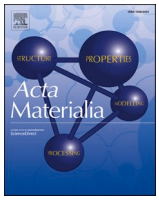
Acta Materialia 274 (2024) 119998

Contents lists available at ScienceDirect

Acta Materialia

journal homepage: www.elsevier.com/locate/actamat



Thermodynamic rationale for transformation-induced dislocations in shape memory alloys

Ahmed Sameer Khan Mohammed , Huseyin Sehitoglu *

Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, 1206W Green St, Urbana, IL 61801, USA

ARTICLE INFO

Keywords:
Shape memory alloys
Phase transformations
Micromechanics
Martensitic transformations

ABSTRACT

Shape Memory Alloys (SMAs) undergo stress-induced martensitic phase-transformation affording a "superelastic" behavior with functional applications. Such behavior is undermined by the formation and accumulation of irreversible residual strains in each cycle. These residual strains arise from transformation-induced dislocation-emission and current understanding has postulated the microstructural role of emitted dislocations to accommodate lattice-mismatch while also observing a preference to occur during reverse-transformation. This study develops a thermodynamic framework to offer a causal explanation for dislocation-emission from Gibbs' free energy considerations. Superelastic stress-strain curves for a reversible pathway without emitted-dislocations and for an irreversible pathway with emitted-dislocations are derived. The role of emitted-strain in relaxing the transformation-strain of the martensitic-inclusion and in accruing residual strain is proposed. It is shown that both pathways obey the first law of thermodynamics but it is the second law of thermodynamics that dictates the path preference. It is shown that the irreversible path achieves the critical condition for spontaneous reverse-transformation at a higher stress-level than the reversible path. Thus, the irreversible pathway initiates earlier during unloading and is thermodynamically selected during reverse-transformation. The driving-forces associated with the irreversible pathway are analyzed to establish the cause of emitted-dislocations and why it is thermodynamically preferred despite offering a higher lattice-friction barrier. Consequently, a new approach to target fatigue-resistant SMA properties is offered, focusing on the interplay of the individual driving-forces coming from the elastic strain-energy, work-interaction, and lattice-friction, as revealed by the thermodynamic framework.

Introduction

Shape Memory Alloys are materials that exhibit the capacity to sustain large reversible strains under the influence of one or multiple physical stimuli such as temperature, mechanical load, or magnetic fields [1–4]. This capacity is attributed by a microstructural mechanism of diffusionless phase-transformation between a parent phase, named austenite, and a product phase, named martensite. Consequently, SMAs find utility as actuating functional materials in biomedical, aerospace, and automotive domains [5–9]. However, the reversibility of deformation in SMAs is not perfect and following each cycle of deformation there is a finite irreversibility that remains in the form of residual strain within the material. The accumulation of such residual strain over several actuating cycles leads to functional failure, where the remnant magnitude of reversible strain is significantly diminished compared to the first cycle. This phenomenon is termed functional fatigue and it critically dictates the useful life of SMAs in application. Therefore, from a technological standpoint, it is crucial to understand functional fatigue of SMAs. The current study focuses on the thermodynamic origin of such irreversible residual strains from cyclic phase-transformation.

In addition to the technological importance, the problem of functional fatigue offers scientific challenges that are necessary to understand. It has therefore been a widely researched topic over decades [10–21]. Experimental studies have established that the accumulating irreversible strains are due to the formation and accumulation of transformation-induced dislocations in the austenite phase [22–28]. Such dislocation activity occurs at an applied stress-level of phase-transformation that is significantly lower than the plastic yield stress of each of the phases (austenite and martensite) [29–31]. Two dominant schools of thought have emerged. They are presented below along with a brief discussion of their merits and challenges:

* Corresponding author.
E-mail address: huseyin@illinois.edu (H. Sehitoglu).

https://doi.org/10.1016/j.actamat.2024.119998
Received 25 February 2024; Received in revised form 9 May 2024; Accepted 10 May 2024
Available online 11 May 2024

i. Dislocations accommodating interface lattice-mismatch: During phase transformation, the transformation front or the "habit-plane" of the transformation sustains large misfit strains to accommodate the lattice-mismatch between the austenite and martensite phases on either side. Due to this localized effect it is postulated that dislocations from the slip-system of the austenite phase form to relieve the high magnitude of interface strains [32]. In fact, it was shown that if SMA composition is engineered to improve interface compatibility between the phases, the magnitude of residual strain is reduced [33,34]. Continuum constitutive models have also been developed based on the phenomenology that localized plasticity develops around the transformation front [35,36].
ii. Dislocation-emission during reverse-transformation: It is postulated that transformation-induced dislocations form only during the reverse transformation. Results from scanning and transmission electron microscopy mapped the traces of emitted dislocations at the receding transformation front [22,23,37,38]. These traces were organized in parallel loops showed a strong alignment with the internal twin boundaries in the martensitic phase. A dislocation-reaction resulting from the interaction of internal twinning partials to form an emissary dislocation was proposed [37], more recently rationalized through molecular-statics simulations [39,40].

Consider the merits and challenges of both propositions (i) and (ii) above. Since the development of the field of topological modeling of crystalline interfaces [41], multiple habit planes of SMAs have been analyzed and studied experimentally through TEM as well. While the necessity of interface disconnections (interface dislocations with a stepped character) has been proposed at the habit-plane interface, no necessity of austenite dislocations has been shown or observed in high-resolution atomic-scale images. Furthermore, if it is the local interface stresses that activate dislocation activity, then residual strain accumulation is expected during the forward phase-transformation from austenite-to-martensite as well. This is in conflict with experimental observations where the dislocation-traces seem to be emitted selectively in the reverse-transformation [22,23,25,37,38]. Two experimental TEM images from refs. [25,37] are reproduced in Fig. A1, in the Appendix, showing the emanating dislocation-traces from the transformed martensite. However, to the merit of proposition (i), irrespective of whether or not dislocations are needed to constitute the transformation front, it makes sense that their presence can still act to relieve the strain-energy of the martensitic inclusion.

Consider the nature of strain induced by transformation-induced dislocations, henceforth referred to as the emitted strain, required to relax the inclusion's strain energy. The strain-energy of the martensitic inclusion is dictated by its eigen-strain that quantifies the lattice-mismatch strain between the austenite and martensite phases. In the absence of emitted-strain, refer to the martensite's eigen-strain as the nascent transformation-strain. To relax the inclusion's strain energy, the emitted strain must reduce the eigen-strain of the martensite inclusion. In other words, the emitted strain must strain the austenite matrix to reduce lattice-mismatch and better accommodate the martensitic inclusion. Consequently, the emitted strain must be of the same sense as that of the nascent transformation-strain of the martensite. In this manner, the eigen-strain of the martensitic inclusion is diminished, starting from its magnitude as the nascent transformation-strain to an "effective" eigen-strain undermined by the magnitude of the emitted strain in the austenitic matrix. For instance, consider an example of a superelastic mechanical response of an SMA under the influence of a tensile load along a certain crystal orientation. The favored martensitic variant during transformation would be one whose transformation-strain can maximize this tensile component (in other words, a Schmid-selected martensitic variant). From the arguments above, the emitted strain must also introduce a tensile strain aligned with the transformation-strain. Such emitted strain accumulates as residual strain upon unload, after the completion of one cycle of transformation. Thus, the residual strain caused by the transformation-induced dislocations is tensile in nature, or more generally, introduces the same nature of strain along the loading-direction as introduced by the martensite's transformation-strain. Such behavior of the residual strain is well known phenomenologically, where the residual strain is aligned with the direction of the applied load, or more generally, of the same sense as the functional transformation strain (see in refs. [42–44] for example). Conversely, given the phenomenological evidence of the nature of residual strain in superelasticity, it is plausible that the microstructural role of the emitted strain is indeed to relax the strain-energy of the transformation.

With regards to proposition (ii), the open question remaining is why the reverse-transformation pathway (from martensite to austenite) is energetically preferred for dislocation-emission. This study focuses on providing a thermodynamic explanation of why transformation-induced dislocations form. The developed framework reconciles both propositions above, showing that it is thermodynamically favorable for the strain-energy of the transformed-phase to relax, following proposition (i) above, and for the transformation-induced-dislocations to occur selectively in the reverse-transformation path, following proposition (ii). Consequently, there is a net accumulated residual strain along the direction of applied load, after one cycle of superelastic transformation. The developed framework addresses the general "driving force" of the transformation involving the thermodynamic stability of the phases, the continuum interaction energies with external load, and frictional barriers to be overcome from the motion of the transformation front to the emission of dislocations. Both the first and second laws of thermodynamics are invoked. By calculating the Gibbs free-energy of a perfectly reversible transformation cycle and that of an irreversible transformation cycle with emitted-strain, it is shown that the irreversible pathway is thermodynamically preferred, providing for the first time a causal explanation for why transformation-induced dislocations occur spontaneously in SMAs.

Methodology and results

Basic definitions

The micromechanical model of the transformation is presented in this subsection. The material parameters of the model are defined along with the analytical definitions essential to the thermodynamic framework developed in succeeding subsections. The scenario considered in this study is the phenomenon of "superelasticity" where the martensite nucleates within the austenitic phase under an external applied load. The austenite matrix is taken to be a single crystal with the direction of applied load, $\hat{\nu}$, parallel to the [011] crystallographic direction. Therefore, the applied load is given by:

$$\underline{\underline{\sigma^0}} = \sigma_0(\hat{\nu} \otimes \hat{\nu}); \sigma^0_{ij} = \sigma_0 \nu_i \nu_j \tag{1}$$

where σ_0 is the magnitude of the applied load. The component of the strain-tensor parallel to the applied load is given by ε_0. The matrix is further idealized as a homogeneous isotropic medium with Young's modulus E and Poisson's ratio ν. The martensite is treated as a penny-shaped Eshelby inclusion nucleating within the austenitic matrix [45]. The aspect ratio of the inclusion is given by the ratio of diameters across the major and minor axes i.e. a_3/a (refer Fig. 1). The $a_1 - a_2$ plane of the penny-inclusion (with $a_1 = a_2 = a$) is parallel to the habit plane of the nucleating martensitic variant. The transformation strain of the inclusion is given by:

$$\underline{\underline{\varepsilon^{tr}}} = (\vec{s} \otimes \hat{m}); \varepsilon^{tr}_{ij} = s_i m_j \tag{2}$$

where $\hat{m}$ is the normal to the habit plane and $\vec{s}$ is the transformation shear. For a given SMA system, the phenomenological theory of

