PERGAMON

Journal of the Mechanics and Physics of Solids
47 (1999) 531–576

JOURNAL OF THE
MECHANICS AND
PHYSICS OF SOLIDS

A theory of thin films of martensitic materials with applications to microactuators[1]

K. Bhattacharya[a], R.D. James[b,*]

[a] *Division of Engineering and Applied Sciences, Mail Code 104-44, California Institute of Technology, Pasadena CA 91125, U.S.A.*

[b] *Department of Aerospace Engineering and Mechanics, 107 Akerman Hall, University of Minnesota, Minneapolis MN 55455, U.S.A.*

Received 1 October 1997; in revised form 7 February 1998

Abstract

A direct derivation is given of a theory for single crystal thin films, starting from three-dimensional nonlinear elasticity theory augmented by a term for interfacial energy. The derivation involves no *a priori* choice of asymptotic expansion or ansatz. It yields a frame-indifferent Cosserat membrane theory with one Cosserat vector field. The theory is applied to multi-well energy functions appropriate to martensitic materials. It is found that, unlike in bulk materials, which generally only support finely twinned austenite/martensite interfaces as energy minimizing states, the thin film theory predicts the existence of exact, untwinned austenite/martensite interfaces. These are used to construct some simple energy minimizing deformations—"tents" and "tunnels"—that could possibly be the basis of simple large-deformation microactuators. Explicit results are given for martensitic materials in the systems NiMnGa, NiTi, NiTiCu, and NiAl. A certain alloy of precise composition $Ni_{30.5}Ti_{49.5}Cu_{20.0}$ is predicted to support a four-sided "tent" on an (001) film, which furthermore is predicted to collapse to the substrate upon heating. A formal derivation is given of higher order theories, which yields two additional Cosserat vectors and an explicit form of the bending energy. The derivation indicates an approach to plate-shell-thin film theories that is rather different from the ones usually followed.

Keywords: A. Phase transformation; Shape memory effect; B. Thin films; Microactuator; C. Variational calculus

1. Introduction

In this paper we give a direct derivation of a theory of deformable thin films beginning from three-dimensional nonlinear elasticity, augmented with a classical

* Corresponding author. Fax: 001 612 626 1588.
E-mail address: james@aem.umn.edu. (R. D. James)

[1] Dedicated to the memory of Juan Simo.

0022–5096/99/$ - see front matter
PII: S0022–5096(98)00043–X

term for interfacial energy, and apply the theory to the behavior of thin films of martensitic material.

Martensitic materials undergo a diffusionless phase transformation at a certain temperature θ_c. Typically, a symmetric high temperature phase (austenite) experiences a spontaneous distortion upon cooling through θ_c; the low temperature phase (martensite) consists of different variants which are symmetry-related via symmetry operations of the austenite. Crystals that undergo a reversible martensitic phase transformation often exhibit the shape-memory effect. Here, an imposed deformation at low temperature, that rearranges the variants of martensite, is recovered upon heating as the crystal returns to the unique austenitic structure.

As discussed by Krulevitch et al. (1995, 1996), actuators which utilize the shape-memory effect exhibit the largest work output per cycle per volume among a variety of actuator systems, including ordinary electromagnetic, piezoelectric, giant magnetostrictive, liquid–solid phase change and muscle. This fact, together with the enhanced rate of heat transfer in thin films, makes shape-memory actuators attractive for miniaturization. Thin films of the shape-memory material NiTi and closely related alloys have been made by magnetron sputtering by several groups (Grummon et al., 1995; Krulevitch et al., 1996; Miyazaki et al., 1995; Mathews, 1996, and references therein). This method produces polycrystalline films of special texture, dominated by (110) in NiTi. From studies of the shape-memory effect in polycrystals (Bhattacharya and Kohn, 1996), it is well-known that, unless the martensite is of very low symmetry or the polycrystal has special texture, the deformations possible in a sample will be highly restricted. In fact, Shu and Bhattacharya (1997) show theoretically that the (110) texture in films of NiTi is somewhat unfavorable for having large recoverable strains.

This raises the question of whether other methods yielding different textures in thin films would be interesting to pursue. The ideal texture that eliminates altogether "fighting between the grains" is single crystalline. Single crystal films can potentially be produced by epitaxial growth on a lattice-matched single crystal substrate, and released by a subsequent back etch. In fact, by patterning the back of the substrate, the film can be released on precisely defined regions. This possibility partly motivated the present study. In Section 5 we suggest several designs for certain "tents" and "tunnels" that take advantage of patterning single crystal films, and our calculations indicate that this line of research will be interesting to pursue. Very recently, Bensaoula et al. (1997) have grown single crystals of the martensitic material Ni_2MnGa on $Ga_{(1-x}In_{x)}As$ by MBE.

Since it is essential for these applications to know exactly how the thin film theory is related to the 3-D theory, and also because we were unsure which of the wide array of different plate theories would be applicable in this case, we adopted the viewpoint explored in recent years by several authors (e.g. Acerbi et al., 1991; Anzellotti et al., 1994; LeDret and Raoult, 1993, 1995, 1996; Fonseca and Francfort, 1998) in related cases. That is, we begin by writing the total energy per unit thickness $e^{(h)}$ of a film of thickness h and change variables so that the competing deformations are defined on a fixed domain. Our energy contains a contribution for interfacial energy, which is expected to possibly be important for very thin films. Calling a minimizer of this

energy $\mathbf{y}^{(h)}$, we try to extract from the sequence $\mathbf{y}^{(h)}$ various quantities that are necessary and sufficient for the determination of the limiting energy $\lim_{h\to 0} e^{(h)}$, under a wide class of boundary conditions and for a general form of the bulk energy density. In our case it turns out [eqn (3.8)] that the limiting energy e^0 is determined by two vector fields $\mathbf{y}$ and $\mathbf{b}$ defined on a plane sheet, and has the form of a Cosserat membrane theory with $\mathbf{y}$ being associated with deformations of a "middle surface" and the Cosserat vector $\mathbf{b}$ with transverse shear and normal compression. The energy e^0 is frame-indifferent and inherits a certain energy-well structure from the original theory which we exploit in the applications. This method of derivation is associated with Γ-convergence, but we do not use any of the abstract techniques of Γ-convergence in this paper.

