Deformation physics of shape memory alloys – Fundamentals at atomistic frontier

Piyas Chowdhury, Huseyin Sehitoglu *  
Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, 1206 W. Green St., Urbana, IL 61801, USA

A R T I C L E   I N F O  
Article history:  
Received 13 January 2017  
Received in revised form 21 March 2017  
Accepted 24 March 2017  
Available online 27 March 2017

Keywords:  
Shape memory  
Martensitic transformation  
Phase reversibility  
Density functional theory  
Molecular dynamics

A B S T R A C T

Application spectrum of shape memory alloys (SMA) is expanding rapidly and proportionately so is the engineering demand for superior materials. An essential prerequisite to developing novel SMAs is a clear perception of the deformation physics underlying their extraordinary shape recoverability. To that end, modern atomistic simulation tools have proffered state-of-the-art models, which usher in new clarifications for SMA deformation properties. It was found, for example, that ab initio energy pathways are at the core of dictating the extent of shear and shuffle for both phase transformation and variant formation at atomic lengthscale. These important revelations are accomplished by addressing inherent solid-state effects, which underpin the natural tendency to seek the energetic ground state. Moreover, empirical potential based models, benefitting from ab initio calculations, have allowed an atomic-resolution view into the phase evolution and the concurrent twinning phenomena relating directly to constitutive properties. Here, we revisit salient examples of these cutting-edge theoretical discoveries regarding SMA deformation along with discussions on pertinent experimental evidences.

© 2017 Elsevier Ltd. All rights reserved.

C O N T E N T S

1. Background .................................................................................................................. 50
   1.1. Perspective on SMA literature .................................................................................. 50
   1.2. Significance of atomistics ....................................................................................... 52

2. Overview of general deformation behaviors of SMAs .................................................... 53

3. Case study: equiatomic NiTi SMA ............................................................................. 53
   3.1. Stability of phases from first principles ...................................................................... 53
   3.1.1. Energy pathway .................................................................................................... 53
   3.1.2. B19° versus B33 .................................................................................................... 54
   3.2. Calculation of elastic moduli ...................................................................................... 54
   3.2.1. Predictions based on single crystal and their significance .......................................... 54
   3.2.2. Modified elastic anisotropy in a twinned lattice .................................................... 55
   3.2.3. Elastic anisotropy of Ni4Ti3 precipitate ................................................................. 57
   3.3. Energetics of twinning .............................................................................................. 57
   3.3.1. Brief overview of various twinning modes in NiTi .................................................. 57
   3.3.2. Type II twinning in martensite (B19°) phase ......................................................... 59

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

http://dx.doi.org/10.1016/j.pmatsci.2017.03.003  
0079-6425/© 2017 Elsevier Ltd. All rights reserved.
1. Background

1.1. Perspective on SMA literature

Nowadays shape memory alloys are finding diversified applications in a wide array of industries (e.g. biomedical, aerospace, automotive) owing to their extraordinary deformation recoverability [1–3]. The mechanical healing ability of an SMA-made component (e.g. a cardiovascular stent) is principally rooted upon reversible martensitic transformations at microstructural level [4]. Deformation micro-mechanisms, however, could be rather complex, involving internal twinning, precipitation, intermediate phase nucleation, etc. [5–8]. Variables across multiple spatial scales (ranging from sub-nanometer quantum forces within a single grain to micron-level multi-grain interactions) collectively contribute to the overall inter-crystal transformability. Given their technological importance, predicting SMA behavior remains a dedicated discipline, currently employing theoretical tools spanning atomistics to continuum [9,10]. This article provides an overview of the atomic lengthscale mechanisms pertaining to molecular dynamics and density functional theory studies. More discussions follow on these two approaches. Fig. 1 puts the current topics into perspective.

The extent of SMA research, both experimental and theoretical, is vast. It is instructive to categorize them lengthscale-wise, as illustrated by Fig. 2, to develop a proper perspective. It follows that the majority of studies concerns continuum scale behaviors, namely, thermodynamics [11], constitutive modeling [12–14], finite element simulations [15–18]; experimentally, thermo-mechanical characterizations and component performance assessment remain the primary emphases [19,20]. At the mesoscale (i.e. grain level), digital image correlation (for measuring strain localizations [21,22]), X-ray diffraction (for phase identification [23–25]), electron microscopy (for studying microscopic defects [26]) and electron backscatter diffraction (for texture determination [27]) are the common experimental techniques for microstructure characterization. On the other hand, theorization of mesoscale variables (e.g. roles of grain size, texture, precipitates) is approached commonly with phenomenological assumptions [28–32]. Such models are tied back-to-back with empirical observations from which are extracted the requisite material constants i.e. the fitting parameters essential for accurate prediction of macroscopic constitutive responses [33]. Phase field models consider the evolution of the martensitic phase in terms of free energy functionals [34–36] (per the Ginzburg-Landau theory [37]). The advent of molecular dynamics [38] in the SMA context is a recent development (over last decade) as a promising mesoscopic tool, capable of addressing sub-micron phenomena within a single grain. The quantum lengthscale tools include density functional theory based predictions of sub-nanometer physics [39].
Fig. 1. A broad perspective on the applications, microstructure and atomistic modeling of SMAs. This figure summarily expresses the background on the current topic.

Fig. 2. The pyramidal nature of the diagram indicates the current extent of SMA literature based on physical lengthscale. Continuum level models and experimentations are the most widely undertaken approaches. Mesoscale experimental characterizations remain quite common in research while molecular dynamics simulations are also emerging in the recent decade. No experimental method exists to address the quantum scale nuances, which density functional theory (DFT) can supplement.
which no experimental technique can directly reach. The atomistic studies altogether have brought forth important physical insight regarding inherent deformation propensity.

1.2. Significance of atomistics

Strain recovery attributes of SMAs are principally related with reversible phase transformation between austenite and martensite accompanied by twinning processes (nucleation and migration) [6,40]. The deterioration of shape recoverability, by contrast, is attributed to slip-mediated plastic deformation [41,42]. Mechanistically, the phase transformation process in SMAs is diffusionless and displacive in character [10,43]. The mechanism involves atoms competing for shear- and/or shuffle-type movements to various degrees (i.e. few angstroms) in search of the most favored crystal structure. The natural predisposition for such phenomena is rooted upon the exact outcome of sub-lattice level interactions as governed by quantum forces. Similarly, material’s tendency to seek energetically preferred ground state(s) gives rise to concurrent twinning phenomenon. Given that SMAs consist of transition metals, factors such as partial/full d-orbital occupancy, electron spin, and phonon spectra constitute the important regulators of both crystal transformability [44] and twinning propensity. These considerations are essentially non-trivial, in that they collectively foster the energetics conducive to inter-lattice conversions as well as atoms arranging into mirrored positions thus forming twins. Evidently, investigations of such ultra-small scale nuances ought to be conducted on predictive grounds since the accuracy of experimental techniques halts at tens of nanometers spatial resolutions.

It is well-known that long-/short-range ordering, atomic staking, solid solution, elastic stiffness, etc. are the most influential variables at the discrete lattice level underpinning the macroscale SMA properties [45–49]. The DFT method relies on the inherent solid-state attributes (e.g. the nature of phonon spectra, Fermi surface) to reproduce these crystallographic nuances [50,51]. Its central premise involves the iterative solution of valence electron densities in a crystal (from a simplified time-independent Schrodinger’s equation) [52]. Material-specific qualities such as the type of magnetism (i.e. the electronic spin) and chemical potential are also accounted for in order to precisely reconstruct the bonding landscape [53,54]. Early evidence of the reliability of DFT calculations in the SMA context was noted in the accurate predictions of elastic moduli [55,56]. The transformation problem-wise, the stability of various crystal phases has lately been determined from an energetic perspective, and thus their relative predilection was rationalized.

The most recent undertakings have reported the energy pathways and in several SMAs (NiTi, NiTiPt, NiTiPd, Ni$_3$FeGa, Ni$_2$MnGa). The predicted energy routes confirmed the existence of various local maxima and minima, thus individuating intermediate (unstable or metastable) structures en route to the formation of more stable ones [57–59]. In the process, the relative contributions of atomic shear and shuffle were established (e.g. in Ni$_3$FeGa). Moreover, an energetic validation of various crystal structures under defect-free condition has emerged (e.g. in Ni$_3$FeGa, Fe-Mn-Al-Ni). This helped to identify their relative propensity for transformation and twinning as would be expected during the actual deformation. Nonetheless,
the DFT methods are computationally expensive, confining a typical simulation supercell to several hundred atoms only [60]. This essentially precludes theorization of mesoscale deformation mechanisms, say, in an isolated yet prototypical grain consisting of multi-million atoms.

Molecular dynamics (MD) simulations have fulfilled the need of predictions based on larger systems [61]. The increased computational efficiency in MD, albeit at the expense of accuracy, is accomplished mainly with the aid of improved “potentials” [62,63]. A potential is an empirical curve-fitted database of select experimental and/or ab initio properties. Here the valence electron densities, instead of being solved directly, are pre-defined and ready for use. The most popular examples of SMA potentials include embedded atom method (EAM) and modified EAM (MEAM) types [64–68]. The range of SMA materials thus-investigated has remained rather scarce due to the unavailability of potentials, the development of which remains an open avenue of research today. The MD models contributed to the fundamental understanding of the dynamic phase evolution process as the deformation proceeds [69,70].

2. Overview of general deformation behaviors of SMAs

It is helpful to open discussions by briefly overviewing the generic mechanical behaviors of SMAs as observed in experimental studies. This would enable the reader to comprehend the significance of atomistic theories, to be recounted next, from the correct perspective. To that end, we provide Fig. 3, which illustrates the most widely reported thermomechanical constitutive responses and the associated sub-structural phase changes. In general, deformation recovery attributes in SMAs can be manifested in three distinct forms: (a) shape memory effect, (b) superelasticity and (c) isobaric shape memory effect.

The shape memory effect refers to a full recuperation of inelastic mechanical strains by means of unloading, heating and cooling. At temperatures below Mf, the initial twinned martensite, when stressed externally, becomes a fully de-twinned one, which remains so upon unloading. Now, heating the unloaded de-twinned martensite above Af would transform it into an austenitic crystal, which can then be cooled to regain the original twinned martensite phase. In the center inset of Fig. 3, the pathway along the green-shaded curve, the red and the blue lines (c → d → a → c) refers to the shape memory effect. Superalasticity is the full recovery of transformation-induced inelastic strain by mere unloading under isothermal condition at temperatures above Af. In the process, an austenitic crystal transforms into an internally twinned martensite. Upon unloading, the microstructural change is reversible, where the martensite reverts to the original austenite again (following the path a → b → c → a on the orange curve). On the other hand, the isobaric shape memory behavior is related with temperature cycling and the associated straining at a constant stress. During cooling (below Mf), the material becomes twinned martensite from austenite with an associated strain, which can be recovered by heating above Af (path: a → d → a).

Phenomenology of the entire deformation scenario (including reversible and irreversible phenomena) has also been modeled mathematically [71–73]. Theoretically, the austenite-to-martensite transformation can be mapped by the distortion matrix, U, whose determinant is close to unity (with eigenvalues ordered as: λ1 ≤ λ2 ≤ λ3). From the U tensor, significant implications regarding the hysteresis behavior can be extracted. For instance, λ2 = 1 means very low hysteresis, which microstructurally corresponds to high geometric compatibility between austenite and martensite [74]. Physically, this could be attributed to a considerable lack of interfacial defects i.e., say, transformation in absence of twin boundaries. Now, given that the U tensor is a function of lattice parameters of individual phases, it calls for a precise evaluation thereof, which can be achieved through DFT simulations [75]. In addition, many other issues have been elucidated both from DFT based predictions, and MD based deformation simulations. As follows, we discourse on materials-specific case studies. In doing so, the atomistic models as well as experimental findings on phase transformation and twinning mechanisms, stiffness effects, etc. will be analyzed. Specific differences among them will be highlighted.

