to appear in *Nonlinear Elasticity: Theory and Application, edited by Y. Fu and R.W. Ogden, Cambridge University Press, 2000.*

Strain-Energy Functions with Multiple Local Minima: Modeling Phase Transformations using Finite Thermoelasticity

by

R. Abeyaratne[1], K. Bhattacharya[2] and J. K. Knowles[2]

August 28, 2000

Abstract

This chapter provides a brief introduction to the following basic ideas pertaining to thermoelastic phase transitions: the lattice theory of martensite, phase boundaries, energy minimization, Weierstrass-Erdmann corner conditions, phase equilibrium, nonequilibrium processes, hysteresis, the notion of driving force, dynamic phase transitions, nonuniqueness, kinetic law, nucleation condition, and microstructure.

1. Introduction. This chapter provides an introduction to some basic ideas associated with the modeling of solid-solid phase transitions within the continuum theory of finite thermoelasticity. No attempt is made to be complete, either in terms of our selection of topics or in the depth of coverage. Our goal is simply to give the reader a flavor for some selected ideas.

This subject requires an intimate mix of continuum and lattice theories, and in order to describe it satisfactorily one has to draw on tools from crystallography, lattice dynamics, thermodynamics, continuum mechanics and functional analysis. This provides for a remarkably rich subject which in turn has prompted analyses from various distinct points of view. The free-energy function has multiple local minima, each minimum being identified with a distinct phase, and each phase being characterized by its own lattice. Crystallography plays a key role in characterizing the lattice structure and material symmetry, and restricts deformations through geometric compatibility. The thermodynamics of irreversible processes provides the framework for describing evolutionary processes. Lattice dynamics describes

[1]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, USA
[2]Division of Engineering and Applied Science, California Institute of Technology, Pasadena, USA

the mechanism by which the material transforms from one phase to the other. And eventually all of this needs to be described at the continuum scale. Since the present volume is concerned with finite elasticity, we shall strive to describe solid-solid phase transitions from this point of view and using this terminology.

The article by Schetky (1979) provides a general introduction to the subject while the book by Duerig et al. (1990) describes many engineering applications. For a more extensive treatment of the crystallographic and microstructural aspects, the reader is referred to Bhattacharya (2000) and James and Hane (2000); for details on dynamics, see the forthcoming monograph by Abeyaratne and Knowles (2001). An introduction from the materials science point of view can be found in Otsuka and Wayman (1999), while Christian (1975) provides a broader treatment of phase transformations. A general discussion of configurational forces can be found in Gurtin (2000); a particular example of this, the driving force on an interface, will play a major role in our discussion.

This chapter is organized as follows: Section 2 is devoted to explaining why we are interested in strain-energy functions with multiple local minima; the motivation we give is based on the lattice theory of martensitic transformations. Other reasons for studying such energy functions include, for example, the van der Waals type theory of two-phase fluids. In finite elasticity, the strain-energy function characterizes the material, e.g., the neo-Hookean and Mooney-Rivlin models of rubber. Accordingly in Section 3 we construct an explicit strain-energy function which describes a class of crystalline solids with cubic and tetragonal phases.

In preparation for solving initial-boundary-value problems, in Section 4 we specialize the three-dimensional theory of finite elasticity to the special case of uniaxial motions of a slab. In this way we derive a convenient one-dimensional mathematical model. Sections 5 and 6 are devoted to studying, respectively, the statics and dynamics of this slab. The static problem illustrates a number of phenomena, in particular the presence of surfaces in the body across which the deformation gradient is discontinuous. The problem also illustrates the distinction between mechanical equilibrium in the sense of balance of forces and moments, and phase equilibrium in the sense of energy extremization. The dynamic problem demonstrates the severe lack of uniqueness of solution to classically formulated initial-boundary-value problems. Attention is drawn to two distinct types of nonuniqueness. The notions of a nucleation condition and a kinetic law are then imported from materials science to show how they select the physically relevant solution from among the totality of available solutions. The quasi-static loading of a slab is also discussed in Section 6 and this too illustrates the role of nucleation and kinetics. Section 7 is devoted to trying to understand the notion of kinetics from a more general point of view. Three-dimensional

thermodynamic processes are considered and the various thermodynamic driving forces and conjugate fluxes contributing to entropy production are identified. In the present setting, the thermodynamic notion of kinetics of irreversible processes is identified with the kinetics of phase transitions.

Finally in Section 8 we return to static problems, this time in three-dimensions. We examine various piecewise homogeneous deformations and illustrate the role played by geometric compatibility in characterizing various microstructures.

2. Strain-energy functions with multiple local minima: motivation from the lattice theory of martensitic transformations. (Born and Huang 1954, Ericksen 1978, 1980, 1984, Ball and James 1992, James 1992, Bhattacharya 2000) Typically, the strain-energy function $W(\boldsymbol{F})$ in finite elasticity has an energy-well at the reference configuration, i.e., W has a local minimum at $\boldsymbol{F} = \boldsymbol{I}$. In this section we shall use a specific example to illustrate how, for certain materials, W has *multiple* local minima. The example pertains to a crystalline solid that can exist in a number of distinct crystalline forms, each crystal structure (lattice) being identified with a different "phase" of the material. Examples of such materials include the alloys Indium-Thallium (cubic and tetragonal phases), Gold-Cadmium (cubic and orthorhombic phases), Nickel-Titanium (cubic and monoclinic phases) and Copper-Aluminum-Nickel (cubic, orthorhombic and monoclinic phases).

Lattice model: A lattice refers to a periodic arrangement of points in space. The simplest lattice, called a *Bravais lattice*, is an infinite set of points in $\mathbb{R}^3$ that can be generated by the translation of a single point $\boldsymbol{o}$ through three linearly independent vectors $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$. Thus if $\mathcal{L}(\boldsymbol{e}, \boldsymbol{o})$ denotes a Bravais lattice, then

$$\mathcal{L}(\boldsymbol{e}, \boldsymbol{o}) = \Big\{ \boldsymbol{X} \,:\, \boldsymbol{X} = \nu^i \boldsymbol{e}_i + \boldsymbol{o} \text{ where } \nu^1,\ \nu^2,\ \nu^3 \text{ are integers} \Big\} \tag{2.1}$$

where summation on the repeated index i is implied. The vectors $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ are called the lattice vectors.

It is important to keep in mind that, because of the inherent symmetry of the lattice, more than one set of lattice vectors may generate the same Bravais lattice. For example, the two-dimensional lattice shown in Figure 2.1 is generated by both $\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ and $\{\boldsymbol{f}_1, \boldsymbol{f}_2\}$. In general, two sets of lattice vectors $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ and $\{\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3\}$ generate the same lattice if and only if

$$\boldsymbol{f}_i = \mu_i^j \boldsymbol{e}_j, \qquad [\mu] \in \mathit{I\!I}; \tag{2.2}$$

here $\mathit{I\!I}$ denotes the set of 3×3 matrices whose elements are integers and whose determinant is ± 1:

$$\mathit{I\!I} = \{\ [\mu]\ :\ [\mu]\ = 3 \times 3 \text{ matrix},\ \mu_i^j = \text{integer},\ \det [\mu] = \pm 1\}. \tag{2.3}$$

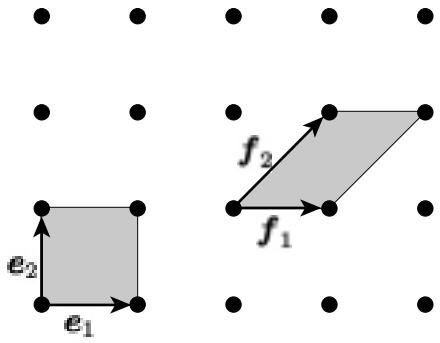

Figure 2.1: A two-dimensional lattice with lattice vectors.

Next consider the *energy* stored in a Bravais lattice. Under the assumption that there are identical atoms located at each lattice point, it is natural to assume that ψ, the free-energy of the lattice, depends on the lattice vectors $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ and the temperature $\theta > 0$:

$$\psi = \psi(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3; \theta) \ ; \tag{2.4}$$

the function ψ is defined for all triplets of linearly independent vectors and temperatures. Note that due to thermal expansion of the lattice, the lattice vectors themselves depend on the temperature: $\{\boldsymbol{e}_1(\theta), \boldsymbol{e}_2(\theta), \boldsymbol{e}_3(\theta)\}$. We will assume that the free-energy function ψ satisfies the following two requirements pertaining to frame indifference and symmetry:

1. Frame-indifference: A rigid rotation of the lattice should not change its free-energy and so
$$\psi(\boldsymbol{Q}\boldsymbol{e}_1, \boldsymbol{Q}\boldsymbol{e}_2, \boldsymbol{Q}\boldsymbol{e}_3; \theta) = \psi(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3; \theta) \qquad \text{for all rotations } \boldsymbol{Q} \ . \tag{2.5}$$

2. Lattice symmetry: The free-energy should depend on the lattice but not on the specific choice of lattice vectors, i.e., two sets of lattice vectors that generate the same lattice must have the same free-energy, and so
$$\psi(\mu_1^j \boldsymbol{e}_j, \mu_2^j \boldsymbol{e}_j, \mu_3^j \boldsymbol{e}_j; \theta) = \psi(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3; \theta) \qquad \text{for all } [\mu] \in I\!I. \tag{2.6}$$

Two-phase martensitic materials: Now consider a *two-phase material* where the lattice has one structure at high temperatures and a different one at low temperatures. As the temperature is changed, the lattice vectors vary continuously due to thermal expansion except at the *transformation temperature* θ_T where they change *discontinuously* (a "first order" martensitic phase transformation). The high temperature lattice is characterized by the vectors $\{\boldsymbol{e}_1^a(\theta), \boldsymbol{e}_2^a(\theta), \boldsymbol{e}_3^a(\theta)\}$ and is associated with one phase of the material (*austenite*); the low temperature lattice is described by $\{\boldsymbol{e}_1^m(\theta), \boldsymbol{e}_2^m(\theta), \boldsymbol{e}_3^m(\theta)\}$ and corresponds to a second phase

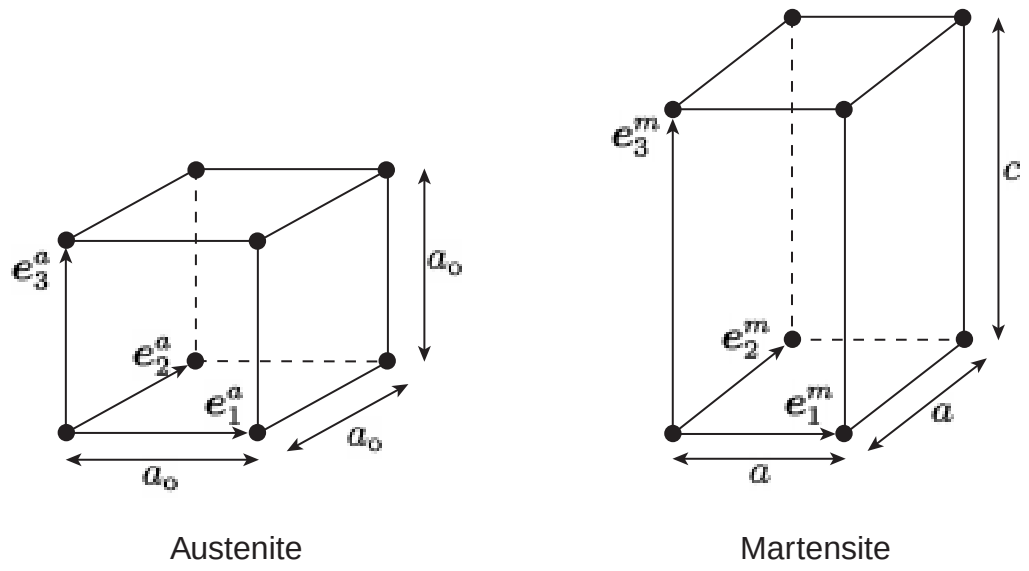

Figure 2.2: Simple cubic lattice (austenite) with lattice parameters $a_o \times a_o \times a_o$ and lattice vectors $\{\boldsymbol{e}_1^a, \boldsymbol{e}_2^a, \boldsymbol{e}_3^a\}$; and simple tetragonal lattice (martensite) with lattice parameters $a \times a \times c$ and lattice vectors $\{\boldsymbol{e}_1^m, \boldsymbol{e}_2^m, \boldsymbol{e}_3^m\}$.

of the material (*martensite*). The discontinuous change in the lattice at the transformation temperature implies that $\{\boldsymbol{e}_1^a(\theta_T), \boldsymbol{e}_2^a(\theta_T), \boldsymbol{e}_3^a(\theta_T)\} \neq \{\boldsymbol{e}_1^m(\theta_T), \boldsymbol{e}_2^m(\theta_T), \boldsymbol{e}_3^m(\theta_T)\}$. Figure 2.2 illustrates a material with cubic and tetragonal phases (lattices).

We will be concerned with relatively modest changes in temperature from the transformation temperature. Consequently the change in the lattice due to thermal expansion can be neglected in comparison with that due to phase transformation. Thus we shall take both sets of lattice vectors $\{\boldsymbol{e}_1^a, \boldsymbol{e}_2^a, \boldsymbol{e}_3^a\}$ and $\{\boldsymbol{e}_1^m, \boldsymbol{e}_2^m, \boldsymbol{e}_3^m\}$ to be independent of temperature.

Consider the austenitic and martensitic lattices $\mathcal{L}(\boldsymbol{e}^a, \boldsymbol{o})$ and $\mathcal{L}(\boldsymbol{e}^m, \boldsymbol{o})$. Since the associated lattice vectors $\{\boldsymbol{e}_1^a, \boldsymbol{e}_2^a, \boldsymbol{e}_3^a\}$ and $\{\boldsymbol{e}_1^m, \boldsymbol{e}_2^m, \boldsymbol{e}_3^m\}$ are each linearly independent, there is a nonsingular tensor $\boldsymbol{U}_1$ which relates them:

$$\boldsymbol{e}_i^m = \boldsymbol{U}_1 \boldsymbol{e}_i^a \ . \tag{2.7}$$

The lattice $\mathcal{L}(\boldsymbol{e}^m, \boldsymbol{o})$ can be viewed as the image of the lattice $\mathcal{L}(\boldsymbol{e}^a, \boldsymbol{o})$ under the linear transformation $\boldsymbol{U}_1$ which is called the *Bain stretch tensor*. A characteristic of a martensitic transformation is that the change of crystal structure is diffusionless, i.e., there is no rearrangement of atoms associated with the lattice change. Consequently the kinematics of the transformation is completely characterized by $\boldsymbol{U}_1$. Given the two lattices, $\boldsymbol{U}_1$ can be calculated from (2.7); we will assume that $\boldsymbol{U}_1$ is symmetric though this is not essential.

The fact that the austenite phase is usually observed above the transformation temperature and the martensite phase is observed below it, does *not* imply that these phases cease

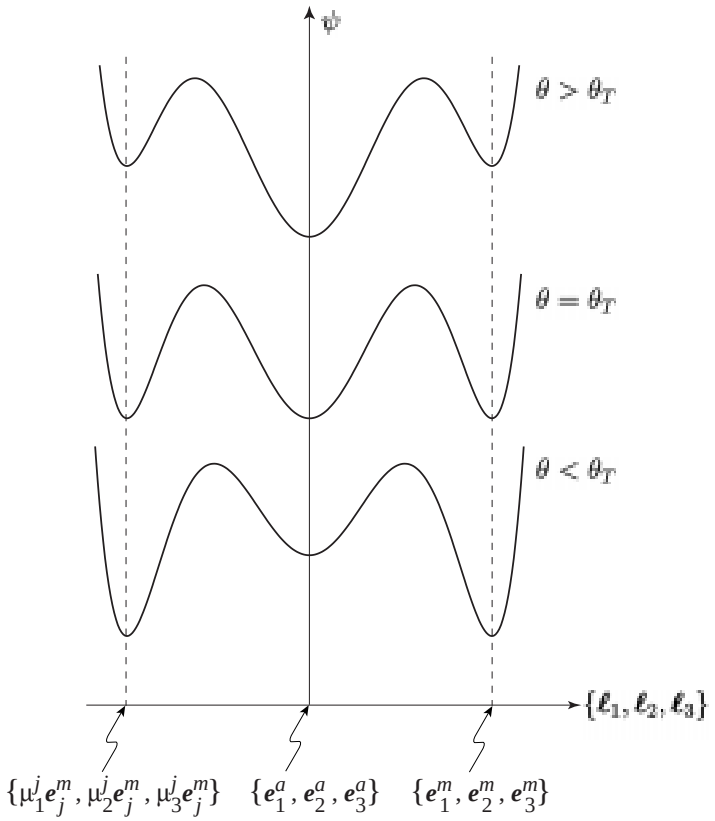

Figure 2.3: Free-energy ψ as a function of lattice vectors $\{\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3\}$ at three different temperatures θ; the transformation temperature is denoted by θ_T. The energy has local minima at the austenite lattice vectors $\{\boldsymbol{e}_1^a, \boldsymbol{e}_2^a, \boldsymbol{e}_3^a\}$ and the martensite lattice vectors $\{\boldsymbol{e}_1^m, \boldsymbol{e}_2^m, \boldsymbol{e}_3^m\}$. For any $[\mu] \in \mathit{I\!I}$, the vectors $\{\mu_1^j \boldsymbol{\ell}_j, \mu_2^j \boldsymbol{\ell}_j, \mu_3^j \boldsymbol{\ell}_j\}$ and $\{\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3\}$ describe the same lattice, and so the energy has additional minima at certain symmetry related lattice vectors; this is illustrated in the figure by the third energy-well at the martensitic "variant" $\{\mu_1^j \boldsymbol{e}_j^m, \mu_2^j \boldsymbol{e}_j^m, \mu_3^j \boldsymbol{e}_j^m\}$. Observe that because we have omitted the effect of thermal expansion, the location of each local minimum does not change with temperature.

to exist, respectively, below and above, this temperature. Rather, it implies an exchange in stability between these phases: the austenite lattice is stable for $\theta > \theta_T$, while the martensite lattice is stable for $\theta < \theta_T$. Indeed, one often observes (metastable) austenite below θ_T and (metastable) martensite above θ_T. Assuming that the notion of stability here is that of energy minimization, the preceding stability properties imply that the free-energy $\psi(\cdot, \theta)$ is minimized by the lattice vectors $\{\boldsymbol{e}_1^a, \boldsymbol{e}_2^a, \boldsymbol{e}_3^a\}$ for $\theta > \theta_T$ and by $\{\boldsymbol{e}_1^m, \boldsymbol{e}_2^m, \boldsymbol{e}_3^m\}$ for $\theta < \theta_T$. At the transformation temperature θ_T, both lattices have equal energy. Thus we shall assume

that

$$\begin{aligned} \psi(\boldsymbol{e}_1^a,\boldsymbol{e}_2^a,\boldsymbol{e}_3^a;\theta) &\leq \psi(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2,\boldsymbol{\ell}_3;\theta) \qquad \text{for } \theta > \theta_T, \\ \psi(\boldsymbol{e}_1^a,\boldsymbol{e}_2^a,\boldsymbol{e}_3^a;\theta) = \psi(\boldsymbol{e}_1^m,\boldsymbol{e}_2^m,\boldsymbol{e}_3^m;\theta) &\leq \psi(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2,\boldsymbol{\ell}_3;\theta) \qquad \text{for } \theta = \theta_T, \\ \psi(\boldsymbol{e}_1^m,\boldsymbol{e}_2^m,\boldsymbol{e}_3^m;\theta) &\leq \psi(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2,\boldsymbol{\ell}_3;\theta) \qquad \text{for } \theta < \theta_T, \end{aligned} \tag{2.8}$$

for all triplets of linearly independent vectors $\{\boldsymbol{\ell}_1,\boldsymbol{\ell}_2,\boldsymbol{\ell}_3\}$; see Figure 2.3. Observe from $(2.8)_2$ that $\psi(\cdot,\theta_T)$ has (at least) two distinct energy-wells, and therefore by continuity, so does $\psi(\cdot,\theta)$ for temperatures close to θ_T.

In summary, the free-energy function ψ, which is a function of the lattice vectors and temperature, is to satisfy the requirements of frame-indifference (2.5), lattice symmetry (2.6) and energy minimization (2.8).

Continuum model: We now proceed from the preceding lattice model to a related continuum model. Consider a crystalline solid, modeled as a continuum, which occupies some region $\mathcal{R}_o$ in a reference configuration. Suppose that underlying each point $\boldsymbol{X} \in \mathcal{R}_o$, there is a Bravais lattice with lattice vectors $\{\boldsymbol{e}_1^o(\boldsymbol{X}),\boldsymbol{e}_2^o(\boldsymbol{X}),\boldsymbol{e}_2^o(\boldsymbol{X})\}$. Now suppose that the continuum undergoes a deformation $\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{X})$ and let $\boldsymbol{F}(\boldsymbol{X}) = \nabla\boldsymbol{x}(\boldsymbol{X})$ be the associated deformation gradient. Let $\{\boldsymbol{e}_1(\boldsymbol{X}),\boldsymbol{e}_2(\boldsymbol{X}),\boldsymbol{e}_3(\boldsymbol{X})\}$ be the lattice vectors of the deformed lattice at the same material point $\boldsymbol{X}$. The Cauchy-Born hypothesis says that the lattice vectors deform according to the macroscopic deformation gradient, i.e.,

$$\boldsymbol{e}_i(\boldsymbol{X}) = \boldsymbol{F}(\boldsymbol{X})\,\boldsymbol{e}_i^o(\boldsymbol{X})\ . \tag{2.9}$$

In other words, the lattice vectors behave like "material filaments"; see Figure 2.4

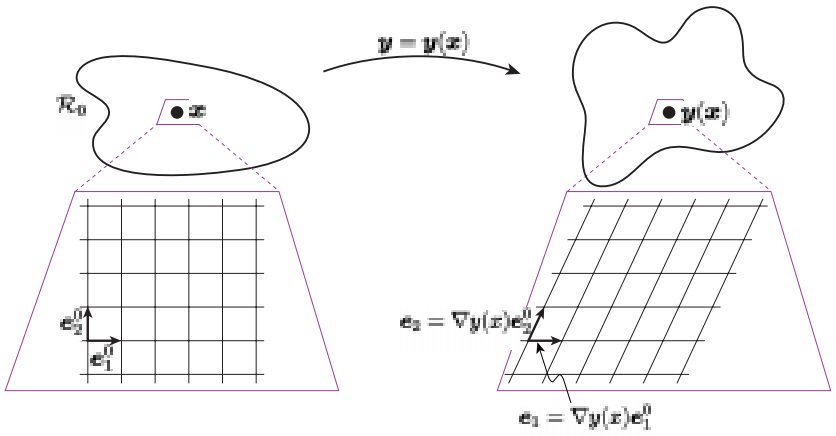

Figure 2.4: The relation between the continuum and the lattice.

