

An Origin of Functional Fatigue of Shape Memory Alloys

Y.Gao^a, L. Casalena^a, M. Bowers^a, R. Noebe^b, M.J.Mills^a, Y.Wang^{a,*}

^a*Department of Materials Science and Engineering, The Ohio State University, Columbus, OH 43210, United States*

^b*NASA Glenn Research Center, Materials and Structures Division, Cleveland, OH 44135, United States*

Abstract

Functional fatigue (FF) during thermal and mechanical cycling, which leads to the generation of macroscopic irrecoverable strain and the loss of dimensional stability, is a critical issue that limits the service life of shape memory alloys (SMAs). Although it has been demonstrated experimentally that such a phenomenon is related to microstructural changes, a fundamental understanding of the physical origin of FF is still lacking, especially from a crystallographic point of view. In this study, we show that in addition to the normal martensitic phase transformation pathway (PTP), there is a symmetry-dictated non-phase-transformation pathway (SDNPTP) during phase transformation cycling, whose activation could play a key role in leading to FF. By investigating crystal symmetry changes along both the PTPs and SDNPTPs, the characteristic types of defects (e.g., dislocations and grain boundaries) generated during transformation cycling can be predicted systematically, and agree well with those observed experimentally in NiTi. By analyzing key materials parameters that could suppress the SDNPTPs, strategies to develop high performance SMAs with much improved FF resistance through crystallographic design and transformation pathway engineering are suggested.

Keywords: Shape memory alloys; Martensitic phase transformations; Crystal symmetry; Alloy design; Defects.

1. Introduction

Since its discovery [1], shape memory alloys (SMA) have found many advanced applications in medical and engineering devices [2-5]. However, one of the critical issues limiting the application of SMAs is functional fatigue (FF) during repeated actuation through either thermal or load cycling (i.e., transformation cycling). It is reported that macroscopic irrecoverable strain (also called open-loop strain) can be accumulated in commercial SMAs, resulting in the loss of dimensional stability [6-18]. Even though such

FF dictates the functionality, durability and service life of SMAs, its physical origin is far from clear and, thus, a systematic way to characterize and control it is still lacking. This is reflected partly by the various loose terminologies introduced in the literature to describe the phenomenon over the years, including “functional fatigue” and “low temperature creep” [6, 7]. Fatigue is commonly associated with crack initiation and propagation during cyclic loading while creep is related to time-dependent deformation at elevated temperatures, and neither of them is necessarily related to phase transformations. In contrast, FF found in SMAs is associated exclusively with structural phase transformations. Even though it is similar to fatigue and creep with regard to irrecoverable deformation caused by external loads below the yield strength, it is transformation-cycle-dependent rather than just stress-cycle- or time-dependent and does not necessarily happen at elevated temperatures. Keeping these differences in mind we formulate in this study a new theoretical framework based on crystallographic theory of structural phase transformations to reveal a unique physical origin and distinctive features of FF in at least some SMAs. However, to keep consistency with existing literature we continue to use the term “functional fatigue”.

Ample experimental results have shown that a considerable amount of crystalline defects are generated during load- or temperature-cycling in a number of typical SMA systems [6-8, 16-19]. In comparison with what happens in conventional dislocation plasticity, defect generation during FF has several distinctive features. As observed in experiments [14-18], the irrecoverable strain can be accumulated during thermal cycling around the martensitic transformation (MT) temperature with a bias-load much lower than the yield stress, and the total irrecoverable strain depends strongly on the number of cycles. This suggests a strong correlation between defect generation and accumulation and transformation cycling. More interestingly, in addition to the generation of dislocations, grain refinement (without apparent recrystallization) and formation of special grain boundaries (Σ boundaries) have been observed during the transformation cycling [19]. These recent observations imply not only a crystallographic origin of the defects, but also a unique generation mechanism, different from that in conventional dislocation plasticity. At a higher level, these results suggest a fundamental relationship between FF and the unique structural change (i.e., symmetry change) during phase transformation cycling.

Several crystallographic studies have attempted to understand defect generation during MTs. Bhattacharya, et al. [20] proposed that if the symmetry groups of the parent and product phases cannot be included in a common finite group, dislocations are inevitably generated. On the other hand, if a group-subgroup relationship exists, the martensitic transformation should be fully reversible and no other lattice defects such as dislocations should be generated, assuming that local stresses produced by the MT would not cause plastic yielding. However, for the most widely used commercial SMA, the NiTi system,

even if it undergoes a B2 to B19' MT that satisfies the group-subgroup relationship and self-accommodation among 12 martensitic variants can be accomplished by 192 twinning modes, defect generation is still widely observed during thermal cycling especially under a biased-load that is much lower than the yield stress [6, 7, 11-14]. On the other hand, another crystallographic criterion called the “cofactor condition” was proposed recently for the design of SMAs, and Zn-Au-Cu and Ni-Ti-Cu alloys have been identified for enhanced reversibility using (or partially using) this criterion [21, 22]. However, such a criterion is rather strict and only the lattice parameters of specific systems can satisfy coincidentally the cofactor condition. Furthermore, to study the FF phenomenon, the effect of biased load on the generation of defect structures has to be considered.

In this article, the physical origin of FF is investigated by phase transformation pathway (PTP) and symmetry-dictated non-phase-transformation pathway (SDNPTP) analyses. It is demonstrated that during transformation cycling, defects including both dislocations and special grain boundaries are generated due to symmetry breaking along the SDNPTPs that are easily activated by either internal or external stresses. By investigating the symmetry change along the SDNPTPs in NiTi, the types of defects induced by the transformation cycling are predicted systematically, revealing $\langle 100 \rangle \{011\}$ type dislocations and Σ grain boundaries, which agree with the experimental observations [19]. Through the PTP and SDNPTP analyses, key material parameters are identified and strategies to improve FF resistance of SMAs are developed. A systematic way of designing SMAs with improved FF resistance is also discussed from the crystallographic point of view.

2. Crystallographic origin of defect generation during martensitic transformation with group-subgroup relationship

As a classical description of the energetics of phase transformations with symmetry breaking in Landau theory, the free energy landscape is represented by a Landau polynomial in which the free energy of nearby low symmetry states (referred to as the Ericksen-Pitteri Neighborhood (EPN) [23-24]) are described by a series expansion with respect to the high symmetry parent phase state. In particular, when the order parameter characterizing the structural change during the phase transformation is chosen as the uniform lattice distortion (i.e., inelastic strain), the PTPs described by the Landau polynomial could be taken as *localized* deformation pathways. On the other hand, because of the translational symmetry of a crystal that leads to an infinite symmetry group (i.e., space group), there are infinite numbers of ways of lattice invariant deformations. As a consequence, there are infinite numbers of deformation states describing the same lattice in the deformation space, and all of these states are connected by *non-localized* deformation pathways (lattice invariant deformation) dictated by the

crystal symmetry, leading to an infinite pathway network in the deformation space [20]. In general, all the above non-localized deformation pathways are intrinsic features of a given crystal lattice, not necessarily involved in any phase transformations. Even though both PTPs and NPTs are in the deformation space, they are usually treated independently in the literature, which is a good simplification when the deformation caused by phase transformation is much smaller than the lattice invariant deformation. However, if the lattice distortion associated with a phase transformation is large enough (i.e., in most SMAs), the PTPs and NPTP could be interconnected and hence new pathways may appear, which has to be taken into account in analyzing domain and defect structures. Here we refer to the new pathways as SDNPTs in order to distinguish them from either PTPs or the deformation pathways related to conventional dislocation and twinning plasticity.

To illustrate the interconnection between the PTPs and SDNPTP as well as the connection between SDNPTs and FF in SMAs at the intuitive level, a simplified 2D example is presented. For MT from a rectangular lattice (α , point group: 2mm) to a parallelogram lattice (β martensite, point group: 2), the transformation mechanism could be described as a simple shear deformation (Fig. 1). As a result, each atom in the α lattice corresponds to an atom in the β lattice, which establishes a lattice correspondence between the two lattices as illustrate by the red frames in Fig. 1. If the absolute value of the shear strain is smaller than 0.5, the free energy landscape among the parent phase structural state and the two nearby crystallographically equivalent and energetically degenerate martensitic structural states (i.e., β -1 and β -2 deformation variants [25]) can be described by a Landau polynomial with the transformation strain (shear strain) being the order parameter (Fig. 1).

However, such a description of PTPs is local and the non-localized lattice-invariant deformation pathways are not captured [20]. For example, α -1 can be transformed to α -2 by a lattice invariant shear (with Burgers vector $[10]_{\alpha}$ on slip plane (01)) (Fig. 2(a)). Since α -1 and α -2 are identical structural states of the α phase, there also exist two identical structural states of the martensite (β -3 and β -4) associated with α -2 (Fig. 2(a)). Even without any energetic consideration, by simply connecting the two local energy landscapes, we find that the structural states of β -2 and β -3 are close to each other (Fig. 2(b)) and the energy barrier between them should be low especially when the transformation strain is approaching 0.5 (when the transformation strain is exactly 0.5, the structural states β -2 and β -3 become identical – centered rectangle). Here a SDNPTP can be defined between β -2 and β -3. Furthermore, if the system is under an applied shear stress, the transition from β -2 to β -3 could become barrierless, or the double-well energy landscape (Fig. 2(a)) could become a single-well (Fig. 2(b)). As a consequence, if starting from α -1 and going through a transformation cycle, both α -1 and α -2 become accessible

and a dislocation with Burgers vector $[10]_\alpha$ can be generated after the $\alpha-1 \rightarrow \beta-2 \rightarrow \beta-3 \rightarrow \alpha-2$ transformation cycle. The dislocation represents a lattice invariant shear connecting $\alpha-1$ with $\alpha-2$. During this process, the activation of the SDNPTP between $\beta-2$ and $\beta-3$, either by an external or internal stress arising from the structural transformation, is critical. Note that from the crystallographic point of view, the deformation provided by the SDNPTP is fundamentally different from that due to general dislocation plasticity. As shown in Fig. 2(a), the deformation pathway through SDNPTP (from $\beta-2$ to $\beta-3$) connects two states associated with the point symmetry only (e.g., a vertical mirror), while that through dislocation plasticity (e.g., $\beta-1$ and $\beta-3$) connects two states associated with the translational symmetry (e.g., a lattice invariant shear).

It is important to note that there is a high symmetry structural state between $\beta-2$ and $\beta-3$ (Fig. 2(a)) when the shear strain is exactly 0.5 (a centered rectangular structure with point group $2mm$), which is located on the boundary between the two EPNs (referred to as CR structure in Fig. 2(a)) and its energy corresponds to an extreme on the free energy landscape (in general a maximum or saddle point, since it has relatively high symmetry in its vicinity). Let us refer to this unstable high symmetry structural state as “X-structure”. Since the crystallographic equivalence between $\beta-2$ and $\beta-3$ originate from the symmetry operations in the “X-structure”, the “X-structure” inevitably exists between the two states, which is critical for the activation of the SDNPTPs. In terms of dislocation plasticity, the generation of dislocations requires a localized shear as large as b/d (from $\alpha-1$ to $\alpha-2$), e.g., a shear strain over 100%. With the assistance of the MT that is driven by undercooling, however, this large deformation can be realized partially through forward and backward MTs (e.g., from $\alpha-1$ to $\beta-2$ and from $\beta-3$ to $\alpha-2$). The remaining part of the deformation is realized through the SDNPTP (from $\beta-2$ to $\beta-3$), which can be activated easily by a stress that is much smaller than what would be required for direct nucleation of a shear loop because of the relatively small amount of deformation along the SDNPTP. This leads to an easy path of defect generation. In general, the closer the martensite structure is to the “X-structure” (e.g., comparing $\beta-2$ and the CR structure in Fig. 2(a)), the lower is the activation energy barrier along the SDNPTP, and the barrier asymptotically approaches zero when the martensitic structure is exactly the “X-structure”.

3. Case study: NiTi system

3.1 Defect generation during transformation cycling in NiTi

In order to derive a rigorous and general approach to analyzing the crystallographic origin of FF, we chose the B2 to B19' MT in NiTi as an example because (a) NiTi is the most widely used commercial SMA system and (b) rich experimental and theoretical data are available in the literature. We first propose a possible SDNPTP, followed by the

prediction of characteristic defect structures. These defect structures are then compared with experimental observations.