An interesting feature of the thin film energy (with surface energy put equal to zero) is that it admits a large family of exact energy minimizing austenite/martensite interfaces (Section 5). This contrasts with results on austenite/martensite interfaces in bulk (Ball and James, 1987) which, without restrictive conditions on lattice parameters, are necessarily finely twinned. These exact austenite/martensite interfaces form the basis of the construction of the tents and tunnels. Necessary and sufficient conditions on the lattice parameters, film normal and geometry for these to be possible are given in Sections 5.3. and 5.4. Some martensitic materials that satisfy the various conditions, with their associated film orientations, are given in Section 7.

If we begin with our e^0, put the interfacial energy terms equal to zero, minimize out the Cosserat vector $\mathbf{b}$, and then relax, we get a membrane energy having the same form as that of LeDret and Raoult. Because we have included interfacial energy of the van der Waals type, our derivation of this thin film theory is in fact easier than previous treatments, but also applies to a wider class of bulk energies. In addition we are able to obtain formally the bending terms at higher order (Section 8).

We also quote the results of a similar derivation in the case of a film and substrate of comparable thickness (Section 4). The limiting energy in this case involves a deformation $\mathbf{y}$ and two Cosserat vectors $\mathbf{b}_f$ and $\mathbf{b}_s$. By varying the thickness ratio of the film and substrate, the energy-well structure of the composite can be changed considerably. We propose this as a method of satisfying special conditions on lattice parameters that, for example, allow a "tent" to be energy minimizing at low temperature.

The derivation of the bending energy and its associated variational principle begins by considering the difference $(e^{(h)}[\mathbf{y}^{(h)}] - e^0[\bar{\mathbf{y}}, \bar{\mathbf{b}}])/h^v$, where $(\bar{\mathbf{y}}, \bar{\mathbf{b}})$ minimize the thin film energy. The first nontrivial terms occur at $v = 2$, and represent bending and higher order interfacial energy. The theory that emerges is frame-indifferent and has two new Cosserat vectors $(\mathbf{c}, \mathbf{d})$. By repeating the calculation using certain kinds of competitors, we also obtain a variational principle. The derivation says that one should do something quite different from the usual procedure: first minimize the membrane energy e^0 over pairs $(\mathbf{y}, \mathbf{b})$ in an appropriate space, use the result to evaluate the coefficients in the bending energy, and then minimize the bending energy over the additional Cosserat vectors $(\mathbf{c}, \mathbf{d})$. In some sense such a procedure is implicitly used when one adopts assumptions of inextensibility, or rigidity of cross-sections, in special plate theories. Our derivation of the bending theory is presented as a formal cal-

culation to keep the length of this paper manageable, but it appears that it can be made rigorous by introducing (unfortunately rather strong) conditions of growth.

Some of our results were given without proof in Bhattacharya and James (1996). Basic notation: $\alpha, \beta, \gamma, x_1, x_2, \ldots$, are scalars; $\mathbf{a}, \mathbf{b}, \mathbf{c}, \ldots$, and the term "vector" denote members of $\mathbb{R}^3$; $\mathbf{A}, \mathbf{B}, \mathbf{C}, \ldots$, and the term "matrix" denote elements of $M^{3\times 3}$, that is, 3×3 matrices, and a superimposed T is transpose; $\mathbf{A} = (\mathbf{a}|\mathbf{b}|\mathbf{c})$ denotes a matrix whose first column is $\mathbf{a}$, second column is $\mathbf{b}$ and third column is $\mathbf{c}$; $SO(3) = \{\mathbf{R} \in M^{3\times 3} : \mathbf{R}^T\mathbf{R} = \mathbf{I}, \det \mathbf{R} = +1\}$ and elements of $SO(3)$ are called rotations; $\nabla^2\mathbf{y}$ is the full matrix of second derivatives of $\mathbf{y}$, not just the Laplacean.

2. Bulk theory of martensite

We assume that the behavior of the film is governed by a free energy of the geometrically nonlinear theory of martensite (Ball and James, 1992, Section 2), augmented by a standard form of the interfacial energy. The use of geometrically nonlinear theory is essential for two reasons: (1) we wish to treat shape-memory materials like NiTi and CuZnAl where geometrically linear theories can make very substantial errors (Bhattacharya, 1993), and (2) we want to apply the resulting theory to films that may deform by undergoing large rotations, as for example in the tunnels described in Section 5. The interfacial energy is also essential for thin film problems; in particular, we are interested in the effects of scale, within the continuum framework.

