PERGAMON

Acta mater. 48 (2000) 197–222

Acta MATERIALIA

www.elsevier.com/locate/actamat

MARTENSITIC TRANSFORMATIONS AND SHAPE-MEMORY MATERIALS☆

R. D. JAMES† and K. F. HANE

Department of Aerospace Engineering and Mechanics, University of Minnesota, Minneapolis, MN 55455, USA

(Received 1 June 1999; accepted 15 June 1999)

Abstract—The authors review theoretical research on martensitic phase transformations in shape-memory materials, with emphasis on recently derived theory and predictions of interest for alloy development. Research on special lattice parameters corresponding to certain microstructures, complex crystal structures and 6M martensite, the relation of micro-scale to macro-scale deformations, ferromagnetic and ferroelectric martensites, and martensite at small scales is covered.

Keywords: Crystallography; Magnetostrictive effects; Microstructure; Phase transformations; Martensite; Shape memory

1. INTRODUCTION

Over the past decade nonlinear thermoelasticity has been developed for martensitic transformations, especially for the study of reversible martensitic transformations in shape-memory materials. The theory has produced specific quantitative predictions, some of which are being used as the basis of alloy development programs and others which await experimental verification. Some of these predictions are rather unexpected and have led to a revision of the fundamentals of martensitic transformations.

Except for the review of Bhattacharya [1], which itself is not readily available to many researchers, the results and techniques presented here are otherwise scattered throughout the literature of materials science, continuum mechanics, mathematics and physics. Some of the results are written in a form that is not easily accessible to working materials' scientists. The purpose of this review is to assemble these results in a succinct, approachable presentation, with a focus on the most recent developments.

There are many outstanding reviews of martensite, shape-memory, and related areas. In particular, the books of Nishiyama [2] and Otsuka and Wayman [3] cover the classical developments in martensitic transformations, the latter including modern research on shape-memory polymers and ceramics. The forthcoming article of Miyazaki and Ishida [4] is a recent review of research and applications of sputtered thin films, with particular emphasis on the TiNi system. The review by Miyazaki and Otsuka [5] is also a valuable source. These reviews, however, do not treat theoretical issues.

To introduce the present approach, it is useful to trace the historical background of the present line of thought. In the 1950s fundamental advances on martensitic transformations—crystallography, mechanism, kinetics, and macroscopic properties—were made by Nishiyama, Kurdyumov, Christian, Read, and others. As explained to us by Lieberman, at the advice of Read, Lieberman and Wechsler took the course of Mindlin on continuum mechanics to learn some large deformation kinematics that Read thought might be useful for understanding the curious irrationality of the austenite/martensite interface. The result was one version of the crystallographic theory of martensite. While this has been reviewed and applied hundreds, perhaps thousands, of times, no further advance along that line was made. The modern work reviewed here continues precisely that line of thought. With advances in continuum mechanics that occurred in the intervening years, it was an easy step to write a free energy function that would produce the austenite/martensite interface by energy minimization, relate it to crystal structure, and then to go on to

† To whom all correspondence should be addressed.

☆ The Millennium Special Issue — A Selection of Major Topics in Materials Science and Engineering: Current status and future directions, edited by S. Suresh.

1359-6454/00/$20.00
PII: S1359-6454(99)00295-5

the many other microstructures that are observed in martensite and to investigate how these relate to behavior.

A main result of the theory is the recognition that some of the common microstructures in shape-memory materials are only possible (as energy-minimizing microstructures) with exceedingly special lattice parameters. These results are collected in Section 4. There is extremely good agreement between measured and predicted lattice parameters on materials that clearly show those microstructures. Advances of the understanding of the relation between microscopic and macroscopic deformation also played a key role, reviewed later in Section 5. In recent months, workers in this area have turned attention to ferromagnetic and ferroelectric martensites (Section 6) and the behavior of martensite at small scales (Section 7). For the former, the theory has directly guided the development of these materials. Difficulties with the theoretical treatment of 18R martensites have also been recently overcome, and there appears now to be the beginning of a satisfactory theory (Section 3).

This review is unfortunately not comprehensive even regarding research related to nonlinear thermoelasticity. The notable omissions include important work on polycrystals [6–8], on geometrically linear theory [9, 10], on hysteresis [11], on constitutive equations (see for example Refs [12–17]), on the simulation of martensitic transformations [18, 19], and on Density Functional Theory computations of atomic structure [20].

We use the following notation. Greek letters are scalars, lower case bold letters are vectors in $\mathbb{R}^3$, and upper case bold letters are 3×3 matrices. Unit vectors have a superimposed hat. $\mathbf{y}(\mathbf{x}) = \mathbf{G}\mathbf{x}$ is the direct form of the formula $y_i = G_{ij}x_j$, where summation over the repeated indexes is assumed. A superscript T denotes the transpose $[[(\mathbf{A}^{\mathrm{T}})_{ij} = (\mathbf{A})_{ji}]]$, and tr is the trace ($\mathrm{tr}\mathbf{A} = A_{ii}$). The symbol ε_{ijk} denotes the permutation symbol, defined uniquely by the two requirements: (i) $\varepsilon_{123} = 1$; and (ii) ε_{ijk} switches sign whenever any pair of indices are switched (used in Section 5). The tensor product of the vectors $\mathbf{a}$ and $\mathbf{b}$ is $\mathbf{a} \otimes \mathbf{b}$, which in components is the matrix $(\mathbf{a} \otimes \mathbf{b})_{ij} = a_i b_j$. If $\mathbf{A}$ and $\mathbf{B}$ satisfy $\mathbf{B} - \mathbf{A} = \mathbf{a} \otimes \mathbf{n}$, we say that they are "rank-1 connected". Rotation matrices (called simply rotations) are denoted by the letters $\mathbf{Q}$ or $\mathbf{R}$, sometimes adorned with superscripts, etc. The set of all rotation matrices is $\mathrm{SO}(3) = \{\mathbf{R}:\mathbf{R}\mathbf{R}^{\mathrm{T}} = \mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I}$ and $\det(\mathbf{R}) = +1\}$. A rotation of ψ counterclockwise degrees with axis $\hat{\mathbf{p}}$ is written $\mathbf{Q} = \mathbf{Q}[\psi,\hat{\mathbf{p}}]$, so $\mathbf{Q}[\psi,\hat{\mathbf{p}}]\hat{\mathbf{p}} = \hat{\mathbf{p}}$.

2. MICROSTRUCTURE BY ENERGY MINIMIZATION

Over the past decade, nonlinear thermoelasticity has been used to study various problems associated with martensitic transformations in shape-memory materials (for example Refs [21–24] and the references therein). One of the successes of this theory is its ability to predict detailed microstructures which are observed in materials. Once the properly invariant free energy function is defined, then there are no further assumptions, and all of the common information about martensitic transformations follows by direct calculation: the twins in the martensite, their types, all austenite/martensite interfaces, more complex microstructures, the effect of stress or electromagnetic field on transformation temperature.

In the simplest case, one begins with a Bravais lattice determined by three linearly independent vectors $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\}$. In particular, a Bravais lattice is the set of all points in three dimensions given by

$$\nu^i\mathbf{e}_i = \nu^1\mathbf{e}_1 + \nu^2\mathbf{e}_2 + \nu^3\mathbf{e}_3 \tag{1}$$

where ν^i are integers. Among all lattice vectors, there are special ones associated with the unstressed austenite $\{\mathbf{e}_1^{\mathrm{a}},\mathbf{e}_2^{\mathrm{a}},\mathbf{e}_3^{\mathrm{a}}\}$ and with (one variant of) the unstressed martensite $\{\mathbf{e}_1^{\mathrm{m}},\mathbf{e}_2^{\mathrm{m}},\mathbf{e}_3^{\mathrm{m}}\}$. From these lattice vectors, one can calculate the point groups of austenite and martensite, defined as the set of orthogonal transformations of the lattice that restore the lattice. For reasons that will be clear below, one needs only the subsets of these groups consisting of elements with positive determinant; we call these groups $\wp^{\mathrm{a}}$ and $\wp^{\mathrm{m}}$. These lattice vectors change slightly with temperature, so the given ones correspond to the transformation temperature. An atomic scale free energy per unit reference volume is postulated, which is a function of lattice vectors and temperature only. Further, it is assumed that there is a neighborhood (called the Ericksen–Pitteri neighborhood [22, 25–27]) of the lattice vectors of the parent phase and that this neighborhood contains the lattice vectors of the product phase as well. In particular, this implies that the point group of the martensite is a subgroup of the point group of the austenite, $\wp^{\mathrm{m}} \subset \wp^{\mathrm{a}}$.

A continuum theory is obtained from the atomic theory by using the Cauchy–Born rule [28–30]. In particular, a reference configuration $\Omega \subset \mathbb{R}^3$ is defined which represents the domain occupied by the body in unstressed austenite. Deformations of the body, due to either transformation or elastic distortion, are described by functions $\mathbf{y}:\Omega \rightarrow \mathbb{R}^3$. The deformation gradient has positive determinant for physically realizable deformations, $\det\nabla\mathbf{y} > 0$. The Cauchy–Born rule states that if $\mathbf{F} = \nabla\mathbf{y}(\mathbf{x})$ is the deformation gradient at $\mathbf{x}$, then the underlying lattice vectors in the deformed configuration at $\mathbf{y}(\mathbf{x})$ are given by

$$\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\} = \left\{\mathbf{F}\mathbf{e}_1^{\mathrm{a}},\mathbf{F}\mathbf{e}_2^{\mathrm{a}},\mathbf{F}\mathbf{e}_3^{\mathrm{a}}\right\} \tag{2}$$

This rule allows one to pass back and forth between

the lattice and continuum pictures. For example, using equation (2) one can take a continuum deformation for compatible martensite variants and find out immediately if an interface between variants represents a type I twin, a type II twin, a compound twin, or none of these.

The basic assumption is that the free energy per unit volume in Ω is a function of the lattice vectors and temperature, as is consistent with a typical atomic level calculation of total energy based on the Born–Oppenheimer approximation. By the Cauchy–Born rule, the free energy can be expressed as a function of the deformation gradient and temperature, $\varphi(\mathbf{F},\theta)$, where $\mathbf{F} = \nabla\mathbf{y}(\mathbf{x})$.

Lattices that are more complex than Bravais lattices (so-called multilattices) are described generally as the union of a finite number n of identical Bravais lattices, displaced relative to the origin by vectors $\{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$. In this case, the Cauchy–Born rule applies to each of these Bravais sublattices, and the free energy is a function of the deformation gradient and $n-1$ "shifts": $\varphi(\mathbf{F},\mathbf{p}_n - \mathbf{p}_1, \ldots, \mathbf{p}_2 - \mathbf{p}_1)$. Transformations for which $\mathbf{p}_i^{\mathrm{m}} - \mathbf{p}_1^{\mathrm{m}} \neq \mathbf{F}(\mathbf{p}_i^{\mathrm{a}} - \mathbf{p}_1^{\mathrm{a}})$ are associated with shuffling. Often, for martensitic transformations in complex lattices, it is sufficient to focus on one Bravais sublattice and assume that the shifts have been minimized out of the free energy (see e.g. Refs [22, 23, 31, 32] and Section 3).

The free energy density φ has two fundamental invariance properties. The first arises from consideration of the Ericksen–Pitteri neighborhood and concerns symmetry: for each temperature and deformation gradient, the free energy density φ satisfies

$$\varphi(\mathbf{FQ},\theta) = \varphi(\mathbf{F},\theta) \text{ for all } \mathbf{Q} \in \wp^{\mathrm{a}} \tag{3}$$

where $\wp^{\mathrm{a}}$ is the point group of the austenite lattice. The second is invariance under all superposed rigid body rotations:

$$\varphi(\mathbf{RF},\theta) = \varphi(\mathbf{F},\theta) \text{ for all } \mathbf{R} \in \mathrm{SO}(3) \tag{4}$$

Note that from equation (3) the symmetry of the austenite dominates the free energy.

Let $\mathbf{U}$ be the unique, linear transformation that maps the austenite lattice vectors to the martensite lattice vectors:

$$\{\mathbf{e}_1^{\mathrm{m}},\mathbf{e}_2^{\mathrm{m}},\mathbf{e}_3^{\mathrm{m}}\} = \{\mathbf{U}\mathbf{e}_1^{\mathrm{a}},\mathbf{U}\mathbf{e}_2^{\mathrm{a}},\mathbf{U}\mathbf{e}_3^{\mathrm{a}}\} \tag{5}$$

By rigidly rotating the martensite lattice vectors (if necessary), we can assume that $\mathbf{U}$ is positive-definite and symmetric. The Bain strain matrix $\mathbf{B}$ is $\mathbf{U}-\mathbf{I}$. A number of facts about the relation between $\mathbf{U}$, $\wp^{\mathrm{a}}$, and $\wp^{\mathrm{m}}$ follow directly from the assumption that $\{\mathbf{e}_i^{\mathrm{m}}\}$ is in the Ericksen–Pitteri neighborhood of $\{\mathbf{e}_i^{\mathrm{a}}\}$ (see Ref. [22] for details). First, it follows that $\wp^{\mathrm{m}}$ is a subgroup of $\wp^{\mathrm{a}}$. Second, $\wp^{\mathrm{m}}$ consists of exactly those elements of $\wp^{\mathrm{a}}$ that leave $\mathbf{U}$ fixed in the following sense:

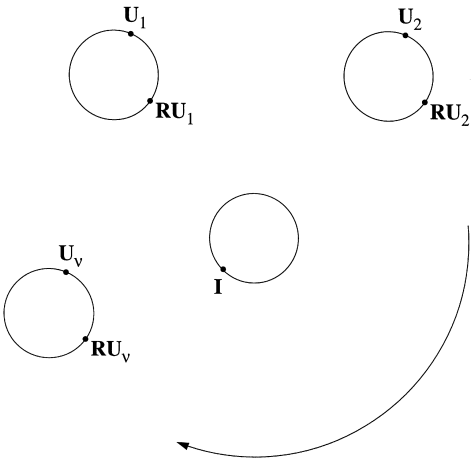

Fig. 1. Schematic of the energy wells at θ_{c}.

$$\wp^{\mathrm{m}} = \{\mathbf{Q} \in \wp^{\mathrm{a}} : \mathbf{QUQ}^{\mathrm{T}} = \mathbf{U}\} \tag{6}$$

Given the point groups, this is a very strong restriction on the form of the matrix $\mathbf{U}$, and this is how the specific forms given below for various transformations were determined. Third, if we calculate all distinct matrices given by $\{\mathbf{U}_1, \ldots, \mathbf{U}_\nu\} = \{\mathbf{QUQ}^{\mathrm{T}} : \mathbf{Q} \in \wp^{\mathrm{a}}\}$, then (by Lagrange's theorem), $\nu = \#\wp^{\mathrm{a}}/\#\wp^{\mathrm{m}}$, where $\#\wp$ stands for the number of elements in the group $\wp$. Below, it will be clear that $\{\mathbf{U}_1, \ldots, \mathbf{U}_\nu\}$ define the variants of martensite.

Let θ_{c} be the transformation temperature, i.e. the temperature at which the austenite and martensite have the same free energy density. The structure- and history-sensitive temperatures M_{s}, M_{f}, A_{s}, and A_{f} are not built into the definition of the free energy, but follow by studying metastable states, e.g. relative (as opposed to absolute) minimizers of the total free energy. At θ_{c}, the matrices $\mathbf{U}$ and $\mathbf{I}$ (the latter because of the choice of reference configuration) are assumed to minimize the free energy density φ. By equations (3) and (4), if $\mathbf{F}$ minimizes φ at any given temperature then so does $\mathbf{RFQ}$ for every rotation $\mathbf{R} \in \mathrm{SO}(3)$ and every $\mathbf{Q} \in \wp^{\mathrm{a}}$. Given one minimizer, we always get multiple minimizers. The quantity $\mathbf{RFQ}$ can be written $\mathbf{RQQ}^{\mathrm{T}}\mathbf{FQ}$, because $\mathbf{QQ}^{\mathrm{T}} = \mathbf{I}$, and, again $\mathbf{RQ}$ is a typical element of SO(3). In summary, if $\mathbf{F}$ minimizes φ then so does $\mathrm{SO}(3)\mathbf{F}_1, \ldots, \mathrm{SO}(3)\mathbf{F}_k$, where $\{\mathbf{F}_1, \ldots, \mathbf{F}_k\} = \{\mathbf{Q}^{\mathrm{T}}\mathbf{FQ}, \mathbf{Q} \in \wp^{\mathrm{a}}\}$ and $\mathrm{SO}(3)\mathbf{F}$ stands for all matrices of the form $\mathbf{RF}$, $\mathbf{R} \in \mathrm{SO}(3)$. Hence, $\varphi(\mathbf{F},\theta_{\mathrm{c}})$ is minimized for $\mathbf{F}$ belonging to the set

$$\mathrm{SO}(3)\mathbf{I}, \mathrm{SO}(3)\mathbf{U}_1, \ldots, \mathrm{SO}(3)\mathbf{U}_\nu \tag{7}$$

and these energy wells given by equation (7) are depicted in Fig. 1.

At θ_{c}, there is the typical exchange of stability associated with a first-order phase transformation. Above the temperature θ_{c}, φ exhibits only the austenite well, and below θ_{c}, φ exhibits only the martensite wells. The matrices $\{\mathbf{U}_1, \ldots, \mathbf{U}_\nu\}$ change

slightly with temperature, and there is a positive-definite symmetric matrix $\mathbf{U}^a(\theta)$, $\mathbf{U}^a(\theta_c)=\mathbf{I}$, that describes thermal expansion of the austenite. (From the assumption that there is only a single well above θ_c, we have $\mathbf{U}^a(\theta)=\mathbf{Q}^T\mathbf{U}^a(\theta)\mathbf{Q}$ for $\mathbf{Q}\in\wp^a$: the classical restriction on thermal expansion matrices.)

The total free energy is

$$\int_\Omega \varphi[\nabla\mathbf{y}(\mathbf{x}),\theta]d\mathbf{x} + \text{loading device energy.} \qquad (8)$$

Whenever possible, the first term of equation (8) can be minimized by putting $\nabla\mathbf{y}(\mathbf{x})$ on the energy wells at the appropriate temperature θ, insuring compatibility by making $\mathbf{y}$ continuous. Sometimes certain constructions, e.g. the austenite/martensite interface, involve energy minimizing sequences $\nabla\mathbf{y}^{(1)}$, $\nabla\mathbf{y}^{(2)}$, ..., $\nabla\mathbf{y}^{(k)}$, ... with transition layers of volume $1/k$ and with $\nabla\mathbf{y}^{(k)}$ bounded. If loads are applied, then the minimization of equation (8) generally leads to a difficult nonlinear elasticity problem.