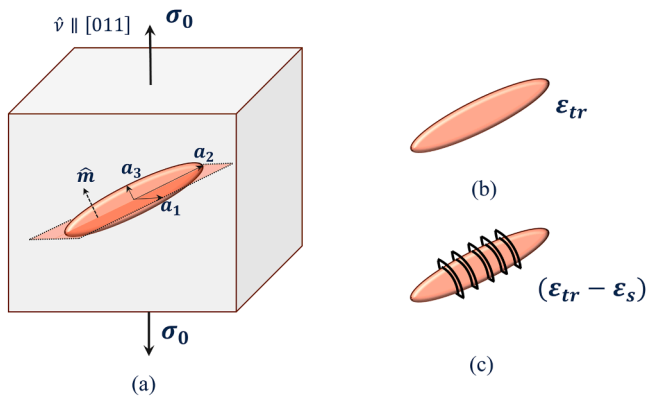


Fig. 1. Schematic of a ellipsoidal martensitic inclusion nucleating inside an austenitic matrix indicating (a) the direction of loading is $\hat{\nu} \parallel [011]$, and magnitude of applied load σ_0, normal to the habit plane of the transformation $\hat{m}$, the axes of the penny-shape ellipsoidal inclusion with $a_1 = a_2 = a$ and $a_3 < a$ (b) transformation strain of the inclusion ε_{tr} (b) effective transformation strain of the inclusion $(\varepsilon_{tr} - \varepsilon_s)$ diminished by the strain induced by presence of transformation-induced dislocations ε_s.

martensite [46–49] can be used to solve for the normal and transformation-shears of all martensitic variants that can nucleate within the austenite phase. In this study, it is assumed that the nucleating martensitic variant belongs to a family of 24 Habit-Plane-Variants commonly nucleating in NiTi under applied load [50]. In this family, the habit plane normal is given by $\hat{m} \in \{\overline{0.8889},\ \overline{0.4044},\ 0.2152\}$ and the direction of transformation shear by $\vec{s} \in \langle 0.4114,\ \overline{0.4981},\ 0.7633\rangle$. Under the influence of the applied load, the martensite variant exhibiting the maximum Schmid factor $SF_m = \left(\underline{\underline{\sigma_0}} : \underline{\underline{\varepsilon_{tr}}}\right)/\sigma_0$ is selected (where ":" represents the tensorial dot-product). For $\hat{\nu} \parallel [011]$, the selected martensite variant has $\hat{m}$ normal to habit plane $(0.8889,\ 0.4044,\ 0.2152)$ and $\vec{s} = [\overline{0.4114},\ 0.4981,\ 0.7633]$. The component of the transformation strain along the direction of the applied load is given by:

$$\varepsilon_{tr} = \left|\underline{\underline{\varepsilon_{tr}}} : (\hat{\nu} \otimes \hat{\nu})\right| = (s_i \nu_i)(m_j \nu_j) \tag{3}$$

Next, we setup the thermodynamic framework for superelastic phase-transformation, similar to what was employed in refs. [50–52]. Only key elements of the model are described here, and the reader is referred to the aforementioned references for a more detailed exposition. The nucleation and growth of the Schmid-selected martensitic variant, having volume fraction f and transformation strain $\underline{\underline{\varepsilon_{tr}}}$ (Eq. (2)), is considered under applied load at a constant temperature T. Note that the volume fraction f can vary continuously between the limits 0 and 1, with $f = 0$ representing a state with no martensite and $f = 1$ representing a state with all martensite and no austenite phase. And now consider the thermodynamics of a unit volume of the matrix. The energy-balance for the system is given by the first law of thermodynamics:

$$E_{EL}(\sigma_0, f) + E_{THERM}(f) - W = 0 \tag{4}$$

where $E_{EL}(f)$ is the internal elastic energy of the system, $E_{THERM}(f)$ is the difference in chemical bonding energy between the martensite and austenite phases, and W is the applied work input. In other words, the applied work is stored in the form of internal elastic energy and chemical energy of the transformed phase. In this study, we focus on the superelastic phase transformation and do not consider any heat transfer in and out of the system. Therefore heat-energy is not involved in Eq. (4).

The chemical energy difference between the martensite and austenite phase at the temperature T is given by:

$$E_{THERM}(f) = B(T - T_0)f \tag{5}$$

The chemical energy arises from a difference in bonding energy between the martensite and austenite phases. The quantity B is proportional to the Clausius–Clapeyron slope of the SMA and T_0 is the equilibrium temperature of the austenite and martensite phases, both of which are intrinsic material parameters for the SMA. The elastic energy is given by:

$$E_{EL}\left(\sigma_{ij}^0, f\right) = \frac{1}{2}\sigma_{ij}^0 \varepsilon_{ij}^{EL} - \frac{1}{2} f \varepsilon_{ij}^{tr}\left(C_{ijkl}\left(S_{klmn}\varepsilon_{mn}^{tr} - \varepsilon_{kl}^{tr}\right)\right) \tag{6}$$

where S is the Eshelby-tensor for the nucleated martensitic inclusion [45], and ε_{ij}^{EL} is the elastic strain sustained by the austenitic matrix under the presence of external stress σ_{ij}^0. They are, therefore, related through the constitutive relation given by:

$$\varepsilon_{ij}^{EL} = \left(C^{-1}\right)_{ijkl}\sigma_{kl}^0 \tag{7}$$

where C is the elastic-constant tensor. The work input is a path-dependent term given by the integral:

$$W = \int \sigma_{ij}^0 d\varepsilon_{ij}^0 \tag{8}$$

where ε_{ij}^0 is the total strain. During the loading stage, the work term simplifies to:

$$W = \frac{1}{2}\sigma_{ij}^0 \varepsilon_{ij}^{EL} + \sigma_{ij}^0 \varepsilon_{ij}^{tr} f \tag{9}$$

However, Eq. (9) does not hold after unloading, and the integral of the loading history, as given in Eq. (8), must be used to evaluate the work input. To predict the system state at which phase-transformations initiate or progress, the second law of thermodynamics is to be invoked. At any given state of the system, all the contributors to the Gibbs' free energy of the system are listed. The elastic self-interaction energy of the system is given by:

$$E_{SELF-EL}\left(\sigma_{ij}^0, f\right) = -\frac{1}{2} f \varepsilon_{ij}^{tr}\left(C_{ijkl}\left(S_{klmn}\varepsilon_{mn}^{tr} - \varepsilon_{kl}^{tr}\right)\right) \tag{10}$$

The work-interaction energy between the external load and the martensitic inclusion is given by:

$$W_{inter}\left(\sigma_{ij}^0, f\right) = \sigma_{ij}^0 \varepsilon_{ij}^{tr} f \tag{11}$$

Comparing Eqs. (10) and (11) with Eqs. (6) and (9) respectively, note that the interaction energy terms contributing to the Gibbs' free energy are a part of the total energy exchanges in the system. Both the work-interaction energy and elastic self-interaction energy can be simplified in terms of the scalar components of the applied load σ_0, transformation strain ε_{tr} as follows:

$$W_{inter}(\sigma_0, f) = \sigma_0 \varepsilon_{tr} SF_m f \tag{12}$$

$$E_{SELF-EL}(\sigma_0, f) = \frac{1}{2} f \varepsilon_{tr}^2 C_{ES}$$

where SF_m is the Schmid factor for the transforming inclusion computed previously, E is the Young's modulus of the matrix and C_{ES} is a positive constant term dependent on the aspect ratio a_3/a of the inclusion through the Eshelby tensor. The Gibbs' free energy for the system is determined, given by:

$$G(\sigma_0, f) = E_{THERM}(f) + E_{SELF-EL}(f) - W_{inter}(\sigma_0, f) \tag{13}$$

All model parameters of this study are listed in Table 1 at the end of the section.

Table 1
Model parameters used for the thermodynamic framework in this study.

Young's Modulus, E	10 GPa
Poisson's ratio, ν	0.3
Chemical energy constant, B	0.1 MPa/K
Equilibrium temperature, T_0	273 K
Critical driving force of transformation, F_c	2 MPa
Room temperature, T	298 K
Maximum applied strain, ε_{max}	5 %
Direction of applied load, $\hat{\nu}$	[011]
Transformation shear, $\vec{s}$	$[\overline{0.4114},\ 0.4981,\ 0.7633]$
Habit plane with normal $\hat{m}$	$(0.8889,\ 0.4044,\ 0.2152)$
Transformation strain, ε_{tr} (determined from $\vec{s}$, $\hat{m}$, $\hat{\nu}$ in Eq. (3))	5.11 %
Emission-rate hardening constant, k	10^4
Emission-rate hardening exponent, n	1.0

Reversible pathway of phase-transformation without emitted dislocations

First the simpler case of a perfectly reversible phase-transformation without transformation-induced dislocations is examined. This section outlines the derivation of the stress–strain curve OABCDO plotted in Fig. 2. Starting from the austenite phase at O ($f = 0$), increasing magnitude of external load is applied to the material. The energy balance of Eq. (4) is continually satisfied. The mechanical response remains linear-elastic, until a critical point is reached where the following critical condition is achieved at A:

$$F(f=0) = -\left.\frac{\partial G}{\partial f}\right|_{f=0} = F_c \tag{14}$$

where F is the applied driving force obtained from the gradient of reduction of the Gibb's free energy and F_c is the critical driving force for the transformation to be overcome. Define the individual contributors to the driving force as follows:

$$\begin{aligned} F_W &= \frac{\partial W_{inter}}{\partial f} \\ F_{SELF} &= \frac{\partial E_{SELF-EL}}{\partial f} \\ F_{THERM} &= \frac{\partial E_{THERM}}{\partial f} \end{aligned} \tag{15}$$

Then, from Eqs. (14) and (13), the applied driving force during the forward transformation, $F(f)$ is given by:

$$F = F_W - F_{THERM} - F_{SELF} \tag{16}$$

The physical interpretation is as follows. The net driving force for the forward transformation comes from the applied work, countering the driving forces arising from an increase of chemical energy and self-elastic energy. The critical driving force to be overcome, F_c, corresponds to barriers associated with the lattice friction of the transformation front and is hence an intrinsic material parameter. At this point, the reduction of Gibbs' free energy for unit increase in the volume-fraction of martensite crosses the lattice friction for the transformation. From the second law of thermodynamics, this criticality condition indicates that the state-variable f will spontaneously increase signaling the nucleation of the martensitic phase within the austenitic phase. Therefore, martensitic transformation initiates at A and the corresponding level of stress is the critical transformation stress $\sigma_0 = \sigma_F$, obtained by solving Eq. (14). An analytical expression can be derived for the forward transformation stress, given by:

$$\sigma_F = \frac{B(T - T_o) + 0.5\varepsilon_{tr}^2 C_{ES} + F_c}{\varepsilon_{tr} S F_m} \tag{17}$$

The strain at this point is given by $\varepsilon_A = \frac{\sigma_F}{E}$. The physical interpretation is as follows. The required stress for transformation is one that can provide enough work-interaction energy to overcome barriers provided from the lattice friction, the chemical energy barrier, and the self-energy of the inclusion in order to form martensite. Consider the energy-balance (first law of thermodynamics) on the pathway AB. In this case, Eq. (4) is re-defined as:

$$E_{TOT} = E_{THERM}(f) + E_{EL}(\sigma_0, f) + E_F - W = 0 \tag{18}$$

where E_F is the energy dissipated through lattice friction, given by:

$$E_F = \int_0^f F_c df \tag{19}$$

Note that this dissipated energy is not a state-dependent variable but a path-dependent. In other words, the frictional energy dissipated depends on the transformation history until the current state. Next step is to determine how the volume fraction of the evolving martensite is determined during the forward transformation. Along AB, since the second-law (Eq. (14)) is satisfied at each point, a consistency condition can be derived given by:

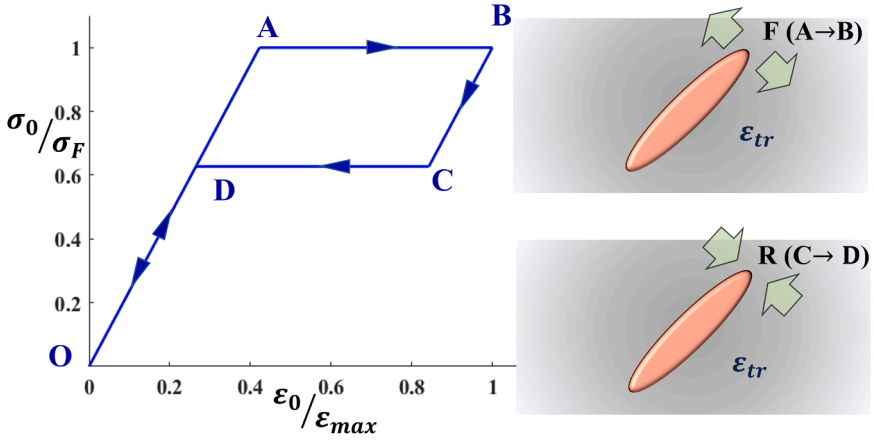


Fig. 2. Superelastic mechanical response in the absence of transformation-induced dislocations, schematically illustrating growth of the martensitic inclusion during forward (F) transformation, from point A to B, and subsequent shrinkage of the inclusion during reverse (R) transformation from C to D; the stress is normalized by the forward transformation stress σ_F (Eq. (17)), the strain is normalized by the maximum applied strain ε_{max} and the transformation strain of the inclusion is given by ε_{tr}.

$$\Delta F = \frac{\partial F}{\partial \sigma_0}\Delta\sigma_0 + \frac{\partial F}{\partial T}\Delta T + \frac{\partial F}{\partial f}\Delta f = \Delta F_c \tag{20}$$

It is presumed that there is no hardening associated with the critical driving-force of the transformation, therefore $\Delta F_c = 0$ on AB (Fig. 2). Also, no temperature change is assumed during the transformation for simplicity, i.e. $\Delta T = 0$. The strain from A to B increments as:

$$\Delta\varepsilon_0 = \frac{\Delta\sigma_0}{E} + \varepsilon_{tr}\Delta f \tag{21}$$

For a given strain-increment $\Delta\varepsilon_0$, the volume fraction increment Δf and stress-increment $\Delta\sigma$ can be solved for, from Eqs. (20) and (21). Without hardening, it turns out that $\Delta\sigma_0 = 0$ And therefore the strain at any point from A to B is given by:

$$\varepsilon_0 = \varepsilon_{tr} f + \frac{\sigma_F}{E} \tag{22}$$

where f is the increasing volume fraction of martensite. And when the maximum strain reaches $\varepsilon_0 = \varepsilon_{max}$ at B, the SMA is unloaded. The maximum transformed volume fraction at B is thus given by (22) as $f_{max} = \frac{1}{\varepsilon_{tr}}\left(\varepsilon_{max} - \frac{\sigma_B}{E}\right)$. Upon unload, the material unloads elastically until the critical point for reverse transformation is reached (point C in Fig. 1), characterized by:

$$-F(f = f_{max}) = \left.\frac{\partial G}{\partial f}\right|_{f=f_{max}} = F_c \tag{23}$$

In other words, the reduction in the Gibbs' free energy for a unit reduction in the volume fraction of martensite must equal the critical driving force to reverse the transformation. In terms of the driving forces, Eq. (23) can be re-written as:

$$F_{THERM} + F_{SELF} - F_W = F_c \tag{24}$$

Physically interpreted, the driving force from the chemical energy and self-elastic energy of the inclusion overcomes the driving force from the work-interaction energy and the lattice-friction to initiate reverse transformation. The critical driving force corresponds to the same lattice-friction for motion of the transformation front as was encountered during the forward-transformation path, thus the same F_c is employed. And here again, note that the energy-balance corresponding to the first law of thermodynamics is given by Eq. (18). The only change being that, during the reverse transformation, the dissipated frictional energy is given by,

$$E_F = \int_0^{f_{max}} F_c df + \int_f^{f_{max}} F_c df \tag{25}$$

which includes the dissipation during the forward transformation, from $f = 0$ to $f = f_{max}$, and during the reverse-transformation, from $f = f_{max}$ to any point during the reversal with a volume fraction $f \geq 0$. Denote E_F^{max} as the total frictional energy dissipated at completion of the superelastic stress–strain cycle. The trajectories of each energy term and driving-force term respectively involved in the energy-balance Eq. (18) and driving-force equilibrium Eq. (24), are plotted in Fig. 3 for the reverse-transformation from B to O.

The critical reverse-transformation stress at C can be solved for from Eq. (23). Similar to the forward transformation, an analytical expression can be obtained for this reverse-transformation stress, given by:

$$\sigma_R = \frac{\left(B(T - T_o) + 0.5\varepsilon_{tr}^2 C_{ES} - F_c\right)}{\varepsilon_{tr} S F_m} \tag{26}$$

The physical interpretation is as follows. The required stress for reverse-transformation is one at which the chemical energy and self-elastic energy of the inclusion is sufficient to overcome the work-interaction energy of the inclusion and critical driving force, causing the inclusion to recede in size or volume fraction. The consistency condition during the reverse-transformation also comes out to be $\Delta\sigma_0 = 0$ as in the forward case, and path CD is parallel to AB. At point D, $f = 0$ is achieved, implying the material has fully transformed back to austenite, and the SMA unloads elastically back to O. Thus the path OABCDO defines a perfectly reversible transformation pathway without dislocation-emission. With this foundation, the obvious question is to ask if transformation-induced dislocations have a thermodynamic basis. If the SMA has a symmetric forward and reverse path without involvement of transformation-induced dislocations why does it not prefer such a path?

Irreversible pathway of phase-transformation with emitted dislocations

Next, an analytical thermodynamic framework for transformation-induced dislocations is proposed. This section outlines the derivation of the stress–strain curve $OABC_SD_SO_S$ plotted in Fig. 4. For this derivation, only one phenomenological supposition is made, noted as follows:

Phenomenological supposition: There are no transformation-induced dislocations during the forward transformation. This assumption follows the school of thought (ii) discussed in section 1.

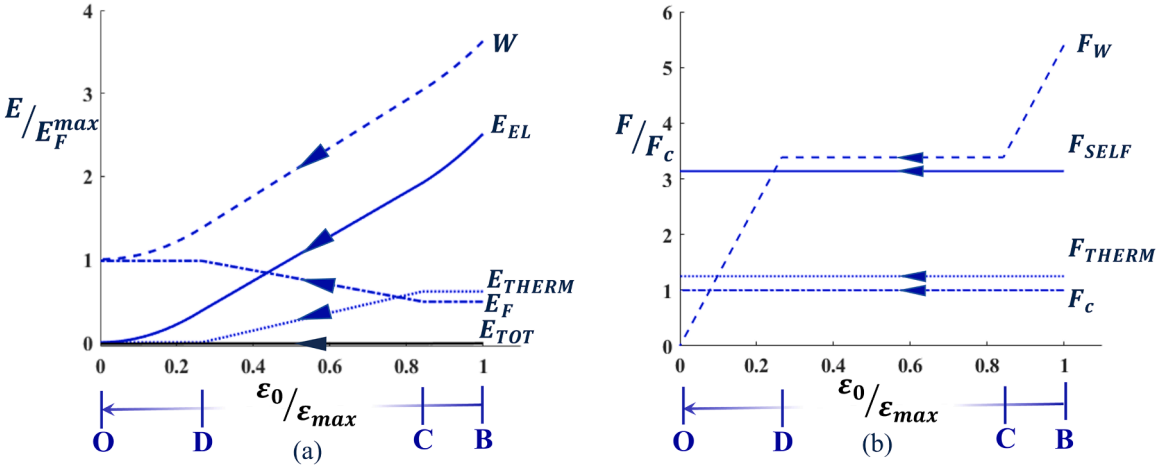


Fig. 3. Trajectories of individual (a) energy-terms and (b) driving-force terms during the reverse-transformation of the martensitic inclusion; the energies are normalized by the total frictional energy E_F^{max} dissipated during the full superelastic stress-strain cycle and the driving-forces are normalized against the critical driving force for reverse-transformation F_c.

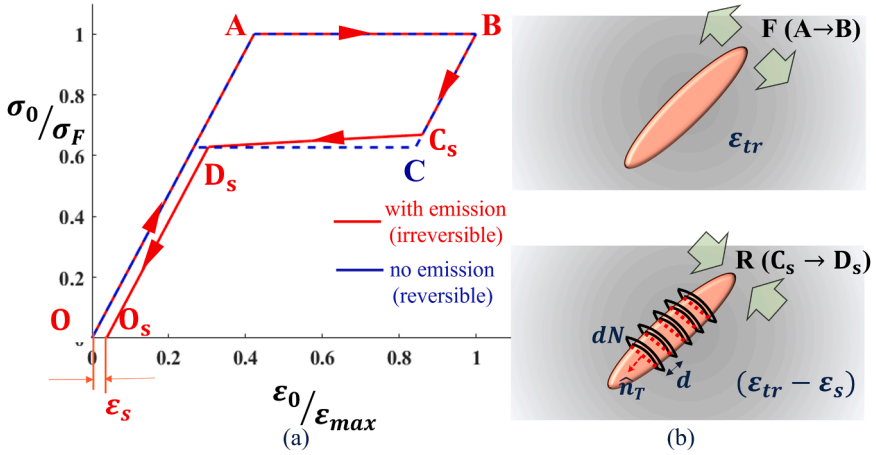


Fig. 4. Superelastic mechanical response with transformation-induced dislocations with schematic illustration of growth of the martensitic inclusion during forward (F) transformation, from point A to B, and subsequent shrinkage of the inclusion during reverse (R) transformation from C_s to D_s with emitted dislocations (a) the point of reverse transformation, C_S, occurs at a higher stress than with the reversible pathway (at C), thereby thermodynamically preferring transformation-induced dislocations ending with irreversible residual strain ε_P at O_s; the stress is normalized by the forward transformation stress σ_F (Eq. (17)), the strain is normalized by the maximum applied strain ε_{max} (b) Schematic illustration of the incrementally emitted dN dislocation loops, from internal twin interfaces with normal $\hat{n}_T$ and spacing d, effectively relaxing the transformation strain of the inclusion to $(\varepsilon_{tr} - \varepsilon_s)$.