3. Case study: equiatomic NiTi SMA

The binary NiTi SMA is a widely applied and well-researched material. Given the scope of the early undertakings, we discuss them at length in the following sections.

3.1. Stability of phases from first principles

3.1.1. Energy pathway

Empirically, the strain recovery in binary NiTi has always been associated with a reversible austenite (B2) ↔ martensite (B19′) transformation. The ordered B2 phase (austenite) possesses high structural energy (as well as entropy) compared to that of the martensite lattice. Hence, the austenite phase is stable only at high temperature (i.e. above the austenite finish temperature, Af) while the martensite is stabilized at lower temperature with lowered symmetry, structural energy and entropy. Depreciating temperature or applying mechanical stress leads to the natural transformation of the austenitic B2 lattice into a monoclinic (martensitic) B19′ as observed empirically. Intriguingly, counterintuitive to the experimental observations, Huang et al. first suggested [45] that B19′ is not the global energy ground state along the entire spectrum of transformation energy pathway. With ab initio calculations, they demonstrated that energetically most favorable martensitic crystal is a body-centered orthorhombic (BCO) structure, also known as the B33 phase. Subsequently, this notion...
was further corroborated by independently-undertaken DFT calculations by other researchers [57,59,76,77]. It is discovered that the energy reduces monotonically from B2 to B19 with respect to transformative modifications of structural parameters (i.e. atomic displacement or primitive cell angle). Although electron microscopy confirmed a B2-to-B33 conversion in other materials such as Co-Ni-Zr alloys [78], no experimental evidence thereof exists for NiTi SMAs. A possible rationale for not reaching the B33 phase, despite being the most favored state, is the presence of as substantial energy barrier, which can be accentuated due to internal stresses inherent to martensite microstructure. We further explain below.

3.1.2. B19\(^0\) versus B33

It is imperative to note that the ab initio predictions are performed considering pristine lattice condition i.e. in absence of interfacial defects (twin or grain boundaries). The NiTi martensitic microstructure in reality, however, consists of self-accommodated matrix-twin pairs (also called the correspondence variant pair, CVP). It is worth recalling that any interfacial defect (e.g. a grain boundary) is associated with a highly localized stress field in its immediate vicinity [79]. This is due to the local misfit of atoms shared by two adjoining grains, which constitute the boundary. Consequently, closely placed twin boundaries (i.e. in a self-accommodated martensite) would generate internal stress concentrations (even though the very formation of twins results in an overall stress reduction macroscopically). Thus, it is reasonable to conjecture that the substantial presence of twin boundaries in the martensite would foster an environment which may promote certain phase(s) while suppressing others. In other words, it is worthwhile to investigate the sensitivity of transformation energy pathway would be a function of the stress state.

DFT based predictions suggest that the preference of B19\(^0\) over B33 can be directly affected by internal hydrostatic stresses. Quantitative evidence to support this argument has recently been forwarded from first principles [58] while accounting for: (a) the sensitivity of high-resolution shear steps on the transformation pathways and (b) the effects of hydrostatic tension on the energy wells thereof. Atomic movements are subjected to highly non-linear topography of quantum bonding energy with angstrom-scale sensitivity. Thus, to capture local peaks and saddle points on the energy curve, the lattice ought to be sheared both incrementally and infinitesimally. Similarly, hydrostatic forces would essentially modify the valence electron density and thus the effective bonding length among atoms. By accounting for these factors, new insight is obtained regarding the stabilization of B19. The earlier studies suggested a barrier-less transformation path from B2 to B33 via B19. However, with minute shear increment, it was found that (Fig. 4) the transformation from B19\(^0\) to B33 requires overcoming an energy barrier albeit almost of insubstantial magnitude (8 meV/atom). Interestingly, hydrostatic tension (of 10 GPa) is found to further accentuate this barrier by raising it to 25 meV/atom. This strongly points to the role of triaxiality in precluding the B19\(^0\)-to-B33 transformation by enhancing the resistance. Further justification of this discovery remains an open question subjected to future independent investigations.

3.2. Calculation of elastic moduli

3.2.1. Predictions based on single crystal and their significance

Determination of elastic moduli in NiTi phases remains a topic of ongoing research both on experimental and theoretical grounds [56,80–88]. The NiTi austenite phase, due to its cubic symmetry, is described by three elastic constants (C\(_{11}\), C\(_{12}\) and C\(_{44}\)) [89]. On the other hand, the moduli in the monoclinic martensite phase having thirteen constants pose greater challenge for accurate experimental determination. The usefulness of moduli for either phase is, nonetheless, imperative, for example,
in continuum micromechanical models [11,30,90], which have motivated several DFT-based endeavors. Wagner and Windl [56] first computed the thirteen stiffness constants, $C_{ij}$ ($C_{11}, C_{12}, C_{13}, C_{15}, C_{22}, C_{23}, C_{25}, C_{33}, C_{35}, C_{44}, C_{46}, C_{55}, C_{66}$) for NiTi B19° as well as B33 considering pristine single crystal configurations. In DFT calculations, lattice ground states were established through energy minimization iterations with respect to lattice parameters, angles and fractional coordinates of atomic nuclei. Subsequently, Hatcher et al. [76] and Wang et al. [84] reported further DFT simulations, and predictions of elastic constants. Fundamentally, there are some methodological differences in all these approaches. For instance, Wagner and Windl imparted full internal relaxation during energy minimization, which was later adopted by Wang et al. By contrast, Hatcher et al. did not allow any coordinate or stress-tensor optimization, which preserved the metastable states during iteration steps. It should be noted that elastic constants are experimentally determined at temperatures levels such as 300 K or 400 K [87,88] while DFT-based constants are predicted at 0 K. Raising temperature is essentially associated with increased degree of lattice vibration (manifested in terms of phonon spectrum) and entropic state, which would ultimately reduce the moduli. Further research is necessary to examine these subtle effects.

The potential application of these DFT elastic constants is considerable. For instance, the ad hoc approach to estimating elastic moduli in micromechanical models [31] can certainly benefit from using the lattice-specific $C_{ij}$ constants. Moreover, they can also contribute to enhancing the accuracy of mesoscale atomistic simulations. For example, these constants are exploited in the development of refined molecular dynamics potential. Embedded atom method (EAM) or modified EAM potentials are well known to be as accurate as the curve-fitted constants [91,92]. Predicted elastic moduli have paved the way for enriching the EAM or MEAM potentials with crystal anisotropy information [93,94]. Since the majority of MD studies consider single crystals, thus-improved potentials can lead to better predictions of constitutive response (more discussion provided in following sections).

Regarding the agreement between the ab initio predictions and the experimental response, there are several important points to consider. The NiTi components used in typical engineering applications are of polycrystalline microstructure. In essence, the manifestation of elastic anisotropy, say, as Young’s modulus in uniaxial stress-strain curves, would depend largely on texture. For instance, elastic response of a full random-textured material would be isotropic in nature. Thus, the ab initio predictions representative of single-grain anisotropy needs to be translated into corresponding polycrystalline aggregate. To that end, it has long been customary to adopt formalisms proposed by Hill, Voigt and Reuss [95–99] for calculating the macroscale equivalents. These formalisms are, however, based on isotropic i.e. complete random texture assumption. In addition to texture effects, individual grains would contain twinned variants, which are subjected to detwinning. Moreover, elastic shear resistance on the twinning plane is an important variable, dictating the initiation of detwinning process. In essence, these issues bring into attention the need to study the directional nature of moduli in presence of mirror symmetry of lattice.

### 3.2.2. Modified elastic anisotropy in a twinned lattice

From an atomistic perspective, the anisotropy of elastic forces in a metallic lattice originates from the directional nature of bonding strength. Under applied mechanical forces, the atoms tend to stretch in certain direction(s) with more ease, which ultimately gives rise to different elastic constants. Similar argument could apply to elastic deformation in the presence of twin boundaries, which are ubiquitous in the undeformed martensitic microstructure. Prior to any form of inelastic deformation (i.e. detwinning or slip), the atoms constituting twinned martensite are bound to undergo elastic stretching (under...
tension) or squeezing (under compression). DFT calculations have demonstrated that the nature of elastic anisotropy is indeed different for a twinned lattice than the monolithic B19*.

To illustrate, let us consider one cycle of loading and unloading of a martensite crystal (in absence of interfacial defects other than twin boundaries) as shown schematically in Fig. 5. The initial martensite structure consists of multiple correspondence variant pairs (i.e. matrix-twin pairs). The initial slope during loading is thus related with the elastic stretching of twinned lattice. Unloading of fully detwinned martensite is controlled by the elastic moduli of a single-variant crystal. The difference in the lattice structure at the atomic level is illustrated with two primitive unit cells where the twin is of (001) compound type. A comparison is provided in terms of direction-dependent Young's moduli, $E_{hkl}$, for a twinned single crystal (adapted from [84]) and for a detwinned one (from [56]). Notice that the twinned lattice has significantly lowered moduli compared to the single variant one for any crystallographic direction (as indicated by the scale bar). For example, the magnitudes $E_{100}$ and $E_{001}$ for the twinned case are nearly one-third of the detwinned lattice. The trends are verified from experiments [84]. It follows that the presence of reflective symmetry ultimately lowers the interatomic forces against elastic

![Fig. 6. Comparison between the direction-dependent moduli of B2 NiTi matrix (inner, brown) and rhombohedral Ni$_4$Ti$_3$ precipitate lattice (adapted from [55]).](image)

![Fig. 7. Evolution of various twinning modes during the NiTi martensite deformation. See text for full description.](image)
This explains why a twinned martensite is preferred as the ground state over a detwinned one (which is stress-induced).

These results certainly point to the need for further research specifically addressing the roles of interface density, different types of interface, etc. in altering the moduli. It is noteworthy that many a scenario can be concocted in terms of these variables, which would lead to different magnitudes of constants. The macroscale response would essentially be a derivative of the ensemble response of all possibilities. Nonetheless, given the importance availability of DFT tools, it is worthwhile to examine these trends, which could potentially suggest ways to improve moduli (e.g. by addition of ternary elements).

3.2.3. Elastic anisotropy of Ni₄Ti₃ precipitate

Rhombohedral Ni₄Ti₃ precipitates of lenticular geometry have significant effects on the mechanical properties of NiTi SMAs [100–102]. Noteworthy among them are reduced transformation strain, stress, hysteresis and suppression of slip-mediated plasticity. Overall, their presence, which is achieved through appropriate aging treatments, is considered beneficial. The microscopic mechanism by which precipitates influence macroscale properties is related with the creation of internal stresses, which are essentially directional in nature. Given the uni-directionality of transformation twinning, the presence of direction-dependent stress-field would alter the propensity of different variant nucleation, [103]. Calculation of elastic moduli (C₁₁, C₁₂, C₁₄, C₁₅, C₃₃, and C₄₄) by Wagner and Windl [55] from first principles brings into attention some interesting insight in this regard. The predictions are compared in terms of direction-dependent Young’s modulus, E₁₁₁, between B2 NiTi (austenite matrix) and the rhombohedral Ni₄Ti₃ (precipitate) in Fig. 6. They first noted that both the precipitate and the embedding matrix have constant moduli on all other radial directions perpendicular to the (111) plane, the precipitate equatorial plane. The ½Σ₁₁₁ direction is parallel to ½Σ₁₁₁, rhombohedral. The highest modulus difference between the rhombohedral precipitate and B2 matrix was found along the direction perpendicular to the precipitate equatorial plane i.e. E₁₁₁. It should be noted that the material always seeks to minimize its strain energy in absence of any external stimuli. This explains why the particle grows as a flat lenticular geometry (thereby reducing elastic strain energy), and has a circular cross-section on the equatorial plane. Per the Eshelby theory [104,105], the modulus mismatch between the embedded particle and the surrounding matrix is an important variable. Both internal and external distortion fields are direct functions of the degree of modulus differences. Thus, in addition to providing physical elucidation, the computed moduli hold considerable promise for use in micromechanical models, which are of considerable recent interests [106].