In many cases, materials scientists use a geometrically linear version of the present theory. This can be obtained expanding φ about a matrix on each of its energy wells, i.e. for variant i, expand $\varphi[(\mathbf{I}+\varepsilon\mathbf{H}+\ldots)\mathbf{U}_i,\theta]$ to second-order in ε. Then from equation (4), it is found that the free energy density only depends on the linear strain matrix $\mathbf{E}=(\nabla\mathbf{u}+\nabla\mathbf{u}^T)/2$, where $\nabla\mathbf{u}=\varepsilon\mathbf{H}$ and $\mathbf{u}(\mathbf{x})=\mathbf{y}(\mathbf{x})-\mathbf{x}$ is the displacement. The wells are defined by linear strain matrices $\{\mathbf{E}_1, \ldots, \mathbf{E}_v\}$ and rotational invariance is replaced by the condition that if $\mathbf{E}_i$ minimizes the linearized free energy, then so does $\mathbf{E}_i+\mathbf{W}$, where $\mathbf{W}^T=-\mathbf{W}$. Schematically, each martensite well in Fig. 1 is replaced by a line passing through $\mathbf{U}_i$, and from the approximation of the rotations, one can imagine the possibility of errors. These are discussed in detail in Ref. [33].

The variants and consequently the energy wells for a number of different transformations have been determined. Some of the more common ones in shape-memory alloys are as follows (all matrices are given in the orthonormal cubic basis).

Cubic to trigonal transition: by assuming that the length of the sides of the cubic and trigonal unit cells are the same, then this transition is described by the trigonal angle ψ solely. This is very nearly the case in Ti–Ni (the R-phase) and Au–Cd alloys (see Ref. [34] and the references therein) and $TbDyFe_2$ (see Ref. [35] and the references therein). There are four variants, which have components

$$\mathbf{U}_1=\begin{pmatrix}\alpha&\beta&\beta\\ \beta&\alpha&\beta\\ \beta&\beta&\alpha\end{pmatrix},\quad \mathbf{U}_2=\begin{pmatrix}\alpha&-\beta&\beta\\ -\beta&\alpha&-\beta\\ \beta&-\beta&\alpha\end{pmatrix},$$

$$\mathbf{U}_3=\begin{pmatrix}\alpha&\beta&-\beta\\ \beta&\alpha&-\beta\\ -\beta&-\beta&\alpha\end{pmatrix},\quad \mathbf{U}_4=\begin{pmatrix}\alpha&-\beta&-\beta\\ -\beta&\alpha&\beta\\ -\beta&\beta&\alpha\end{pmatrix}$$

where $\alpha=(\sqrt{1+2\cos\psi}+2\sqrt{1-\cos\psi})/3$ and $\beta=(\sqrt{1+2\cos\psi}-\sqrt{1-\cos\psi})/3$.

Cubic to tetragonal transition: many martensitic materials undergo the cubic to tetragonal transition, such as Fe–Ni–C, Fe–Pd, In–Tl, Ni–Al, Ni–Mn, and Ni_2MnGa, for example (see Refs [21–23, 36] and the references therein). There are three variants, which have components

$$\mathbf{U}_1=\begin{pmatrix}\beta&0&0\\ 0&\alpha&0\\ 0&0&\alpha\end{pmatrix},\quad \mathbf{U}_2=\begin{pmatrix}\alpha&0&0\\ 0&\beta&0\\ 0&0&\alpha\end{pmatrix},$$

$$\mathbf{U}_3=\begin{pmatrix}\alpha&0&0\\ 0&\alpha&0\\ 0&0&\beta\end{pmatrix},$$

respectively. The transformation stretches α and β are proportional to the ratio of the tetragonal lattice parameters to the cubic one.

Cubic to orthorhombic transition: this transition can occur in one of two different ways [22, 23, 37, 38]. All six variants may have diagonal components exclusively, "cube-edge" variants [37]. This occurs rarely, but one of the stress-induced phases in Ni_2MnGa is of this type. The other type is defined by

$$\mathbf{U}_1=\begin{pmatrix}\beta&0&0\\ 0&\dfrac{\alpha+\gamma}{2}&\dfrac{\alpha-\gamma}{2}\\ 0&\dfrac{\alpha-\gamma}{2}&\dfrac{\alpha+\gamma}{2}\end{pmatrix},$$

$$\mathbf{U}_2=\begin{pmatrix}\beta&0&0\\ 0&\dfrac{\alpha+\gamma}{2}&\dfrac{\gamma-\alpha}{2}\\ 0&\dfrac{\gamma-\alpha}{2}&\dfrac{\alpha+\gamma}{2}\end{pmatrix},$$

$$\mathbf{U}_3=\begin{pmatrix}\dfrac{\alpha+\gamma}{2}&0&\dfrac{\alpha-\gamma}{2}\\ 0&\beta&0\\ \dfrac{\alpha-\gamma}{2}&0&\dfrac{\alpha+\gamma}{2}\end{pmatrix},$$

$$\mathbf{U}_4=\begin{pmatrix}\dfrac{\alpha+\gamma}{2}&0&\dfrac{\gamma-\alpha}{2}\\ 0&\beta&0\\ \dfrac{\gamma-\alpha}{2}&0&\dfrac{\alpha+\gamma}{2}\end{pmatrix},$$

$$\mathbf{U}_5 = \begin{pmatrix} \frac{\alpha+\gamma}{2} & \frac{\alpha-\gamma}{2} & 0 \\ \frac{\alpha-\gamma}{2} & \frac{\alpha+\gamma}{2} & 0 \\ 0 & 0 & \beta \end{pmatrix},$$

$$\mathbf{U}_6 = \begin{pmatrix} \frac{\alpha+\gamma}{2} & \frac{\gamma-\alpha}{2} & 0 \\ \frac{\gamma-\alpha}{2} & \frac{\alpha+\gamma}{2} & 0 \\ 0 & 0 & \beta \end{pmatrix}, \tag{9}$$

respectively, where α, β, and γ are the transformation stretches, $\alpha \neq \gamma$, and they are determined as for the stretches in the cubic to tetragonal transition. This transition occurs in the Cu–Al–Ni shape-memory alloy (see Refs [23, 32, 33] and the references therein), and these variants are called "face-diagonal" variants in Ref. [37].

Cubic to monoclinic transition: this transition can also occur in two different ways [37, 39]: (a) "Face-diagonal" variants have a unique twofold axis along a face-diagonal of the original cubic unit cell. From Ref. [31], there are twelve variants of the form

$$\mathbf{U}_1 = \begin{pmatrix} \xi & \rho & \rho \\ \rho & \sigma & \tau \\ \rho & \tau & \sigma \end{pmatrix}, \quad \mathbf{U}_2 = \begin{pmatrix} \xi & -\rho & -\rho \\ -\rho & \sigma & \tau \\ -\rho & \tau & \sigma \end{pmatrix},$$

$$\mathbf{U}_3 = \begin{pmatrix} \xi & -\rho & \rho \\ -\rho & \sigma & -\tau \\ \rho & -\tau & \sigma \end{pmatrix}, \quad \mathbf{U}_4 = \begin{pmatrix} \xi & \rho & -\rho \\ \rho & \sigma & -\tau \\ -\rho & -\tau & \sigma \end{pmatrix},$$

$$\mathbf{U}_5 = \begin{pmatrix} \sigma & \rho & \tau \\ \rho & \xi & \rho \\ \tau & \rho & \sigma \end{pmatrix}, \quad \mathbf{U}_6 = \begin{pmatrix} \sigma & -\rho & \tau \\ -\rho & \xi & -\rho \\ \tau & -\rho & \sigma \end{pmatrix},$$

$$\mathbf{U}_7 = \begin{pmatrix} \sigma & -\rho & -\tau \\ -\rho & \xi & \rho \\ -\tau & \rho & \sigma \end{pmatrix}, \quad \mathbf{U}_8 = \begin{pmatrix} \sigma & \rho & -\tau \\ \rho & \xi & -\rho \\ -\tau & -\rho & \sigma \end{pmatrix},$$

$$\mathbf{U}_9 = \begin{pmatrix} \sigma & \tau & \rho \\ \tau & \sigma & \rho \\ \rho & \rho & \xi \end{pmatrix}, \quad \mathbf{U}_{10} = \begin{pmatrix} \sigma & \tau & -\rho \\ \tau & \sigma & -\rho \\ -\rho & -\rho & \xi \end{pmatrix},$$

$$\mathbf{U}_{11} = \begin{pmatrix} \sigma & -\tau & \rho \\ -\tau & \sigma & -\rho \\ \rho & -\rho & \xi \end{pmatrix}, \quad \mathbf{U}_{12} = \begin{pmatrix} \sigma & -\tau & -\rho \\ -\tau & \sigma & \rho \\ -\rho & \rho & \xi \end{pmatrix},$$

where the specific components are

$$\xi = \frac{\alpha\big(\alpha + \gamma\sin(\theta)\big)}{\sqrt{\alpha^2 + \gamma^2 + 2\alpha\gamma\sin(\theta)}},$$

$$\rho = \frac{\alpha\gamma\cos(\theta)}{\sqrt{2}\sqrt{\alpha^2 + \gamma^2 + 2\alpha\gamma\sin(\theta)}},$$

$$\sigma = \frac{1}{2}\left(\frac{\gamma\big(\gamma + \alpha\sin(\theta)\big)}{\sqrt{\alpha^2 + \gamma^2 + 2\alpha\gamma\sin(\theta)}} + \beta\right),$$

$$\tau = \frac{1}{2}\left(\frac{\gamma\big(\gamma + \alpha\sin(\theta)\big)}{\sqrt{\alpha^2 + \gamma^2 + 2\alpha\gamma\sin(\theta)}} - \beta\right)$$

The transformation stretches are $\alpha = a/a_0$, $\beta = b/(\sqrt{2}a_0)$, and $\gamma = c/(\sqrt{2}a_0)$ where the lattice parameter of the cubic unit cell is a_0 and the lattice parameters of the monoclinic unit cell are a, b, and c, and θ is the angle between the edges with lengths a and c. More traditional notation labels the monoclinic angle θ as β. This transition occurs in the Ti–Ni shape-memory alloys (see Refs [31, 39] and references therein). (b) "Cube-edge" variants have a unique twofold axis along an edge of the original cubic unit cell. There are twelve variants with components

$$\mathbf{U}_1 = \begin{pmatrix} \beta & 0 & 0 \\ 0 & \rho & \sigma \\ 0 & \sigma & \tau \end{pmatrix} \quad \mathbf{U}_2 = \begin{pmatrix} \beta & 0 & 0 \\ 0 & \rho & -\sigma \\ 0 & -\sigma & \tau \end{pmatrix},$$

$$\mathbf{U}_3 = \begin{pmatrix} \beta & 0 & 0 \\ 0 & \tau & \sigma \\ 0 & \sigma & \rho \end{pmatrix} \quad \mathbf{U}_4 = \begin{pmatrix} \beta & 0 & 0 \\ 0 & \tau & -\sigma \\ 0 & -\sigma & \rho \end{pmatrix},$$

$$\mathbf{U}_5 = \begin{pmatrix} \rho & 0 & \sigma \\ 0 & \beta & 0 \\ \sigma & 0 & \tau \end{pmatrix} \quad \mathbf{U}_6 = \begin{pmatrix} \rho & 0 & -\sigma \\ 0 & \beta & 0 \\ -\sigma & 0 & \tau \end{pmatrix},$$

$$\mathbf{U}_7 = \begin{pmatrix} \tau & 0 & \sigma \\ 0 & \beta & 0 \\ \sigma & 0 & \rho \end{pmatrix} \quad \mathbf{U}_8 = \begin{pmatrix} \tau & 0 & -\sigma \\ 0 & \beta & 0 \\ -\sigma & 0 & \rho \end{pmatrix},$$

$$\mathbf{U}_9 = \begin{pmatrix} \rho & \sigma & 0 \\ \sigma & \tau & 0 \\ 0 & 0 & \beta \end{pmatrix} \quad \mathbf{U}_{10} = \begin{pmatrix} \rho & -\sigma & 0 \\ -\sigma & \tau & 0 \\ 0 & 0 & \beta \end{pmatrix},$$

$$\mathbf{U}_{11} = \begin{pmatrix} \tau & \sigma & 0 \\ \sigma & \rho & 0 \\ 0 & 0 & \beta \end{pmatrix} \quad \mathbf{U}_{12} = \begin{pmatrix} \tau & -\sigma & 0 \\ -\sigma & \rho & 0 \\ 0 & 0 & \beta \end{pmatrix}, \tag{10}$$

where

$$\rho = \frac{\alpha^2+\gamma^2+2\alpha\gamma(\sin(\theta)+\cos(\theta))}{2\sqrt{\alpha^2+\gamma^2+2\alpha\gamma\sin(\theta)}},$$

$$\sigma = \frac{\alpha^2-\gamma^2}{2\sqrt{\alpha^2+\gamma^2+2\alpha\gamma\sin(\theta)}},$$

$$\tau = \frac{\alpha^2+\gamma^2+2\alpha\gamma(\sin(\theta)-\cos(\theta))}{2\sqrt{\alpha^2+\gamma^2+2\alpha\gamma\sin(\theta)}},$$

and $\beta = b/a_0$, $\alpha = \sqrt{2}a/a_0$, $\gamma \propto c/a_0$ (see Section 3 for explicit expressions for the stretch γ), and θ is the monoclinic angle between the edges with lengths a and c (see Ref. [40] and the references therein).

Other transformations that have been analyzed in detail are the tetragonal to monoclinic [41, 42] and the orthorhombic to monoclinic [22] (see also Ref. [37]).

The construction of energy-minimizing microstructures often reduces to considering continuous deformations with piecewise constant gradients on the austenite and martensite energy wells. In particular, consider a body Ω which undergoes a homogeneous deformation with piecewise constant gradient of the form

$$\mathbf{y} = \begin{cases} \mathbf{R}\mathbf{F}_1\mathbf{x} + \mathbf{c}_1 & \text{for } \mathbf{x}\cdot\hat{\mathbf{n}} \le 0,\ \mathbf{x}\in\Omega, \\ \mathbf{F}_2\mathbf{x} + \mathbf{c}_2 & \text{for } \mathbf{x}\cdot\hat{\mathbf{n}} > 0,\ \mathbf{x}\in\Omega, \end{cases} \tag{11}$$

where $\mathbf{c}_1$ and $\mathbf{c}_2$ are constant vectors, $\mathbf{F}_1 \neq \mathbf{F}_2$, $\mathbf{R} \in \mathrm{SO}(3)$, and $\hat{\mathbf{n}}$ is the normal to the surface dividing the body into regions with either deformation gradient. Necessary and sufficient conditions that the deformation given in equation (11) is continuous are: $\mathbf{c}_1 = \mathbf{c}_2$, $\hat{\mathbf{n}}$ is a constant vector, and

$$\mathbf{R}\mathbf{F}_1 - \mathbf{F}_2 = \mathbf{a}\otimes\hat{\mathbf{n}} \tag{12}$$

Therefore, the regions undergo a common translation, the dividing surface is a plane, and the deformation gradients differ by a rank-one matrix. The Hadamard compatibility condition [equation (12)] commonly arises in the construction of microstructures, and solutions to this equation can be found as follows: if the deformation gradients $\mathbf{F}_1$ and $\mathbf{F}_2$ are known, then define the symmetric stretch $\mathbf{C}$ to be $(\mathbf{F}_1\mathbf{F}_2^{-1})^{\mathrm{T}}(\mathbf{F}_1\mathbf{F}_2^{-1})$, which from equation (12) can also be written as $\mathbf{C} = (\mathbf{I}+\mathbf{a}\otimes\mathbf{F}_2^{-\mathrm{T}}\hat{\mathbf{n}})^{\mathrm{T}}(\mathbf{I}+\mathbf{a}\otimes\mathbf{F}_2^{-\mathrm{T}}\hat{\mathbf{n}})$. Necessary and sufficient conditions that solutions to equation (12) for the vectors $\mathbf{a}$ and $\hat{\mathbf{n}}$ exist are given by Proposition 1:

Proposition 1 [21, 22]. Necessary and sufficient conditions for a symmetric 3×3 matrix $\mathbf{C} \neq \mathbf{I}$ with eigenvalues $\lambda_1 \le \lambda_2 \le \lambda_3$ to be expressible in the form

$$\mathbf{C} = (\mathbf{I}+\hat{\mathbf{m}}\otimes\mathbf{b})(\mathbf{I}+\mathbf{b}\otimes\hat{\mathbf{m}})$$

with $1+\mathbf{b}\cdot\hat{\mathbf{m}} > 0$ and $\mathbf{b}\neq\mathbf{0}$, $\hat{\mathbf{m}}\neq\mathbf{0}$ are that $\lambda_1 > 0$ and $\lambda_2 = 1$. All solutions are given by

$$\mathbf{b} = \frac{\rho}{\sqrt{\lambda_3-\lambda_1}}\Big(\sqrt{\lambda_3(1-\lambda_1)}\hat{\mathbf{e}}_1 + \kappa\sqrt{\lambda_1(\lambda_3-1)}\hat{\mathbf{e}}_3\Big) \tag{13}$$

$$\hat{\mathbf{m}} = \frac{1}{\rho}\left(\frac{\sqrt{\lambda_3}-\sqrt{\lambda_1}}{\sqrt{\lambda_3-\lambda_1}}\right)\Big(-\sqrt{1-\lambda_1}\hat{\mathbf{e}}_1 + \kappa\sqrt{\lambda_3-1}\hat{\mathbf{e}}_3\Big) \tag{14}$$

with ρ a nonzero constant, and $\hat{\mathbf{e}}_1$ and $\hat{\mathbf{e}}_3$ are the eigenvectors of $\mathbf{C}$ corresponding to the eigenvalues λ_1 and λ_3, respectively, and $\kappa = \pm 1$.

From Proposition 1, there are at most two solutions to equation (12), and for each solution, the rotation $\mathbf{R}$ can be found by direct substitution of the vectors $\mathbf{a}$ and $\hat{\mathbf{n}}$ back into equation (12).