The bulk free energy density shall be a smooth function $\varphi : M^{3\times 3} \times (0, \infty) \to (0, \infty)$, where $\varphi(\mathbf{A}, \theta)$ represents the free energy per unit reference volume of the material as a function of the deformation gradient $\mathbf{A}$ and temperature θ. The free energy function is assumed to be frame-indifferent:

$$\varphi(\mathbf{QA}, \theta) = \varphi(\mathbf{A}, \theta) \quad \forall \mathbf{Q} \in SO(3), \quad \mathbf{A} \in M^{3\times 3}, \quad \theta > 0, \tag{2.1}$$

and to satisfy the following conditions of material symmetry:

$$\varphi(\mathbf{AH}, \theta) = \varphi(\mathbf{A}, \theta) \quad \forall \mathbf{H} \in P, \quad \mathbf{A} \in M^{3\times 3}, \quad \theta > 0, \tag{2.2}$$

where P is the point group of the austenite phase at the transformation temperature θ_c. Here, θ_c can be taken to be the temperature at which the austenite and martensite have the same free energy density. The justification of these conditions comes from an underlying crystallographic model (Ball and James, 1992). This involves embedding a lattice model into the continuum model via the Cauchy–Born rule, and following through the symmetry conditions that arise from the underlying lattice. For our purposes here, we shall not need anything from the basic theory beyond (2.1) and (2.2).

We assume growth hypotheses

$$c_1(|\mathbf{A}|^2 - 1) \leqslant \varphi(\mathbf{A}, \theta) \leqslant c_2(|\mathbf{A}|^q - 1) \tag{2.3}$$

where $2 < q < 6$. While (2.3) forbids the natural physical hypothesis $\varphi(\mathbf{A}, \theta) \to \infty$ as $\det \mathbf{A} \to 0$, this would appear to be more of a technicality than a substantive objection.

To model the presence of different phases, φ shall be assigned energy wells. We assume that there are a finite number of positive-definite symmetric matrices $\mathbf{U}_0, \mathbf{U}_1, \ldots, \mathbf{U}_n$ such that

(i) For $\theta > \theta_c$, $\varphi(\cdot, \theta)$ is minimized on

$$SO(3)\mathbf{U}_0. \tag{2.4}$$

(ii) For $\theta \leqslant \theta_c$, $\varphi(\cdot, \theta)$ is minimized on

$$SO(3)\mathbf{U}_1 \cup SO(3)\mathbf{U}_2 \cup \cdots SO(3)\mathbf{U}_n. \tag{2.5}$$

The matrices $\mathbf{U}_1, \ldots, \mathbf{U}_n$ are the distortion matrices, representing the linear transformations that take the undistorted austenite lattice at θ_c to the lattices associated with the n variants of martensite. The notation $SO(3)\mathbf{U}$ stands for the set of all rotation matrices post-multiplied by $\mathbf{U}$, i.e. $SO(3)\mathbf{U} = \{\mathbf{A} \in M^{3\times 3} : \mathbf{A} = \mathbf{QU}, \mathbf{Q}^T\mathbf{Q} = \mathbf{I}, \det \mathbf{Q} = +1\}$. Though it is not reflected by the notation, all the matrices $\mathbf{U}_0, \mathbf{U}_1, \ldots, \mathbf{U}_n$ are assumed to depend (weakly) on temperature, representing ordinary thermal expansion of the austenite or of the individual variants of martensite, and we have by assumption that $\mathbf{U}_0 = \mathbf{I}$ at $\theta = \theta_c$. We also assume that (2.4) and (2.5) are consistent with the symmetries (2.1) and (2.2) in the sense that,

$$\mathbf{U}_0 = \mathbf{R}\mathbf{U}_0\mathbf{R}^T \quad \text{for all } \mathbf{R} \in P, \quad \{\mathbf{U}_1, \ldots, \mathbf{U}_n\} = \{\mathbf{R}\mathbf{U}_1\mathbf{R}^T : \mathbf{R} \in P\}. \tag{2.6}$$

The forms of the matrices $\mathbf{U}_0, \mathbf{U}_1, \ldots, \mathbf{U}_n$ for various symmetry changes have been worked out by Pitteri and Zanzotto (1996). To illustrate our results, we shall focus on four common cases having cubic austenite (all matrices in (2.7)–(2.11) below are expressed in an orthonormal basis parallel to the cubic axes, the *cubic basis*):

(A) Cubic to tetragonal transformations. Here there are two lattice parameters $\eta_1 > 0$ and $\eta_2 > 0$, $\eta_1 \neq \eta_2$, and three variants of martensite with distortion matrices of the form:

$$\begin{pmatrix} \eta_2 & & \\ & \eta_1 & \\ & & \eta_1 \end{pmatrix}, \quad \begin{pmatrix} \eta_1 & & \\ & \eta_2 & \\ & & \eta_1 \end{pmatrix}, \quad \begin{pmatrix} \eta_1 & & \\ & \eta_1 & \\ & & \eta_2 \end{pmatrix}. \tag{2.7}$$

The measured values of these parameters for $Ni_{64}Al_{36}$ are $\eta_1 = 0.9392, \eta_2 = 1.1302$ and for Ni_2MnGa are $\eta_1 = 0.9512, \eta_2 = 1.130$.