Even though it is not straightforward to determine whether there is an “X-structure” close to the B19’ martensite structure, several theoretical and experimental studies suggest that one may exist. Based on *ab initio* calculations, for example, it has been reported [26] that the B19’ structure is both geometrically (~10% shear strain) and energetically close to a base-centered orthorhombic (BCO) structure, which has relatively high symmetry in its vicinity (compared with its nearby monoclinic structures). As shown in Fig. 3(a), a monoclinic cell can be drawn in a B2 lattice, which can be deformed uniformly into a BCO unit cell. During the backward deformation from BCO to B2, there are two crystallographically equivalent B2 states, which are labelled as initial B2 and new B2, respectively, in Fig. 3. In addition, according to the neutron diffraction study [27], plastic deformation of the B19’ phase in NiTi can be induced by a compressive stress as low as 33 MPa, with inelastic deformation at small strains/stresses in NiTi confirmed in both tension and compression by Stebner, et al. and most likely due to slip or mode D twinning [28]. These easy modes of deformation suggest low activation energy along a possible SDNPTP that could be easily activated during the phase transformation. Similar pathways originating from the geometrical closeness between the monoclinic and BCO structures have also been proposed in the MT in U-Nb [29].

To further validate the existence of the SDNPTP, characteristic defect structures generated along this pathway are analyzed through the crystallographic theory of structural transformations [24, 30-32] and compared with experimental observations. From a methodological point of view, even though no direct evidence about the deformation pathway could be obtained from experiments, resulting domain and defect structures could be taken as the characteristic markers of the pathways. This is similar to the classical example of validating the Bain path during the FCC to BCC transformation through the phenomenological theory of martensite crystallography (PTMC) [30-32].

Because the types of defects induced by the transformation are dictated by the broken symmetry along all the deformation pathways, both the PTPs (between B2 and B19’) and SDNPTPs (between two B19’ states) need to be taken into account in analyzing the defect structures observed in experiments. The following analysis is similar to the crystallographic analysis in PTMC [24, 30-32], which predicts habit planes and orientation relationship (OR) between austenite and martensite. However, it should be noted that the initial PTMC is limited to analyzing PTPs and the associated martensite domain structures. Using a similar mathematical approach, we analyze SDNPTPs and the associated crystalline defect structures. As shown in Fig. 3(b), if the {011} cross-section of a B2 structure is considered, a simple rectangular structure in 2D can be found. Note

that only one set of sub-lattice atoms (blue ones) are shown in Fig. 3(b) for simplicity. The BCO structure, which can be taken as a centered rectangular structure in 2D, is formed through a shear along the $[100]$ direction of B2. In this sense, the deformation pathway between B2 and BCO is exactly parallel to the 2D example discussed above (Fig. 2). Note that the energy landscape in Fig. 3(b) shows only the pathway leading to one BCO state. Because there are six crystallographic equivalent $\{011\}$ planes and the shear direction $[100]$ could be either positive or negative, there are twelve total BCO states around each B2, which is consistent with the twelve correspondence martensitic variants (or deformation variants [25]) determined for NiTi in the literature. Considering a transformation cycle from B2 to B19' and back to B2, a new B2 state could be generated if the SDNPTP between the two B19' states is activated. Considering the deformation between the two B2 structural states, we can choose two sets of corresponding lattice vectors in the two B2 structures (initial B2 and new B2 in Fig. 3):

$$\mathbf{e}_1^{initial} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{e}_2^{initial} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{e}_3^{initial} = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix} \quad (1a)$$

$$\mathbf{e}_1^{new} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{e}_2^{new} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{e}_3^{new} = \begin{bmatrix} -1 \\ -2 \\ 2 \end{bmatrix} \quad (1b)$$

And the deformation gradient matrix can be determined by solving:

$$\mathbf{e}_i^{new} = \mathbf{T}_1 \mathbf{e}_i^{initial}, \quad i = 1 \sim 3 \quad (2)$$

where the initial B2 state is taken as the reference. The deformation gradient matrix for the new B2 state with respect to the initial B2 is:

$$\mathbf{T}_1 = \begin{bmatrix} 1 & 0.5 & -0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

Considering two neighboring B2 domains, one of which is in the initial B2 state while the other is the new B2 state, their boundary is constrained by a geometrical compatibility condition [24, 30-32],

$$\mathbf{Q}\mathbf{T}_1 - \mathbf{I} = \mathbf{a} \otimes \mathbf{n} \quad (4)$$

where $\mathbf{Q}$ is a rigid body rotation, and $\mathbf{a}$ and $\mathbf{n}$ are the shear vector and boundary plane (also shear plane) normal, respectively. Two solutions can be obtained,

$$\left\{ \begin{array}{l} \mathbf{a}_1 = \frac{1}{2} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \\ \mathbf{n}_1 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \end{array} \right\} \quad \left\{ \begin{array}{l} \mathbf{a}_2 = \frac{1}{18} \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix} \\ \mathbf{n}_2 = \begin{pmatrix} -4 \\ -1 \\ 1 \end{pmatrix} \end{array} \right\} \quad (5)$$

The first solution describes a $[100](011)$ dislocation, which is the typical type of lattice dislocations observed in B2 NiTi. The second solution suggests a relatively high index plane $\{411\}$. By calculating the rotation $\mathbf{Q}$ corresponding to the second solution, it is found that $\mathbf{Q}$ is a 38.94° rotation around the $[011]$ direction, which is the characteristic angle of a $\Sigma 9$ boundary in BCC structures. The formation of $\Sigma 9$ boundary is illustrated in Fig. 4 (the (011) cross-section of B2 is considered). Before the transformation, there is a B2 perfect crystal without any defect (Fig. 4(a)). After the transformation, a compound twin of B19' martensite can be formed with twin boundary $(100)_{\text{B19}'}$ (corresponding to $(100)_{\text{B2}}$, Fig. 4(b)). If only the center part (colored by green) goes back to the initial B2 state, while the upper and lower parts go to two new B2 states (by nucleation and growth during the B19' to B2 inverse MT), three different B2 domains will appear after the transformation cycle. In Fig. 4(c), the red domain corresponds to the deformation of $\mathbf{T}_1$, while the blue domain corresponds to $\mathbf{T}_2$ (Eqns. (3) & (7a)). It can be found that both the domain boundaries (green/red, green/blue) are exactly $\Sigma 9$ boundaries, which indicates a misorientation of 38.94° .

Because the above two solutions originate from the same deformation mode (i.e., the same deformation gradient in Eqn. (3)), they offer the same amount of strain. Thus during transformation cycling generation of both dislocations and grain boundaries should be expected and both contribute to the total irrecoverable strain. However, since the crystallographic characters of dislocations and grain boundaries are different, the total amount of irrecoverable strain should depend on the density of generated dislocations (solution 1) and grain boundaries (solution 2).

As a direct extension, similar calculations can be applied to determine defect structures after multiple transformation cycles. Starting from one initial B2 state, there are twelve crystallographically equivalent B19' variants that may lead to twelve new B2 states after one transformation cycle, the defect structures formed on the boundaries between two new B2 states can be determined through [24, 30-32]

$$\mathbf{Q}\mathbf{T}_j - \mathbf{T}_i = \mathbf{a} \otimes \mathbf{n} \quad i, j = 1 \sim 12 \quad (6)$$

For example, we consider the deformation gradient matrix for another two new B2 states ($\mathbf{T}_2$ and $\mathbf{T}_3$) after one cycle (crystallographically equivalent form as $\mathbf{T}_1$), and another two new B2 states ($\mathbf{T}_{21}$ and $\mathbf{T}_{22}$) after two cycles.

$$\mathbf{T}_2 = \begin{bmatrix} 1 & -0.5 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (7a)$$

$$\mathbf{T}_3 = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (7b)$$

$$\mathbf{T}_{21} = \begin{bmatrix} 1 & 0.5 & -0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & -0.5 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1.25 & 0.5 & -0.75 \\ 0.5 & 1 & -0.5 \\ 0 & 0 & 1 \end{bmatrix} \quad (7c)$$

$$\mathbf{T}_{22} = \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0.5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1.25 & 0.75 \\ 0 & 0 & 1 \end{bmatrix} \quad (7d)$$

By using $\mathbf{T}_2$ and $\mathbf{T}_1$ in Eqn. (6), a $\Sigma 3$ boundary can be obtained (Fig. 4(d)). If $\mathbf{T}_3$ and $\mathbf{T}_1$ are inserted in Eqn. (6), a $\Sigma 5$ boundary can be obtained. Similarly, it can be found that $\Sigma 9$, $\Sigma 5$, $\Sigma 3$, $\Sigma 7$ and $\Sigma 13$ boundaries can be generated after one cycle, while higher order special grain boundaries such as $\Sigma 11$ (using $\mathbf{T}_{21}$ and $\mathbf{T}_{22}$ in Eqn. (6)), $\Sigma 15$, $\Sigma 17$, ..., $\Sigma 37$, $\Sigma 39$, ... can be generated after two and more successive cycles.

3.2 Experimental observations of defects generation during transformation cycling in NiTi

As shown in previous theoretical analysis, various kinds of defects structure including dislocations and grain boundaries can be generated during transformation cycling, and the evolution of these defects will play a role in subsequent cycles. In order to further understand the physical origin of FF and reveal the defect generation mechanism during transformation cycling, we examined experimentally a number of samples of NiTi under different conditions by using several characterization techniques. Three typical experimental observations are shown in the followings. The domain structure and special grain boundary in a sample after 2 load-biased (150 MPa) thermal cycles are presented first, which can be directly compared with previous crystallographic predictions. Then the dislocation structure generated after 20 load-biased (100 MPa) cycles are shown, followed by the observation of a sample after 20438 load-bias (100 MPa) cycles.

Crystallographic orientation mapping was performed using Transmission Kikuchi Diffraction (TKD) methods [33-35] on a FEI XL-30 ESEM scanning electron microscope with an EDAX Hikari camera. A custom sample holder was used to hold focused ion beam (FIB) prepared transmission electron microscope (TEM) foils at -20° tilt. Data were collected at a 4 mm working distance using 30 kV and a 100 μm aperture. Corresponding microstructural analysis was performed using scanning transmission electron microscopy on a FEI Tecnai F-20. Unlike prior work using ASTAR TEM-OIM systems [19], TKD allows for the sampling of much larger scan areas (sometimes 50x) which allow for more representative orientation maps at comparable resolutions.

3.2.1. TKD results of domain structures and grain boundaries generated after 2 cycles

Orientation mapping of the 2-cycle 150 MPa sample showed a variety of low angle, high angle, and special boundaries. A portion of a TKD map containing sub-grain boundaries with 60° misorientations is shown in Fig. 5(a). The double headed arrow represents the region chosen for a misorientation line profile shown in Fig. 5(b) alongside. From the grain orientations, we can determine that the arrow crosses two $\Sigma 3$ boundaries, giving rise to 60° misorientations.

From the OIM analysis, the orientation information, $(hkl)\langle uvw \rangle$, of grain A and B is determined as the following:

$$\text{A: } (0\ 1\ 2)\ [18\ -10\ 5]$$

$$\text{B: } (-1\ -4\ 8)\ [24\ 4\ 5].$$

By taking $[100]_A$, $[010]_A$, $[001]_A$ (crystal directions in grain A) as x, y, z axes of a new Cartesian coordinate, grain B can be regarded as a 60.98° rotation along $[0.476, -0.619, -0.650]$ direction. This rotation axis is deviated from $[1\ \bar{1}\ \bar{1}]$ by 6.88° (perfect $\Sigma 3$ boundary is a 60° rotation along $[1\ \bar{1}\ \bar{1}]$). According to the cubic symmetry of the B2 lattice, the misorientation for $\Sigma 3$ could be described equivalently in three ways, which leads to the possible deformation gradient matrices for the two neighboring grain A and B (Table 1).

Table 1. Comparison of possible deformation modes for grain A and B

	Rotation Q	Deformation gradient T_A/T_B	Interaction energy ($10^6\ \text{J/m}^3$)
(i)	70.5° rotation along $[101]$	$\begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & -0.5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -0.5 & 1 & 0.5 \\ 0 & 0 & 1 \end{bmatrix}$	-45.0 15.0

(ii)	70.5° rotation along $[01\bar{1}]$;	$\begin{bmatrix} 1 & 0.5 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -0.5 & -0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	-33.75 -33.75
(iii)	70.5° rotation along $[110]$;	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0.5 & -0.5 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -0.5 & 0.5 & 1 \end{bmatrix}$	-33.75 26.26

Note both grain A and grain B should be energetically favorable under the applied load, which suggests that the interaction energy with the applied load should be negative:

$$E^{\text{int}} = \sigma^{\text{load}} : \varepsilon_i \quad (8)$$

where $:$ is the double dot product and ε_i is the Lagrangian strain for each deformation mode,

$$\varepsilon_i = \frac{T_i^T T_i - I}{2} \quad (9)$$

Note that the loading direction is the normal direction of Fig. 5(a), so that it is along $[012]$ in the new coordinates. The tensor form of the applied load in the new coordinate is,

$$\sigma^{\text{load}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 30 & 60 \\ 0 & 60 & 120 \end{bmatrix} \text{MPa} \quad (10)$$

As listed in Table 1, only case (ii) can provides two negative values of the interaction energy, so that formation of the $\Sigma 3$ boundary in Fig. 5 by T_A and T_B is uniquely determined by the crystallographic information provided by the TKD results, as illustrated in Fig. 4(d).