2.1. Twinning

A feature of the present theory is that all twins and their modes are predicted directly. A macroscopic picture of a twin is shown in Fig. 2. The compatibility equation is

$$\mathbf{R}\mathbf{U}_i - \mathbf{U}_j = \mathbf{a}\otimes\hat{\mathbf{n}} \tag{15}$$

where $\mathbf{R}$ is the twin rotation, $\mathbf{a}$ is parallel to the twin shear, and $\mathbf{U}_j^{-1}\hat{\mathbf{n}}$ is parallel to the twin plane normal in the deformed configuration. The magnitude of the twin shear s is $|\mathbf{U}_j^{-1}\hat{\mathbf{n}}||\mathbf{a}|$. From Refs [24, 43], *if* there exists a rotation $\mathbf{Q} = \mathbf{Q}[180^\circ,\hat{\mathbf{p}}]$ in the austenite Laue group $\wp^{\mathrm{a}}$ (from equation (6), only those rotations in the full austenite point group with positive determinant need to be considered) such that $\mathbf{U}_j = \mathbf{Q}\mathbf{U}_i\mathbf{Q}$, then the solutions to the twinning equation (15) are

$$\hat{\mathbf{n}}^{\mathrm{I}} = \hat{\mathbf{p}}, \quad \mathbf{a}^{\mathrm{I}} = 2\left(\frac{\mathbf{U}_j^{-1}\hat{\mathbf{p}}}{\left|\mathbf{U}_j^{-1}\hat{\mathbf{p}}\right|^2} - \mathbf{U}_j\hat{\mathbf{p}}\right),$$
$$\mathbf{R}^{\mathrm{I}} = \left(-\mathbf{I} + \frac{2}{\left|\mathbf{U}_j^{-1}\hat{\mathbf{p}}\right|^2}\mathbf{U}_j^{-1}\hat{\mathbf{p}}\otimes\mathbf{U}_j^{-1}\hat{\mathbf{p}}\right)\mathbf{Q} \tag{16}$$

and

$$\hat{\mathbf{n}}^{\mathrm{II}} = \frac{2}{\rho}\left(\hat{\mathbf{p}} - \frac{\mathbf{U}_j^2\hat{\mathbf{p}}}{\left|\mathbf{U}_j\hat{\mathbf{p}}\right|^2}\right), \quad \mathbf{a}^{\mathrm{II}} = \rho\mathbf{U}_j\hat{\mathbf{p}},$$
$$\mathbf{R}^{\mathrm{II}} = \left(-\mathbf{I} + \frac{2}{\left|\mathbf{U}_j\hat{\mathbf{p}}\right|^2}\mathbf{U}_j\hat{\mathbf{p}}\otimes\mathbf{U}_j\hat{\mathbf{p}}\right)\mathbf{Q} \tag{17}$$

where ρ is a nonzero constant that can be chosen to normalize the twin plane normal. It can be immediately seen from the form of the rotation matrix $\mathbf{R}^{\mathrm{I}}$

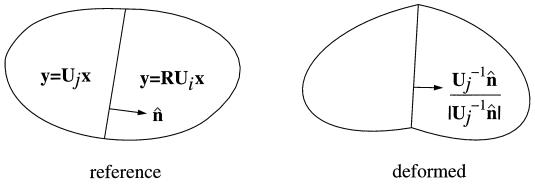

Fig. 2. Schematic of the twin microstructure.

and the Cauchy–Born rule [equation (2)] that the first solution [equation (16)] is a type I twin, while the second solution [equation (17)] is a type II twin. If there are two 180° rotations in $\wp^a$ relating the variants, then the solutions are both compound twins. Also, it can be shown that the twin shear s is the same for both twin solutions [equations (16) and (17)] and is given by

$$s = 2\sqrt{|\mathbf{U}_j\hat{\mathbf{p}}|^2|\mathbf{U}_j^{-1}\hat{\mathbf{p}}|^2 - 1}.$$

Moreover, the Cauchy–Born rule can be used to draw the lattice picture of the twins, and further, all of the twinning elements, K_1, K_2, η_1, and η_2, can be found for each of the twin solutions above (see Refs [29, 36]).

Table 1 gives the results of the calculations for four of the five transformations given in the previous section. For the cubic to monoclinic transition with "cube-edge" variants [equation (10)] consult Ref. [39]. We can see from this table that all of the twin planes and twin shears can be found, and that the predictions agree with experimentally observed twins.

Of course, in some cases it is not true that there exists a rotation $\mathbf{Q} = \mathbf{Q}[180^\circ,\hat{\mathbf{p}}]$ in the austenite Laue group $\wp^a$ such that $\mathbf{U}_j = \mathbf{Q}\mathbf{U}_j\mathbf{Q}$. But in some of these cases (rarely), equation (15) still admits solutions (to find them, use directly Proposition 1) [39, 41, 42]. In such cases, the two variants are perfectly compatible across a pair of interfaces, but there is no mirror symmetry. An example occurs in $LaNbO_4$ [42]. In this case, all of the crystallographic information including the normals, relative rotations, atomic positions, etc., arise perfectly naturally from the present theory.

Table 1. Twin microstructures for various transitions, where the type of twins is indicated along with either a twin plane or twin shear, which are given relative to a basis parallel to the edges of a cubic unit cell. The unique number of possible realizations of a particular twin solution is indicated as well, along with some of the alloys exhibiting the various twins

Transition	Twin type	Number	Observed
Cubic to trigonal [34, 35]	Compound, {100} and {110} twin planes	24	Au–Cd, Tb–Dy–Fe_2, Ti–Ni, Ti–Ni–Al, Ti–Ni–Fe
Cubic to tetragonal [21–23, 36]	Compound, {110} twin planes	12	$BaTiO_3$, Fe–Ni–C, In–Tl, Ni–Al, Ni–Mn, Ni_2MnGa
Cubic to orthorhombic [23, 32, 33, 38, 44]	Compound, {100} twin planes	12	Cu–Al–Ni
	Type I, {110} twin planes	24	Cu–Al–Ni
	Type II, ⟨110⟩ twin shears	24	Cu–Al–Ni
Cubic to monoclinic [31, 33, 39]	Compound, {100} and {110} twin planes	24	Ti–Ni, Ti–Ni–Cu
	Type I, {100} and {110} twin planes	72	Ti–Ni, Ti–Ni–Cu
	Type II, ⟨100⟩ and ⟨110⟩ twin shears	72	Ti–Ni
	Not standard type, possible for special lattice parameters	96	Not possible for lattice parameters of Ti–Ni

2.2. *Austenite–martensite microstructures*

At the critical temperature θ_c, both phases can co-exist in a specimen giving rise to the austenite–martensite microstructures. These microstructures provide low-energy paths through which a specimen can transform.

2.3. *Austenite–single variant of martensite interface*

The simplest austenite–martensite microstructure is two adjacent regions one with gradient on the

Table 2. Austenite–single variant of martensite interfaces. The unique number of possible realizations of this particular interface is also indicated

Transition	Number/Restrictions	Observed
Cubic to trigonal [34]	Not possible	
Cubic to tetragonal [36]	3 if $\alpha = 1$	
Cubic to orthorhombic [38]	12 if $\alpha = 1$, $\beta < 1$, and $\gamma > 1$ 12 if $\beta = 1$, $\alpha < 1$, and $\gamma > 1$	Ti–Ta, Ti–Ni–Cu
Cubic to monoclinic in Ti–Ni alloy [31]	Not possible for lattice parameters of Ti–Ni	
Cubic to monoclinic in 6M martensites [40]	24 if 6M unit cell	Cu–Al–Ni, Cu–Zn, Cu–Zn–Al, Cu–Zn–Ga

austenite energy well and the other with gradient on a single martensite energy well. From equation (12), the compatibility equation with variant $\mathbf{U}_i$ is

$$\mathbf{R}\mathbf{U}_i - \mathbf{I} = \mathbf{b} \otimes \hat{\mathbf{m}} \tag{18}$$

where the unknowns are the rotation $\mathbf{R}$ and the vectors $\mathbf{b}$ and $\hat{\mathbf{m}}$. The vector $\mathbf{b}$ is the shape strain, and $\hat{\mathbf{m}}$ is the habit plane normal. From Proposition 1, the necessary and sufficient condition that solutions to equation (18) exist is that the symmetric stretch $\mathbf{U}_i^2$ has ordered eigenvalues $\lambda_1 \leq \lambda_2 = 1 \leq \lambda_3$. If this condition is satisfied, then the shape strain $\mathbf{b}$ and habit plane normal $\hat{\mathbf{m}}$ are given by equations (13) and (14), respectively (see Ref. [36]).

For all of the symmetry lowering transformations, the austenite–single variant of martensite interface is possible with special lattice parameters, namely lattice parameters that give the needed eigenvalues. Two alloys that do exhibit this type of interface are certain special compositions of Ta–Ti [45] and Ti–Ni–Cu [46]. For both of these material systems, the composition was varied in order to satisfy the condition on eigenvalues. Some 9R and 18R structures also very nearly satisfy this condition (see Section 3). Table 2 gives the restrictions on the transformation stretches for several transitions in order for this interface to be possible.

2.4. Austenite–twinned martensite microstructure

As is well known, the austenite–twinned martensite microstructure is governed by the crystallographic theory of martensite. So, our only purpose here is to show how this structure emerges from energy minimization and to summarize the less well-known restrictions on lattice parameters that permit this microstructure. The basic structure in the reference configuration is shown in Fig. 3. In order for the deformation to be continuous, a transition layer is introduced between the austenite region and the twinned martensite region. This transition layer necessarily involves deformations

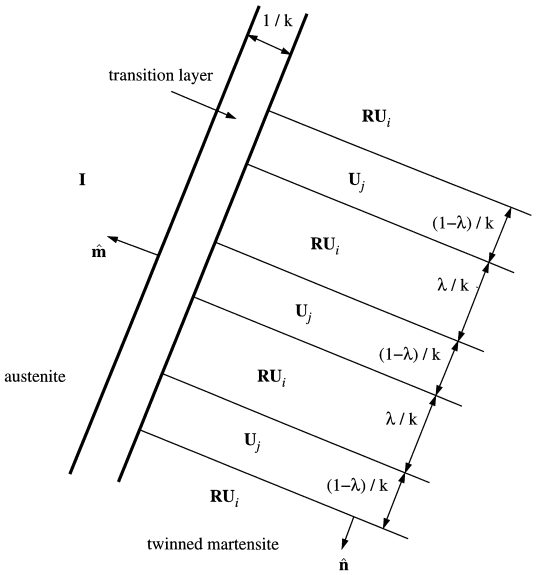

Fig. 3. Austenite–twinned martensite microstructure.

with gradients not on the energy wells, and so, the microstructure is not an energy minimizer. If one scales the width of the martensite layers and the width of the transition layer by $1/k$ as shown in Fig. 3, one gets a sequence of deformations; in the limit as $k \to \infty$, the energy of this sequence converges to the absolute minimum, and the limiting interface between the two phases is a plane [21]. Necessary and sufficient conditions that the energy of the transition layer tends to zero are that the twinning equation (15) and the equation

$$\bar{\mathbf{R}}\big(\lambda \mathbf{R}\mathbf{U}_i + (1-\lambda)\mathbf{U}_j\big) - \mathbf{I} = \mathbf{b} \otimes \hat{\mathbf{m}} \tag{19}$$

have a solution, the latter to be solved for ($\bar{\mathbf{R}} \in$ SO(3), $\mathbf{b}$, $\hat{\mathbf{m}}$). This is equivalent to the standard equation of the crystallographic theory of martensite [47]. The translation to the notation of the crystallographic theory follows by substituting for $\mathbf{R}\mathbf{U}_i$ using equation (15), then writing:

$$\underbrace{\mathbf{R}}_{\bar{\mathbf{R}}} \quad \overbrace{(\mathbf{I} + \lambda \mathbf{a} \otimes \mathbf{U}_j^{-1}\hat{\mathbf{n}})}^{\mathbf{P}_2} \quad \underbrace{\mathbf{B}+\mathbf{I}}_{\mathbf{U}_j} = \overbrace{\mathbf{I} + \mathbf{b} \otimes \hat{\mathbf{m}}}^{\mathbf{P}_1} \tag{20}$$

But, conceptually, there is a major advantage of writing the equation in the form of equation (19). That is, the matrix $\lambda \mathbf{R}\mathbf{U}_i + (1-\lambda)\mathbf{U}_j$ in parentheses in equation (19) is precisely the macro-scale deformation gradient of the twinned martensite. Therefore, equation (19) is, just like equation (15), simply the compatibility equation between two deformation gradients. Once this is noticed, the realization that many more complicated energy minimizing microstructures can be constructed by checking similar conditions becomes clear.

An austenite–twinned martensite microstructure is possible with pairs of distinct variants $\mathbf{U}_i$ and $\mathbf{U}_j$

Table 3. Austenite–twinned martensite microstructure. The unique number of possible realizations of this particular microstructure is given as well

Transition	Twin type	Number	Observed
Cubic to trigonal [34]	Compound, {100} and {110} twins	36 for $\psi > 90°$	
	Compound, {110} twins	12 for $\psi < 90°$	Au–Cd, Ti–Ni
Cubic to tetragonal [21, 23, 36]	Compound	24	Fe–Ni–C, In–Tl, Ni–Al, Ni–Mn
Cubic to orthorhombic [23, 32, 38]	Compound	24	
	Type I	48	Cu–Al–Ni
	Type II	48	Cu–Al–Ni
Cubic to monoclinic in Ti–Ni alloy [31, 33]	Type I, {100} and {110} twins	96	
	Type II, ⟨100⟩ and ⟨110⟩ twins	96	Ti–Ni

if a solution exists to the twinning equation (15) and the habit plane equation (19). Again use is made of Proposition 1. From Theorem 7 of Ref. [21], necessary and sufficient conditions that there is a solution are

$$\delta \stackrel{\text{def}}{=} \mathbf{a} \cdot \mathbf{U}_j \left(\mathbf{U}_j^2 - \mathbf{I} \right)^{-1} \hat{\mathbf{n}} \leq -2 \tag{21}$$

and

$$\eta \stackrel{\text{def}}{=} \operatorname{tr}\left(\mathbf{U}_j^2 \right) - \det\left(\mathbf{U}_j^2 \right) - 2 + \frac{|\mathbf{a}|^2}{2\delta} \geq 0. \tag{22}$$

The volume fraction λ of variant i is found from

$$\lambda = \frac{1}{2}\left(1 - \sqrt{1 + \frac{2}{\delta}} \right) \tag{23}$$

with the function δ from equation (21). Formulas for the solutions are given in Ref. [21]. Detailed studies of lattice parameters satisfying equations (21) and (22) can be found in Refs [21, 23, 24, 31, 33, 34, 36, 38, 41]. The occasional statement (used in the literature on constitutive equations) about the "24 habit planes" in nontetragonal martensites is incorrect (Table 3).

3. COMPLEX CRYSTAL STRUCTURES: 18R MARTENSITES

Recently, the 18R martensites have been included within the present theory by Hane [40], following a revision of the nomenclature and choice of the unit cell by Otsuka *et al.* [48]. These transformations presented a long-standing difficulty for the theory, which was especially problematic because the alloys exhibiting this structure, such as Cu–Al–Ni, Cu–Zn, Cu–Zn–Al, and Cu–Zn–Ga [49], comprise an important class of shape-memory alloys.

The reason that the old view of 18R presented a difficulty arises from a fundamental problem in the interpretation of X-ray data for martensitic materials (see Ref. [50] for further discussion). In the present theory, the deformation plays a basic role, the energy wells being defined by special stretch matrices $\{\mathbf{U}_1, \ldots, \mathbf{U}_v\}$. But X-ray pictures do not give a deformation, they only give a crystal structure. Having crystal structures for austenite and a variant of martensite does not determine uniquely $\mathbf{U}_1$; one needs also the correspondence, and the problem is exacerbated in complex lattices because there can be many different correspondences that give similar values for $\mathbf{U}_1$.

The old description of 18R (with its correspondence) gave a specific value for $\mathbf{U}_1$, and therefore by monoclinic symmetry for $\{\mathbf{U}_1, \ldots, \mathbf{U}_{12}\}$. In the phenomenological theory, in order to achieve compatibility between austenite and a single variant of martensite, a shear $\mathbf{K}$ was introduced which was defined solely by a shear parameter k (see Refs [51–55]). This shear parameter was adjusted so that the deformation gradients $\mathbf{I}$ and $\mathbf{RKU}_1$, with $\mathbf{R}$ some rotation matrix, satisfy a compatibility equation. Formally, the parameter k was interpreted as being related to the density of internal defects in the martensite phase (for example Refs [51–53, 55, 56]). The only way to make this fit into the present theory was to "extend the wells" by including with them a path of energy minimizers, parameterized by k, leading away from each well $\mathrm{SO}(3)\mathbf{U}_i$: by frame-indifference, this gave a huge set of energy-density-minimizing states, which seemed inconsistent with the mechanical behavior of these alloys.

In their study of Cu–Al, Nishiyama and Kajiwara [57] originally proposed two unit cells to describe the long-period stacking ordered structure of the martensite: 6M and 18R. Recently, Otsuka *et al.* [48] reconsidered the structure 6M. They favored 6M over 18R because the former correctly accounts for the monoclinic symmetry and stacking of the lattice, and this choice of a different unit cell implies a different correspondence. To understand the revision introduced by Otsuka *et al.* [48] we refer to Fig. 4. The parent phase from which is obtained all of the long-period stacking ordered structures under consideration is either a B_2 or a DO_3 ordered structure. The DO_3 structure is obtained by stacking the planes A_1 and B_1 as shown in Fig. 4 in the sequence A_1B_1, while the B_2 structure is contained in the DO_3 structure by considering the same planes, but with lattice parameter $a_0/2$. Both of these lattices are cubic if the ordering is neglected. Following Nishiyama and Kajiwara [57], the transformation from the parent phase to the long-period stacking ordered structures is assumed to take place by a contraction along the [100] direction and an

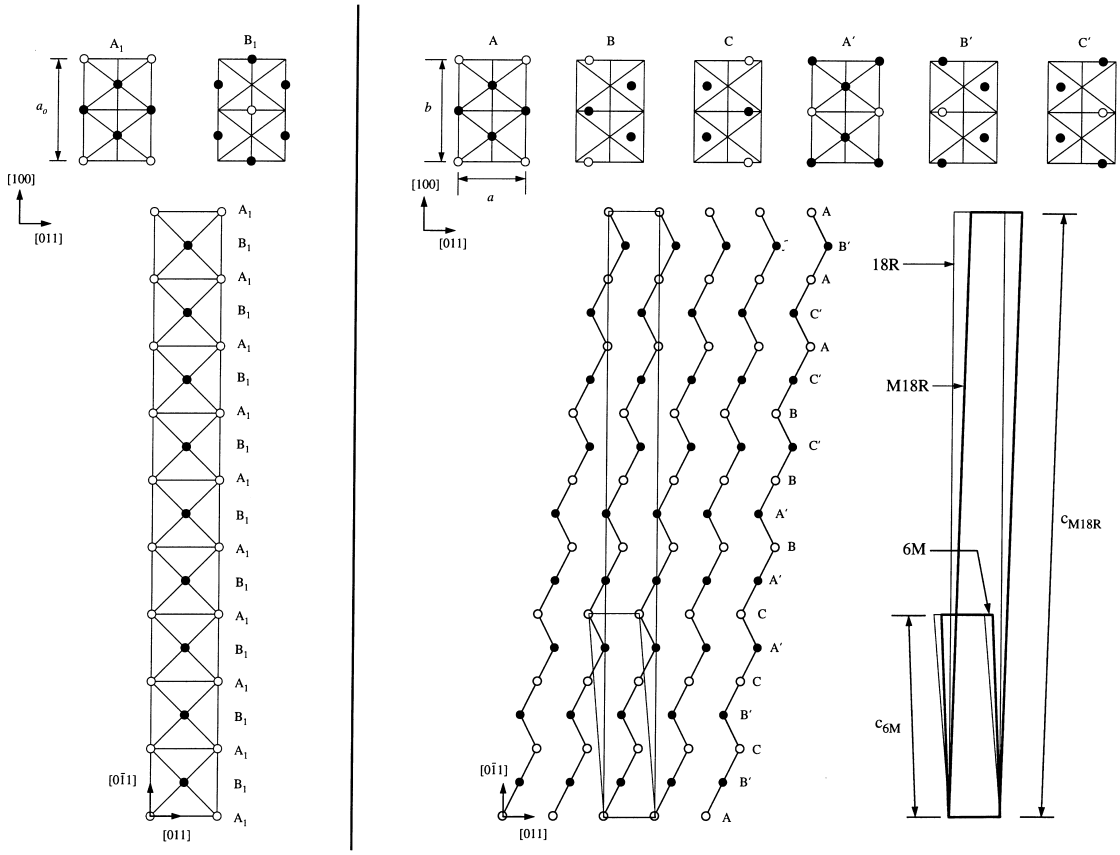

Fig. 4. At the left is shown the DO_3 crystal structure along with the $(01\bar{1})$ planes and its stacking sequence. In the middle is the 18R martensite, where the $(01\bar{1})$ close-packed planes with ideal stacking are given at the top. The rectangle outlined by the thin line is the 18R unit cell and its corresponding 6M unit cell is indicated as well. At the right is shown the relationship between the 18R and M18R unit cells; while, the relationship between a 6M unit cell (thin) for 18R and a 6M unit cell (thick) for M18R is also indicated.

expansion along the [011] direction, in order to create $(0\bar{1}1)$ close-packed planes as shown in Fig. 4, and an expansion along the $[0\bar{1}1]$ direction. Certain sites within this unit cell become unstable and the atoms undergo shuffles in the [011] direction to more stable positions. It is this shuffling that creates the internal defects within the martensite phase, and it may happen that errors occur in the shuffling leading to stacking faults [58]. Long-period stacking ordered structures are made up of stacks of the $(0\bar{1}1)$ close-packed planes. Various stacking sequences of these planes are found in a number of shape-memory alloys. Some of the observed structures are denoted as 2H, 9R, and 18R in the Ramsdell notation [49].