We can now use (2.9) to define the continuum free-energy function $W(\boldsymbol{F},\theta)$ in terms of the lattice free-energy ψ through

$$W(\boldsymbol{F},\theta) = \psi(\boldsymbol{F}\boldsymbol{e}_1^o, \boldsymbol{F}\boldsymbol{e}_2^o, \boldsymbol{F}\boldsymbol{e}_3^o;\theta) \ . \tag{2.10}$$

Note that the free-energy function W depends on the choice of reference configuration because the right hand side of (2.10) involves the lattice vectors associated with that configuration. It follows from (2.5) and (2.10) that W inherits the frame-indifference property that was previously built into ψ:

$$\text{Frame-indifference:} \qquad W(\boldsymbol{Q}\boldsymbol{F},\theta) = W(\boldsymbol{F},\theta) \qquad \text{for all rotations } \boldsymbol{Q} \ . \tag{2.11}$$

Turning next to material symmetry, recall that the lattice free-energy ψ has the property (2.6) which ensures that it depends on the lattice but not the choice of lattice vectors. It is convenient to rewrite (2.6) in the equivalent form

$$\psi(\boldsymbol{H}\boldsymbol{e}_1, \boldsymbol{H}\boldsymbol{e}_2, \boldsymbol{H}\boldsymbol{e}_3;\theta) = \psi(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3;\theta) \qquad \text{for all } \boldsymbol{H} \in \mathcal{G}(\mathcal{L}), \tag{2.12}$$

where

$$\mathcal{G}(\mathcal{L}) = \left\{ \boldsymbol{H} \,:\, \boldsymbol{H}\boldsymbol{e}_i = \mu_i^j \boldsymbol{e}_j, \quad [\mu] \in \mathit{I\!I} \right\} \ . \tag{2.13}$$

Observe that $\mathcal{G}(\mathcal{L})$ denotes the set of all linear transformations that map the lattice back into itself, in the sense that $\mathcal{L}(\boldsymbol{e},\boldsymbol{o}) = \mathcal{L}(\boldsymbol{f},\boldsymbol{o})$ if and only if $\boldsymbol{f}_i = \boldsymbol{H}\boldsymbol{e}_i$ for some $\boldsymbol{H} \in \mathcal{G}(\mathcal{L})$. The group $\mathcal{G}(\mathcal{L})$ characterizes the symmetry of the lattice and one can verify that it depends on the lattice $\mathcal{L}$ and not on the particular choice of lattice vectors on the right hand side of (2.13). In view of (2.10) and (2.12), the continuum free-energy inherits the symmetry property

$$W(\boldsymbol{F}\boldsymbol{H},\theta) = W(\boldsymbol{F},\theta) \qquad \text{for all } \boldsymbol{H} \in \mathcal{G}(\mathcal{L}_o) \tag{2.14}$$

where $\mathcal{L}_o$ is the lattice associated with the reference configuration. However, note that in addition to rotations and reflections, the symmetry group $\mathcal{G}$ contains finite shears as well; see, for example, Figure 2.1. Such shears cause large distortions of the lattice and are usually associated with lattice slip and plasticity. In situations where thermoelasticity provides an adequate mathematical model, it is natural therefore to exclude these large shears and restrict attention to a suitable subgroup of $\mathcal{G}$; see Pitteri (1984) and Ball and James (1992) for a precise analysis of this issue. The appropriate sub-group is the *Laue group* $\mathcal{P}$ which is the set of rotations that map a lattice back into itself:

$$\mathcal{P}(\mathcal{L}) = \{ \boldsymbol{R} \,:\, \boldsymbol{R} \text{ is proper orthogonal, } \boldsymbol{R} \in \mathcal{G}(\mathcal{L}) \} \ . \tag{2.15}$$

For example, the Laue group of a simple cubic lattice consists of the 24 rotations that map the unit cube back into itself. Thus instead of (2.14), we shall require the continuum free-energy to conform to the less stringent requirement

$$\text{Material Symmetry:} \qquad W(\boldsymbol{F}\boldsymbol{R},\theta) = W(\boldsymbol{F},\theta) \qquad \text{for all } \boldsymbol{R} \in \mathcal{P}(\mathcal{L}_o). \tag{2.16}$$

It is worth emphasizing that it is the symmetry of the reference lattice $\mathcal{L}_o$ that enters here.

Now consider a two-phase martensitic material. Suppose that the reference configuration coincides with a stress-free crystal of austenite at the transformation temperature. Then the reference lattice is the austenite lattice ($\boldsymbol{e}^o = \boldsymbol{e}^a$); moreover, $\boldsymbol{F} = \boldsymbol{I}$ describes stress-free austenite, and $\boldsymbol{F} = \boldsymbol{U}_1$ characterizes stress-free martensite where $\boldsymbol{U}_1$ is the Bain stretch tensor introduced in (2.7). Thus it follows from (2.7), (2.8) and (2.10) that W inherits the following multi-well structure:

$$\begin{aligned} W(\boldsymbol{I},\theta) &\leq W(\boldsymbol{F},\theta) && \text{for } \theta > \theta_T, \\ W(\boldsymbol{I},\theta) = W(\boldsymbol{U}_1,\theta) &\leq W(\boldsymbol{F},\theta) && \text{for } \theta = \theta_T, \\ W(\boldsymbol{U}_1,\theta) &\leq W(\boldsymbol{F},\theta) && \text{for } \theta < \theta_T, \end{aligned} \tag{2.17}$$

for all nonsingular $\boldsymbol{F}$.

In summary, a continuum-scale free-energy function $W(\boldsymbol{F},\theta)$ that characterizes a two-phase material must satisfy the requirements of frame-indifference (2.11), material symmetry (2.16) and energy minimization (2.17). Observe that for such a material, the function $W(\cdot,\theta_T)$, and therefore by continuity $W(\cdot,\theta)$ for θ close to θ_T, has (at least) two distinct energy-wells.

Variants: The three properties – frame-indifference, material symmetry and energy minimization – imply certain important characteristics of W. For example for $\theta > \theta_T$, it follows from frame-indifference and the fact that $\boldsymbol{F} = \boldsymbol{I}$ minimizes W, that necessarily $\boldsymbol{F} = \boldsymbol{Q}$ also minimizes W for any rotation $\boldsymbol{Q}$. Similarly for $\theta < \theta_T$, the free-energy is minimized by $\boldsymbol{Q}\boldsymbol{U}_1\boldsymbol{R}$ for an arbitrary rotation $\boldsymbol{Q}$ and any rotation $\boldsymbol{R} \in \mathcal{P}(\mathcal{L}_o)$.

In order to explore this systematically, note first by combining frame-indifference (2.11) and material symmetry (2.16) that

$$W(\boldsymbol{R}^T\boldsymbol{F}\boldsymbol{R},\theta) = W(\boldsymbol{F},\theta) \qquad \text{for all } \boldsymbol{R} \in \mathcal{P}_a \tag{2.18}$$

where $\mathcal{P}_a = \mathcal{P}(\mathcal{L}(\boldsymbol{e}^a))$ is the Laue group of the reference configuration – unstressed austenite. It follows from this that, since $\boldsymbol{U}_1$ minimizes the free-energy for $\theta < \theta_T$, so does the tensor $\boldsymbol{R}^T\boldsymbol{U}_1\boldsymbol{R}$ for all $\boldsymbol{R} \in \mathcal{P}_a$. As $\boldsymbol{R}$ varies over the set of all rotations in $\mathcal{P}_a$, the tensor $\boldsymbol{R}^T\boldsymbol{U}_1\boldsymbol{R}$

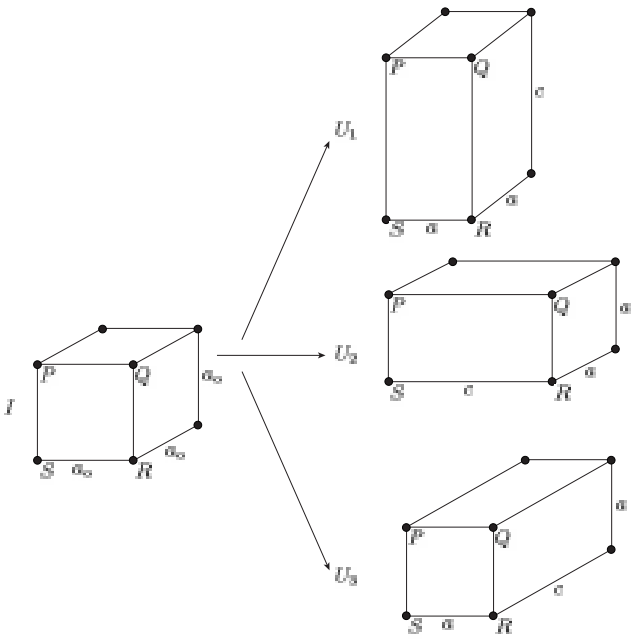

Figure 2.5: Cubic austenite and three variants of tetragonal martensite. Observe that one *cannot* rigidly rotate a martensite variant in such a way as to make the atoms P,Q,R,S coincide with the locations of these same atoms in any other martensite variant.

takes on various symmetric tensor values. Suppose, as is the case for most martensitic transformations, that the austenite lattice has greater symmetry than the martensite lattice in the sense that $\mathcal{P}_m \subset \mathcal{P}_a$. Then, one can show that $\boldsymbol{R}^T\boldsymbol{U}_1\boldsymbol{R} = \boldsymbol{U}_1$ for all $\boldsymbol{R} \in \mathcal{P}_m$. However, as $\boldsymbol{R}$ varies over the rotations in $\mathcal{P}_a$ which are not in $\mathcal{P}_m$, then $\boldsymbol{R}^T\boldsymbol{U}_1\boldsymbol{R}$ takes on a certain finite number of distinct, symmetric tensor values, which we denote by $\boldsymbol{U}_2, \boldsymbol{U}_3, \ldots, \boldsymbol{U}_N$; these stretch tensors are said to describe the *variants* of martensite; the number of variants, N, is given by the formula

$$N = \frac{\text{the order of } \mathcal{P}_a}{\text{the order of } \mathcal{P}_m} \ . \tag{2.19}$$

For example, in the case of a material which is cubic in the austenite phase and tetragonal in the martensite phase, the orders of $\mathcal{P}_a$ and $\mathcal{P}_m$ are 24 and 8 respectively; thus there are three variants of martensite characterized by three stretch tensors $\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{U}_3$; see Figure 2.5. Note because of (2.18) that all variants have the same energy:

$$W(\boldsymbol{U}_1, \theta) = W(\boldsymbol{U}_2, \theta) = \cdots = W(\boldsymbol{U}_N, \theta) \ . \tag{2.20}$$

Thus W has N distinct martensitic local minima at $\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{U}_3, \ldots, \boldsymbol{U}_N$ all of which have the same energy. A schematic graph of W versus $\boldsymbol{U}$ would therefore have the same generic form as the curve shown in Figure 2.3, with the minimum at $\{\boldsymbol{e}_1^m, \boldsymbol{e}_2^m, \boldsymbol{e}_3^m\}$ corresponding to one at $\boldsymbol{U}_1$ and the minimum at $\{[\mu]\boldsymbol{e}_1^m, [\mu]\boldsymbol{e}_2^m, [\mu]\boldsymbol{e}_3^m\}$ corresponding to one at some $\boldsymbol{U}_k$.

Finally, returning to the notion of frame-indifference, it follows from (2.11) that austenite corresponds not only to the identify tensor $\boldsymbol{I}$, but also to all rotations of the identity. Similarly, the first variant of martensite corresponds to all tensors of the form $\boldsymbol{Q}\boldsymbol{U}_1$ where $\boldsymbol{Q}$ is an arbitrary rotation, and so on. Therefore, the following *sets* of tensors comprise the bottoms of the *austenite energy-well* $\mathcal{A}$ and the N *martensitic energy-wells* $\mathcal{M}_k$:

$$\left.\begin{array}{rcl} \mathcal{A} &=& \{\boldsymbol{F}\colon\ \boldsymbol{F}=\boldsymbol{Q} \text{ for all rotations } \boldsymbol{Q}\}\ , \\ \mathcal{M}_k &=& \{\boldsymbol{F}\colon\ \boldsymbol{F}=\boldsymbol{Q}\boldsymbol{U}_k \text{ for all rotations } \boldsymbol{Q}\}, \quad k=1,2,\dots N, \end{array}\right\} \tag{2.21}$$

Let $\mathcal{M}$ denote the collection of all martensite wells:

$$\mathcal{M}=\mathcal{M}_1\bigcup\mathcal{M}_2\bigcup\cdots\bigcup\mathcal{M}_N\ . \tag{2.22}$$

In *summary*, the continuum-scale free-energy $W(\boldsymbol{F},\theta)$ associated with a two-phase martensitic material has the property that W is minimized on the austenite energy-well $\mathcal{A}$ at high temperatures, it is minimized on the martensite energy-well $\mathcal{M}$ at low temperatures, and on both the austenite and martensite wells at the transformation temperature θ_T, i.e.,

$$\begin{array}{rcl} W(\boldsymbol{F},\theta) &>& W(\boldsymbol{F}',\theta) \quad \text{for all} \quad \boldsymbol{F}\notin\mathcal{A}, \quad \boldsymbol{F}'\in\mathcal{A}, \quad \theta>\theta_T\ , \\ W(\boldsymbol{F},\theta) &>& W(\boldsymbol{F}',\theta) \quad \text{for all} \quad \boldsymbol{F}\notin\mathcal{A}\bigcup\mathcal{M}, \quad \boldsymbol{F}'\in\mathcal{A}\bigcup\mathcal{M}, \quad \theta=\theta_T\ , \\ W(\boldsymbol{F},\theta) &>& W(\boldsymbol{F}',\theta) \quad \text{for all} \quad \boldsymbol{F}\notin\mathcal{M}, \quad \boldsymbol{F}'\in\mathcal{M}, \quad \theta<\theta_T\ . \end{array} \tag{2.23}$$

3. A strain-energy function with multiple local minima: a material with cubic and tetragonal phases. (Ericksen, 1978, 1980, 1986) Perhaps the simplest example of a two-phase martensitic material is one that occurs as a face-centered cubic austenite phase and a face-centered tetragonal martensite phase; examples of such materials include In-Tl, Mn-Ni and Mn-Cu. Let $\{\boldsymbol{c}_1,\boldsymbol{c}_2,\boldsymbol{c}_3\}$ denote fixed *unit* vectors in the cubic directions and let the lattice vectors associated with the austenite lattice be denoted by $\{\boldsymbol{e}_1^a,\boldsymbol{e}_2^a,\boldsymbol{e}_3^a\}$. Observe from Figure 3.1 that (in contrast to a simple cubic material) for a face-centered cubic material one *cannot* choose the austenite lattice vectors to be a scalar multiple of $\{\boldsymbol{c}_1,\boldsymbol{c}_2,\boldsymbol{c}_3\}$. Instead, if $a_o\times a_o\times a_o$ denotes the lattice parameters of the cubic phase (see figure) one acceptable choice for the lattice vectors is

$$\boldsymbol{e}_1^a=\frac{a_o}{2}\ (\boldsymbol{c}_2+\boldsymbol{c}_3)\,, \quad \boldsymbol{e}_2^a=\frac{a_o}{2}\ (-\boldsymbol{c}_2+\boldsymbol{c}_3)\,, \quad \boldsymbol{e}_3^a=\frac{a_o}{2}\ (\boldsymbol{c}_1+\boldsymbol{c}_3)\,. \tag{3.1}$$

Similarly for the tetragonal phase, if the lattice parameters are denoted by $a\ \times a\ \times c$, the lattice vectors $\{\boldsymbol{e}_1^m,\boldsymbol{e}_2^m,\boldsymbol{e}_3^m\}$ can be taken to be

$$\boldsymbol{e}_1^m=\frac{a}{2}\ (\boldsymbol{c}_2+\boldsymbol{c}_3)\,, \quad \boldsymbol{e}_2^m=\frac{a}{2}\ (-\boldsymbol{c}_2+\boldsymbol{c}_3)\,, \quad \boldsymbol{e}_3^m=\frac{c}{2}\ \boldsymbol{c}_1+\frac{a}{2}\ \boldsymbol{c}_3. \tag{3.2}$$

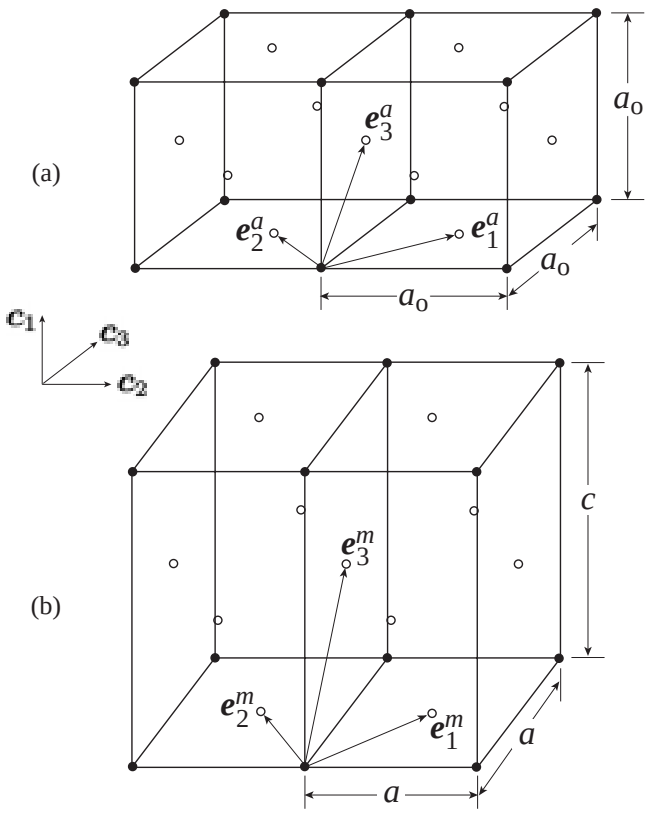

Figure 3.1: Face-centered cubic lattice with lattice parameters $a_o \times a_o \times a_o$ and lattice vectors $\{\boldsymbol{e}_1^a, \boldsymbol{e}_2^a, \boldsymbol{e}_3^a\}$, and face-centered tetragonal lattice with lattice parameters $a \times a \times c$ and lattice vectors $\{\boldsymbol{e}_1^m, \boldsymbol{e}_2^m, \boldsymbol{e}_3^m\}$. The unit vectors $\{\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3\}$ are associated with the cubic directions. Solely for purposes of clarity, the atoms at the vertices are depicted by filled dots while those at the centers of the faces are shown by open dots.

One can readily verify that the mapping $\boldsymbol{e}_i^a \rightarrow \boldsymbol{e}_i^m = \boldsymbol{U}_1 \boldsymbol{e}_i^a$ between the two sets of lattice vectors is described by the Bain stretch tensor

$$\boldsymbol{U}_1 \;=\; \beta \boldsymbol{c}_1 \otimes \boldsymbol{c}_1 + \alpha \boldsymbol{c}_2 \otimes \boldsymbol{c}_2 + \alpha \boldsymbol{c}_3 \otimes \boldsymbol{c}_3 \tag{3.3}$$

where the stretches α and β are given by

$$\alpha \;=\; \frac{a}{a_o}, \qquad \beta \;=\; \frac{c}{a_o}. \tag{3.4}$$

Thus the transformation from the cubic lattice to the tetragonal lattice is achieved by stretching the parent lattice equally in two of the cubic directions and unequally in the third direction, the stretches associated with this being α, α and β.

As noted in the previous section, there are three variants of martensite associated with a cubic/tetragonal material. This was illustrated in Figure 2.5 which showed that there are three possible ways in which to stretch the cubic lattice in order to obtain the tetragonal

lattice, the aforementioned one being the one where the unequal stretching occurs in the $\boldsymbol{c}_1$ direction. More generally, one can stretch the cubic lattice by the ratio β in the $\boldsymbol{c}_k$ direction and by α in the remaining two cubic directions; this leads to the three Bain stretch tensors

$$\boldsymbol{U}_k \;=\; \alpha\boldsymbol{I} + [\beta - \alpha]\boldsymbol{c}_k \otimes \boldsymbol{c}_k, \qquad k = 1, 2, 3, \tag{3.5}$$

whose components in the cubic basis are

$$[U_1] = \begin{pmatrix} \beta & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \alpha \end{pmatrix}, \quad [U_2] = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \alpha \end{pmatrix}, \quad [U_3] = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \beta \end{pmatrix}. \tag{3.6}$$

We now construct an explicit strain-energy function which can be used to describe a material which possesses a cubic phase and a tetragonal phase. In view of frame indifference, W depends on $\boldsymbol{F}$ only through the Cauchy-Green deformation tensor $\boldsymbol{C} = \boldsymbol{F}^T\boldsymbol{F}$, or equivalently the Lagrangian strain tensor $\boldsymbol{E} = (1/2)(\boldsymbol{C} - \boldsymbol{I})$. As before, suppose that we take the reference configuration to coincide with an unstressed state of austenite at the transformation temperature. Then the reference configuration, and therefore the strain-energy function $W(\boldsymbol{E}, \theta)$, must possess cubic symmetry. It is known (see, for example, Smith and Rivlin, 1958 and Green and Adkins, 1970) that, to have cubic symmetry, W must be a function of the "cubic invariants"

$$\begin{aligned} &I_1 = E_{11} + E_{22} + E_{33}, && I_2 = E_{11}E_{22} + E_{22}E_{33} + E_{33}E_{11} \\ &I_3 = E_{11}E_{22}E_{33}, && I_4 = E_{12}E_{23}E_{31}, \\ &I_5 = E_{12}^2 + E_{23}^2 + E_{31}^2, && \ldots \text{etc.}, \end{aligned} \tag{3.7}$$

where E_{ij} refers to the i,j-component of $\boldsymbol{E}$ in the cubic basis $\{\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3\}$. The number of invariants in this list depends on the particular cubic class under consideration (e.g. section 1.11 of Green and Adkins, 1970) but the analysis below is valid for all of these classes.