3.3. Energetics of twinning

3.3.1. Brief overview of various twinning modes in NiTi

Twinning and detwinning are important physical phenomena underling shape memory effects, which comprise martensite deformation followed by its recovery via heating and cooling. The martensite microstructure of NiTi is characterized by complex self-accommodated twinned structures consisting of multiple variants [107]. Twins give rise to lattice invariant strain. The significance of twinning-mediated deformation mechanism is that the strains are essentially recoverable as long as concurrent dislocation activities are insubstantial. Some types of twins are prevalent at very early stage of deformation (at low stress) while the activation of others occurs at more advanced stages of straining. To illustrate more specifically, we provide Fig. 7, where a typical martensite (B19₀) load/unload curve (green shaded) followed by heating above A₁ (red curve) are shown along with the microstructure evolution. This graph and the inset schematics are constructed based on the existing literature evidence, which we will discuss in more details.

In NiTi martensite, three types of twins are prevalent, namely, Type I, Type II and compound twins [108–110]. The initial undeformed martensite (as having transformed from austenite by cooling below Mₜ) primarily consists of Type II and Type I twins (point “a”). The Bilby-Crocker theory [111] of twin classification is adopted in terms the so-called twin elements (planes K₁, K₂; directions η₁, η₂; shear magnitude, s). During deformation, one variant of Type II twin continues to grow...
in thickness at the expense of others, a process known as “detwinning” (points “b” and “c”). At higher applied stress, another frequent twinning mode in NiTi martensite is of \(\{001\}\) type [23], the so-called compound twin. This twin can undergo deformation-induced growth while the associated strains remain mostly recoverable (points “d” and “e”). At later stages of deformation, another compound twinning of type \(\{100\}\) is observed. Furthermore, \(\{201\}\) and \(\{113\}\) types of twins are also observed [112,113] (points “f” through “h”) at later stages of martensite deformation. It is important to note that numerous experimental studies have reported a substantial presence of dislocation activities in conjunction with twinning (typically at high strain levels, that is, beyond f) [26,114]. The presence of slip would essentially lead to irreversible deformation, which would be manifested in the form of residual strains upon unloading (and also heating above \(A_t\)). The role of slip on the deterioration of the shape memory effects is a pressing issue, which merits a separate volume. Due to limited scope of the current paper, interested readers are referred to the authors’ earlier publication dealing exclusively with slip in SMAs [42].

After unloading the martensite and subsequently heating it above \(A_t\), the strain is recovered by dint of transformation into an austenitic crystal. In the austenite phase, twins of types \(\{114\}\) and \(\{112\}\) [115–118] are reported. It is worth mentioning that the twins in B2 phase could be an outcome of plastic deformation as well as residual products originating from highly deformed martensite. In essence, these advanced twinning modes reportedly contribute to additional strain accommodation

---

**Fig. 9.** Arrangement of atoms on individual ledges is shown, which are part of a Type II twin boundary in the NiTi B19' lattice.

**Fig. 10.** (Left) Atomistic configuration of a single variant designated “B” (green → Ni and silver → Ti). (Right) The twinned structure between variants A and B. Arrows indicate shuffle of Ni atoms in creating reflective lattice. Notice the partial shuffle at the twin boundary, where all the atoms are coplanar.
when slip systems are unavailable. More specific discussions on the experimental discoveries and the relevant atomistic predictions follow.

3.3.2. Type II twinning in martensite (B19\(^0\)) phase

The (0 1 1) Type II twin is described in terms of the twinning elements as follows: \(K_1 = (0.7205 \bar{1} 1)\), \(\eta_1 = (0 1 1)\), \(K_2 = (0 1 1)\), \(\eta_2 = (1.5727 1 \bar{1})\) and a twinning shear of 0.2804 \(107\). The two variants (i.e. the matrix and the twin) have 180° rotational symmetry about [0 1 1] direction. Experimentally, the Type II type twins are most commonly observed for solutionized NiTi \([112,119]\). The study of Type II twins in NiTi martensite has remained a topic of considerable debate, and a general consensus regarding its unique interface morphology is yet to emerge \([120]\). Specifically, phenomenological models \([121–123]\) widely considers the twinning plane as irrational \([111]\) i.e. \(K_1 = (0.7205 \bar{1} 1)\), while high resolution TEM analyses \([124,125]\) suggest a rational yet stepped geometry. The origin of rather unusual boundary geometry is attributed to a need to minimize the total interfacial energy \([126]\). In Fig. 8, one such evidence is presented from \([125]\). We present the argument for stepped geometry in view of the pertinent atomic theories (to be discussed next). The interface between two variants (A and B) can be noticed as consisting of multiple ledges. Each ledge is essentially of the (1 1 1) type plane, and its width is on the order of several atomic diameters. As indicated earlier, the Type II twinning can experience detwinning process, where one of the variants is consumed while the other growing in thickness. For the case of the stepped geometry, individual ledges would undergo layer-by-layer growth during detwinning. During the process, it is intuitive that the material would seek the mechanism with minimum energy. Specifically, the DFT calculations by Ezaz et al. \([127]\) suggest that a combination of shuffle and shear can facilitate the twin growth in a most energetically favorable manner. The significance of combined shear and shuffle as a viable mechanism was noted in earlier twinning literature \([128,129]\).
Prediction of the energy pathway depends on accurate atomic arrangement on each ledge, which is illustrated in Fig. 9. Notice the orientation of the irrational plane, as widely used in the classical theory. Based on this geometry, the problem reduces to computing the energy density per unit area corresponding to layer-by-layer growth starting with a single ledge. Such a scenario is elucidated in Fig. 10 based on the approach adopted in [127]. On the left, five stacks of atoms on the (111/C221) plane are shown, which constitute variant “B”. The twinned structure is formed when layers are sheared and atoms are shuffled to a mirrored position to create another variant “A”. Ni atoms (green) are shuffled partially on the twin boundary itself (black). A physical rationale of shuffle has been proposed on the basis of electronic charge density (which is a measure of bonding strength) evolution during the twinning process. Fig. 11 (reconstructed based on data from [127]) shows interatomic charge density difference, $Dq$, between two atomic layers, one belonging to an undeformed variant B and another the boundary of a four layer twin. Notice the highly asymmetric nature of the $Dq$ variations among Ni-Ni, Ni-Ti and Ti-Ti pairs. The calculations indicate that Ti-Ti possesses a weak directional bonding while both Ni-Ni and Ni-Ti are essentially nondirectional. This scenario is particularly conducive to the movement of Ni atoms, which can squeeze through Ti atoms like soft spheres and eventually lower the overall structural energy. These results are particularly useful, in that they provide a strong physical argument for the genesis of shuffle during twinning. As will be discussed in next sections, the atomic shuffle in other twinning modes could also be originating from similar effects.

Upon accounting for both shear and shuffle, the computed twinning energy pathway as in Fig. 12 was computed [127]. Starting with the variant B (lower, blue), consecutive layers are rigidly displaced by increasing magnitudes (i.e. $u = 1b, 2b, 3b, 4b$) from the initial boundary. Consequently, the variant A (upper, red) continues to growth. It was found that the requisite

---

**Fig. 13.** Bright-field TEM images of various types of compound twins in the martensite (B19’) of NiTi [113].

**Fig. 14.** DFT-based twinning energy pathways for (100) and (001) compound twins. The atomistic configurations considered in the calculations are also shown (re-plotted based on [133]).
energy cost to add another layer to the existing twin is considerably smaller (14 mJ m\(^{-2}\)) compared to twin nucleation. The addition of new layers is essentially equivalent to the growth of one variant, which is in principle similar to the detwinning. It is concluded from these results that the shuffling facilitates the Type II detwinning by obviating partial slip gliding as precursor, which is normally the case in conventional non-transforming alloys [129]. This model essentially assumed the stepped geometry based on the earlier TEM finding, and is able to explain the ease of Type II detwinning. The low energy cost clarifies why this particular deformation mode is dominant during early martensite deformation. It remains to be seen how further independent research will assist in understanding the formative process of the boundary itself in the light of a similar energetic argument. Specifically, the ongoing debate regarding the interface geometry could benefit from further atomistic calculations.

3.3.3. Compound twinning in martensite (B19\(^{0}\)) phase

Typically, the stress plateau during the martensite deformation is attributed to the detwinning of Type II twins [107]. On the other hand, detailed microstructure studies revealed that the activation of (001) and (100) twinning modes, also knows the compound twins, in the deformed martensite can accommodate additional straining [23,112,113,130–132]. It should be mentioned here that the (001) and (100) twinning modes are not the same, despite the conjugate nature, owing to the asymmetry of the monoclinic B19\(^{0}\) lattice. One example of electron microscopy evidence is presented in Fig. 13 (from [113]). Notice the presence of various types of compound twins. Interestingly, another (201) type twins are also presented, which we cover in the next section. The presence of these twins assists in additional ductility of the deformed martensite. In addition, dislocation activities can also accompany the deformation at this stage [26,114]. It is important to note here that the preservation of shape memory effects entails twinning as the primary deformation mechanism while minimizing slip. In this regard, atomistic calculations have shed light into the energy barriers for the compound twins as well as slip, which clarified their relative propensity.

Fig. 14 presents a comparison between the twinning energy landscapes for the (001) and (100) modes (based on [133]). As in the inset crystal configurations indicate, these two deformation planes would result in completely different mirror symmetry upon shearing. Another important difference was noted between these two systems that the reflective symmetry for (100) case requires combined shear and shuffle while (001) twins can be created by shear alone. The energy profiles represented displacement of three layers. The computational results indicate that the (001) twinning would be preferred due to its lower fault energy. The twin migration energy costs (i.e. the energy for displacing one layer from a pre-twinned lattice) are 7.6 mJ m\(^{-2}\) and 41 mJ m\(^{-2}\) respectively for (001) and (100) twins. It is worth mentioning that the twinning shear magnitudes for both cases are 0.2385. However, due to asymmetry of B19\(^{0}\) lattice, the equal amount of shearing requires overcoming different levels of energy barriers.

In addition, for the same crystallographic systems (i.e. plane and direction), the energy barrier for slipping was also investigated. It was found that a high energy barrier (448 mJ m\(^{-2}\)) is required to initiate (100) slip. This magnitude considerably higher than the twinning energy, and hence precludes the possibility of the (100) slip nucleation [133]. However, only a 20 mJ m\(^{-2}\) of barrier magnitude is predicted for (001) slip, which is consistent with its experimentally-observed prevalence in the deformed martensite. This prediction explains the TEM observation of slip in the deformed microstructure in conjunction with compound twins. It is likely that the high internal stresses (stemming from a network of interface) might assist in overcoming the slip nucleation energy barrier [42].