For definiteness, consider those alloys with a high-temperature parent phase of a DO_3 ordered structure as shown in Fig. 4. The low-temperature product phase is a long-period stacking ordered structure, which is shown in Fig. 4 as well, where the $(0\bar{1}1)$ close-packed planes with ideal stacking are depicted at the top right. Note that all of the planes A, B, and C are equivalent, while all of the planes A′, B′, and C′ are equivalent. In Fig. 4, the thin line outlines the 18R unit cell. It is, however, found in experiments that the stacking positions deviate from the ideal positions; thus, a modified 18R structure, called M18R, is needed. Tadaki *et al.* [59] rationalize the deviation from the ideal stacking positions as being due to the differing radii of the atoms in the alloy. The M18R unit cell is outlined at the extreme right in Fig. 4 by the thick line, and the M18R is distinguished from the 18R by a monoclinic angle θ different from 90°.

On the other hand, each of the lattices, 18R and M18R, can be indexed by different unit cells and two such were proposed by Nishiyama and Kajiwara [57] and Otsuka *et al.* [48]. The unit cells called 6M are consistent with the accepted monoclinic symmetry of the lattice. One such 6M unit cell is indicated in Fig. 4 for each of the lattices 18R and M18R, and the correspondence for both M18R and 6M is evident from the figure. Both correspondences yield stretch matrices $\{\mathbf{U}_1, \ldots, \mathbf{U}_{12}\}$ of the "cube-edge" type given above in equation (10), with

Table 4. Lattice parameters for various alloys exhibiting the faulted martensites. The parameters given are for the 6M unit cells which are found from data for either the M9R or M18R cells using equation (24). The angle $\tilde{\theta}$ is that needed for compatibility from equation (25)

Alloy	Parent phase	Product phase
Cu–Zn–Al, 15 at.% Zn, 17 at.% Al [60] ($DO_3 \rightarrow 6M$)	$a_0 = 5.996$ Å	$a = 4.553$ Å, $b = 5.452$ Å, $c = 13.014$ Å, $\theta = 94.2°$ and $\tilde{\theta} = 94.5°$
Cu–Al–Ni, 14 wt% Al, 4 wt% Ni [61] ($DO_3 \rightarrow 6M$)	$a_0 = 5.836$ Å	$a = 4.430$ Å, $b = 5.330$ Å, $c = 12.79$ Å, $\theta = 95.6°$ and $\tilde{\theta} = 95.2°$
Cu–Zn–Ga, 20 at.% Zn, 12 at.% Ga [56] ($DO_3 \rightarrow 6M$)	$a_0 = 5.86$ Å	$a = 4.40$ Å, $b = 5.33$ Å, $c = 12.78$ Å, $\theta = 94.9°$ and $\tilde{\theta} = 94.5°$
Cu–Zn, 39.3 at.% Zn [59] ($B_2 \rightarrow 6M$)	$a_0 = 2.94$ Å	$a = 4.412$ Å, $b = 2.678$ Å, $c = 12.84$ Å, $\theta = 95.1°$ and $\tilde{\theta} = 94.5°$

the parameters $\alpha = \sqrt{2}a/a_0$ and $\beta = b/a_0$ and stretch γ and angle θ indicated below.

Typically, the lattice parameters a, b, c, and θ are reported for the M18R unit cell. Unfortunately, a common practice is for experimentalists to ignore the monoclinic angle (M18R) and report the structure as orthorhombic (18R). For both the unit cells, M18R and its corresponding 6M cell, the lattice parameters a and b are the same, while the lattice parameter c along the $[0\bar{1}1]$ direction and the monoclinic angle θ are different. Let c_{M18R} and θ_{M18R} denote the length c and angle θ, respectively, for the M18R unit cell, then the corresponding length and angle for its 6M unit cell are

$$c_{6\mathrm{M}} = \frac{c_{\mathrm{M18R}}\sin(\theta_{\mathrm{M18R}})}{3\sin(\theta_{6\mathrm{M}})} \quad \text{and}$$
$$\theta_{6\mathrm{M}} = \theta_{\mathrm{M18R}} + \tan^{-1}\left(\frac{a}{c_{\mathrm{M18R}}\sin(\theta_{\mathrm{M18R}})}\right), \tag{24}$$

respectively [40, 48]. The energy well structure obtained from M18R has the difficulty mentioned above, while that obtained from its corresponding 6M is reasonable, as we now explain.

Using the energy wells for cubic to monoclinic transformations of "cube-edge" type given in equation (10) and the 6M unit cell, the austenite is exactly compatible with martensite across an interface if and only if [40],

$$\cos^2\theta = \frac{(1-\alpha^2)(1-\gamma^2)}{\alpha^2\gamma^2} \tag{25}$$

where $\alpha = \sqrt{2}a/a_0$ and $\gamma = \sqrt{2}c_{6\mathrm{M}}/3a_0$. In fact, four particular alloys have lattice parameters of the cubic parent and monoclinic phases which very nearly satisfy the condition given by equation (25). To show this, let $\tilde{\theta}$ be the monoclinic angle that exactly satisfies equation (25), i.e. $\cos^2(\tilde{\theta}) = (1-\alpha^2)(1-\gamma^2)/\alpha^2\gamma^2$, using the measured values of α and γ for these alloys. From Table 4, we see that the measured θ is extremely close to $\tilde{\theta}$. If, after accounting for experimental error, stacking faults are needed at all to secure compatibility between austenite and martensite, then it will be a very low density. Using the measured lattice parameters a_0, a, b, and c, and the monoclinic angle $\tilde{\theta}$ given by equation (25), Hane [40] has calculated all of the exact austenite/martensite interfaces for these alloys, and has shown good agreement with experiment.† This nearly exact compatibility between austenite and martensite is a remarkable feature of the 6M structure.

These calculations indicate the importance of taking into account the deviation from the ideal stacking positions when measuring all of the lattice parameters of these martensites. Further, it appears that the deviation from the ideal is required in order to have compatibility between austenite and martensite, and that a nearly ideal stacking can be obtained after the completion of the austenite to martensite transformation and variant rearrangement under stress.

Recent experimental observations by Sun *et al.* [62] and Shield [63] have revealed unusual deformations under stress. In the former case, smooth inhomogeneous deformations are observed, having gradients apparently quite far from the energy wells. This could support the idea that the deformation gradients associated with the shear parameter k, if not the lowest, are still rather low energy deformations. The interesting measurements of elastic moduli by Rodriguez *et al.* [64] also support this idea: these authors report the moduli associated with the shear path as having value zero (albeit with an error bar of ± 20 GPa). If a low-energy path departs transversally from a point on the energy wells, then (by linearization near that point on the well), some particular wave speed must vanish for linear theory, and it would be interesting to measure directly that wave speed. So, in summary, the 6M correspondence seems now to be the best, but a full understanding of the nature of this apparent low-energy valley on the energy surface, $\varphi = \varphi(\mathbf{F},\theta)$, awaits further study.

4. SPECIAL LATTICE PARAMETERS AND DESIGN OF MATERIALS

Of special interest in the study of martensitic transformations are microstructures known as the wedge, triangle, and diamond, which are shown

† It is interesting to note that the predictions of habit planes in fact agree with earlier predictions based on the M18R structure and the shear parameter, and its associated high density of internal faults [51–53, 55, 56]. The reason for this is that in the earlier analysis by the phenomenological theory, the value of the shear parameter is precisely chosen so that the low-energy path goes from the old M18R wells to the correct 6M wells.

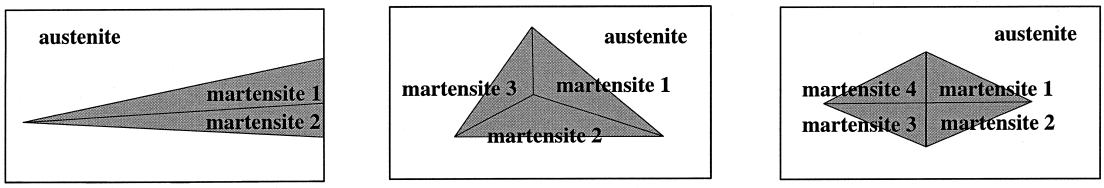

Fig. 5. From left to right, schematic of the wedge, triangle, and diamond microstructures. The gray regions can be twinned or single variant.

schematically in Fig. 5. Here we review the special relations among lattice parameters which are necessary and sufficient that such microstructures are compatible and energy minimizing.

Certain metallic alloys, which undergo a martensitic transformation, exhibit a spear or wedge microstructure. A wedge of martensite, which grows in a sea of austenite, consists of two regions of martensite separated by a planar interface called the midrib, and separated individually from the austenite phase by habit planes. The wedge nucleates and grows as a coherent structure with both habit planes and midrib plane appearing and growing together. On heating, the reverse transformation takes place with the wedge shrinking and finally disappearing (see Refs [23, 32, 36] and the references therein). The wedge is thought to be important for the thermoelasticity and reversibility of the transformation, because it provides a means by which the specimen can transform from a point on its boundary or grain boundary.

Within the context of the present theory, the wedge is constructed by fitting two austenite–martensite microstructures together coherently at the midrib, using only deformation gradients from the energy wells except on transition layers whose energy can be reduced to zero by refinement. Two different kinds of wedges have been considered: those in which both martensite regions are single variants of martensite; and those in which both martensite regions are twinned. Bhattacharya [23] gives necessary and sufficient conditions that the wedge is compatible and energy minimizing: the shape strains [$\mathbf{b}$ in equation (18) or equation (19)] of the two martensite plates making up each half of the wedge are parallel and the corresponding habit plane normals [$\hat{\mathbf{m}}$ in equation (18) or equation (19)] are not parallel. These conditions place rather strong restrictions on the lattice parameters. The most complete picture of these lattice parameter restrictions in various cases is given by Hane and Shield [36]. Among materials that clearly show the wedge, these restrictions are satisfied to a remarkable degree of accuracy (see Fig. 6).

The restrictions on lattice parameters are summarized in Tables 5 and 6. Notice that the wedges with single variants of martensite are only possible in materials with extremely special lattice parameters, but such materials have the potential for good shape memory, as discussed below. The general restrictions on the lattice parameters for a single variant wedge in both of the cubic to monoclinic transitions (''face-diagonal'' and ''cube-edge'' types) have not been worked out, but it is expected that such microstructures are possible in materials with lattice parameters which lie on surfaces in the space of lattice parameters: two restrictions on the lattice parameters are required.

The conditions for the twinned wedge are nearly satisfied in a Ti–Ni shape-memory alloy, but to date, they have not been unambiguously observed (see Ref. [31] and the references therein). One reason for this may be the well-known problem that the size scale of microstructures in Ti–Ni tends to be intermediate between optical and electron microscopy. Also, the calculations in Ref. [31] indicate that the conditions for the wedge are only approximately satisfied, and a small amount of additional elastic energy is required in order to make the de-

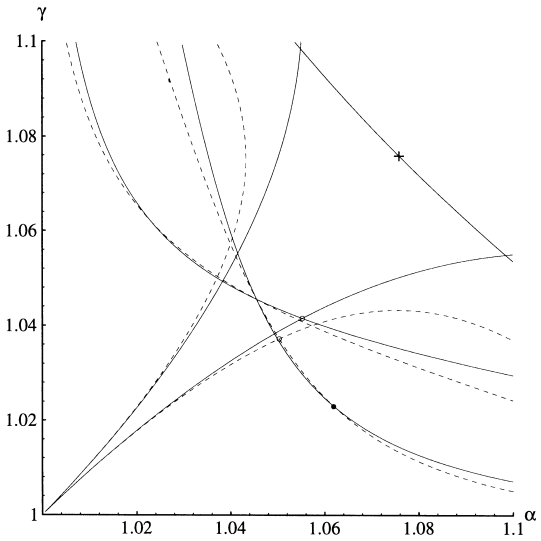

Fig. 6. Plot of the projections of the surfaces on which the twinned wedge microstructure is possible for the cubic to orthorhombic transition with variants given in equation (9). The transformation stretch β is fixed at the value for a Cu–Al–Ni shape-memory alloy, $\beta = 0.9178$ (see Refs [23, 32, 44] and the references therein). The solid curves are with type I twins and the dashed curves with type II twins. The filled circle indicates the other two transformation stretches for Cu–Al–Ni. The open circles are those stretches at which the diamond microstructure with twins is possible; and the point marked by a cross is an inadmissible point (see Ref. [32] for more details).

Table 5. Single variant wedges. The unique number of possible realizations of this microstructure is also indicated

Transition	Number/Restrictions	Observed
Cubic to trigonal [34]	Not possible	
Cubic to tetragonal [36]	Not possible	
Cubic to orthorhombic [38]	12 if $\alpha = 1$, $\gamma = \beta/\sqrt{3\beta^2 - 1}$	
	18 if $\beta = 1$, $\gamma = \alpha/\sqrt{2 - \alpha^2}$	
Cubic to monoclinic in Ti–Ni alloy [31]	Not possible for Ti–Ni	
Cubic to monoclinic in 6M martensites [40]	Not possible for Cu–Al–Ni, Cu–Zn, Cu–Zn–Al, Cu–Zn–Ga	

formation compatible. In any case, "wedge-like" structures are observed as part of the triangle morphology. Further, the 6M martensites in the specific alloys of Cu–Al–Ni, Cu–Zn, Cu–Zn–Al, and Cu–Zn–Ga with lattice parameters given in Table 4 cannot support the wedge; in fact, these materials have parameters that are far from the special lattices parameters for the wedge.

Other special microstructures involve fitting multiple wedges together coherently [36], so that they are completely surrounded by austenite. Two such microstructures are the triangle and the diamond, which are depicted in Fig. 5. In fact, the diamond is just two wedges back-to-back. These microstructures are possible only in materials with very special lattice parameters: typically, the lattice parameters must satisfy two restrictions which are summarized in Tables 7 and 8. Since the single variant wedge microstructure is not possible in the 6M martensites in the particular alloys considered, then neither are the triangle nor the diamond.

The diamond microstructure discussed above can be contrasted with a similar microstructure proposed by Schroeder and Wayman [65] and Saburi and Wayman [66]. They call this morphology a self-accommodating plate group. The plate group is considered to be common to many shape-memory alloys. In particular, a plate group is formed by four habit planes symmetrically arranged about a pole in such a manner that the average shape deformation is nearly the identity. The compatibility equations between the martensite regions are, however, not considered; in fact, it can be shown using the data given in Ref. [66] that none of the inner compatibility equations is satisfied for their diamond microstructures. Perhaps the observed microstructures are more complicated than as described. Similarly, in Refs [67, 68], the authors propose a triangle morphology to model a microstructure that appears in Ti–Ni shape-memory alloys. Hane and Shield [31] show that the proposed microstructure cannot be energy minimizing according to the present theory.

A material which undergoes, say, the cubic to orthorhombic transition and which has the special lattice parameters in order for a single variant wedge microstructure to be possible (Table 5) has the potential for good shape-memory properties. One reason is that such materials can form the triangle and diamond as well, which provides a simple mechanism for self-accommodation. Also, the absence of fine twinning and transition layers means much less energy needs to be used to create austenite–martensite interfaces, which should make the hysteresis in such alloys small, as is consistent with other low hysteresis alloys (certain 6M alloys and Ti–Ni–Cu alloys), which also have untwinned austenite–martensite interfaces. (For connection between low hysteresis and exact austenite/martensite interfaces, see Ref. [11].) In addition, such materials have the ability to form austenite–twinned martensite interfaces *for any volume fraction and any fineness of the twins*! This can be imagined by drawing several neighboring parallel single-variant wedges, so a jagged interface exists between austenite and martensite. Such materials could form an

Table 6. Twinned wedges. The unique number of possible realizations of this microstructure is indicated as well

Transition	Twin type	Number/Restrictions	Observed
Cubic to trigonal [34]	Compound, {100} twins	12 at $\psi = 116.4^\circ$	
Cubic to tetragonal [23, 36]	Compound	12 if $\alpha^2 = \dfrac{1 + 2\beta^2 + 5\beta^4}{1 - 2\beta^2 + 9\beta^4}$	Fe–Ni–C, Ni–Al, Ni–Mn
Cubic to orthorhombic [23, 32, 38]	Compound	12 if $\gamma^2 = \dfrac{\alpha^2\beta^2}{4\alpha^2\beta^2 - 2\alpha^2 - \beta^2}$	
	Compound	12 if $\gamma^2 = \dfrac{\alpha^2\beta^2 - 2\beta^2}{2\alpha^2 - 2\alpha^2\beta^2 - \beta^2}$	
	Type I	12 on surfaces	Cu–Al–Ni
	Type II	12 on surfaces	Cu–Al–Ni
	Types I and II (mixed twin)	24 on curve	
Cubic to monoclinic in Ti–Ni alloy [31]	Type I, {100} twins	12	Both possible for Ti–Ni, not observed
	Type II, ⟨110⟩ twins	12	

Table 7. Triangle microstructures. The unique number of possible realizations of this microstructure is indicated as well

Transition	Number/Restrictions	Observed
Cubic to trigonal [36]	Not possible	
Cubic to tetragonal [34]	4 if $\alpha = \sqrt{5/3}$, $\beta = \sqrt{1/3}$	
Cubic to orthorhombic [38]	4 if $\alpha = 1$, $\gamma = \beta/\sqrt{3\beta^2 - 2}$	
	12 if $\beta = 1$, $\gamma = \alpha/\sqrt{2 - \alpha^2}$	
	Not possible with twins	
Cubic to monoclinic in Ti–Ni alloy [31]	Not possible for Ti–Ni	
Cubic to monoclinic in 6M martensites	Not possible for Cu–Al–Ni, Cu–Zn, Cu–Zn–Al, Cu–Zn–Ga	

infinite variety of different microstructures at essentially zero energy.

It would be extremely interesting, from both practical and theoretical points of view, to seek alloys that satisfy the special lattice parameter relationships given here.