For temperatures close to the transformation temperature, Ericksen has argued based on experimental observations, that as a first approximation, all of the shear strain components (in the cubic basis) vanish, and additionally, that the sum of the normal strains also vanishes. Accordingly he suggested a geometrically constrained theory based on the two constraints

$$\begin{aligned} I_1 &= E_{11} + E_{22} + E_{33} = 0, \\ I_5 &= E_{12}^2 + E_{23}^2 + E_{31}^2 = 0. \end{aligned} \tag{3.8}$$

In this case, the only two nontrivial strain invariants among those in the preceding list are I_2 and I_3 and so $W = W(I_2, I_3, \theta)$. Suppose that W is a polynomial in the components

of $\boldsymbol{E}$. In order to capture the desired multi-well character, W must be at least a quartic polynomial. It follows from (3.7) and (3.8) that the most general quartic polynomial of the form $W = W(I_2, I_3, \theta)$ is

$$W = c_0 + c_2 I_2 + c_3 I_3 + c_{22} I_2^2, \tag{3.9}$$

where the coefficients c_i may depend on temperature: $c_i = c_i(\theta)$.

In order to describe the cubic phase, this function W must have a local minimum at $\boldsymbol{E} = \boldsymbol{O}$, whereas to capture the tetragonal variants, W must have local minima at $\boldsymbol{E} = \boldsymbol{E}_k = (1/2)(\boldsymbol{U}_k^2 - \boldsymbol{I})$, $k = 1, 2, 3$, where the $\boldsymbol{U}_k$'s are given by (3.5). We now impose these requirements. The function W given in (3.9) automatically has an extremum at $\boldsymbol{E} = \boldsymbol{O}$, and this is a local minimum if $c_2 < 0$. Next, recall from the preceding section that, in the presence of material symmetry, if W has a local minimum at any one of the tensors $\boldsymbol{E}_k$, then W will automatically have energy-wells at the remaining two $\boldsymbol{E}_k$'s. In order for W to have an extremum at $\boldsymbol{E} = \boldsymbol{E}_1$ it is necessary that $c_2 = -pc_3 + 6p^2 c_{22}$, where we have set

$$p = \frac{1}{2}(\alpha^2 - 1). \tag{3.10}$$

It should be kept in mind that the constraint $I_1 = 0$ implies that tr $\boldsymbol{E}_1 = 0$ and therefore the lattice stretches α and β must be related by $2\alpha^2 + \beta^2 = 3$. If the extremum at $\boldsymbol{E} = \boldsymbol{E}_1$ is to be a local minimum one must also have $12pc_{22} > c_3 > 0$ where we are considering the case $p > 0$. Combining these various requirements leads to

$$c_2 = -pc_3 + 6p^2 c_{22}, \qquad 12pc_{22} > c_3 > 6pc_{22} > 0. \tag{3.11}$$

Finally, the values of W at the bottoms of the two energy-wells are $W(\boldsymbol{O}, \theta) = c_o(\theta)$ and $W(\boldsymbol{E}_1, \theta) = c_o(\theta) + p^3[c_3(\theta) - 9pc_{22}(\theta)]$. Since the martensite minimum must be lower than the austenite minimum for low temperatures, vice versa for high temperatures, and the values of W at the two minima must be the same at the transformation temperature, we require that

$$c_3'(\theta) - 9pc_{22}'(\theta) > 0, \qquad c_3(\theta_T) = 9pc_{22}(\theta_T), \tag{3.12}$$

where the prime denotes differentiation with respect to the argument. Thus, in summary, by collecting all of the preceding requirements, and introducing a more convenient pair of coefficients $d(\theta)$ and $e(\theta)$, the strain-energy function W can be expressed as

$$\begin{gathered} W = c_o(\theta) + d(\theta)\left[\frac{I_3}{p^3} - \frac{I_2}{p^2}\right] + e(\theta)\left[\left(\frac{I_2}{p^2}\right)^2 - 3\,\frac{I_2}{p^2} + 9\,\frac{I_3}{p^3}\right], \\ e(\theta) > 0, \quad 3e(\theta) > d(\theta) > -3e(\theta), \quad d'(\theta) > 0, \quad d(\theta_T) = 0. \end{gathered} \tag{3.13}$$

Figure 3.2 shows a contour plot of the strain-energy (3.13). In view of the constraints (3.8) there are only two independent strain components E_{11}, E_{22}. The figure shows the contours of

$W(E_{11}, E_{22}, \theta)$, as given by (3.13), on the E_{11}, E_{22}-plane. It shows the austenitic energy-well at the origin $\boldsymbol{E} = \boldsymbol{O}$ surrounded by the three martensitic energy-wells at $\boldsymbol{E} = \boldsymbol{E}_1, \boldsymbol{E}_2$ and $\boldsymbol{E}_3$. The values of the material parameters associated with this figure where chosen solely on the basis of obtaining a fairly clear contour plot.

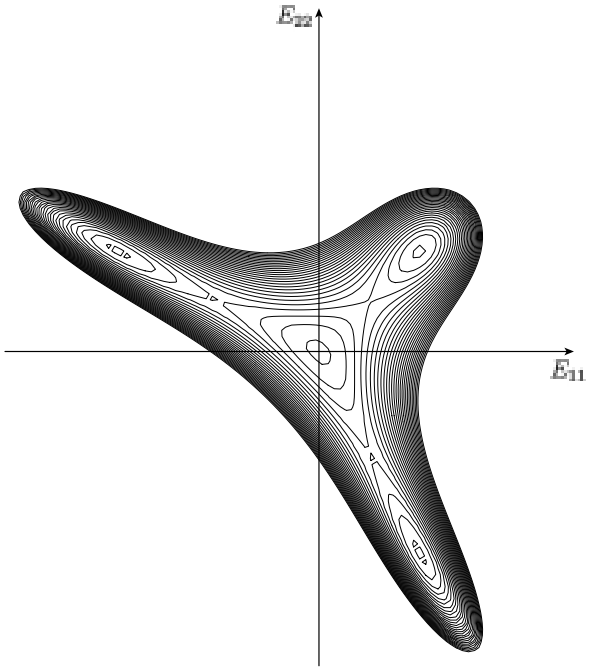

Figure 3.2: Contours of constant strain-energy on the E_{11}, E_{22}-plane. Observe the presence of the austenitic energy-well at $\boldsymbol{E} = \boldsymbol{O}$ surrounded by the three martensitic energy-wells at $\boldsymbol{E} = \boldsymbol{E}_1, \boldsymbol{E}_2, \boldsymbol{E}_3$.

In order to draw a one-dimensional cross-section of the energy (3.13), consider the path in strain space $\boldsymbol{E} = \boldsymbol{E}(\varepsilon),\ -1.5 < \varepsilon < 1.5$, where

$$[E(\varepsilon)] \ = \ \begin{pmatrix} p\varepsilon^2 & 0 & 0 \\ 0 & p\varepsilon(3-\varepsilon)/2 & 0 \\ 0 & 0 & -p\varepsilon(3+\varepsilon)/2 \end{pmatrix}. \tag{3.14}$$

Observe that $\boldsymbol{E}(0) = \boldsymbol{O},\ \boldsymbol{E}(-1) = \boldsymbol{E}_2, \boldsymbol{E}(1) = \boldsymbol{E}_3$ so that this path passes through the austenite well and two of the martensite wells. Figure 3.3 shows the variation of energy along this path at three different temperatures: the figure displays $w(\varepsilon) = W(\boldsymbol{E}(\varepsilon), \theta)$ versus ε. The three graphs correspond to three temperatures greater than, equal to, and less than, the transformation temperature θ_T.

The strain-energy function (3.13) captures the key qualitative characteristics of a martensitic material which exists in cubic and tetragonal phases. However, due to the restrictive

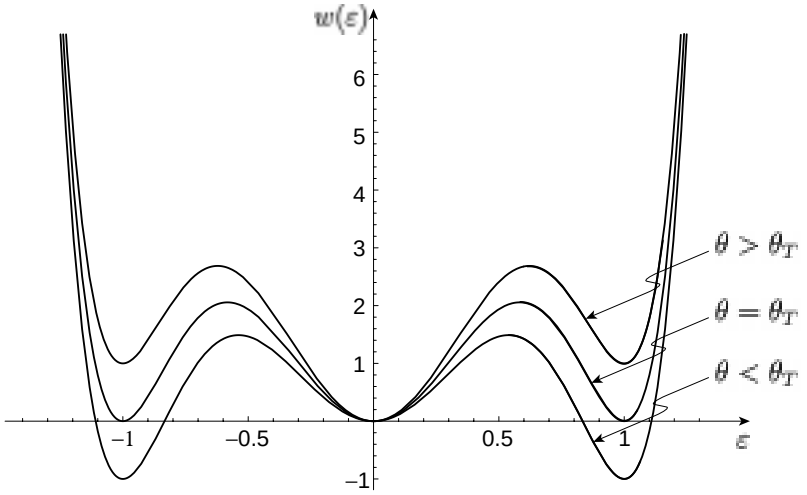

Figure 3.3: A one-dimensional cross-section of the energy (3.13) along the path (3.14) in strain-space. The figure plots $w(\varepsilon) = W(\boldsymbol{E}(\varepsilon))$ versus strain ε at three different temperatures. Because we have omitted the effect of thermal expansion, the location of the local minima do not change with temperature. The material parameters underlying this plot are the same as those associated with the previous figure.

nature of the kinematic constraints (3.8), it fails to provide a *quantitatively* accurate model. The natural generalization of (3.13) is therefore to relax these constraints by using Lagrange multipliers c_5 and c_{11} and to replace (3.9) by

$$W = c_0 + c_2 I_2 + c_3 I_3 + c_{22} I_2^2 + c_5 I_5 + c_{11} I_1^2, \qquad c_i = c_i(\theta). \tag{3.15}$$

James (see, for example, Kloućek and Luskin, 1994) has shown that the response predicted by this generalized form of W is in reasonable agreement with the observed behavior of In-Tl.

4. Uniaxial motion of a slab. Formulation. In order to explain some of the key ideas with a minimum of mathematical complexity, it is convenient to work in an essentially one-dimensional isothermal setting. In this section we formulate the basic equations pertaining to such a setting, and in the following two sections we shall make use of it.

Consider a slab, which in an unstressed reference configuration occupies the region $0 < X_1 < L, -\infty < X_2, X_3 < \infty$, and consider uniaxial motions of the form

$$\boldsymbol{x} = \boldsymbol{X} + u(X_1, t)\boldsymbol{e}_1. \tag{4.1}$$

Here $\boldsymbol{e}_1$ is a unit vector that is normal to the faces of the slab. Since only the coordinate X_1 plays a central role in what follows, from hereon we shall set

$$X_1 = x. \tag{4.2}$$

The deformation gradient tensor associated with (4.1) is

$$\boldsymbol{F}(x,t) = \boldsymbol{I} \ + \ \gamma(x,t)\ \boldsymbol{e}_1 \otimes \boldsymbol{e}_1, \qquad \text{where} \qquad \gamma(x,t) = \frac{\partial u}{\partial x}; \tag{4.3}$$

γ is a measure of the normal strain. Note that $\det \boldsymbol{F} = 1 + \gamma$, and since we need $\det \boldsymbol{F} > 0$, it is necessary that

$$\gamma > -1. \tag{4.4}$$

The particle velocity associated with (4.1) is

$$\boldsymbol{v} = v(x,t)\boldsymbol{e}_1 \qquad \text{where} \qquad v(x,t) = \frac{\partial u}{\partial t}. \tag{4.5}$$

Suppose that the slab is composed of an elastic material characterized by a strain-energy function $W(\boldsymbol{F})$. It is convenient for our present purposes to consider the restriction of the energy to uniaxial motions, and therefore to introduce the function w defined by

$$w(\gamma) = W(\boldsymbol{I} \ + \ \gamma \boldsymbol{e}_1 \otimes \boldsymbol{e}_1), \qquad \gamma > -1. \tag{4.6}$$

The first Piola-Kirchhoff stress tensor $\boldsymbol{S}$ is given by the constitutive relation

$$\boldsymbol{S} = \left. \frac{\partial W}{\partial \boldsymbol{F}} \right|_{\boldsymbol{F} = \boldsymbol{I} \ + \ \gamma \boldsymbol{e}_1 \otimes \boldsymbol{e}_1}. \tag{4.7}$$

We assume that the symmetry of the material is such that, in the uniaxial motion (4.1), all shear stress components calculated from (4.7) vanish identically (where the components are in an orthonormal basis $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ with $\boldsymbol{e}_1$ normal to the slab faces). This would be true, for example, if the material was isotropic, or even for certain anisotropic materials provided the material is suitably oriented with respect to the slab. Let σ denote the normal stress component S_{11}. It is then readily seen from (4.6), (4.7) that

$$\sigma = \hat{\sigma}(\gamma) \qquad \text{where} \qquad \hat{\sigma}(\gamma) = w'(\gamma), \quad \gamma > -1. \tag{4.8}$$

The only nonvanishing stresses in the slab are the normal stress components $S_{11}(= \sigma)$, S_{22} and S_{33} and they only depend on x and t. Therefore, the equation of motion in the absence of body forces, $\mathrm{Div}\boldsymbol{S} = \rho \dot{\boldsymbol{v}}$, leads to the single scalar equation

$$\frac{\partial \sigma}{\partial x} = \rho \frac{\partial^2 u}{\partial t^2} \tag{4.9}$$

where ρ denotes the mass density in the reference configuration.

In the formulation above, it has been assumed that the motion was sufficiently smooth. Suppose now that the motion is smooth on either side of a planar surface $x = s(t)$ but that the deformation gradient $\boldsymbol{F}$ and particle velocity $\boldsymbol{v}$ suffer jump discontinuities across this surface; we assume that the displacement itself remains continuous. In the context of this Chapter, $x = s(t)$ denotes the location of either a shock wave or a phase boundary, the distinction between them being that particles on either side of a shock wave belong to the

same phase of the material while at a phase boundary they are in different phases. In this less smooth setting, the balance of linear momentum and the continuity of the displacement field lead, in general, to the respective jump conditions

$$[\![\boldsymbol{S}\boldsymbol{n}]\!] + \rho V_n [\![\boldsymbol{v}]\!] = \boldsymbol{o}, \qquad [\![\boldsymbol{v}]\!] + V_n [\![\boldsymbol{F}\boldsymbol{n}]\!] = \boldsymbol{o}, \tag{4.10}$$

where $\boldsymbol{n}$ is a unit normal vector to the surface of discontinuity and $V_n = \boldsymbol{V} \cdot \boldsymbol{n}$ is the normal velocity of propagation of this surface. The notation used here is that for any field quantity $g(\boldsymbol{X}, t)$, $\overset{\pm}{g}$ denote its limiting values at a point on the surface of discontinuity, the limits being taken from either side of the surface, and $[\![g]\!] = \overset{+}{g} - \overset{-}{g}$ is the associated jump in g. When specialized to uniaxial motions, (4.10) reduces to

$$[\![\sigma]\!] = -\rho \dot{s} [\![v]\!], \qquad [\![v]\!] = -\dot{s} [\![\gamma]\!], \tag{4.11}$$

where we have used the fact that now $V_n = \dot{s}$. By eliminating $[\![v]\!]$ between the two equations in (4.11) and using (4.8) one obtains

$$\rho \dot{s}^2 = \frac{[\![\hat{\sigma}(\gamma)]\!]}{[\![\gamma]\!]} \tag{4.12}$$

indicating that $\rho \dot{s}^2$ equals the slope of the chord joining the points $(\overset{-}{\gamma}, \overset{-}{\sigma})$ and $(\overset{+}{\gamma}, \overset{+}{\sigma})$ on the stress-strain curve. Observe that in the special case when the stress response is linear, i.e., when $\hat{\sigma}(\gamma) = \mu\gamma$, equation (4.12) gives the propagation speed to be $\dot{s} = \pm\sqrt{\mu/\rho}$.

It is well-known in the classical theory of gas dynamics, that even though an inviscid fluid cannot dissipate energy, when the flow involves a shock wave there will in fact be dissipation localized to the shock surface. A propagating strain discontinuity in an elastic solid is similar. Thus, now consider the energetics of the slab and let $\Delta(t; \xi_1, \xi_2)$ denote the *rate of dissipation* associated with a slice of the slab between $x = \xi_1$ and $x = \xi_2$:

$$\Delta(t; \xi_1, \xi_2) = \sigma(\xi_2, t) v(\xi_2, t) - \sigma(\xi_1, t) v(\xi_1, t) \; - \; \frac{d}{dt} \int_{\xi_1}^{\xi_2} \left[w(\gamma(x,t)) + \frac{1}{2} \rho \dot{v}^2 \right] \, dx; \tag{4.13}$$

Δ represents the difference between the rate of external working and the rate at which the kinetic and stored energies increase (both per unit area in the X_2, X_3-plane). It is natural to require that the dissipation rate be non-negative, at all instants, and during all processes:

$$\Delta(t; \xi_1, \xi_2) \geq 0. \tag{4.14}$$

This *dissipation inequality* corresponds to a mechanical version of the second law of thermodynamics. When the fields are smooth, one can readily show that the dissipation rate Δ vanishes automatically, reflecting the conservative nature of an elastic material in a smooth

process. On the other hand when the motion involves a propagating strain discontinuity, this is no longer true. Suppose that there is a strain discontinuity at some location $x = s(t)$ within the slice $\xi_1 < x < \xi_2$. Then by using the field equations and jump conditions one can rewrite (4.13) and express the dissipation rate in the slab *solely* in terms of quantities *at* the strain discontinuity:

$$\Delta \;=\; [\![\sigma v]\!] + [\![w]\!]\dot{s} + [\![\frac{1}{2}\rho v^2]\!]\dot{s} \geq 0. \tag{4.15}$$

Thus if there is any dissipation of mechanical energy in the elastic slab, it can only occur at surfaces of strain discontinuity. Finally, by substituting the momentum and kinematic jump conditions (4.11) into the identity

$$\overset{+}{\sigma}\,\overset{+}{v} - \overset{-}{\sigma}\,\overset{-}{v} = \frac{1}{2}(\overset{+}{\sigma} + \overset{-}{\sigma})(\overset{+}{v} - \overset{-}{v}) + \frac{1}{2}(\overset{+}{\sigma} - \overset{-}{\sigma})(\overset{+}{v} + \overset{-}{v}) \tag{4.16}$$

one finds that

$$[\![\sigma v]\!] = \frac{1}{2}(\overset{+}{\sigma} + \overset{-}{\sigma})[\![\gamma]\!]\dot{s} + \frac{1}{2}\rho\dot{s}[\![v^2]\!]. \tag{4.17}$$

Therefore the dissipation rate (4.15) can be expressed as

$$\Delta(t) = f\dot{s} \geq 0 \tag{4.18}$$

where

$$f \;=\; w(\overset{+}{\gamma}) - w(\overset{-}{\gamma}) - \frac{1}{2}\left[\hat{\sigma}(\overset{+}{\gamma}) + \hat{\sigma}(\overset{-}{\gamma})\right](\overset{+}{\gamma} - \overset{-}{\gamma}). \tag{4.19}$$

The quantity f is called the *driving force* (per unit area) on the surface of discontinuity and is an example of Eshelby's notion of a "configurational force on a defect", (Eshelby 1956, 1970). The dissipation inequality (4.18) must hold at all discontinuities.

In summary, in a *dynamic problem* we shall seek velocity, strain and stress fields, $v(x,t), \gamma(x,t)$, $\sigma(x,t)$, which conform to given initial and boundary conditions, obey the constitutive relation $\sigma = \hat{\sigma}(\gamma) = w'(\gamma)$ and satisfy the following field equations and jump conditions:

$$\begin{array}{lllll} \dfrac{\partial\sigma}{\partial x} = \rho\,\dfrac{\partial v}{\partial t}, & \dfrac{\partial v}{\partial x} = \dfrac{\partial\gamma}{\partial t} & & & \text{where the fields are smooth,} \\[2ex] \rho\dot{s}^2 = \dfrac{\overset{+}{\sigma} - \overset{-}{\sigma}}{\overset{+}{\gamma} - \overset{-}{\gamma}}, & \overset{+}{v} - \overset{-}{v} = -\dot{s}(\overset{+}{\gamma} - \overset{-}{\gamma}), & f\dot{s} \geq 0 & & \text{at each discontinuity.} \end{array} \tag{4.20}$$

In the case of a *static problem*, we set all time derivatives in the preceding formulation to vanish and so the appropriate field equations and jump conditions controlling $u(x), \gamma(x)$ and $\sigma(x)$ are

$$\begin{array}{llll} \dfrac{d\sigma}{dx} = 0, & \gamma = \dfrac{du}{dx}, & & \text{where the fields are smooth,} \\[2ex] \overset{+}{\sigma} = \overset{-}{\sigma}, & \overset{+}{u} = \overset{-}{u} & & \text{at each discontinuity.} \end{array} \tag{4.21}$$

This completes the statement of the various field equations and jump conditions. Before turning to specific boundary-initial value problems, we need also to prescribe the constitutive relation of the material. Since our interest is in two-phase materials, the strain-energy function w must possess a two-well structure. Accordingly we shall assume that $w(\gamma)$ has *two energy-wells* located at $\gamma = 0$ and $\gamma = \gamma_T > 0$ and an intermediate local maximum at $\gamma = \gamma_*$:

$$w'(0) = w'(\gamma_*) = w'(\gamma_T) = 0; \qquad w''(0) > 0, \quad w''(\gamma_*) < 0, \quad w''(\gamma_T) > 0. \tag{4.22}$$

Further, we assume that there are exactly two points $\gamma = \gamma_{max}$ and $\gamma = \gamma_{min}$, ordered such that

$$0 < \gamma_{max} < \gamma_* < \gamma_{min} < \gamma_T, \tag{4.23}$$

at which the curvature of w changes sign:

$$\left.\begin{aligned} w''(\gamma) > 0 \quad &\text{for} \quad -1 < \gamma < \gamma_{max}, \\ w''(\gamma) < 0 \quad &\text{for} \quad \gamma_{max} < \gamma < \gamma_{min}, \\ w''(\gamma) > 0 \quad &\text{for} \quad \gamma > \gamma_{min}. \end{aligned}\right\} \tag{4.24}$$

It follows that the stress response function $\hat{\sigma}(\gamma) = w'(\gamma)$ increases for $-1 < \gamma < \gamma_{max}$, decreases for $\gamma_{max} < \gamma < \gamma_{min}$, and increases again for $\gamma > \gamma_{min}$; and that the stress vanishes at $\gamma = 0, \gamma_*$ and γ_T:

$$\left.\begin{aligned} &\hat{\sigma}(0) = \hat{\sigma}(\gamma_*) = \hat{\sigma}(\gamma_T) = 0, \\ \hat{\sigma}'(\gamma) > 0 \quad &\text{for} \quad -1 < \gamma < \gamma_{max}, \\ \hat{\sigma}'(\gamma) < 0 \quad &\text{for} \quad \gamma_{max} < \gamma < \gamma_{min}, \\ \hat{\sigma}'(\gamma) > 0 \quad &\text{for} \quad \gamma > \gamma_{min}. \end{aligned}\right\} \tag{4.25}$$

The two phases of this material are a *low-strain phase* corresponding to the strain interval $-1 < \gamma < \gamma_{max}$ and a *high-strain phase* corresponding to $\gamma > \gamma_{min}$. (We shall not refer to the low-strain phase as austenite, and the high-strain phase as martensite because our discussion is also valid, for example, if the low- and high-strain phase correspond to two variants of martensite.) Figure 4.1 shows a particular example of such functions w and $\hat{\sigma}$; in this example w has been taken to be piecewise quadratic and consequently $\hat{\sigma}$ is piecewise linear.