3.3.4. (201) type twinning in B19\(^{0}\) martensite phase

The (201) type twinning can occur in conjunction with the (001) compound twinning. The twinning elements are: \(K_1 = (201), \eta_1 = [10 \bar{2}], K_2 = (001), \eta_2 = [100]\) and a twinning shear of 0.425 [107]. This mode is typically observed beyond the stress plateau (at a strain level >3%) [134,135]. A closer inspection on its interface is shown in Fig. 15 obtained

![Fig. 15. High-resolution TEM evidence of (201) type twin boundary in the B19\(^{0}\) lattice of NiTi martensite [113].](image-url)
from high resolution TEM (from [113]). Note the reflective symmetry about the twin plane. The presence of the $\{2\ 0\ 1\}$ twinning mode has been associated with additional ductility during the martensite deformation [134,136]. Moreover, the $\{2\ 0\ 1\}$ twins are observed to intercept other twins, giving rise to hardening benefits. The strengthening attribute combined with the additional straining capability ultimately has the effect of extending the utility of NiTi SMA. As will discuss in next section, the significance of $\{2\ 0\ 1\}$ twins is that they are retained as residues even after the material is reverse-transformed into austenite phase. While the morphology of the twin has been studied ex-situ, the mechanism of its formation is later elucidated using ab initio calculations.
DFT-based energy path calculations were undertaken [135] to examine the mechanism of (2 0 1)[1 0 2] twinning system via combined shear and shuffle (Fig. 16). The shuffles are reported to be bi-directional in nature as indicated in the inset figure. First, the formation of a three-layer twin is illustrated. The same mechanism would govern the twin migration process i.e. addition of the fourth layer. The predicted fault energy surface for the fourth layer of the twin starting from a three-layer one is presented alongside. Note the complex coupling of shear and shuffle occurs simultaneously during the twin growth. The shuffle of atoms occurs about their motif points (which is the middle point between the pair of atoms as indicated). The creation of the four-layer twin would occur along the minimum energy pathway (marked by the white dotted line). It follows that an energy barrier (i.e. the peak $\gamma$ magnitude) of 61 mJ m$^{-2}$ needs to be overcome to reach the metastable position (i.e. a local energy minimum). With continued shear and shuffle, another transition state is reached. Compared to the energy barriers during the (0 0 1) and (1 0 0) twin growth (discussed earlier), (2 0 1)[1 0 2] twinning possesses a higher magnitude.

From detailed DFT analyses of the monoclinic B19$^\prime$ atomic structure [135], it was noted that the mere shear of the atomic lattice is not sufficient to create mirror symmetry between the parent and the displaced crystals for the (2 0 1)[1 0 2] system. This is attributable to the non-cubic nature (i.e. lower symmetry) of the B19$^\prime$ martensite lattice. That is why additional movements of atoms (i.e. in the form of shuffle) are necessitated in conjunction with homogeneous shear. The role of shuffle is quite significant, in that it reduces the degree of required shear in order to achieve the desired mirror symmetry for nucleating the $f_{20}/c_{22}^1$ twin. As the DFT calculations confirm, a shear magnitude of 0.338 accompanied by atomic shuffles gives rise to an energetically stable twin embryo. One may recall that the phenomenological theories dictate a higher magnitude of shear i.e. 0.425 (without shuffle) [107,111,137,138] to create the (2 0 1) twins. In essence, the twinning shear in the atomic lattice is lowered as a result of shuffle-assisted compensation of the necessary atomic movements.

In addition, Ezaz et al. [135] reported an energy barrier of 1390 mJ m$^{-2}$ for slip belonging to the same crystallographic system. Due to significantly high magnitude, the possibility of slip was ruled out.

### 3.3.5. $f_{211}$ twinning in austenite (B2) phase

Most literature on the austenitic NiTi phase consisting of B2 lattice is centered on the discussion of phase transformation behavior. Nonetheless, there is experimental evidence of twins in the B2 phase. Their presence could be attributed to: (a) residual twins originating from transformation of highly deformed martensite, which contains higher order twins and (b) twinning-assisted straining that occurs at the latter stage of austenite deformation when requisite slip and/or transformation planes and directions are unavailable [115,116,118]. The bright-field TEM in Fig. 17(a) shows a mixed presence of B2 and B19$^\prime$ in NiTi specimen, which has undergone equal channel angular extrusion process [136]. The austenite is characterized by the presence (1 1 2) twins. The regions circled with “b” and “c” are where the diffraction patterns in Fig. 17(b) and (c) are taken, confirming the lattice types. From lattice correspondence relationship, it was suggested that the B2 (1 1 2) twin in fact could be a residual form of (1 1 3) type twin in the B19$^\prime$ martensite phase. The significance of (1 1 2) twins in B2 is related with plastic deformation of austenite in absence of dislocations. It is interesting to note that twinning is preferred as the plastic deformation mode rather than slip. The origin of such behavior is subsequently rationalized by a lower energy barrier to twinning than to slip as computed from first principles.

![Fig. 18](image_url)

**Fig. 18.** (Left) The twinning mechanisms based on combined shear and shuffle, where a four-layer twin in the NiTi austenite. (Right) Minimum energy pathway (highlighted by white lines) for the proposed twinning mechanism as a function of hydrostatic stress under shear- and shuffle-type atomic displacements (normalized by $\frac{\alpha}{\sqrt{2}}$ and $\frac{\alpha}{\sqrt{3}}$ respectively) [140].
Mechanistically, the (2 1 1) twinning was envisioned as the so-called pseudo-twinning type with shear occurring in the \( \frac{1}{2} [1 1 1] \) direction \([139]\) in the early literature. However, this mechanism was ruled out for ordered NiTi B2 lattices due to high energy cost (of 500 mJ m\(^{-2}\)) \([140,141]\). As a more feasible mechanism, the incorporation of shuffle in conjunction with shear was proposed \([142]\). Subsequently, detailed calculation on the energy pathways were undertaken to establish the precise extent of shear and shuffle considering \( \{2 1 1\} [1 1 1] \) twinning system \([140]\). In Fig. 18, the mechanism is illustrated along with its energetics in the presence and absence of hydrostatic pressure (adapted from \([140]\)). The amount of twinning shear is \( \frac{1}{\sqrt{2}} \approx 0.707 \) for this system. The shown energy contours (plotted against normalized shear and shuffle) correspond to the creation of the fifth layer from a four-layer twin. It is noted that shear and shuffle of atoms are co-planar as shown by arrows. In this case, the neighboring distance between atoms is an important parameter, which dictates the resistance of shuffling. The white lines represent the minimum energy paths. Note that in order for the twin migration to occur with least applied force the shear and shuffle ought to occur simultaneously. The energy barrier under zero hydrostatic stress was reported to be 79 mJ m\(^{-2}\). The role of hydrostatic stress is examined since it can affect SMA deformation behavior profoundly \([143]\). From the DFT energy pathways, it follows that the magnitude of the energy barrier decreases by 12% under tension, and increases by 19% under compression. More, the degree of shear and shuffle are also modified as a result of hydrostatic forces as the location of the transition state (where the energy is the maximum on the minimum energy pathway) is shifted in the contour plots.

3.3.6. \( \{1 1 4\} \) twinning in B2 austenitic phase

In the austenitic phase of polycrystalline microstructure, irreversible (plastic) deformation can occur in the form of slip or twinning, depending on the orientations of grains. Since a minimum of five independent active straining systems are required for arbitrary deformations, the origin of large strain accommodation can be traced back to the presence of additional twin systems along with slip. A number of research groups have confirmed the existence of the \( \{1 1 4\} \) twin system \([115,116,138,142,144–149]\), a result of irreversible (plastic) deformation of B2 austenite. Interestingly, the manifestation of
twinning can also occur as the residual product of \{201\} twin from the B19\textsuperscript{0} martensite phase. Fig. 19 presents high resolution TEM evidence of the direct correspondence between these twinning modes \cite{138,150}. The region covered in the TEM image contains both B2 and B19\textsuperscript{0} lattices. Notice the continuity of the twinning planes (K\textsubscript{1}). The B2 portion of the lattice has originated from the reverse transformation of the B19\textsuperscript{0} martensite. DFT calculations have elucidated the mechanism of \{114\} twinning, and quantified the associated energy pathways.

Atomic scale simulations revealed that in addition to combined shear and shuffle, an interface shift may contribute to the nucleation of \{114\} twin \cite{151}. We illustrate the mechanism in Fig. 20 and also present the associated energy pathway as a function of shear and shuffle (normalized) (based on data from \cite{151}). Beginning with a four-layer (114) twin, four more consecutive layers are first rigidly displaced (shear). The shuffle occurs unidirectionally on every second plane as shown. The combined shear and shuffle then culminate in an eight-layer twin. The minimum energy pathway is highlighted with dotted lines. The “TS” denotes the transition state where the fault energy (\gamma) is the maximum (148 mJ m\textsuperscript{-2}). A further lowering of the total structural energy was noted by allowing a boundary shift as depicted. It should be noted that the process of shuffle creates the reflective mirror symmetry which is not destroyed by the boundary shift. Only the atoms belonging to the twin boundaries (upper and lower) are displaced resulting in an energetically favorable structure. In addition, the feasibility of slipping was also examined by Ezaz et al. \cite{151} with a predicted energy barrier of 2681 mJ m\textsuperscript{-2}, which rules out the possibility of slip and further establishes the preference for twinning instead. These findings, in particular the relatively small energy barrier (148 mJ m\textsuperscript{-2}), provides a rationale for the occurrence of \{114\} twinning as the preferred mode of plastic straining. As reported earlier, the lowest possible energy barrier for slip is 142 mJ m\textsuperscript{-2} \cite{42,152}; thus, the \{114\} twinning systems becomes the next available plastic deformation system.

Overall, the foregoing computational results can be useful to understand the irreversibility of SMA deformation as well as its hysteresis. For instance, when austenite-to-martensite transformation occurs, a certain volume fraction of austenite is likely to be untransformed, which in turn may undergo plastic deformation. A competition between available slip and twinning systems would ensue. From the atomistic assessment, one can examine how to lower the relevant energy barriers to create a preference for a certain deformation mode, say, via alloying with other elements. Depending on the activation slip or twinning, the irreversibility and hysteresis would differ.

3.4. Molecular dynamics models

3.4.1. Early endeavors on empirical potential

In the last decade, seminal MD works were conducted on NiTi, which eventually set the course for the present state of the art. It is instructive to recount the substantiality of their contributions here. Noting that an EAM (embedded atom method) potential is mostly as good as the curve-fitted materials properties, Farkas et al. \cite{153} incorporated experimental and DFT based fault energies of NiTi B2 therein. This work particularly brought forth the challenges associated with reproducing Fig. 21. Stress-strain response of NiTi martensite (nano-sized and at 1 K temperature) of various compositions \cite{62}. Atomic snapshots taken at different stages of deformation demonstrate how a multi-variant martensite, when stressed continually, becomes a single variant (which remains so after unloading). Heating the single-variant crystal results in the original multi-variant structure.
the stability of multiple phases (e.g. B2, B19') simultaneously. Considering high temperature single phase data, they were able to capture the cohesive energy (4.93 eV/atom) and lattice constant of stable B2 NiTi \(a_{B2} = 3 \text{ Å}\) comparable to experimental values (4.95 eV/atom and 3.01 Å respectively [154]). The issue of predicting more than one phase using a single potential was addressed next.