The above supposition will be revisited and a causal explanation will be provided from a thermodynamics standpoint, through a Gibbs' free energy formulation. With no emitted-dislocations during the forward transformation, the forward path OAB is identical to that in section 2.1. The unloading regime is considered with special focus on the reverse phase transformation. With dislocations emitted as parallel dislocation loops, consider the form of residual strain that they introduce into the austenitic matrix. This process must be considered incrementally. During the reverse-transformation, if the volume-fraction of martensite reduces infinitesimally by $-df$, an increase in the emitted residual strain $d\varepsilon_s$ is introduced. This residual strain arises from both the formation of loops around the inclusion and any plastic strain accrued by the infinitesimal motion of each of these dislocations around the inclusion. Such an increment in emitted residual strain $d\underline{\underline{\varepsilon_S}}$ per unit volume of the austenite phase can be analytically expressed in the following form:

$$d\underline{\underline{\varepsilon_s}} = \left(\frac{dN}{d} + d\varepsilon^P\right)\left(\vec{b}_A \otimes \hat{n}_T\right) \tag{27}$$

where dN is the number of dislocations emitted at each loop, d is the spacing between the loops, $\vec{b}_A$ is the Burgers vector of the emitted dislocations and $\hat{n}_T$ is the normal to the slip-plane of the dislocations, and $d\varepsilon^P$ is the incremental plastic strain from the slip of these loops. Following the mechanistic proposition in school of thought (ii), the dN dislocations are emitted from one of the internal twin interfaces within the martensite, d aligns with the spacing between consecutive twin interfaces and $\hat{n}_T$ is the normal to the internal twin boundary (refer Fig. 4 (b)). The strain from the emitted dislocation loops, $d\underline{\underline{\varepsilon_S}}$, accumulates incrementally as a residual strain within the matrix during the reverse transformation. The analytical expression (27) is forwarded solely from a physical standpoint to ground the causes of residual strain arising from the emitted dislocation loops. Going forward, a simpler form will be adopted for the thermodynamic formulation.

With the addition of residual strain into the matrix, the net transformation strain of the martensitic inclusion is effectively reduced, given by:

$$\varepsilon^*_{ij} = \varepsilon^{tr}_{ij} - \varepsilon^S_{ij} \tag{28}$$

Note that this is consistent with the school of thought (i) discussed in section 2. In other words, the emission of the dislocation loops around the inclusion effectively relaxes the lattice mismatch between the matrix and the inclusion, manifesting as a reduction in the effective transformation strain. The increment in residual strain per unit volume of austenite is then given by the slip-emission rate:

$$\varepsilon^D_{ij} = \frac{d\varepsilon^S_{ij}}{df} \tag{29}$$

where f is the volume-fraction of martensite. Note that the strain arises from these dislocation loops forming around the inclusion, and potentially from their plastic-slip around the martensite (Eq. (27)). This understanding is necessary to quantify the driving force associated with the formation of this residual strain and how it interacts with the transformation strain of the martensite in dictating the thermodynamics. For simplicity, it is assumed that $\underline{\underline{\varepsilon_S}}$, as a tensor, is parallel to the transformation strain $\underline{\underline{\varepsilon_{tr}}}$. This is given by:

$$\underline{\underline{\varepsilon_S}} = \varepsilon_S\ (\vec{s} \otimes \hat{m})/\varepsilon_{tr} \tag{30}$$

where ε_S is the accumulating residual strain in the direction of applied load (i.e. $\varepsilon_S = \underline{\underline{\varepsilon_S}} \cdot (\hat{\nu} \otimes \hat{\nu})$ consistent with its definition in Eq. (30)). The effective transformation strain along the direction of applied load is given by $\varepsilon_* = (\varepsilon_{tr} - \varepsilon_s)$. Correspondingly, the scalar emission-rate, ε_D, for the residual strain, ε_s, given by:

$$\varepsilon_D = \frac{d\varepsilon_S}{df} \tag{31}$$

Since the dislocation-emission occurs upon reverse transformation i.e. with reducing volume fraction f, the emission-rate ε_D is always of opposite sense to the emitted strain ε_S. In other words, for positive accumulating emitted strain ε_S during reverse-transformation, the emission-rate is negative i.e. $\varepsilon_D \leq 0$.

Next, the elastic-energy terms and work terms, E^S_{EL}, $E^S_{SELF-EL}$, W^S, W^S_{inter} must be re-defined. The elastic energy terms have the same form as defined previously in section 2, with the only difference being that the effective transformation strain is given by ε^*_{ij} and not ε^{tr}_{ij}. In other words, the elastic energy terms are now relaxed by the transformation-induced emitted strain, defined by:

$$E^S_{EL}\left(\sigma^0_{ij}, f, \varepsilon^S_{ij}\right) = \frac{1}{2}\sigma^0_{ij}\varepsilon^{EL}_{ij} - \frac{1}{2}f\varepsilon^*_{ij}\left(C_{ijkl}\left(S_{klmn}\varepsilon^*_{mn} - \varepsilon^*_{kl}\right)\right) \tag{32}$$

$$E^S_{SELF-EL}\left(\sigma^0_{ij}, f, \varepsilon^S_{ij}\right) = -\frac{1}{2} f \varepsilon^*_{ij}\left(C_{ijkl}\left(S_{klmn}\varepsilon^*_{mn} - \varepsilon^*_{kl}\right)\right) \tag{33}$$

The total work input, W^S, is still defined as the history-dependent integral in Eq. (8), with the only change that the stress-strain history corresponds to the irreversible path with dislocation emission (denoted by subscript S):

$$W^S = \int_S \sigma^0_{ij} d\varepsilon^0_{ij} \tag{34}$$

The work-interaction energy is re-defined based on the modified contribution from the effective transformation strain ε^*_{ij} and an additional contribution from the emitted strain ε^S_{ij}. This is given by:

$$W^S_{inter} = \sigma^0_{ij}\varepsilon^*_{ij} f + \sigma^0_{ij}\varepsilon^S_{ij} \tag{35}$$

The above energy terms can be simplified in terms of scalars $\sigma_0, f, \varepsilon_*$ as follows:

$$E^S_{EL} = \frac{\sigma_0^2}{2E} + \frac{1}{2} f \varepsilon_*^2 C_{ES} \tag{36}$$

$$E^S_{SELF-EL} = \frac{1}{2} f \varepsilon_*^2 C_{ES}$$

$$W^S_{inter} = \sigma_0 \varepsilon_* f S F_m + \sigma_0 \varepsilon_s S F_m$$

Furthermore, the driving force with dislocation-emission increases to:

$$F^S_c(\varepsilon_D) = F_c + F_S(\varepsilon_D) \tag{37}$$

where $F_S(\varepsilon_D)$ is the additional driving-force barrier to be overcome for a given emission-rate ε_D. The additional barrier F_S corresponds to the atomistic misfit energy in the core of the emitted dislocations, required to be able to form around the inclusion, and dissipated lattice-frictional energy for motion of these dislocations. For simplicity, it is only assumed that the driving force depends only on the emission-rate ε_D. And motivated from the power-law hardening approach used in continuum plasticity models, the following functional form is chosen for $F_S(\varepsilon_D)$:

$$F_S(\varepsilon_D) = F_c . k \varepsilon_D^{2n} \tag{38}$$

where $k \geq 0$ and $n \geq 0$ are material parameters. Only the misfit energy is added to the driving force and not the additional strain-energy arising from the dislocation strain fields. This is because the strain-energy of the dislocations are accounted for in the expressions for E^S_{EL} and $E^S_{SELF-EL}$ in Eqs. (32) and (33) respectively, by considering the change in effective transformation strain. The physical interpretations corresponding to the modified definitions of energy terms is as follows:

a. The elastic energies E^S_{EL} and $E^S_{SELF-EL}$ reduce due to the relaxation of the transformation strain.
b. The work-interaction term W^S_{inter} reduces due to the relaxing transformation strain while being supplemented by an additional contribution from the added emitted-strain in the system.
c. The driving force F^S_c is increased by the magnitude of core-misfit energy required to form the dislocation loops around the inclusion.

Therefore, the first-law of thermodynamics for reverse-transformation with dislocation-emission is given by:

$$E^S_{TOT} = E^S_{EL} + E_{THERM} + E^S_F - W^S = 0 \tag{39}$$

where E^S_F is the energy dissipated during reverse-transformation with dislocation-emission (analogous to E_F in Eq. (18) during forward transformation). It is given by,

$$E^S_F = \int_0^{f_{max}} F_c df + \int_f^{f_{max}} F^S_c(\varepsilon_D) df \tag{40}$$

indicating the cumulative contribution of the dissipated energy during the forward transformation (without dislocation-emission) and of the dissipated energy during reverse transformation (with slip-emission). With emitted dislocations, the dissipated frictional energy includes contributions from the lattice friction for the motion of the transformation front, the core-misfit energy of the emitted dislocations that are formed and lattice-friction associated with the plastic slip of these dislocations. The Gibbs' free energy becomes:

$$G^S(\sigma_0, f, \varepsilon_S) = E^S_{THERM}(f) + E^S_{SELF-EL}(f, \varepsilon_S) - W^S_{inter}(\sigma_0, f, \varepsilon_S) \tag{41}$$

From the second law of thermodynamics, the critical condition for reverse-transformation with transformation-induced dislocations is given by:

$$-F_S(f = f_{max}) = \left.\frac{\partial G^S}{\partial f}\right|_{f_{max}} = F_c\left(1 + k\varepsilon_D^{2n}\right) \tag{42}$$

The driving forces involved are defined as follows:

$$\begin{aligned} F^S_W &= \frac{\partial W^S_{inter}}{\partial f} \\ F^S_{SELF} &= \frac{\partial E^S_{SELF-EL}}{\partial f} \\ F^S_{THERM} &= \frac{\partial E_{THERM}}{\partial f} \end{aligned} \tag{43}$$

In terms of the driving forces, Eq. (42) can be re-written as:

$$F_{THERM} + F^S_{SELF} - F^S_W = F^S_c \tag{44}$$

The critical condition, given by Eq. (42), is checked at each decremental load σ_0 starting from the point of unloading at B. At each decremented load σ_0, the emission-rate ε_D is solved for from condition (42). Only when a negative solution for the emission-rate $\varepsilon_D \leq 0$ is found, reverse-transformation with dislocations is initiated. If a solution does not exist or an unphysical positive solution i.e. $\varepsilon_D > 0$ is found, reverse-transformation is not initiated and elastic unloading is continued. At point C_S, the critical condition (42) is satisfied for a physical emission-rate $\varepsilon_D \leq 0$. An analytical expression for the reverse-transformation stress at C_S can be obtained, given by:

$$\sigma^S_R = \frac{\left(B(T - T_o) + \left(0.5\varepsilon_{tr}^2 C_{ES} - f_{\max}\varepsilon_{tr} C_{ES}\bar{\varepsilon}_D\right) - F^S_c(\bar{\varepsilon}_D)\right)}{(\varepsilon_{tr} + (1 - f_{\max})\bar{\varepsilon}_D) S F_m} \tag{45}$$

where $\bar{\varepsilon}_D \leq 0$ is the emission-rate at C_S. Compare respective terms in Eq. (45) with those in Eq. (26) from the reversible transformation scenario. Note the change in each term and a comparison of their relative magnitudes are provided in Eqs. (46), (47) and (48). The terms from Eq. (26) will be listed in the Left-Hand-Side (LHS) and the corresponding terms from Eqs. (45) on the Right-Hand-Side (RHS).