3.2.2. TEM results of dislocations generated after 20 cycles

A typical dislocation structure generated after 20 load-biased (100 MPa) cycles is shown in Fig. 6. As indicated by Eqns. (4) & (5), the two types of defects, i.e., dislocation and grain boundary, are degenerate solutions, so that both of them should be expected in experimental observations. In Fig. 7, a complex dislocation structure next to a triple point in the original austenite grain structure is shown. The lower magnification image (Fig. 7(b)) show how the structure changes when tilting, indicating such a triple point could be a location with new boundary content. As revealed in the literature, dislocation density increases dramatically with the number of cycling, especially under a biased-load, which

strongly suggests dislocation generation mechanism occurring at each cycle [9, 10, 14]. Meanwhile, dislocation multiplication and evolution would lead to complex dislocation substructures (Fig. 6). Since there are plenty of experimental results reported in the literature regarding transformation induced dislocations [9-16], we will not further discuss it in this work.

3.2.3. TKD results of multi-domain structure after more than 20,000 cycles

More complicated domain structures are generated after 20,438 cycles. In Fig. 8, significant sub-grain formation within the original B2 grains can be observed. Most of the boundaries among sub-grains are general grain boundaries rather than special grain boundaries, which is in contrast different from the observed grain structure after a small number of cycles (e.g., Fig. 5). There could be two reasons for such a change: (1) our previous theoretical analysis shows that higher order special grain boundaries will be generated with more number of cycles, which would be geometrically close to general grain boundaries; (2) similar to the formation of dislocation cells and sub-grains during plastic deformation, the dislocation structures tend to evolve into an energetically favorable state, which is not only related to the dislocation density but also the direction of applied stress [36, 37]. Theoretical analysis of the formation of sub-grains after a large number of cycles is beyond the scope of the purely crystallographic approach presented in this paper, and further investigations will be performed in future work.

4. General design strategies for FF resistant shape memory alloys

Based on the crystallographic origin of defect generation identified above, we discuss qualitatively parameters that would play key roles in controlling FF and that should be considered in the development of FF-resistant SMAs, which include (a) activation energy of the SDNPTP, (b) internal stress state, and (c) external applied load. With the ever-advancing capabilities of computational and experimental techniques, such analyses will become quantitative. The first parameter is an intrinsic material parameter dictated by the nature of the MT and determined by crystal structures and lattice parameters of the austenite and martensite and their lattice correspondence. As shown schematically in Fig. 2(b), the activation energy can be characterized by a critical stress, σ^c , required to make the transition between β -2 and β -3 barrierless, which is equivalent to the maximum slope along the SDNPTP on the free energy surface between the two states. This critical stress acts in parallel to the critical resolved shear stress introduced in plastic yielding of crystals. The internal stress state is an extrinsic parameter because it is associated with microstructural inhomogeneity in a system, which is determined not only by the nature of the MT (e.g., self-accommodation) but also by the thermomechanical processing history (grain size, texture, dislocation substructure and residual stress). The external applied load on the sample is self-explanatory.

If we use the maximum internal stress, σ^i , to represent the second parameter and σ^a for the external load, then the sufficient and necessary condition for the occurrence of defect generation and thus onset of FF can be formulated as

$$\sigma^c < \sigma^i + \sigma^a \quad (11)$$

which indicates that if the critical activation stress is smaller than the sum of internal and external stresses, FF will happen inevitably. The above inequality also represents an interplay between the intrinsic and extrinsic parameters because σ^c is the intrinsic resistance to FF while $\sigma^i + \sigma^a$ are the extrinsic driving force for FF and the balance between them determines whether FF happens or not.

Note that the above concept for the inequality is simple and parallels the criterion for plastic yielding of crystals, sharing the same nature as that of the critical resolved shear stress to initiate plastic yielding. In this case, σ^c is a critical stress for the activation of the SDNPTP associated with point symmetry as discussed above. Both quantities are intrinsic material properties. However, as shown in the NiTi system, deformation through the SDNPTP is both geometrically and energetically favored compared with plastic yielding (e.g., the SDNPTP between two nearby B19' states requires ~20% shear strain while dislocation involves ~100% shear strain). The introduction of σ^i in the above inequality is necessary in analyzing both plastic yielding and activation of SDNPTPs of SMAs since local stresses depend strongly on the self-accommodation of multi-domain structures of martensite.

It is important to note that the FF resistance of an SMA is determined not only by the yield strength of the austenite and martensite crystals, but also by σ^c , which is determined by crystal symmetry [20]. For example, if the symmetry groups of the parent and product phases cannot be included in a common finite group [20], $\sigma^c=0$ and the inequality is always satisfied, which means that defect generation always occurs even with small values of σ^i and σ^a . On the other hand, if the strict geometrical constraint, i.e., the cofactor condition [21, 22], is satisfied, σ^i could be reduced greatly and FF should not be expected without applied stress, which is consistent with the experimental observation of enhanced reversibility in Zn-Au-Cu and Ni-Ti-Cu systems [21, 22, 38]. However, there is no guarantee of good fatigue resistance under applied load unless the intrinsic material parameter σ^c is relatively large. As a result, any design criterions formulated for high performance SMAs have to consider both σ^c and σ^i . Previous treatment in the literature considers these independently with the group theory criterion [20] (related to σ^c) and the cofactor condition criterion [21, 22] related to σ^i , without accounting for the interplay

between the two mechanisms. That is also why we clearly distinguish them as “intrinsic” and “extrinsic” and define two parameters with the same units (stress), which makes their interplay mathematically tractable.

Theoretically, σ^c can be determined through atomistic calculations using, e.g., the *ab initio* method [39], and σ^i can be determined through mesoscopic and macroscopic stress field calculations that account for formation of multi-variant domain structures during MTs in a single crystal or polycrystalline specimen, such as the phase field microelasticity method and finite element method [40, 41]. In addition, if $\sigma^c > \sigma^i$, the critical applied stress (σ^{ac}) to trigger FF is:

$$\sigma^{ac} = \sigma^c - \sigma^i \quad (12)$$

where σ^{ac} is the critical applied stress under which functional fatigue of the SMA starts. Note that in the above discussion the stress parameters are in scalar forms for simplicity. The generalization to tensor forms is required for detailed applications since their values may depend strongly on crystallographic directions, which is straightforward.

The balance between functionality and durability of SMAs can be represented by the parameter of maximum work output that can be determined by:

$$w^{\max} = \sigma^{ac} \cdot \varepsilon^r = (\sigma^c - \sigma^i) \cdot \varepsilon^r \quad (13)$$

where ε^r is the recoverable strain. Work output is one of the most important parameters considered in SMA design since it reflects both FF resistance and functionality (i.e., recoverable strain). According to Eqn. (13), there are three ways to increase $w^{\max}$: (i) increasing σ^c , (ii) increasing ε^r and (iii) reducing σ^i .

Based on the physical origin of FF presented above, strategies to increase σ^c can be proposed. From a crystal symmetry point of view, the SDNPTP leading to defect generation connects two B19' states, which are degenerate states with respect to the high symmetry BCO structure. Such degeneracy is originated from the translational symmetry of the crystals. Thus in order to prevent the activation of SDNPTPs and improve FF resistance, the key is to break the translational symmetry as well as the degeneracy of states that are connected by SDNPTPs. Two typical mechanisms have already been proven effective: introducing long range atomic ordering and/or nano-particles in the matrix phase [42-55]. The effect of long range atomic ordering is illustrated in Fig. 9(a) through a 2D example. After long range ordering is introduced, the states of β -2 and β -3 are not degenerate any more (in Fig. 9(a), the distance between the two green atoms are

closer in the β -3 state as compared to that in the β -2 state) and the transition from β -2 and β -3 becomes energetically unfavorable. As a result, a larger σ^c and better FF resistance are expected. As a matter of fact, most commercial SMA systems are based on ordered austenite structures, which strongly suggest the importance of atomic ordering on FF resistance. Of course, dislocations are relatively difficult to operate as well in ordered intermetallics as compared to their disordered solid solution counterparts. It should be emphasized that it is not accurate to claim that atomic ordering is a necessary condition for good FF resistance in SMAs, even though most commercialized SMAs are based on ordered austenite as suggested in the literature [37-39, 51]. Instead it is important to find a system with both good fatigue resistance and relatively large recoverable strain. Similar to atomic ordering, nano-precipitates in austenite [38, 45-55] will also break the degeneracy of β -2 and β -3 as illustrated schematically in Fig. 9(b), because of the geometrical constraint and local stress concentration caused by the non-transformable precipitates. Of course, nano-precipitates will also strengthen the austenite effectively.

Note that even though both means are effective in terms of enlarging σ^c , they could have different influences on ε^r and σ^i . For example, if the lattice parameters are not strongly dependent on atomic ordering, then the change of ε^r and σ^i could be negligible and the benefit of an enlarged σ^c in improving FF resistance will be maintained. In contrast, the drawback of introducing nano-precipitates is that the precipitates may also make the PTPs between austenite and martensite more difficult to activate and reduces the overall recoverable strain.

Theoretically, the recoverable strain, ε^r , is bounded by the transformation strain, which is determined entirely by the lattice parameters of and lattice correspondence between the austenite and martensite. In a well-aligned single crystal, the uniaxial recoverable strain could approach the transformation strain [24, 45]. For polycrystals or with the presence of precipitates, ε^r will be reduced. For NiTi, the recoverable strain is $\sim 10\%$ in certain single crystal orientations (close to the magnitude of the transformation strain, 10.7%), while it is usually less than 5% in polycrystals [24, 45]. Creation of appropriate texture in austenite grains may help to maximize the recoverable strain.

Even though numerical calculations are required to determine σ^i for a specific system, several general principles can be utilized to make qualitative estimations. For example, σ^i should depend strongly on the volume change during MT. As shown in the literature [24], most SMA systems have a volume change smaller than 1%. Otherwise σ^i would become larger than σ^c , leading to irreversible defect generation. In addition, it is necessary to have enough martensitic variants as fundamental building blocks for self-accommodation, especially in polycrystals. For example, if the crystal structure of austenite is cubic, the

martensite could be orthorhombic (6 variants), monoclinic (12 variants) or triclinic (24 variants), which satisfies the theoretical requirement of self-accommodation in polycrystals [24]. Besides the above criteria, σ^i can be further reduced if the lattice parameters of the parent and product phases satisfy some special geometrical constraints [22, 24, 30-32]. For example, σ^i could be reduced greatly in ZnAuCu system, where the lattice parameters of the austenite and martensite satisfy the co-factor condition. Note that even though a strict geometrical condition could make σ^i approach zero, it cannot guarantee a high critical applied stress (σ^{ac}) to activate FF because σ^c could also be small, which is an intrinsic material property independent of the self-accommodation mechanism.

Based on the above discussions, the reversibility and possible shape memory properties of several well-known systems can be qualitatively predicted according to Eqns. (11) ~ (13). For iron and iron-based alloys undergoing the FCC to BCC/BCT MT, σ^c is small (σ^c approaches zero as the tetragonality of BCT approaches 1) while σ^i is large (the volume change is $\sim 2\%$), which results in the generation of significant amount of crystalline defects during the MT. For Ni-Al and Ni-Mn systems undergoing the cubic to tetragonal MT, σ^i is large in polycrystals because of the limited number of martensitic variants (only 3 variants). Comparing the Cu-Al-Ni system with the Au-Cd system undergoing the cubic to orthorhombic transformation, the functionality of the former is better because of a larger ε^r . The NiTi system has a reasonable balance among all the above parameters, which makes it one of the most widely used commercial SMAs.

In the literature, extensive experimental efforts have been made to test the properties of a large number of SMAs, including the alloy systems mentioned above, and several rules of thumb have been proposed for the design of SMAs [24], which seem to agree with our theoretical analysis based on Eqn. (11) and Eqn. (13) presented above. For example, a promising SMA system usually requires (1) atomic ordering in austenite and (2) twinning substructure in martensite. In addition, a systematic and quantitative evaluation for potential SMA systems can be established if the database of necessary parameters (e.g., critical stress to activate SDNPTP, maximum internal stress, maximum recoverable strain) is available, which could accelerate and reduce the cost of design and development of SMAs in the future.