The crystalline-to-amorphous transition behavior subjected to ion irradiation was first studied by Lai and Liu [67,155] using a new EAM model. They succeeded in establishing the correct inter-species bonding functionals, which enabled predictions of solid solubilities of disordered Ni-Ti intermetallic compounds. Here it should be noted that the shape recovery properties of Ni-Ti alloys is observed only for the composition in the neighborhood of equiatomic stoichiometry. By way of exploring the non-stoichiometric amorphization behavior, the necessary incentive for further theorization of the NiTi SMA behaviors was essentially set in motion. Kastner and co-workers [9,61,156] conducted extensive studies on the phase transformation behavior of using Lennard-Jones potential (which provides generic pair-wise interatomic interactions excluding the effects of embedding atoms into pervasive electronic cloud). Sato et al. [157] later employed Lai and Liu’s potential, and predicted multiple transformation pathways between B2 and B19' phases for equiatomic composition. Concurrently, Ishida and Hiwatari [65] considered angular dependency of bonds with a modified EAM (MEAM) approach. They conducted a quantitative analysis of thermally induced reversible B2 ↔ B19' transformation. These works essentially emphasized the need for more robust modeling ability for analyzing detailed crystallography of dynamic phase evolution, which came next.

Fig. 22. (Top) Molecular dynamics superelastic stress-strain response of NiTi nano-wire at 400 K with a transformation path (during loading) of B2 → B19 → B19' [63]. (Middle) Superelasticity at 350 K characterized by a loading path: B2 → B19 → BCO → BCO twin. (Bottom) Strain recovery via shape memory effect (i.e. by unloading, heating and cooling). Notice the formation of twinned regions as local bands.
3.4.2. Recent developments

The interatomic energy functional in the Lai and Liu’s original potential was formulated as an exponential expression with a preset cutoff radius of 4.2 Å. Other research groups introduced cut-off functions instead thereby mitigating the diverging forces, which caused undesired large displacements of atoms [62,68,158,159]. The outcomes of such adaptation were remarkable in terms of predicting phase stability and transition properties. For instance, Mutter and Nielaba [158,160] first succeeded in predicting the temperature-driven two-way phase transformations and the hysteresis of total structural energy. More recently, they simulated the application of stress on a nano-sized multi-variant (i.e. twinned) martensitic crystal (B19\textsuperscript{0}), which resulted in a single variant B19\textsuperscript{0} via de-twinning process [62]. Subsequently, heating was found to induce twinning and (followed by cooling) a complete reversion to the initial configuration with full strain recovery. The entire process is shown in Fig. 21. The unusual circumvention of commonly observed austenite (B2) formation was attributed to a smaller energy barrier for twinning than for the B19\textsuperscript{0}-to-B2 conversion as inherent in the adapted potential. It is noteworthy that in the Mutter and Nielaba model, only B2 and B19\textsuperscript{0} phases were stabilized and observed during the structural transitions. One can compare the predicted results with experimental behaviors, where twinned martensite undergoes detwinning when stressed [161]. From the TEM evidence, it is noticed that the original martensite first reorients and then de-twins to become a single variant. It was noted that at higher strains, mono-variant martensite might undergo plastic deformation (via slipping). The predictions and empirical observations altogether pose interesting questions regarding the apparent exclusiveness of twinning and transformation, which may be clarified upon further research.

Intriguingly, some recent literature noted the occurrence of transient structures such as B19, B33 (i.e. body centered orthorhombic, BCO) during NiTi nanowire deformation using a similarly modified version of Lai and Liu’s potential.
Fig. 22 presents three distinct situations of NiTi nanowire deformation from Zhong et al. [63,68]. In the first case (topmost), the superelastic behavior at 400 K is related with a transformation path: B2 \to B19 \to B19', and its complete reversal upon unloading. Superelasticity is also found to occur via B2 \to B19 \to BCO \to twinned BCO at 350 K. It was noted that at higher temperature (450 K), the B19' lattice forms prior to the BCO phase. The same transformation route was found to result in the shape memory type recovery (i.e. additionally aided by heating) due to the formation of "macroscopic twin-like" deformation of the entire nanowire (as evident from top view). Another interesting point to consider is the nature of MD based stress-strain responses. Typical to MD stress-strain curves is an abrupt stress drop when elastic deformation is followed by a single nucleation event of highly localized nature with no effective obstruction thereto. This is the case in the Zhong et al.'s model.

A hardening mechanism would result when multiple events (e.g. many isolated transformed regions) interact and thus resist each other. Therefore, B2-to-B19' conversion in multiple regions would preclude the descent of stress. Such observations are noted by Chowdhury et al. [64,93] (Fig. 23) as well as Ko et al. (Fig. 24) [66]. In the former case, the original austenite (B2) is compressed along (0 1 1) direction, resulting in a twinned martensite consisting of B19'. An absence of stress drop can be attributed to wide-spread transformation events (as opposed to a localized band). Ko et al., using a modified potential (MEAM) predicted martensitic transformation giving rise to hardening behaviors. The stress-strain response and corresponding atomic snapshots are shown in Fig. 24. A visualization of atoms via common neighbor analysis clarified how the B19' martensite (dark brown) nucleates in the B2 matrix (light blue). The superelastic behavior (simulated) is associated with a reversible B2 \leftrightarrow B19' transformation. In addition, Ko et al. [163] also predicted stable stress plateau in nanocrystalline NiTi.

It is noteworthy that Zhong’s model could capture intermediate structures with the sequence: B2 \to B19 \to B19' \to B33 consistent with DFT predictions. As explained earlier section, the B19', despite not being the ground state martensite, would be dominated by B33 in pristine defect-free condition i.e. in absence of internal stresses. Due to localization of transformed material, no additional resistance was present in Zhong’s case, thus facilitating the existence of B19 and B33. This idea is further reinforced by works by Ko et al. and Chowdhury et al. where only B2 \leftrightarrow B19' was captured. Concurrent nucleation of multiple twinned regions presumably generated sufficiently large stress-fields, thereby ensuring the dominance of B19'.

It is also worth mentioning here that intermediate crystals in binary NiTi, although quantum mechanically predicted along the transformation paths, are not manifested experimentally. Only through additional stimuli (e.g. alloying with Cu or Fe, precipitation) can some of the transition states be stabilized. For instance, alloying with ternary elements such as Cu [164,165] and Fe [166] makes the B19 or R phases experimentally observable. That these intermediary structures occur during deformation simulations under pristine conditions provides new mechanistic insight into their stabilization condition. It remains to be seen if new experiments particularly on single crystals can generate further evidence therein.

**Fig. 25.** The presence of a nano-sized coherent precipitate in austenitic NiTi matrix creates unique internal stress distribution [69]. The [1 1 0] compressive stress-strain response of a precipitated single crystal is dictated by activation of a complete different set of variants (lower insets) compared to the un-precipitated material.
3.4.3. Promising developments

Other promising utilization of MD simulations is manifested recently, for example, in uncovering the role of coherent precipitates [69,157]. Suitable aging treatments can generate precipitated microstructure, which is empirically associated with lower transformation stress, strain and hysteresis. Several intricate effects of precipitation are experimentally identified, namely, (a) creation of disturbance field of precipitate, which alters the variant preference locally, and (b) depletion of Ni

Fig. 26. (Upper left) TEM image showing two parallel precipitates in NiTi austenite 40 nm apart; (upper right) a highly non-linear type strain profile between the two precipitates measured from experiments (along the white arrow from the left inset) [167]. (Bottom) TEM evidence of localization of martensite plates near precipitate with a clarifying schematic [168].

Fig. 27. (a) The twinned structure in the Ni-Mn-Ga martensite [184,185] (color contrast among the variants achieved through cross-polarization), (b) the variants of martensitic phase [188] and (c) magnetic domains in the martensite phase colored according to magnetization with the arrows indicating domains of similar direction (adapted from [188]).

3.4.3. Promising developments

Other promising utilization of MD simulations is manifested recently, for example, in uncovering the role of coherent precipitates [69,157]. Suitable aging treatments can generate precipitated microstructure, which is empirically associated with lower transformation stress, strain and hysteresis. Several intricate effects of precipitation are experimentally identified, namely, (a) creation of disturbance field of precipitate, which alters the variant preference locally, and (b) depletion of Ni
from surrounding matrix, which forms the Ni-rich Ni$_4$Ti$_3$ particle. Prediction of the former effect has recently been undertaken by Chowdhury et al. (Fig. 25). A disturbance field of non-linearly decaying nature has been predicted as a direct outcome of the inter-lattice misfit at the periphery of matrix and the particle. This in turn was found to set local preference for variant nucleation, which is different than that of the un-precipitated lattice.

The results of the model can be directly compared to empirical evidence. As shown in the upper inset of Fig. 26, the measured strain field (with high resolution TEM) between two precipitate is of highly non-linear nature, which is consistent with the prediction [167]. The consequence of such distortion near the precipitate has been predicted to be a cause for preferred variant nucleation. This can be confirmed from TEM microscopy (bottom inset) where a prevalence of martensite plate can be noticed near the precipitate [168]. In this model, the effects related to the composition gradient near the particle were not incorporated. It can be conjectured that the neighboring materials with reduced Ni would possess less transformability. Furthermore, it is arguable that, unlike the simulated single crystal, in reality, there would be multitudes of precipitates with overlapping network of disturbance fields. With the success of the ideal situation of a single precipitate case, these issues pose promising future endeavors.

4. Case study: ferromagnetic Ni$_2$MnGa SMA

4.1. Empirical observation of structure-property relations

Ni$_2$MnGa alloys are the most promising candidates as magnetic SMAs [169–171] and hence the most widely researched ones. This material possesses the unique attribute of deformation recovery via applied magnetic stimuli in addition to thermomechanical ones. These SMAs are also very composition-sensitive. Specifically, addition of the species Mn has a strong effect on the local magnetic moment [172,173]. The shape-recovery properties in these alloys originate from a complex synergy of phase transformation, magnetic moment ordering, and deformation twinning [174]. Under applied magnetizing fields, the twin boundary motion can occur driven by the impetus of internal magnetic moments to align with the applied one. In conjunction with reversible martensitic transformation, the variant growth via twin migration process contributes to the unique deformation healing characteristics. Microstructurally, the austenite is of ordered face centered cubic lattice (L2$_1$) [175,176] while the martensite can be tetragonal or orthorhombic, which can also undergo modulation into the so-called 5M and 7M structures [177–179]. For instance, atoms can reshuffle with a period of 5 and 7 lattice parameters giving rise to 5M and 7M structures (which are also denoted as 10M and 14M respectively by doubling the periodicity). “M” indicates monoclinicity. The modulated martensite phases are of lower structural energy [180–182] (as predicted) and, further promote twin boundary mobility. The phase transformation sequence is as follows [183]: L2$_1$ → 10M → 14M → L1$_a$.

It is important to understand the microstructural processes underlying the unique deformation propensity. In Fig. 27(a), the twinned structure in the Ni-Mn-Ga martensite can be seen [184,185] (color contrast among the variants achieved through cross-polarization). Differences among the variants in terms of magnetic domain fields make them sensitive to applied magnetizing fields, which in turn generate driving forces necessary for twin boundary movement. Using Fresnel (Lorentz) microscopy, magnetic domains can be characterized [186,187]. The distribution of microscopic magnetization domains helps understand the origin of one variant tending to grow at the expense of others. In Fig. 27(b), for instance, the reconstructed herring-bone morphology of martensitic phase is shown [188]. The dark and gray areas represent different variants. The colored plot in Fig. 27(c) indicates various magnetic domains (inset color wheel indicating magnetization directions). Areas identical in coloration possess the same magnetization direction as indicated by the arrows. Under the application of external magnetic fields, the magnetic moment of each variant would be forced to re-orient. Consequently, mechanical forces would develop leading ultimately to detwinning process. Specifically, the modulated martensite structure has been attributed to an enhanced ease of magnetically-stimulated interfacial movement. Fig. 27(d) provides evidence of layer-by-layer modulated structure based on high resolution electron microscopy [189]. Due to its technological importance, the Ni$_2$MnGa SMSs have remained an interesting topic both for phenomenological and atomistic studies [190–192]. DFT-based computations have uncovered the electronic properties underpinning the observed phenomenology of Ni$_2$MnGa deformation as discussed next.

