Nanoscale phase transition behavior of shape memory alloys – closed form solution of 1D effective modelling

M.P. Li\textsuperscript{a,b}, Q.P. Sun\textsuperscript{b,*}

\textsuperscript{a}Department of Engineering Mechanics, Wuhan University, Wuhan, China
\textsuperscript{b}Department of Mechanical and Aerospace Engineering, The Hong Kong University of Science and Technology, Hong Kong, China

**A R T I C L E   I N F O**

Article history:
Received 1 July 2017
Revised 15 September 2017
Accepted 15 September 2017
Available online 20 September 2017

Keywords:
Grain size effects
Nanoscale phase transition
Internal length scales
Energy dissipation
Two-phase coexistence

**A B S T R A C T**

We investigate the roles of grain size ($l_g$) and grain boundary thickness ($l_b$) on the stress-induced phase transition (PT) behaviors of nanocrystalline shape memory alloys (SMAs) by using a Core-shell type “crystallite-amorphous composite” model. A non-dimensionalized length scale $l'_b = l'_g/l_b$ is identified as the governing parameter which is indicative of the energy competition between the crystallite and the grain boundary. Closed form analytical solutions of a reduced effective 1D model with embedded microstructure length scales of $l'_g$ and $l_b$ are presented in this paper. It is shown that, with $l'_b$ reduction, the energy of the elastic non-transformable grain boundary will gradually become dominant in the phase transition process, and eventually bring fundamental changes of the deformation behaviors: breakdown of two-phase coexistence and vanishing of superelastic hysteresis. The predictions are supported by experimental data of nanocrystalline NiTi SMAs.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Shape memory alloys (SMAs), as a typical ferroelastic material undergoing first-order martensitic phase transition (PT), are of great interest in research of physics, materials sciences and engineering due to their wide applications (Otsuka et al., 1979; Miyazaki et al., 1981; Ahluwalia et al., 2004; Saxena and Aeppli, 2009; He and Sun, 2009a,b; Salje, 2012; Sun et al., 2012). With the fast development of nanomaterials and nanotechnology, new phenomena and properties of nanocrystalline (nc) SMAs emerge, e.g., the large elastic reversible deformation (5–6%) with controllable hysteresis and the Invar and Elinvar effects over wide temperature window have been observed in nc SMAs (Waitz et al., 2004, 2009; Kim et al., 2006; Ye et al., 2010; Ahadi and Sun, 2013, 2015; Yin et al., 2016; Xia and Sun, 2017a,b; Ahadi et al., 2017). Compared with the coarse-grained counterparts, one of the key features of nanostructured SMAs is the large volume fraction of grain boundaries (GBs) and amorphous phase (AP), both of which are atomic disordered regions away from thermodynamic equilibrium (see Fig. 1(a) and (b)) (Gleiter, 1988; Meyers et al., 2006; Sun et al., 2014). The GBs and AP have profoundly different structure and property from the transformable crystallites and will significantly influence the PT behaviors of nc materials (Ueland and Schuh, 2013; Ko et al., 2017). To understand and get insight into these emergent phenomena by modelling and simulation, it is necessary to treat the material as a structure hierarchy with multiple relevant length scales.

Nucleation and growth of micro-domain inside discrete grains are the most widely used paradigms in theoretical modelling of the stress-induced PT process in coarse-grained SMA polycrystals. The nucleation events involve intrinsic instability of the crystallite, cause dissipation of elastic energy and lead to the typical large stress hysteresis in the loading-unloading...
Fig. 1. (a) GS and GB thickness in microstructure hierarchy of nanostructured SMAs and (b) the Core-shell type “crystallite-amorphous composite” model characterized by two internal length scales $l_\text{G}$ (GS) and $l_\text{B}$ (GB thickness).

superelastic cycles (Truskinovsky and Zanzotto, 1996; Vainchtein et al., 1998; Puglisi and Truskinovsky, 2000; Brinson et al., 2004; Sun and He, 2008). However, recent experiments of nc NiTi show that the superelastic hysteresis of the polycrystal decreases rapidly with the grain size (GS) reduction and tends to vanish as GS approaches ~10 nm (Ahadi and Sun, 2013). Furthermore, the in-suit XRD experiment of nc NiTi demonstrates that the evolution of XRD profile gradually changes from the typical two-peak mode (i.e., B2 and B19’ phases coexist) to a continuous and reversible single-peak shift mode (i.e., continuous non-uniform distortion of crystallite and lack of two-phase coexistence) when GS is reduced from 1500 nm to 18 nm (Ahadi and Sun, 2015). It is conjectured that the significantly increased volume fraction of GBs has strong effects (Sun et al., 2014; Ko et al., 2017) on the energetic landscape of the system and brings the above fundamental changes in the PT behaviors of the nc materials.

Many models have been proposed to describe the deformation configurations and the hysteresis in solid-to-solid PT within the framework of nonlinear elasticity theory. Fundamental understanding of the relationship between the metastable equilibrium configurations and the hysteresis has been achieved from both the discrete and continuous aspects (Abeyaratne and Knowles, 1990; Ball et al., 1995; Abeyaratne et al., 1996; Bhattacharya, 1999; Truskinovsky and Zanzotto, 1996; Vainchtein et al., 1998; James, 2000; Puglisi and Truskinovsky, 2000). It is now generally accepted that, the defects (such as grain boundaries, dislocations and second-phase particles) and the imposed internal length scales from the material heterogeneity make the system trapped in metastable configurations and lead to the hysteresis dissipation during the switching between metastable states. Besides, a number of continuum modelling works have addressed size-effects on PT by either using a material model having a length scale or directly imposing length scales into the modelling (Cai and Dai, 2006; Waitz et al., 2007; Salje, 2008; Sun and He, 2008; Porta et al., 2009; Petryk et al., 2010; Stupkiewicz and Petryk, 2010; Qiao et al., 2011; Ueland and Schuh, 2013; Qiao and Radovitzky, 2016; Tan and Bhattacharya 2016). These works captured some features of the size-dependent stress hysteresis and strain hardening in polycrystalline SMAs.

For the PT in nc SMAs, one of the key issues in the modelling is how to quantify the interfacial dominance in the responses as the GS is reduced down to nanoscale. There are two types of interfaces: grain boundary (GB) and phase boundary (PB). As GS is reduced down to the nanoscale, the volume fraction of atoms in both GBs and PBs increases significantly. The GBs and PBs, both treated as sharp interfaces in coarse-grained polycrystals, become bulky in nc SMAs. For example, the GBs occupy approximately 27% of the total atoms when GS = 10 nm (assuming GB layer thickness to be 1 nm which is a generally accepted value). To quantify the structure and energy of the PB, a gradient energy term of the order parameter (it is strain for SMAs) is added to the nonconvex free energy density in the spirit of the Ginzburg–Landau theory (Falk, 1982, 1983; Saxena et al., 1997; Sun and He, 2008). The gradual increase in the energy of PB relative to the crystallite was used to interpret the observed GS effects (Sun and He, 2008; Ahadi and Sun, 2013). Recent meso-scale (Ahlulwala et al., 2015) and large-scale MD (Ko et al., 2017) simulations show that there is strong competition between PBs and GBs in nanoscale PT and that the strong mechanical constraint of the 3D network of GBs indeed plays a critical role in suppressing the PT
and contributes to an increase in the free energy of the system. To quantify such strong GB constraint at nanoscale, different versions of core-shell type crystallite-amorphous composite models are proposed to explain various size-dependent properties (e.g., Kim, 1998; Meyers et al., 2006; Hahn and Meyers, 2015). The basic idea of the core-shell model, as illustrated in Fig. 1, is that the inhomogeneous system consists of two parts: (1) the transformable “core” (crystallite) and (2) the non-transformable “shell” (grain boundary). Although including both GB and PB into the theoretical modelling makes the analytical solution much more challenging, it is a necessary part of the quantitative understanding.

From modelling point of view, the nc SMAs can be described as a core-shell type “crystallite-amorphous composite”. However, to get analytical solution of the response so that explicit clear insight of the length scale effect phenomena can be obtained, further simplification of the model is necessary. For example, we may focus on the response of a representative 3D grain and such 3D grain of the “composite” needs to be further reduced into an effective 1D grain (see Fig. 2(a) and (b)). Such 3D→1D effective modelling is expected not only to contain the key length scales of the transformable crystallite and the non-transformable GB, but also to retain the key features of the 3D polycrystal such as the volume fraction of GB and its mechanical constraints to the crystallite. From mathematical solution point of view, strong care should be taken in handling the minimization of such inhomogeneous system, the solution scheme of which may quite different from that of the homogeneous system with nonconvex and nonlocal Ginzburg–Landau type functional (Falk, 1983; Carr et al., 1984; Vainchtein et al., 1998; Cai and Dai, 2006). It is believed that such effective 1D model, although much simplified, still has the advantage of providing scaling of the GS-related PT behaviors but at the same time with clear physical insight and understanding of the roles of each internal length scales.