A warning about the literature. Often in the literature a picture is drawn of a microstructure, and corresponding strain matrices are given. It is then shown that the sum of these strains weighted by the volume fractions is zero, as an argument for self-accommodation (see Section 5, especially equation (36) for further discussion). The example often chosen is the diamond morphology. Above, we have explained that the diamond morphology is only possible with exceedingly special lattice parameters. But in the literature the strains that are given usually do not satisfy these restrictions, even approximately. Therefore, *the strains that are given do not correspond to the picture that is drawn*. That is, if pictures of reference and deformed configurations were generated using those strains (and introducing rotations as necessary to achieve compatibility where possible), the deformed picture would necessarily have gaps, surfaces with discrete slips, or would exhibit interpenetration of matter. In short, it would not be compatible. This practice is widespread and occurs in even some recent papers and reviews. Of course, the experimental observations do not justify the use of those ideal pictures. This practice is unfortunate for materials science, i.e. there are often compelling physical arguments given to suggest that these special microstructures are desirable for some interesting phenomenon such as the shape-memory effect. We feel that the use of special relations like the ones reviewed here could be used as a basis to search for new alloys, which would then be likely to exhibit those microstructures. While raising this criticism, the present authors are well aware that the methods reviewed here may not be so easily accessible, and the present treatment is intended to remedy this situation.

5. MICROSTRUCTURE AND THE MINORS RELATIONS

The main advances described in this review can be attributed to two developments: improved methods of constructing microstructures in the geometrically exact case, and the development of general restrictions on microstructure. Here we explain the latter in simple terms. As asserted above, we believe that the widespread use of these restrictions would enhance the quantitative understanding of a variety of materials in which "deformation" and "microstructure" play a role (e.g. materials undergoing coherent diffusional phase transformations, plastic deformation, or magnetostrictive/piezoelectric processes).

The minors relations are identities that connect microstructural deformation to macroscopic deformation. Physically, it is clear there must be some connection: if the maximum strain on the micro-scale is ε, then, no matter how the microstructure is arranged, the macro-scale strain cannot exceed ε. The minors relations apply only to the case of coherent (i.e. continuous) deformations, but the size of the deformation on the micro-scale is completely unrestricted, as is the complexity of the microstructure. To describe the relations, we first have to give a precise interpretation of the terms macro-scale and micro-scale. In the simplest view, we consider a region Ω in three-dimensional space. Suppose that the boundary of Ω, written as $\partial\Omega$, is subject to a

Table 8. Diamond microstructures. The unique number of possible realizations of this microstructure is also given

Transition	Number/Restrictions	Observed
Cubic to trigonal [34]	Not possible	
Cubic to tetragonal [36]	Not possible	
Cubic to orthorhombic [38]	3 if $\beta = 1$, $\gamma = \alpha/\sqrt{2 - \alpha^2}$	
	18 on curve with type I twins	
	18 on curve with type II twins	
Cubic to monoclinic in Ti–Ni alloy [31]	Not possible for Ti–Ni	
Cubic to monoclinic in 6M martensites	Not possible for Cu–Al–Ni, Cu–Zn, Cu–Zn–Al, Cu–Zn–Ga	

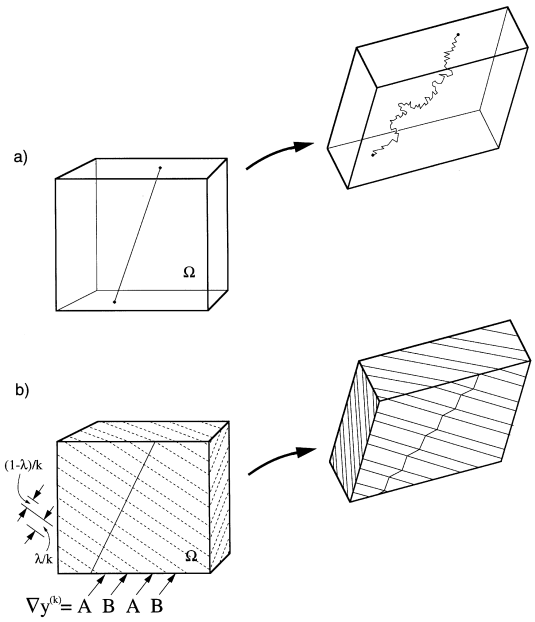

Fig. 7. Illustration of macro-scale and micro-scale deformations. (a) The macro-scale deformation is linear, and the micro-scale deformationis indicated for the inclined "scratch". (b) A layered micro-deformation consisting of compatible deformation gradients $\mathbf{A}$ and $\mathbf{B}$ with volume fraction λ and $(1-\lambda)$, respectively, and macroscopic deformation $\mathbf{y}_{\text{macro}}(\mathbf{x})=[\lambda\mathbf{A}+(1-\lambda)\mathbf{B}]\mathbf{x}$.

deformation $\mathbf{y}(\mathbf{x}) = \mathbf{G}\mathbf{x}$ for $\mathbf{x} \in \partial\Omega$ as shown in Fig. 7. The line passing through Ω represents a material line. After deformation, it becomes jagged, as shown in Fig. 7(a), but its deformed end positions are determined by the boundary deformation. Hence, we have two deformations $\mathbf{y}_{\text{macro}}(\mathbf{x})$ and $\mathbf{y}_{\text{micro}}(\mathbf{x})$. In Fig. 7(a), $\mathbf{y}_{\text{macro}}(\mathbf{x}) = \mathbf{G}\mathbf{x}$, and $\mathbf{y}_{\text{micro}}(\mathbf{x})$ describes the detailed deformation of the scratch. We ask the question: are there functions $\varphi(\mathbf{F})$ of the deformation gradient $\mathbf{F}$ that behave particularly well under averaging, in the sense that

$$\int_\Omega \varphi\big[\nabla\mathbf{y}_{\text{micro}}(\mathbf{x})\big]\mathrm{d}\mathbf{x} = \int_\Omega \varphi\big[\nabla\mathbf{y}_{\text{macro}}(\mathbf{x})\big]\mathrm{d}\mathbf{x} = \varphi(\mathbf{G}) \times \text{volume of } \Omega \tag{26}$$

holds for all $\mathbf{y}_{\text{micro}}(\mathbf{x})$ that satisfy the boundary conditions $\mathbf{y}_{\text{micro}}(\mathbf{x}) = \mathbf{y}_{\text{macro}}(\mathbf{x}) = \mathbf{G}\mathbf{x}$ on $\partial\Omega$? Such functions φ are called null Lagrangians. We can easily find a necessary condition on the null Lagrangians, by testing the relation of equation (26) against the most common micro-deformation observed in martensite [Fig. 7(b)]. Consider two compatible deformation gradients $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{B}-\mathbf{A} = \mathbf{a}\otimes\mathbf{n}$, in a simple layering with the volume fraction of $\mathbf{A}$ being $\lambda \in (0,1)$ as shown in Fig. 7(b). In this case, $\mathbf{y}_{\text{macro}}(\mathbf{x}) = [\lambda\mathbf{A} + (1-\lambda)\mathbf{B}]\mathbf{x}$. To make such a micro-deformation meet exactly these boundary conditions, one needs to add a small transition layer, but this can be done (as explained above) with a bounded gradient; and by refining the layers, the volume of this transition layer can be made arbitrarily small. Since the deformation gradient in the transition layer is bounded and its volume is arbitrarily small, the transition layer makes a negligible contribution to the left-hand side of equation (26). When equation (26) is evaluated for this particular micro/macro-deformation, we get (after dividing by the volume of Ω), the necessary condition

$$\lambda\varphi(\mathbf{A}) + (1-\lambda)\varphi(\mathbf{B}) = \varphi\big[\lambda\mathbf{A} + (1-\lambda)\mathbf{B}\big] \tag{27}$$

which must hold for all rank-one-connected matrices $\mathbf{A}$ and $\mathbf{B}$. Equation (27) is an algebraic condition on φ that was solved by Ericksen [69]: $\varphi(\mathbf{G})$ satisfies equation (27) if and only if it is a linear combination of the three functions $\mathbf{G}$, cof$\mathbf{G}$, and det$\mathbf{G}$. In components (repeated indices summed),

$$\varphi(\mathbf{G}) = \alpha_{ij}\mathbf{G}_{ij} + \beta_{ij}(\text{cof}\mathbf{G})_{ij} + \gamma\text{det}\mathbf{G} \tag{28}$$

Here, cof$\mathbf{G}$ stands for the 3×3 matrix of cofactors of $\mathbf{G}$ and α_{ij}, β_{ij} and γ are any constants. If $\mathbf{G}$ is invertible (the typical case for martensite), then cof$\mathbf{G}$ is given by the simple formula $\text{cof}\mathbf{G} = (\text{det}\mathbf{G})\mathbf{G}^{-\mathrm{T}}$. Also, we have the component formulas,

$$(\text{cof}\mathbf{G})_{kr} = \frac{1}{2}\varepsilon_{ijk}\varepsilon_{pqr}G_{ip}G_{jq} \text{ and}$$

$$\text{det}\mathbf{G} = \frac{1}{6}\varepsilon_{ijk}\varepsilon_{pqr}G_{ip}G_{jq}G_{kr}. \tag{29}$$

So far these are only necessary conditions that φ enjoys the averaging property of equation (26). But a little calculation shows that they are indeed sufficient. One method is to show by differentiation that

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \varphi\big[\mathbf{G} + t\nabla\mathbf{u}(\mathbf{x})\big]\mathrm{d}\mathbf{x} = 0$$

when φ is given by equation (28) and $\mathbf{u}$ satisfies zero boundary conditions but otherwise is arbitrary. Hence, the value of $\int_\Omega \varphi[\mathbf{G} + t\nabla\mathbf{u}(\mathbf{x})]\mathrm{d}\mathbf{x}$ at $t = 0$ is the same as its value at $t = 1$. If we put $\mathbf{y}_{\text{micro}}(\mathbf{x}) = \mathbf{G}\mathbf{x} + \mathbf{u}(\mathbf{x})$, we get equation (26).

Suppose now that a compatible micro-deformation assumes only the values $\mathbf{F}_1, \ldots, \mathbf{F}_n$ having corresponding (positive) volume fractions $\lambda_1, \ldots, \lambda_n$, with $\sum_{i=1}^n \lambda_i = 1$. Here, λ_i is defined as the volume of the subset of Ω where $\nabla\mathbf{y} = \mathbf{F}_i$, divided by the volume of Ω. We write equation (26) using equation (28) first with $\beta_{ij} = \gamma = 0$, then with $\alpha_{ij} = \gamma = 0$, and then with $\alpha_{ij} = \beta_{ij} = 0$, so as to isolate the three terms in equation (28), and then we use the arbitrariness of the coefficients. We get the following *minors relations*:

$$\mathbf{G} = \sum_{i=1}^n \lambda_i\mathbf{F}_i \tag{30}$$

$$\text{cof}\mathbf{G} = \sum_{i=1}^{n} \lambda_i \,\text{cof}\,\mathbf{F}_i \tag{31}$$

$$\det\mathbf{G} = \sum_{i=1}^{n} \lambda_i \det\mathbf{F}_i. \tag{32}$$

These are remarkable relations. With a few unimportant exceptions they are the only relations known that precisely relate the micro-deformation gradient to the macro-deformation gradient. Their remarkableness arises from the fact that one only needs to know the micro-volume fractions and gradients to impose them: they hold no matter what the pattern of microstructure. They have been used to deduce some facts about martensite that are quite non-obvious at the outset.

In applications to martensite, $\mathbf{F}_1, \ldots, \mathbf{F}_n$ typically come from the energy wells. For example, for any two compatible variants of martensite $\text{SO}(3)\mathbf{U}_1$ and $\text{SO}(3)\mathbf{U}_2$, it is possible to choose a basis such that they are diagonal with the forms $\text{diag}[\alpha,\beta,\gamma]$ and $\text{diag}[\beta,\alpha,\gamma]$ (see Ref. [22]). The relations given by equations (30)–(32) have been used to find all macro-deformations that can be produced by all possible microstructures consisting of these two variants of martensite [22]. The answer is $\mathbf{y}_{\text{macro}}(\mathbf{x}) = \mathbf{G}\mathbf{x}$ where $\mathbf{G}^{\text{T}}\mathbf{G}$ has the form:

$$\begin{pmatrix} C_{11} & C_{12} & 0 \\ C_{12} & C_{22} & 0 \\ 0 & 0 & \gamma^2 \end{pmatrix} \tag{33}$$

and C_{11}, C_{12}, and C_{22} satisfy the inequalities,

$$\begin{aligned} &C_{11}C_{22} - C_{12}^2 = \alpha^2\beta^2, \\ &C_{11} + C_{22} + 2C_{12} \le \alpha^2 + \beta^2, \quad \text{and} \\ &C_{11} + C_{22} - 2C_{12} \le \alpha^2 + \beta^2. \end{aligned} \tag{34}$$

Despite a lot of work, the set of all macro-deformations that can be obtained by using three tetragonal variants of martensite is still unknown (this has become known as "the three-well problem").

To explain in a very simple case how the minors relations are exploited, let us analyze the relation between macro-strain and micro-strain. A convex function is a function satisfying

$$f(\lambda_1\mathbf{A}_1 + \ldots + \lambda_n\mathbf{A}_n) \le \lambda_1 f(\mathbf{A}_1) + \ldots + \lambda_n f(\mathbf{A}_n)$$

for all choices of $\mathbf{A}_1, \ldots, \mathbf{A}_n$ and $\lambda_1, \ldots, \lambda_n$ with $\lambda_i \ge 0$ and $\sum \lambda_i = 1$. It can easily be seen that if $\mathbf{e}$ is a unit vector, then $f(\mathbf{A}) = |\mathbf{A}\mathbf{e}| - 1$ is a convex function. But, if $\mathbf{A}$ is a deformation gradient, then $|\mathbf{A}\mathbf{e}| - 1$ is the strain experienced by an elementary line in the direction $\mathbf{e}$, i.e. its [(deformed length)−(original length)]/original length. Applying this f to the first minors relation [equation (30)] and using directly the definition of convexity, we obtain:

$$\begin{aligned} |\mathbf{G}\mathbf{e}| - 1 &\le \lambda_1(|\mathbf{F}_1\mathbf{e}| - 1) + \ldots + \lambda_n(|\mathbf{F}_n\mathbf{e}| - 1) \\ |\mathbf{G}\mathbf{e}| - 1 &\le \max_{i\in\{1,\ldots,n\}} (|\mathbf{F}_i\mathbf{e}| - 1). \end{aligned} \tag{35}$$

That is, the macro-strain of any line element cannot exceed the largest micro-strain of the same line element. No such statement holds for the minimum strain: even if every micro-strain in the direction $\mathbf{e}$ is zero, then the macro-strain in the same direction can be arbitrarily close to -1. (For an example, plot the macro-deformation corresponding to Fig. 7(b) with $\mathbf{A} = \mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2 + \mathbf{f}^+ \otimes \mathbf{e}$ and $\mathbf{B} = \mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2 + \mathbf{f}^- \otimes \mathbf{e}$, $\lambda = 1/2$, where $|\mathbf{f}^{\pm}| = 1$ and $|\mathbf{f}^+ + \mathbf{f}^-| \ll 1$, $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}\}$ an orthonormal basis [70].) Analysis of the minors relations often involves eliminating $\mathbf{G}$ between two minors relations and using convexity in some way.

Self-accommodation refers to the existence of a microstructure of martensite completely surrounded by unstressed austenite. It has an obvious importance for the ease of transformation, especially in polycrystals. The framework for questions of self-accommodation is exactly the one adopted here, specialized to $\mathbf{G} = \mathbf{I}$. Using the minors relations, Bhattacharya ([24], Table 3.1) has found necessary and sufficient conditions on the matrices $\mathbf{U}_1, \ldots, \mathbf{U}_\nu$ that permit self-accommodation. All of these arguments follow the same pattern: use the minors relations to derive some restrictions on $\mathbf{U}_1, \ldots, \mathbf{U}_\nu$, then explicitly construct a family of micro-deformations that satisfies the restrictions.

In even the recent literature on martensite, the property of self-accommodation is analyzed using a procedure that is essentially misleading. It is useful to describe its limitations here. The minors relations apply to any gradient, not just the deformation gradient. So we can apply them to the displacement gradient $\nabla\mathbf{u}$ of geometrically linear theory, with the understanding of the inherent errors in geometrically linear theory described above (see Ref. [33]). So, let the micro-scale displacement have gradients $\mathbf{H}_1, \ldots, \mathbf{H}_n$ with corresponding volume fractions $\lambda_1, \ldots, \lambda_n$, belonging to the ν energy wells $\{\mathbf{E}_1, \ldots, \mathbf{E}_\nu\}$. (Of course, here and in equations (30)–(32), n can be much larger than ν because a variant can have different rotations.) In geometrically linear theory, the macroscopic displacement associated with self-accommodation is $\mathbf{u}_{\text{macro}} = 0$. Each $\mathbf{H}_i$ is of the form $\mathbf{E}_i + \mathbf{W}_i$, where $\mathbf{E}_i$ belongs to the set $\{\mathbf{E}_1, \ldots, \mathbf{E}_\nu\}$ and $\mathbf{W}_i$ is skew. The first minors relation is then,

$$0 = \lambda_1\mathbf{H}_1 + \lambda_2\mathbf{H}_2 + \ldots + \lambda_n\mathbf{H}_n. \tag{36}$$

Add the transpose of equation (36) to itself to get rid of all the skew matrices, then collect the volume fractions corresponding to a single variant (i.e. put $\bar{\lambda}_i = \sum \lambda_j$ where the sum is taken over all λ_j corresponding to variant i). We get,

$$0 = \bar{\lambda}_1 \mathbf{E}_1 + \ldots + \bar{\lambda}_\nu \mathbf{E}_\nu. \quad (37)$$

In the literature on martensite, equation (37) is often used to judge whether a microstructure is self-accommodating. That is not true. The truth is, even in geometrically linear theory, one can have strains $\{\mathbf{E}_1, \ldots, \mathbf{E}_\nu\}$ satisfying equation (37), but there does not exist a compatible microstructure having those strains.

For a deeper analysis let us return to geometrically exact theory. The use of equation (37) is essentially like using only the first minors relation equation (30). For variants of martensite, the third minors relation [equation (32)] is automatically satisfied if the variants are chosen to have determinant 1, and everyone recognizes that this must be satisfied for self-accommodation. So what is really being omitted in confining attention to equation (37) is the second minors relation of equation (31). But equation (31) is a very strong restriction! For example, the second minors relation is used in a crucial way to derive the results of equations (33) and (34). Also, Bhattacharya [24] shows using the minors relations that the plate group associated with a diamond morphology does not form a coherent deformation satisfying homogeneous boundary conditions.

That raises a natural question. Suppose all three minors relations [equations (30)–(32)] are satisfied by $\mathbf{G}$, $\mathbf{F}_1, \ldots, \mathbf{F}_n$ and corresponding volume fractions $\lambda_1, \ldots, \lambda_n$. Does that mean there is a compatible micro-deformation with gradients $\mathbf{F}_1, \ldots, \mathbf{F}_n$ on regions with volume fractions $\lambda_1, \ldots, \lambda_n$, meeting boundary conditions $\mathbf{y}_{\text{micro}}(\mathbf{x}) = \mathbf{G}\mathbf{x}$, allowing for the possibility of vanishingly small transition layers? Unfortunately, the answer is no. They are only necessary conditions. But in surprisingly many cases of interest they are effective. The study of what are the necessary and sufficient conditions on ($\mathbf{G}$, $\mathbf{F}_1, \ldots, \mathbf{F}_n$) has led to a rapidly growing subfield of mathematics which grew from studies of martensitic transformations [71].