4.2. Atomistic investigation from first principles

A number of DFT-based studies have been undertaken to elucidate various aspects of sub-lattice effects in Ni$_2$MnGa SMAs [193–202]. Topics of interests include phonon dispersion properties, structural stabilities, determination of equilibrium lattice constants, Fermi surface, energy pathways, magnetism, composition effects, etc., to name a few. The DFT literature in this regard is quite substantial. For instance, Uijtewaal et al. [194] studied the temperature-dependence of free energy by accounting for interplay between the vibrational (phonon) and the magnetic (magnon) excitations spanning austenite to martensite (with pre-martensitic structure in-between) phases. Kart et al. [203] reported elastic moduli of austenite, modulated and non-modulated martensite phases. The DFT results based on single crystal configuration has also been extended towards understanding magnetocaloric effects in polycrystalline structures [204]. Similarly, important results regarding martensite have also been obtained from first principles. We discuss in detail specific case studies (below), exhibiting the efficacy of DFT simulations in uncovering Ni$_2$MnGa deformation behavior.
Fig. 28. (Top left), The martensite crystal configurations (NN, 10 and 14M) are shown (adapted from [205]). (Top right) Total energy profile as a function of
tetragonality [205,206]. (Bottom) Total energy contour as a function of c and a (right figure is a closer inspection of where austenite (A) and martensite 
structures reside) (adapted from [209]).

Fig. 29. Step-by-step shearing of L1₀ martensite in stoichiometric Ni₂MnGa and the associated changes in the energy and magnetization (adapted from 
[193]).
4.2.1. Energetics-tetragonality relationships in austenite and martensite phases

The shape recovery of Ni$_2$MnGa is achieved via the reversible transformation between the ordered L2$_1$ austenite and the L1$_o$ martensite phases. The lattice structure of the L1$_o$ martensite could be non-modulated or modulated with a certain periodicity (10M and 14M). In Fig. 28 (top left), the martensite crystal configurations are shown (adapted from [205]). The modulation gives rise to nano-twinned structure with reflective symmetry of atomic arrangements. The number of stacked planes differs for the 10M and 14M cases. The non-modulated (NM) L1$_o$ unit cell has the perfect tetragonal structure (the lattice parameters, a, b and c are shown). The total energy profile (black) as a function of tetragonality (i.e. c/a ratio) indicates the lowering of energy state (with local saddle points) for the martensitic transformation into a non-modulated structure [205,206]. The austenite phase is reportedly stable at c/a = 1, while the NM L1$_o$ martensite at c/a ≈ 1.26. Notice that for c/a < 1, the energy upsurges monotonically, indicating increasing instability of corresponding structures. Through a number of DFT Investigations, the trend that the martensite phase possesses a lower energy (by an amount of 20–30 meV per formula unit) is confirmed [181,193,201,207,208]. Notice the presence of a small energy barrier between the austenitic and the martensitic grounds states (i.e. the local energy minima). It follows that the modulated configurations are also energetically close to the non-modulated one, however, with reduced tetragonality (i.e. c/a ratio). On the other hand, some DFT reports have noted austenite to martensite transformation while preserving volume and at a c/a ratio less than unity [209]. The lower insets of Fig. 28 presents the contour plot of total energy as a function of lattice constants (in atomic units) with two lines highlighting constant volume and constant c/a ratio (at 0.97). The zoomed view (right) shows that the austenite and martensite reside at energy minima along the constant volume line with martensitic c/a being 0.97. Earlier, quite a many experimental studies have reported a c/a ratio greater than unity for stressed and non-stoichiometric alloys using neutron powder and X-ray diffraction methods [24,172,210–212]. Further studies in future particularly addressing the sensitivity of triaxiality on the energy values could reconcile these apparent disparities. Overall, the results indicate that the austenite-to-martensite transformation can occur from L2$_1$ lattice to 10M or 14M structure, without invoking much monoclinic distortion. The NM structure can result from the layer-by-layer coarsening of the nanotwins. The inter-conversion between NM and modulated structures have been noted in experimental studies [213,214]. Other DFT studies have addressed the energetics of twin boundary motion (as discussed next section).

Fig. 30. (Upper left) Superelastic stress–strain curve (schematic) of Ni-Fe-Ga with multiple stress plateaus [222]; first plateau is due to the transition from austenite (L2$_1$) to modulated martensite (10M and 14M) while the second one is associated with transformation into tetragonal (L1$_o$) martensite. TEM evidence of each phase (10M, 14M and L1$_o$ [227]) are shown.
4.2.2. Shearing energetics of martensite variants

Ease of twin boundary motion is very important for magnetically induced strain recovery [215–217]. Variants on either side of the twin boundary (i.e. matrix and twinned lattices) would have different magnetic anisotropy, which under applied magnetizing field would attempt to align with it. Entel and co-workers have predicted the energetics thereof via DFT simulations. Considering stoichiometric Ni$_2$MnGa, DFT simulations established L1$_o$ as the martensitic ground state [218,219], which is found to lie at the absolute minimum of energy landscape (c/a ratio tending to 1.25). In Fig. 29, the DFT results on shearing of non-modulated L1$_o$ martensite are presented (adapted from [193]). Upon establishing the basic twinned structure (i.e. consisting of two variants), rigid shear displacement is applied incrementally, which is accompanied by internal relaxation at each step. The associated changes in the magnetic properties (in terms of electronic spin contributions) are also provided. It follows that increasing degree of shear encounters monotonically increasing energy cost until reaching a peak. Upon overcoming the energy barrier, the shear-displaced structure arrives at an even lower energy state (after >11% shear strain). The crystal structure snapshots (with a closer inspection near the twin boundary) at four different points show the degree of shear and the corresponding changes in the atomic positions. The magnetization (per formula unit) reportedly continued to increase with shear displacement. It was concluded that the simple coherent motion of the interface induces a high energy barrier to overcome. The process could be made energetically more favorable by incorporating dislocation- and/or disconnection-assisted migration [220,221], resulting in easily-occurring variant growth. Overall, a large magnetic anisotropy alongside a low energy barrier for twin boundary motion is desirable for recoverable straining as concluded. Twinning energy cost is required to be low for modulated martensite, a problem which remains to be explored using ab initio calculations.

5. Case study: ferromagnetic Ni$_2$FeGa SMA

5.1. A rationale for martensite modulation

Similar to Ni$_2$MnGa, the case of Ni$_2$FeGa presents itself as a unique conduit for research. The most noteworthy feature of this material’s constitutive response is the presence of dual stress plateaus in its pseudoelastic curves as presented schematically in Fig. 30 [222]. The origin thereof is attributed to the existence of multiple distinct stages of phase transformation [223–229]: L2$_1$ → 10M → 14M → L1$_o$. The most crucial phenomenon there has been identified as the modulation of monoclinic martensite, which occurs during the very first step. This stage involves the cubic austenitic structure (L2$_1$) converting to a modulated monoclinic structure (10M). This is a significant intermediate juncture, in that the early modulation facilitates subsequent transformation to the 14M, and then finally to tetragonal martensite (L1$_o$). The modulation process involves...
atoms undergoing complex shuffle-type motions, while distorting the crystal volumetrically, to ultimately reach the 10M phase.

The austenitic phase of this alloy is of L2$_1$ lattice type (Fig. 31). A unit cell of L2$_1$ can be thought of an ordered conglomeration of eight B2 cells. During the martensitic transformation of the L2$_1$, the lattice structure first becomes monoclinic with an angle of 91.49°. As the DFT calculations suggest, a simple volume distortion (via rigid shear-type atomic displacements i.e. the Bain strain) results in a ten-layer non-modulated structure (10NM). This structure does not constitute the energy ground state. Thus, the inherent energetic instability of the 10NM lattice generates an impetus for atoms to undergo further shuffle-type movements to seek the lowest energy configuration. This notion can be verified by studying the associated energy landscape [230]. The energy landscape is computed by considering multiple possible pathways of L2$_1$ → 10M by exploring all combinations of distortion and shuffle. Thus, the least resistance path i.e. the minimum energy one is established (highlighted with a dotted line). It thus follows that the least resistance path i.e. the minimum energy route in nature would be achieved through the simultaneous occurrence of shear and shuffle. It remains an interesting future endeavor to predict the rest of the transformation pathway i.e. 10M → 14M → L1$_0$.

5.2. Implications for deformation reversibility

The predicted energetics bears important implications for the Ni-Fe-Ga SMA properties. From the atomistic standpoint, the reversibility of phase transformation in SMAs depends on several factors, namely, (a) low energy barriers for both forward and reverse transformations, and (b) high energy barrier for irreversible deformations such as dislocation slip. The extent of strain recovery as well as the associated hysteresis is affected by these variables. In that regard, the aforementioned calculations suggest that there exists a rather small barrier 8.5 mJ m$^{-2}$ at the early state of shear and shuffle in Ni$_2$FeGa. The corresponding lattice structure exists as a transition state (marked as ‘TS’ in Fig. 31). This explains why the transformation stress is also considerably low (<50 MPa), which makes the two-way process occur with least resistance. However, it is not sufficient for the parent lattice to be able to only transform. It is crucial to ensure the absence of factors that might hamper the two-way conversion e.g. considerable slip activities. An enhanced slipping propensity may adversely affect the reversibility of transformation. The plastic resistance ought to be substantially high to retain the reversibility between the parent and the transformed phases. Hand-in-hand with the transformation energetics, the slipping tendency can also be assessed. As with the case of Ni$_2$FeGa, the slip energy barrier was found significantly high, thus ensuring its superior transformability.

5.3. Devising potential strategies for alloy improvement

Further insight can be gleaned from examining certain simulation pre-conditions for the foregoing energy path calculations, which can suggest new possibilities of improved alloy. One major factor is the ordered structure of the SMA lattice, which demonstrably decides the thermal hysteresis in a range of SMAs. For example, the Fe–C or Fe–Ni steels (with no long-range order) are known for their very large thermal hysteresis (400°C). Also, Fe-Ni-Co-Ti and Fe-Mn-Si SMAs having no long-range ordering exhibit high hysteresis levels (200°C). On the other hand, the ordered Fe–Pt shape memory alloys possess hysteresis levels as low as 10°C. Similarly NiTiX (X = Cu, Fe) type alloys with long-range order (B2 type) have hysteresis typically on the order of 10°C to 20°C. In Ni$_2$FeGa, where the austenite phase is long-range ordered cubic L2$_1$ and the martensite is monoclinic 10M, the thermal hysteresis is reported quite low i.e. 1°C. It is important to note the fact that the low transformation barrier alongside the least resistance path is found under the ordered lattice condition. This essentially

![Fig. 32](image-url) MD simulations of single crystals of Co-Ni-Al SMA [94,233]. (Left) Simulated heating and cooling induce a hysteresis of total energy (potential and kinetic) as the material undergoes a reversible transformation between a B2 austenite and an L1$_0$ martensite during the thermal cycle. The origin of the hysteresis is attributed to the twinning and detwinning processes. (Right) Above the austenite finish temperature (A$_f$), a distinct pseudoelastic constitutive response is noted when the single crystal is compressed isothermally along [100] direction, and then unloaded.
suggests that one can explore possibilities of ordering a certain lattice on a hypothetical yet educated basis to investigate the associated energetics. This would in turn provide important information regarding the novel alloy composition.