5. Design strategies for Ni-Ti-Pt, Ni-Ti-Cu and Ni-Ti-Au SMAs

The general SMA design strategies outlined above are applied to Ni-Ti-based ternary alloy systems that undergo B2 to B19 MT. Two typical alloy systems, Ni-Ti-Pt and Ni-Ti-Cu, have been widely investigated in the literature. Even though both of the systems undergo the same cubic to orthorhombic transformation, from a crystal symmetry point

of view, their properties are found to be quite different. In terms of functionality (i.e., recoverable strain), Ni-Ti-Pt is superior to Ni-Ti-Cu, while in terms of durability (i.e., FF resistance), Ni-Ti-Cu is much better than Ni-Ti-Pt [21, 38]. Since the maximization of work output is usually the main concern, it is critical to find a way to balance the recoverable strain and FF resistance.

Even though the exact values of σ^c , ε^r and σ^i for both Ni-Ti-Pt and Ni-Ti-Cu systems are currently unavailable (as they require theoretical calculations as described in Section 4), some qualitative conclusions can still be made based on the crystal structures and lattice parameters of the austenite and martensite in these systems. For example, based on the lattice parameters of the B2 and B19 phases in Ni-Ti-Pt and Ni-Ti-Cu systems listed in Table 2 [56, 57], the principal values of the transformation strain found in Ni-Ti-Pt (-11%, 2%, 8%) are much larger than those found in Ni-Ti-Cu (-5%, 0, 5%). Since the transformation strain is the upper limit of recoverable strain, it could be expected that $\varepsilon_{Ni-Ti-Pt}^r > \varepsilon_{Ni-Ti-Cu}^r$. Furthermore, the crystal structure of B19 martensite in Ni-Ti-Pt is geometrically close to an “HCP” structure. Similar to the analysis of the Ni-Ti binary system, the “HCP” structure can be proposed as the high symmetry X-structure in Ni-Ti-Pt and Ni-Ti-Cu (similar to the role of BCO in NiTi), and the SDNPTP related to this structure is parallel to the well-known Burgers path proposed for the BCC to HCP transformation. As a result, the difficulty of SDNPTP activation and hence the FF resistance can be indicated by the geometrical closeness between the B19 and the “HCP” X-structure. Because of the b/a and c/a ratios of martensite in Ni-Ti-Pt are 1.71 and 1.61, respectively, which are close to those in a perfect HCP structure (1.732 and 1.633), while they are 1.61 and 1.49 in Ni-Ti-Cu, it should be expected that the SDNPTP in Ni-Ti-Pt is easier to activate than that in Ni-Ti-Cu, i.e., $\sigma_{Ni-Ti-Pt}^c < \sigma_{Ni-Ti-Cu}^c$. In addition, because one of the principal transformation strains is zero in Ni-Ti-Cu, which facilitates the self-accommodation of martensitic variants, the internal stress concentration should also be smaller in Ni-Ti-Cu, thus $\sigma_{Ni-Ti-Pt}^i > \sigma_{Ni-Ti-Cu}^i$. According to these arguments, a much better reversibility and FF resistance should be expected for Ni-Ti-Cu in contrast to Ni-Ti-Pt, which is consistent with experimental observations.

Following the above design criteria for SMAs with balanced properties, we propose focusing on a relatively unstudied alloy system, Ni-Ti-Au. The lattice parameters for the Ni-Ti-Au system are presented in Table 2 and compared with those of Ni-Ti-Pt and Ni-Ti-Cu systems. By calculating the transformation strain, it is found that $\varepsilon_{Ni-Ti-Pt}^r \sim \varepsilon_{Ni-Ti-Au}^r > \varepsilon_{Ni-Ti-Cu}^r$, which suggests that the system with Au could have a potential functionality as good as that of Ni-Ti-Pt, which is better than that of Ni-Ti-Cu. In the Ni-Ti-Au system, the b/a and c/a ratios are 1.66 and 1.57, respectively, which is almost in the middle of the previous numbers for Ni-Ti-Pt and Ni-Ti-Cu. As a result, it

can be expected that $\sigma_{Ni-Ti-Pt}^{act} < \sigma_{Ni-Ti-Au}^{act} < \sigma_{Ni-Ti-Cu}^{act}$. Therefore, Ni-Ti-Au alloys should have a functionality as good as that of Ni-Ti-Pt with better FF resistance, in addition to its significantly higher temperature capability than Ni-Ti-Cu [45].

Table 2. Crystallographic comparisons between Ni-Ti-Pt, Ni-Ti-Cu and Ni-Ti-Au systems

	Ni-Ti-Pt	Ni-Ti-Cu	Ni-Ti-Au
a_B2 (nm)	0.308	0.303	0.322
a_B19	0.275	0.288	0.294
b_B19 (nm)	0.470	0.451	0.490
c_B19	0.444	0.428	0.463
Transformation strain (principal value)	-0.11 0.02 0.08	-0.05 0.00 0.05	-0.09 0.02 0.08
b_B19/a_B19	1.71	1.61	1.66
c_B19/a_B19	1.61	1.49	1.57

6. Summary

The physical origin of functional fatigue (FF) in shape memory alloys (SMAs) is analyzed from the crystallographic point of view and defect generation mechanisms in several typical SMA systems, including NiTi binary and Ni-Ti-based ternary systems, are investigated through a combination of theoretical and experimental studies. The major findings are:

- A symmetry-dictated non-phase-transformation pathway (SDNPTP) is identified, whose activation leads to defect generation resulting in the loss of dimensional stability during transformation cycling, thus could be responsible for FF in SMAs.
- The SDNPTP is easier to activate if a high-symmetry structure is crystallographically close to the martensitic structure.
- FF in NiTi is most likely to be associated with the activation of a SDNPTP through a base-centered orthorhombic structure in between two corresponding variants of the B19' martensite, resulting in the generation of B2 matrix dislocations and Σ grain boundaries, which have been observed experimentally.
- A critical stress is proposed to characterize the activation of SDNPTPs, which is parallel to the criteria for critical resolved shear stress for yielding. FF occurs if the local stress in the system determined by the sum of the internal (residual) and external (applied) stress exceeds this critical stress during the martensitic transformation.

Strategies to improve FF resistance and systematic ways to develop SMAs having balanced properties through crystallographic design and transformation pathway engineering are proposed and used to evaluate Ni-Ti-Pt, Ni-Ti-Cu and Ni-Ti-Au SMA systems.

Acknowledgement:

The work was supported by the Department of Energy Grant No. DE-SC0001258 and Natural Science Foundation Grant No. DMR-141032. RDN gratefully acknowledges support from the NASA Transformative Aeronautics Concepts Program (TACP), Transformational Tools & Technologies Project.

Reference:

- [1] W. J. Buehler WJ, J. V. Gilfrich, R. C. Wiley. Effect of Low-Temperature Phase Changes on the Mechanical Properties of Alloys near Composition TiNi. *J. Appl. Phys.* 34 (1963) 1475-1477.
- [2] T. W. Duerig. The use of superelasticity in modern medicine. *MRS Bulletin* 27 (2002), 101-104.
- [3] J. San Juan, M. L. No, C. A. Schuh. Superelasticity and shape memory in micro- and nanometer-scale pillars. *Adv. Mater* 20 (2008), 272-278.
- [4] C. Mavroidis. Development of advanced actuators using shape memory alloys and electrorheological fluids. *Res. Nondestr. Eval.* 14(2002), 1-32.
- [5] Y. Q. Fu, J. K. Luo, M. Hu, H. J. Du, A. J. Flewitt, W. I. Milne. Micromirror structure actuated by TiNi shape memory thin films. *J Micromech. Microeng* 15 (2005), 1872-1877,
- [6] G. Eggeler, E. Hornbogen, A. Yawny, A. Heckmann, M. Wagner. Structural and functional fatigue of NiTi shape memory alloys. *Mater. Sci. Eng. A* 378 (2004), 24-33.
- [7] S. V. Raj SV & R. D. Noebe. Low temperature creep of hot-extruded near-stoichiometric NiTi shape memory alloy part II: Effect of thermal cycling. *Mater Sci. Eng. A* 581 (2013), 154-163.
- [8] R. Hamilton, H. Sehitoglu, Y. Chumlyakov, H. Maier. Stress dependence of the hysteresis in single crystal NiTi alloys. *Acta Mater.* 52 (2004), 3383-3402.
- [9] 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 (2009), 3549-3561.
- [10] T. Simon, A. Kroger, C. Somsen, A. Dlouhy, G. Eggeler. On the multiplication of dislocations during martensitic transformations in NiTi shape memory alloys. *Acta Mater.* 58 (2010), 1850-1860.
- [11] S. Miyazaki, Y. Igo, K. Otsuka. Effect of thermal cycling on the transformation temperatures of Ti-Ni alloys. *Acta Metall. Mater.* 34 (1986), 2045-2051.
- [12] A. Yawny, M. Sade, G. Eggeler. Pseudoelastic cycling of ultra-fine-grained NiTi shape-memory wires. *Zeitschrift fur Metallkunde* 96 (2005), 608-618.
- [13] G. S. Bigelow, A. Gary, S. A. Padula II, D. J. Gaydosh, R. D. Noebe. Load-biased shape-memory and superelastic properties of a precipitation strengthened high-temperature Ni_{50.3}Ti_{29.7}Hf₂₀ alloy. *Scripta Mater.* 64 (2011), 725-728.
- [14] 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.
- [15] M. L. Bowers, X. Chen, M. De Graef, P. M. Anderson, M. J. Mills. Characterization and modeling of defects generated in pseudoelastically deformed NiTi microcrystals. *Scripta Mater.* 78 (2014), 69-72.

- [16] C. Grossmann, J. Frenzel, V. Sampath, T. Depka, G. Eggeler. Elementary transformation and deformation processes and the cyclic stability of NiTi and NiTiCu shape memory spring actuators. *Metall. Mater. Trans. A* 40A (2009), 2530-2544.
- [17] O. Benafan, R. D. Noebe, S. A. Padula II, D. W. Brown, S. Vogel, R. Vaidyanathan. Thermomechanical cycling of a NiTi shape memory alloy-macroscopic response and microstructural evolution. *Int. J. Plast.* 56 (2014), 99-118.
- [18] O. Benafan, R. D. Noebe, S. A. Padula II, D. W. Brown, S. Vogel, R. Vaidyanathan. Temperature dependent deformation of the B2 austenite phase of a NiTi shape memory alloy. *Int. J. Plast.* 2013; 51:103-121.
- [19] M. L. Bowers, Y. Gao, L. Yang, D. J. Gaydos, M. De Graef, R. D. Noebe, Y. Wang, M. J. Mills. Austenite grain refinement during load-biased thermal cycling of a Ni_{49.9}Ti_{50.1} shape memory alloy. *Acta Mater.* 91 (2015), 318-329.
- [20] K. Bhattacharya, S. Conti, G. Zanzotto, J. Zimmer. Crystal symmetry and the reversibility of martensitic transformations. *Nature* 428 (2004), 55-59.
- [21] J. Cui, Y. S. Chu, O. O. Famodu, Y. Furuya, J. Hattrick-Simpers, R. D. James, A. Ludwig, S. Thienhaus, M. Wuttig, Z. Zhang, I. Takeuchi. Combinatorial search of thermoelastic shape-memory alloys with extremely small hysteresis width. *Nat. Mater.* 5 (2006), 286-290.
- [22] Y. Song, X. Chen, V. Dabade, T. W. Shield, R. D. James. Enhanced reversibility and unusual microstructure of a phase-transforming material. *Nature* 502 (2013), 85-88.
- [23] J. L. Ericksen. Some phase transitions in crystals. *Arch. Ration. Mech. Anal.* 73 (1980), 99-124.
- [24] K. Bhattacharya. *Microstructure of martensite*, Oxford University Press, New York, 2004.
- [25] Y. Gao, R. Shi, S. A. Dregia, Y. Wang. Group theory description of transformation pathway degeneracy in structural phase transformations. *Acta Mater.* 109 (2016), 353-363.
- [26] X. Huang, G. J. Ackland, K. M. Rabe. Crystal structures and shape-memory behaviour of NiTi. *Nat. Mater.* 2 (2003), 307-311.
- [27] S. Rajagopalan, A. L. Little, M. A. M. Bourke, R. Vaidyanathan. Elastic modulus of shape-memory NiTi from in situ neutron diffraction during macroscopic loading, instrumented indentation, and extensometry. *Appl. Phys. Lett.* 86 (2005), 081901.
- [28] A.P. Stebner, S. C. Vogel, R. D. Noebe, T. A. Sisneros, B. Clausen, D. W. Brown, A. Gary, L. C. Grinson. Micromechanical quantification of elastic, twinning, and slip strain partitioning exhibited by polycrystalline, monoclinic nickel–titanium during large uniaxial deformations measured via in-situ neutron diffraction. *J. Mech. Phys. Solids* 61 (2013), 2302-2330.
- [29] C. N. Tupper, D. W. Brown, R. D. Field, T. A. Sisneros, B. Clausen. Large strain deformation in uranium 6 wt pct niobium. *Metall. Trans. A* 43A (2012), 520-530.