In this section, we have only allowed very simple macro-deformations, just linear ones. In fact, it is of interest to be able to treat nonlinear macro-deformations, too. In fact, as can be imagined, the minors relations in the form of equations (30)–(32) hold in the general case when there is a separation of scales. This means that typical micro-scale oscillations of the deformation gradient occur on a scale that is much smaller than the length scale oscillations of the macro-scale deformation gradient. $\mathbf{G}$ is in that case the macro-scale deformation gradient.

6. FERROMAGNETIC AND FERROELECTRIC MARTENSITES

In recent years, people have begun exploring martensitic materials that are also either ferromagnetic or ferroelectric. Most of the research has focused on the ferromagnetic case. The presence of ferromagnetism or ferroelectricity offers a new "handle" on martensitic microstructure: by applying a field, there exists the possibility either of inducing the transformation between austenite and martensite or of rearranging the variants of martensite.

Below we shall review explicitly the magnetic case and point out as we go along the modifications that are necessary in the case of ferroelectric martensites.

It has been known for some time that martensite in Fe–Ni, Fe–Ni–C, Fe–Mn–C, Fe–Ni–Co–Ti, and Fe–Pt alloys is ferromagnetic, and that M_s is shifted by a magnetic field. As shown by Shimizu and Kakeshita [72], this shift is described by a version of the Clausius–Clapeyron (C–C) equation (in the literature on martensite, the C–C equation is attributed to Patel and Cohen [73], who first applied it to martensitic transformations). In the Fe–Ni materials, very large fields are required to shift the transformation, on the order of 10 T for a 20°C shift. The recent interest has focused on the search for materials in which a shape change can be induced by small fields, the magnetic version of shape-memory materials.

As will be explained in more detail below, the field-induced redistribution of martensite and the field-induced austenite/martensite transformation can both be considered a form of magnetostriction. That is, the shape change is accompanied by a change of the local state of magnetization. In this sense, they are not completely different from giant magnetostrictive materials such as $TbDyFe_2$ (Refs [35, 74] exploit this similarity to analyze its domain structures). The main difference is that the martensitic transformation is first-order, and the austenite can be ferromagnetic, while the transition to the ferromagnetic state in $TbDyFe_2$ is second-order.

Consider a material that undergoes both a first-order martensitic transformation and a ferromagnetic transition. Ferromagnetic transitions are typically second-order: from a theoretical viewpoint, they are discovered by linearizing the equilibrium equations about the unstressed austenitic or martensitic state and looking for a bifurcation. From a physical viewpoint, the ferromagnetic state collapses continuously to the nonferromagnetic state as the temperature is raised to the Curie point. Therefore, there are three fundamental temperatures in a martensitic material: the austenite/martensite transformation temperature, the Curie temperature for the austenite and the Curie temperature for the martensite.† These three temperatures can be ordered in six different ways, leading to the various qualitative

† The austenite/martensite transformation temperature can be further expanded to the four temperatures M_s, M_f, A_s, and A_f, but we shall focus only on the temperatures with thermodynamic significance.

behaviors [75, 76]. For example, it is theoretically possible to have a martensitic transformation in which the austenite is ferromagnetic and the martensite is not, or vice versa. If the Curie temperature for martensite is above the austenite/martensite transformation temperature, then upon heating, this Curie temperature would not be observed because the specimen would first transform to austenite. In such cases, it could be observed by first stabilizing the martensite using stress, then heating.

The magnetic state of either the austenite or the martensite is characterized by its magnetization $\mathbf{m}$. This is a vector-field defined on the deformed configuration, $\mathbf{y}(\Omega)$. The direction of the magnetization is affected by applied fields and stress, but its magnitude is a function of temperature only: $|\mathbf{m}(\mathbf{y})| = m_s(\theta)$†, where m_s is the saturation magnetization. In ferromagnetic shape-memory materials, there are in general different saturation magnetizations $m_s^a(\theta)$ and $m_s^m(\theta)$, respectively, for the austenite and for the martensite, and each is expected to obey approximately the modified Curie–Weiss law. The condition $|\mathbf{m}(\mathbf{y}(\mathbf{x}))| = m_s^a(\theta)$ applies for $\mathbf{x}$ such that $\nabla\mathbf{y}(\mathbf{x})$ is near the austenite well, and $|\mathbf{m}(\mathbf{y}(\mathbf{x}))| = m_s^m(\theta)$ applies for $\mathbf{x}$ such that $\nabla\mathbf{y}(\mathbf{x})$ is near the martensite wells. Thus, there is typically a jump in the saturation magnetization at the austenite/martensite transformation temperature. This is true in the widely studied alloy systems Ni_2MnGa and Fe_3Pd. The presence of this jump implies the possibility of field-induced transformation.

To understand quantitatively what can happen in such a material, we need an expression for its free energy. The general form can be adapted from Brown [77] and James and Kinderlehrer [35, 74]. The free energy density includes a dependence on magnetization, as well as deformation gradient and temperature:

$$\varphi(\mathbf{F},\mathbf{m},\theta). \tag{38}$$

The condition of frame-indifference is $\varphi(\mathbf{RF},\mathbf{Rm},\theta) = \varphi(\mathbf{F},\mathbf{m},\theta)$, which holds for all rotation matrices $\mathbf{R}$ in SO(3) and all values of $(\mathbf{F},\mathbf{m},\theta)$. Hence, energy wells will always have the form $(\mathbf{RU},\mathbf{Rm}),\mathbf{R} \in \mathrm{SO}(3)$. (Important: the same $\mathbf{R}$ goes in front of both $\mathbf{U}$ and $\mathbf{m}$, so deformation gradient and magnetization are rotated simultaneously.) Care has to be exercised in writing the total free energy, because the deformation is defined on the reference configuration, Ω, while the magnetization is defined on the deformed configuration; we follow Refs [35, 74].

Suppose that the applied field is $\mathbf{h}$. $\mathbf{h}$ is the field that would be present if the specimen were removed. In tests, it is usually arranged to be constant. The applied field energy (or Zeeman energy) is:

$$-\int_{\mathbf{y}(\Omega)} \mathbf{h}\cdot\mathbf{m}(\mathbf{y})\,d\mathbf{y} = -\int_{\Omega} \mathbf{h}\cdot\mathbf{m}(\mathbf{y}(\mathbf{x}))\left|\det\nabla\mathbf{y}\right|d\mathbf{x}. \tag{39}$$

Finally, there is a contribution to the energy that has no counterpart for ordinary martensite, namely the demagnetization energy (abbreviated, demag.). This is a nonlocal energy. That is, its value for the union of two disjoint regions is not typically equal to its sum evaluated for the two regions separately. It is obtained by first solving the magnetostatic equation

$$\mathrm{div}(-\nabla\zeta + 4\pi\mathbf{m}) = 0 \quad \text{(cgs units)} \tag{40}$$

for the *magnetostatic potential* ζ on all of space, where $\mathbf{m}$ has been extended to be zero outside $\mathbf{y}(\Omega)$. Given such a magnetization $\mathbf{m}$, this equation has a unique solution ζ up to an additive constant. The demag. energy is then,

$$\frac{1}{8\pi}\int_{\mathbb{R}^3}\left|\nabla\zeta\right|^2 d\mathbf{y} = \frac{1}{2}\int_{\mathbf{y}(\Omega)} \nabla\zeta\cdot\mathbf{m}\,d\mathbf{x} \tag{41}$$

where the second form follows from equation (40) (multiply by ζ and integrate by parts). Finally, there are exchange and strain-gradient energies, which we omit because they are unimportant for large bodies (cf. Ref. [78]). Collecting the various contributions, the total energy is

$$\int_{\Omega}\left\{\varphi\left[\nabla\mathbf{y}(\mathbf{x}),\mathbf{m}(\mathbf{y}(\mathbf{x})),\theta\right] - \mathbf{h}\cdot\mathbf{m}(\mathbf{y}(\mathbf{x}))\left|\det\nabla\mathbf{y}\right|\right\}d\mathbf{x} + \frac{1}{8\pi}\int_{\mathbb{R}^3}\left|\nabla\zeta\right|^2 d\mathbf{y}. \tag{42}$$

We note the following: the demag. energy is zero if $\mathrm{div}\,\mathbf{m} = 0$, which is to be interpreted in the following sense:

$$\begin{aligned}&\mathrm{div}\,\mathbf{m} = 0, \quad \text{where } \mathbf{m} \text{ is smooth, and}\\ &[\mathbf{m}_2 - \mathbf{m}_1]\cdot\mathbf{n}' = 0, \quad \text{at the interfaces with normal } \mathbf{n}'.\end{aligned} \tag{43}$$

Here, $\mathbf{m}_1$ and $\mathbf{m}_2$ are the values of the magnetization on each side of the interface. The latter has the interpretation that there are no poles on the interfaces of discontinuity of the magnetization.

There is an important distinction between ferroelectric and ferromagnetic martensites. In ferroelectric martensites, contributions to the demag. energy from discontinuities of the polarization at the boundary $\partial\mathbf{y}(\Omega)$ can be eliminated by the use of

† Brown [77] advocates the saturation condition $|\det\nabla\mathbf{y}(\mathbf{x})|\mathbf{m}(\mathbf{y}(\mathbf{x})) = m_s(\theta)$, representing the condition that the dipole moment per unit deformed volume is only a function of temperature. In the case of ferromagnetic shape memory, this may at best apply near each of the energy wells. Currently, the data are not sufficiently accurate near the wells to decide between $|\mathbf{m}(\mathbf{y})| = m_s(\theta)$ and $|\det\nabla\mathbf{y}(\mathbf{x})|\mathbf{m}(\mathbf{y}(\mathbf{x})) = m_s(\theta)$ so for simplicity we omit the factor $|\det\nabla\mathbf{y}(\mathbf{x})|$.

conductors. On such conductors, the electric potential ζ is constant, and in this case, equation (40) must be solved subject to boundary conditions that ζ is equal to a constant on the conductors and additional conditions on the total charge. Even with conductors absent, in ferroelectric martensites, free charges will eventually leak onto $\partial\mathbf{y}(\Omega)$ and effectively decrease the last term of equation (42). There are no free magnetic charges, but in ferromagnetic martensites, the same can be achieved by yokes of soft magnetic material. This is treated in equation (42) by enlarging Ω to include the yoke, and using an appropriate $\varphi(\mathbf{F},\mathbf{m},\theta,\mathbf{x})$ for the yoke material where $\mathbf{x}$ belongs to the yoke. Both yokes and conductors can be problematic for ferromagnetic or ferroelectric martensites because the latter may undergo large changes of shape.

We now use equation (42) to predict some aspects of the behavior of ferromagnetic shape-memory materials. We follow all of the developments of Section 2, but now recognize that each atom also has a magnetic moment, its dipole moment per unit volume, whose local spatial average is represented by $\mathbf{m}$. So any transformation that restores the lattice gives the same value of φ, as long as we use *the same value of* $\mathbf{m}$. Combining this condition with the condition of frame-indifference, we get that if $(\mathbf{U}, \mathbf{m})$ minimizes the energy density φ, then so does $(\hat{\mathbf{R}}\mathbf{U}\hat{\mathbf{Q}}, \pm\hat{\mathbf{R}}\mathbf{m})$, where $\hat{\mathbf{Q}}$ is any member of the Laue group of the austenite and $\hat{\mathbf{R}} \in \mathrm{SO}(3)$. Here, we have also included the $\pm$ representing the fundamental invariance of magnetism with respect to time reversals (see Refs [35, 74]). In ferroelectric martensites, one omits the $\pm$ invariance. For convenience, we choose $\mathbf{U}$ to be positive-definite and symmetric ($\mathbf{U}-\mathbf{I}$ is the Bain strain). Then it is convenient to put $\mathbf{Q} = \hat{\mathbf{Q}}^{\mathrm{T}}$ and $\mathbf{R} = \hat{\mathbf{R}}\hat{\mathbf{Q}}^{\mathrm{T}}$, and recognize that $\mathbf{Q}$ and $\mathbf{R}$ are arbitrary elements of the Laue group of austenite and SO(3), respectively. It follows that the energy wells of φ are of the form $(\mathbf{RQUQ}^{\mathrm{T}}, \pm\mathbf{RQm})$. In particular, they are given by

$$\begin{aligned}&\{(\mathbf{U}_1, \pm\mathbf{m}_1), \ \ldots\ , (\mathbf{U}_\nu, \pm\mathbf{m}_\nu)\}\\ &\quad = \{(\mathbf{QU}_1\mathbf{Q}^{\mathrm{T}}, \pm\mathbf{Qm}_1):\\ &\qquad \mathbf{Q} \in \text{Laue group of austenite}\}\end{aligned} \tag{44}$$

together with all rotations of these, i.e. $(\mathbf{RU}_i, \pm\mathbf{Rm}_i)$, $\mathbf{R} \in \mathrm{SO}(3)$. We wish to emphasize that, while not yet common in materials science, the precise statement of the conditions of invariance and the precise placement of the energy wells is absolutely crucial, since these determine all the subsequent twin modes, their types, the possible interfaces of magnetic domains, and more complicated microstructures of these.

For various transformations, the stretch matrices will have forms like those given in Section 2. For each stretch matrix, there is often a single easy axis

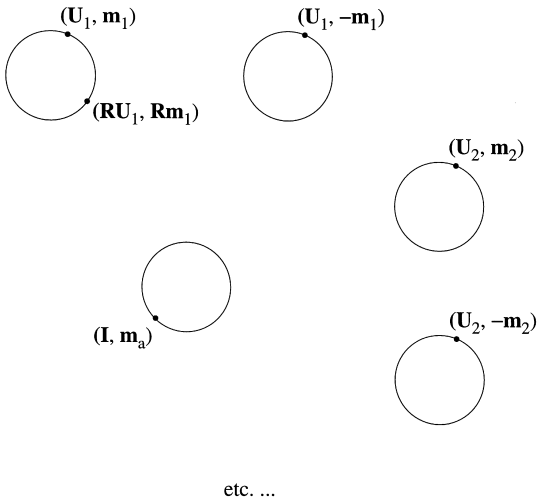

Fig. 8. Schematic of the energy wells for a ferromagnetic material.

of magnetization as indicated in equation (44), but it is consistent with equation (44) that, say, $\mathbf{QU}_1\mathbf{Q}^{\mathrm{T}} = \mathbf{U}_1$ but $\mathbf{Qm}_1 \neq \mathbf{m}_1$, in which case variant 1 would have a pair of rather particular easy axes. From the information that is currently available, this appears to happen in Fe_3Pd. The structure of the energy wells is illustrated schematically in Fig. 8.

First, consider redistribution of martensite variants induced by applying a field. Ideally, we start in variant 1 and apply a field so as to induce a transformation to variant 2. As expected and as observed, this is accompanied by the presence of a banded twin structure in which the volume fraction of variant 2 gradually increases (Fig. 9). An important question is whether there is a low-energy path connecting variants 1 and 2. We consider a simple banded structure as shown in Fig. 10, with alternating states $(\hat{\mathbf{F}}_1,\hat{\mathbf{m}}_1)$ $(\hat{\mathbf{F}}_2,\hat{\mathbf{m}}_2)$, and volume fraction λ. A low-energy path avoids poles on the interfaces (i.e. no contribution to demag. energy), has the states $(\hat{\mathbf{F}}_1,\hat{\mathbf{m}}_1)$ and $(\hat{\mathbf{F}}_2,\hat{\mathbf{m}}_2)$ on the energy wells, and has the deformation gradients compatible. That is $(\hat{\mathbf{F}}_2,\hat{\mathbf{m}}_2) = (\mathbf{RU}_2,\mathbf{Rm}_2)$ and $(\hat{\mathbf{F}}_1,\hat{\mathbf{m}}_1) = (\mathbf{U}_1,\mathbf{m}_1)$ with

$$\begin{aligned}&\mathbf{RU}_2 - \mathbf{U}_1 = \mathbf{a} \otimes \mathbf{n} \quad \text{and}\\ &(\mathbf{Rm}_2 \pm \mathbf{m}_1) \cdot \mathbf{U}_1^{-\mathrm{T}}\mathbf{n} = 0.\end{aligned} \tag{45}$$

Here, $\mathbf{n}$ is the reference normal; $\mathbf{n}' = \mathbf{U}_1^{-\mathrm{T}}\mathbf{n}\,[\|(\mathbf{RU}_2)^{-\mathrm{T}}\mathbf{n}]$ is the normal in the deformed configuration; and we have without loss of generality put the rotation in front of $(\mathbf{U}_1, \mathbf{m}_1)$ equal to $\mathbf{I}$.

Even though equation (45) is the twinning equation plus an additional equation and therefore, apparently very restrictive, it has solutions in many cases, which we now explain. In Section 2.1, we showed that if $\mathbf{U}_2 = \mathbf{QU}_1\mathbf{Q}^{\mathrm{T}}$ with $\mathbf{Q}$ a 180° rotation in the Laue group of the austenite, then the first

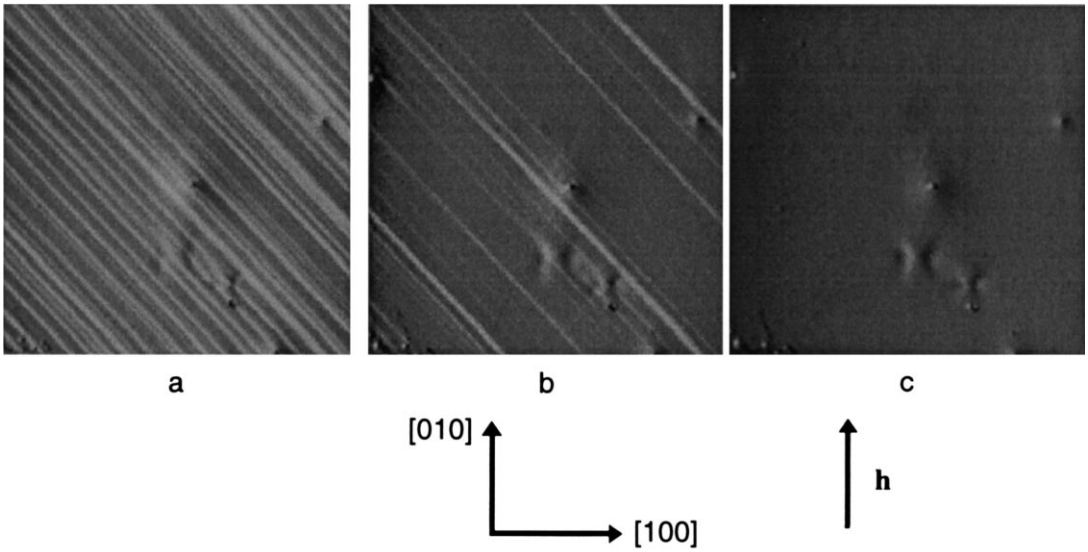

Fig. 9. Evolution of microstructure with increasing field in Ni_2MnGa. The field values are (a) 6000 Oe, (b) 10 000 Oe, and (c) 12 000 Oe, respectively [79].

expression in equation (45) has a pair of solutions $(\mathbf{R}^{\mathrm{I}},\mathbf{a}^{\mathrm{I}},\hat{\mathbf{n}}^{\mathrm{I}})$ and $(\mathbf{R}^{\mathrm{II}},\mathbf{a}^{\mathrm{II}},\hat{\mathbf{n}}^{\mathrm{II}})$ interpretable as type I/type II twins. Formulas for the solutions are given in equations (16) and (17). From equation (44), we get that for such $\mathbf{Q}$, $\mathbf{m}_2 = \mathbf{RQm}_1$. The forms of $\mathbf{RQ} = \mathbf{RQ}^{\mathrm{T}}$ for type I/type II twins are given in equations (16) and (17), from which we get immediately that,

$$\mathbf{m}_2 \cdot \mathbf{n}' = \mathbf{RQm}_1 \cdot \mathbf{n}' = \begin{cases} \mathbf{m}_1 \cdot \mathbf{n}', \text{ type I} \\ -\mathbf{m}_2 \cdot \mathbf{n}', \text{ type II} \end{cases}. \quad (46)$$

Therefore, by choosing the $\pm$ sign properly in equation (45), we have simultaneously satisfied both expressions of equation (45). Briefly: if the deformation gradients are compatible and the magnetizations are crystallographic, then there are no poles on the twin boundaries in a simple laminated structure. This remarkable result is understandable when we pass to the crystallographic picture. Since (in the deformed configuration) the lattices on each side of the twin boundary are related by a 180° rotation about the normal $\mathbf{n}'$ (type I) or shear vector $\mathbf{a}$ (type II), then corresponding crystallographic vectors have the same projection on the normal, if the sign of the each of the vectors is chosen properly. The magnetizations in neighboring twin bands typically point in quite different directions.