6. Case study: Co-Ni-Al SMA

The Co-Ni-Al SMAs are known for their narrow thermal hysteresis (<30 °C) and superior ferromagnetic properties [231]. Experimentally, the strain recovering attributes of Co-Ni-Al SMAs are found to be related with reversible B2-to-L1₂ (i.e. cubic to tetragonal) phase transformations upon thermomechanical loading and unloading [232]. An atomistic framework for studying the underlying phase transformation could be particularly useful to rationalize its empirical behaviors. To that end, notable achievements are reported in the literature. Recently, Mishin, Yamakov and co-workers proposed a MD potential for Co-Ni-Al SMAs, fitted with experimental and DFT data, which captures the inter-species bonding landscape accurately [94,233]. As a result, reversible thermal hysteresis and superelastic strain recovery were predicted in agreement with empirical observations. It is interesting to note that the EAM potential itself was not geared for transformation. Nonetheless, the thermal and/or mechanical forces gave rise to the reversible transformation in any case. This could be an outcome of higher energy cost for plastic deformation than the phase transformation.

In Fig. 32, the simulated energy hysteresis upon cooling and heating is presented. At high temperature, the material is of B2 austenite type. This structure is created via Monte Carlo equilibration of the parent crystal. The total energy of the single crystal was tracked during the temperature cycling. The origin of the hysteresis is attributed to first twinning during incremental cooling followed by a completely reversible detwinning process.

On the other hand, distinct pseudoelastic stress-strain response was predicted under tensile and compressive loads. Several interesting physical phenomena were noted in these simulations. The material showed a marked tension-compression asymmetry as demonstrated by loading the single crystal along the [1 1 0] and [0 0 1] crystallographic directions. In both instances, the deformed material fully recovered the strain. The structural evolution that brought about such recovery was rooted on the reversible transformation between the original B2 and the internally twinned L1₂ lattices. The identification of different lattice types was conducted through common neighbor analysis (Fig. 33).

Incidentally, it might be relevant to recall recent computational works (based on first principles) on a similar ferromagnetic SMA (Co-Ni-Ga), by Arróyave and co-workers [234–236]. They examined the stability of the austenite phase as partial ordered B2 or fully ordered L2₁ structure. They reported that an increased magnetism results in a decreased energy differential between the austenite and martensite. Although MD potential does not address the magnetism, as in the case of Co-Ni-Al, a similar effect is expected to be inherent in this material too.

7. Case study: Ti-Nb based SMAs

Ti-Nb based SMAs were developed as a result of the need to eliminate Ni-related toxicity in orthopedic applications arising from the usage of NiTi SMAs [237–240]. Given the superior biocompatibility, the study of shape recovering attributes and the underlying transformation phenomena in Ti-Nb alloys are pursued rigorously through experiments. The nature of the martensitic transformation process for these materials is found to be a strong function of the alloy content [241,242]. By alloying with a third species (e.g. Ta, Zr, Al, Sn, Mo, N) to various degrees, the further property modulation (e.g. stress

![Fig. 33. Atomistic snapshot of B2 austenite at 700 K and L1₂ martensite at 200 K (simulated deformation of Co-Ni-Al SMA using molecular dynamics) [94,233].](image-url)
hysteresis, plastic strength, transformation strain) can be achieved [243–247]. One unique feature of Ti-Nb based alloys is that depending on composition, the parent austenite of disordered bcc structure (β phase at high temperature) can transform into an hcp (α' phase) for low Nb content or a BCO lattice (α00 phase) for higher Nb concentration. Thermodynamically, the bcc β phase and another hcp structure (α phase) are the equilibrium phases while the α' and α00 phases are metastable. Strain recovery attributes have been associated with the reversible transformation between the β phase and the α00 phase [248]. It is empirically found that alloys with Nb content less than 7.2 at% (approximately) typically undergo bcc-to-hcp transformation [249,250]. From first principles, the addition of more Nb to the base Ti lattice has been related with a modification to the electron charge density [251], to which the propensity for a certain phase is attributed. These findings essentially motivate further research geared towards isolating the specific role of individual alloying chemical species. One of the important steps would be a detailed analysis of the transformation mechanism.

7.1. Atomistic mechanism of transformation

Using molecular dynamics simulations, the hcp-to-bcc transformation process in Ti-Nb SMA is investigated in details by Yang et al. [252]. A summary of the simulated mechanism is presented in Fig. 34. It was found that the evolution of phase occurs according to the following order: initial hcp → intermediate BCO → intermediate face-centered orthorhombic (FCO) → final bcc. Moreover, it is noted that the internal twinning of martensite is initiated at the stage of FCO phase formation. It is quite interesting to note that BCO phase occurs as an intermediate step, which according to the experimental observations is in fact the predominant martensite. Since the MD simulations are conducted considering pristine lattice, it is possible that stabilization of the orthorhombic structure in nature may be a result of additional stimuli. No direct evidence exists to support such a notion; nonetheless, these results motivate further clarifications. For instance, as one possibility, alloy content may impart preference for the martensite lattice type. A DFT-based energetic analysis could provide further quantitative substantiation, which poses a promising research.
Similarly, the mechanism of the reverse transformation process (i.e. bcc-to-hcp) can be understood by revisiting the concept forwarded by Burgers [253]. The importance of the mechanism lingers to-date due to the potential application of DFT calculations to energetically assess the possible atomic routes. We illustrate the mechanism with the aid of Fig. 35 given its significance in understanding Ti-Nb transformation mechanism. Co-planar atoms are colored the same; for example, silver atoms are situated on every alternate $\{011\}_{bcc}$ plane while the orange ones belong to parallel alternate plane of the same family. The whole transformation process can be understood as: (i) a collinear shuffle of (silver) atoms along $h_{110}$ direction on every alternate $\{011\}_{bcc}$ plane and (ii) a concurrent contraction and dilation of alternate $\{011\}_{bcc}$ planes along $h_{001}$ and $h_{110}$ directions respectively. As uncovered in Fig. 35.

Similarly, the mechanism of the reverse transformation process (i.e. bcc-to-hcp) can be understood by revisiting the concept forwarded by Burgers [253]. The importance of the mechanism lingers to-date due to the potential application of DFT calculations to energetically assess the possible atomic routes. We illustrate the mechanism with the aid of Fig. 35 given its significance in understanding Ti-Nb transformation mechanism. Co-planar atoms are colored the same; for example, silver atoms are situated on every alternate $\{011\}_{bcc}$ plane while the orange ones belong to parallel alternate plane of the same family. The whole transformation process can be understood as: (i) a collinear shuffle of (silver) atoms along $h_{110}$ direction (indicated by red arrows), contraction and dilatation of atoms on adjacent parallel $\{011\}$ plane along perpendicular directions, $h_{001}$ and $h_{110}$ respectively (based on [253]).
molecular dynamics simulations, slip dissociation and their subsequent motion can facilitate the formation of hcp nucleus [254,255]. A 1/2(1 1 1) type slip can dissociate into three partials as follows:

$$\frac{1}{2}[1 1 1] \rightarrow \frac{1}{8}[0 1 1] + \frac{1}{8}[1 1 1] + \frac{1}{4}[2 1 1]$$

In particular, the 1/8(0 1 1) type slip can create hcp stacking when gliding on alternate (0 1 1)_{bcc} planes. Consideration of dislocation gliding to bring about phase transformation indeed is a most likely driving mechanism, which has more directly been noted in other alloys as we discuss next.

8. Case study: Fe-Mn-Si based SMAs

Fe-based SMAs are well-known for low costs compared to other classes of alloys. Additionally, the Fe-Mn-Si alloys or their hybrids are particularly useful for their potential for room temperature applications [256,257]. In terms of deformation mechanism, strain recovery in these alloys is achieved by reversible fcc ↔ hcp transformation [258,259]. One interesting aspect of this transformation process has been noted in the form of gliding partial dislocations, which apparently brings about the conversion of one crystal type into another [260]. This unique type of dislocation-assisted transformation constitutes a significant conduit for modeling.

8.1. Transformation mechanism and avenues for atomistic modeling

The fcc-to-hcp transformation mechanism can be best understood by considering alternate glissile motions of partial dislocations. The process is illustrated in Fig. 36. When Shockley partial dislocations of 1/6(1 1 2)_{fcc} type start gliding on every other (1 1 1)_{fcc} plane, an alternation of slipped and un-slipped stacking of planes is created. This essentially transforms the...
original fcc stacking (ABCABC...) into a hcp stacking (ABAB...) in the wake of the glissile partials. The atoms (red) from the slipped plane in original fcc lattice re-position to the hcp basal plane. On the other hand, the positions of the atoms from the adjacent (un-slipped) \{1 1 1\}_h, plane remain unchanged. Consequently, the hcp lattice structure is created. Electron microscopy evidence of fcc-to-hcp transformation by means of layered stacking fault formation is presented in Fig. 37 [256,261].

The movements of partial dislocations and the resultant shear of atoms bear considerable significance for modeling purpose the transformation mechanics. It can be reasonably deduced that the transformation would occur by essentially overcoming the total energy expenditure by the applied work. The energy cost can originate from continuum fields of slip and the lattice resistance for the atomic movements from fcc to hcp stacking. The forces governing slip motion subjected to mutual attraction and/or repulsion is well established [79]. Atomistically, the motion of atoms would require overcoming energy barriers originating from the discrete lattice, which are possible to obtain from first principles.

9. Case study: Fe-Mn-Ni-Al SMA

In the recent years, the Fe-Mn-Ni-Al alloys [262–264] have reportedly demonstrated considerable promise. For instance, superior superelastic attributes (with a transformation strains of >8% and a low hysteresis) are noted at room temperature. In particular, the transformation stress is found to have negligible temperature dependence over a substantial range of −196 °C to 240 °C as indicated by a $\frac{\partial\sigma}{\partial T}$ magnitude of 0.53 MPa/°C [265] (e.g. compared to NiTi having $\frac{\partial\sigma}{\partial T} = 6 – 8$ MPa/°C [266]). The superelasticity of Fe-Mn-Al-Ni SMAS is controlled by reversible transformation between bcc austenite and fcc martensite lattices. The mechanistic process of its transformation at the atomic level can be understood by revising earlier models in the literature. Bogers and Burgers [253,267] first theorized the atomic shear displacements needed for fcc-to-bcc conversion. Subsequently, Olson and Cohen [268,269] proposed the glide of partial dislocations as precursor to the transformative displacements of atoms. Intuitively, the inherent propensity to seek the energetic ground state would govern the atomic re-positioning and their association with slip. The necessary degree of combined shear and shuffle can thus be conveniently modeled after the ab initio energy pathways. As we discuss next, these mechanistic assumptions can prove useful for predicting critical stress parameter. Moreover, the slip-based geometric considerations have also been justified in high-resolution electron microscopy.

