^{\text{fwd}} = C_M(T - M_s), \quad \sigma_f^{\text{fwd}} = C_M(T - M_f),$$
$$\sigma_s^{\text{rev}} = C_A(T - A_s), \quad \sigma_f^{\text{rev}} = C_A(T - A_f), \quad C_A = C_M. \tag{120}$$

To derive the relations in (120), the difference between the elastic moduli of austenite and martensite is ignored.

The nonlocal model developed in "The Simplified SMA Gradient Model Including Only Terms with $\nabla \xi$" section introduces six energetic and two dissipative length scales in addition to the SMA material constants already being used in the classical local Boyd–Lagoudas SMA model [56]. A simplified version of the model considering only the three independent length scales ℓ_1, ℓ_d^{fwd}, and ℓ_d^{rev} was used to study the behavior of SMA beams with various thicknesses. Based on the nonlocal constants, two parameters $\mathcal{M}_1^{\text{fwd}}$ and $\mathcal{M}_2^{\text{rev}}$ can be identified in the transformation differential equation. The material constants used in nonlocal modeling for SMAs are given in Table 4. The dissipative length scales, ℓ_d^{fwd} and ℓ_d^{rev}, are excluded from this table since they do not *independently* appear in the solution of the SMA structures studied. This is because the solutions of the transformation surfaces, e.g., Eq. (68), rather than the rate forms are considered for those one-dimensional problems.

The relation between the local material parameters expressed in the course of model development and the more common SMA properties listed in Table 4 was established in this section using the problem of isothermal uniaxial stretching of an SMA bar ("Extension of SMA Prismatic Bars" section). Solution to this problem showed that the nonlocal model reduces to the original local SMA model for the cases in which there are no spatial gradients of the MVF.

Therefore, calibration of the nonlocal parameters, e.g., $\mathcal{M}_1^{\text{fwd}}$ and $\mathcal{M}_1^{\text{rev}}$ for the case of "Pure Bending of an SMA Beam" section, must be performed using specially designed experiments which provide spatial gradients. To this end, bending of SMA beams with various thicknesses was discussed in the results section. Also torsion of SMA thin wires with various diameters, or compression of tapered SMA pillars with different diameters are promising for nonlocal material parameter calibration purposes. To the best knowledge of the authors, there are no experimental results published for bending of SMA thin films with different thicknesses, although this has been done for the case of dislocation plasticity using aluminum foils [39]. Experimental observations for compression of SMA micropillars, on the other hand, are available and can be used for calibration of the current SMA gradient model.

References

1. Benard WL, Kahn H, Heuer AH, Huff MA (1998) Thin-film shape-memory alloy actuated micropumps. J Microelectromech Syst 7(2):245–251
2. Kahn H, Huff MA, Heuer AH (1998) The TiNi shape-memory alloy and its applications for MEMS. J Micromech Microeng 8(3):213
3. Fu YQ, Luo JK, Ong SE, Zhang S, Flewitt AJ, Milne WI (2008) A shape memory microcage of TiNi/DLC films for biological applications. J Micromech Microeng 18(3):035026
4. Shin DD, Lee DG, Mohanchandra KP, Carman GP (2006) Thin film NiTi microthermostat array. Sens Actuators, A 130:37–41
5. Arzt E (1998) Size effects in materials due to microstructural and dimensional constraints: a comparative review. Acta Mater 46(16):5611–5626
6. Bazant ZP, Planas J (1997) Fracture and size effect in concrete and other quasibrittle materials. New directions in civil engineering. Taylor & Francis, London
7. Chen CQ, Shi Y, Zhang YS, Zhu J, Yan YJ (2006) Size dependence of young's modulus in ZnO nanowires. Phys Rev Lett 96:075505
8. Yang F (2004) Size-dependent effective modulus of elastic composite materials: spherical nanocavities at dilute concentrations. J Appl Phys 95(7):3516–3520
9. Fleck NA, Muller GM, Ashby MF, Hutchinson JW (1994) Strain gradient plasticity: theory and experiment. Acta Metall Mater 42(2):475–487
10. Hutchinson JW (2000) Plasticity at the micron scale. Int J Solids Struct 37(1):225–238
11. Morrison JLM (1939) The yield of mild steel with particular reference to the effect of size of specimen. Proc Inst Mech Eng 142(1):193–223
12. Frommen C, Wilde G, Rösner H (2004) Wet-chemical synthesis and martensitic phase transformation of Au–Cd nanoparticles with near-equiatomic composition. J Alloy Compd 377(1–2):232–242
13. Glezer AM, Blinova EN, Pozdnyakov VA, Shelyakov AV (2003) Martensite transformation in nanoparticles and nanomaterials. J Nanopart Res 5(5–6):551–560
14. Frick CP, Orso S, Arzt E (2007) Loss of pseudoelasticity in nickel–titanium sub-micron compression pillars. Acta Mater 55(11):3845–3855
15. San Juan J, Nó ML (2013) Superelasticity and shape memory at nano-scale: size effects on the martensitic transformation. J Alloy Compd 577(Supplement 1):S25–S29
16. Norfleet DM, Sarosi PM, Manchiraju S, Wagner MF, Uchic MD, Anderson PM, Mills MJ (2009) Transformation-induced plasticity during pseudoelastic deformation in Ni–Ti microcrystals. Acta Mater 57(12):3549–3561
17. Ozdemir N, Karaman I, Mara NA, Chumlyakov YI, Karaca HE (2012) Size effects in the superelastic response of $Ni_{54}Fe_{19}Ga_{27}$ shape memory alloy pillars with a two stage martensitic transformation. Acta Mater 60(16):5670
18. San Juan J, Nó ML, Schuh CA (2009) Nanoscale shape-memory alloys for ultrahigh mechanical damping. Nat Nanotechnol 4(7):415–419
19. Chen Y, Schuh CA (2011) Size effects in shape memory alloy microwires. Acta Mater 59(2):537–553
20. Babanly MB, Lobodyuk VA, Matveeva NM (1993) Size effect in martensite transformation in TiNiCu alloys. Fiz Met Metalloved 75(5):89–95