- [30] J. K. Mackenzie & J. S. Bowles. The crystallography of martensite transformations II. *Acta Metall.* 2 (1954), 138-147.
- [31] M. S. Wechsler, D. S. Lieberman & T. A. Read. On the theory of the formation of martensite. *Trans. AIME* 197 (1953), 1503-1515.
- [32] C. M. Wayman. Introduction to the crystallography of martensitic transformation Collier-Macmillan, New York, 1964.
- [33] J. Orsborn, G. Lee, S. A. Romo, T. F. Broderick, A. J. Ramirez, D. W. McComb, H. L. Fraser. Practical considerations for high-resolution transmission Kikuchi diffraction mapping and analysis in titanium alloys. *Microsc. Microanal.* 22 (Suppl 3) (2016), 636-637.
- [34] P. W. Trimby. Orientation mapping of nanostructured materials using transmission Kikuchi diffraction in the scanning electron microscope. *Ultramicroscopy* 120 (2012), 16-24.
- [35] S. Suzuki. Features of transmission EBSD and its application. *JOM* 65 (2013), 1254-1263.
- [36] R. Sedlacek, W. Blum, J. Kratochvil, S. Forest. Subgrain formation during deformation: Physical origin and consequences. *Metall. Mater. Trans. A* 33A (2002), 319.
- [37] E. I. Galindo-Nava, P. E. J. Rivera-Diaz-del-Castillo. A thermodynamic theory for dislocation cell formation and misorientation in metals. *Acta Mater.* 60 (2012), 4370-4378.
- [38] C. Chluba, W. Ge, R. Lima de Miranda, J. Strobel, L. Kienle, E. Quandt, M. Wuttig. Ultralow-fatigue shape memory alloy films. *Science* 348 (2015), 1004-1007.
- [39] Y. Mishin, M. Asta, J. Li. Atomistic modeling of interfaces and their impact on microstructure and properties. *Acta Mater.* 58 (2010), 1117-1151.
- [40] Wang Y, Li J. Phase field modeling of defects and deformation. *Acta Mater.* 58 (2010), 1212-1135.
- [41] F. Roters, P. Eisenlohr, L. Hantcherli, D. D. Tjahjanto, T. R. Bieler, D. Raabe. Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite-element modeling: Theory, experiments, applications. *Acta Mater.* 58 (2010), 1152-1211.
- [42] J. S. Lee Pak, C. Y. Lei & C. M. Wayman. Atomic ordering in Ti-V-Al shape memory alloys. *Mater. Sci. Eng. A* 132 (1991), 237-244.
- [43] K. Otsuka & K. Shimizu. On the crystallographic reversibility of martensitic transformations. *Scripta Metall.* 11 (1977), 757-760.
- [44] C. M. Wayman. On memory effect related to martensitic transformations and observations in b-brass and Fe₃Pt. *Scripta Metall* 5 (1971), 489-492.
- [45] K. Otsuka, X. Ren. Physical metallurgy of Ti-Ni-based shape memory alloys. *Prog. Mater. Sci.* 50 (2005), 511-678.

- [46] J. Khalil-Allafi, A. Dlouhy, G. Eggeler. Ni₄Ti₃-precipitation during aging of NiTi shape memory alloys and its influence on martensitic phase transformations. *Acta Mater* 50 (2002), 4255-4274.
- [47] K. Gall, H. Sehitoglu, Y. I. Chumlyakov, I. V. Kireeva, H. J. Maier. The influence of aging on critical transformation stress levels and martensite start temperatures in NiTi: Part II—discussion of experimental results. *J. Eng. Mater. Technol.* 121 (1999), 128-37.
- [48] L. Patriarca, H. Sehitoglu, E. Yu. Panchenko, Y. I. Chumlyakov. High-temperature functional behavior of single crystal Ni_{51.2}Ti_{23.4}Hf_{25.4} shape memory alloy. *Acta Mater.* 106 (2016), 333-343.
- [49] W. Tirry, D. Schryvers. Linking a completely three-dimensional nanostrain to a structural transformation eigenstrain. *Nat. Mater.* 8 (2009), 752-757.
- [50] W. Tirry, D. Schryvers. Quantitative determination of strain fields around Ni₄Ti₃ precipitates in NiTi. *Acta Mater.* 53 (2005), 1041-1049.
- [51] Y. Gao, N. Zhou, F. Yang, Y. Cui, L. Kovarik, N. Hatcher, R. D. Noebe, M. J. Mills, Y. Wang. P-phase precipitation and its effect on martensitic transformation in (Ni, Pt) Ti shape memory alloys. *Acta Mater.* 60 (2012), 1514-1527.
- [52] F. Yang, D. R. Coughlin, P. J. Phillips, L. Yang, A. Devaraj, L. Kovarik, R. D. Noebe, M. J. Mills. Structure analysis of a precipitate phase in an Ni-rich high-temperature NiTiHf shape memory alloy. *Acta Mater.* 61 (2013), 3335-3346.
- [53] H. E. Karaca, E. Acar, G. S. Ded, B. Basaran, H. Tobe, R. D. Noebe, G. Bigelow, Y. I. Chumlyakov. Shape memory behavior of high strength NiTiHfPd polycrystalline alloys. *Acta Mater.* 61 (2013), 5036-5049.
- [54] B. C. Hornbuckle, T. T. Sasaki, G. S. Bigelow, R. D. Noebe, M. L. Weaver, G. B. Thompson. Structure–property relationships in a precipitation strengthened Ni–29.7 Ti–20Hf (at%) shape memory alloy. *Mater. Sci. Eng.* 637 (2015), 63-69.
- [55] K. Otsuka, A. Saxena, J. Deng, X. Ren. Mechanism of the shape memory effect in martensitic alloys: an assessment. *Phil. Mag.* 91 (2011), 4514-4535.
- [56] T. Saburi, Y. Watanabe & S. Nenno. Morphological characteristics of the orthorhombic martensite in a shape memory Ti–Ni–Cu alloy. *ISIJ Int.* 29 (1989), 405-411.
- [57] S. K. Wu, C. M. Wayman. Martensitic transformations and the shape memory effect in Ti₅₀Ni₁₀Au₄₀ and Ti₅₀Au₅₀ alloys. *Metallography* 20 (1987), 359-376.

Figure Captions:

Fig. 1. Schematic drawing of a square to parallelogram transformation in 2D and the local energy landscape as described through a Landau polynomial. (The lattice correspondence is shown by the red frames.)

Fig. 2. Schematic drawing of a square to parallelogram transformation in 2D: (a) Global free energy landscape; (b) Free energy landscape biased by an external load. (Unit cell in each lattice is illustrated by the dashed lines, while the lattice correspondence is shown by the red frames.)

Fig. 3. (a) B2 and BCO structures in NiTi; (b) schematic illustration of the structural states global energy landscape in NiTi. The lattice correspondence during the deformation (structure transformation) are shown by the red frames.

Fig. 4. Schematic drawing of the formation of special grain boundaries in B2 during transformation cycling: (a) perfect crystal in initial B2 state; (2) compound twinning in B19' states; (3) $\Sigma 9$ boundaries between the initial B2 and new B2 states; (4) $\Sigma 3$ boundary between two new B2 states.

Fig. 5. (a) Transmission Kikuchi Diffraction (TKD) map showing a 60 degree misorientation between sub-grains in a 2-cycle 150 MPa specimen of NiTi. The misorientation profile of the double-headed arrow is plotted in (b), crossing two $\Sigma 3$ boundaries.

Fig. 6. Bright-field image showing multiple dislocations in a 20-cycle 100 MPa sample crossing a bend contour.

Fig. 7. Bright-field image showing the dislocation structure near a triple point in the original austenite grain: (a) high magnitude image; (b) low magnitude image with 5° tilting.

Fig. 8. (a) Scanning transmission electron micrograph of a 20,438 cycle 100 MPa specimen after in-situ heating and achieving a fully B2 structure. (b) TKD mapping of the same region showing significant sub-grain formation within the original B2 grains. Data points having a confidence index (CI) below 0.05 have been removed and are shown as black pixels.

Fig. 9. Schematic illustrations of effects of (a) long range atomic ordering and (b) nano-precipitates on the free energy landscape of martensitic transformations.