(B) Cubic to orthorhombic transformation as in the $\beta_1 \to \gamma_1'$ transformation in CuAlNi. Here there are three lattice parameters $\alpha > 0$, $\beta > 0$, $\gamma > 0$, $\alpha \neq \gamma$, and six variants of martensite with distortion matrices of the form:

$$\begin{pmatrix} \frac{\alpha+\gamma}{2} & \frac{\alpha-\gamma}{2} & 0 \\ \frac{\alpha-\gamma}{2} & \frac{\alpha+\gamma}{2} & 0 \\ 0 & 0 & \beta \end{pmatrix}, \quad \begin{pmatrix} \frac{\alpha+\gamma}{2} & 0 & \frac{\alpha-\gamma}{2} \\ 0 & \beta & 0 \\ \frac{\alpha-\gamma}{2} & 0 & \frac{\alpha+\gamma}{2} \end{pmatrix}, \quad \begin{pmatrix} \beta & 0 & 0 \\ 0 & \frac{\alpha+\gamma}{2} & \frac{\alpha-\gamma}{2} \\ 0 & \frac{\alpha-\gamma}{2} & \frac{\alpha+\gamma}{2} \end{pmatrix},$$

$$\begin{pmatrix} \frac{\alpha+\gamma}{2} & \frac{\gamma-\alpha}{2} & 0 \\ \frac{\gamma-\alpha}{2} & \frac{\alpha+\gamma}{2} & 0 \\ 0 & 0 & \beta \end{pmatrix}, \quad \begin{pmatrix} \frac{\alpha+\gamma}{2} & 0 & \frac{\gamma-\alpha}{2} \\ 0 & \beta & 0 \\ \frac{\gamma-\alpha}{2} & 0 & \frac{\alpha+\gamma}{2} \end{pmatrix}, \quad \begin{pmatrix} \beta & 0 & 0 \\ 0 & \frac{\alpha+\gamma}{2} & \frac{\gamma-\alpha}{2} \\ 0 & \frac{\gamma-\alpha}{2} & \frac{\alpha+\gamma}{2} \end{pmatrix} \tag{2.8}$$

The measured values of these parameters for CuAlNi (14.2wt%Al–4.3wt%Ni) are $\alpha = 1.0619$, $\beta = 0.9178$, $\gamma = 1.0230$ and for $Ni_{30.5}Ti_{49.5}Cu_{20.0}$ are $\alpha = 1.0$, $\beta = 0.9579$, $\gamma = 1.0583$ to within ± 0.0001 (Moberly, 1989).

(C1) Cubic to monoclinic transformation as in the $DO_3 \rightarrow 6M$ (formerly called $18R$) transformation in CuZnAl. Here there are four lattice parameters $\alpha > 0$, $\beta > 0$, $\gamma > 0$, $\delta \neq 0$, $\alpha\gamma - \delta^2 > 0$, and twelve variants of martensite with distortion matrices of the form:

$$\begin{pmatrix} \beta & 0 & 0 \\ 0 & \alpha & \delta \\ 0 & \delta & \gamma \end{pmatrix}, \begin{pmatrix} \beta & 0 & 0 \\ 0 & \alpha & -\delta \\ 0 & -\delta & \gamma \end{pmatrix}, \begin{pmatrix} \beta & 0 & 0 \\ 0 & \gamma & \delta \\ 0 & \delta & \alpha \end{pmatrix}, \begin{pmatrix} \beta & 0 & 0 \\ 0 & \gamma & -\delta \\ 0 & -\delta & \alpha \end{pmatrix},$$

$$\begin{pmatrix} \alpha & 0 & \delta \\ 0 & \beta & 0 \\ \delta & 0 & \gamma \end{pmatrix}, \begin{pmatrix} \alpha & 0 & -\delta \\ 0 & \beta & 0 \\ -\delta & 0 & \gamma \end{pmatrix}, \begin{pmatrix} \gamma & 0 & \delta \\ 0 & \beta & 0 \\ \delta & 0 & \alpha \end{pmatrix}, \begin{pmatrix} \gamma & 0 & -\delta \\ 0 & \beta & 0 \\ -\delta & 0 & \alpha \end{pmatrix},$$

$$\begin{pmatrix} \alpha & \delta & 0 \\ \delta & \gamma & 0 \\ 0 & 0 & \beta \end{pmatrix}, \begin{pmatrix} \alpha & -\delta & 0 \\ -\delta & \gamma & 0 \\ 0 & 0 & \beta \end{pmatrix}, \begin{pmatrix} \gamma & \delta & 0 \\ \delta & \alpha & 0 \\ 0 & 0 & \beta \end{pmatrix}, \begin{pmatrix} \gamma & -\delta & 0 \\ -\delta & \alpha & 0 \\ 0 & 0 & \beta \end{pmatrix}. \tag{2.9}$$

The measured values of these parameters for $Cu_{68}Zn_{15}Al_{17}$ (Chakravorty and Wayman, 1977; Hane, 1997) are $\alpha = 1.087$, $\beta = 0.9093$, $\gamma = 1.010$, $\delta = 0.0250$. This corresponds to a monoclinic angle of 94.2°.

(C2) Cubic to monoclinic transformation as in NiTi. Here, there are four lattice parameters $\alpha > 0$, $\beta > 0$, δ, $\varepsilon \neq 0$, $\alpha^2 - \delta^2 > 0$, $\alpha\beta - \varepsilon^2 > 0$,

$$\begin{pmatrix} \alpha & \delta & \varepsilon \\ \delta & \alpha & \varepsilon \\ \varepsilon & \varepsilon & \beta \end{pmatrix}, \begin{pmatrix} \alpha & \delta & -\varepsilon \\ \delta & \alpha & -\varepsilon \\ -\varepsilon & -\varepsilon & \beta \end{pmatrix}, \begin{pmatrix} \alpha & -\delta & -\varepsilon \\ -\delta & \alpha & \varepsilon \\ -\varepsilon & \varepsilon & \beta \end{pmatrix}, \begin{pmatrix} \alpha & -\delta & \varepsilon \\ -\delta & \alpha & -\varepsilon \\ \varepsilon & -\varepsilon & \beta \end{pmatrix},$$