21. Busch JD, Johnson AD, Lee CH, Stevenson DA (1990) Shape-memory properties in Ni-Ti sputter-deposited film. J Appl Phys 68(12):6224–6228

22. Ishida A, Sato M (2003) Thickness effect on shape memory behavior of Ti-50.0at.%Ni thin film. Acta Mater 51(18):5571–5578

23. Wan D, Komvopoulos K (2005) Thickness effect on thermally induced phase transformations in sputtered titanium-nickel shape-memory films. J Mater Res 20:1606–1612

24. Guimarães JRC (2007) Excess driving force to initiate martensite transformation in fine-grained austenite. Scr Mater 57(3):237–239

25. Kim Y, Cho G, Hur S, Jeong S, Nam T (2006) Nanocrystallization of a Ti–50.0Ni(at.%) alloy by cold working and stress/strain behavior. Mater Sci Eng, A 438–440:531–535

26. Kockar B, Karaman I, Kim JI, Chumlyakov YI, Sharp J, Yu CJ (2008) Thermomechanical cyclic response of an ultrafine-grained NiTi shape memory alloy. Acta Mater 56(14):3630–3646

27. Malygin GA (2008) Nanoscopic size effects on martensitic transformations in shape memory alloys. Phys Solid State 50(8):1538–1543

28. Waitz T, Antretter T, Fischer FD, Simha NK, Karnthaler HP (2007) Size effects on the martensitic phase transformation of NiTi nanograins. J Mech Phys Solids 55(2):419–444

29. Waitz T, Kazykhanov V, Karnthaler HP (2004) Martensitic phase transformations in nanocrystalline NiTi studied by {TEM}. Acta Mater 52(1):137–147

30. Eringen AC (1978) Line crack subject to shear. Int J Fract 14(4):367–379

31. Eringen AC (1983) On differential equations of nonlocal elasticity and solutions of screw dislocation and surface waves. J Appl Phys 54(9):4703–4710

32. Lazar M, Maugin GA (2005) Nonsingular stress and strain fields of dislocations and disclinations in first strain gradient elasticity. Int J Eng Sci 43(13–14):1157–1184

33. Choi K, Kuhn JL, Ciarelli MJ, Goldstein SA (1990) The elastic moduli of human subchondral, trabecular, and cortical bone tissue and the size-dependency of cortical bone modulus. J Biomech 23(11):1103–1113

34. Kakunai S, Masaki J, Kuroda R, Iwata K, Nagata R (1985) Measurement of apparent young's modulus in the bending of cantilever beam by heterodyne holographic interferometry. Exp Mech 25(4):408–412

35. Lakes RS (1986) Experimental microelasticity of two porous solids. Int J Solids Struct 22(1):55–63

36. Fleck NA, Hutchinson JW (1997) Strain gradient plasticity. Adv Appl Mech 33:295–361

37. Jirasek M (2004) Nonlocal theories in continuum mechanics. Acta Polytech 44(5–6):16–34

38. McElhaney KW, Vlassak JJ, Nix WD (1998) Determination of indenter tip geometry and indentation contact area for depth-sensing indentation experiments. J Mater Res 13:1300–1306

39. Stölken JS, Evans AG (1998) A microbend test method for measuring the plasticity length scale. Acta Mater 46(14):5109–5115

40. Bazant ZP, Jirásek M (2002) Nonlocal integral formulations of plasticity and damage: survey of progress. J Eng Mech 128(11):1119–1149

41. Borino G, Fuschi P, Polizzotto C (1999) A thermodynamic approach to nonlocal plasticity and related variational principles. J Appl Mech 66(4):952–963

42. Jirasek M, Bazant ZP (2002) Inelastic analysis of structures. Wiley, New York

43. Engelen RAB, Geers MGD, Baaijens F (2003) Nonlocal implicit gradient-enhanced elasto-plasticity for the modelling of softening behaviour. Int J Plast 19(4):403–433

44. Fleck NA, Hutchinson JW (1993) A phenomenological theory for strain gradient effects in plasticity. J Mech Phys Solids 41(12):1825–1857