This paper aims to quantify and to understand the effects of material internal length scales of grain size (GS), grain boundary (GB) and phase boundary (PB) on the stress-induced PT behaviors of nc SMAs through closed form analytical solutions of an effective 1D Core-shell type model. In Section 2, 3D→1D effective modelling procedure is described and the solution schemes are provided. We give the theoretical predictions of the GS effects on the deformation mode, superelastic hysteresis and energy partition in Section 3 and compare them with the available experimental data. The conclusions are given in Section 4.
2. Modeling scheme

To quantify the GS dependency of the stress-induced PT responses at the level of a single grain, we model the grain as a “Core-Shell” type mixture of crystallite and amorphous phases (Figs. 1 and 2). We first convert the representative 3D cubic grain into a 1D effective grain with embedded length scales of \( l_g \) and \( l_b \) in Section 2.1 and give the energy density expressions of the 1D effective grain. Then the governing equations for the deformation of the 1D effective grain are derived based on the energy minimization principle in Section 2.2 with the solution schemes outlined. The solutions for nano-grain PT are given in Section 2.3.

2.1. From 3D cubic grain to effective 1D grain – free energy function

Without losing the generality and for the purpose of simplicity, the nc material is idealized as a collection of uniform cubic grains. We consider a representative cubic grain under uniform applied stress \( \sigma \) (see Fig. 2(b)). The volume fraction of GB (or AP) \( \xi \) for a cubic grain is \( \xi = V_b/V_0 = 1 - (1 - 1/l_b)^3 \), where \( l_g = l_g/l_b \geq 1 \). \( l_g \) is the GS and \( l_b \) is the GB thickness is a non-dimensionalized length scale and \( V_0 = l_b^3 \) is the volume of the grain.

By assuming uniform distribution of axial strain \( \varepsilon(x, y, z) \) in the square cross section perpendicular to the loading axis \( x \) (i.e., \( \varepsilon(x, y, z) = \varepsilon(x) \)), the axial deformation \( \varepsilon(x) \) of the representative cubic grain (3D) under an external axial stress \( \sigma \) can be reduced to that of an effective 1D grain or a 1D inhomogeneous structure. The deformation of the 1D grain will be described in the framework of nonconvex and nonlocal elasticity theory by introducing the strain-gradient energy term \( \alpha \varepsilon^2 / 2 \) (\( \alpha \) is the gradient coefficient) to quantify the structure and the energy of phase boundary (PB) of the crystallite (see Figs. 1 and 2(b)). From theory we know that the PB is a smooth transition layer of thickness proportional to \( \sqrt{\alpha E} \) (\( E \) is the elastic modulus of austenite phase). From experiment the characteristic thickness of PB in NiTi SMAs is of the order of 1 nm. Different from some macroscopic 3D to 1D nonlocal elasticity model (Coleman and Newman, 1988; Dai and Cai, 2006) where the length scale \( \alpha \) is associated with the characteristic length scale of the macroscopic system (such as the diameter of a long fiber) under consideration, here the PB in a nano-grain contains only several layers of atoms due to the misfit between martensite (B19\(^{\prime} \)) lattice and the austenite (B2) lattice (Falk, 1983). Also, the thickness of PB is often of the same order of magnitude as the thickness of the GB. Furthermore, there is no twinning of B19\(^{\prime} \) lattice observed in such extremely small GS under externally applied stress (also see Ko et al., 2017).

This 1D grain is effective in the sense that it retains the key features of the 3D cubic grain such as the volume fraction of GB and its mechanical constraints to the crystallite along both the loading and transverse directions. The advantage of the above effective 1D model is that closed form analytical solutions can be obtained so that clear quantitative understanding of the roles of the length scales of \( l_g \) and \( l_b \) as well as the properties of the crystallite and GB in the PT responses can be obtained.

The free energy density function of the 1D element is

\[
\psi(\varepsilon, \varepsilon_b) = \psi(b) + \alpha \varepsilon_b^2 / 2
\]

(2.1)

where \( \psi(b) \) is the strain energy density function, and the term \( \alpha \varepsilon_b^2 / 2 \) is to account the elastic strain gradient energy and \( \alpha(>0) \) is the gradient coefficient. Since the PB in the nonlocal model is a smooth transition layer of length scale proportional to \( \sqrt{\alpha E} \) (\( E \) is the elastic modulus), we denote \( \sqrt{\alpha E} = l_b \) or \( \alpha = \sqrt{E} / l_b \). The length scale of PB can be nondimensionalized by \( l_b \) as \( \sqrt{\alpha / (E l_b^2)} \) (\( = l_b / l_b \)). (Remark: the effect of \( \sqrt{E} / l_b \) on PB energy and energy partition will be discussed in Section 3. We find that, the energy partition and energy dissipation are insensitive to the value of \( \alpha / E l_b^2 \) (see Fig. 9(b) where \( \alpha / E l_b^2 = 0.1, 1, 10 \) and Eq. (3.5)) in nanoscale PT. For convenience, \( \alpha / E l_b^2 \) is often replaced with a nondimensionalized quantities \( \beta (= 8\alpha / (E l_b^2)) \) in the text.)

Due to the inhomogeneity of the 1D structure, \( \psi(\varepsilon) \) of (2.1) is expressed as

\[
\psi(\varepsilon) = \begin{cases} 
\psi_0(\varepsilon) = E\varepsilon^2 / 2, & \text{for grain boundary phase} \\
\psi_c(\varepsilon) = \eta \psi_0(\varepsilon) + (1 - \eta) \psi_c(\varepsilon), & \text{for effective crystallite} 
\end{cases}
\]

(2.2)

where \( \psi_0(\varepsilon) \) is the strain energy density function for the non-transformable elastic GB phase. The deformation of GB phase is reasonably treated as linear elastic, since the GB is an atomic disordered region (see Fig. 1(a) and (b)) and can exhibit large linear elastic deformation as that of metallic glass. \( \psi_c(\varepsilon) \) is the strain energy density function of the effective crystallite, \( \eta = A_0 / A_0 = 1 - (1 - 1/l_b)^2 \) is the area fraction of GB in the cross section \( A_0 = l_b^2 \) of the grain (see Fig. 2). \( \psi_c(\varepsilon) \) is the non-convex strain energy density function of the transformable crystallite (see Figs. 1(b) and 2(c))

\[
\psi_c(\varepsilon) = \begin{cases} 
E\varepsilon^2 / 2 (\varepsilon < \varepsilon_s) \\
-\frac{E l_g E t_s}{2} \left( \varepsilon - \sigma_M / E - \frac{\varepsilon_t}{2} \right)^2 + \frac{E t_s^2}{8(\varepsilon_s + E)} - \frac{\sigma_M^2}{2E} \left( \varepsilon_s \leq \varepsilon \leq \varepsilon_c \right) \\
E(\varepsilon - \varepsilon_t)^2 / 2 + \varepsilon_t \sigma_M, \quad (\varepsilon \geq \varepsilon_c)
\end{cases}
\]

(2.3)

where \( \sigma_M \) is the temperature dependent Maxwell stress of the crystallite, \( \varepsilon_t \) is the transformation strain, \( \varepsilon_s = \sigma_N / E \) and \( \varepsilon_c = \varepsilon_t + (\sigma_M - \sigma_N) / E \). \( \sigma_N = E t_s E t_s / [2(E_s + E)] + \sigma_M \) is the nucleation stress for forward PT and the modulus of unstable interm
phase in non-convex region is negative, denoted as \(-E_x\). For the effective crystallite, the modulus of interim phase of the non-convex region is \(E_g - \eta E - (1 - \eta)E_x\). When we set the Maxwell stress \(\sigma_M = 0\), \(\psi_\varepsilon(x)\) has two-energy wells representing Austenite (A) phase \((\varepsilon = 0)\) and Martensite (M) phase \((\varepsilon = \varepsilon_M)\), respectively (Fig. 2(c)). Then we have constructed the model of the effective 1D grain characterized by the fundamental properties of crystallite and GB and the length scales of \(l_g\) and \(l_b\). Fig. 2(d) shows the change of the \(\psi_\varepsilon(x)\) and \(\sigma(\varepsilon)\) of the effective crystallite with \(\bar{l}_g\).

2.2. Energy minimization and solution schemes

Under a given external stress \(\bar{\sigma}\), the total free energy \(G\) of the effective 1D grain is a functional of its strain configuration \(\varepsilon(x)\)

\[
G = A_0 \int_0^{l_g} \varphi(\varepsilon(x), \varepsilon_x(x)) dx
\]

(2.4)

where \(\varphi(\varepsilon, \varepsilon_x) = f(\varepsilon, \varepsilon_x) - \bar{\sigma} \varepsilon\) is the Gibbs free energy density.