$$0.5\varepsilon_{tr}^2 C_{ES} \leq \left(0.5\varepsilon_{tr}^2 C_{ES} - f_{max}\varepsilon_{tr} C_{ES}\bar{\varepsilon}_D\right) \tag{46}$$

$$F_c \leq F^S_c(\bar{\varepsilon}_D) = F_c . \left(1 + k\bar{\varepsilon}_D^{2n}\right) \tag{47}$$

$$\varepsilon_{tr} S F_m \geq (\varepsilon_{tr} + (1 - f_{max})\bar{\varepsilon}_D) S F_m \tag{48}$$

Consider the physical interpretation of the change in each term. Eq. (46) represents an effective increase in the driving force coming from the self-elastic energy. The emitted dislocations serve to relax the self-elastic strain energy of the inclusion. Therefore, the corresponding driving force favors the reverse transformation in order to favor such a reduction in self-elastic energy, supplementing the driving force with the summand $-f_{max}\varepsilon_{tr} C_{ES}\bar{\varepsilon}_D \geq 0$. Eq. (47) represents an increase in the driving

force to initiate reverse-transformation with emission of dislocations. There is an additional cost for the reverse-transformation to initiate coming from the misfit-energy of the dislocation cores, given by $F_c^S = F_c k \bar{\varepsilon}_D^{2n} \geq 0$. Finally, Eq. (48) physically implies that the effective transformation strain of the inclusion is dropping due to the presence of emitted dislocation loops. This causes a reduction in the work-interaction energy of the inclusion, and a reduction from its corresponding driving force. Taken together, the physical interpretation for the critical-stress of reverse-transformation at C_S, given by Eq. (45) is as follows. The required stress for reverse-transformation with transformation-induced dislocations is one at which the driving-forces from the chemical energy and the *relaxing* self-elastic energy of the inclusion is sufficient to overcome the *reduced* work-interaction energy of the inclusion and the *increased* critical driving force, causing the inclusion to recede in size or volume fraction.

Then reverse-transformation initiates along with dislocation-emission, accumulating the strain ε_S incrementally as the volume fraction of martensite f reduces. And throughout this process, the first law of thermodynamics holds as given by Eq. (39). The consistency condition during the reverse transformation is as follows:

$$\frac{\partial F_s}{\partial \sigma_0}\Delta\sigma_0 + \frac{\partial F_s}{\partial T}\Delta T + \frac{\partial F_s}{\partial f}\Delta f = \Delta F_c^S \tag{49}$$

In addition, the consistency condition for the stress and strain decrements must be satisfied, given by:

$$\Delta\varepsilon_0 = \frac{\Delta\sigma_0}{E} + \varepsilon_{tr}\Delta f + \varepsilon_D \Delta f \tag{50}$$

During unloading the stress is decremented, and for a given decrement $\Delta\sigma_0$, the change in volume-fraction Δf, and change in nominal strain $\Delta\varepsilon_0$ can be solved for from Eqs. (49) and (50). At point D_S, $f = 0$ is achieved implying the material has fully transformed back to austenite, and the SMA unloads elastically back to O_S. Thus, the path $OABC_SD_SO_S$ defines the irreversible transformation pathway with dislocation-emission (Fig. 4(a)). The trajectories of each energy term and driving-force term respectively involved in the energy-balance Eq. (39) and driving-force equilibrium Eq. (44), are plotted in Fig. 5 for the reverse-transformation from B to O_S.

In summary, note that both the reversible pathway (OABCDO path in dashed blue lines) and irreversible pathway ($OABC_SD_SO_S$ path in solid red lines), in Fig. 4(a) satisfy their respective conditions of energy balance, satisfying the first law of thermodynamics. But the question to ask is which of the two is preferred in reality. This is dictated by the second-law of thermodynamics. If the condition for criticality or the *spontaneity condition* for irreversible-transformation at C_S is achieved before the reversible condition C, the irreversible pathway is thermodynamically preferred and is observed in reality. In this study, the analytical framework does indeed find the critical point C_S (irreversible pathway with dislocation emission) to occur at a stress-level above C (reversible pathway with no emission), as shown in Fig. 4(a). It is therefore asserted that the irreversible pathway is indeed the thermodynamically preferred one and therefore the reason why dislocation-emission is observed in reality. The results are discussed further in section 3.

Discussion

The purpose of this study was to provide a rationale for transformation-induced dislocations from a thermodynamic standpoint. Two pathways of phase-transformation under external applied load were considered: (1) Reversible pathway OABCDO without emitted dislocations and (2) Irreversible pathway $OABC_SD_SO_S$ with transformation-induced dislocations resulting in residual strain. A methodology to derive the stress–strain history of both pathways was established in section 2. It was argued and shown that both pathways individually satisfy the first law of thermodynamics throughout (refer Figs. 3(a) and 5(a)). But the deciding factor that preferentially selects one pathway over the other is the condition of criticality derived from the second law of thermodynamics. Therefore, if the condition for transformation-induced dislocations, at C_S is achieved before the condition for reversible transformation, at C, the irreversible pathway is preferred. It is asserted that such a scenario occurs in reality, due to which it is plausible for reverse phase-transformation to be accompanied with dislocation-emission in SMAs.

Note that the critical driving-force barrier for transformation-with-dislocations, F_c^S, is higher than the corresponding critical barrier without dislocations, F_c, due to the additional cost associated with the dislocation-core energies of the emitted dislocations and lattice-friction associated with their slip. Given this higher barrier, why is the critical condition for transformation (point C_S) achieved earlier than that of reversible transformation? In other words, why is the stress-level corresponding to C_S higher than that of C? To understand this, the energy exchanges and driving forces in both pathways must be considered in further detail. Fig. 6 illustrates the change in energy terms and driving-force terms close to the point of initiation of the reverse-transformation. Only contributions from applied work, elastic energy and critical driving-force are considered. The contribution from the chemical energy difference E_{THERM} is not considered since its contribution to the energy balance and driving-force (through F_{THERM}) is the same for both the reversible and irreversible pathways. For the other contributors, the following behaviors are noted:

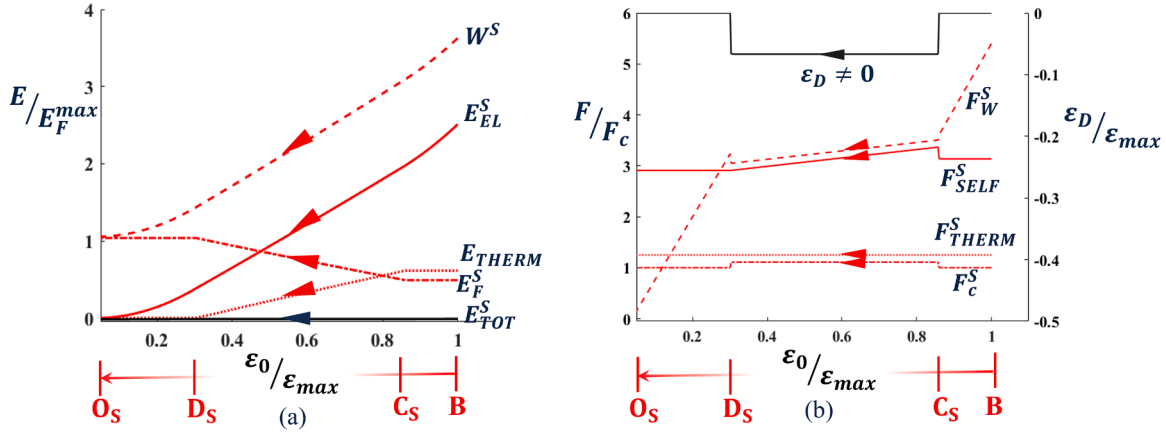


Fig. 5. Trajectories of individual (a) energy-terms and (b) driving-force terms during the reverse-transformation of the martensitic inclusion with dislocation-emission; the emission-rate ε_D is also plotted, showing a constant non-zero emission during the reverse-transformation regime C_S to D_S.