Figure 1

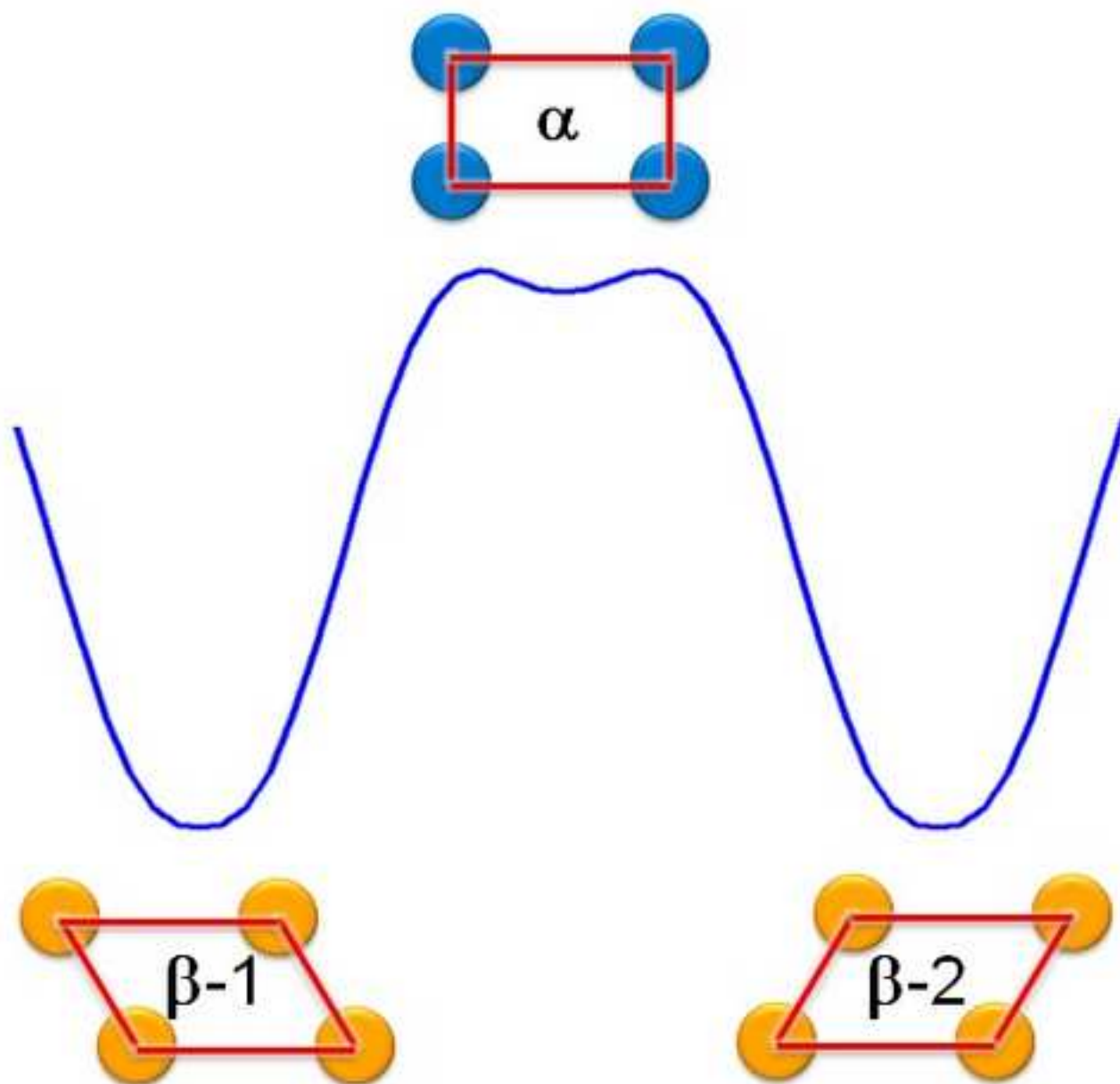


Figure 2

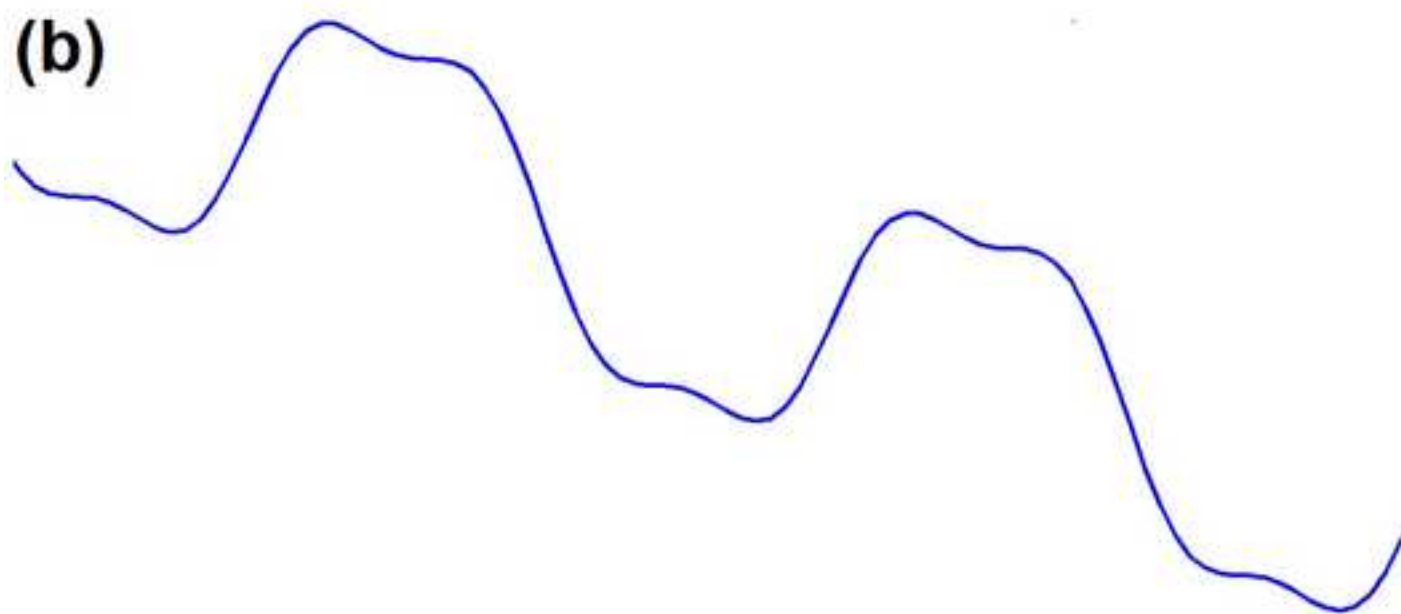
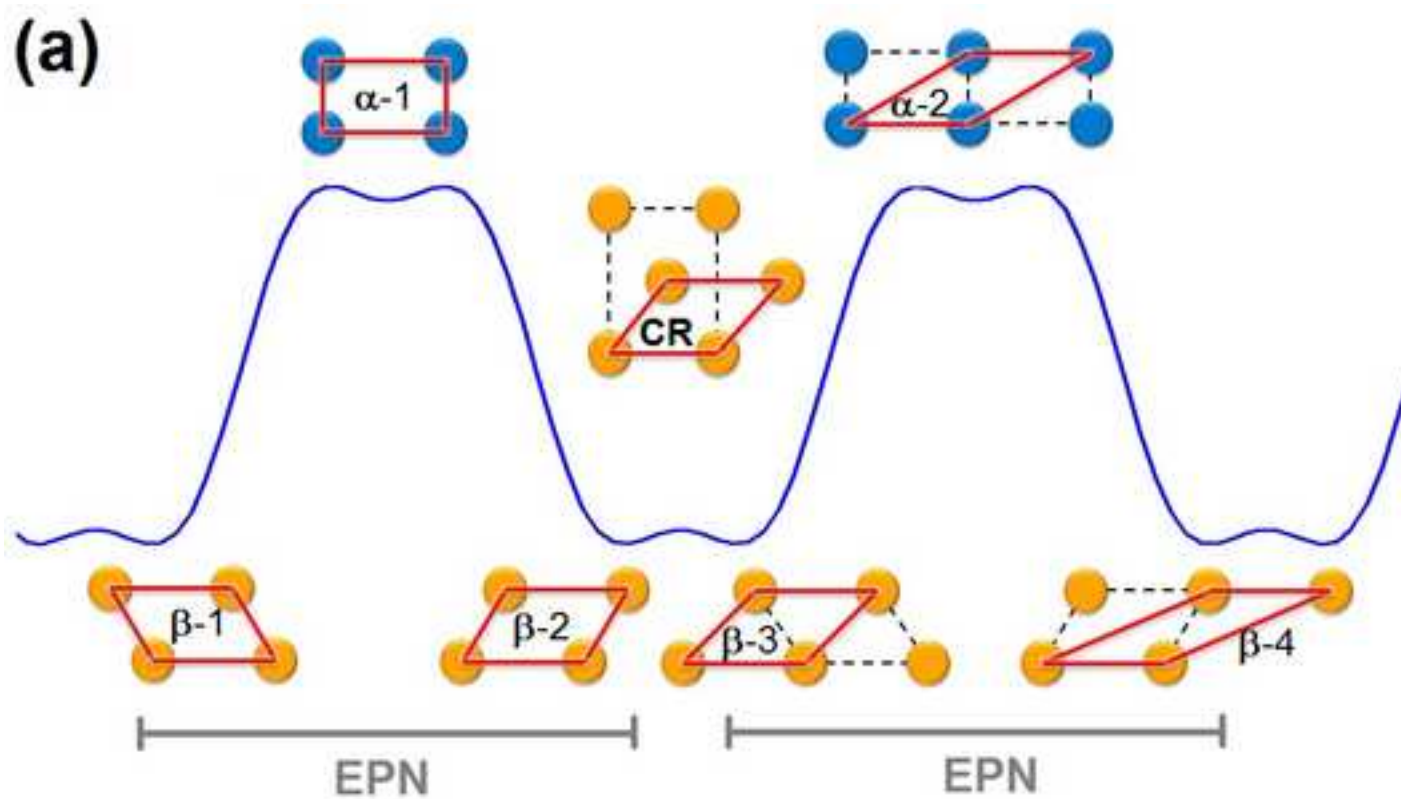