$$\begin{pmatrix} \alpha & \varepsilon & \delta \\ \varepsilon & \beta & \varepsilon \\ \delta & \varepsilon & \alpha \end{pmatrix}, \begin{pmatrix} \alpha & -\varepsilon & \delta \\ -\varepsilon & \beta & -\varepsilon \\ \delta & -\varepsilon & \alpha \end{pmatrix}, \begin{pmatrix} \alpha & -\varepsilon & -\delta \\ -\varepsilon & \beta & \varepsilon \\ -\delta & \varepsilon & \alpha \end{pmatrix}, \begin{pmatrix} \alpha & \varepsilon & -\delta \\ \varepsilon & \beta & -\varepsilon \\ -\delta & -\varepsilon & \alpha \end{pmatrix},$$

$$\begin{pmatrix} \beta & \varepsilon & \varepsilon \\ \varepsilon & \alpha & \delta \\ \varepsilon & \delta & \alpha \end{pmatrix}, \begin{pmatrix} \beta & -\varepsilon & -\varepsilon \\ -\varepsilon & \alpha & \delta \\ -\varepsilon & \delta & \alpha \end{pmatrix}, \begin{pmatrix} \beta & -\varepsilon & \varepsilon \\ -\varepsilon & \alpha & -\delta \\ \varepsilon & -\delta & \alpha \end{pmatrix}, \begin{pmatrix} \beta & \varepsilon & -\varepsilon \\ \varepsilon & \alpha & -\delta \\ -\varepsilon & -\delta & \alpha \end{pmatrix}. \tag{2.10}$$

The measured values of these parameters for $Ni_{50}Ti_{50}$ are $\alpha = 1.0243$, $\beta = 0.9563$, $\delta = 0.058$, $\varepsilon = 0.0427$ (Knowles and Smith, 1981).

Now we turn to the description of the surface energy. By far, the most popular method for modeling surface energy is to add a term of the form $\nabla^2\mathbf{y} \cdot \mathbf{A}\nabla^2\mathbf{y}$ to the energy density, where $\mathbf{A}$ is positive–definite and satisfies conditions of crystallographic symmetry appropriate to the austenite. Here $\nabla^2\mathbf{y}$ denotes the full $3 \times 3 \times 3$ matrix of the second derivatives, not just the Laplacean. There are indications (Barsch and Krumhansl, 1984, 1988) that such a model, when fitted carefully with measured material constants, predicts twin boundaries that are more diffuse than the corresponding observed interfaces. However, this model has one property in its favor: it scales exactly like interfacial energy under uniform dilatations. Since we mainly rely on its scaling and smoothing properties, we use the simple form $\kappa|\nabla^2\mathbf{y}|^2$ in our subsequent calculations.

The thin film occupies a reference domain

$$\Omega_h = \left\{\mathbf{x} \in \mathbb{R}^3 : (x_1, x_2) \in \mathrm{S}, -\frac{h}{2} < x_3 < \frac{h}{2}\right\}, \tag{2.11}$$

where S is, for simplicity, a bounded Lipschitz domain (i.e., corners allowed but no cusps) with unit area. Deformations of the film are described by functions $\mathbf{y} : \Omega_h \to \mathbb{R}^3$. In (2.11) the components x_i are relative to an orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ (the *film basis*). The relation between the film basis and the cubic basis describes the orientation of the film.

The total free energy is assumed to have the final form

$$e^{(h)}[\mathbf{y}; \theta] = \int_{\Omega_h} \{\kappa|\nabla^2\mathbf{y}|^2 + \varphi(\nabla\mathbf{y}, \theta)\}\, \mathrm{d}x. \tag{2.12}$$

For a typical deformation $\mathbf{y} : \Omega_h \to \mathbb{R}^3$ satisfying our boundary conditions (see below) the energy $e^{(h)}$ will scale like h as $h \to 0$, so we shall be interested in the limiting behavior of the energy per unit reference thickness

$$(1/h)e^{(h)}[\mathbf{y}; \theta]. \tag{2.13}$$

In the following arguments, θ shall be held fixed, so we suppress it from the notation.

3. Derivation of the thin film theory

Consider the energy $(1/h)e^{(h)}$ of (2.13). It is convenient to work on a fixed domain, so we change variables (Fig. 1):

$$z_1 = x_1, \quad z_2 = x_2, \quad z_3 = \frac{1}{h}x_3, \quad \mathbf{x} \in \Omega_h. \tag{3.1}$$

To each deformation $\tilde{\mathbf{y}} : \Omega_h \to \mathbb{R}^3$ we associated a deformation $\mathbf{y} : \Omega_1 \to \mathbb{R}^3$ via

$$\mathbf{y}(\mathbf{z}(\mathbf{x})) = \tilde{\mathbf{y}}(\mathbf{x}), \quad \mathbf{x} \in \Omega_h. \tag{3.2}$$