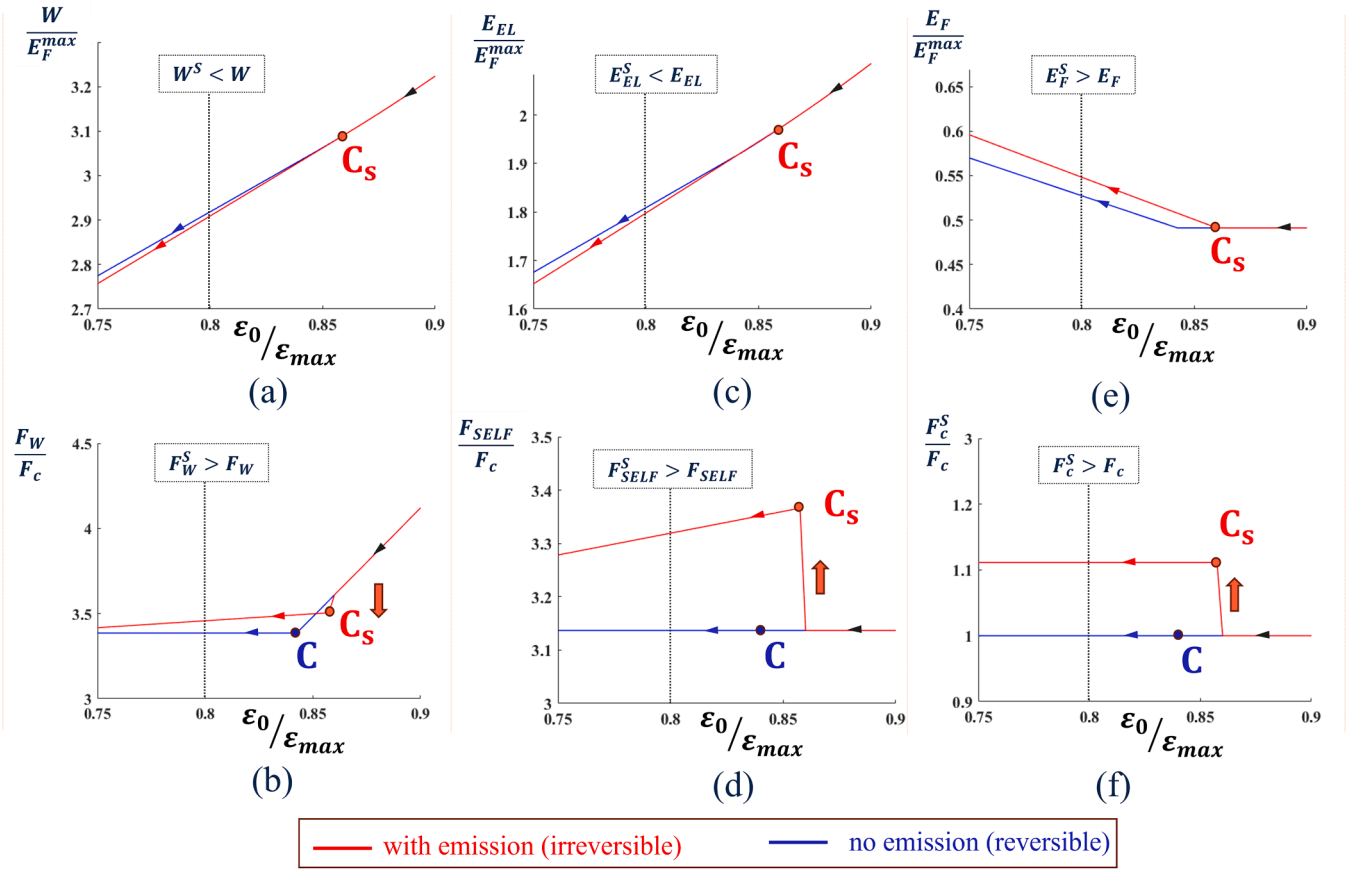


Fig. 6. Comparing the trajectories of energy-components and driving-force-components during reverse-transformation: (a) Applied work (b) Driving-force corresponding to work interaction energy (c) Elastic-energy of the inclusion (d) Driving-force corresponding to the self-elastic energy of the inclusion (e) Dissipated frictional energy (f) Critical driving force to be overcome for reverse-transformation; the arrows indicate jumps in the driving-force term at the point of reverse-transformation C_S in the irreversible pathway $OABC_SD_SO_S$.

a. Applied work: From Fig. 6(a), note that the applied work in the irreversible pathway W^S is lower in comparison to W. This initial drop in W^S is caused by the earlier initiation of the reverse-transformation at C_S, effectively reducing the area integral (Eq. (34)) in comparison to the W incurred in the reversible pathway. The corresponding driving force F_W^S also exhibits a drop at C_S due to the relaxing transformation strain of the inclusion (indicated by the arrow in Fig. 6(b)). However, the magnitude of the irreversible driving-force F_W^S is still higher than F_W due to the higher reverses-transformation stress at C_S, compared to that at C.
b. Elastic energy: The elastic energy of the inclusion reduces in the irreversible pathway, i.e. $E_{EL}^S < E_{EL}$ (Fig. 6(c)), due to the relaxation of transformation-strain caused by the emitted dislocation loops and consequent relaxation of the strain-energy. The propensity to reduce the strain-energy through such a mechanism contributes an increase to the driving-force for reverse-transformation. Such an increase can be seen in the jump in F_{SELF}^S at C_S (indicated by the arrow in Fig. 6(d)) exceeding the driving-force F_{SELF} enforced by the reversible pathway.
c. Dissipated frictional energy: The emission of dislocations contributes an energy cost to E_F^S in the form of core-energy of the emitted dislocations and frictional dissipation from their slip around the inclusion. The corresponding driving-force F_c^S is, therefore, also higher than F_c to accommodate this additional energy cost (Fig. 6(e) and (f)).

Now consider the coupled effect of the above driving forces in Eq. (44). There are three changes in the driving-force terms F_W^S, F_{SELF}^S, and F_c^S that occur with dislocation-emission (indicated by arrows in Fig. 6(b), (d) and (f)), that need to be considered simultaneously. As the material unloads, there is a certain point at which there is a potential gain in driving-force caused by the relaxing elastic-energy, given by F_{SELF}^S. This is accompanied by a potential drop in the driving-force from the applied work, F_W^S. Both changes favor the reverse-transformation to initiate. If both these favoring contributions are large enough to overcome the increased driving-force F_c^S accompanying dislocation-emission, then reverse-transformation initiates earlier at C_S, with dislocation-emission, instead of at C. Dislocation-emission is therefore thermodynamically favored to initiate spontaneously with the reverse-transformation, making the irreversible pathway the observed behavior during superelastic phase-transformation.

Next, we revisit the phenomenological supposition made at the beginning of section 2.3 and provide a thermodynamic argument explaining the phenomenology. The developed framework with dislocation-emission will now be applied to a hypothetical scenario of dislocation-emission during forward-transformation. Revisiting the loading phase from O, consider the possibility of initiating the forward-transformation with dislocation-emission. The energy terms in Eqs. (12) are redefined as follows:

$$W_{inter}(\sigma_0, f) = \sigma_0 \varepsilon_* SF_m f + \sigma_0 \varepsilon_S SF_m \tag{51}$$

$$E_{SELF-EL} = \frac{1}{2} f \varepsilon_*^2 C_{ES}$$

where $\varepsilon_* = (\varepsilon_{tr} - \varepsilon_S)$ is the relaxing transformation strain during forward transformation, and ε_S is the strain from dislocation-emission. Note that these functional forms are identical to that in Eq. (36) with the only

difference that the transformation strain is relaxing in the forward direction. In other words, the emission rate ε_D defined by Eq. (31) is positive, i.e. $\varepsilon_D \geq 0$. Now, following the derivation in section 2.2, particularly using Eqs. (13) and (14) with the modified energy definitions in (51), the critical point of forward-transformation with dislocation-emission, denoted by A_S, can be solved for. This critical stress is given by:

$$\sigma_F^S = \frac{\left(B(T - T_o) + 0.5\varepsilon_{tr}^2 C_{ES} + F_c^S(\widehat{\varepsilon}_D)\right)}{(\varepsilon_{tr} + \widehat{\varepsilon}_D)SF_m} \tag{52}$$

where $\varepsilon_D = \widehat{\varepsilon}_D$ is the emission-rate at the initiation of forward transformation. The critical stress in Eq. (52) is compared against the forward-transformation stress of the reversible pathway in Eq. (17) for a range of emission-rates $\widehat{\varepsilon}_D > 0$. The result is plotted in Fig. 7(a). Note that forward transformation stress σ_F^S is always larger than σ_F for any emission-rate. This is because the higher driving-force to transform-with-dislocations, given by $F_c^S(\widehat{\varepsilon}_D)$, has a dominating effect. In other words, during the loading phase, the critical condition to initiate forward transformation without dislocation-emission is achieved first, thereby initiating transformation without dislocations (refer Fig. 7(b)). Therefore, even though it is plausible to expect a relaxation-mechanism and consequent transformation-induced slip to be active during the forward-transformation, the thermodynamics of the process renders the mechanism inadmissible. Alternatively, dislocation-emission is thermodynamically preferred only in the reverse-transformation, causally explaining the phenomenological supposition made at the beginning of section 2.3.

The physical interpretation for the discussion in the previous paragraph is as follows: the higher slip-resistance of SMAs relative to the forward-transformation stress favors the forward-transformation to proceed without dislocation-emission. In the reverse-transformation, however, the increased barrier F_c^S is overcome by the relaxing strain-energy of the martensitic inclusion. Therefore, dislocation-emission is thermodynamically preferred only during the reverse-transformation. Note that the above statements rationalize both schools of thought (i) and (ii) discussed in section 1. The emission of dislocations does indeed serve to relax the strain-energy of the inclusion by accommodating the lattice-mismatch and reducing the effective transformation strain, consistent with (i). The relaxing strain-energy provides the driving-force necessary to induce dislocation-emission, thermodynamically preferring transformation-induced-dislocations to form in the reverse-transformation, consistent with (ii). In the absence of such a relaxation mechanism during the nascent forward-transformation phase, the increased barrier for dislocation-emission dominates and thermodynamically prefers forward-transformation without dislocations, also consistent with (ii).

The proposed framework offers a starting point to thermodynamically examine further characteristics of SMA behavior in functional fatigue. A future extension of the proposed framework would be to continue the analysis beyond point O_S, examining the influence of accumulating residual strain on the forward-transformation stress of subsequent cycles. Such an analysis would carry over the residual strain ε_S and consider the effective eigen-strain $\varepsilon_{ij}^* = \varepsilon_{ij}^{tr} - \varepsilon_{ij}^S$ instead of the nascent transformation strain ε_{ij}^{tr}. In addition, it is proposed that there can be an erasure-rate ε_{ij}^R (similar to the emission-rate ε_{ij}^D) accompanying a reduction of the residual strain during forward transformation as the martensite sweeps over the austenitic matrix containing residual dislocations. Such a framework could provide a thermodynamic explanation for the change of forward transformation stress or the upper plateau stress as observed over successive cycles of superelastic transformation. On a different note, the current study leveraged Habit-Plane Variant (HPV) solutions from the Phenomenological Theory of Martensite Crystallography (PTMC) to develop the thermodynamics of transformation. In doing so, an inherent assumption is that the characteristics of the transforming inclusion, namely the habit plane normal and transformation shear, remain independent of the state of stress in both forward/reverse transformation. Recent research on NiTi has proposed the need to consider elastic distortions of the austenite and martensite phases in deriving admissible HPVs with a strain-compatible interface between the phases [53–55]. The coupled influence of emitted strain on evolving the crystallography of reverse-transforming martensite has also been proposed [56]. It is plausible that the crystallography of the habit-plane interface and internal twin volume fraction of the martensite can evolve in the presence of elastic stress and emitted strain, ultimately influencing the thermodynamics of the process. In the forward transformation, the state of stress can dictate the crystallography of admissible HPVs and consequently the transformation strain. The critical point to initiate forward transformation, A, would therefore occur when the favored habit-plane variant has a transformation strain, at the current