The periodic boundary condition applies at the two ends of 1D grain \(\varepsilon_x(0) = \varepsilon_x(l_g) = 0\) since it is extracted from a series of identical grains (see Fig. 2(a)). Furthermore, we assume a symmetric deformation in the effective 1D system thus the boundary conditions can be further written as \(\varepsilon_x(0) = \varepsilon_x(l_g/2) = 0\). This assumption is reasonable for a representative nano-grain. On one hand, only the symmetric deformation can minimize the total energy of the 1D symmetric system (see detail in Li, 2017). On the other hand, recent large-scale MD simulations of nc SMAs (Ko et al., 2017) clearly shows that the martensite phase nucleates at the grain interior and grows towards the grain boundaries during the loading process, which is consistent with and supports our symmetric deformation assumption.

The equilibrium configuration \(\varepsilon(x)\) is determined by the first variation of the total Gibbs free energy functional

\[
\delta G(\varepsilon(x)) = 0
\]

(2.5)

which gives the Euler–Lagrange equation for \(\varepsilon(x)\) as

\[
\frac{d}{dx} \frac{\partial \varphi}{\partial \varepsilon_x} - \frac{\partial \varphi}{\partial \varepsilon} = 0
\]

(2.6)

or

\[
\frac{d}{dx} \left( \varphi - \varepsilon_x \frac{\partial \varphi}{\partial \varepsilon_x} \right) = 0
\]

(2.7)

We denote

\[
\phi = \varphi - \varepsilon_x \frac{\partial \varphi}{\partial \varepsilon_x} = \psi(\varepsilon) - \bar{\sigma} \varepsilon - \frac{1}{2} \alpha \varepsilon^2
\]

(2.8)

Then \(\phi\), the first integral constant, is actually the Hamiltonian of the system.

For our 1D inhomogeneous system, the Hamiltonians (or the integral constants) of the effective crystallite and the grain boundary are different, i.e.,

\[
\phi = \begin{cases} 
\phi_a(\varepsilon) = \psi_a(\varepsilon) - \bar{\sigma} \varepsilon - \alpha \varepsilon^2/2, & \varepsilon \in [0, l_b/2] \cup [l_g - l_b/2, l_g] \\
\phi_c(\varepsilon) = \psi_c(\varepsilon) - \bar{\sigma} \varepsilon - \alpha \varepsilon^2/2, & \varepsilon \in [l_b/2, l_g - l_b/2]
\end{cases}
\]

(2.9)

For a homogeneous system with Ginzburg–Landau type energy density, the minimization of its functional is a classical topic and has already been well studied in mathematics and physics. Often, the Hamiltonian of the system and the phase-plane analysis are used to get all kinds of solutions such as solitary-wave (soliton), kink/antikink or periodic (periodon) (Falk, 1983; Cai and Dai, 2006). However, for the inhomogeneous system here, the Hamiltonians of the effective crystallite and of the GB are different, so the traditional phase-plane analysis in the literature for the homogeneous systems (Carr et al., 1984; Cai and Dai, 2006) may become ineffective and has difficulty to apply directly to get the solutions of the system.

To handle this problem, we adopt an inverse method, i.e., to use the piecewise hyperbolic functions as the solution that can exactly satisfy the Euler–Lagrange equation and the boundary conditions. It is noticed that, the Euler–Lagrange equation (2.6) is a second-order ODE

\[
\alpha \varepsilon_{xx} - \sigma(\varepsilon) + \bar{\sigma} = 0
\]

(2.10)

where \(\sigma(\varepsilon) = \partial \psi / \partial \varepsilon\). Since the stress-strain relation is linear for GB phase and is tri-linear for effective crystallite (see Fig. 2(c) and (d)), the solution \(\varepsilon(x)\) satisfying Euler–Lagrange equation and boundary conditions is a piecewise hyperbolic as follows

\[
\varepsilon(x) = \begin{cases} 
\varepsilon_1(x) = C_1 \cosh \left( \sqrt{\frac{E}{\alpha x}} \right) + \bar{\sigma} / E, & x \in [0, l_b/2] \\
\varepsilon_2(x), & x \in [l_b/2, l_g/2]
\end{cases}
\]

(2.11)
where $\varepsilon_2(x)$ is the equilibrium strain profile in effective crystallite and can take a periodic form (with $n$ domains) or take a single domain ($n = 1$). For example, for the case $\varepsilon_s < \min\{\varepsilon_2(x)\} \leq \max\{\varepsilon_2(x)\} < \varepsilon_e$ ($\varepsilon_s$ and $\varepsilon_e$ are boundaries of spinodal region as shown in Eq. (2.3) and Fig. 2(c)), we have periodic form for $E_g < 0$ and $n = \frac{[\varepsilon_s - \varepsilon_e]}{2\pi} \sqrt{\frac{\varepsilon_2}{\varepsilon_e}}$.

$$
\varepsilon_2(x) = C_2 \cosh\left(\frac{E_0}{\alpha} \frac{2x - l_x}{2}\right) + \frac{\sigma - \sigma_N}{E} + \frac{\sigma_N}{E}.
$$

(2.12)

for the case $\min\{\varepsilon_2(x)\} \leq \max\{\varepsilon_2(x)\} < \varepsilon_s$, we only need to replace $E_g$ in Eq. (2.12) with $E$; for the case $\varepsilon_s < \min\{\varepsilon_2(x)\} \leq \max\{\varepsilon_2(x)\}$, we need to replace $E_g$ in Eq. (2.12) with $E$ and add a constant term $(1 - \eta)\varepsilon_{tr}$ to the Eq. (2.12).

In the above cases, the solution in Eqs. (2.11) and (2.12) only contains two independent parameters $C_1$ and $C_2$. Substituting Eqs. (2.11) and (2.12) into Eq. (2.4), the minimization of the energy functional $G(\varepsilon(x))$ becomes the minimization of the function $G(C_1, C_2)$.

For the other cases, i.e., $\varepsilon_s < \min\{\varepsilon_2(x)\} < \varepsilon_e < \max\{\varepsilon_2(x)\}$, we need to consider the expression of $\varepsilon_2(x)$ can take the form of piecewise hyperbolic with several parameters. In the following, we first give a proof that there is only one independent parameter for a homogeneous system and then move on to the solution of inhomogeneous system.

For a homogeneous system (the strain energy density function is $\psi_1(\varepsilon)$) with boundary condition $\varepsilon_s(0) = \varepsilon_s(l_x) = 0$, according to the works by Cai and Dai (2006), we can denote the minimum and maximum values ($\varepsilon_{min}$, $\varepsilon_{max}$) of the non-trivial equilibrium solution $\varepsilon(x)$ by $g_1(= \varepsilon_{min})$ and $g_2(= \varepsilon_{max})$, respectively. Then, from (2.8), we have

$$
\psi_1(g_1) - \sigma g_1 = \psi_1(g_2) - \sigma g_2 = \phi
$$

(2.13)

The ranges of $g_1$ and $g_2$ are (Cai and Dai, 2006)

$$
\max\{\varepsilon_1, \varepsilon_{min}\} < g_1 < \varepsilon_e < g_2 < \min\{\varepsilon_3, \varepsilon_{max}\}
$$

(2.14)

where $\varepsilon_1 = \varepsilon_2/\gamma + \varepsilon_3$, $\varepsilon_2 = \varepsilon_1/\gamma - \varepsilon_3$, and $\varepsilon_3 = \varepsilon_2/\gamma - \varepsilon_1$ are the three roots of the equation $d\psi_1(\varepsilon)/d\varepsilon = \sigma$, $\varepsilon_{min} = \varepsilon_2 - \sqrt{(\varepsilon_2 - \varepsilon_3)(\varepsilon_2 - \varepsilon_1)}$ and $\varepsilon_{max} = \varepsilon_2 + \sqrt{(\varepsilon_2 - \varepsilon_3)(\varepsilon_2 - \varepsilon_1)}$, which is the point at which the homoclinic orbit crosses the $x$-axis on phase planes (Carr et al., 1984; Cai and Dai, 2006), $\varepsilon_1$, $\varepsilon_2$ and $\varepsilon_3$ are boundaries of spinodal region as shown in Eq. (2.3) and Fig. 2(c). Thus the relation between $g_1$ and $g_2$ can be explicitly expressed as a single valued function $g_1 = f_1(g_2)$ according to Eqs. (2.13) and (2.14).