5. A static problem and the role of energy minimization. (Ericksen,1975) Suppose that the slab described in the preceding section is in equilibrium. The basic equations

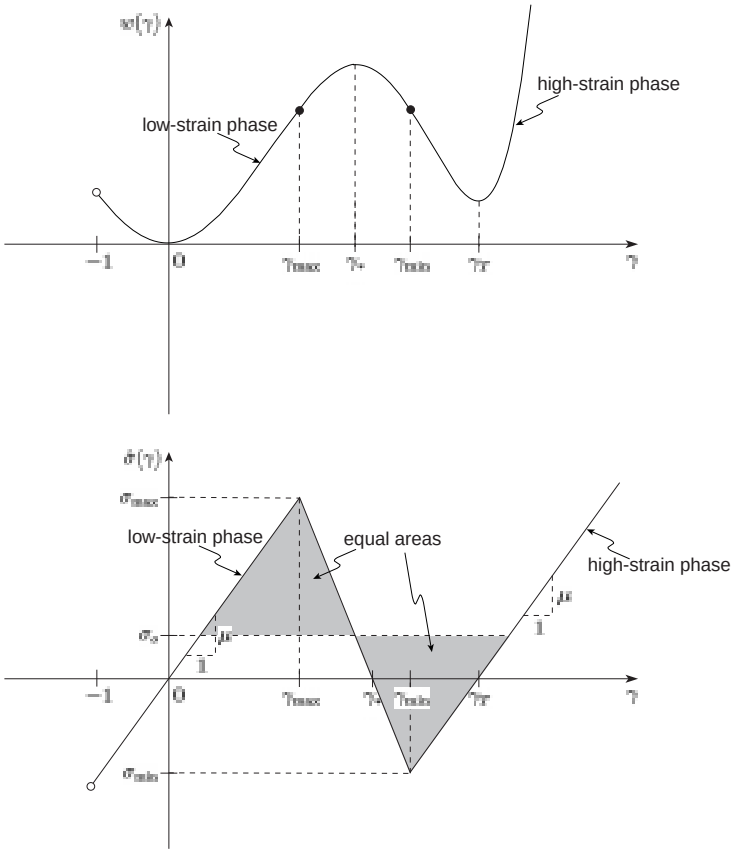

Figure 4.1: Graphs of a piecewise quadratic strain-energy function $w(\gamma)$ and the corresponding piecewise linear stress-strain relation $\sigma = \hat{\sigma}(\gamma) = w'(\gamma)$.

governing its kinematics, constitutive response and equilibrium are

$$\gamma(x) = u'(x), \quad \sigma(x) = \hat{\sigma}(\gamma(x)), \quad \sigma'(x) = 0 \tag{5.1}$$

which must hold at each $x \in (0, L)$ except at locations of strain discontinuity; if there is a strain jump at $x = s$, then the jump conditions

$$\overset{+}{\sigma} \; = \; \bar{\sigma}, \qquad \overset{+}{u} = \bar{u} \tag{5.2}$$

apply there. On integrating $(5.1)_3$ and using $(5.1)_2$, and if necessary $(5.2)_1$ as well, one finds

$$\hat{\sigma}(\gamma(x)) = \sigma, \qquad 0 < x < L, \tag{5.3}$$

where the constant σ represents the (as yet unknown) value of normal stress in the slab. Note that (5.3) must hold independently of whether or not the strain field is continuous.

Since our aim is to illustrate certain points with a minimum of mathematical complexity, it will be convenient for us to take the energy $w(\gamma)$ to be a piecewise quadratic, continuously differentiable function, so that $\hat{\sigma}(\gamma)$ is then a piecewise linear, continuous function.

Specifically, we will take

$$w(\gamma) = \begin{cases} \frac{\mu}{2}\gamma^2, & \\ \\ \frac{\mu}{2}(\gamma-\gamma_T)^2 + \sigma_o\gamma_T, & \end{cases} \qquad \hat{\sigma}(\gamma) = \begin{cases} \mu\gamma, & -1 \le \gamma \le \gamma_{max}, \\ \\ \mu(\gamma-\gamma_T), & \gamma \ge \gamma_{min}, \end{cases} \tag{5.4}$$

where the modulus μ and material parameter σ_o are positive constants. Graphs of $w(\gamma)$ and $\hat{\sigma}(\gamma)$ are shown in Figure 4.1. The maximum and minimum values of stress, σ_{max} and σ_{min}, shown in the figure correspond to

$$\sigma_{max} = \hat{\sigma}(\gamma_{max}) = \mu\gamma_{max} > 0, \qquad \sigma_{min} = \hat{\sigma}(\gamma_{min}) = \mu(\gamma_{min}-\gamma_T) < 0. \tag{5.5}$$

The declining branch of the stress-strain curve corresponds to unstable states of the material and so in statics we will require the strain to lie outside the interval $(\gamma_{max}, \gamma_{min})$; for this reason we have not displayed the formulae pertaining to the interval $\gamma_{max} < \gamma < \gamma_{min}$ in (5.4). The inequality $\sigma_{min} < 0$ is not essential; it simply ensures that w has two energy-wells rather than being merely non-convex.

Suppose that the left boundary of the slab is fixed and the right boundary is subjected to a prescribed displacement δ in the x_1-direction:

$$u(0) = 0, \qquad u(L) = \delta. \tag{5.6}$$

Given δ and the strain-energy function $w(\gamma)$, we are to determine the displacement, strain and stress fields such that the equations (5.1), jump conditions (5.2) and boundary conditions (5.6) hold. We call this the "equilibrium problem". [3]

Solution-1: Suppose first that the slab involves only the low-strain phase so that $\gamma(x) \in (-1,\ \gamma_{max})$ at every particle $x \in (0, L)$. The relevant constitutive relation is then, from $(5.4)_2$, $\sigma = \mu\gamma$. This, together with, (5.3), $(5.1)_1$ and $(5.6)_1$ leads to

$$\sigma(x) = \sigma, \qquad \gamma(x) = \frac{\sigma}{\mu}, \qquad u(x) = \frac{\sigma}{\mu}x, \qquad 0 < x < L. \tag{5.7}$$

The remaining boundary condition $(5.6)_2$ requires that the stress and elongation be related by

$$\sigma = \mu\frac{\delta}{L}. \tag{5.8}$$

Since we have assumed that the entire slab is in the low-strain phase, i.e., that $-1 < \gamma(x) < \gamma_{max}$ for $0 < x < L$, the solution (5.7), (5.8) is valid only if

$$-1 < \frac{\delta}{L} < \frac{\sigma_{max}}{\mu}. \tag{5.9}$$

[3] It would be more natural to consider a bar, rather than a slab, since that would allow us to compare the theory with experiment. Here we have chosen to work with a slab because, in contrast to a (three-dimensional) bar, the setting of the slab leads to an exact solution within the theory of finite elasticity.

Thus in summary, if the prescribed elongation lies in the interval (5.9), the stress, strain and displacement fields in the slab are given by (5.7), (5.8).

Solution-2: Next consider the case where the slab involves the high-strain phase only, so that now $\gamma(x) > \gamma_{min}$ for $0 < x < L$. The constitutive relation is thus $\sigma = \mu(\gamma - \gamma_T)$ and we now find that

$$\sigma(x) = \sigma, \qquad \gamma(x) = \frac{\sigma}{\mu} + \gamma_T, \qquad u(x) = \left(\frac{\sigma}{\mu} + \gamma_T\right) x, \qquad 0 < x < L. \tag{5.10}$$

By using the boundary condition $(5.6)_2$, the stress and elongation are found to be related by

$$\sigma = \mu\left(\frac{\delta}{L} - \gamma_T\right). \tag{5.11}$$

Finally, the assumption that the solution involves only the high-strain phase imposes the requirement

$$\frac{\delta}{L} > \frac{\sigma_{min}}{\mu} + \gamma_T. \tag{5.12}$$

Summarizing, when the prescribed elongation is sufficiently small, i.e., when $-1 < \delta/L < \sigma_{max}/\mu$, there is a configuration of the slab in which the strain is uniform and every particle is in the low-strain phase. Likewise when the elongation is sufficiently large, i.e., $\delta/L > \sigma_{min}/\mu + \gamma_T$, there is a uniform configuration associated with the high-strain phase. Note that we have *not* yet found a solution for values of elongation in the intermediate range $\sigma_{max}/\mu \le \delta/L \le \sigma_{min}/\mu + \gamma_T$. For δ in this middle interval, there is of course a solution with uniform strain that is associated with the declining branch of the stress-strain curve; but that solution is unstable. In order to find an alternative solution when δ is in this intermediate range, it is natural to consider the possibility of a *mixture* of the low- and high-strain phases.

Solution-3: For some $s \in (0, L)$, suppose that the segment $0 < x < s$ of the slab is composed of the high-strain phase while the segment $s < x < L$ is composed of the low-strain phase. The interface at $x = s$ separates these two phases and is therefore a phase boundary. The stress is again constant throughout the slab (see (5.3)): $\sigma(x) = \sigma$, $0 < x < L$. The strains are determined by making use of the relevant stress-strain relations, i.e., $\sigma = \mu\gamma$ for the low-strain segment and $\sigma = \mu(\gamma - \gamma_T)$ for the high-strain segment. The displacement field then follows by integration, keeping in mind the continuity requirement $(5.2)_2$. In this way we find

$$\gamma(x) = \begin{cases} \bar{\gamma} = \dfrac{\sigma}{\mu} + \gamma_T, & 0 < x < s, \\ \overset{+}{\gamma} = \dfrac{\sigma}{\mu}, & s < x < L. \end{cases} \qquad u(x) = \begin{cases} \bar{\gamma}\, x, & 0 < x < s, \\ \overset{+}{\gamma}\, x + (\bar{\gamma} - \overset{+}{\gamma})s, & s < x < L. \end{cases} \tag{5.13}$$

From $(5.13)_2$ and the boundary condition $(5.6)_2$ we obtain

$$\sigma = \mu\left(\frac{\delta}{L} - \frac{s}{L}\gamma_T\right) \tag{5.14}$$

which relates the stress, elongation and phase boundary location. Finally, recall that this solution was obtained under the assumption that the strains $\bar{\gamma}$ and $\overset{+}{\gamma}$ lie in the intervals $\bar{\gamma} > \gamma_{min}$, $-1 < \overset{+}{\gamma} < \gamma_{max}$, corresponding, respectively, to the high-strain and low-strain phases. It follows by using (5.13), (5.14), that this two-phase solution exists whenever

$$\frac{\sigma_{min}}{\mu} + \frac{s}{L}\gamma_T \;<\; \frac{\delta}{L} \;<\; \frac{\sigma_{max}}{\mu} + \frac{s}{L}\gamma_T \;. \tag{5.15}$$

Thus, given any $s \in (0, L)$, if the prescribed elongation δ lies in the range (5.15), then there exists a two-phase solution involving a mixture of both phases; it is given by (5.13), (5.14).

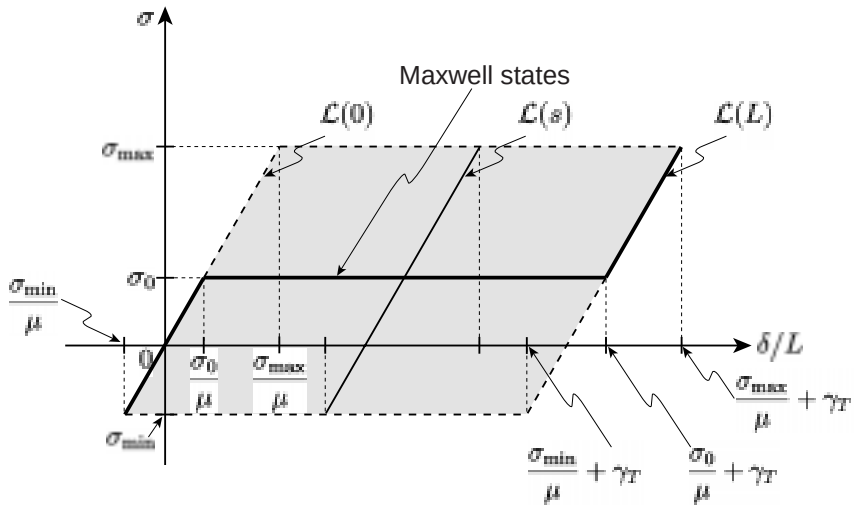

Figure 5.1: The set of all two-phase equilibria described on the δ, σ-plane. Each point in the parallelogram can be associated with a two-phase solution (5.13) - (5.15). The left and right hand boundaries of the parallelogram correspond, respectively, to the low-strain and high-strain phase solutions. Energy extremizing two-phase equilibria, i.e., solutions which describe phase equilibrium, are associated with the horizontal line labeled Maxwell states.

The arbitrariness of the phase boundary location in Solution-3 reflects a lack of uniqueness to the equilibrium problem as posed. A useful way in which to examine the nature of this nonuniqueness is to consider the δ, σ-plane. In Figure 5.1, the straight line segment $\mathcal{L}(s)$ is defined by (5.14) and (5.15), and represents the stress-elongation response of Solution-3 at some fixed value of s. Corresponding to each $s \in (0, L)$ there is a corresponding line segment $\mathcal{L}(s)$, and as s ranges over the interval $0 < s < L$ this one-parameter family of lines fill

the parallelogram shown in the figure. Each point (δ, σ) in the interior of this parallelogram can be uniquely associated with a solution of the type under consideration. The left and right boundaries of the parallelogram correspond to $s = 0$ and $s = L$ respectively. Observe by setting $s = 0$ in (5.14) and comparing it with (5.8) that the left boundary, $\mathcal{L}(0)$, in fact describes a portion of the σ, δ-curve associated with Solution-1. Similarly, the right boundary, $\mathcal{L}(L)$, describes a portion of the σ, δ-curve associated with Solution-2. It can be seen from the figure that the equilibrium problem has a two-phase solution (Solution-3) whenever the elongation δ lies in the range $\sigma_{min}/\mu < \delta/L < \sigma_{max}/\mu + \gamma_T$; moreover, for each value of δ in this interval, the problem has a one parameter family of such solutions, the (unknown) phase boundary location s being the parameter. Had we allowed for the presence of an arbitrary number of phase boundaries, the positions of each of them would have been indeterminate.

Variational approach. In an attempt to resolve this nonuniqueness, and in particular to determine the phase boundary location, we turn to an alternative notion of equilibrium, viz. equilibrium in the sense of energy extremization. The total potential energy (per unit area in the x_2, x_3-plane) is

$$E(u) = \int_0^L w(u'(x))dx, \qquad u \in \mathcal{A}, \tag{5.16}$$

where the set $\mathcal{A}$ of kinematically admissible displacement fields consists of all continuous, piecewise smooth functions $u : [0, L] \to I\!R$ which satisfy the boundary conditions (5.6):

$$\mathcal{A} = \{u \ : \ u \in C_p^2(O, L) \cap C(0, L), \ u(0) = 0, \ u(L) = \delta \ \}. \tag{5.17}$$

Suppose that E is minimized by a particular function $u \in \mathcal{A}$, and suppose that u' is discontinuous at a location $x = s$. Consider a one-parameter family of admissible displacement fields $\bar{u}(x, \varepsilon)$, $-\varepsilon_o < \varepsilon < \varepsilon_o$, which has the property that $\bar{u}(x, 0) = u(x)$. Suppose that for each ε, the first derivative of $\bar{u}(\cdot, \varepsilon)$ is discontinuous at the location $x = \bar{s}(\varepsilon)$ and that $\bar{s}(0) = s$. Then, by setting

$$\frac{d}{d\varepsilon} \ E(\bar{u}(x, \varepsilon))\bigg|_{\varepsilon=0} = \frac{d}{d\varepsilon} \ \int_0^{\bar{s}(\varepsilon)} w(\bar{u}'(x, \varepsilon))dx\bigg|_{\varepsilon=0} + \frac{d}{d\varepsilon} \ \int_{\bar{s}(\varepsilon)}^L w(\bar{u}'(x, \varepsilon))dx\bigg|_{\varepsilon=0} = 0 \tag{5.18}$$

one finds in the usual way that it is necessary that u satisfy the equilibrium equation

$$\frac{d}{dx}\hat{\sigma}(u'(x)) = 0, \qquad x \neq s, \ 0 < x < L. \tag{5.19}$$

In addition, the arbitrariness of $\bar{u}_\varepsilon(s-, 0)$ requires that u' satisfy the jump condition

$$[\![\hat{\sigma}(u')]\!] = 0 \qquad \text{at} \quad x = s; \tag{5.20}$$

and finally, the arbitrariness of $\bar{s}'(0)$ requires that u' satisfy a second jump condition

$$[\![w(u') - \hat{\sigma}(u')u']\!] = 0 \qquad \text{at} \quad x = s. \tag{5.21}$$

This pair of jump conditions are the so-called Weierstrass-Erdmann corner conditions of the calculus of variations (see, for example, Gelfand and Fomin, 1963). The first of these is precisely the stress continuity condition $(5.2)_1$, whereas the second condition is equivalent to the vanishing of the driving force on the phase boundary:

$$f = 0 \quad \text{where} \quad f = w(\overset{+}{\gamma}) - w(\overset{-}{\gamma}) - \sigma(\overset{+}{\gamma} - \overset{-}{\gamma}); \tag{5.22}$$

cf. equation (4.19).

Thus the energy extremization notion of equilibrium is more stringent than the force balance point of view in that it requires all of the same conditions *plus* an extra condition. This additional requirement of vanishing driving force is often referred to as a characterization of "phase equilibrium". It can be written, after using $(4.8)_2$ as

$$\int_{\overset{-}{\gamma}}^{\overset{+}{\gamma}} \hat{\sigma}(\gamma)\, d\gamma = \sigma(\overset{+}{\gamma} - \overset{-}{\gamma}). \tag{5.23}$$

Since $\sigma = \hat{\sigma}(\overset{+}{\gamma}) = \hat{\sigma}(\overset{-}{\gamma})$, this can be described geometrically as saying that the stress σ cuts off equal areas of the stress-strain curve; see Figure 4.1. The special value of stress for which this is true is called the *Maxwell stress*.

In the case of the piecewise quadratic energy (5.4), substituting $(5.13)_1$ and (5.14) into (5.23) shows that the Maxwell stress equals σ_o, where σ_o is the material constant introduced earlier in (5.4). Thus, if a two-phase configuration is to extremize the energy, the stress must equal the Maxwell stress throughout the slab: $\sigma(x) = \sigma_o,\ 0 < x < L$; the strain and displacement fields are found by replacing σ in (5.13) by σ_o:

$$u(x) = \begin{cases} \overset{-}{\gamma}\, x, & 0 < x < s, \\ \overset{+}{\gamma}\, x + (\overset{-}{\gamma} - \overset{+}{\gamma})s, & s < x < L, \end{cases} \qquad \text{where} \quad \overset{-}{\gamma} = \frac{\sigma_o}{\mu} + \gamma_T,\ \overset{+}{\gamma} = \frac{\sigma_o}{\mu}. \tag{5.24}$$

Setting $\sigma = \sigma_o$ in (5.14) allows us to solve for the previously unknown location s of the phase boundary:

$$s = \left(\frac{\delta}{L} - \frac{\sigma_o}{\mu} \right) \frac{L}{\gamma_T}. \tag{5.25}$$

Finally, the inequality (5.15) defining the range of values of δ for which a two-phase solution exists now specializes to

$$\frac{\sigma_o}{\mu} \ < \ \frac{\delta}{L} \ < \ \frac{\sigma_o}{\mu} + \gamma_T \ . \tag{5.26}$$

From among all points in the parallelogram in Figure 5.1, the energy-extremizing two-phase solutions are associated with points on the horizontal line $\sigma = \sigma_o$ labeled "Maxwell states".

Uniqueness. The preceding discussion concerned Solution-3. Collecting all of the solutions at hand, we now have a low-strain phase solution (Solution-1), a high-strain phase solution (Solution-2), and a two-phase (Maxwell) solution (Solution-3 with $\sigma = \sigma_o$) associated with values of δ in the respective ranges (5.9), (5.12) and (5.26). Observe from this, and perhaps more clearly from Figure 5.1, that the equilibrium problem now has a unique solution if the elongation lies in the interval $(-1,\ \sigma_o/\mu] \cup [\sigma_{max}/\mu,\ \sigma_{min}/\mu + \gamma_T] \cup [\sigma_o/\mu + \gamma_T,\ \infty)$.

However for each $\delta \in (\sigma_o/\mu,\ \sigma_{max}/\mu)$ we have two solutions, a low-strain phase solution and a two-phase solution (corresponding to parts of the lines $\mathcal{L}(0)$ and $\sigma = \sigma_o$ in Figure 5.1); and similarly associated with each $\delta \in (\sigma_{min}/\mu + \gamma_T,\ \sigma_o/\mu + \gamma_T)$ we have a high-strain phase solution and a two-phase solution (corresponding to $\mathcal{L}(L)$ and $\sigma = \sigma_o$). In order to distinguish between these solutions let us compare their energies. Evaluating (5.16) at each of the three solutions (5.7), (5.10) and (5.24) leads to the corresponding energies

$$E_1 = \frac{\mu}{2}\left(\frac{\delta}{L}\right)^2, \qquad E_2 = \frac{\mu}{2}\left(\frac{\delta}{L} - \gamma_T\right)^2 + \sigma_o\gamma_T, \qquad E_3 = -\frac{\sigma_o^2}{2\mu} + \sigma_o\frac{\delta}{L}, \tag{5.27}$$

respectively. It can be readily verified that

$$E_3 - E_1 = -\frac{\mu}{2}\left(\frac{\delta}{L} - \frac{\sigma_o}{\mu}\right)^2 \leq 0, \quad \text{and} \quad E_3 - E_2 = -\frac{\mu}{2}\left(\frac{\delta}{L} - \gamma_T - \frac{\sigma_o}{\mu}\right)^2 \leq 0. \tag{5.28}$$

Therefore, whenever a two-phase solution and a single phase solution exist at the same value of δ, the two-phase solution has less energy.