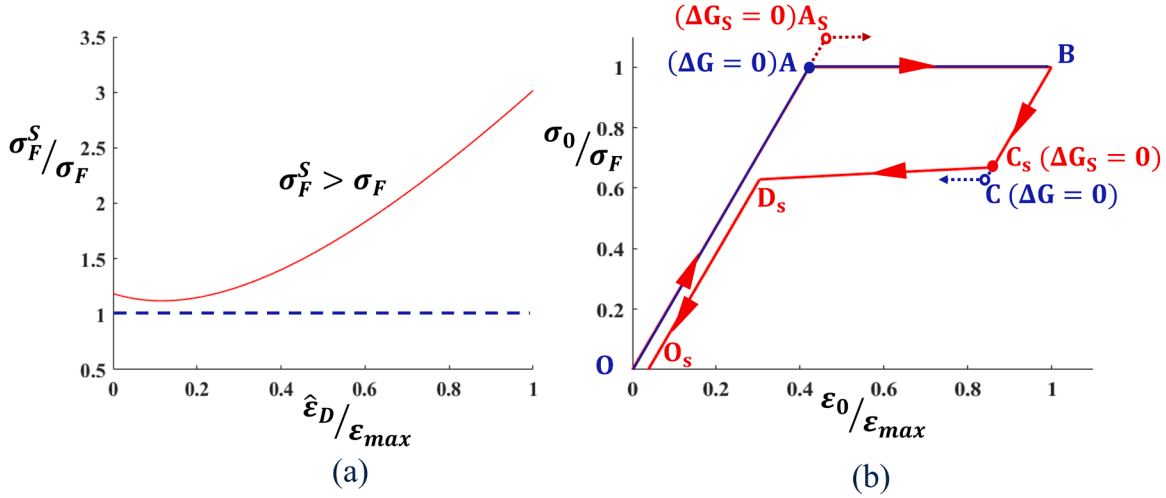


Fig. 7. (a) Plot of the critical stress for forward-transformation with dislocation-emission, σ_F^S at a range of emission-rates $\widehat{\varepsilon}_D$, establishing that σ_F^S is higher than the forward-transformation stress without emission, σ_F (b) Plot of the irreversible pathway of the superelastic stress–strain curve highlighting the critical Gibbs' free energy conditions achieved during forward and reverse-transformation. The critical condition for forward-transformation without emission is achieved at A, at a lower-stress than for A_S where emission occurs; for the reverse-transformation, the critical condition for transformation-induced dislocations is achieved earlier at C_S as compared to the critical-condition without emission at C, consequently preferring dislocation-emission during the reverse-transformation; the irreversible transformation pathway $OABC_SD_SO_S$ is therefore selected by thermodynamics and consequently observed in reality.

state of stress, for which the driving-force from work-interaction overcomes the resisting driving-forces from elastic self-energy of the variant, chemical energy, and lattice-friction (refer section 2.2). In the reverse-transformation, it is hypothesized that the martensite microstructure could evolve to improve strain-compatibility with the austenite-phase in the presence of emitted strain. Such an evolution would serve to bring down the elastic self-energy of the inclusion, in effect providing an additional microstructural driving force (supplementing F^S_{SELF}) to favor emission of transformation-induced dislocations during the reverse-transformation. The proposed framework could therefore be extended to incorporate the habit-plane normal and transformation-shear as state-dependent characteristics that are solved for in thermodynamically deriving the transformation pathway.

In closing, it is worthwhile to consider how the above thermodynamic understanding of transformation-induced dislocations provides a route to achieve fully reversible phase-transformation eliminating dislocation-emission and ultimately functional fatigue. Firstly, the framework provides a quantifiable stress-gap between the two critical points in the reverse transformation, σ^S_R (at C_S) and σ_R (at C, also refer Fig. 7(b)). If the gap $(\sigma^S_R - \sigma_R)$ can be reduced to zero or made negative, the reversible pathway gains thermodynamic preference over the irreversible pathway, ultimately achieving reverse-transformation without dislocation-emission. Consider the driving-forces involved and relevant material properties that would dictate this stress-gap:

a. Increased barrier associated with dislocation-emission, $\boldsymbol{F^S_c(\varepsilon_D)}$: If the jump in the resisting driving force coming from transformation-induced-dislocations (refer Fig. 6(f)) is reduced, the reverse-transformation stress σ^S_R lowers and diminishing the gap to σ_R. Physically, this implies enhancing the slip-resistance of SMAs.
b. Reduced driving-force from elastic-relaxation, $\boldsymbol{F^S_{SELF}}$: If the dislocation-emission does not relax the strain-energy significantly, the associated jump in the driving force (Fig. 6(d)) is reduced, undermining the positive driving-force available to cause dislocation-emission. Physically, this points to reducing lattice-mismatch between the austenitic and martensitic phases such that the need for dislocation-induced strain-relaxation is either minimized or eliminated.

Note that the above two approaches have been the traditional options pursued to enhance reversibility of phase-transformation in SMAs. However, to the best of the author's knowledge, their coupled effect has not been considered till date. In other words, the goal is to not pursue each of the above approaches independently but to rather consider their coupled influence. To that end, the proposed framework offers to shift the paradigm in approaching reversibility:

a. Coupled influence of $\boldsymbol{F_{SELF}}$ and $\boldsymbol{F^S_c}$: The approach should be to begin with the equilibrium of driving-forces at the point of reverse-transformation coming from Gibbs' free energy considerations (Eq. (44)). If the increased barrier F^S_c coming from the SMA's slip-resistance is high enough to undermine the driving-force arising from the relaxing strain-energy F_{SELF}, the propensity for dislocation-emission can be reduced. Note that this approach does not focus on the absolute contributions from each driving-force component but rather their relative influence.
b. Influence of work-interaction $\boldsymbol{F^S_W}$: Additionally, the driving-force-equilibriums involves a contribution from the work-term F^S_W, that has not been considered before and introduced by this study. This driving-force comes from the work-interaction energy between the external applied load, the relaxing transformation-strain and the emitted strain (refer Eqs. (36) and (43)). This driving-force supplements the resistance against F_{SELF}, also bringing in a loading-orientation dependence through the involvement of the Schmid factor SF_m. If the incipient drop in F^S_W (refer Fig. 6(b)) caused by dislocation-emission is reduced (again, relative to changes in F^S_{SELF}, F^S_c), the driving-force from the applied work offers increased resistance to dislocation-emission.

Thus, the proposed framework not only offers a thermodynamic explanation for transformation-induced-dislocations but leverages the same thermodynamic basis to (1) quantify the propensity for dislocation-emission, and (2) propose directions to engineer targeted material properties for improved transformation-reversibility and ultimately better resistance to functional fatigue.

Conclusions

This study offers a thermodynamic framework for transformation-induced-dislocations during superelastic phase-transformaton of SMAs. The physical origin of emitted strain is proposed, arising from the formation of loops around the transforming inclusion and from incremental slipping around the inclusion. An emission-rate is defined to couple the emitted strain to the progression of the phase-transformation, subsequently deriving the superelastic stress-strain curves without (reversible) and with (irreversible) transformation-induced dislocations. Both reversible and irreversible transformation-pathways satisfy the energy-balance of first law of thermodynamics. It is shown that it is thermodynamically favorable for dislocation-emission to occur during reverse-transformation as the critical spontaneity condition, coming from the second law of thermodynamics, is achieved earlier than in the scenario of reversible transformation without dislocation-emission. The driving-force for dislocation-emission is shown to come from the propensity of the elastic strain-energy of the inclusion to relax, overcoming the lattice-friction barrier and the driving-force from work-interaction. The absence of such a mechanism during the forward-transformation is shown, justifying the selective thermodynamic preference for dislocation-emission to occur only in the reverse-transformation. The proposed framework quantifies the propensity for dislocation-emission through a stress-gap between points on the irreversible pathway and reversible pathway, at which the respective spontaneity condition for reverse-transformation are satisfied. Leveraging the quantified stress-gap and the derived driving-force equilibrium in the irreversible pathway, new approaches to target SMA properties for improved functional-fatigue resistance are proposed.

CRediT authorship contribution statement

Ahmed Sameer Khan Mohammed: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Huseyin Sehitoglu:** Funding acquisition, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work is supported by the NSF DMR-2104971 MMN program, which is gratefully acknowledged.

Appendix A: Experimental evidence of transformation-induced dislocations emanating from the martensite

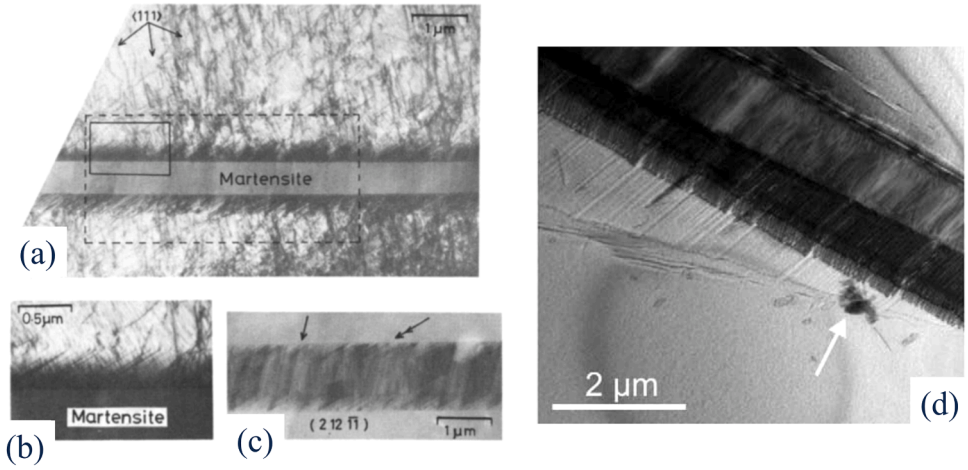


Fig. A1. TEM evidence from literature illustrating transformation-induced dislocations emanating from the martensite front during the reverse-transformation: Figures (a–c) are taken from ref. [37] showing oriented traces of emitted dislocations in (a and b), that are aligned with (c) traces of transformation twins within the martensite plate, retained after reverse-transformation (the study target was a Cu–Zn alloy); Figure (d) is an in situ TEM image from ref. [25] showing a reverse-transforming martensitic phase (dark color) in NiTi having fine twins and leaving behind traces of emitted dislocations aligned the internal twin boundaries; the reader is referred to the respective references for more details.

References

[1] H. Sehitoglu, G. Anlas, A.S.K. Mohammed, Shape memory alloys – frontier developments, in: M.H.F. Aliabadi, W.O. Soboyejo (Eds.), Comprehensive Structural Integrity, Second Edition, Elsevier, Oxford, 2023, pp. 610–679.

[2] K. Otsuka, C.M. Wayman, Shape Memory Materials, Cambridge University Press, 1999.

[3] E. Patoor, D.C. Lagoudas, P.B. Entchev, L.C. Brinson, X. Gao, Shape memory alloys, Part I: general properties and modeling of single crystals, Mech. Mater. 38 (5) (2006) 391–429.

[4] P. Müllner, V.A. Chernenko, M. Wollgarten, G. Kostorz, Large cyclic deformation of a Ni-Mn-Ga shape memory alloy induced by magnetic fields, J. Appl. Phys. 92 (11) (2002) 6708–6713.

[5] T.W. Duerig, The use of superelasticity in modern medicine, MRS Bull. 27 (2) (2002) 101–104.

[6] T. Duerig, A. Pelton, D. Stöckel, An overview of nitinol medical applications, Mater. Sci. Eng.: A 273-275 (1999) 149–160.

[7] C. Mavroidis, Development of advanced actuators using shape memory alloys and electrorheological fluids, Res. Nondestruct. Eval. 14 (1) (2002) 1–32.

[8] M. Gueltig, H. Ossmer, M. Ohtsuka, H. Miki, K. Tsuchiya, T. Takagi, M. Kohl, High frequency thermal energy harvesting using magnetic shape memory films, Adv. Energy Mater. 4 (17) (2014) 1400751.