The total Gibbs free energy $G$ for the homogeneous system (the strain energy density function is $\psi_1(\varepsilon)$) is

$$
G = A_0 \int_0^{l_x} \left\{ \psi_1(\varepsilon(x)) - \sigma \varepsilon(x) + \frac{1}{2} \alpha [\varepsilon_s(x)]^2 \right\} dx
$$

(2.15)

After some rearrangement by using Eqs. (2.8) and (2.13), $G$ can be expressed as

$$
G = A_0 \int_0^{l_x} \left\{ \psi_1(\varepsilon_2) - \sigma g_2 \right\} + \gamma_{interface} \right\} dx
$$

(2.16)

where

$$
\gamma_{interface} = \int_0^{l_x} \alpha [\varepsilon_s(x)]^2 dx
$$

(2.17)

is the specific interfacial free energy (Cahn and Hilliard, 1958) of the effective “phase boundary (PB)”. Applying Eq. (2.8) to the homogeneous system, we have

$$
dx = \sqrt{\frac{\sigma}{2 \rho_0(\varepsilon)}} \frac{d\varepsilon}{\rho_0(\varepsilon)}
$$

(2.18)

where $\rho_0(\varepsilon) = \sqrt{\psi_1(\varepsilon) - \sigma \varepsilon + \sigma g_2 - \sigma g_2}$. We can substitute Eq. (2.18) into Eq. (2.17) and change the variable of integration form $x$ to $\varepsilon$ and express the specific interfacial free energy as

$$
\gamma_{interface} = N \gamma_{\varepsilon}(g_2) = N \sqrt{2 \alpha} \int_{f_1(g_2)}^{f_2} \rho_0(\varepsilon) d\varepsilon
$$

(2.19)

where $N$ is the number of interfaces and $0 \leq N \leq N_{max}$ (Cahn and Hilliard, 1958; Falk, 1983), $N_{max}$ is the maximum number of interfaces and $N_{max} \leq \frac{l_x}{2} \sqrt{\frac{\varepsilon_2}{\sigma}}$ according to the perturbation analysis (Vainchtein et al., 1998) on spinodal region. $\gamma_{\varepsilon}(g_2) = \sqrt{2 \alpha} \int_{f_1(g_2)}^{f_2} \rho_0(\varepsilon) d\varepsilon$ is the specific interfacial energy of per interface. Substituting Eq. (2.19) into Eq. (2.16), we have

$$
G(g_2) = A_0 \left\{ \int_0^{l_x} \psi_1(\varepsilon_2) - \sigma g_2 \right\} + N \gamma_{\varepsilon}(g_2)
$$

(2.20)

Thus for the homogeneous system, the resulting expression of the energy $G(g_2)$ (Eq. (2.20)) for a given $N$ only depends on the parameter $g_2$ (the maximum strain of the equilibrium strain profile).

For our inhomogeneous system, the Hamiltonians (or the integral constants) in the effective crystallite and the grain boundary regions are different as shown in Eq. (2.9) and should be discussed separately. We denote $g_0 = \varepsilon(0)$ in grain
boundary and $g_1$ and $g_2$ as the minimum and maximum values of the solution $\varepsilon_2(x)$ in effective crystallite. Since $\varepsilon_s(0) = 0$, $g_0$ can be either the minimum or the maximum value of the solution $\varepsilon_1(x)$ in grain boundary region. We can show that $g_0$ is always the minimum strain in grain boundary. If $g_0$ is the maximum strain in grain boundary, then the continuity of strain field gives $g_0 > \varepsilon(l_0/2) > g_1$. On the other hand the Eqs. (2.10) and (2.14) give $g_0 < \bar{\sigma}/E < g_1$, which is contrary with $g_0 > g_1$. So we have $\varepsilon_{\min} = g_0$ and $\varepsilon_{\max} = g_2$, and $\phi_0$ and $\phi_e$ are directly related to $\varepsilon_{\min}$ and $\varepsilon_{\max}$ as

$$\begin{align*}
\phi_0(\varepsilon_{\min}) &= \psi_0(\varepsilon_{\min}) - \bar{\sigma}\varepsilon_{\min} \\
\phi_e(\varepsilon_{\max}) &= \psi_e(\varepsilon_{\max}) - \bar{\sigma}\varepsilon_{\max}
\end{align*}$$

(2.21)

Thus we can solve the Eq. (2.8) as

$$dx = \sqrt{\frac{\alpha}{2}} \frac{d\varepsilon}{\rho(\varepsilon)}$$

(2.22)

where $\rho(\varepsilon) = \left\{ \begin{array}{ll}
\sqrt{\psi_0(\varepsilon) - \bar{\sigma}\varepsilon - \phi_0(\varepsilon_{\min})}, & \varepsilon \in [\varepsilon_{\min}, \varepsilon_p] \\
\sqrt{\psi_e(\varepsilon) - \bar{\sigma}\varepsilon - \phi_e(\varepsilon_{\max})}, & \varepsilon \in [\varepsilon_p, \varepsilon_{\max}]
\end{array} \right.$ and $\varepsilon_p$ is determined by the condition $\psi_0(\varepsilon_p) - \psi_e(\varepsilon_p) = \phi_0(\varepsilon_{\min}) - \phi_e(\varepsilon_{\max})$. According to Eqs. (2.21) and (2.9), the total Gibbs free energy $G$ of the effective 1D system (substituting Eq. (2.1) into Eq. (2.4))

$$G = 2A_0 \left\{ \int_0^{l_{0}/2} \left[ \psi_0(\varepsilon) - \bar{\sigma}\varepsilon + \frac{1}{2} \alpha \varepsilon_0^2 \right] dx + \int_{l_{0}/2}^{l_{\max}} \left[ \psi_e(\varepsilon) - \bar{\sigma}\varepsilon + \frac{1}{2} \alpha \varepsilon^2 \right] dx \right\}$$

(2.23)

$G$ can be further expressed as

$$G = A_0 \left\{ l_0 \bar{\sigma} \phi_e(\varepsilon_{\max}) + l_b \phi_0(\varepsilon_{\min}) + \gamma_{\text{interface}} (\varepsilon_{\min}, \varepsilon_{\max}) \right\}$$

(2.24)

where the interfacial energy

$$\gamma_{\text{interface}} = 2 \int_0^{l_{0}/2} \alpha \varepsilon_0^2 dx = 2\sqrt{2\alpha} \int_{\varepsilon_{\min}}^{\varepsilon_{\max}} \rho(\varepsilon) d\varepsilon + 2(n-1) \gamma \varepsilon(\varepsilon_{\max})$$

(2.25)

where $n$ is the number of domains and $1 \leq n \leq n_{\text{max}}$, $n_{\text{max}}$ is the maximum number of domains and we have $n_{\text{max}} \leq \frac{l_b}{\pi} \sqrt{\frac{\tau}{\alpha}}$.

The trivial solution can be obtained for $n = 1$ and $\varepsilon_{\min} = \varepsilon_{\max}$.

Finally, the equilibrium solution of the energy functional (Eq. (2.4)) for the inhomogeneous system is now converted into the stationary condition of the function $G(\varepsilon_{\min}, \varepsilon_{\max})$ as

$$\begin{align*}
\frac{\partial G}{\partial \varepsilon_{\min}} &= 0 \\
\frac{\partial G}{\partial \varepsilon_{\max}} &= 0
\end{align*}$$

(2.26)

For each $n$ and the corresponding $(\varepsilon_{\min}, \varepsilon_{\max})$, the solution $\varepsilon(x)$ can be integrated from the Eq. (2.22) in the form of the piecewise hyperbolic functions.

The stability of the above equilibrium solutions can be analyzed by the following two approaches:

1. In mathematically strict approach, a solution $\varepsilon^*(x)$ is called stable if the second variation of the energy functional is positive definite (on phase transition related problems, see Carr et al., 1984; Lifshitz and Rybikov, 1985; Vainchtein et al., 1998; Healey and Miller, 2007; Healey and Sipos, 2013), i.e.

$$\delta^2 G(\varepsilon^*; \zeta) = 2A_0 \left\{ \int_0^{l_{0}/2} \left( E \zeta \varepsilon + \alpha \zeta^2 \varepsilon \right) dx + \int_{l_{0}/2}^{l_{\max}} \left[ \psi_e(\varepsilon^*) \zeta^2 + \alpha \zeta^2 \right] dx \right\} > 0$$

(2.27)

for all admissible variations $\zeta(x) \in H^1(0, l_b/2)$ satisfying $\zeta(0) = \zeta_b(l_b/2) = 0$, where $\psi^*_e(\varepsilon^*) = \frac{d^2 \psi_e(\varepsilon)}{d\varepsilon^2} |_{\varepsilon=\varepsilon^*}$ is piecewise constant. The argument is similar to the proof of Theorem 8.2 in the paper by Carr et al. (1984). If $n > 2$, there must be two points $x_1$ and $x_2$ satisfying $l_b/2 < x_1 < x_2 < l_b/2$ such that

$$\varepsilon^*_x(x_1) = \varepsilon^*_x(x_2) = 0$$

(2.28)

Let

$$\zeta(x) = \zeta_0(x) + \chi \zeta_1(x)$$

(2.29)

where

$$\zeta_0(x) = \begin{cases} 
0 & 0 \leq x \leq x_1 \\
\varepsilon^*_x(x) & x_1 \leq x \leq x_2 \\
0 & x_2 \leq x \leq l_b/2
\end{cases}$$

(2.30)

and $\zeta_1(x)$ is any $H^1(0, l_b/2)$ function that satisfies $\zeta_{1,x}(0) = \zeta_{1,x}(l_b/2) = 0$ and

$$\zeta_1(0) = 0, \ z_1(x_1) = 1 \text{ and } \zeta_1(x) = 0 \text{ for } x_2 \leq x \leq l_b/2$$

(2.31)
We notice that, approximately, the system is transformable if $\eta_2 = (2) = \alpha < 20\bar{l}/E_s$.