Consequently from among all equilibrium solutions, the one with the least energy for the various ranges of δ is:

$$\begin{array}{rcll} -1 & < \delta/L & \leq \sigma_o/\mu & \text{low} - \text{strain phase solution}, \\ \sigma_o/\mu & < \delta/L & < \sigma_o/\mu + \gamma_T & \text{two} - \text{phase (Maxwell) solution}, \\ \sigma_o/\mu + \gamma_T & \leq \delta/L & & \text{high} - \text{strain phase solution}, \end{array} \tag{5.29}$$

From (5.8), (5.11) and $\sigma = \sigma_o$, the corresponding relationships between stress σ and elongation δ are

$$\sigma = \begin{cases} \mu\dfrac{\delta}{L} & \text{for} \quad -1 < \delta/L \leq \sigma_o/\mu, \\ \sigma_0 & \text{for} \quad \sigma_o/\mu \leq \delta/L \leq \sigma_o/\mu + \gamma_T, \\ \mu\left(\dfrac{\delta}{L} - \gamma_T\right) & \text{for} \quad \delta/L \geq \sigma_o/\mu + \gamma_T. \end{cases} \tag{5.30}$$

The σ, δ-relation (5.30) is depicted in Figure 5.1 by the bold portions of $\mathcal{L}(0)$ and $\mathcal{L}(L)$ and the horizontal line $\sigma = \sigma_o$. It should be emphasized that this graph simply depicts the set of energy minimizing equilibrium solutions; as we shall discuss later, it does *not* imply that this is the curve that is traversed during a loading-unloading process.

Potential energy. Before concluding our discussion of static equilibrium it is worth making some brief observations about the *potential energy function* of the material. Recall that the local minima of the strain-energy function $w(\gamma)$ inform us about the various *stress-free* phases of the material and their relative stability. The *potential energy function* $G(\gamma, \sigma)$ plays a similar role when the stress does not vanish. In particular, given the stress σ, different phases of the material at that value of stress will be associated with different local minima of $G(\cdot, \sigma)$. It is worth pointing out that the energy E that was considered in the preceding discussion was the total (macroscopic) energy of the slab; here we are concerned with a local energy associated with the material.

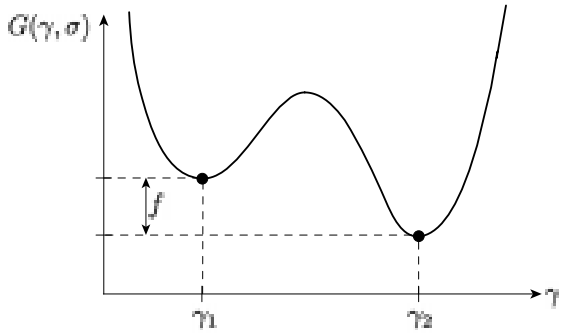

Figure 5.2: The potential energy G as a function of strain at a fixed value of stress. When the stress lies in the range $\sigma_{min} < \sigma < \sigma_{max}$, $G(\cdot, \sigma)$ has two local minima. The minimum on the left is the global minimum when $\sigma_{min} < \sigma < \sigma_o$ while the one on the right is the global minimum for $\sigma_o < \sigma < \sigma_{max}$; here σ_o is the Maxwell stress.

The potential energy G is defined by

$$G(\gamma, \sigma) = w(\gamma) - \sigma\gamma \tag{5.31}$$

where σ and γ are treated as independent variables, not necessarily related through the stress-strain relation. Observe that

$$\frac{\partial G}{\partial \gamma} = 0 \quad \Leftrightarrow \quad \sigma = w'(\gamma) = \hat{\sigma}(\gamma), \tag{5.32}$$

and that

$$\frac{\partial^2 G}{\partial \gamma^2} = w''(\gamma) = \hat{\sigma}'(\gamma). \tag{5.33}$$

Consequently given σ, a strain γ corresponds to a local minimum of the potential energy $G(\cdot, \sigma)$ if $\sigma = \hat{\sigma}(\gamma)$ and $\hat{\sigma}'(\gamma) > 0$; it corresponds to a local maximum if $\sigma = \hat{\sigma}(\gamma)$ and

$\hat{\sigma}'(\gamma) < 0$. Thus any point on any rising branch of the stress-strain curve can be associated with a local minimum of the potential energy function. If the stress-strain curve has the rising, falling, rising form shown in Figure 4.1, then G has multiple local minima for certain ranges of σ.

Consider for example the material characterized by (5.4) and described in Figure 4.1. For this material, when $\sigma < \sigma_{min}$ the function $G(\cdot, \sigma)$ has a single minimum and it is located in the low-strain phase at $\gamma = \gamma_1 = \sigma/\mu$; for $\sigma > \sigma_{max}$, $G(\cdot, \sigma)$ again has a single minimum but now it is located in the high-strain phase at $\gamma = \gamma_2 = \sigma/\mu + \gamma_T$. On the other hand when $\sigma_{min} < \sigma < \sigma_{max}$ the potential energy $G(\cdot, \sigma)$ has local minima at both γ_1 and γ_2. The height $G(\gamma_1, \sigma) - G(\gamma_2, \sigma)$ between these minima is

$$G(\gamma_1, \sigma) - G(\gamma_2, \sigma) = w(\gamma_1) - w(\gamma_2) - \sigma(\gamma_1 - \gamma_2). \tag{5.34}$$

Even though our discussion here has not concerned a phase boundary, (5.34) is precisely what would be the driving force if there was a a phase boundary separating the strain γ_1 from the strain γ_2, cf. $(5.22)_2$. In terms of stress, one can write this difference in height after using (5.4) as

$$f = (\sigma - \sigma_o)\gamma_T. \tag{5.35}$$

It follows from (5.34) and (5.35) that the minimum at γ_1 is lower than the minimum at γ_2 when $f < 0$ (i.e., $\sigma_{min} < \sigma < \sigma_o$), and vice versa; both minima have the same height when $f = 0$ (i.e., $\sigma = \sigma_o$). Suppose that a particle is always associated with the global minimum of G. In this case, for $\sigma < \sigma_o$ the particle will be in the low-strain phase, for $\sigma > \sigma_o$ it will be in the high-strain phase, and at $\sigma = \sigma_o$ it could be associated with either phase. Returning from this local view to the global setting for the slab, e.g., (5.30), we observe that in any configuration of the slab which minimizes the total energy E, the individual particles are associated with states that minimize the potential energy G.

6. A dynamic problem and the role of kinetics and nucleation. (Abeyaratne and Knowles, 1991a, 1993b, 2000b) Thus far we have restricted attention to statics and consequently, even though we have encountered both phases of the two-phase material, we have not considered the *transformation* from one into the other. The process of phase transformation is inherently time-dependent, and in order to examine it we now turn to a dynamic problem for a slab. From Section 4 we recall the basic differential equations

$$\frac{\partial}{\partial x}\hat{\sigma}(\gamma) = \rho\,\frac{\partial v}{\partial t}, \qquad \frac{\partial v}{\partial x} = \frac{\partial \gamma}{\partial t}, \tag{6.1}$$

which must hold wherever the fields are smooth. If the strain and particle velocity suffer jump discontinuities at $x = s(t)$, the jump conditions

$$\rho\dot{s}^2 = \frac{\hat{\sigma}(\overset{+}{\gamma}) - \hat{\sigma}(\overset{-}{\gamma})}{\overset{+}{\gamma} - \overset{-}{\gamma}}, \qquad \overset{+}{v} - \overset{-}{v} = -\dot{s}(\overset{+}{\gamma} - \overset{-}{\gamma}), \tag{6.2}$$

and the dissipation inequality,

$$f\dot{s} \geq 0 \qquad \text{where} \qquad f \;=\; w(\overset{+}{\gamma}) - w(\overset{-}{\gamma}) \;-\frac{1}{2}\left[\hat{\sigma}(\overset{+}{\gamma}) + \hat{\sigma}(\overset{-}{\gamma})\right](\overset{+}{\gamma} - \overset{-}{\gamma}), \tag{6.3}$$

must hold there.

Suppose that the slab is semi-infinite ($L = \infty$) and that it is initially undeformed and at rest:

$$\gamma(x,0) = 0, \quad v(x,0) = 0 \qquad \text{for } x > 0; \tag{6.4}$$

thus the slab is composed of the unstressed low-strain phase at time $t = 0$. Finally, suppose that the boundary $x = 0$ is subjected to a constant normal velocity $-\dot{\delta}$ for all positive time:

$$v(0,t) = -\dot{\delta} \qquad \text{for } t > 0 \qquad \text{where } \dot{\delta} > 0. \tag{6.5}$$

The problem at hand involves finding $\gamma(x,t), v(x,t)$ by solving the differential equations (6.1) subject to the jump conditions (6.2), the dissipation inequality (6.3), the initial conditions (6.4) and the boundary condition (6.5). Since we take $\dot{\delta} > 0$ the slab is being subjected to a sudden elongational loading. We shall refer to this as the "dynamic problem".

Several general results pertaining to the Cauchy problem associated with quasilinear systems of equations of the form (6.1) are known. In particular, if the stress response function $\hat{\sigma}(\gamma)$ is monotonically increasing, and in addition it is *either* strictly convex or strictly concave, then it is known that the Cauchy problem has a unique (weak) solution (Oleinik,1957). In the present setting $\hat{\sigma}(\gamma)$ is characterized by (4.22) - (4.25) or more specifically by (5.4). It is neither monotonic nor does it have the convexity property required for this uniqueness result to hold. Thus we expect that the solution to the dynamic problem may be non-unique. Nevertheless, it is worth attempting to solve this problem as posed above, with the aim of finding *all* of its solutions. The hope would then be that some characteristic of the totality of solutions might reveal the (physical) cause of the non-uniqueness and suggest a remedy.

Before attempting to solve the dynamic problem it is useful to make some preliminary observations pertaining to the character of propagating discontinuities. In contrast to the setting of statics, here a strain discontinuity can correspond to either a phase boundary or a shock wave. If the material on both sides of the discontinuity belong to the same phase then the discontinuity is a *shock wave*. If $\overset{-}{\gamma}$ and $\overset{+}{\gamma}$ are both associated with the low-strain phase, then from (5.4) we have $\overset{\pm}{\sigma} = \mu\overset{\pm}{\gamma}$, whereas if they are both associated with the high-strain phase then $\overset{\pm}{\sigma} = \mu(\overset{\pm}{\gamma} - \gamma_T)$. In either case, the propagation speed of the shock wave is given by $(6.2)_1$ which now simplifies to

$$\dot{s} = \pm c, \qquad c = \sqrt{\frac{\mu}{\rho}}. \tag{6.6}$$

Thus (for the material at hand), shock waves propagate at a speed c that is independent of the strains on either side of it; this degeneracy is special to the piecewise linear stress-strain relation. Next, if the strains on the two sides of the discontinuity belong to distinct phases, the discontinuity is a *phase boundary*. For example, suppose that $\bar{\gamma}$ is associated with the high-strain phase and $\overset{+}{\gamma}$ is associated with the low-strain phase; then $\bar{\sigma} = \mu(\bar{\gamma} - \gamma_T)$ and $\overset{+}{\sigma} = \mu\overset{+}{\gamma}$. Substituting these into $(6.2)_1$ allows us to express the propagation speed as

$$\frac{\dot{s}^2}{c^2} \;=\; 1 \;-\; \frac{\gamma_T}{\bar{\gamma} - \overset{+}{\gamma}}. \tag{6.7}$$

Note that $\bar{\gamma} > \overset{+}{\gamma}$ since $\bar{\gamma}$ is in the high-strain and $\overset{+}{\gamma}$ is in the low-strain phase. Thus, according to (6.7), a phase boundary propagates at a subsonic speed $|\dot{s}| < c$.

Next consider the driving force f on a discontinuity. By using the constitutive relation (5.4) and evaluating $(6.3)_2$, we find that $f = 0$ at a shock wave; again this degeneracy is peculiar to a material with a piecewise linear stress-strain relation. On the other hand for a phase boundary where $\bar{\gamma}$ is associated with the high-strain phase and $\overset{+}{\gamma}$ is associated with the low-strain phase, $(6.3)_2$ yields

$$f \;=\; \left(\frac{\overset{+}{\sigma} + \bar{\sigma}}{2} - \sigma_o\right)\gamma_T \;=\; \left(\mu\,\frac{\overset{+}{\gamma} + \bar{\gamma} - \gamma_T}{2} - \sigma_o\right)\gamma_T\,. \tag{6.8}$$

Note that (6.8) specializes to (5.35) when $\overset{+}{\sigma} = \bar{\sigma} = \sigma$.

We now return to the dynamic problem. Since there is no length scale in the problem it is natural to seek a scale-invariant solution of the form

$$\gamma(x,t) = \gamma\,(x/t)\,, \qquad v(x,t) = v\,(x/t)\,. \tag{6.9}$$

Substituting this assumed solution form into the differential equations (6.1) and using the constitutive relation (5.4) shows that $\gamma(x,t)$ and $v(x,t)$ must necessarily be constant on regions of the x,t-plane where γ and v are smooth. Thus, on the x,t-plane, the solution will involve various curves corresponding to propagating shock waves and phase boundaries, and the strain and particle velocity fields will suffer jump discontinuities across these curves but will otherwise remain constant. Necessarily therefore the states on either side of a particular discontinuity are constant, and consequently by $(6.2)_1$ its propagation speed must remain constant as well. The various discontinuities are therefore described on the x,t-plane by straight lines.

Consequently the solution of the dynamic problem must have the following general structure (Figure 6.1): there is a family of n rays, $x = c_i t,\ i = 1, 2, \ldots n$, which separate the first

quadrant of the x, t-plane into $n+1$ wedges. Between every pair of rays the strain and particle velocity are constant:

$$\gamma(x,t) = \gamma_j, \quad v(x,t) = v_j \qquad \text{for} \qquad c_j t < x < c_{j+1} t, \ j = 0, 1, 2, \ldots n. \tag{6.10}$$

Each ray corresponds to either a shock wave or a phase boundary and $0 = c_0 < c_1 < c_2 \ldots < c_n < c_{n+1} = \infty$. Determining the solution thus reduces to finding all of the constants γ_k, v_k and c_k.

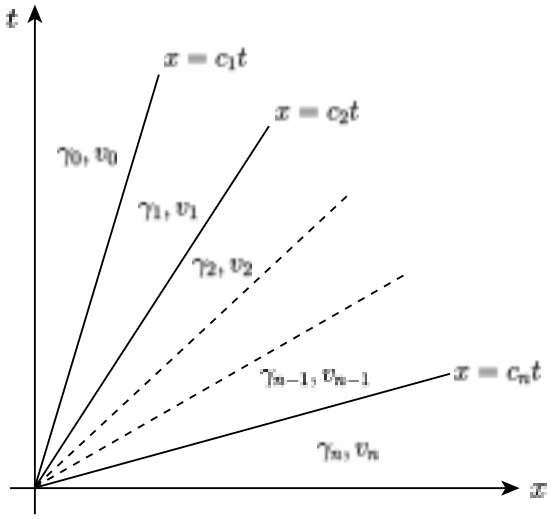

Figure 6.1: General form of solution to sudden elongation problem. Piecewise homogeneous states (γ_k, v_k) separated by shock waves and phase boundaries propagating at speeds c_k.

By adapting the arguments made in Abeyaratne and Knowles (1991a) for a Riemann problem, one can establish the following useful results pertaining to the solution (6.10):

1. No strain γ_j is associated with the declining branch of the stress-strain curve.

2. The solution involves at most one shock wave.

3. The solution involves at most one phase boundary.

Thus in particular, a solution can exhibit at most two rays bearing discontinuities. The argument used by Abeyaratne and Knowles (1991a) in obtaining these conclusions depends heavily on the dissipation inequality. It follows from these results that the solution must necessarily have one of the two forms shown in Figure 6.2: Solution-A involves no phase boundaries and therefore no change of phase; in contrast Solution-B does involve a change of phase.

Solution-A (Figure 6.2(a)): This solution involves a single discontinuity which is a shock wave propagating at the speed $c = \sqrt{\mu/\rho}$ and has the form

$$\gamma(x,t),\ v(x,t) \ = \ \begin{cases} 0,\ 0 & \text{for} \quad x > ct, \qquad (\text{low} - \text{strain phase}) \\ \overset{+}{\gamma},\ \overset{+}{v} & \text{for} \quad 0 < x < ct, \qquad (\text{low} - \text{strain phase}) \end{cases} \tag{6.11}$$

where we have made use of the initial conditions to determine the state ahead of the shock. Since the slab is initially in the low-strain phase, it must necessarily remain in the low-strain phase for all time since there are no phase boundaries in this solution. By enforcing the jump condition $(6.2)_2$ and the boundary condition (6.5) we find the state behind the shock to be

$$\overset{+}{\gamma} = \frac{\dot{\delta}}{c}, \qquad \overset{+}{v} = -\dot{\delta}. \tag{6.12}$$

Since the driving force on the shock wave vanishes, the dissipation inequality imposes no restrictions. Finally, since the slab is, by assumption, entirely in the low-strain state, we must have $-1 < \overset{+}{\gamma} < \gamma_{max}$. This requires the elongation-rate to be sufficiently small:

$$\frac{\dot{\delta}}{c} \ < \ \frac{\sigma_{max}}{\mu}. \tag{6.13}$$

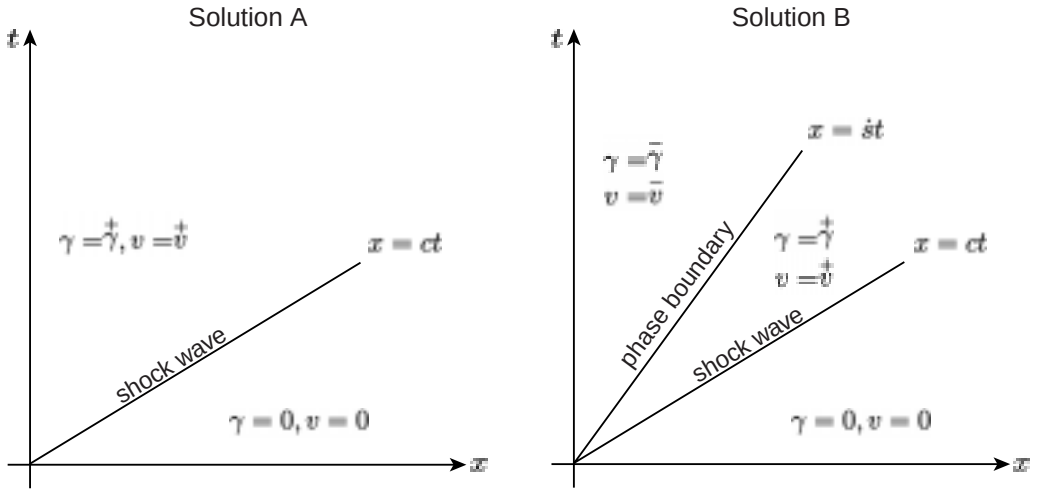

Figure 6.2: Solution forms with and without phase transition.

Solution-B (Figure 6.2(b)): Here the solution involves a front running shock wave at $x = ct$ behind which there is a propagating phase boundary $x = \dot{s}t$. The material ahead of the shock wave is in the initial state (i.e., the undeformed low-strain phase at rest); the material between the shock wave and the phase boundary is also in the low-strain phase but

it involves a nonvanishing strain and particle velocity; finally, the material behind the phase boundary is in the high-strain phase. Thus the solution has the form

$$\gamma(x,t),\ v(x,t) \ = \ \begin{cases} 0,\ 0 & \text{for} \quad x > ct, & (\text{low} - \text{strain phase}) \\ \overset{+}{\gamma},\ \overset{+}{v} & \text{for} \quad \dot{s}t < x < ct, & (\text{low} - \text{strain phase}) \\ \overset{-}{\gamma},\ \overset{-}{v} & \text{for} \quad 0 < x < \dot{s}t, & (\text{high} - \text{strain phase}) \end{cases} \tag{6.14}$$

where, again, we have already made use of the initial conditions. Applying the jump conditions (6.2) at both the shock wave and the phase boundary, and using the boundary condition (6.5) determines the unknown strains and particle velocities in terms of the phase boundary speed $\dot{s}$:

$$\overset{+}{\gamma} = \frac{\dot{\delta}}{c} - \frac{c\dot{s}\gamma_T}{c^2 - \dot{s}^2}, \qquad \overset{-}{\gamma} = \frac{\dot{\delta}}{c} + \frac{c\gamma_T}{c + \dot{s}}, \qquad \overset{+}{v} = -c\overset{+}{\gamma}, \qquad \overset{-}{v} = -\dot{\delta}. \tag{6.15}$$

Since the material ahead of the phase boundary is in the low-strain phase while the material behind it is in the high-strain phase, we must have $-1 < \overset{+}{\gamma} < \gamma_{max}$ and $\overset{-}{\gamma} > \gamma_{min}$; therefore the elongation-rate must necessarily lie in the range

$$\frac{\sigma_{min}}{\mu} + \gamma_T \frac{\dot{s}}{c + \dot{s}} \ < \frac{\dot{\delta}}{c} \ < \ \frac{\sigma_{max}}{\mu} + \gamma_T \frac{c\dot{s}}{c^2 - \dot{s}^2} \ . \tag{6.16}$$

Finally consider the dissipation inequality $(6.3)_1$. It is automatically satisfied at the shock wave since the driving force there vanishes identically. On the other hand at the phase boundary, since $\dot{s} > 0$ we must demand $f \geq 0$. By using (6.8) and (6.15) we find that the driving force is given by

$$f \ = \ \mu\gamma_T \left\{ \frac{\dot{\delta}}{c} \ - \ \frac{\sigma_o}{\mu} \ - \ \frac{\gamma_T}{2} \frac{\dot{s}(2c - \dot{s})}{c^2 - \dot{s}^2} \right\}. \tag{6.17}$$

The dissipation inequality therefore requires that

$$\frac{\dot{\delta}}{c} \ \geq \ \frac{\sigma_o}{\mu} \ + \ \frac{\gamma_T}{2} \frac{\dot{s}(2c - \dot{s})}{c^2 - \dot{s}^2}. \tag{6.18}$$

In summary, for any arbitrarily chosen phase boundary speed $\dot{s} \in (0, c)$, if the elongation-rate $\dot{\delta} > 0$ satisfies the inequalities (6.16) and (6.18), an admissible solution to the dynamic problem is given by (6.14), (6.15). Observe that the phase boundary propagation speed $\dot{s}$ remains indeterminate. Thus, when a phase changing solution exists, there is in fact a *one-parameter family of such solutions* (parameter $\dot{s}$).