[9] A. Villoslada, A. Flores, D. Copaci, D. Blanco, L. Moreno, High-displacement flexible shape memory alloy actuator for soft wearable robots, Rob. Auton. Syst. 73 (2015) 91–101.

[10] G. Eggeler, E. Hornbogen, A. Yawny, A. Heckmann, M. Wagner, Structural and functional fatigue of NiTi shape memory alloys, Mater. Sci. Eng.: A 378 (1) (2004) 24–33.

[11] W. Abuzaid, H. Sehitoglu, Superelasticity and functional fatigue of single crystalline FeNiCoAlTi iron-based shape memory alloy, Mater. Des. 160 (2018) 642–651.

[12] R. Sidharth, Y. Wu, F. Brenne, W. Abuzaid, H. Sehitoglu, Relationship between functional fatigue and structural fatigue of iron-based shape memory alloy FeMnNiAl, Shape Mem. Superelast. 6 (2) (2020) 256–272.

[13] P. Šittner, O. Molnárová, L. Kadeřávek, O. Tyc, L. Heller, Deformation twinning in martensite affecting functional behavior of NiTi shape memory alloys, Materialia 9 (2020) 100506.

[14] C. Damiani, M. Sade, F.C. Lovey, Fatigue in Cu-Zn-Al single crystals during pseudoelastic cycling: in situ observations by SEM and optical microscopy, J. Phys. IV France 112 (2003) 623–626.

[15] K.N. Melton, O. Mercier, Fatigue of NITI thermoelastic martensites, Acta Metallurgica 27 (1) (1979) 137–144.

[16] M. Sade, R. Rapacioli, M. Ahlers, Fatigue in Cu·Zn·Al single crystals, Acta Metallur. 33 (3) (1985) 487–497.

[17] H. Tobushi, T. Hachisuka, T. Hashimoto, S. Yamada, Cyclic deformation and fatigue of a TiNi shape-memory alloy wire subjected to rotating bending, J. Eng. Mater. Technol. 120 (1) (1998) 64–70.

[18] D.C. Lagoudas, D.A. Miller, L. Rong, P.K. Kumar, Thermomechanical fatigue of shape memory alloys, Smart Mater. Struct. 18 (8) (2009) 085021.

[19] C. Maletta, E. Sgambitterra, F. Furgiuele, R. Casati, A. Tuissi, Fatigue of pseudoelastic NiTi within the stress-induced transformation regime: a modified Coffin–Manson approach, Smart Mater. Struct. 21 (11) (2012) 112001.

[20] Y. Gao, L. Casalena, M.L. Bowers, R.D. Noebe, M.J. Mills, Y. Wang, An origin of functional fatigue of shape memory alloys, Acta Mater. 126 (2017) 389–400.

[21] J. Frenzel, On the importance of structural and functional fatigue in shape memory technology, Shape Mem. Superelast. 6 (2) (2020) 213–222.

[22] S. Kajiwara, W.S. Owen, Substructure of austenite formed by a partial reverse martensitic transformation in an Fe-Pt alloy, Metall. Trans. 4 (8) (1973) 1988–1990.

[23] X. Jiang, M. Hida, Y. Takemoto, A. Sakakibara, H. Yasuda, H. Mori, In situ observation of stress-induced martensitic transformation and plastic deformation in TiNi alloy, Mater. Sci. Eng.: A 238 (2) (1997) 303–308.

[24] D.M. Norfleet, P.M. Sarosi, S. Manchiraju, M.F.X. Wagner, M.D. Uchic, P. M. Anderson, M.J. Mills, Transformation-induced plasticity during pseudoelastic deformation in Ni–Ti microcrystals, Acta Mater. 57 (12) (2009) 3549–3561.

[25] T. Simon, A. Kröger, C. Somsen, A. Dlouhy, G. Eggeler, On the multiplication of dislocations during martensitic transformations in NiTi shape memory alloys, Acta Mater. 58 (5) (2010) 1850–1860.

[26] R. Delville, B. Malard, J. Pilch, P. Sittner, D. Schryvers, Transmission electron microscopy investigation of dislocation slip during superelastic cycling of Ni–Ti wires, Int. J. Plast. 27 (2) (2011) 282–297.

[27] A.R. Pelton, G.H. Huang, P. Moine, R. Sinclair, Effects of thermal cycling on microstructure and properties in Nitinol, Mater. Sci. Eng.: A 532 (2012) 130–138.

[28] J. Zhang, C. Somsen, T. Simon, X. Ding, S. Hou, S. Ren, X. Ren, G. Eggeler, K. Otsuka, J. Sun, Leaf-like dislocation substructures and the decrease of martensitic start temperatures: a new explanation for functional fatigue during thermally induced martensitic transformations in coarse-grained Ni-rich Ti–Ni shape memory alloys, Acta Mater. 60 (5) (2012) 1999–2006.

[29] P. Chowdhury, H. Sehitoglu, A revisit to atomistic rationale for slip in shape memory alloys, Prog. Mater. Sci. 85 (2017) 1–42.

[30] S. Nemat-Nasser, J.-Y. Choi, W.-G. Guo, J.B. Isaacs, Very high strain-rate response of a NiTi shape-memory alloy, Mech. Mater. 37 (2) (2005) 287–298.

[31] H. Sehitoglu, I. Karaman, R. Anderson, X. Zhang, K. Gall, H.J. Maier, Y. Chumlyakov, Compressive response of NiTi single crystals, Acta Mater. 48 (13) (2000) 3311–3326.

[32] H.M. Paranjape, M.L. Bowers, M.J. Mills, P.M. Anderson, Mechanisms for phase transformation induced slip in shape memory alloy micro-crystals, Acta Mater. 132 (2017) 444–454.

[33] H. Gu, L. Bumke, C. Chluba, E. Quandt, R.D. James, Phase engineering and supercompatibility of shape memory alloys, Mater. Today 21 (3) (2018) 265–277.

[34] R. Zarnetta, R. Takahashi, M.L. Young, A. Savan, Y. Furuya, S. Thienhaus, B. Maaß, M. Rahim, J. Frenzel, H. Brunken, Y.S. Chu, V. Srivastava, R.D. James, I. Takeuchi, G. Eggeler, A. Ludwig, Identification of quaternary shape memory alloys with near-zero thermal hysteresis and unprecedented functional stability, Adv. Funct. Mater. 20 (12) (2010) 1917–1923.

[35] V.M. Dornelas, S.A. Oliveira, M.A. Savi, A macroscopic description of shape memory alloy functional fatigue, Int. J. Mech. Sci. 170 (2020) 105345.

[36] P.B. Entchev, D.C. Lagoudas, Modeling of Transformation-Induced Plasticity in SMAs, Shape Memory Alloys: Modeling and Engineering Applications, Springer US, Boston, MA, 2008, pp. 233–277.

[37] S. Kajiwara, T. Kikuchi, Dislocation structures produced by reverse martensitic transformation in a Cu-Zn alloy, Acta Metall. 30 (2) (1982) 589–598.
[38] H. Akamine, A. Heima, Y. Soejima, M. Mitsuhara, T. Inamura, M. Nishida, Where and when are dislocations induced by thermal cycling in Ti–Ni shape memory alloys? Acta Mater. 244 (2023) 118588.
[39] A.S.K. Mohammed, H. Sehitoglu, Martensitic twin boundary migration as a source of irreversible slip in shape memory alloys, Acta Mater. 186 (2020) 50–67.
[40] A.S.K. Mohammed, H. Sehitoglu, Modeling the interface structure of type II twin boundary in B19′ NiTi from an atomistic and topological standpoint, Acta Mater. 183 (2020) 93–109.
[41] R.C. Pond, X. Ma, Y.W. Chai, J.P. Hirth, Chapter 74 topological modelling of martensitic transformations, in: F.R.N. Nabarro, J.P. Hirth (Eds.), Dislocations in Solids, Elsevier 2007, pp. 225–261.
[42] R. Sidharth, A.S.K. Mohammed, H. Sehitoglu, Functional fatigue of NiTi shape memory alloy: effect of loading frequency and source of residual strains, Shape Memory Superelast. 8 (4) (2022) 394–412.
[43] Y. Zhang, Z. Moumni, Y. You, W. Zhang, J. Zhu, G. Anlas, Multiscale TRIP-based investigation of low-cycle fatigue of polycrystalline NiTi shape memory alloys, Int. J. Plast. 115 (2019) 307–329.
[44] P. Hua, K. Chu, F. Ren, Q. Sun, Cyclic phase transformation behavior of nanocrystalline NiTi at microscale, Acta Mater. 185 (2020) 507–517.
[45] T. Mura, Isotropic inclusions, in: T. Mura (Ed.), Micromechanics of Defects in Solids, Springer Netherlands, Dordrecht, 1987, pp. 74–128.
[46] J.M. Ball, R.D. James, Fine phase mixtures as minimizers of energy, Arch. Ration. Mech. Anal. 100 (1) (1987) 13–52.
[47] X. Zhang, H. Sehitoglu, Crystallography of the B2 → R → B19′ phase transformations in NiTi, Mater. Sci. Eng.: A 374 (1) (2004) 292–302.
[48] J.S. Bowles, J.K. Mackenzie, The crystallography of martensite transformations I, Acta Metall. 2 (1) (1954) 129–137.
[49] M.S. Wechsler, M.S. Wechsler, D.S. Lieberman, T.A. Read, Trans. AIME 197 (1953) 1503.
[50] K. Gall, H. Sehitoglu, The role of texture in tension–compression asymmetry in polycrystalline NiTi, Int. J. Plast. 15 (1) (1999) 69–92.
[51] M. Huang, L.C. Brinson, A Multivariant model for single crystal shape memory alloy behavior, J. Mech. Phys. Solids 46 (8) (1998) 1379–1409.
[52] R.F. Hamilton, H. Sehitoglu, Y. Chumlyakov, H. Maier, Stress dependence of the hysteresis in single crystal NiTi alloys, Acta Mater. 52 (11) (2004) 3383–3402.
[53] A.N. Bucsek, D. Dale, J.Y.P. Ko, Y. Chumlyakov, A.P. Stebner, Measuring stress-induced martensite microstructures using far-field high-energy diffraction microscopy, Acta Crystallogr. Sect. A 74 (5) (2018) 425–446.
[54] C. Cayron, What EBSD and TKD tell us about the crystallography of the martensitic B2-B19′ transformation in NiTi Shape memory alloys, Crystals (2020).
[55] L. Heller, P. Šittner, On the habit planes between elastically distorted austenite and martensite in NiTi, Acta Mater. 269 (2024) 119828.
[56] L. Heller, H. Seiner, P. Šittner, P. Sedlák, O. Tyc, L. Kadeřávek, On the plastic deformation accompanying cyclic martensitic transformation in thermomechanically loaded NiTi, Int. J. Plast. 111 (2018) 53–71.