Substituting this into Eq. (2.26), we get

$$\eta_2 = \frac{\eta_2}{\alpha} - \alpha \left( \frac{\eta_2}{\alpha} \right)^2 > 0$$

and

$$\eta_2 > 0$$

at the point $\eta_{\min} = \min(\eta(x))$ and $\eta_{\max} = \max(\eta(x))$. We should notice that there might exist limitations in the above Hessian condition. For example, it may lead to an overestimation of the stability range of metastable states (see Zurlo et al., 2017). This issue definitely needs further investigation in the future study.

### 2.3. Solutions for nano-grain

It is noticed that the maximum number of domains ($n_{\max} \leq \frac{\eta_2}{\alpha} \sqrt{\frac{-\eta_2}{\alpha}}$) in a solution $\eta(x)$ decreases with grain size $l_g$ and $n_{\max} = 1$ when $l_g$ is small enough. When $\frac{l_g}{\alpha} \sqrt{\frac{-\eta_2}{\alpha}} < 2$, we have $n_{\max} < 2$ (or $n_{\max} = 1$), i.e., $\frac{l_g}{\alpha} < 1 + \frac{E_s}{E_g} + \sqrt{\frac{E_s}{E_g}} \left( \frac{\eta_2}{\alpha} + \frac{\eta_2}{\alpha} \right)$. For instance, $n_{\max} = 1$ holds for $l_g < 40\text{nm}$ when the material parameters are $E/E_s = 4$, $\alpha/\sqrt{E_b} = 1$ and $l_g = 1\text{nm}$. To study the PT response in nc SMAs, we can focus on the case of $n_{\max} = 1$ to find the basic features of PT at nanoscale.

By setting $n_{\max} = 1$, the total Gibbs free energy $G(\eta_{\min}, \eta_{\max})$ can be expressed as

$$G = A_0 \left[ (l_g - l_b)\Phi_\eta(\eta_{\max}) + l_g \Phi_\eta(\eta_{\min}) + \gamma_{\text{interface}}(\eta_{\min}, \eta_{\max}) \right]$$

(2.32)

where

$$\gamma_{\text{interface}}(\eta_{\min}, \eta_{\max}) = 2 \int_0^{\eta_{\min}^{1/2}} \alpha \eta_\max^2 d\eta = 2\sqrt{2}\alpha \int_{\eta_{\min}}^{\eta_{\max}} \rho(\eta)d\eta$$

(2.33)

We can clearly see from the Eq. (2.32) that, the three energy terms (bulk energy, GB energy and PB energy) of the 1D effective grain are linked to the three length scales $l_g$, $l_b$ and $\sqrt{\alpha/E}$. Fig. 3 shows the landscape of the total Gibbs free energy $G$ as the function of $\eta_{\min}$ and $\eta_{\max}$ for different values of $l_g$ ($= l_g/l_b$) under a given external stress $\sigma = 0.9\sigma_n + \sigma_M$. It is seen that the landscape of the energy function changes from non-convex to convex as $l_g$ decreases, the consequence of which will be discussed in detail in the Section 3.2.

Substituting Eqs. (2.32) and (2.33) into Eq. (2.26), we can numerically determine the values of $\eta_{\min}$ and $\eta_{\max}$ as $\eta^*_{\min}$ and $\eta^*_{\max}$. By integrating Eq. (2.22), the solution $\eta(x)$ of the nano-grain can be obtained in the form of the three piecewise
hyperbolic functions as

$$\varepsilon(x) = \begin{cases} 
\varepsilon_1(x) = C_1 \cosh \left( \sqrt{\frac{E}{\alpha}} x \right) + \frac{\bar{\sigma}}{E}, & x \in [0, l_b] \\
\varepsilon_2(x) = C_2 \cosh \left( \sqrt{\frac{E}{\alpha}} \left( \frac{2x - C_3}{2} \right) \right) + \frac{\bar{\sigma} - \sigma_N}{E_g} + \frac{\sigma_N}{E}, & x \in [l_b, l_q] \\
\varepsilon_3(x) = C_4 \cosh \left( \sqrt{\frac{E}{\alpha}} \left( \frac{2x - l_q}{2} \right) \right) + \frac{\bar{\sigma}}{E} + (1 - \eta) \varepsilon_{tr}, & x \in \left[ l_q, \frac{l_g}{2} \right] 
\end{cases} \tag{2.34}$$

where the coefficients $C_1, C_2, C_3$ and $C_4$ of the hyperbolic functions can be determined as

$$\begin{align*}
C_1 &= \varepsilon_{\text{min}}^* - \frac{\bar{\sigma}}{E} \\
C_2 &= \sqrt{A^2 + B^2} \\
C_3 &= 2l_q - 2 \sqrt{\frac{\alpha}{E_g}} \arctanh \frac{A}{B} \\
C_4 &= \varepsilon_{\text{max}}^* - (1 - \eta) \varepsilon_{\text{tr}} - \frac{\bar{\sigma}}{E}
\end{align*}$$

In the above, $A$ and $B$ are

$$\begin{align*}
A &= \left( \varepsilon_{\text{min}}^* - \frac{\bar{\sigma}}{E} \right) \sqrt{\frac{E}{E_g}} \sinh \left( \sqrt{\frac{E}{\alpha} l_q} \right) \\
B &= \left( \varepsilon_{\text{max}}^* - \frac{\bar{\sigma}}{E} \right) \cosh \left( \sqrt{\frac{E}{\alpha} l_q} \right) + \left( \frac{1}{E} - \frac{1}{E_g} \right) (\bar{\sigma} - \sigma_N)
\end{align*}$$

$l_b$ and $l_q$ are the boundaries of $\varepsilon_2(x)$ which can be calculated case by case as shown in Fig. 4 (see Li, 2017 for detail). According to the sorting results of $\varepsilon_{\text{min}}^*, \varepsilon_{\text{max}}^*, \varepsilon_s$ and $\varepsilon_c$ ($\varepsilon_s$ and $\varepsilon_c$ are strains at the boundary of spinodal region as shown in Eq. (2.3) and Fig. 2(c)), all the six possible equilibrium strain profiles are summarized in Fig. 4.

The approximate analytical expressions of $\varepsilon_{\text{max}}^*$ and $\varepsilon_{\text{min}}^*$ can be obtained if the specific interfacial energy $\gamma_{\text{interface}}(\varepsilon_{\text{max}} - \varepsilon_{\text{min}})$ is approximated as the following. For large $l_g$ (i.e., $l_g \gg l_b$), $\gamma_{\text{interface}} \sim \sqrt{E\alpha} (\varepsilon_{\text{max}} - \varepsilon_{\text{min}})^2$, see Cahn and Hilliard, 1958) is independent of $l_g$ and the total interfacial energy term $A_0 \gamma_{\text{interface}}$ is ignorable compared with
the bulk energy terms. For small \( l_g \) (i.e., \( l_g > l_g^\bullet \)), the strain changes smoothly along the whole effective grain as shown in Fig. 4(d)-(f) where the interfacial structure and energy depend strongly on \( l_g \) at nanoscale. Such increasing constraints of GB with decreasing GS has been used in explaining size-dependent PT behaviors in recent experiments and MD simulations (Sun et al., 2014; Ahadi and Sun, 2015; Qiao and Radovitzky, 2016; Ko et al., 2017). For small \( l_g \), the strain gradient \( \varepsilon_x \) is of the order of \( (\epsilon_{\text{max}} - \epsilon_{\text{min}})/l_g \), i.e.,

\[
\varepsilon_x \sim \Delta \epsilon / \Delta x = 2(\epsilon_{\text{max}} - \epsilon_{\text{min}})/l_g
\]  

(2.35)

From Eq. (2.33) it is seen that, \( \gamma_{\text{interface}} \) is of the order of \( \alpha \epsilon_x^2 l_g \), i.e.,

\[
\gamma_{\text{interface}} \approx 4\alpha (\epsilon_{\text{max}} - \epsilon_{\text{min}})^2/l_g
\]  

(2.36)

The above estimation is acceptable as shown in Fig. 7(b), where the value of \( \gamma_{\text{interface}} \) by Eq. (2.36) is very close to the exact value of \( \gamma_{\text{interface}} \) calculated from Eq. (2.33) for small \( l_g \).

Using Eqs. (2.36) and (2.32), we can have approximate analytical solutions of \( (\epsilon^*_{\text{min}}, \epsilon^*_{\text{max}}) \) from Eq. (2.26) as follows:

- **Solution I** – uniform stable solution: \( \epsilon^*_{\text{min}} = \epsilon^*_{\text{max}} = \bar{\sigma}/E \).