Nature of nonuniqueness. Need for nucleation and kinetics. In order to delineate the precise nature of this non-uniqueness, it is convenient to consider the $\dot{\delta}, \dot{s}$-plane. The curves

labeled $\overset{+}{\gamma} = \gamma_{max}, \overset{-}{\gamma} = \gamma_{min}$ and $f = 0$ in Figure 6.3 are defined by (6.16) and (6.18) with the inequalities replaced by equalities. Each point $(\dot{\delta}, \dot{s})$ in the shaded region $\mathcal{B}$ satisfies all three inequalities in (6.16), (6.18). Each such point can therefore be associated uniquely with a solution involving a phase change (Solution-B). The bold segment $\mathcal{A}$ of the $\dot{\delta}$-axis corresponds to the interval (6.13) and delineates points which can be associated with a solution involving no phase change (Solution-A).

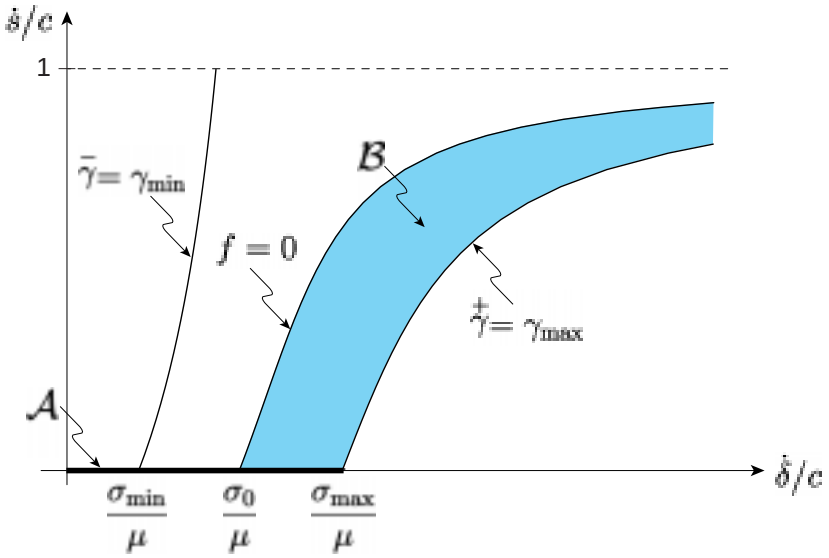

Figure 6.3: The set of all solutions to the dynamic problem displayed on the $\dot{\delta}, \sigma$-plane. Each point in the shaded region $\mathcal{B}$ can be uniquely associated with a Solution-B which involves a phase change. Points on the line segment $\mathcal{A}$ can be associated with Solutions-A which do not involve a phase change. Observe the severe nonuniqueness of solution when the prescribed elongation-rate $\dot{\delta}/c$ exceeds σ_o/μ.

Figure 6.3 allows us to reach the following conclusions about the totality of (scale-invariant) solutions:

1. If the elongation-rate is sufficiently small, i.e., if $0 < \dot{\delta}/c \leq \sigma_o/\mu$, there is a unique solution (Solution-A) and it involves no phase change.

2. If the elongation-rate is sufficiently large, i.e., if $\dot{\delta}/c \geq \sigma_{max}/\mu$, then corresponding to each prescribed $\dot{\delta}$ there is a one-parameter family of solutions, each of which involves a phase change (Solution-B); the parameter is the propagation speed $\dot{s}$. There are no solutions which do not involve a phase change.

3. Finally if the elongation-rate has a value in the intermediate range $\sigma_o/\mu < \dot{\delta}/c < \sigma_{max}/\mu$, then, corresponding to each prescribed $\dot{\delta}$, there exists a one parameter family

of solutions involving a phase change (Solution-B) *as well as* a solution that does not involve a phase change (Solution-A).

The lack of uniqueness in a mathematical model of a physical problem usually reflects an incompleteness in the modeling. Observe that in the present context we have *two distinct types of nonuniqueness*: first, when a phase transforming solution exists, we need to determine the unknown propagation speed $\dot{s}$ of the phase boundary; and second, when both types of solutions exist at the same $\dot{\delta}$, we need to choose between them. The former reflects a need for information describing the propagation speed, which physically is equivalent to a description of the *rate* at which the material transforms from one phase to the other; in the materials science literature this is described by a "kinetic (or evolution) law". The second type of non-uniqueness reflects a need for information about *when* the transformation process commences, i.e., a characterization of when a new phase first "nucleates". The suggestion therefore is that the dynamic problem as posed previously should be supplemented with two additional conditions: a nucleation condition and a kinetic law.

Taking a purely continuum mechanical approach to this, it is natural to suppose that the propagation of a phase boundary is governed by the state of the material on either side of it, i.e., by a relation of the form

$$\dot{s} = V(\overset{+}{\gamma}, \overset{-}{\gamma}). \tag{6.19}$$

The function V is determined by the material and it provides the continuum theory with an appropriate characterization of the lattice-scale dynamics underlying the transformation process. The pair of equations (6.7) and (6.8) can be solved to determine the two strains $\overset{+}{\gamma}, \overset{-}{\gamma}$ in terms of $\dot{s}$ and f. Substituting the result into (6.19) leads to an equation of the form $H(\dot{s}, f) = 0$. Assuming solvability for f, the propagation law (6.19) can therefore be written in the equivalent form

$$f = \varphi(\dot{s}) \tag{6.20}$$

relating the propagation speed and the driving force. Equation (6.20) provides the simplest example of a *kinetic law*. In order to be consistent with the dissipation inequality $f\dot{s} \geq 0$ the function φ must have the property

$$\dot{s}\varphi(\dot{s}) \geq 0 \tag{6.21}$$

but is otherwise unrestricted by the continuum theory. It is part of the characterization of the material and needs to be determined through a combination of lattice-scale modeling and laboratory experiments.

When (6.20) is enforced at the phase boundary in our dynamic problem, we find by using

(6.17) that the elongation-rate and the propagation speed are related by

$$\frac{\dot{\delta}}{c} \;=\; \frac{\sigma_o}{\mu} \;+\; \frac{\gamma_T}{2}\frac{\dot{s}(2c-\dot{s})}{c^2-\dot{s}^2} \;+\; \frac{\varphi(\dot{s})}{\mu\gamma_T}. \tag{6.22}$$

This describes a curve $\mathcal{K}$ in the $\dot{\delta}, \dot{s}$-plane which lies in the shaded region $\mathcal{B}$. If φ increases monotonically, $\mathcal{K}$ is a monotonically rising curve as shown in Figure 6.4. Thus from among all Solutions-B the only ones which are permitted by the kinetic law are those associated with the curve $\mathcal{K}$; they are characterized by (6.14), (6.15) and (6.22).

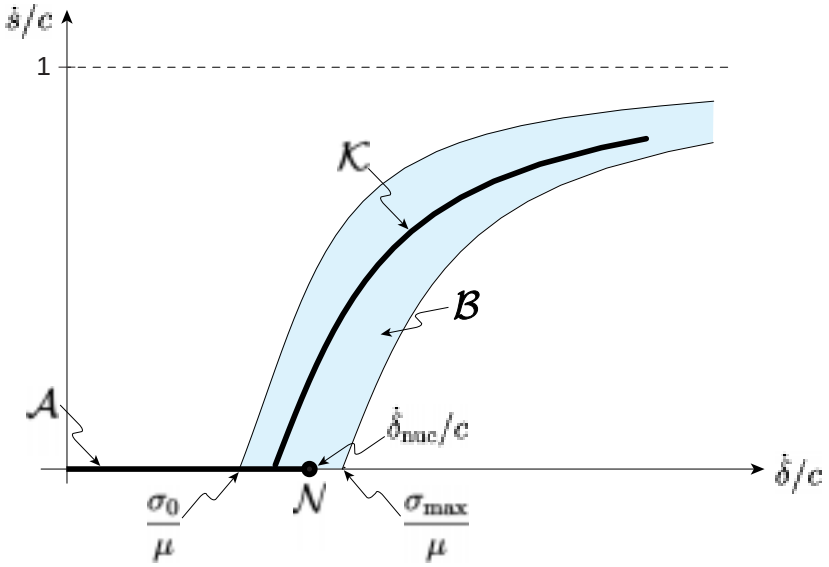

Figure 6.4: Solutions to the dynamic problem which conform to the kinetic law and nucleation condition. Only points on the curve $\mathcal{K}$ can be associated two-phase solutions which conform to the kinetic law. The point $\mathcal{N}$ is determined by the nucleation condition and represents the boundary between phase change and no phase solutions: if the elongation-rate $\dot{\delta} < \dot{\delta}(\mathcal{N})$ the slab does not change phase and responds according to Solution-A; if $\dot{\delta} < \dot{\delta}(\mathcal{N})$ the slab changes phase and its response is described by Solution-B together with the kinetic law.

In the presence of the kinetic law, the totality of solutions has been reduced to those associated with the two curves $\mathcal{A}$ and $\mathcal{K}$ of the $\dot{\delta}, \dot{s}$-plane. Thus the remaining nonuniqueness pertains to selecting between them whenever they both exist at the same $\dot{\delta}$, i.e., when the elongation-rate lies in the range $\sigma_o/\mu + \varphi(0)/\mu\gamma_T < \dot{\delta}/c < \sigma_{max}/\mu$; see Figures 6.3 and 6.4. In order to make this determination, consider the following *nucleation condition*: a phase boundary will be nucleated provided that the driving force on it soon after it nucleates exceeds some critical value $f_{nuc}(> 0)$. We can apply this to the dynamic problem by calculating $\dot{s}_{nuc}$ from $f_{nuc} = \varphi(\dot{s}_{nuc})$ and then calculating $\dot{\delta}_{nuc}$ from (6.22). This determines the threshold value of the elongation-rate for nucleation, which is depicted by the point $\mathcal{N}$

in Figure 6.4. In the special case where $f_{nuc} = \varphi(0)$, we have $\dot{s}_{nuc} = 0$ and

$$\frac{\dot{\delta}_{nuc}}{c} \;=\; \frac{\sigma_o}{\mu} \;+\; \frac{f_{nuc}}{\mu\gamma_T} \tag{6.23}$$

so that $\mathcal{N}$ then lies at the point of intersection of the curves $\mathcal{A}$ and $\mathcal{K}$.

In summary, the *unique solution* to the dynamic problem that is consistent with the kinetic law and nucleation condition is the following: if the elongation-rate lies in the range $0 < \dot{\delta} \leq \dot{\delta}_{nuc}$ the slab does not change phase and the solution is given by (6.11), (6.12). On the other hand when $\dot{\delta} > \dot{\delta}_{nuc}$, the slab does change phase and the solution is given by (6.14), (6.15), (6.22).

Quasi-static processes. Before ending this section it is worth briefly considering quasi-static processes of a finite slab, $0 < x < L$. Inertial effects are now neglected and the governing problem is identical to that which described the equilibrium problem of Section 5 except that now all fields depend on the time as a parameter. The elongation history $\delta(t)$ is prescribed and the fields $u(x,t), \gamma(x,t), \sigma(x,t)$ are to satisfy the (static) equations, jump conditions and boundary conditions (5.1), (5.2), (5.6) at each instant t. The three families of solutions, Solutions-1, -2 and -3, carry over immediately to the present quasi-static setting, and the totality of solutions can again be described in the δ, σ-plane by the parallelogram with the two extended lines shown in Figure 6.5 (recall Figure 5.1). However, in contrast to the static case, there is no reason now to restrict attention to energy minimizing solutions unless the loading-rate is infinitesimally slow. Consequently we do not have the additional phase equilibrium condition $f = 0$ at our disposal; instead we should enforce the kinetic law and nucleation condition just as in the inertial case. The driving force f associated with quasi-static processes is given by (5.35) and the dissipation inequality $f\dot{s} \geq 0$ must hold.

First, consider an instant t_1 at which the state of the slab is described by Solution-3 and is associated with a point (δ_1, σ_1) in the interior of the parallelogram (Figure 6.5). The slab involves a phase boundary at the location s_1 given by (5.14). Given the loading history $\delta(t),\ t > t_1$, our task is to find the resulting loading path, $(\delta(t), \sigma(t)),\ t > t_1$, emanating from the point (δ_1, σ_1). This path is determined by equations (5.14), (6.20), (5.35) which involve the three unknown functions $\sigma(t), f(t)$ and $s(t)$. Eliminating f and σ between them leads to

$$\frac{\delta}{L} \;=\; \frac{\varphi(\dot{s})}{\mu\gamma_T} + \frac{s\gamma_T}{L} + \frac{\sigma_o}{\mu}. \tag{6.24}$$

Given $\delta(t)$ for $t > t_1$, this ordinary differential equation, together with the initial condition $s(t_1) = s_1$, can be solved for $s(t)$. Hence $\sigma(t)$ can be determined from (5.14). Thus the kinetic law helps determine the loading path unambiguously.

Second, consider an instant t_2 at which the state of the slab is described by Solution-1 and is associated with a point (δ_2, σ_2) on the left boundary of the parallelogram. Note that

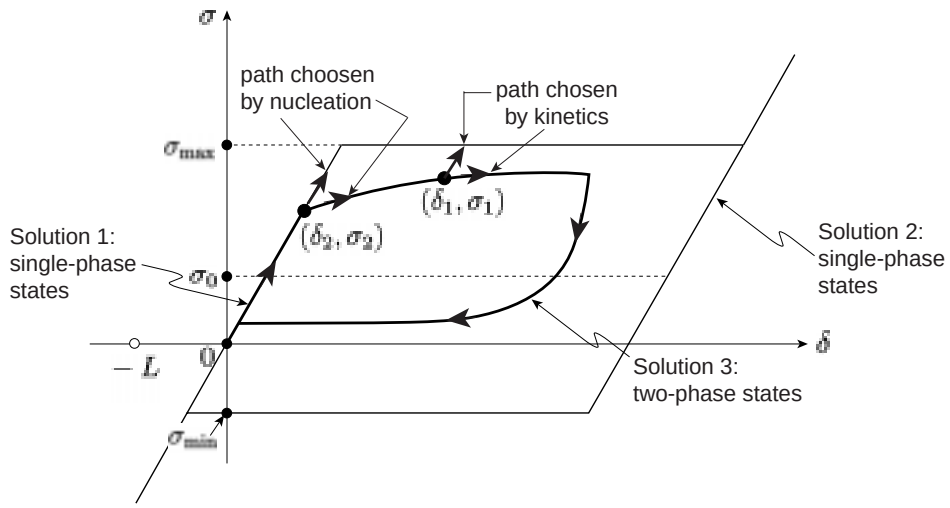

Figure 6.5: The set of all solutions involved in quasi-static processes of the slab. At a two-phase state such as (δ_1, σ_1) the kinetic law determines the path to be followed during subsequent loading/unloading. At a single-phase state such as (δ_2, σ_2) the nucleation condition determines whether during subsequent loading the path is to continue up the left boundary and therefore remain a single phase solution, or whether it should enter into the parallelogram and thus nucleate a phase transformation.

the slab is entirely in the low-strain phase at this instant and that therefore, it does not involve a phase boundary at which the kinetic law can be used. Again, given the elongation history, $\delta(t), t > t_2$, we wish to determine the resulting loading path. If it is known that the high-strain phase is not nucleated at the instant t_2+, then the slab remains entirely in the low-strain phase and therefore the loading path continues up the left boundary of the parallelogram. On the other hand if we knew that the high-strain phase is nucleated at t_2+, then a phase boundary appears, say at $s = 0$, and now we can use the kinetic law (6.24) to determine the loading path as in the preceding paragraph. Thus the decision that must be made is whether or not the new phase is nucleated at the instant t_2+. The nucleation condition answers this question and we find from (5.35) and (5.14) that the the critical values of stress and elongation at nucleation during this quasi-static process are $\sigma_{nuc} = \sigma_o + f_{nuc}/\gamma_T$ and

$$\frac{\delta_{nuc}}{L} = \frac{\sigma_o}{\mu} + \frac{f_{nuc}}{\mu\gamma_T} \tag{6.25}$$

respectively. The response during an arbitrary loading program can now be readily pieced together. The response is hysteretic and rate-dependent in general.

Conclusion. The dynamic problem describing the sudden elongation of a slab has served to illustrate the fact that when the strain-energy function has multiple minima, the usual formulation of the problem leads to a severe lack of uniqueness of solution. This lack of uniqueness was traced to the absence of information in the model describing when and how

fast the transformation from one energy-well to the other takes place. This in turn pointed to the need to supplement the formulation with a nucleation condition and a kinetic law which characterize these two events. Simple models of the nucleation condition and kinetic law were presented and the role they play in determining the (unique) solution was described. Finally it has been shown how these same conditions come into play in the quasi-static loading of a slab.

Additional insight into quasi-static processes can be gained by considering the *potential energy function* of the material introduced earlier. Recall from the discussion surrounding Figure 5.2 that the driving force f corresponds to the height between the two potential energy-wells, and moreover, that the low-strain energy-well lies below the high-strain energy-well when $f < 0$ and vice versa. In view of the dissipation inequality $f\dot{s} \geq 0$, this means that when the low-strain energy-well in the lower one, then $\dot{s} \leq 0$ and so the phase boundary cannot move to the right. Thus particles cannot transform from the phase on the right of the phase boundary to that on the left. Throughout this discussion we have taken the high-strain phase to be on the left of the phase boundary and the low-strain phase to be on the right (e.g., see discussions surrounding (5.13) or (6.7)). The dissipation inequality thus implies that a particle cannot jump from the lower energy-well to the higher one. Observe in addition that the kinetic law is in fact a relation between the rate of transformation and the height between the two wells of the potential energy G; this is a familiar picture in the material science characterization of rate-processes.

The specific forms of the nucleation condition and kinetic law involve detailed modeling of micromechanical processes and experimentation, a topic which we shall not address here. Explicit models of kinetic laws based on various mechanisms such as thermal activation, viscosity and strain-gradient effects, and wiggly energies, can be found, for example, in Christian (1975), Truskinovsky (1985), Abeyaratne and Knowles (1991b, 1993a), Müller and Xu (1991), and Abeyaratne et al. (1996). Some experimental studies of nucleation and kinetics are described, for example, in Grujicic at al. (1985), Chu (1993), Escobar and Clifton (1993, 1995), Escobar (1995) and Kyriakides and Shaw (1995).

7. Nonequilibrium thermodynamic processes. Kinetics. (Heidug and Lehner, 1985 and Abeyaratne and Knowles, 1990, 2000a) The dynamic and quasi-static problems analyzed in the preceding section both pointed to a clear need, in the continuum theory of phase transitions, for a law that characterizes the transformation rate. In the present section we take a broad approach to this issue and view it from the perspective of continuum thermodynamics. We will demonstrate, in particular, how the preceding anecdotal evidence for a kinetic law, is in fact a manifestation of the more general (and well-known) need for evolution laws whenever a system undergoes an irreversible thermodynamic process.

For illustrative purposes first consider the case of a smooth thermomechanical process of a thermoelastic body. Balance of linear and angular momentum lead to the field equations

$$\text{Div } \boldsymbol{S} + \boldsymbol{b} = \rho \dot{\boldsymbol{v}}, \qquad \boldsymbol{S}\boldsymbol{F}^T = \boldsymbol{F}\boldsymbol{S}^T, \tag{7.1}$$

and the first law of thermodynamics requires that

$$\boldsymbol{S} \cdot \dot{\boldsymbol{F}} + \text{Div } \boldsymbol{q} + r = \dot{\varepsilon}, \tag{7.2}$$

where the heat flux vector $\boldsymbol{q}$ characterizes the heat flow per unit reference area, ρ is the mass density in the reference configuration, and r and ε are the heat supply and internal energy respectively, both per unit reference volume. The second law of thermodynamics requires that the total rate of entropy production associated with any subregion $\mathcal{D}$ of the body be nonnegative:

$$\Gamma(t; \mathcal{D}) = \frac{d}{dt} \int_{\mathcal{D}} \eta dV \; - \; \int_{\partial \mathcal{D}} \frac{\boldsymbol{q} \cdot \boldsymbol{n}}{\theta} dA - \int_{\mathcal{D}} \frac{r}{\theta} dV \geq 0 \tag{7.3}$$

where η denotes the internal entropy per unit reference volume.

By making use of the field equations (7.1), (7.2) and the thermoelastic constitutive equations

$$\boldsymbol{S} = \frac{\partial W}{\partial \boldsymbol{F}}, \qquad \eta = -\frac{\partial W}{\partial \theta}, \qquad W = W(\boldsymbol{F}, \theta), \tag{7.4}$$

where $W = \varepsilon - \eta\theta$ denotes the Helmholtz free-energy per unit reference volume, one can simplify (7.3) and express the total rate of entropy production in the form

$$\Gamma(t; \mathcal{D}) = \int_{\mathcal{D}} \frac{\boldsymbol{q} \cdot \nabla \theta}{\theta^2} \; dV \; \geq 0. \tag{7.5}$$

This implies the pointwise inequality

$$\boldsymbol{q} \cdot \nabla \theta \; \geq \; 0 \tag{7.6}$$

which must hold at all points in the body. Observe that the entropy production rate involves the product of a flux and a quantity which "drives" that flux: one speaks of the temperature gradient $\nabla\theta$ as the *driving force* conjugate to the heat flux $\boldsymbol{q}$. In the special case of *thermodynamic equilibrium*, the flux and the driving force both vanish:

$$\boldsymbol{q} \; = \; \boldsymbol{o}, \qquad \nabla \theta \; = \; \boldsymbol{o}. \tag{7.7}$$

There is nothing in the description of a thermoelastic material, or the usual principles of continuum mechanics, which provide any information about the driving force $\nabla\theta$ and the flux $\boldsymbol{q}$ when the body is *not* in thermodynamic equilibrium. According to the theory of irreversible processes (e.g., Truesdell, 1969, Kestin, 1979, and Callen, 1985), such information

must be provided empirically, by relating the flux $\boldsymbol{q}$ to the driving force $\nabla\theta$ and other state variables. In the present setting, this would be a relation of the form

$$\boldsymbol{q} = \hat{\boldsymbol{q}}(\nabla\theta; \boldsymbol{F}, \theta), \tag{7.8}$$

where $\hat{\boldsymbol{q}}$ is a constitutive function. We recognize the *kinetic law* (7.8) to be no more than the familiar *heat conduction law*, a common example of which is the linear relation $\hat{\boldsymbol{q}} = \boldsymbol{K}\nabla\theta$ with $\boldsymbol{K}$ the conductivity tensor. In general, the kinetic response function (heat conduction response function) $\hat{\boldsymbol{q}}$ must be such that

$$\hat{\boldsymbol{q}}(\boldsymbol{g}; \boldsymbol{F}, \theta) \cdot \boldsymbol{g} \geq 0 \quad \text{for all vectors } \boldsymbol{g}, \qquad \hat{\boldsymbol{q}}(\boldsymbol{o}; \boldsymbol{F}, \theta) = \boldsymbol{o}; \tag{7.9}$$

the former is a consequence of the entropy inequality (7.6), while the latter reflects (7.7) and in any case follows from the former in the presence of sufficient smoothness.