Fig. 39. (Bottom left) TEM image showing intersecting bands of hcp phase in austenitic (fcc) stainless steel [274]; (top) higher resolution picture of the intersection, at which bcc phase is formed as a result of crisscrossing hcp planes; (bottom right) a schematic illustration of how passage of partial slip in fcc matrix can create hcp layers, the intersection of which gives rise to bcc embryo.
9.1. Transformation mechanism

Building upon the early fundamental concepts [253,267–269], we revisit the atomistic process of bcc-to-fcc transformation in the form of Fig. 38. The entire process could be understood as a dual shearistic mechanism involving \((1\ 1\ 0)(1\ 1\ 0)_{\text{bcc}}\) and \((1\ 1\ 1)(1\ 1\ 2)_{\text{fcc}}\) systems. In the left inset, four blue boxes represent four bcc unit cells, where atoms are colored in silver and red on parallel \((0\ 0\ 1)_{\text{bcc}}\) planes (dark-shaded). Note that only select atoms per primitive cell are shown for visual convenience. When a \(1/8\ (1\ 1\ 0)_{\text{bcc}}\) dislocation sweeps alternately on every second \((0\ 1\ 1)_{\text{bcc}}\) plane in opposite directions, the swept atoms (silver-colored) would be accordingly re-positioned. This is due to the fact the partial dislocation would leave a stacking fault in its wake, which is essentially a planar fault with re-arranged atoms. In a similar shear-like manner, the passage of a \(1/6\ (1\ 1\ 2)_{\text{fcc}}\) type partial dislocation (yellow-colored) on every third \((1\ 1\ 1)_{\text{fcc}}\) i.e. parallel to \((0\ 1\ 1)_{\text{bcc}}\) layer would re-locate in-plane atoms (red-colored) suitable for generating stacking faults. Thus, the fcc unit cell would essentially be created in the region where the layered stacking faults left behind by two different sets of glissile dislocations would intersect. The particular geometry of intersecting stacking faults (or local hcp layers) constitutes an important feature, which has been appropriately utilized in modeling as well as evidenced during experiments most recently.

From modeling standpoint, the mechanics of the bcc-to-hcp transformation essentially reduces to a straightforward problem of balancing the applied work with the overall energy expenditure. The origin of the total energy cost is twofold: (a) overcoming the long-range elastic interaction among layer-by-layer dislocations and (b) surpassing the barriers on the fault energy landscape. While the former can be modeled after elastic solutions of dislocation strain energy considerations [79], the latter is obtained from first principles. The feasibility of combining continuum slip and atomistic fault energies within Peierls-Nabarro framework [270,271] has proven useful for predicting slip and twinning stresses earlier both for conventional and shape memory materials [42,272,273]. Similarly, by way of extending such approaches, the transformation problem can potentially be disintegrated into a synergy of classical slip mechanics and \(\gamma\) energy. In addition, it is worth mentioning here that interchangeable symmetry is at the core of the reversibility of phase transformation in shape memory alloys as noted by Bhattacharya et al. [71]. However, the foregoing transformation mechanisms (i.e. fcc \(\leftrightarrow\) hcp, fcc \(\leftrightarrow\) bcc and bcc \(\leftrightarrow\) hcp) do not hold any group/subgroup correlations in terms the symmetry of various Bravais lattice systems.

9.2. Experimental evidence of similar mechanism

Fig. 39 (adapted from [274]) presents high-resolution TEM observation of passing partial slip forming hcp layers (stacking faults), which forms bcc phase at their intersection in austenitic stainless steel. It should be noted here that the superelastic deformation process is essentially a two-way transformation phenomenon for the SMA case. Thus, although the material in question is not an SMA, the understanding of its stress-induced martensitic transformation (i.e. fcc-to-bcc) is directly applicable to the behavior of Fe-Mn-Al-Ni SMAs. At lower left, a low magnification image of intersection hcp bands can be noticed. With higher resolution (top inset), the fcc matrix, hcp structure, and bcc phase in the junction are visible. It is inferred that partial slip form in the fcc matrix and then glide on consecutive parallel \((1\ 1\ 1)_{\text{fcc}}\) planes, thereby creating bands of local hcp structures. As the schematic (lower right) shows, when two such bands traverse each other, the material at the junction transform to bcc lattice. The occurrence of requisite atomic shear and/or shuffle can be envisioned as the reverse mechanism of that illustrated in Fig. 38. That the same mechanism could possibly be operable during the SMA deformation has been justifiably demonstrated subsequently on predictive ground.

10. Other emerging studies: Ni-Ti-Hf, Ti-Ni-Cu, Ni-Mn-Sn SMAs

10.1. Ni-Ti-Hf based alloys

Ni-Ti-Hf SMAs hold considerable promise for high temperature applications, especially in aerospace industries (e.g. as variable-geometry engine outlet) [3,275]. To enhance properties to that end, aging treatments are employed to tweak microstructure by, for example, refining grains, imposing precipitates [276–279]. The beneficial roles of these precipitates have been related with increased superelastic attributes, strength, work output and stability [280–282]. Atomistic simulations have brought forth important structural properties of precipitates, which have shed more light into empirical understanding of the microstructure [283,284]. Yang et al. [276] conducted extensive microstructure characterization on the precipitates through high resolution microscopy along with DFT calculations. Specifically, the lattice structure was examined in detail. Starting with an unrelaxed face-centered orthorhombic crystal using lattice constants from [285], they proceeded to seek the ground-state crystal structure via energy minimization. What is unique about this study is that they conducted a thorough analysis on the one-to-one correspondence between the (simulated) intensities of high-resolution scanning TEM and the DFT-based atomic positions. It was found that due to relaxation of the (initial) tentative positions, the atomic motif points were shifted, leading to a better agreement with experiments. Thus-computed lattice constants were reported to be: \(a = 1.264\) nm, \(b = 0.882\) nm and \(c = 2.608\) nm for a chemical formula of \(\text{Ni}_{0.5}\text{Ti}_{0.17}\text{Hf}_{0.33}\). These results are important, in that they can be utilized in the phenomenological stretch tensor based models [71–73]. Similarly, slip analysis of the Ni-Ti-Hf SMAs has also been investigated from first principles [284]. It remains to be seen how new studies emerge to shed more light into other outstanding issues.
such as the elastic moduli, the misfit distortion field, resistance against slip, which are the principle microscopic factors affecting macroscale SMA attributes.

10.2. Ti-Ni-Cu based SMAs

The main prospects of Ti-Ni-Cu SMAs are related with high temperature applications, which demand a desired amalgam of many characteristics. For instance, elevation of transformation temperature, improved fatigue life and narrow thermal hysteresis are among the noteworthy attributes [164,165,286–289]. These important improvements in the mechanical properties are noted as a direct result of Cu alloying. Addition of Cu also alters the transformation pathway, whereby, depending on Cu content, intermediate B19 phase can be prevalent in the microstructure. With Cu content <7.5 at%, the transformation occurs from B2 to B19′; for Cu between 7.5 at% and 15 at%, the path becomes B2 → B19 → B19′, and B2 → B19 for Cu more than 15 at%. While this material has attracted numerous experimental investigations, recently DFT studies focusing on lattice structure and stability have also emerged [290–292].

Teng et al. [292], using DFT calculations, predicted a decline in lattice constants with increasing Cu content. Subsequently, Gou et al. [290,291] further corroborated that the addition of Cu has the effect of reducing the monoclinic angle of the martensite phase. They noted the increasing instability (manifested in terms of rising free energy) of monoclinic (B190) structure with increasing Cu addition. Specifically, a threshold range was predicted, noting that for 0 at% < Cu < 18.75 at% martensite is monoclinic whereas at 20% it becomes orthorhombic. Altogether, these findings rationalize the predominance of the B19 phase (having orthorhombic structure) with high Cu content in the experimental observations.

10.3. Ni-Mn-Sn based SMAs

The class of Ni-Mn-Sn is principally known for their magnetically recoverable deformation [293–295]. Additionally, their potential as refrigerant resulting from significant elastocaloric effect has also been examined [296,297]. Similar to other Heusler alloys, the stoichiometric alloy (N2MnSn) is characterized by an ordered austenitic phase of cubic L21 lattice, while the ordered martensite being the L10 type [298] (which can also exist as 10M or 14M modulated structure). It was reported that while the austenite is ferromagnetic, the martensite becomes antiferromagnetic even below the Curie temperature. Ab-initio calculations have provided some rationale to these observations [299].

From first principles, fundamental differences in the sub-lattice attributes, as affected by Mn addition, were studied [196,198,300,301]. In the non-stoichiometric composition, Mn atoms randomly occupy the sites of Sn (tin) atoms (i.e. Ni1Mn(x)xSn1−x) [302]. This has several consequences as the calculations suggest. For one, the magnetic moment between the Mn-occupying Sn and the “regular-lattice” Mn was found to be anti-parallel to that of the regular nearest-neighbor Mn-Mn pair [301,302]. Consequently, the collective magnetization of the alloy was predicted to have experienced a considerable decline with excess Mn, an effect also noted in experiments [302]. Secondly, it was suggested [300] that the L21 structure becomes increasingly unstable with additional Mn solutes, which results in a natural stabilization process via splitting of quantum energy bands. This reportedly has the effect of easing the austenite-to-martensite transformation. Similar physical effects are also noted in Ni-Mn-In based magnetic shape memory alloys from ab initio predictions [303].

11. Concluding remarks

It is worthwhile to make the most of the existing computational tools to develop a synergy among different modeling approaches across lengthscales. On critiquing the recent predictive literature, we have highlighted the applicability and core inferences of atomic models. The current discourse contains a brief narrative of our current understanding of microscopic SMA behaviors, which also pose thought-provoking questions warranting future research. From engineering standpoint, a lattice-scale insight can potentially accelerate the pursuit of developing improved predictive capabilities. The grain level atomic simulations via MD are essentially the bridge between the continuum assumptions and the discrete lattice physics. These developments essentially motivate extension of the atomistic approach towards more practical engineering issues. To elaborate, let us consider the theorization of fatigue-induced damage of SMA components.

The study of metal fatigue is treated with special emphasis given the adverse consequences of the associated microscopic alterations, which deteriorates desired mechanical attributes and component life [304–306]. Fatigue of shape memory alloys can be generally understood in two broad categories [307–310]: (a) functional fatigue and (b) structural fatigue. Functional fatigue refers to the drifting of superelastic responses over cycles, which results from persistent accumulation of slip and martensite. Structural fatigue, on the other hand, means nucleation of a crack (due to excessive localization of plasticity and residual martensite) and its gradual progression to catastrophic fracture. In order for an SMA to retain its strain recoverability i.e. the phase reversibility, it is not sufficient for the parent lattice to be able to only transform. It is crucial to ensure the absence of factors that might hamper the two-way conversion e.g. considerable slip activities. An enhanced slipping propensity may adversely affect the reversibility of transformation. Thus, the sector to which the atomistic simulations could contribute is to unveil the actual mechanism of concurrent plasticity and transformation and their mutual interaction [311].

The extent of literature on MD simulations covering nano-scale cracking mechanism is considerable for conventional (non-transforming) materials [312–321]. It remains to be seen with existing pace in developing new MD potentials as well
as faster DFT simulations how new atomistic models will emerge. For instance, transformation energy pathways for the fcc-to-bcc mechanism described in section 9.1 has very recently been addressed; also, a mechanics model is proposed which predicted a critical stress level of 191 MPa [322] in close agreement with experimental value (200 MPa). Similarly, computed energy pathway for bcc-to-orthorhombic transformation mechanism (in Ti-Nb based SMAs) has recently helped evaluate critical stress thereof and compare it with critical slip and twinning stresses [323]. These critical stresses bear considerable promise as important input into fracture models.

Acknowledgement

This research was financially supported by the Nyquist Chair funds. We thank the reviewers for important input.

References


Bönisch M. Structural properties, deformation behavior and thermal stability of martensitic Ti-Nb alloys; 2016.


Burgers W. On the process of transition of the cubic-body-centered modification into the hexagonal-close-packed modification of zirconium. Physica 1934;1:561–86.


Burgers W. On the process of transition of the cubic-body-centered modification into the hexagonal-close-packed modification of zirconium. Physica 1934;1:561–86.