This solution exists only when \( \epsilon^*_{\text{max}} < \epsilon_c \Rightarrow \bar{\sigma} < \sigma_N \) (i.e., \( \epsilon^*_{\text{max}} = E\bar{E}_{\text{tr}}/[2(E_N + E) + \sigma_M] \), \( \sigma_N \) is nucleation stress), \( \bar{\sigma} \) is the external applied stress. This stationary solution is always stable since \( \Delta = E^2 l_g^2 (l_g - 1) \geq 0 \).

- **Solution II** – nonuniform stable solution:

\[
\begin{cases}
\epsilon^*_{\text{min}} = \frac{\bar{\sigma}}{E} + \frac{(1 - \xi)\epsilon_{\text{tr}}}{(l_g - 1)/\beta + 1} \\
\epsilon^*_{\text{max}} = \frac{\bar{\sigma}}{E} + \epsilon_{\text{tr}} - 2\mu(l_g)\sigma_N/E
\end{cases}
\]  

(2.37)

where \( \mu(l_g) = (E/E_N + 1)[(\xi - (1 - \xi)/(l_g - 1 + \beta)]/E + \epsilon_{\text{tr}} - 2\mu(l_g)\sigma_N/E \) and \( \beta = 8\alpha/(E l_g^2) \) are non-dimensional quantities. The role of \( \mu(l_g) \) will be discussed in Section 3.

This solution exists only when \( \epsilon^*_{\text{max}} > \epsilon_c \Leftrightarrow \bar{\sigma} > \sigma_N - 2(1 - \mu(l_g))(\sigma_N - \sigma_M) \). It is always stable since \( \Delta = E^2 l_g^2 (l_g - 1 + \beta) \geq 0 \).

- **Solution III**:

\[
\begin{cases}
\epsilon^*_{\text{min}} = \frac{\bar{\sigma}}{E} + \frac{1/E_N}{1 - \mu(l_g)} + \frac{1}{E} \left( \frac{\sigma_N - \sigma_N}{(l_g - 1)/\beta + 1} \right) \\
\epsilon^*_{\text{max}} = \frac{\bar{\sigma}}{E} + \frac{1/E_N}{1 - \mu(l_g)} + \frac{1}{E} \left( \frac{\sigma_N - \sigma_N}{(l_g - 1)/\beta + 1} \right)
\end{cases}
\]  

(2.38)

The solution exists only when \( \epsilon_c \leq \epsilon^*_{\text{max}} \leq \epsilon_c \). Thus for the case \( \mu(l_g) < 1 \), the solution exists when \( \sigma_N - 2(1 - \mu(l_g))(\sigma_N - \sigma_M) \leq \bar{\sigma} \leq \sigma_N \); for the case \( \mu(l_g) > 1 \), the solution exists when \( \sigma_N \leq \bar{\sigma} \leq \sigma_N - 2(1 - \mu(l_g))(\sigma_N - \sigma_M) \). The stability of this stationary solution is judged by the condition of convexity as \( \Delta = EE_N l_g^2 ((l_g - 1)/\beta + 1)/\mu(l_g - 1) > 0 \). Thus this solution is stable for \( \mu(l_g) > 1 \), while unstable for \( \mu(l_g) < 1 \). It must be noticed that the effective 1D system considered here displays a strong GS dependent Gibbons free energy landscape in configuration space with multiple stable states separated by energy barriers. Once the energy barrier is removed by the applied stress, the PT takes place from a metastable state to a more stable state with energy dissipation, leading to the stress hysteresis in a loading-unloading cycle. Thus the free energy of the two metastable states (i.e., the energy difference of the two local minima) under the external stress is essential to the stress-strain hysteresis loop. In fact, as shown in Section 3.2, when GS is reduced below a critical value, the free energy function become convex with only one minimum (unique stable state) for any given external stress. As such, the deformation of the system is stable and thermodynamically reversible with no hysteresis and no dissipation.

In summary, for \( \mu(l_g) < 1 \), we have only one stable solution in the external applied stress range of \( \bar{\sigma} > \sigma_N \) and in the range of \( \bar{\sigma} < \sigma_N - 2(1 - \mu(l_g))(\sigma_N - \sigma_M) \), two stable solutions and one unstable solution in the range of \( \sigma_N - 2(1 - \mu(l_g))(\sigma_N - \sigma_M) < \bar{\sigma} < \sigma_N \). For \( \mu(l_g) > 1 \), there is only one stable solution under any external applied stress \( \bar{\sigma} \).

Applying the definition of the average strain \( \bar{\varepsilon} = \frac{1}{l_g} \int_0^{l_g} \varepsilon(x) dx \) and Eq. (2.34) to the 1D grain, we can obtain the nominal stress-strain relation \( (\bar{\sigma} \sim \bar{\varepsilon}) \) for different values of \( l_g \) under stress-controlled loading-unloading cycle (as shown in Fig. 5). It is seen that with the reduction of \( l_g \), three new features of the responses can be observed: 1) from the plateau type \( \bar{\sigma} \sim \bar{\varepsilon} \) (snap-through instability) to hardening type (stable) response and energy dissipation (hysteresis loop area) decreases and finally vanishes; 2) transformation strain decreases; 3) the reverse transformation stress \( \sigma_M \rightarrow A \) increases and eventually becomes far above the Maxwell stress \( \sigma_M \), indicating that there is no two-phase coexistence during deformation. The detailed analysis of the GS dependence of energy dissipation, two-phase coexistence and the underlying energy partition will be given in Section 3.
Fig. 5. The stress-strain relation ($\bar{\sigma}$ ~ $\bar{\varepsilon}$) for different values of $\bar{l}_g$ under stress-controlled loading-unloading cycle. (a) $\bar{l}_g = 100$; (b) $\bar{l}_g = 50$; (c) $\bar{l}_g = 20$; (d) $\bar{l}_g = 10$. The material parameters used are: $E/\bar{\varepsilon}_s = 4$ and $\bar{\sigma}/\bar{\varepsilon}_s^0 = 1$.

3. Model predictions of GS dependence of PT responses

In all coarse-grained polycrystalline SMAs and bulk single crystals, the first order PT process is characterized by the two important signatures: (1) nucleation of new phase and two-phase coexistence and (2) stress hysteresis or energy dissipation. In this section, we will give the model predictions of the important length scales effects on the two-phase coexistence in Section 3.1 and on energy dissipation in Section 3.2. The critical length scale $\bar{l}_g$ for the breakdown of two-phase coexistence and $\bar{l}_g^0$ for the vanishing of hysteresis are identified. We analyze the energy partition and competition in the effective 1D PT system and provide insight to such fundamental changes of PT behaviors in Section 3.3.

3.1. Breakdown of A-M two-phase coexistence

In the case of coarse grain, the PB is characterized as a transient zone of nanometer thickness separating high and low strain regions (phases). The calculated strain profiles by Eq. (2.34) for $\bar{l}_g = 1500$ and $\bar{l}_g = 18$ are shown in Fig. 6(a) and (b). It is seen that, due to the increased GB constraint with $\bar{l}_g$ reduction, phase transformation or deformation inside the 1D grain indeed changed its deformation mode from the coexistence of well-defined high (A) and low (M) strain phases separated by “sharp” PB to the continuous but more “smooth” strain profile containing a relatively “thick” PB. This model prediction is supported by both the recent MD simulation (see Fig. 10 in Ko et al., 2017) and the recent in-suit XRD profiles (Ahadi and Sun, 2015) during loading-unloading for coarse-grained sample (GS = 1500nm, Fig. 6(c)) and nano-grained sample (GS = 18nm, Fig. 6(d)). The former shows two typical peaks representing Austenite and Martensite phases, but the later only shows a reversible shift and broadening of a single peak, indicating a continuous and nonuniform distortion of the crystal.

To be more quantitative, we use $\Delta \varepsilon = \varepsilon_{max}^* - \varepsilon_{min}^*$ to define the condition for two-phase coexistence as

$$\varepsilon_s - \varepsilon_L \leq \varepsilon_{max}^* - \varepsilon_{min}^* \leq \varepsilon_{tr}$$

(3.1)

where $\varepsilon_s = \sigma_N/E$ and $\varepsilon_L = (\sigma_N - 2\sigma_M)/E$ (see Figs. 2(c), 6(a) and (b)). The Eq. (3.1) comes from the physical conditions that $\varepsilon_{max}^*$ of the high strain region should be no less than $\varepsilon_s$ (i.e., $\varepsilon_{max}^* \geq \varepsilon_s$) and $\varepsilon_{min}^*$ of the low strain region should be no more than $\varepsilon_L$ (i.e., $\varepsilon_{min}^* \leq \varepsilon_L$).