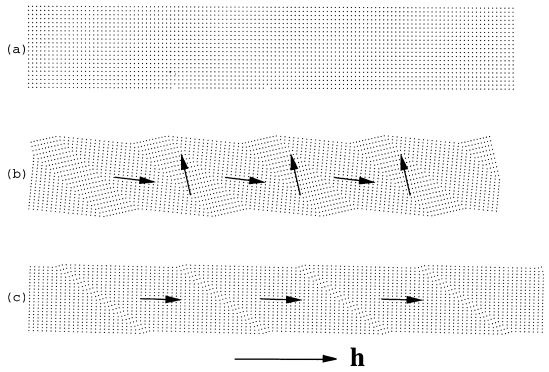

Fig. 10. Redistribution of two magnetically hard martensite variants produced by an applied field.

To induce a reversible change of shape in ferromagnetic martensites, one needs to compete two forces. Large-scale rearrangement of martensites have been produced by alternating two fields in different directions [80], which utilizes the effect of the second term of equation (42), or by competing a field and a stress [80]. [The stress affects the loading device energy which has been omitted from equation (42).] Researchers have also noticed the effect of demag. energy [the last term of equation (42)] in various tests, and a reversible change of shape can also be produced by competing a single field against demag. energy: in a specimen of suitable elongated or flattened shape, the demag. energy favors one variant and the applied field favors a different variant.

The most important magnetic measurement for ferromagnetic shape-memory materials, in which the change of shape is to be induced in the martensite, is the measurement of magnetic anisotropy. Magnetic anisotropy quantifies the difficulty of rotating the magnetization away from an easy axis. One can appreciate from Fig. 10 that if the nearly vertical magnetization rotates into the direction of the field, then there will be no driving force on the twin boundaries arising from the applied field energy [equation (39)].

Two alloy systems have been studied most extensively, though many questions remain even about these systems. These are alloys near the compo-

Table 9. Typical magnetic properties of Ni_2MnGa and Fe_3Pd

Property	Ni_2MnGa	Fe_3Pd
Energy wells	$\begin{pmatrix}\beta & 0 & 0\\ 0 & \alpha & 0\\ 0 & 0 & \alpha\end{pmatrix}$ $m_s(100)$	$\begin{pmatrix}\beta & 0 & 0\\ 0 & \alpha & 0\\ 0 & 0 & \alpha\end{pmatrix}$ $m_s(010)$ and $m_s(001)$
	$\begin{pmatrix}\alpha & 0 & 0\\ 0 & \beta & 0\\ 0 & 0 & \alpha\end{pmatrix}$ $m_s(010)$	$\begin{pmatrix}\alpha & 0 & 0\\ 0 & \beta & 0\\ 0 & 0 & \alpha\end{pmatrix}$ $m_s(100)$ and $m_s(001)$
	$\begin{pmatrix}\alpha & 0 & 0\\ 0 & \alpha & 0\\ 0 & 0 & \beta\end{pmatrix}$ $m_s(001)$	$\begin{pmatrix}\alpha & 0 & 0\\ 0 & \alpha & 0\\ 0 & 0 & \beta\end{pmatrix}$ $m_s(100)$ and $m_s(010)$
Stretches	$\alpha = 1.013$, $\beta = 0.952$	$\alpha = 1.01$, $\beta = 0.98$ (M_s–40°C)
Saturation magnetization	$m_s = 600$ emu/cm^3 ($Ni_{51.3}Mn_{24.0}Ga_{24.7}$ at M_s)	$m_s = 1390$ emu/cm^3 ($Fe_{70}Pd_{30}$ at M_s)
Anisotropy constants of austenite	$\kappa_1 = 4 \times 10^4$ ergs/cm^3, $\kappa_2 = -10 \times 10^4$ ergs/cm^3	$\kappa_1 = -0.5 \times 10^5$ ergs/cm^3 (approx. [91])
Anisotropy constants of martensite	$\kappa_u = 2.5 \times 10^6$ ergs/cm^3	?

sitions Ni_2MnGa and Fe_3Pd. The Heusler alloy Ni_2MnGa exhibits several stress-induced martensitic phases [81–83] and an intermediate premartensitic phase [84–86]. We just consider here the thermal martensite, which is tetragonal [87]; apparently, the other martensites have not been induced using only an applied field. Magnetic measurements in the austenite (Curie temperature, susceptibility, saturation magnetization, magnetization vs field and anisotropy constants) are straightforward and are available for several compositions (see e.g. Refs [86–90], and citations therein). The saturation magnetization of the martensite is also widely available (e.g. Refs [88–90]). Measurements of anisotropy constants and magnetization curves of the martensite in the literature are widely divergent. This is because typical measurements have been done on specimens with microstructure; under such conditions, the true easy axis can appear to be the hard axis [90]. Magnetization vs field on single crystal specimens of austenite and on a single variant of martensite (obtained by detwinning the specimen before the test) in an alloy near Ni_2MnGa are given in Ref. [90]; the resulting anisotropy constants are given in Table 9. Note that the martensite is two orders of magnitude magnetically harder than the austenite.

Magnetization curves relate to the energy density $\varphi(\mathbf{F},\mathbf{m},\theta)$, with the temperature θ held fixed at the appropriate value, in the following way. The "free" magnetic energy of the austenite and martensite, defined by,

$$\varphi_{\text{aus}}(\mathbf{m}) = \min_{\mathbf{U}\in\mathcal{N}_{\text{aus}}(\mathbf{I})} \varphi(\mathbf{U},\mathbf{m},\theta)$$
$$\varphi_{\text{mart}}(\mathbf{m}) = \min_{\mathbf{U}\in\mathcal{N}_{\text{mart}}(\mathbf{U}_1)} \varphi(\mathbf{U},\mathbf{m},\theta) \qquad (47)$$

where the minimization is taken over positive-definite symmetric matrices in suitable neighborhoods $\mathcal{N}_{\text{aus}}$ and $\mathcal{N}_{\text{mart}}$, are invariant under the corresponding point groups of austenite and martensite, e.g. $\varphi_{\text{aus}}(\mathbf{Qm}) = \varphi_{\text{aus}}(\mathbf{m})$ for $\mathbf{Q}$ in the point group of the austenite. Physically, these give the magnetic energy at the relaxed strain. There are simple, standard properly invariant forms of these functions, the most common for cubic austenite and tetragonal martensite (c-axis is easy) being,

$$\varphi_{\text{aus}}(\mathbf{m}) = \kappa_1\left(\hat{m}_1^2\hat{m}_2^2 + \hat{m}_1^2\hat{m}_3^2 + \hat{m}_2^2\hat{m}_3^2\right) + \kappa_2\ \hat{m}_1^2\hat{m}_2^2\hat{m}_3^2 \qquad (48)$$
$$\varphi_{\text{mart}}(\mathbf{m}) = \kappa_u \sin^2\theta$$

where $\hat{\mathbf{m}} = \mathbf{m}/|\mathbf{m}|$ and θ is the angle between $\hat{\mathbf{m}}$ and the c-axis of the martensite. The constants κ_1, κ_2, and κ_u are the anisotropy constants and can be obtained directly from magnetization curves. A summary of the properties of Fe_3–Pd and Ni_2–Mn–Ga is given in Table 9.

The other method of inducing a change of shape is by field-induced transformation. It is expected to be governed by a magnetic version of the Clausius–Clapeyron equation. But, in fact, the presence of the nonlocal demag. energy, in general, forbids the derivation of a local Clausius–Clapeyron equation along classical lines. To proceed, normalize equation (40) by dividing by m_s and recall the form of equation (48). It is then seen that the three terms in the total energy equation (42) (magnetoelastic: applied field: magnetostatic) scale like (κ or E: hm_s: m_s^2), where κ is a typical anisotropy constant and E is a typical elastic modulus. It is seen from Table 9 that all three of these terms are roughly equally important for Ni_2MnGa. However, *if we neglect the magnetostatic energy*, we can consider homogeneous states $[\nabla\mathbf{y}(\mathbf{x}) = \mathbf{F}_a(h), \mathbf{m} = \mathbf{m}_a(h)]$ and $[\nabla\mathbf{y}(\mathbf{x}) = \mathbf{F}_m(h), \mathbf{m} = \mathbf{m}_m(h)]$, corresponding to austenite and martensite, respectively, and parameterized by the field strength h, that equi-minimize the total energy [equation (42)] at a temperature $\theta(h)$. Here, the applied field has been assumed to be spatially homogeneous with fixed direction, $\mathbf{h} = h\mathbf{e}$. When

$h = 0$, these must belong to the appropriate energy wells for austenite and martensite. Now, in the standard way, one writes the condition that the two states have the same total free energy (equation (42) with demag. omitted), differentiates this condition with respect to h, and evaluates at $h = 0$:

$$\theta'(h)\big|_{h=0} = \frac{\mathbf{e}\cdot(\mathbf{m}_a(0) - \mathbf{m}_m(0)\det\mathbf{U}_1)}{\eta_m - \eta_a}. \qquad (49)$$

Here, the integrals have been removed using the homogeneity of the states, and the entropy densities of austenite and martensite are defined by $\eta_a = -\partial\varphi[\mathbf{F}_a(0),\mathbf{m}_a(0),\theta(0)]/\partial\theta$ and $\eta_m = -\partial\varphi[\mathbf{F}_m(0),\mathbf{m}_m(0),\theta(0)]/\partial\theta$. Using equation (49) and evaluating the entropy difference in terms of the latent heat obtained from DSC measurements (7.5×10^4 ergs/cm^3) and transformation temperature (375 K), and a typical value for the jump of the saturation magnetization of 10 emu/g [87], we estimate for Ni_2MnGa a temperature rise of 0.5 K/T [92]. Given the approximation of dropping the demag. energy, we conclude that the effect will be hardly observable in this alloy at moderate fields.

We finish with a remark on the observations on ferromagnetic martensitic steels summarized by Shimizu and Kakeshita [72]. They are in a rather different regime from the one discussed here. Their field is so large that the second term in equation (42) completely dominates the other terms, by the scaling argument given above. At their fields, the magnetization of austenite and martensite have long since been aligned in the direction of the applied field, and the associated anisotropy energy is relatively negligible. At those large fields, the important energy concerns the breakdown of the constraint $|\mathbf{m}| = m_s(\theta)$. The associated energy is modeled with a high field susceptibility. They show good agreement between shifts of M_s with temperature and predictions of a Clausius–Clapeyron equation based on the applied field energy and the high field susceptibility energy.

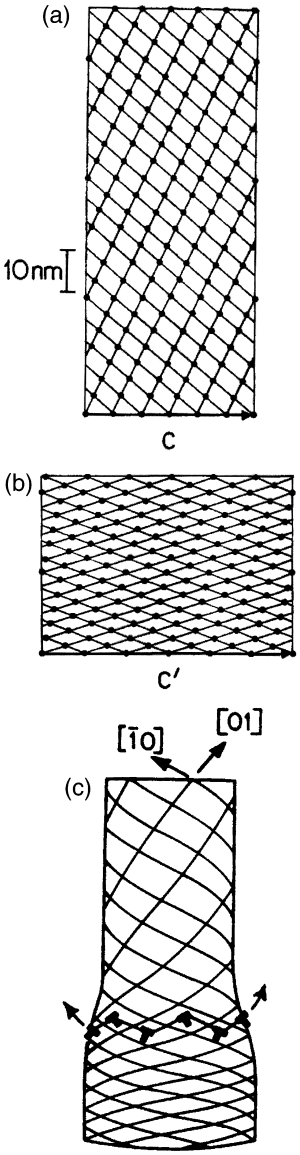

Fig. 11. Plane lattice crystal structure of the tail-sheath of T-4 Bacteriophage extended (a) and contracted (b). (c) dislocation model of coherent interface between "austenite" (extended) and "martensite" (contracted) [102]. Courtesy of G. B. Olson. Originally published in *J. Physique*, 1982, **43**, 855, with permission of EDP Sciences.

7. MARTENSITE AT SMALL SCALES

Phenomena in martensitic materials such as the shape-memory effect are most definitely structure-sensitive. Hence, these phenomena are expected to exhibit unexpected trends when the specimen is scaled smaller and smaller.

For example, consider the habit plane dividing a block of size $L \times L \times L$. A classical estimate gives the twin spacing as proportional to $L^{1/2}$. More recent studies (Kohn and Müller [93] and Schryvers [94]) show that there are parameter regimes in which branching of the twins is predicted to occur, and the twin spacing varies with distance ℓ from the habit plane according to the scaling law $\ell^{2/3}$. Both of these calculations assume that the twin spacing is much smaller than body size. But in either case, $L^{1/2}$ or $L^{2/3}$ exceeds L when L is sufficiently small and there is expected to be unusual small-scale behavior.

As far back as the 1950s, workers (see, e.g. Cech and Turnbull [95]) noticed a scale dependence of the martensitic transformation. By extrapolation, this gave rise to the notion that the martensitic transformation would be suppressed at small scales. The effect was expected to be analogous to the formation of ferromagnetic domains, sufficiently small particles being single domain with coercivity two orders of magnitude larger than bulk material (in, e.g. Fe). The recent evidence both supports and

conflicts with this view. In a study by Kajiwara *et al*. [96] on 20–200 nm particles of Fe–Ni, most particles transform at temperatures near bulk values, but particles that fail to transform do so down to cryogenic temperature. Alternatively, 10–100 nm particles of Fe–Ni prepared by a different method [97], and containing an intermediate volume fraction of martensite, exhibited a (size dependent) martensite to austenite transformation upon heating, but no subsequent austenite to martensite transformation upon cooling. In polycrystalline films, which include the added effect of constraint, the observations exhibit general consistency: ~80 nm thick films of $PbTiO_3$ with 70–100 nm grains are single domain at room temperature (unlike in bulk) [98], while NiAl films of thickness 70–100 nm and grain size 10–20 nm show an absence of twinning well below M_f [99]. At larger sizes, micron scale sputtered films of NiTi exhibit perfect shape memory and properties close to those in the bulk [4, 100], and MBE grown single crystals of Ni_2MnGa of thickness 500 nm also show substructures of fine twins [101]. Even more spectacular is the recognition that what appears to be a martensitic transformation occurs in the nanoscale organism Bacteriophage T-4 [102]. Collectively, these observations show that the current predictive understanding of martensitic transformations at small scale is rudimentary.

More specifically, in the latter example, Olson and Hartman [102] note that a displacive phase transformation of the martensitic type appears to take place in several different cylindrical protein crystals. One occurs during tail-sheath contraction as the organism Bacteriophage T-4 performs life functions. The extended and the contracted forms are shown in Figs 11(a) and (b), respectively. The lengths c and c' in this figure give the circumference of the cylinders, where the outer diameter of the cylindrical sheath is 24 nm extended and 32 nm contracted. A coherent interface between the extended "austenite" phase and the contracted "martensite" phases can be rationalized as shown in Fig. 11(c). The organism also appears to have a built-in mechanism for nucleation.

There has been significant recent developments on sputtered films [4, 100, 103, 104]. The reader is referred to the review of Miyazaki and Ishida [4] for details. The important effect of the constraint of the substrate has been studied in Refs [105, 106].

We discuss what can be expected from a theoretical viewpoint. Consider a single-crystal film of thickness h and cross-section S in its reference state, and an orthonormal basis $(\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3)$ with $\mathbf{e}_3$ normal to the film. Here, x_i is a coordinate in the direction $\mathbf{e}_i$. Bhattacharya and James [107] begin with the free energy equation (8), but including also a term for interfacial energy. They then examine the asymptotic limit of the total free energy as $h \to 0$, extracting along the way a limiting membrane theory (see Ref. [107] for the details of this derivation). To describe this theory, consider the free energy density $\varphi(\mathbf{F},\theta)$ given in equation (8). $\mathbf{F}$ can be written $(\mathbf{a}_1|\mathbf{a}_2|\mathbf{a}_3)$ where the vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ are the three columns of $\mathbf{F}$; when $\mathbf{F} = \nabla\mathbf{y}$ this notation becomes $(\mathbf{y}_{,1}|\mathbf{y}_{,2}|\mathbf{y}_{,3})$, where the comma denotes partial derivative, $\mathbf{y}_{,i} = \partial\mathbf{y}/\partial x_i$. The membrane theory is based on the free energy

$$h\int_S \varphi\big(\mathbf{y}_{,1}|\mathbf{y}_{,2}|\mathbf{b},\theta\big)\mathrm{d}x_1\mathrm{d}x_2 \tag{50}$$

where $\mathbf{y}(x_1,x_2)$ is now a mapping from the "middle surface" S into $\mathbb{R}^3$ and $\mathbf{b}(x_1,x_2)$ describes deformations of the film relative to the middle surface. The function $\mathbf{b}(x_1,x_2)$ is interpreted in the following way: in the reference configuration consider two points A and B, one on the bottom of the film and the other on the top of the film, B directly above A [i.e. same (x_1,x_2)]. Now deform the film in an arbitrary way, so that A and B go to A′ and B′. Then

$$h\mathbf{b}(x_1,x_2) = \overrightarrow{\mathrm{A'B'}}.$$

Thus, $\mathbf{b}$ measures in-plane shear and normal compression of the film.

The energy equation (50) is associated with the "membrane" energy because it is the part of the energy that scales as the thickness of the film. Bending energy can be derived with a higher order asymptotic argument [107]. It arises at order h^3. For example, the bending energy per unit length for a flat cantilever is $EI\kappa^2/2$ where E is a certain elastic modulus (not Young's modulus), κ is the curvature of the film [equal to 1/(radius of curvature)], and the h dependence appears in the moment of inertia I of the cross-section, $I = bh^3/12$. It is the h^3 sensitivity of the bending energy that explains the major recent developments in the use of vibrating cantilevers as sensors. For shape-memory actuators, the opposite is often desired: the change of shape of the film should store as much energy as possible, and therefore the design of the actuator should be based on the membrane energy [equation (50)].