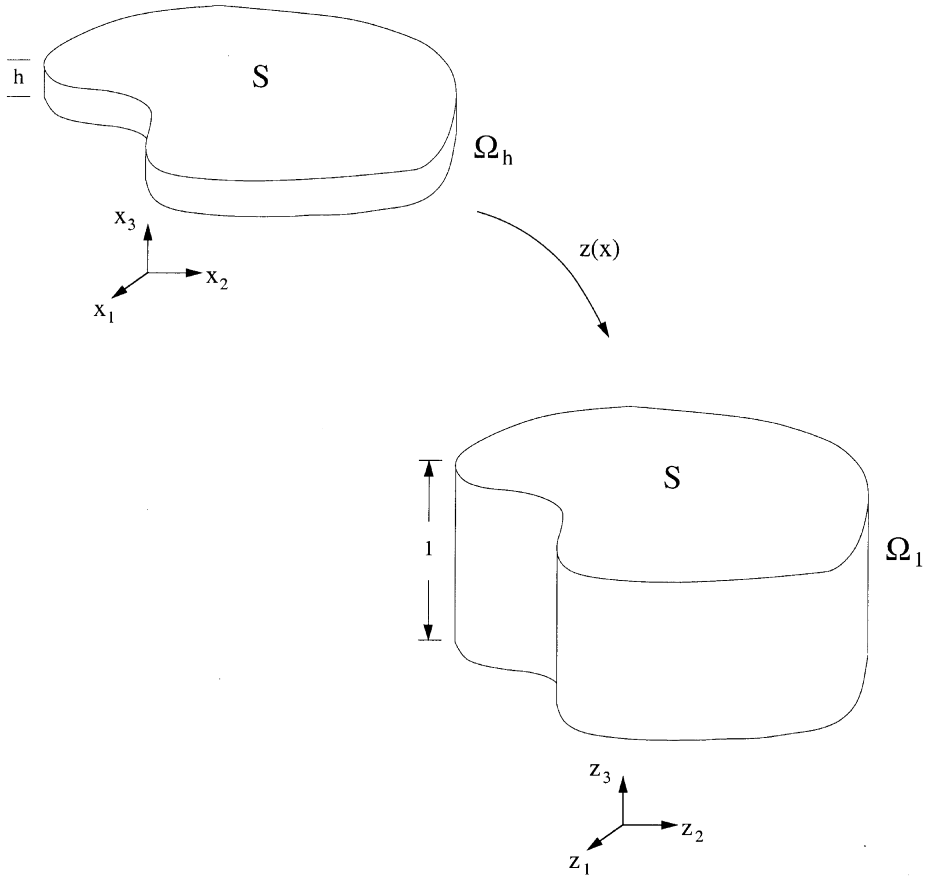

Fig. 1. Change of variables given in (3.1).

We use the notation ∇_p for the gradient in the plane of the film, e.g.

$$\nabla_p \mathbf{y} = \mathbf{y}_{,1} \otimes \mathbf{e}_1 + \mathbf{y}_{,2} \otimes \mathbf{e}_2. \tag{3.3}$$

Also, it is useful to express the bulk free energy density $\varphi(\mathbf{A})$ as a function of the three columns of $\mathbf{A}$: with a slight abuse of notation we shall write $\varphi(\mathbf{a}_1|\mathbf{a}_2|\mathbf{a}_3)$ for $\varphi(\mathbf{a}_1 \otimes \mathbf{e}_1 + \mathbf{a}_2 \otimes \mathbf{e}_2 + \mathbf{a}_3 \otimes \mathbf{e}_3)$.

We now change variables in $(1/h)e^{(h)}$ using (3.1) and (3.2). This gives

$$e_1^{(h)}[\mathbf{y}] := \frac{1}{h} e^{(h)}[\tilde{\mathbf{y}}] = \int_{\Omega_1} \left\{ \kappa \left[|\nabla_p^2 \mathbf{y}|^2 + \frac{2}{h^2} |\nabla_p \mathbf{y}_{,3}|^2 + \frac{1}{h^4} |\mathbf{y}_{,33}|^2 \right] + \varphi\left(\mathbf{y}_{,1} | \mathbf{y}_{,2} | \frac{1}{h} \mathbf{y}_{,3} \right) \right\} d\mathbf{z}. \tag{3.4}$$

We picture a film that has been released from the substrate on part of its lower boundary. The released boundary is here identified with S. A thick substrate is expected to highly constrain the part of the film that adheres to it, so we model its effect on the released film by a displacement boundary condition at the edge of the film, $\partial S \times (-(h/2), h/2)$. For simplicity we impose the linear boundary condition $\tilde{\mathbf{y}}(\mathbf{x}) = \mathbf{A}\mathbf{x}, \mathbf{x} \in \partial S \times [-(h/2), h/2]$. For example, if the adhering part of the film remains in the austenite phase, the simple boundary condition $\tilde{\mathbf{y}}(\mathbf{x}) = \mathbf{x}, \mathbf{x} \in \partial S \times (-(h/2), h/2)$ would be appropriate. In terms of the new variables, the linear boundary condition becomes,

$$\mathbf{y}(\mathbf{z}) = \mathbf{A}^{(h)}\mathbf{z}, \quad \mathbf{z} \in \partial S \times (-\tfrac{1}{2}, \tfrac{1}{2}), \quad \mathbf{A}^{(h)} = (\mathbf{a}_1 | \mathbf{a}_2 | h\mathbf{a}_3). \tag{3.5}$$

The derivation of the thin film theory below is not particularly sensitive to the use of linear boundary conditions, but their presence simplifies the proof (see Remark 3.2). The main property of boundary conditions used in the derivation given below is that they be consistent with the existence of a family of test functions that makes $e_1^{(h)}$ bounded independent of h.