In fact, we only need to consider the nonuniform stable solution II (Eq. (2.37)) for which

$$\Delta \varepsilon = \varepsilon_{max}^* - \varepsilon_{min}^* = \frac{\bar{l}_g(1 - \xi)}{\bar{l}_g - 1 + \beta} \varepsilon_{tr}$$

(3.2)
It is seen that, $\Delta \varepsilon \to \varepsilon_{tr}$ when $\bar{l}_g \to \infty$. Substituting Eq. (3.2) into the Eq. (3.1), we obtain the critical size $\bar{l}_g^o$ above which the two-phase coexistence prevails

$$\bar{l}_g \geq \bar{l}_g^o \approx (\beta + 2)E/E_s + 3.$$  \hspace{1cm} (3.3)

For example, $\bar{l}_g^o = 43$ for the typical material parameters $E/E_s = 4$ and $\beta = 8\alpha/(E_l^2) = 8$. Assuming $l_b = 1$ nm, we will get $\bar{l}_g^o = 43$ nm, which is close to the experimentally observed critical grain size of 32 nm of nc NiTi for the breakdown of the two-phase coexistence (Ahadi and Sun, 2015).

It must be pointed out here that there are two physical origins underlying the loss of two-phase coexistence in a finite system. The first one is the length scale of the phase boundary in the gradient theory of phase transitions, which is proportional to $\sqrt{\alpha}$ ($\alpha$ is the gradient coefficient). It is well known that the phase boundary spreads over the whole domain (which changes the overall behavior) when the gradient coefficient becomes large enough, see for instance (Alikakos and Simpson, 1987; Alikakos and Shaing, 1987). Conceptually similar effects were also studied in micromagnetics (see DeSimone, 1994 and 1995). The second physical origin for the loss of two-phase coexistence, as revealed by this paper, is the nano-scale grain boundary dominance which controls the energy landscape of the considered crystallite-amorphous (or Core-shell) composite system by changing the non-convexity of the energy density function (see Fig. 2(d)). Still using the gradient theory of phase transition, the main focus of this paper is on the consequence of grain size reduction to the overall behavior of nc polycrystal. As will be shown below (see Section 3.3 for details), for given length scales of the grain boundary ($l_b$) and phase boundary ($\sqrt{\alpha/E}$), the energy of the grain boundary gradually becomes dominant over the other energy terms when the GS is decreased down to nano-scale, which eventually brings fundamental changes of the phase transition behaviors including the loss of two-phase coexistence.

Now we further discuss the novel features of the interfacial energy ($\gamma_{interface}$) after the breakdown of two-phase coexistence at nanoscale. We can define the non-dimensionalized interfacial energy as the ratio of interfacial over bulk energy: $\bar{\gamma}_{interface} = A_0 \gamma_{interface}/(V_0 \sigma N E_{tr}) = \gamma_{interface}/(l_b \sigma N E_{tr})$. For the coarse grain ($l_g \gg l_b$), $\gamma_{interface}$ can be taken as a constant (Eq. (2.33)), so $\bar{\gamma}_{interface}$ follows the scaling law of $1/l_g$, which indicates the gradual dominance of $\gamma_{interface}$ over the bulk term with $l_g$ reduction for coarse grain. However, such scenario changes drastically in nc SMAs. As shown in Fig. 7(a) and (b), $\bar{\gamma}_{interface}$ changes non-monotonically with $l_g$ since $\gamma_{interface}$ (Eqs. (2.33) and (2.36)) in the strongly constrained deformation at nanoscale is not constant any more. The breakdown of $1/l_g$ scaling for $\bar{\gamma}_{interface}$ at nanoscale is attributed to the breakdown of two-phase coexistence.
3.2. Vanishing of hysteresis and dissipation

As shown in Fig. 3, the effective 1D system considered here displays a strong GS dependent Gibbs free energy landscape in configuration space with multiple stable states (with high and low strain phases) separated by energy barriers. Once the energy barrier is removed by the applied stress, the PT takes place from a metastable state to a more stable state with energy dissipation, leading to the hysteresis in a loading-unloading cycle. Thus the stress-strain hysteresis loop is directly related to the difference between the free energy values of the two states under the external stresses. In fact, when GS is below a critical value, the free energy function becomes convex with only one minimum (unique stable state) for any given external stress. As such, the deformation of the system is stable and thermodynamically reversible with no hysteresis and no dissipation.

To determine the critical size below which the hysteresis vanishes, we have to make sure the Gibbs free energy function $G$ has unique stationary point ($\varepsilon_{\text{min}}^*, \varepsilon_{\text{max}}^*$) and is convex at this point. According to the analysis in Section 2.2, there are always $\Delta > 0$ and $\partial^2 G / \partial \varepsilon_{\text{min}}^* > 0$ only when $\mu(\bar{l}_g) > 1$. Thus the energy function $G$ is convex when

$$\mu(\bar{l}_g) > 1 \Rightarrow \bar{l}_g < \bar{l}_g^{\text{hy}}$$

(3.4)

In the above criteron, $\bar{l}_g^{\text{hy}}$ is the critical length scales for the vanishing of hysteresis and $\bar{l}_g^{\text{hy}} \approx E/E_s + 3/2 - \beta/2 + \sqrt{(E/E_s + 3/2 - \beta/2)^2 + 3(E/E_s + 1)(\beta - 1)}$. The value of $\bar{l}_g^{\text{hy}}$ is insensitive to $\beta$ ($= 8\sigma / (E\bar{l}_g^3)$) for $0 < \beta < \infty$ since we have

$$2(E/E_s + 1) < \bar{l}_g^{\text{hy}} < 3(E/E_s + 1)$$

(3.5)

For example, $10 < \bar{l}_g^{\text{hy}} < 15$ for the typical material parameters $E/E_s = 4$. Taking $\bar{l}_g = 1\text{nm}$, we have $10\text{nm} < \bar{l}_g^{\text{hy}} < 15\text{nm}$, which is close to the experimental value of GS=10 nm for near vanishing hysteresis in nc NiTi (Ahadi and Sun, 2013).

For $\bar{l}_g > \bar{l}_g^{\text{hy}}$, the dissipated energy can be calculated as the difference of the Gibbs free energies of the two metastable states under the nucleation stress $\bar{\sigma} = \sigma_N$. Thus the dissipated energy per unit volume in forward PT is

$$\Delta g = (G_{m1} - G_{m2})/V_0$$

(3.6)

where $G_{m1} = \frac{-(\sigma_N - \sigma_M)^2V_0/(2E)}{}$ is the Gibbs free energy for the uniform solution I (at $\bar{\sigma} = \sigma_N$, and $G_{m2} = \frac{-(\sigma_N - \sigma_M)^2V_0/(2E) - (\sigma_N - \sigma_M)\varepsilon_N(1 - \xi)^2(1 - \mu(\bar{l}_g))V_0}{\text{for the nonuniform solution II (at} \bar{\sigma} = \sigma_N\text{). Taking the dissipated energy in reverse PT the same as in forward PT, the specific hysteresis dissipation in one loading-unloading cycle is}}$

$$H_g = 2\Delta g = H_0(1 - \xi)[1 - \mu(\bar{l}_g)]$$

(3.7)

where $H_0 = 2(\sigma_N - \sigma_M)\varepsilon_{tr} = EE_N\varepsilon_{tr}^2/(E + E_s)$ is the hysteresis during stress-controlled loading-unloading cyclic in coarse-grained polycrystals ($\bar{l}_g \rightarrow \infty$). The hysteresis dissipation can be calculated approximately as $H_g/H_0 \approx (1 - 3/\bar{l}_g)(1 - \bar{l}_g^{\text{hy}}/\bar{l}_g) \approx$
Fig. 8. (a) The scaling law of hysteresis dissipation and the comparison with the experiment results (Ahadi and Sun, 2015). (b) The characteristic curve family $G_m = G_m(\varepsilon_m, \bar{l}_g)$ of Gibbs free energy with $\bar{l}_g$ as the parameter.

$1 - \bar{l}_g^{hy} / \bar{l}_g$, which is valid when $\bar{l}_g^{hy} \gg 3$. This scaling law well captures the experimentally observed GS dependence of hysteresis loop area (Ahadi and Sun, 2013) as shown in Fig. 8(a). From the above analysis, we can see that the difference between nucleation stress and the Maxwell stress ($\sigma_m - \sigma_M$) is directly related to the energy dissipation during PT. Thus for simplification, we set $\sigma_m = 0$ in the following discussion.

To give a clear picture of the effect of $\bar{l}_g$ on hysteresis dissipation, we choose the characteristic curve $G_m$ that is the intersection between the vertical plane passing through the solution I and the solution II and the free energy function surface $G = G(\varepsilon_{min}, \varepsilon_{max})$. With $\bar{l}_g$ as the parameter, we get a family of the characteristic curve $G_m = G_m(\varepsilon_m, \bar{l}_g)$ where the variable $\varepsilon_m = \sqrt{(\varepsilon_{max} - \bar{\sigma}/E)^2 + (\varepsilon_{min} - \bar{\sigma}/E)^2}$. Fig. 8(b) shows the curve family $G_m$ under the nucleation stress $\bar{\sigma} = \sigma_N$ (the vertical plane is $\beta \varepsilon_{max} - (\bar{l}_g + \beta) \varepsilon_{min} + \frac{1}{2} \sigma_N / E = 0$). It is seen that the dissipated energy ($G_{m1} - G_{m2}$) decreases with $\bar{l}_g$ and eventually vanishes at the critical length scale $\bar{l}_g = \bar{l}_g^{hy}$. It is also seen that at the nanoscale PT behaves like second-order with $\varepsilon_m$ as the order parameter. This is also consistent with the result in Fig. 7(a) that $\gamma_{interface}$ decreases and tends to vanish with $\bar{l}_g$ reduction.