The minimizers of equation (50) have been studied using the energy-well structure for martensitic materials [40, 107, 108]. The main observation is that the presence of $\mathbf{b}$ instead of $\mathbf{y}_{,3}$ means that one of the conditions of compatibility is lost in passing from bulk to thin film. The physical interpretation of this condition is immediate. If strains in the thickness direction differ on each side of an interface cutting through the film, then these can be interpolated by an $h \times h$ layer whose energy contribution (h^2) is negligible relative to membrane energies.

The latter suggests that in single-crystal films there will be the possibility of exact austenite/martensite interfaces, with no fine twinning, according

Table 10. Exact interfaces between austenite and variant 9 of martensite for films of various orientation in $Ni_{50}Ti_{50}$. Interface lines are lines on the film at which austenite and martensite meet. The in-plane principal stretches refer to the martensite [107]

Film normal	Austenite/Martensite interface?	Interface lines			In-plane principal stretches
100	Yes	[0, −0.9639, 0.2664]	or	[0, 0.3841, 0.9233]	(0.9358, 1.0473)
010	Yes	[−0.9639, 0, 0.2664]	or	[0.3841, 0, 0.9233]	(0.9358, 1.0473)
001	Yes	[−0.9728, 0.2317, 0]	or	[0.2317, −0.9728, 0]	(1.0840, 0.9663)
110	No				(0.9582, 0.9663)
$1\bar{1}0$	Yes	[0.1892, 0.1892, 0.9636]	or	[0.6080, 0.6080, −0.5105]	(1.1066, 0.9320)
101	Yes	[−0.3339, 0.8815, 0.3339]	or	[0.5018, 0.7046, −0.5018]	(0.9464, 1.0286)
$10\bar{1}$	Yes	[0.1351, −0.9816, 0.1351]	or	[0.6840, −0.2538, 0.6840]	(1.1005, 0.9574)
011	Yes	[0.8815, −0.3339, 0.3339]	or	[0.7046, 0.5018, −0.5018]	(0.9464, 1.0286)
$01\bar{1}$	Yes	[−0.9816, 0.1351, 0.1351]	or	[−0.2538, 0.6840, 0.6840]	(1.1005, 0.9574)
111	No				(0.9422, 0.9663)
$\bar{1}11$	Yes	[0.3505, 0.8139, −0.4634]	or	[0.5952, −0.1864, 0.7816]	(1.1001, 0.9424)
$1\bar{1}1$	Yes	[0.8139, 0.3505, −0.4634]	or	[−0.1864, 0.5952, 0.7816]	(1.1001, 0.9424)
$11\bar{1}$	Yes	[−0.3238, 0.8110, 0.4872]	or	[0.8110, −0.3238, 0.4872]	(1.0582, 0.9663)

to the minimization of equation (50). This turns out to be true, and the appropriate condition of compatibility is an invariant line condition. The most useful form of this condition is written in terms of the film normal $\mathbf{e}_3$ and the transformation stretch matrix $\mathbf{U} \in \{\mathbf{U}_1, \ldots, \mathbf{U}_\nu\}$. The necessary and sufficient condition for the existence of exact austenite/martensite interfaces is $\mathbf{e}_3 \cdot [\text{cof}\,(\mathbf{U}^2 - \mathbf{I})]\mathbf{e}_3 \leq 0$, where cof is defined above in equation (29). This is often satisfied; Table 10 summarizes the interfaces possible in low index films of NiTi. These predictions would be extremely interesting to explore experimentally.

An advantage of films is that the reduction of macroscopic strain that is a consequence of the fine twinning in the bulk austenite/martensite interface is avoided in films. Notice the large in-plane stretches listed in Table 10. It is therefore interesting to think about other structures that can be made on a film, which utilize the membrane mode. One such structure, the "tunnel" is obtained by releasing the film along a strip whose edges are exact austenite/martensite interfaces. The possibility of a tunnel is more restrictive than a simple austenite/martensite interface, because the latter may produce an in-plane shear. Necessary and sufficient conditions for a tunnel are: $\mathbf{e}_3 \cdot [\text{cof}\,(\mathbf{U}^2 - \mathbf{I})]\mathbf{e}_3 = 0$ and $\text{tr}\mathbf{U}^2 - \mathbf{e}_3 \cdot \mathbf{U}^2\mathbf{e}_3 - 2 \geq 0$. Examples of materials satisfying these conditions are given in Refs [40, 107].

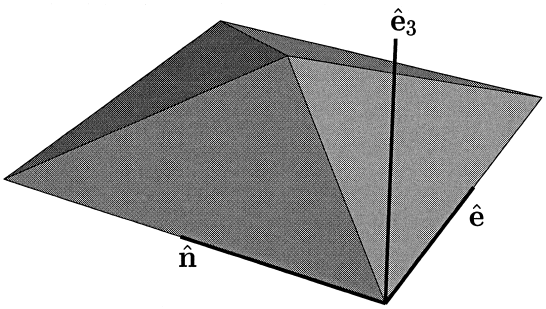

Fig. 12. Tent in a {100} single-crystal film of a Cu–Zn–Al alloy.

More interesting is the possibility of trying to release a film on a compact polygonal region bounded by austenite/martensite interfaces. In general, the associated conditions are very highly restrictive, but solutions become possible at special lattice parameters when the normal of the film is an n-fold axis of symmetry of the austenite ($n \geq 2$). Then an n-sided "tent" is possible if also the conditions given above for the tunnel are satisfied. These conditions hold for the particular $Cu_{68}Zn_{15}Al_{17}$ alloy whose lattice parameters are listed above in Table 4. Figure 12 shows a picture of the deformed shape of this tent, drawn using these measured lattice parameters, so one can get a good idea of the large strains that are predicted.

Acknowledgements—This work was supported by ONR N00014-95-1-1145 and -91-J-4034, ARO DA/DAAG55-98-1-0335, NSF DMS-9505077, and AFOSR/MURI F49620-98-1-0433. We thank R. V. Kohn and D. Schryvers for their comments.

REFERENCES

1. Bhattacharya, K., in *Shape Memory Alloys: From Microstructure to Macroscopic Properties*, ed. G. Airoldi, I. Müller and S. Miyazaki. Trans Tech Publications, 1997.
2. Nishiyama, Z., *Martensitic Transformations*, ed. M. E. Fine, M. Meshii, C. M. Wayman. Academic Press, New York, 1978.
3. Otsuka, K. and Wayman, C. M. (ed.), *Shape Memory Materials*. Cambridge University Press, Cambridge, 1998.
4. Miyazaki, S. and Ishida, A., *Mater. Sci. Engng A*, 1999, in press.
5. Miyazaki, S. and Otsuka, K., *Iron Steel Inst. Japan Int.*, 1989, **29**, 353.
6. Bhattacharya, K. and Kohn, R. V., *Acta metall. mater.*, 1995, **44**, 529.
7. Bhattacharya, K. and Kohn, R. V., *Archs Ration. Mech. Analysis*, 1997, **139**, 99.
8. Shu, Y. C. and Bhattacharya, K., *Acta metall. mater.*, 1998, **46**, 5457.
9. Roytburd, A. L., *Solid St. Phys.*, 1978, **33**, 317.
10. Khachaturyan, A., *Theory of Structural Transformations in Solids*. Wiley, New York, 1983.
11. Ball, J. M., Chu, C. and James, R. D., *J. Physique III, Colloque*, 1995, **5**(C8), 245.

12. Abeyaratne, R. and Knowles, J. K., *J. Mech. Phys. Solids*, 1993, **41**, 541.
13. Brinson, L. C., *J. Intell. Mater. Syst. Struct.*, 1993, **4**, 229.
14. Brinson, L. C. and Huang, H. S., *J. Intell. Mater. Syst. Struct.*, 1995, **6**, 108.
15. Patoor, E., Eberhardt, A. and Berveiller, M., *Pitman Res. Notes Math. Ser.*, 1993, **296**, 38.
16. Leclercq, S. and Lexcellent, C., *J. Mech. Phys. Solids*, 1996, **44**, 953.
17. Boyd, J. G. and Lagoudas, D. C., *Int. J. Plasticity*, 1996, **12**, 805.
18. Wang, Y, Chen, L.-Q. and Khachaturyan, A., in *Proceedings of International Conference on Solid → Solid Phase Transformations in Inorganic Materials '94*, ed. W. Johnson, J. W. Howe, D. E. Laughlin and W. A. Soffia. The Minerals, Metals, Materials Society, 1994, p. 245.
19. Luskin, M., *Acta Numerica*, 1996, **5**, 191.
20. Rhee, J. Y., Harmon, B. N. and Lynch, D. W., *Phys. Rev. B (Condensed Matter)*, 1996, **54**, 17385.
21. Ball, J. M. and James, R. D., *Archs Ration. Mech. Analysis*, 1987, **100**, 13.
22. Ball, J. M. and James, R. D., *Phil. Trans. R. Soc.*, 1992, **A338**, 389.
23. Bhattacharya, K., *Acta metall. mater.*, 1991, **39**, 2431.
24. Bhattacharya, K., *Archs Ration. Mech. Analysis*, 1992, **120**, 201.
25. Ericksen, J. L., *J. appl. Mech.*, 1978, **45**, 740.
26. Ericksen, J. L., *Archs Ration. Mech. Analysis*, 1979, **72**, 1.
27. Pitteri, M., *J. Elasticity*, 1984, **14**, 175.
28. Ericksen, J. L., in *Phase Transformations and Material Instabilities in Solids*, ed. M. E. Gurtin. Academic Press, New York, 1984, p. 61.
29. Zanzotto, G., *Archs Ration. Mech. Analysis*, 1992, **121**, 1.
30. Zanzotto, G., *Acta crystallogr.*, 1996, **A52**, 839.
31. Hane, K. F. and Shield, T. W., *Acta mater.*, 1999, **47**, 2603.
32. Hane, K. F. and Shield, T. W., *Proc. R. Soc. A*, 1999, **455**, 3901.
33. Bhattacharya, K., *Continuum Mech. Thermodynam.*, 1993, **5**, 205.
34. Hane, K. F., Shield, T. W., *Mater. Sci. Engng A*, submitted.
35. James, R. D. and Kinderlehrer, D., *Phil. Mag. B*, 1993, **68**, 237.
36. Hane, K. F. and Shield, T. W., *Phil. Mag. A*, 1998, **78**, 1215.
37. Zanzotto, G., Weak and symmetry-breaking transitions in simple lattices. Preprint, 1996.
38. Hane, K. F., Shield, T. W., *J. Elasticity*, submitted.
39. Pitteri, M. and Zanzotto, G., *Acta metall. mater.*, 1997, **46**, 225.
40. Hane, K. F., *J. Mech. Phys. Solids*, 1999, **47**, 1917.
41. Simha, N. K., *J. Mech. Phys. Solids*, 1997, **45**, 261.
42. Jian, L. and James, R. D., *Acta metall. mater.*, 1997, **45**, 4271.
43. Gurtin, M. E., *Archs Ration. Mech. Analysis*, 1984, **84**, 1.
44. Chu, C. and James, R. D., *J. Physique III, Colloque*, 1995, **5**(C8), 143.
45. Bywater, K. A. and Christian, J. W., *Phil. Mag.*, 1972, **25**, 1249.
46. Moberly, W. J., Proft, J. L., Duerig, T. W. and Sinclair, R., *Mater. Sci. Forum*, 1990, **56–58**, 605.
47. Wayman, C. M., *Introduction to the Crystallography of Martensitic Transformations*. Macmillan, New York, 1964.
48. Otsuka, K., Ohba, T., Tokonami, M. and Wayman, C. M., *Scripta metall. mater.*, 1993, **29**, 1359.
49. Funakubo, H. (ed.), *Shape Memory Alloys*. Gordon and Breach, New York, 1987.
50. Ericksen, J. L., *Archs Ration. Mech. Analysis*, 1997, **139**, 181.
51. Kajiwara, S., *Trans. Japan Inst. Metals*, 1976, **17**, 435.
52. Kajiwara, S., *Trans. Japan Inst. Metals*, 1976, **17**, 447.
53. Escobar, J. C. and Clifton, R. J., *Mater. Sci. Engng. A*, 1993, **170**, 125.
54. Saburi, T. and Nenno, S., *Scripta metall.*, 1975, **9**, 887.
55. Otsuka, K., Wayman, C. M., Nakai, K., Sakamoto, H. and Shimizu, K., *Acta metall.*, 1976, **24**, 207.
56. Saburi, T., Nenno, S., Kato, S. and Takata, K., *J. less-common Metals*, 1976, **50**, 223.
57. Nishiyama, Z. and Kajiwara, S., *Japan. J. appl. Phys.*, 1963, **2**, 478.
58. Andrade, M., Chandrasekaran, M. and Delaey, L., *Acta metall.*, 1984, **32**, 1809.
59. Tadaki, T., Tokoro, M. and Shimizu, K., *Trans. Japan Inst. Metals*, 1975, **16**, 285.
60. Chakravorty, S. and Wayman, C. M., *Acta metall.*, 1977, **25**, 989.
61. Tokonami, M., Otsuka, K., Shimizu, K., Iwata, Y. and Shibuya, I., in *Proc. Int. Conf. of Martensitic Transformation ICOMAT 1979*, Cambridge, MA, 1979, p. 639.
62. Sun, Q.-P., Xu, T. T. and Zhang, X., *J. Engng Mater. Technol.*, 1999, **121**, 38.
63. Shield, T. W., *J. Mech. Phys. Solids*, 1995, **43**, 869.
64. Rodriguez, P. L., Lovey, F. C., Guenin, G., Pelegrina, J. L., Sade, M. and Morin, M., *Acta metall. mater.*, 1993, **41**, 3307.
65. Schroeder, T. A. and Wayman, C. M., *Acta metall.*, 1977, **25**, 1375.
66. Saburi, T. and Wayman, C. M., *Acta metall.*, 1979, **27**, 979.
67. Miyazaki, S., Otsuka, K. and Wayman, C. M., *Acta metall.*, 1989, **37**, 1873.
68. Miyazaki, S., Otsuka, K. and Wayman, C. M., *Acta metall.*, 1989, **37**, 1885.
69. Ericksen, J. L., *Archs Ration. Mech. Analysis*, 1962, **10**, 189.
70. Ball, J. M. and James, R. D., *Proc. R. Soc. A*, 1991, **432**, 93.
71. Müller, S., *Variational Problems for Microstructure and Phase Transitions*, in *Lecture Notes No. 2*. Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig, 1998.
72. Shimizu, K. and Kakashita, T., *Iron Steel Inst. Japan Int.*, 1989, **29**, 97.
73. Patel, J. R. and Cohen, M., *Acta metall.*, 1953, **1**, 531.
74. James, R. D. and Kinderlehrer, D., *J. appl. Phys.*, 1994, **76**, 7012.
75. James, R. D. and Wuttig, M., in *Symposium on "Mathematics and Control in Smart Structures"*, ed. V. V. Varadan and J. Chandra, *Proceedings of SPIE*, 2715. SPIE, 1996, p. 420.
76. James, R. D. and Wuttig, M., *Phil. Mag. A*, 1998, **77**, 1273.
77. Brown, W. F., in *Magnetoelastic Interactions*, ed. C. Truesdell, *Springer Tracts in Natural Philosophy*, Vol. 9. Springer, Berlin, 1966.
78. De Simone, A., *Archs Ration. Mech. Analysis*, 1993, **125**, 99.
79. James, R. D., Tickle, R. and Wuttig, M., *Mater. Sci. Engng*, in press.
80. Tickle, R., James, R.D., Shield, T., Schumacher, P., Wuttig, M. and Kokorin, V. V., *IEEE J. Magnet.* 1999, **35**, 4301.

81. Kokorin, V. V. and Martynov, V. V., *Physics Metals Metallogr*., 1991, **72**, 101.
82. Martynov, V. V. and Kokorin, V. V., *J. Physique III*, 1992, **2**, 739.
83. Martynov, V. V. and Kokorin, V. V., *Trans. Mater. Res. Soc. Japan*, 1994, **18B**, 839.
84. Zheludev, A., Shapiro, S. M., Wochner, P. and Tanner, L. E., *Phys. Rev. B*, 1996, **54**, 15045.
85. Kokorin, V. V., Chernenko, V. A., Cesari, E., Pons, J. and Seguí, C., *J. Phys. Condensed Matter*, 1996, **8**, 6457.
86. Planes, A., Obradó, E., Gonzàlez-Comas, A. and Lluís Mañosa, A., *Phys. Rev. Lett*., 1997, **79**, 3926.
87. Webster, P. J., Ziebeck, K. R. A., Town, S. L. and Peak, M. S., *Phil. Mag. B*, 1984, **49**, 295.
88. David, S., Thesis, Centre National de la Recherche Scientifique, Laboratoire de Magnetisme Louis Néel, 1991.
89. Ullakko, K., Huang, J. K., Kantner, C., O'Handley, R. C. and Kokorin, V. V., *Appl. Phys. Lett*., 1996, **69**, 1966.
90. Tickle, R. and James, R. D., *J. Magnet. Magnetic Mater*., 1999, **195**, 627.
91. Matsui, M., Kuang, J. P., Totani, T. and Adachi, K., *J. Magnet. Magnetic Mater*., 1986, **54–57**, 911.
92. Tickle, R., Thesis, University of Minnesota, 1999.
93. Kohn, R. V. and Müller, S., *Phil. Mag. A*, 1992, **66**, 697.
94. Schryvers, D., *Phil. Mag. A*, 1993, **68**, 1017.
95. Cech, R. E. and Turnbull, D., *J. Metals*, 1956, **8**, 124.
96. Kajiwara, S., Ohno, S. and Honma, K., *Phil. Mag. A*, 1991, **63**, 625.
97. Yunhong, C., Guichang, D., Huaixian, L., Jinhui, W. and Gang, L., *Japan. J. appl. Phys*., 1995, **34**, 113.
98. Ren, S. B., Lu, C. J., Shen, H. M. and Wang, Y. N., *Phys. Rev. B*, 1997, **55**, 3485.
99. Yandouzi, M., Toth, L. and Schryvers, D., *Nanostruct. Mater*., 1998, **10**, 99.
100. Grummon, D., Hou, L. and Pence, T. J., *J. Physique III, Colloque*, 1995, **5**(C8), 665.
101. Dong, J.W., Chen, L.C., James, R.D., McKernan, S. and Palmstrøm, C.J., *Appl. Phys. Lett*., 1999, **75**, 1443.
102. Olson, G. B. and Hartman, H., *J. Physique*, 1982, **43**(C4), 855.
103. Chang, L., Hu-Simpson, C., Grummon, D. S., Pratt, W. and Loloee, R., *Mater. Res. Soc. Symp*., 1990, **187**, 137.
104. Hou, L. and Grummon, D., *Scripta metall. mater*., 1995, **33**, 989.
105. Hou, L., Pence, T. J. and Grummon, D. S., *Mater. Res. Soc. Symp*., 1995, **360**, 369.
106. Roytburd, A. L., Kim, T. S., Su, Q., Slutsker, J. and Wuttig, M., *Acta mater*., 1998, **46**, 5095.
107. Bhattacharya, K. and James, R. D., *J. Mech. Phys. Solids*, 1999, **47**, 531.
108. Bhattacharya, K., De Simone, A., Hane, K.F., James, R.D., Palmstrøm, C.J., *Mater. Sci. Engng*, in press.