The existence of a minimizer $\mathbf{y}^{(h)} \in W^{2,2}(\Omega_1, \mathbb{R}^3)$ of $e_1^{(h)}$ for each fixed $h > 0$ is guaranteed by the direct method of the calculus of variations. To see this, note first that $e_1^{(h)} \geqslant 0$. Let $\mathbf{y}_k^{(h)} \in W^{2,2}(\Omega_1, \mathbb{R}^3)$, $\mathbf{y}_k^{(h)}(\mathbf{z}) = \mathbf{A}^{(h)}\mathbf{z}$, $\mathbf{z} \in \partial S \times (-\frac{1}{2}, \frac{1}{2})$ be a minimizing sequence. Using the growth condition (2.3) on φ and the positivity of κ, we have

$$\begin{aligned} c_h \geqslant e_1^{(h)}[\mathbf{y}_k^{(h)}] &\geqslant \int_{\Omega_1} \varphi\left(\mathbf{y}_{k,1}^{(h)} | \mathbf{y}_{k,2}^{(h)} | \frac{1}{h}\mathbf{y}_{k,3}^{(h)}\right) \mathrm{d}\mathbf{z} \\ &\geqslant c\left(|\mathbf{y}_{k,1}^{(h)}|^2 + |\mathbf{y}_{k,2}^{(h)}|^2 + \frac{1}{h^2}|\mathbf{y}_{k,3}^{(h)}|^2 - 1\right). \end{aligned} \tag{3.6}$$

This gives $\|\nabla \mathbf{y}_k^{(h)}\|_{L^2} < c_h'$. Now using the Poincaré inequality in the form (Morrey, 1966, Theorem 3.6.4)

$$\int_{\Omega_1} |\mathbf{y}_k^{(h)}|^2 \, \mathrm{d}\mathbf{z} \leqslant c_1 \left[\int_{\Omega_1} |\nabla \mathbf{y}_k^{(h)}|^2 \, \mathrm{d}\mathbf{z} + \int_{\partial S \times (-(1/2), 1/2)} |\mathbf{y}_k^{(h)}|^2 \, \mathrm{d}\mathbf{z}\right], \tag{3.7}$$

and the boundary conditions, we conclude that $\|\mathbf{y}_k^{(h)}\|_{W^{2,2}} < c_h''$. Hence, for a suitable subsequence (not relabeled), $\mathbf{y}_k^{(h)} \rightharpoonup \mathbf{y}^{(h)}$ in $W^{2,2}(\Omega_1, \mathbb{R}^3)$ and (using Rellich's Theorem) $\mathbf{y}_k^{(h)} \to \mathbf{y}^{(h)}$ in $W^{1,q}(\Omega_1, \mathbb{R}^3), q < 6$.

Now using the convexity of the second gradient terms and the continuity of φ, together with the bound (2.3), we find that $e_1^{(h)}$ is weakly lower semicontinuous on $\{\mathbf{y}_k^{(h)}\}$. Thus, $\mathbf{y}^{(h)}$ is a minimizer. It satisfies the boundary conditions by virtue of the strong convergence $\mathbf{y}_k^{(h)} \to \mathbf{y}^{(h)}$ in $W^{1,q}$. Now we consider the behavior of the minimizers $\mathbf{y}^{(h)}$ as $h \to 0$.

Theorem 3.1. The family of minimizers $\mathbf{y}^{(h)} \in W^{2,2}(\Omega_1, \mathbb{R}^3)$ has a subsequence (not relabeled) such that

$$\left.\begin{aligned} \nabla_{\mathrm{p}}^2\mathbf{y}^{(h)} &\to \nabla_{\mathrm{p}}^2\bar{\mathbf{y}} \\ \frac{1}{h}\nabla_{\mathrm{p}}\mathbf{y}_{,3}^{(h)} &\to \nabla_{\mathrm{p}}\bar{\mathbf{b}} \\ \frac{1}{h^2}\mathbf{y}_{,33}^{(h)} &\to 0 \end{aligned}\right\} \text{ in } L^2(\Omega_1), \tag{3.8}$$

where $(\bar{\mathbf{y}}, \bar{\mathbf{b}})$ are independent of z_3 and $(\bar{\mathbf{y}}, \bar{\mathbf{b}})$ minimizes the limiting energy

$$e^0[\mathbf{y}, \mathbf{b}] = \int_{\mathrm{S}} \{\kappa(|\nabla_{\mathrm{p}}^2\mathbf{y}|^2 + 2|\nabla_{\mathrm{p}}\mathbf{b}|^2) + \varphi(\mathbf{y}_{,1}|\mathbf{y}_{,2}|\mathbf{b})\}\, \mathrm{d}z_1\, \mathrm{d}z_2 \tag{3.9}$$

among $(\mathbf{y}, \mathbf{b}) \in W^{2,2}(\mathrm{S}, \mathbb{R}^3) \times W^{1,2}(\mathrm{S}, \mathbb{R}^3)$ satisfying the boundary conditions

$$\left.\begin{aligned} \mathbf{y}(z_1, z_2) &= \mathbf{a}_1 z_1 + \mathbf{a}_2 z_2 \\ \mathbf{b}(z_1, z_2) &= \mathbf{a}_3 \end{aligned}\right\} \quad (z_1, z_2) \in \partial\mathrm{S}. \tag{3.10}$$

Proof. Compare the energy of $\mathbf{y}^{(h)}$ with that of $\mathbf{y}(\mathbf{z}) = \mathbf{A}^{(h)}\mathbf{z}$. This gives

$$e_1^{(h)}[\mathbf{y}^{(h)}] \leqslant |\Omega_1|\varphi(\mathbf{a}_1|\mathbf{a}_2|\mathbf{a}_3) \tag{3.11}$$

so that

$$\|\nabla_{\mathrm{p}}^2\mathbf{y}^{(h)}\|_{L^2} \leqslant c, \quad \left\|\frac{1}{h}\nabla_{\mathrm{p}}\mathbf{y}_{,3}^{(h)}\right\|_{L^2} \leqslant c, \quad \left\|\frac{1}{h^2}\mathbf{y}_{,33}^{(h)}\right\|_{L^2} \leqslant c. \tag{3.12}$$