It should be noticed that under certain conditions such as matrix elasticity or mixed loading device (see Budiansky and Truskinovsky, 1993; Vainchtein et al., 1998), phase transition can also proceed without hysteresis. By adopting an energy density function with a coupling term of dilatation and shear strain, it is shown that the phase transition under certain stress paths can only proceed with stable and continuous strain changes. To reflect the effects of matrix elasticity on phase transition, a convex elastic foundation energy term is added to the three-parabola nonconvex model (Vainchtein et al., 1998), then the energy functional of the system can become convex when the elastic foundation stiffness (see page 37-39 of Vainchtein et al., 1998) is large enough.

3.3. Energy partition of 1D effective grain − GB energy dominance

The total Gibbs free energy of the considered system can be divided into three parts $G = G_{crystalite} + G_{GB} + G_{interface}$. According to Eqs. (2.32) and (2.2), we have

$$
G_{crystalite} = V_0[1 - \xi] \psi_c (\varepsilon_{max}^* - \bar{\sigma} \varepsilon_{max}^*)
$$

$$
G_{GB} = V_0 [\xi \psi_0 (\varepsilon_{max}^* - \bar{\sigma} \varepsilon_{max}^*) - \varepsilon_{tr} (\bar{\sigma} \varepsilon_{max}^* - \bar{\sigma}) + \bar{l}_g \beta G_0^2 \varepsilon_{tr}^2 / 2]
$$

$$
G_{interface} = V_0 \beta G_0^2 \varepsilon_{tr}^2 / 2
$$

(3.8)

where $\xi = 1 - (1 - \bar{l}_g^2)^3 \approx 3 \bar{l}_g$ is the volume fraction of GB for the representative cubic grain, $\bar{\sigma} = (\varepsilon_{max}^* - \varepsilon_{min}^*) / (\varepsilon_{tr} \bar{l}_g)$ is the non-dimensionalized strain gradient and $\beta = 8 \alpha / (E \bar{l}_g^2)$ is the non-dimensionalized strain gradient coefficient.

We consider the change of each part during PT, e.g., under the nucleation stress $\bar{\sigma} = \sigma_N$. According to the analytical solutions in Section 2.2, we have $\varepsilon_{max}^* = \sigma_N / E$ and $\bar{\sigma} = 0$ before PT, and $\varepsilon_{max}^* \approx \sigma_N / E + \varepsilon_{tr}$ and $\bar{\sigma} = (1 - \xi) / (\bar{l}_g - 1 + \beta)$ after
PT. Thus the changes (normalized by the quantity $V_0\sigma\Delta N^\xi$) of the three parts are

$$\begin{align*}
\Delta\tilde{g}_{\text{released}} &= -\Delta C_{\text{crystallite}} / (V_0\sigma N^\xi) \approx 1 - \xi \\
\Delta\tilde{g}_{\text{GB}} &= \Delta G_{\text{GB}} / (V_0\sigma N^\xi) \approx \left( \xi - 2\rho \tilde{l}_g g^2 \right) E/E_s \\
\Delta\tilde{g}_{\text{interface}} &= \Delta G_{\text{GB}} / (V_0\sigma N^\xi) = \beta (1 + E/E_s) g^2 
\end{align*}$$

(3.9)

The variations of $\Delta\tilde{g}_{\text{released}}, \Delta\tilde{g}_{\text{GB}}$ and $\Delta\tilde{g}_{\text{interface}}$ with $\tilde{l}_g$ are shown in Fig. 9(a). Three features of the grain size dependence at nanoscale can be found: (1) the amount of interfacial energy change $\Delta\tilde{g}_{\text{interface}}$ is relatively small ($\Delta\tilde{g}_{\text{interface}} \leq 8.5\%$ for different $\alpha/(E/E_s)$ when $E/E_s = 4$ as shown in Fig. 9(b) and is ignorable compared with the bulk energy terms; (2) the energy change in GB $\Delta\tilde{g}_{\text{GB}}$ gradually becomes dominant over the other two parts with the $\tilde{l}_g$ reduction; (3) when $\tilde{l}_g > \tilde{l}_g^\text{PT}$ the PT (instability) can occur spontaneously and the difference $(\Delta\tilde{g}_{\text{released}} - \Delta\tilde{g}_{\text{GB}})$ represents the amount of energy dissipated during the forward PT, but when $\tilde{l}_g < \tilde{l}_g^\text{PT}$, there is no energy dissipation (no instability) and the deformation can only be driven by further stressing.

If further ignoring the interfacial energy by taking $\rho \approx 0$, we have $\Delta\tilde{g}_{\text{GB}} \approx (E/E_s)(3/\tilde{l}_g^\text{PT})$ and $\Delta\tilde{g}_{\text{released}} \approx 1 - 3\tilde{l}_g$ for simplification. Thus we can even more clearly see that the above fundamental changes of stress-induced PT behaviors are mainly attributed to the significantly increased volume fraction $\xi$ of GB ($\xi \approx 3/\tilde{l}_g^\text{PT}$) and the difference between the properties of GB and the crystallite ($E/E_s$).

It should be noticed that all the above analysis and results are based on the continuum modelling with incorporation of microstructural internal length scales of the material. Though the results obtained get support from both experiments (Ahadi and Sun, 2013, 2015) and meso-scale and MD simulations (Ahlulwalia et al., 2015; Ko et al., 2017), it is recognized that such top-down continuum approach will break down at certain limiting length scale. Also, the GB energy in the present model is estimated as a pure linear elastic phase. Other grain boundary mechanisms (e.g., dislocation, sliding and diffusion) should be included in the future modelling of deformation behavior of nc SMAs.

4. Summary and Conclusions

Nucleation and growth of new phase, two-phase coexistence and the hysteresis dissipation are the key signatures of the first-order phase transition in materials and are widely used as the basic paradigms in the continuum modelling of deformation behavior of most polycrystalline shape memory alloys where grain size is much larger than the length scales of grain boundary and phase boundary. The current technology has been able to manufacture nc SMAs with grain size down to 5–10 nm. The unusual phenomena and properties emerged in these nc materials not only open up new possibilities in the application but also provide challenging issues not only in modelling and understanding but also in paradigm building and shift.

We investigated the effects of grain size, grain boundary and phase boundary on the deformation behavior of nc SMAs in the framework of continuum mechanics with incorporation of the internal length scales $l_b, l_g$ and $\sqrt{\alpha/E}$, respectively. The strong constraint of the 3D network of grain boundaries is quantified by a core-shell type of 3D → 1D effective modelling and closed form analytical solutions are obtained. The key novelty and insights of nanoscale phase transition behavior of SMAs obtained in the paper are the following:

- Physical and quantitative pictures of phase transition process in nc SMAs are given, for the first time, by the established 1D effective modelling and the solutions with embedded length scales of $l_b, l_g$ and $\sqrt{\alpha/E}$. These length scales are directly linked to the changes of the released bulk crystallite energy, the grain boundary energy and the phase boundary energy. The non-dimensionalized length scales $l_g = l_b/l_g$ is identified as the key geometrical governing parameter in con-
trolling the response of the 1D microstructured grain under given external stress and intrinsic properties of the crystallite (modulus $E$, $E_s$ and transformation strain $\epsilon_T$).

- With the decrease of $I_g$, the grain boundary energy that is ignorable in the coarse grained situation gradually becomes dominant over the other energy terms, this eventually brings fundamental changes of the phase transition behaviors by convexifying the total free energy of the system. It is revealed that nanoscale phase transition is realized in the manner of the second-order with stable, continuous and reversible distortion of crystallite and that phase boundary energy decreases and tends to vanish with $I_g$ reduction.

- The model predictions of the critical grain size $I^\text{c}_g$ for the breakdown of two-phase coexistence and the critical grain size $I^\text{c}_g$ for the vanishing of hysteresis dissipation are obtained and expressed in terms of the intrinsic properties of the crystallite ($E/E_s$, $\epsilon_T$) and the length scales $I_g$ and $\sqrt{\alpha/E}$ of the material without fitting parameters. The predictions are supported by the available experimental data and recent MD simulations.

Acknowledgments

The financial support to this work by the Hong Kong Research Grants Council (project No. 16209414) and the National Natural Science Foundation of China (project No. 11532010) are gratefully acknowledged.

References