Figure 3

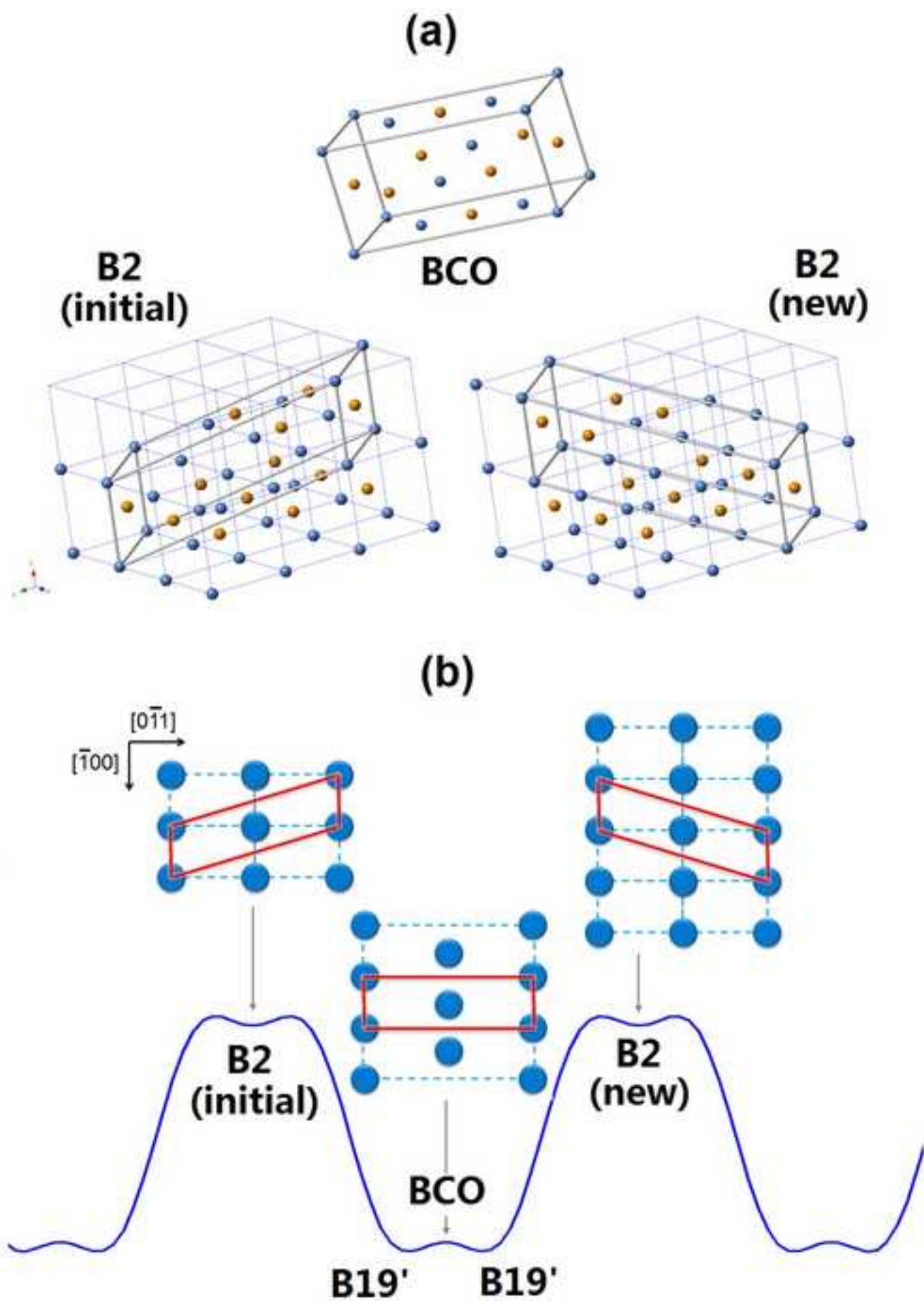


Figure 4

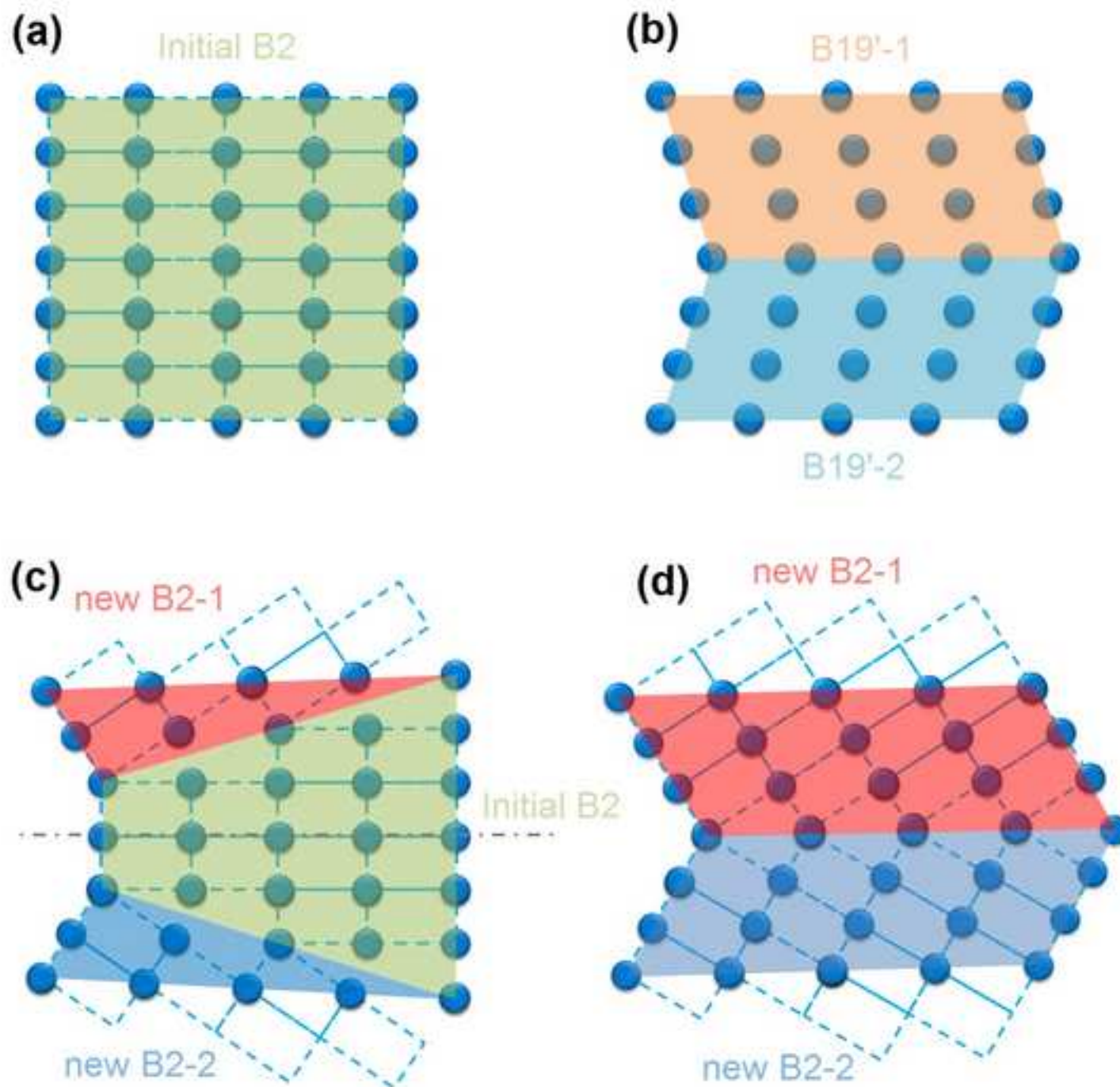


Figure 5

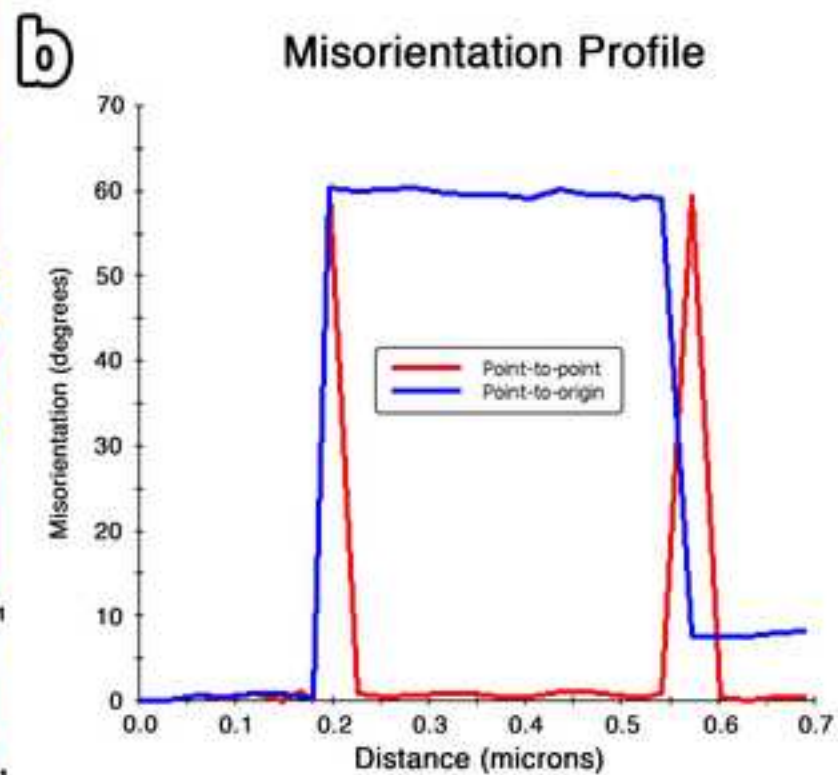
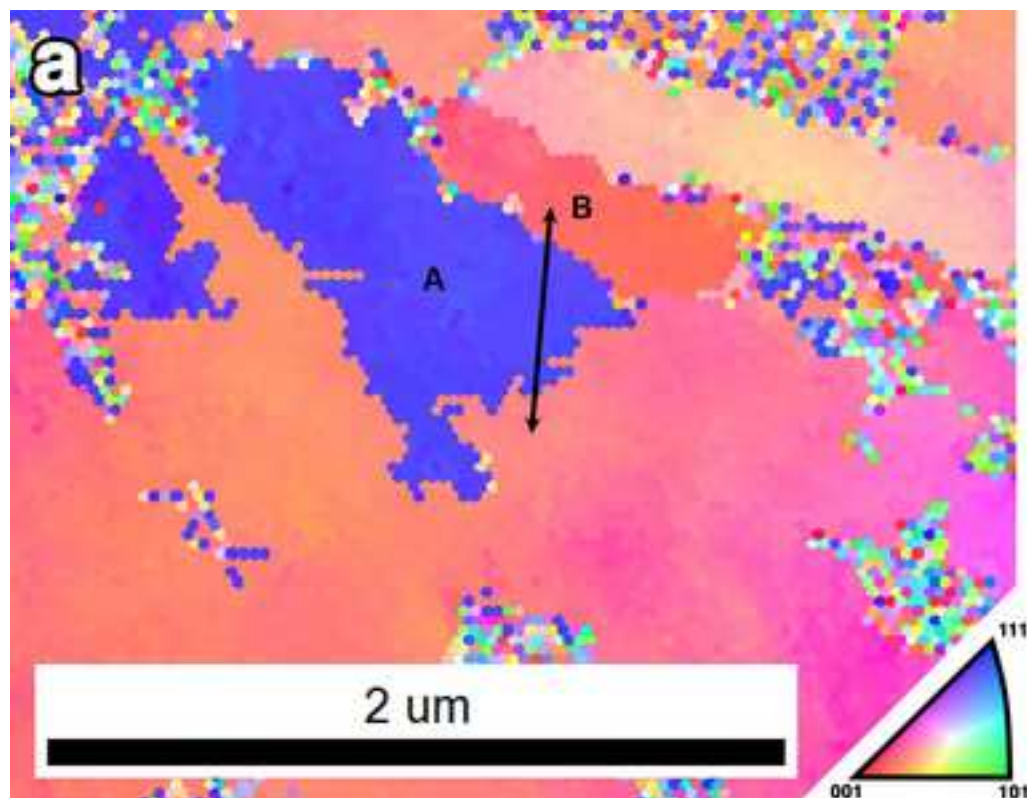