The preceding discussion concerned the particular setting of smooth processes of a thermoelastic body. As discussed for example in Callen (1985), under very general circumstances the rate of entropy production can be expressed as the sum of the products of various thermodynamic driving forces and their conjugate fluxes, and the theory of irreversible processes states that the evolution of the system is governed by kinetic laws relating the driving forces to the fluxes.

The setting of interest to us is the propagation of a phase boundary in a thermoelastic body. Let the surface $\mathcal{S}_t$ denote the location of the phase boundary in the reference configuration and assume that the deformation and temperature remain continuous across this surface but that the deformation gradient, particle velocity and temperature gradient are permitted to suffer jump discontinuities. Let V_n denote the normal velocity of propagation of $\mathcal{S}_t$. In this case, the field equations (7.1), (7.2) must be supplemented by the jump conditions

$$[\![\boldsymbol{S}\boldsymbol{n}]\!] = -\rho V_n [\![\boldsymbol{v}]\!], \qquad [\![\boldsymbol{S}\boldsymbol{n} \cdot \boldsymbol{v}]\!] + [\![\boldsymbol{q} \cdot \boldsymbol{n}]\!] + V_n [\![\varepsilon + \rho \boldsymbol{v} \cdot \boldsymbol{v}/2]\!] = 0, \tag{7.10}$$

which are to hold at points on the phase boundary $\mathcal{S}_t$. The field equations, jump conditions and constitutive relation can be used to write the total rate of entropy production as

$$\Gamma(t; \mathcal{D}) = \int_{\mathcal{D}} \frac{\boldsymbol{q} \cdot \nabla\theta}{\theta^2} \, dV \; + \; \int_{\mathcal{S}_t \,\cap\, \mathcal{D}} \frac{f V_n}{\theta} \, dA, \tag{7.11}$$

where

$$f \;=\; [\![W]\!] \;-\; \frac{\overset{+}{\boldsymbol{S}} + \overset{-}{\boldsymbol{S}}}{2} \cdot [\![\boldsymbol{F}]\!]; \tag{7.12}$$

cf. (4.19). Thus there are now two sources of entropy production: heat conduction and phase transformation. From (7.11) we identify $\nabla\theta$ as the driving force associated with heat

conduction and $\boldsymbol{q}$ as its conjugate flux, and f as the driving force associated with phase transformation and V_n as its corresponding flux. Note that while $\boldsymbol{q}$ represents the flux of heat, ρV_n represents the flux of mass across the propagating phase boundary. If the body is in thermal equilibrium, i.e., $\theta(\boldsymbol{X},t) = \textit{constant}$, the entropy production rate due to heat conduction vanishes; similarly if the body is in phase equilibrium, by which we mean that $f(\boldsymbol{X},t)$ vanishes on $\mathcal{S}_t$, the entropy production rate due to phase transformation vanishes. Thus $\nabla\theta$ and f are the "agents of entropy production" when a thermoelastic body is not in thermal and phase equilibrium.

As noted previously, according to the general theory, the evolution of the various fluxes in an irreversible process are governed by *kinetic laws* relating each flux to its conjugate driving force and other state variables. In the present setting, this implies the kinetic laws

$$\begin{aligned} &\text{Heat conduction}: \quad \boldsymbol{q} = \hat{\boldsymbol{q}}(\nabla\theta; \boldsymbol{F}, \theta), \\ &\text{Phase transition}: \quad V_n = \hat{V}\left(f; \overset{+}{\boldsymbol{F}}, \overset{-}{\boldsymbol{F}}, \theta\right) \quad \text{on} \quad \mathcal{S}_t. \end{aligned} \tag{7.13}$$

The constitutive functions $\hat{\boldsymbol{q}}$ and $\hat{V}$ must satisfy the requirements

$$\hat{\boldsymbol{q}}(\boldsymbol{g}; \boldsymbol{F}, \theta) \cdot \boldsymbol{g} \geq 0 \text{ for all vectors } \boldsymbol{g}, \qquad \hat{V}(f; \overset{+}{\boldsymbol{F}}, \overset{-}{\boldsymbol{F}}, \theta) f \geq 0 \text{ for all scalars } f, \tag{7.14}$$

$$\hat{\boldsymbol{q}}(\boldsymbol{o}; \boldsymbol{F}, \theta) = \boldsymbol{o} \qquad \hat{V}(0; \overset{+}{\boldsymbol{F}}, \overset{-}{\boldsymbol{F}}, \theta) = 0, \tag{7.15}$$

where the inequalities follow from the entropy inequalities $\boldsymbol{q} \cdot \nabla\theta \geq 0$ and $fV_n \geq 0$, and the equalities reflect thermal and phase equilibrium.

Conclusion. The preceding discussion allows us to place our previous analyses in a more general context. In particular we see that the driving force f, which we first encountered in Section 3, has a more general thermodynamic significance; that the dissipation inequality is in fact a consequence of the second law of thermodynamics; that the phase equilibrium condition $f = 0$, first encountered in Section 4, is consistent with the notion of thermodynamic equilibrium; and finally that the kinetic law is required because of the nonequilibrium character of the phase transformation process.

It is worth noting the following general expression for the driving force on a phase boundary,

$$f \;=\; \left(\overset{+}{\psi} - \overset{-}{\psi}\right) - \;\frac{\overset{+}{\boldsymbol{S}} + \overset{-}{\boldsymbol{S}}}{2} \cdot \left(\overset{+}{\boldsymbol{F}} - \overset{-}{\boldsymbol{F}}\right) + \;\;\frac{\overset{+}{\eta} + \overset{-}{\eta}}{2}\left(\overset{+}{\theta} - \overset{-}{\theta}\right), \tag{7.16}$$

which is valid in three-dimensions for any continuum (not necessarily thermoelastic) undergoing an arbitrary thermomechanical process (including an adiabatic one), Abeyaratne and Knowles (2000a). Here $\boldsymbol{S}(\boldsymbol{X},t)$, $\psi(\boldsymbol{X},t)$ and $\eta(\boldsymbol{X},t)$ denote the fields of Piola-Kirchhoff stress, Helmholtz free-energy and specific entropy respectively and no constitutive relation

between them is assumed. Equation (7.16) specializes for a thermoelastic material by using the constitutive characterizations $\psi = W(\boldsymbol{F},\theta),\ \boldsymbol{S} = W_{\boldsymbol{F}},\ \eta = -W_\theta$. In the presence of heat conduction, the temperature is continuous across the phase boundary and so the last term in (7.16) disappears; in an adiabatic process however the temperature can jump across the phase boundary and so this term must be retained.

8. Higher dimensional static problems. The issue of geometric compatibility. (Wechsler et al., 1953, Bowles and Mackenzie, 1954, Ball and James, 1987, and Bhattacharya, 1991.) In the equilibrium theory of finite elasticity we have the constitutive relation

$$\boldsymbol{S} = \frac{\partial W(\boldsymbol{F})}{\partial \boldsymbol{F}}, \tag{8.1}$$

and the equations of force and moment balance

$$\text{Div } \boldsymbol{S} = \boldsymbol{o}, \qquad \boldsymbol{S}\boldsymbol{F}^T = \boldsymbol{F}\boldsymbol{S}^T, \tag{8.2}$$

which must hold at points in the body where the fields are smooth. At a phase boundary, the deformation gradient tensor and stress suffer jump discontinuities but the deformation itself is continuous. The jump conditions associated with force balance and the continuity of the deformation are

$$\overset{+}{\boldsymbol{S}}\,\boldsymbol{n} \ = \overset{-}{\boldsymbol{S}}\,\boldsymbol{n}, \qquad \overset{+}{\boldsymbol{F}}\,\boldsymbol{\ell} = \overset{-}{\boldsymbol{F}}\,\boldsymbol{\ell} \tag{8.3}$$

which must hold at all points $\boldsymbol{X}$ on the phase boundary; the vector $\boldsymbol{n}$ here is a unit normal to the phase boundary at $\boldsymbol{X}$ and $(8.3)_2$ must hold for all vectors $\boldsymbol{\ell}$ which are tangent to the phase boundary at $\boldsymbol{X}$. Finally, phase equilibrium requires that

$$f \ = \ W(\overset{+}{\boldsymbol{F}}) - W(\overset{-}{\boldsymbol{F}}) - \ \overset{\pm}{\boldsymbol{S}} \cdot \left(\overset{+}{\boldsymbol{F}} - \overset{-}{\boldsymbol{F}}\right) \ = \ 0 \tag{8.4}$$

which is one of the Weiestrass-Erdmann corner conditions, the other being $(8.3)_1$ (see Grinfeld, 1981, James, 1981, and Abeyaratne, 1983).

Here we consider the simplest possible static problem involving a phase boundary: consider a body which occupies all of $\mathbb{R}^3$, and suppose that it is subjected to a piecewise homogeneous deformation

$$\boldsymbol{x} = \begin{cases} \overset{+}{\boldsymbol{F}}\boldsymbol{X} & \text{for } \boldsymbol{X}\cdot\boldsymbol{n} \geq 0, \\ \\ \overset{-}{\boldsymbol{F}}\boldsymbol{X} & \text{for } \boldsymbol{X}\cdot\boldsymbol{n} \leq 0, \end{cases} \tag{8.5}$$

where $\overset{\pm}{\boldsymbol{F}}$ are constant tensors belonging to two distinct phases (or variants) of the material. The planar interface $\mathcal{S} = \{\boldsymbol{X} : \ \boldsymbol{X}\cdot\boldsymbol{n} = 0\}$ corresponds to a phase boundary. Furthermore, suppose that the deformation gradient tensors $\overset{\pm}{\boldsymbol{F}}$ lie at the bottoms of two energy-wells

so that $\partial W(\overset{\pm}{\boldsymbol{F}})/\partial \boldsymbol{F} = \boldsymbol{O}$. Finally, assume that the temperature is the transformation temperature so that $W(\overset{+}{\boldsymbol{F}}) = W(\overset{-}{\boldsymbol{F}})$. Thus the equilibrium requirements (8.2), $(8.3)_1$ and the phase equilibrium requirement (8.4) are all satisfied trivially. The only equation that remains to be satisfied is the kinematic requirement $(8.3)_2$, which can be expressed in the equivalent form

$$\overset{+}{\boldsymbol{F}} = \overset{-}{\boldsymbol{F}} + \boldsymbol{a} \otimes \boldsymbol{n} \tag{8.6}$$

where $\boldsymbol{a} \neq \boldsymbol{o}$ is an arbitrary vector.

Keeping in mind the description of the energy-wells in (2.21) - (2.23), the questions we pose are, given two stretch tensors $\overset{+}{\boldsymbol{U}}$ and $\overset{-}{\boldsymbol{U}}$,

(i) under what conditions on $\overset{\pm}{\boldsymbol{U}}$ does (8.6) hold for some vectors $\boldsymbol{a}, \boldsymbol{n}$ and some rotation tensors $\overset{\pm}{\boldsymbol{R}}$ with $\overset{\pm}{\boldsymbol{F}} = \overset{\pm}{\boldsymbol{R}}\overset{\pm}{\boldsymbol{U}}$?, and

(ii) when (8.6) does hold, how does one determine $\boldsymbol{a}, \boldsymbol{n}$ and $\overset{\pm}{\boldsymbol{R}}$?

It is convenient to first rephrase the questions in a form which does not involve the rotation tensors. Observe that if $\boldsymbol{a}, \boldsymbol{n}, \overset{+}{\boldsymbol{R}}, \overset{-}{\boldsymbol{R}}$ satisfy (8.6) for some $\overset{+}{\boldsymbol{U}}$ and $\overset{-}{\boldsymbol{U}}$, then so do $\boldsymbol{Q}\boldsymbol{a}, \boldsymbol{n}, \boldsymbol{Q}\overset{+}{\boldsymbol{R}}, \boldsymbol{Q}\overset{-}{\boldsymbol{R}}$ for any rotation $\boldsymbol{Q}$. Thus with no loss of generality we can pick one of the rotation tensors $\overset{\pm}{\boldsymbol{R}}$ arbitrarily and so we take $\overset{-}{\boldsymbol{R}} = \boldsymbol{I}$. Next, let $\boldsymbol{D}, \boldsymbol{b}$ and $\boldsymbol{m}$ be defined by

$$\boldsymbol{D} = \overset{+}{\boldsymbol{F}}\overset{-}{\boldsymbol{F}}^{-1}, \qquad \boldsymbol{b} = |\overset{-}{\boldsymbol{F}}^{-T}\boldsymbol{n}|\boldsymbol{a}, \qquad \boldsymbol{m} = \overset{-}{\boldsymbol{F}}^{-T}\boldsymbol{n}/|\overset{-}{\boldsymbol{F}}^{-T}\boldsymbol{n}|. \tag{8.7}$$

On using (8.7), equation (8.6) can be written as $\boldsymbol{D} = \boldsymbol{I} + \boldsymbol{b} \otimes \boldsymbol{m}$; since $\det \boldsymbol{D} > 0$, we need $\boldsymbol{b} \cdot \boldsymbol{m} > -1$. Finally, set $\boldsymbol{C} = \boldsymbol{D}^T\boldsymbol{D}$ so that $\boldsymbol{C}$ is symmetric and positive definite and note from $(8.7)_1$ that it can be written as

$$\boldsymbol{C} = \overset{-}{\boldsymbol{U}}^{-1}\overset{+}{\boldsymbol{U}}^2\overset{-}{\boldsymbol{U}}^{-1}. \tag{8.8}$$

The geometric compatibility equation now yields

$$\boldsymbol{C} = (\boldsymbol{I} + \boldsymbol{m} \otimes \boldsymbol{b})(\boldsymbol{I} + \boldsymbol{b} \otimes \boldsymbol{m}). \tag{8.9}$$

The questions asked above can now be posed in the following equivalent form: (i) find conditions on the symmetric positive definite tensor $\boldsymbol{C}$ under which (8.9) holds for some vectors $\boldsymbol{b}$ and $\boldsymbol{m}$, $\boldsymbol{b} \cdot \boldsymbol{m} > -1$, and ($ii$) when these conditions hold find $\boldsymbol{b}$ and $\boldsymbol{m}$. Note that once $\boldsymbol{b}$ and $\boldsymbol{m}$ have been found, $\overset{+}{\boldsymbol{R}}$ can be determined from $\boldsymbol{D} = \overset{+}{\boldsymbol{F}}\overset{-}{\boldsymbol{F}}^{-1} = \overset{+}{\boldsymbol{R}}\overset{+}{\boldsymbol{U}}\overset{-}{\boldsymbol{U}}^{-1} = \boldsymbol{I} + \boldsymbol{b} \otimes \boldsymbol{m}$ while $\boldsymbol{a}$ and $\boldsymbol{n}$ can be determined from $(8.7)_{2,3}$.

Proposition: (Ball and James, 1987) Let $0 < \lambda_1 \leq \lambda_2 \leq \lambda_3$ be the ordered eigenvalues of $\boldsymbol{C}$ and let $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ be a set of corresponding orthonormal eigenvectors. Then a necessary and sufficient condition for there to exist vectors $\boldsymbol{b}$ ($\neq \boldsymbol{o}$) and $\boldsymbol{m}$ with $\boldsymbol{b} \cdot \boldsymbol{m} > -1$ satisfying (8.9) is that

$$\lambda_2 = 1, \qquad \lambda_1 \neq \lambda_3. \tag{8.10}$$

When (8.10) holds, there are two pairs of vectors $\boldsymbol{b}, \boldsymbol{m}$ for which (8.9) holds and they are given by

$$\left.\begin{aligned} \boldsymbol{b} &= g\left\{\sqrt{\frac{\lambda_3(1-\lambda_1)}{\lambda_3-\lambda_1}}\boldsymbol{e}_1 \;+\; \kappa\sqrt{\frac{\lambda_1(\lambda_3-1)}{\lambda_3-\lambda_1}}\;\boldsymbol{e}_3\right\}, \\ \boldsymbol{m} &= \frac{1}{g}\;\frac{\sqrt{\lambda_3}-\sqrt{\lambda_1}}{\sqrt{\lambda_3-\lambda_1}}\;\left\{-\sqrt{1-\lambda_1}\;\boldsymbol{e}_1 \;+\; \kappa\;\sqrt{\lambda_3-1}\;\boldsymbol{e}_3\right\}, \end{aligned}\right\} \tag{8.11}$$

where g is chosen to make $\boldsymbol{m}$ a unit vector and $\kappa = \pm 1$.

The proof of this result may be found in Ball and James (1987). We now apply it to a number of examples, all of which pertain to a material which possesses cubic and tetragonal phases. As discussed in Sections 2 and 3, the cubic austenite phase is carried into the tetragonal martensite phase by subjecting it to stretches α, α, β in the cubic directions $\{\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3\}$; here $\alpha = a/a_0,\ \beta = c/a_o$ where the cubic and tetragonal lattice parameters are $a_o \times a_o \times a_o$ and $a \times a \times c$ respectively. Recall that the martensite phase comes in three variants characterized by the stretch tensors $\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{U}_3$ whose components in the cubic basis are given by (3.6). Since the reference configuration has been taken to coincide with unstressed austenite, it corresponds to the deformation gradient tensor $\boldsymbol{I}$.

In the examples to follow we shall not explicitly write out the solutions $\boldsymbol{b}, \boldsymbol{m}, \boldsymbol{a}, \boldsymbol{n}, \overset{+}{\boldsymbol{R}}$. We will focus instead on the implications of the condition (8.10) for piecewise homogeneous deformations. In particular, the examples will illustrate how (8.10) restricts the possibilities, in some cases disallowing such deformations, in others allowing them always, and in yet others, allowing them under special circumstances.

Example 1: One variant of martensite. Consider the possibility of constructing a piecewise homogeneous deformation involving just one martensitic variant, say $\boldsymbol{U}_1$. Thus we take $\overset{+}{\boldsymbol{F}} = \overset{+}{\boldsymbol{R}}\,\boldsymbol{U}_1$ and $\overset{-}{\boldsymbol{F}} = \boldsymbol{U}_1$. Then $\boldsymbol{C} = \boldsymbol{I}$ and so the eigenvalues of $\boldsymbol{C}$ do *not* satisfy (8.10). Hence a deformation of the assumed form does not exist. The same conclusion is reached if we had taken $\overset{+}{\boldsymbol{F}} = \boldsymbol{R}$ and $\overset{-}{\boldsymbol{F}} = \boldsymbol{I}$ corresponding to a piecewise homogeneous deformation involving austenite alone. These observations correspond to a special case of the following general result: if a continuous and piecewise smooth deformation $\boldsymbol{x}(\boldsymbol{X})$ on a region $\mathcal{R}$ is such that $\nabla\boldsymbol{x}(\boldsymbol{X}) = \boldsymbol{R}(\boldsymbol{X})\boldsymbol{U}$ where $\boldsymbol{U}$ is a *constant* stretch tensor and $\boldsymbol{R}(\boldsymbol{X})$ is a field of rotation tensors, then $\boldsymbol{R}$ is necessarily constant.

Example 2: Two variants of martensite. Next consider a piecewise homogeneous deformation involving two of the three martensitic variants, say $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$. Now $\overset{+}{\boldsymbol{F}} = \overset{+}{\boldsymbol{R}}\,\boldsymbol{U}_2$ and $\overset{-}{\boldsymbol{F}} = \boldsymbol{U}_1$ and so $\boldsymbol{C} = \boldsymbol{U}_1^{-1}\boldsymbol{U}_2^2\boldsymbol{U}_1^{-1}$. From (3.6), $\boldsymbol{C}$ has components

$$[C] = \begin{pmatrix} \alpha^2/\beta^2 & 0 & 0 \\ 0 & \beta^2/\alpha^2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{8.12}$$

whence the ordered eigenvalues of $\boldsymbol{C}$ are $\alpha^2/\beta^2,\ 1,\ \beta^2/\alpha^2$ (if $\alpha < \beta$), or $\beta^2/\alpha^2,\ 1,\ \alpha^2/\beta^2$ (if $\beta < \alpha$). Thus the requirements (8.10) are automatically satisfied thus guaranteeing the existence of a piecewise homogeneous deformation involving two variants. A deformation involving two martensitic variants is called a *twin* or *twinning deformation*; the interface between the variants is a *twin boundary*.

Example 3: Austenite and one variant of martensite. Next consider the possibility of a piecewise homogeneous deformation involving austenite on one side of the interface and one variant of martensite, say $\boldsymbol{U}_1$, on the other side. Thus we take $\overset{+}{\boldsymbol{F}} = \overset{+}{\boldsymbol{R}}\,\boldsymbol{U}_1$ and $\overset{-}{\boldsymbol{F}} = \boldsymbol{I}$ and find $\boldsymbol{C} = \boldsymbol{U}_1^2$. From (3.6) we get

$$[C] = \begin{pmatrix} \beta^2 & 0 & 0 \\ 0 & \alpha^2 & 0 \\ 0 & 0 & \alpha^2 \end{pmatrix}. \tag{8.13}$$

It follows that the requirement (8.10) holds if and only if

$$\alpha = 1, \qquad \beta \neq 1. \tag{8.14}$$

The condition $\alpha = 1$ is a restriction on the lattice and implies that the lattice parameters must obey $a = a_o$. The austenite $\rightarrow$ martensite transformation therefore involves a change in only *one* lattice parameter, the cubic and tetragonal lattice dimensions being $a_o \times a_o \times a_o$ and $a_o \times a_o \times c$ respectively. No known material has a lattice with this characteristic, and this explains why no sharp interfaces separating cubic austenite from one variant of tetragonal martensite have been observed.