45. Aifantis EC (1984) On the microstructural origin of certain inelastic models. J Eng Mater Technol 106(4):326–330

46. Aifantis EC (1999) Strain gradient interpretation of size effects. Int J Fract 95(1–4):299–314

47. Fleck NA, Hutchinson JW (2001) A reformulation of strain gradient plasticity. J Mech Phys Solids 49(10):2245–2271

48. Qiao L, Rimoli JJ, Chen Y, Schuh CA, Radovitzky R (2011) Nonlocal superelastic model of size-dependent hardening and dissipation in single crystal Cu-Al-Ni shape memory alloys. Phys Rev Lett 106(8):085504

49. M Tabesh, JG Boyd, DC Lagoudas (2014) Modeling size effect in the SMA response: a gradient theory. In: Proceedings of SPIE, vol 9058, pp 905803–905803–11

50. Engelen R, Fleck NA, Peerlings RHJ, Geers M (2006) An evaluation of higher-order plasticity theories for predicting size effects and localisation. Int J Solids Struct 43(7):1857–1877

51. Gudmundson P (2004) A unified treatment of strain gradient plasticity. J Mech Phys Solids 52(6):1379–1406

52. Gurtin ME (1996) Generalized Ginzburg-Landau and Cahn-Hilliard equations based on a microforce balance. Phys D 92(3):178–192

53. Aifantis EC (2003) Update on a class of gradient theories. Mech Mater 35(3):259–280

54. Aifantis EC (2011) On the gradient approach—relation to Eringen's nonlocal theory. Int J Eng Sci 49(12):1367–1377

55. Mühlhaus HB, Aifantis EC (1991) A variational principle for gradient plasticity. Int J Solids Struct 28(7):845–857

56. Lagoudas DC (ed) (2008) Shape memory alloys: modeling and engineering applications. Springer, New York

57. Badnava H, Kadkhodaei M, Mashayekhi M (2014) A non-local implicit gradient-enhanced model for unstable behaviors of pseudoelastic shape memory alloys in tensile loading. Int J Solids Struct 51(23):4015–4025

58. Duval A, Haboussi M, Zineb TB (2011) Modelling of localization and propagation of phase transformation in superelastic SMA by a gradient nonlocal approach. Int J Solids Struct 48(13):1879–1893

59. Peultier B, Zineb TB, Patoor E (2006) Macroscopic constitutive law of shape memory alloy thermomechanical behaviour. Application to structure computation by FEM. Mech Mater 38(5):510–524

60. Brinson LC (1993) One-dimensional constitutive behavior of shape memory alloys: thermomechanical derivation with non-constant material functions and redefined martensite internal variable. J Intell Mater Syst Struct 4(2):229–242

61. Gurtin ME, Anand L (2005) A theory of strain-gradient plasticity for isotropic, plastically irrotational materials. Part I: small deformations. J Mech Phys Solids 53(7):1624–1649

62. Sun QP, He YJ (2008) A multiscale continuum model of the grain-size dependence of the stress hysteresis in shape memory alloy polycrystals. Int J Solids Struct 45(13):3868–3896

63. Lubliner J (2008) Plasticity theory. Courier Dover Publications, New York

64. Santaoja K (2004) Gradient theory from the thermomechanics point of view. Eng Fract Mech 71(4):557–566

65. Boehler JP (1987) Representations for isotropic and anisotropic non-polynomial tensor functions. In: Boehler JP (ed) Applications of tensor functions in solid mechanics. International Centre for Mechanical Sciences, vol 292. Springer, Vienna, pp 31–53

66. Spencer AJM (1987) Isotropic polynomial invariants and tensor functions. In: Boehler JP (ed) Applications of tensor functions in solid mechanics. International Centre for Mechanical Sciences, vol 292. Springer, Vienna, pp 141–169

67. Lagoudas DC, Bo Z, Qidwai MA (1996) A unified thermodynamic constitutive model for SMA and finite element analysis of active metal matrix composites. Mech Compos Mater Struct 3(2):153–179

68. Lagoudas DC, Hartl D, Chemisky Y, Machado L, Popov P (2012) Constitutive model for the numerical analysis of phase transformation in polycrystalline shape memory alloys. Int J Plast 32:155–183

69. Gurtin ME (1982) An introduction to continuum mechanics. Academic Press, New York

70. Maugin GA (1999) The thermomechanics of nonlinear irreversible behaviors. World Scientific, Singapore

71. Rivlin RS (1997) On the principles of equipresence and unification. In: Barenblatt GI, Joseph DD (eds) Collected papers of RS Rivlin. Springer, Berlin, pp 1425–1426

72. Gurtin ME, Anand L (2009) Thermodynamics applied to gradient theories involving the accumulated plastic strain: the theories of Aifantis and Fleck and Hutchinson and their generalization. J Mech Phys Solids 57(3):405–421

73. Lele SP, Anand L (2008) A small-deformation strain-gradient theory for isotropic viscoplastic materials. Philos Mag 88(30–32):3655–3689