Figure 6

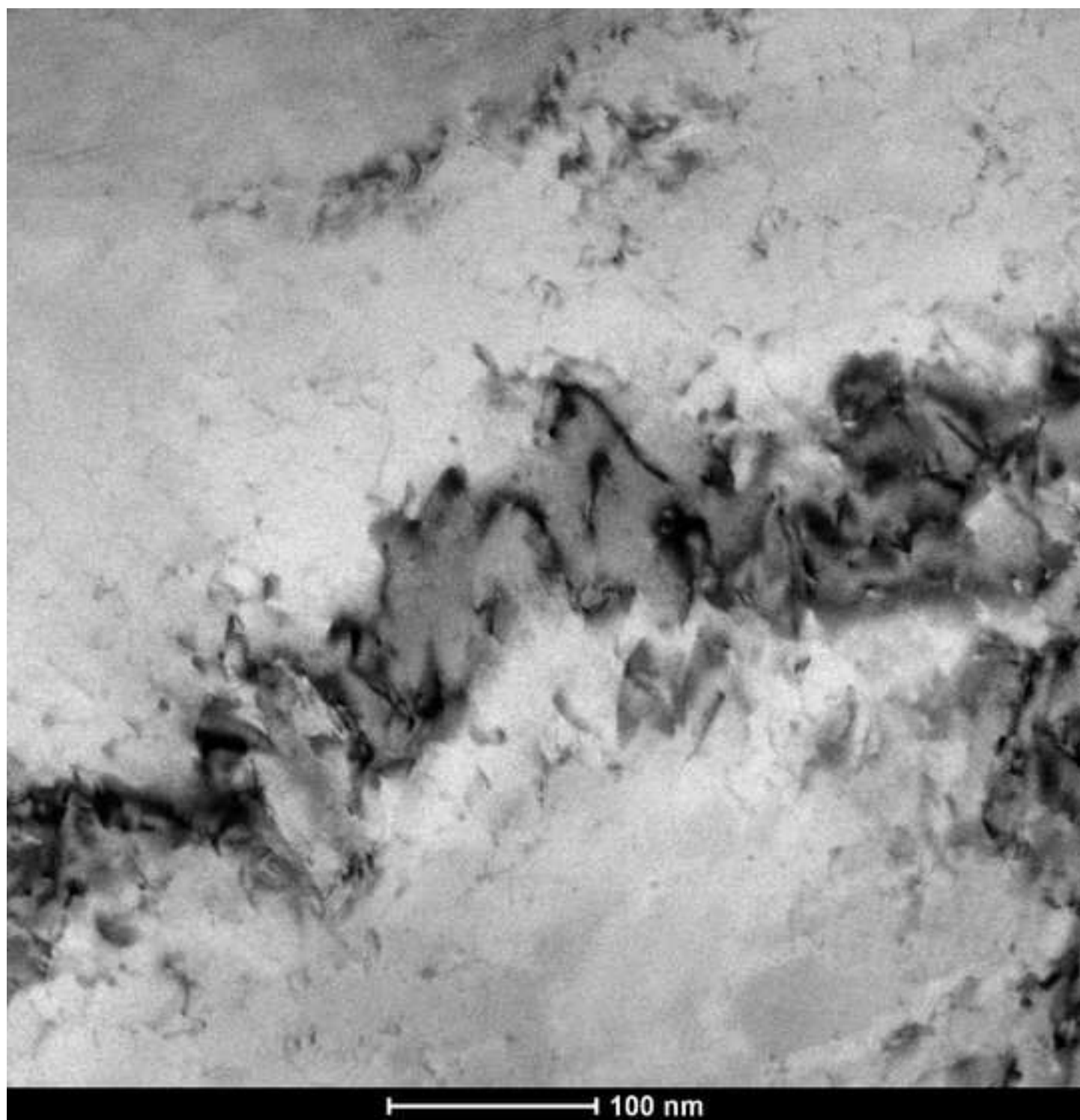


Figure 7

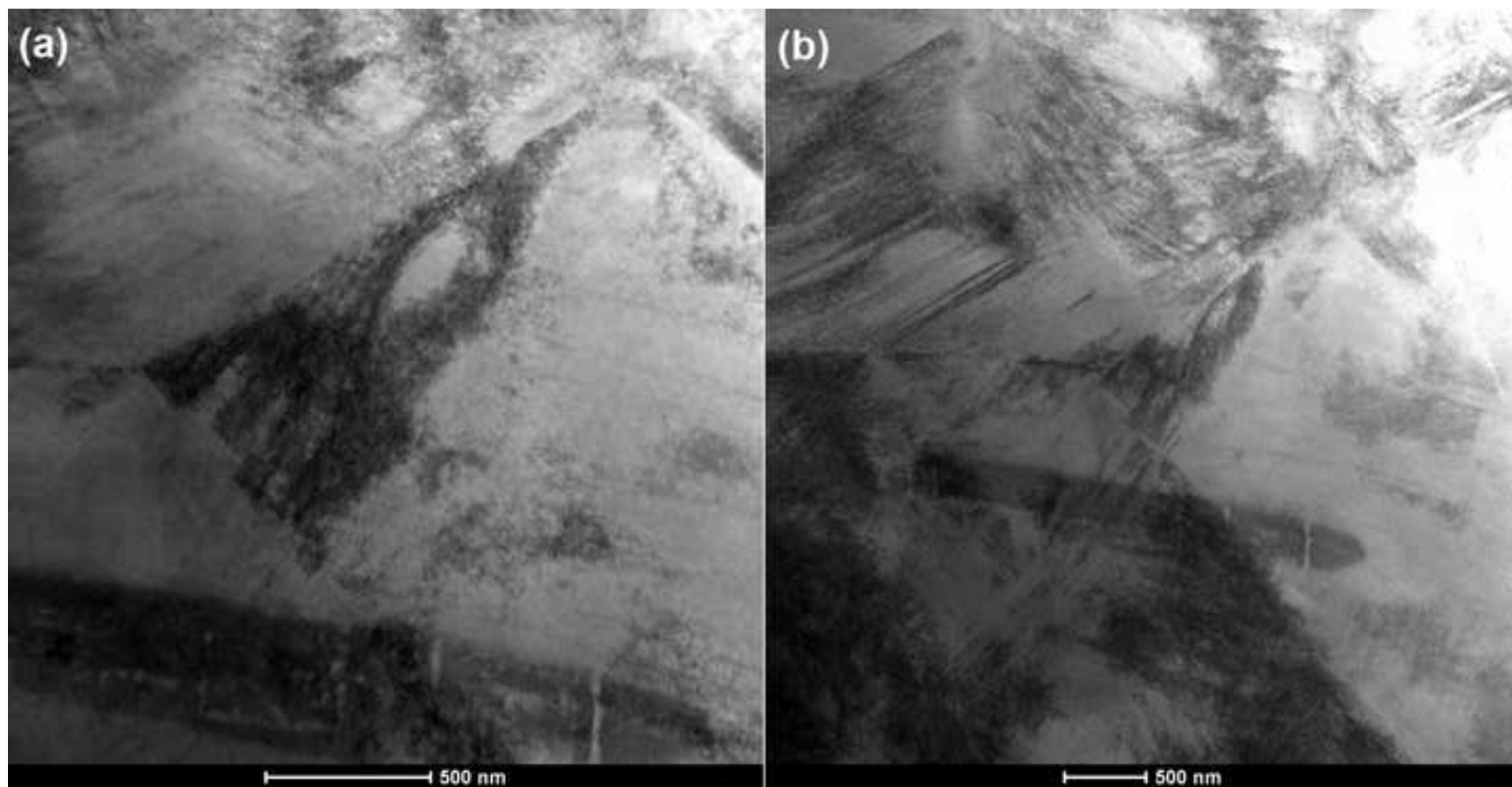


Figure 8

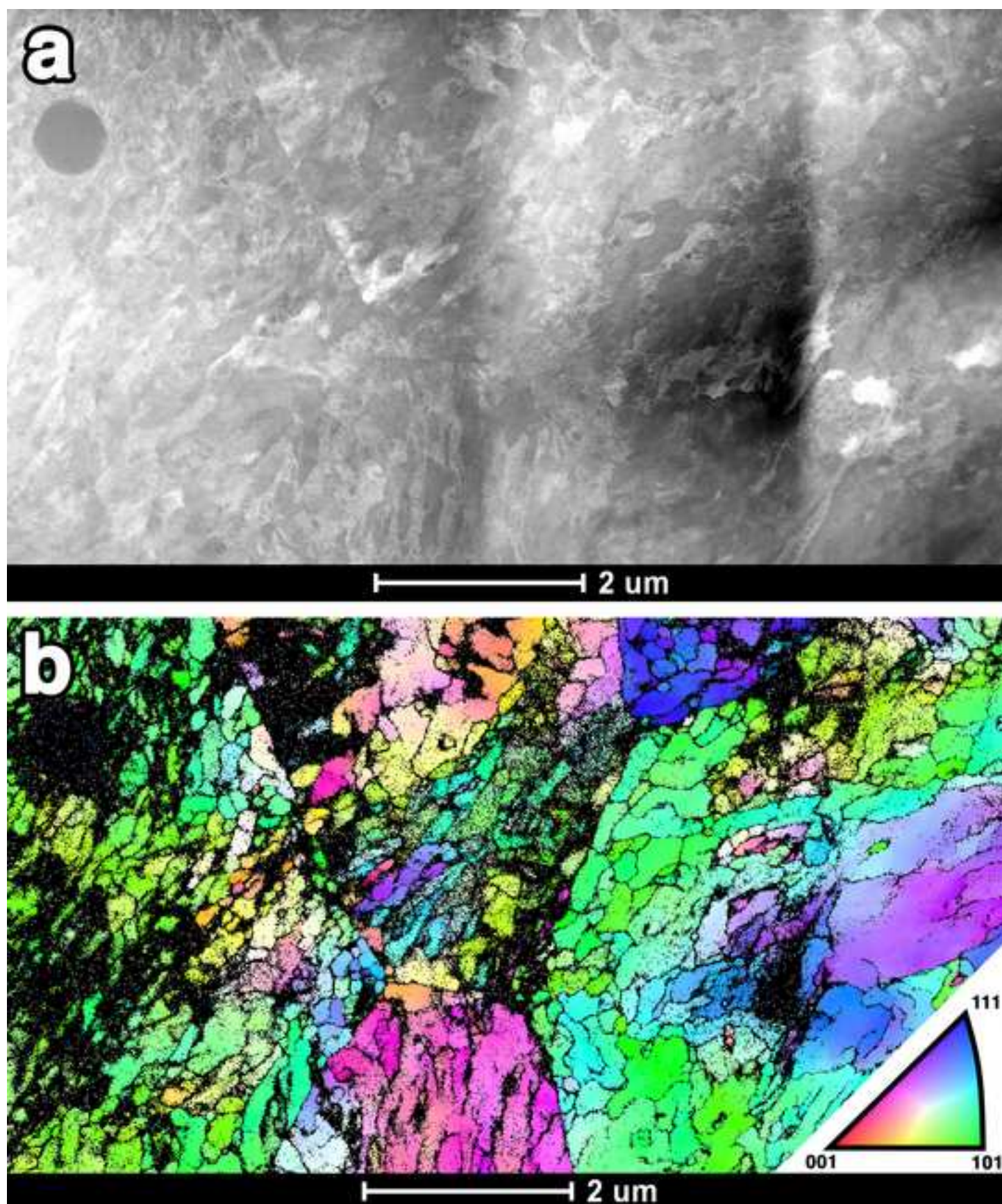
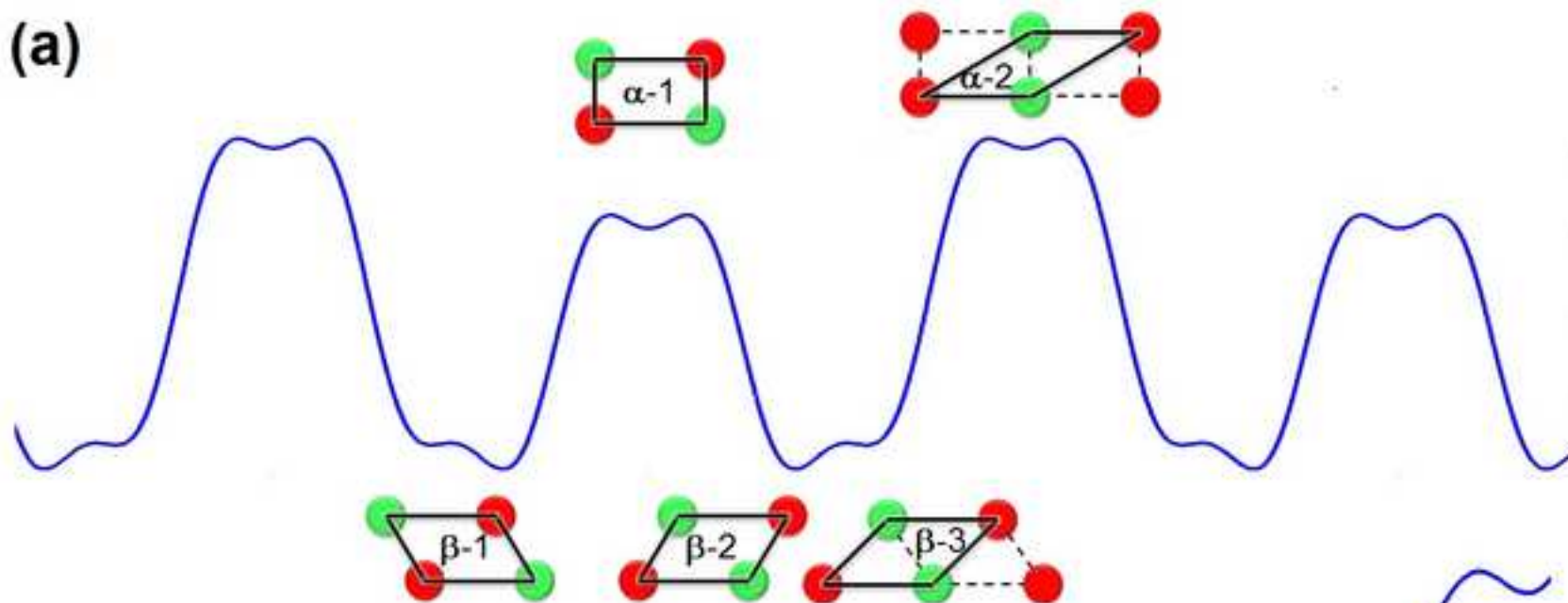
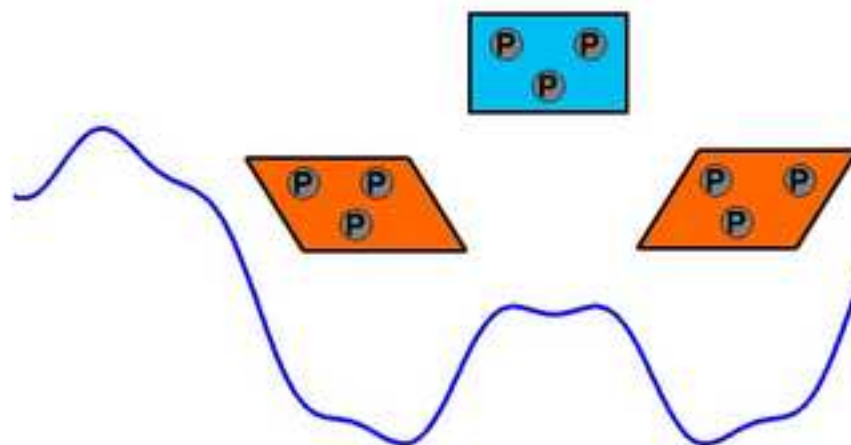


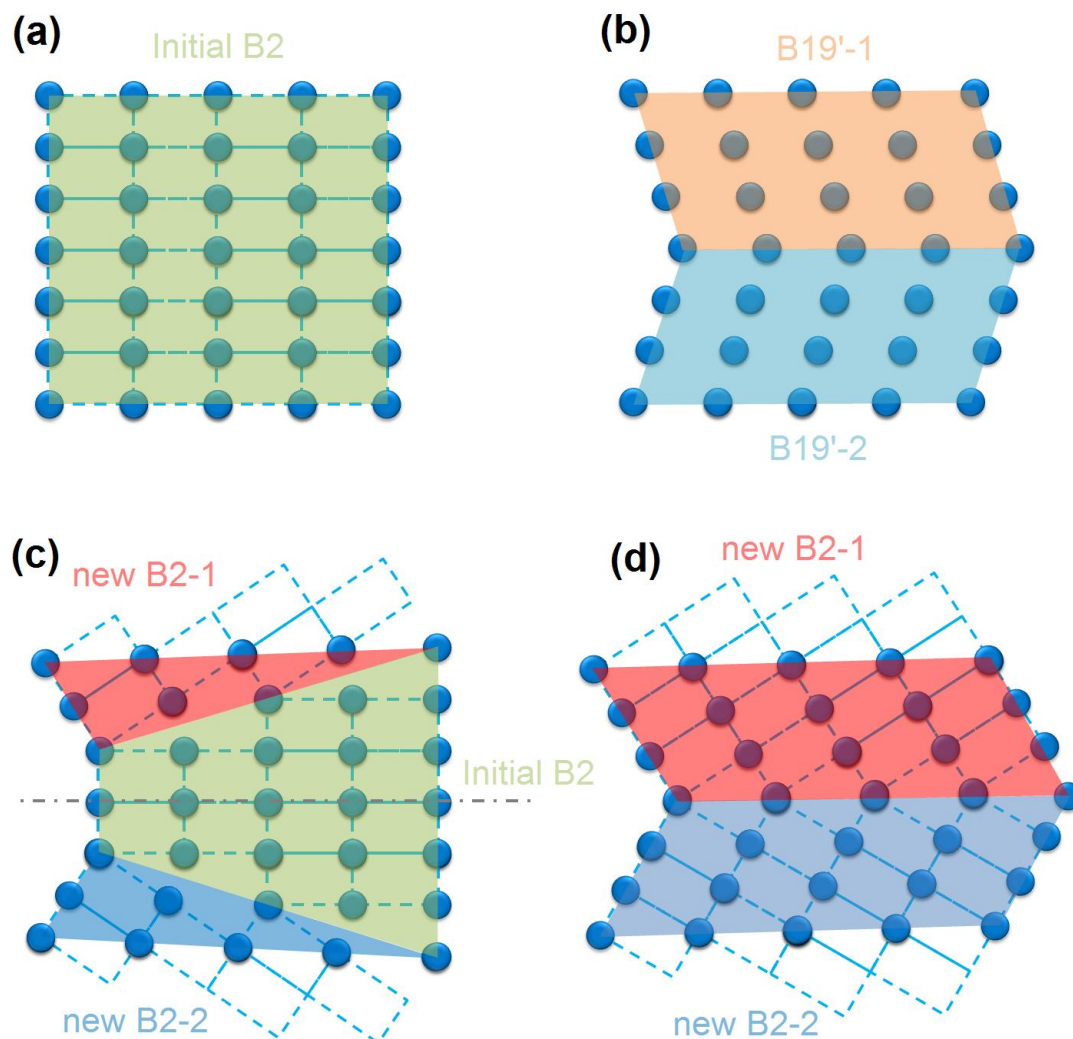
Figure 9

(a)



(b)





Schematic drawing of the formation of special grain boundaries in B2 during transformation cycling ($\{011\}$ cross-section of B2 with only one set of sub-lattice atoms plotted): (a) perfect crystal in initial B2 state; (2) compound twinning in B19' states; (3) $\Sigma 9$ boundaries between the initial B2 and new B2 states; (4) $\Sigma 3$ boundary between two new B2 states.