The negative result of Example 3 suggests that we examine austenite/martensite deformations which have a more complicated structure than that in Example 3. A complete

discussion of the next two examples in beyond the scope of this article since they must be viewed from the point of view of sequences of deformations with increasingly finer microstructure (see Ball and James, 1987, and Ball, 1989). Here we restrict attention to a discussion of the macroscopic kinematics.

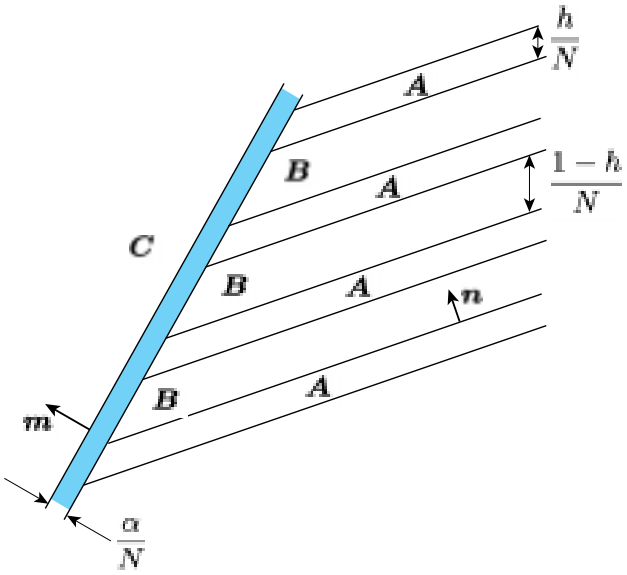

Figure 8.1: An interface separating austenite from twinned martensite. The martensitic variants are characterized by the deformation gradient tensors $\boldsymbol{A}, \boldsymbol{B}$ and have volume fractions h and $1-h$. The thickness of the shaded diffuse interface tends to zero as the fineness of the twin bands increases, i.e., $N \to \infty$.

Example 4: Austenite and twinned martensite. (Ball and James, 1987) Let us now consider the deformation described schematically in Figure 8.1 in which there is an "interface" which separates austenite from martensite. The martensitic half-space involves fine bands of martensite which alternate between variants -1 and -2. Thus the martensitic region is twinned, and Example 2 describes its kinematics. The deformation gradient tensors in the martensitic layers are $\boldsymbol{A} = \boldsymbol{Q}_2\boldsymbol{U}_2$ and $\boldsymbol{B} = \boldsymbol{Q}_1\boldsymbol{U}_1$ and geometric compatibility across the martensite/martensite interfaces requires

$$\boldsymbol{A} - \boldsymbol{B} = \boldsymbol{a}' \otimes \boldsymbol{n}. \tag{8.15}$$

The widths of the martensitic layers are h/N and $(1-h)/N$ where the volume fraction h is as yet unknown and N is a positive integer. We have seen in Example 3 that neither variant of martensite (characterized by $\boldsymbol{A}$ or $\boldsymbol{B}$) is compatible with austenite (characterized by $\boldsymbol{C}$). Therefore, at best, we can only join austenite to martensite through an interpolation layer, corresponding to a diffuse interface of width α/N as depicted by the grey strip in Figure 8.1.

Consider now the sequence of deformations indexed by N. As $N \to \infty$ the diffuse interface collapses to a plane. If the local deformation gradient field is to remain bounded during this limiting process, it is necessary and sufficient that the deformation gradient $\boldsymbol{C}$ in the austenite region be compatible with the *average* deformation gradient $h\boldsymbol{A} + (1-h)\boldsymbol{B}$ in the martensite region. In other words, as $N \to \infty$, the austenite regions "sees" not the individual martensitic variants but the average of the two. Compatibility between the austenite and the averaged martensite across the austenite/martensite interface requires

$$[h\boldsymbol{A} + (1-h)\boldsymbol{B}] - \boldsymbol{C} = \boldsymbol{b} \otimes \boldsymbol{m}. \tag{8.16}$$

Given $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$, we wish to solve this pair of equations with $\boldsymbol{A} = \boldsymbol{Q}_2\boldsymbol{U}_2, \boldsymbol{B} = \boldsymbol{Q}_1\boldsymbol{U}_1, \boldsymbol{C} = \boldsymbol{I}$ and determine $\boldsymbol{a}', \boldsymbol{n}, \boldsymbol{b}, \boldsymbol{m}, \boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$.

In order to put this into the standard form, set $\boldsymbol{R} = \boldsymbol{Q}_1^T\boldsymbol{Q}_2,\ \bar{\boldsymbol{R}} = \boldsymbol{Q}_1,\ \boldsymbol{a} = \boldsymbol{Q}_1^T\boldsymbol{a}'$ so that the two compatibility equations can be rewritten as

$$\boldsymbol{R}\boldsymbol{U}_2 - \boldsymbol{U}_1 = \boldsymbol{a} \otimes \boldsymbol{n}, \qquad \bar{\boldsymbol{R}}\,(\boldsymbol{U}_1 + h\boldsymbol{a} \otimes \boldsymbol{n}) = \boldsymbol{I} + \boldsymbol{b} \otimes \boldsymbol{m}. \tag{8.17}$$

The first of these is the twinning equation which we examined previously in Example 2 and found to be solvable under all conditions. The second equation leads to (8.9) provided we set

$$\boldsymbol{C} = (\boldsymbol{U}_1 + h\boldsymbol{n} \otimes \boldsymbol{a})\,(\boldsymbol{U}_1 + h\boldsymbol{a} \otimes \boldsymbol{n})\,. \tag{8.18}$$

If (8.9) is to be solvable, $\boldsymbol{C}$ must satisfy the conditions (8.10); in particular, it is necessary that one of the eigenvalues of $\boldsymbol{C}$ be unity, and therefore that $\det(\boldsymbol{C} - \boldsymbol{I}) = 0$. An explicit expression for $\boldsymbol{C}$ can be obtained by using (8.18) and substituting into it (3.6) for $\boldsymbol{U}_1$, and the solution of Example 2 for $\boldsymbol{a}$ and $\boldsymbol{n}$. This leads to

$$\begin{aligned} \boldsymbol{C} = \ & \frac{[h(\beta^2-\alpha^2)-\beta^2]^2 + \alpha^2\beta^2}{\alpha^2+\beta^2}\boldsymbol{c}_1 \otimes \boldsymbol{c}_1 \ + \ \frac{[h(\beta^2-\alpha^2)+\alpha^2]^2 + \alpha^2\beta^2}{\alpha^2+\beta^2}\boldsymbol{c}_2 \otimes \boldsymbol{c}_2 \\ & \pm \frac{(\beta^2-\alpha^2)^2}{\beta^2+\alpha^2}\, h(1-h)(\boldsymbol{c}_1 \otimes \boldsymbol{c}_2 + \boldsymbol{c}_2 \otimes \boldsymbol{c}_1) \ + \ \alpha^2 \boldsymbol{c}_3 \otimes \boldsymbol{c}_3 \end{aligned} \tag{8.19}$$

which expresses $\boldsymbol{C}$ in terms of the lattice parameters α, β, the orthonormal cubic basis vectors $\{\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3\}$, and the unknown volume fraction h; the $\pm$ here arises from the fact that the twinning equation $(8.17)_1$ has two solutions. Setting $\det(\boldsymbol{C} - \boldsymbol{I}) = 0$ now leads to a quadratic equation for the volume fraction h whose roots are

$$h = \frac{1}{2}\left(\ 1 \pm \sqrt{1 + \frac{2(\alpha^2-1)(\beta^2-1)(\beta^2+\alpha^2)}{(\beta^2-\alpha^2)^2}}\ \right). \tag{8.20}$$

Note that the sum of these two roots is unity. When (8.20) holds, the eigenvalues of $\boldsymbol{C}$ are found to be $1, \alpha^2, \alpha^2\beta^2$. The necessary and sufficient condition (8.10), together with the

inequality which ensures that (8.20) yields real roots in the interval $[0, 1]$, can now be shown to be equivalent to

$$\begin{array}{llll} \text{either} & \alpha < 1 < \beta & \text{and} & \alpha^{-2} + \beta^{-2} \leq 2, \\ \text{or} & \beta < 1 < \alpha & \text{and} & \alpha^{2} + \beta^{2} \leq 2. \end{array} \tag{8.21}$$

Thus in summary, an austenite-twinned martensite deformation of the assumed form (Figure 8.1) exists provided that the lattice parameters obey (8.21) and the variant volume fraction is given by (8.20).

Example 5: Wedge of twinned martensite in austenite. (Bhattacharya, 1991) As a final example consider the deformation illustrated in Figure 8.2. Here one has a wedge of martensite surrounded by austenite. The wedge involves a mid-rib, on one side of which there is a fine mixture of two martensitic variants characterized by deformation gradients $\boldsymbol{A}$ and $\boldsymbol{B}$ and volume fractions h_1 and $1 - h_1$; on the other side there is second fine mixture of martensitic variants characterized by the deformation gradients $\boldsymbol{C}, \boldsymbol{D}$ and volume fractions h_2 and $1-h_2$.

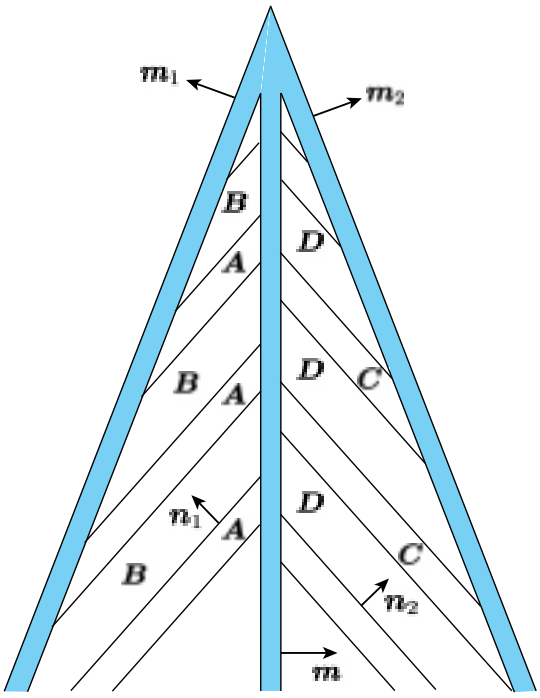

Figure 8.2: Wedge of twinned martensite surrounded by austenite. The microstructure involves fives types of interfaces: two sets of twin boundaries – separating different variants of martensite, $\boldsymbol{A}/\boldsymbol{B}$ and $\boldsymbol{C}/\boldsymbol{D}$ –, two austenite/martensite interfaces, and the mid-rib of the wedge which separates twinned martensite from twinned martensite.

There are five types interfaces to be considered: two sets of twin boundaries, the two austenite-martensite boundaries, and the mid-rib. Enforcing geometric compatibility across these interfaces leads to the following five conditions:

$$\begin{aligned}
&\text{Martensite/Martensite :} && \boldsymbol{A} - \boldsymbol{B} = \boldsymbol{a}_1 \otimes \boldsymbol{n}_1,\\
&\text{Martensite/Martensite :} && \boldsymbol{C} - \boldsymbol{D} = \boldsymbol{a}_2 \otimes \boldsymbol{n}_2,\\
&\text{Austenite/TwinnedMartensite :} && [h_1\boldsymbol{A} + (1-h_1)\boldsymbol{B}] - \boldsymbol{I} = \boldsymbol{b}_1 \otimes \boldsymbol{m}_1,\\
&\text{Austenite/TwinnedMartensite :} && [h_2\boldsymbol{C} + (1-h_2)\boldsymbol{D}] - \boldsymbol{I} = \boldsymbol{b}_2 \otimes \boldsymbol{m}_2,\\
&\text{TwinnedMartensite/TwinnedMartensite :}\\
&\qquad [h_1\boldsymbol{A} + (1-h_1)\boldsymbol{B}] \;-\; [h_2\boldsymbol{C} + (1-h_2)\boldsymbol{D}] = \boldsymbol{b} \otimes \boldsymbol{m}.
\end{aligned} \tag{8.22}$$

The first two of these are twinning equations and we know from Example 2 that they are solvable without the need for any restrictions. The next two equations are precisely of the type analyzed in Example 4. We know that these are solvable provided the volume fractions h_1 and h_2 of the variants are suitably chosen (and the lattice parameters satisfy some mild inequalities). This completely determines all of the free parameters in the problem. The solvability of the final equation therefore imposes a relationship between the lattice parameters which Bhattacharya (1991) has shown to be

$$\alpha^2 = \frac{(1-\beta^2)^2 + 4\beta^2(1+\beta^2)}{(1-\beta^2)^2 + 8\beta^4}. \tag{8.23}$$

Thus we conclude that a deformation of the assumed form involving a wedge of martensite can only exist in special materials where the lattice parameters satisfy the restriction (8.23). Bhattacharya (1991) has conducted an extensive survey of the literature and shown that a variety of cubic-tetragonal martensitic materials do in fact satisfy this restriction and that these materials exhibit martensitic wedges; there are also materials which do not satisfy (8.23) and they do not exhibit wedge shaped martensitic domains.

Acknowledgments. Many of the results described here were obtained in the course of investigations supported by the U.S. Army Research Office, National Science Foundation and Office of Naval Research. We are grateful to Debra Blanchard for preparing all of the figures.

References

1. Abeyaratne, R. 1983 An admissibility condition for equilibrium shocks in finite elasticity. *Journal of Elasticity* **13**, 175–184.

2. Abeyaratne, R., Chu, C. and James, R.D. 1996 Kinetics of materials with wiggly energies: theory and application to the evolution of twinning microstructures in a Cu-Al-Ni shape memory alloy. *Philosophical Magazine A* **73**, 457–497.

3. Abeyaratne, R. and Knowles, J.K. 1990 On the driving traction acting on a surface of strain discontinuity in a continuum. *Journal of the Mechanics and Physics of Solids* **38**, 345–360.

4. Abeyaratne, R. and Knowles, J.K. 1991a Kinetic relations and the propagation of phase boundaries in solids. *Archive for Rational Mechanics and Analysis* **114**, 119–154.

5. Abeyaratne, R. and Knowles, J.K. 1991b Implications of viscosity and strain-gradient effects for the kinetics of propagating phase boundaries in solids. *SIAM Journal of Applied Mathematics* **51**, 1205–1221.

6. Abeyaratne, R. and Knowles, J.K. 1993a A continuum model of a thermoelastic solid capable of undergoing phase transitions. *Journal of the Mechanics and Physics of Solids* **41**, 541–571.

7. Abeyaratne, R. and Knowles, J.K. 1993b Nucleation, Kinetics and Admissibility Criteria for Propagating Phase Boundaries. In *Shock Induced Transitions and Phase Structures in General Media*, eds. J.E. Dunn, R.L. Fosdick and M. Slemrod. Springer-Verlag IMA Volume 52, 3–33.

8. Abeyaratne, R. and Knowles, J.K. 2000a A note on the driving traction acting on a propagating interface: adiabatic and non-adiabatic processes of a continuum. *ASME Journal of Applied Mechanics* accepted August 1997.

9. Abeyaratne, R. and Knowles, J.K. 2000b On a shock-induced martensitic phase transition. *Journal of Applied Physics* **87**, 1123–1134.

10. Abeyaratne, R. and Knowles, J.K. 2001 *Evolution of Phase Transitions*, Cambridge University Press. To appear.

11. Ball, J.M. 1989 A version of the fundamental theorem for young measures. In *Partial Differential Equations and Continuum Models of Phase Transitions Lecture Notes in Physics*, eds. M. Rascle and D. Serre and M. Slemrod. Springer-Verlag **344**, 207–215.

12. Ball, J.M. and James, R.D. 1987 Fine phase mixtures as minimizers of energy. *Archive for Rational Mechanics and Analysis* **100**, 13–52.

13. Ball, J.M. and James, R.D. 1992 Proposed experimental tests of a theory of fine microstructure and the two-well problem. *Phil. Trans. R. Soc. London A* **338**, 389–450.

14. Bhattacharya, K. 1991 Wedge-like microstructure in martensite. *Acta Metallurgica Materialia* **10**, 2431–2444.

15. Bhattacharya, K. 2000 Theory of Martensitic Microstructure and the Shape-Memory Effect, monograph in preparation.

16. Born, M. and Huang, K. 1954 *Dynamical Theory of Crystal Lattices*, Oxford University Press.

17. Bowles, J.S. and Mackenzie, J.K. 1954 The crystallography of martensite transformations. *Acta Metall.* **2**, 129–147.

18. Callen, H.B. 1985 *Thermodynamics and an Introduction to Thermostatistics*, Sections 14-2, 14-3, Wiley, New York.

19. Christian, J.W. 1975 *The Theory of Martensitic Transformations in Metals and Alloys, Part 1*, Pergamon.

20. Chu, C. 1993 Hysteresis and Microstructures: A Study of Biaxial Loading on Compound Twins of Copper-Aluminum-Nickel Single Crystals. Ph.D. dissertation, University of Minnesota, Minneapolis, MN.

21. Duerig, T., Melton, K., Stoeckel, D. and Wayman, C.M. 1990 *Engineering Aspects of Shape Memory Alloys*, Butterworth-Heinemann.

22. Eshelby, J.D. 1956 The continuum theory of lattice defects. In *Solid State Physics*, eds. F. Seitz and D. Turnbull. Academic Press, 79–144.

23. Eshelby, J.D. 1970 Energy relations and the energy-momentum tensor in continuum mechanics. In *Inelastic Behavior of Solids*, eds. by M.F. Kanninen et al. McGraw-Hill, 77–115.

24. Ericksen, J.L. 1975 Equilibrium of bars. *Journal of Elasticity* **5**, 191–201.

25. Ericksen, J.L. 1978 On the symmetry and stability of thermoelastic solids. *ASME Journal of Applied Mechanics* **45**, 740–744.

26. Ericksen, J.L. 1980 Some phase transitions in crystals. *Archive for Rational Mechanics and Analysis* **73**, 99–124.

27. Ericksen, J.L. 1984 The Cauchy and Born hypotheses for crystals. In *Phase Transformations and Material Instabilities in Solids* , ed. M.E. Gurtin. Academic Press, 61–78.

28. Ericksen, J.L. 1986 Constitutive theory for some constrained elastic crystals. *International Journal of Solids Structures* **22**, 951–964.

29. Escobar, J.C. 1995 Plate Impact Induced Phase Transformations in Cu-Al-Ni Single Crystals. Ph.D. dissertation, Brown University, Providence, RI.

30. Escobar, J.C. and Clifton, R.J. 1993 On pressure-shear plate impact for studying the kinetics of stress-induced phase transformations. *Journal of Materials Science and Engineering* **A170**, 125–142.

31. Escobar, J.C. and Clifton, R.J. 1995 Pressure-shear impact-induced phase transitions in Cu-14.4 Al-4.19 Ni single crystals. *SPIE* **2427**, 186–197.

32. Gelfand, I.M. and Fomin, S.V. 1963 *Calculus of Variations*. Prentice-Hall.

33. Green, A.E. and Adkins, J.E. 1970 *Large Elastic Deformations*. Oxford University Press.

34. Grinfeld, M.A. 1981 On heterogeneous equilibrium of nonlinear elastic phases ad chemical potential tensors. *Lett. Appl. Eng. Sci.* **19**, 1031–1039.

35. Grujicic, M., Olson G.B. and Owen, W.S. 1985 Mobility of the $\beta_1 - \gamma_1'$ martensitic interface in Cu-Al-Ni: part I. Experimental measurements. *Metallurgical Transactions A* **16A**, 1723–1734.

36. Gurtin, M.E. 2000 *Configuration Forces as Basic Concepts of Continuum Physics*. Springer.

37. Heidug, W. and Lehner, F.K. 1985 Thermodynamics of coherent phase transformations in non-hydrostatically stressed solids. *Pure Appl. Geophysics* **123**, 91–98.

38. James, R.D. 1981 Finite deformation by mechanical twinning. *Arch. Rational Mech. Anal.* **77**, 143–176.

39. James, R.D. 1992 Lecture Notes for AEM 8589: Mechanics of Crystalline Solids, (unpublished). University of Minnesota.

40. James, R.D. and Hane, K.F. 2000 Martensitic transformations and shape-memory materials. *Acta Materialia* **48**, 197–222.

41. Kestin, J. 1979 *A Course in Thermodynamics*. Section 14.3.1, Volume II. New York: McGraw-Hill.

42. Kloućek, P. and Luskin, M. 1994 The computation of the dynamics of the martensitic transformation. *Continuum Mech. Thermodyn.* **6**, 209–240.

43. Kyriakides, S. and Shaw, J.A. 1995 Thermomechanical Aspects of NiTi. *Journal of the Mechanics and Physics of Solids* **43**, 1243–1281.

44. Müller, I. and Xu, H. 1991 On the pseudoelatstic hysteresis. *Acta Metalurgica Materialia* **39**, 263–271.

45. Oleinik, O.A. 1957 On the uniqueness of the generalized solutions of the Cauchy problem for a nonlinear system of equations occuring in mechanics. *Uspekhi Matematicheskii Nauk (N.S.)* **12**, 169–176.

46. Otsuka, K., and Wayman, C.M. 1999 *Shape-memory Alloys*. Cambridge University Press.

47. Pitteri, M. 1984 Reconciliation of local and global symmetries of crystals. *Journal of Elasticity* **14**, 175–190.

48. Schetky, L.M. 1979 Shape-memory alloys. *Scientific American* **241**, 74–82.

49. Smith, G.F. and Rivlin, R.S. 1958 The strain energy function for anisotropic elastic materials. *Transactions of the American Mathematical Society* **88**, 175–193.

50. Truesdell, C. 1969 *Rational Thermodynamics*. New York: McGraw-Hill.

51. Truskinovsky, L. 1985 Structure of an isothermal discontinuity. *Soviet Phys. Dokl.* **30**, 945–948.

52. Wechsler, M.S., Lieberman, D.S. and Read, T.A. 1953 On the theory of the formation of martensite. *Transactions of the AIME Journal of Metals* **197**, 1503–1515.