It also follows from (3.11) that, on the left hand side of (3.6), we can choose c_h independent of h. Hence, using (3.11) and the Poincaré inequality in the form (3.7), we can add to the list (3.12) the conditions

$$\|\mathbf{y}^{(h)}\|_{L^2} \leqslant c, \quad \|\nabla\mathbf{y}^{(h)}\|_{L^2} \leqslant c, \quad \left\|\frac{1}{h}\mathbf{y}_{,3}^{(h)}\right\|_{L^2} \leqslant c, \tag{3.13}$$

so that $\|\mathbf{y}^{(h)}\|_{W^{2,2}} \leqslant \sqrt{6}c$ and $\|(1/h)\mathbf{y}_{,3}^{(h)}\|_{W^{1,2}} \leqslant 2c$.

Therefore, there is a subsequence (not relabeled) such that

$$\begin{aligned} &\mathbf{y}^{(h)} \rightharpoonup \bar{\mathbf{y}} \text{ in } W^{2,2}, \\ &\frac{1}{h}\mathbf{y}_{,3}^{(h)} \rightharpoonup \bar{\mathbf{b}} \text{ in } W^{1,2}. \end{aligned} \tag{3.14}$$

From $(3.12)_{2,3}$ and the fact that Ω_1 is convex in the z_3 direction it follows that $\nabla\bar{\mathbf{y}}$ and $\bar{\mathbf{b}}$ are independent of z_3. By the trace theorem (Evans and Gariepy, 1992, Sect 4.3) $(\bar{\mathbf{y}}, \bar{\mathbf{b}})$ satisfy the boundary conditions (3.10). Write

$$\nabla_{\mathrm{p}}^2\mathbf{y}^{(h)} = \nabla_{\mathrm{p}}^2\bar{\mathbf{y}} + \mathbf{E}_{\mathrm{p}}^{(h)}, \quad \mathbf{E}_{\mathrm{p}}^{(h)} \rightharpoonup 0 \text{ in } L^2,$$

$$\frac{1}{h}\nabla \mathbf{y}^{(h)}_{,3} = \nabla \bar{\mathbf{b}} + \mathbf{E}^{(h)}_3, \quad \mathbf{E}^{(h)}_3 \rightharpoonup 0 \text{ in } L^2. \tag{3.15}$$

For the next test function we would like to choose $\bar{\mathbf{y}}(\mathbf{z}) + h\bar{\mathbf{b}}(\mathbf{z})z_3$, but $\bar{\mathbf{b}}$ is not smooth enough to do the subsequent second differentiation. Thus, we smooth $\bar{\mathbf{b}}$ by introducing $\bar{\mathbf{b}}_\varepsilon \in C^\infty$ with $\bar{\mathbf{b}}_\varepsilon \to \bar{\mathbf{b}}$ in $W^{1,2}(\Omega_1, \mathbb{R}^3)$, $\bar{\mathbf{b}}_\varepsilon$ satisfying the boundary condition $(3.10)_2$ and $\bar{\mathbf{b}}_\varepsilon$ independent of z_3. Then we compare $e^{(h)}_1[\mathbf{y}^{(h)}]$ with $e^{(h)}_1[\mathbf{y}^{(h)}_\varepsilon]$ where $\mathbf{y}^{(h)}_\varepsilon := \bar{\mathbf{y}} + h\bar{\mathbf{b}}_\varepsilon z_3$, using along the way the definitions in (3.15). This gives

$$\begin{aligned}
\int_{\Omega_1} \Bigg\{ \kappa \Bigg[& |\nabla^2_{\mathrm{p}} \bar{\mathbf{y}}|^2 + 2\nabla^2_{\mathrm{p}} \bar{\mathbf{y}} \cdot \mathbf{E}^{(h)}_{\mathrm{p}} + |\mathbf{E}^{(h)}_{\mathrm{p}}|^2 + 2(|\nabla \bar{\mathbf{b}}|^2 + 2\nabla \bar{\mathbf{b}} \cdot \mathbf{E}^{(h)}_3 + |\mathbf{E}^{(h)}_3|^2) \\
& + \frac{1}{h^4} |\mathbf{y}^{(h)}_{,33}|^2 \Bigg] + \varphi \left(\mathbf{y}^{(h)}_{,1} | \mathbf{y}^{(h)}_{,2} | \frac{1}{h} \mathbf{y}^{(h)}_{,3} \right) \Bigg\} \mathrm{d}\mathbf{z} \\
& \leqslant \int_{\Omega_1} \{ \kappa [|\nabla^2_{\mathrm{p}} \bar{\mathbf{y}} + h\nabla^2_{\mathrm{p}} \bar{\mathbf{b}}_\varepsilon z_3|^2 + 2|\nabla \bar{\mathbf{b}}_\varepsilon|^2] \\
& + \varphi(\bar{\mathbf{y}}_{,1} + h\bar{\mathbf{b}}_{\varepsilon,1} z_3 | \bar{\mathbf{y}}_{,2} + h\bar{\mathbf{b}}_{\varepsilon,2} z_3 | \bar{\mathbf{b}}_\varepsilon) \} \, \mathrm{d}\mathbf{z}
\end{aligned} \tag{3.16}$$

Fix ε and take lim sup of (3.16) as $h \to 0$. Using the smoothness of $\bar{\mathbf{b}}_\varepsilon$, simplify the right hand side. Then cancel the first term on the left with its counterpart on the right, and eliminate the second and fifth terms using (3.15). The eighth term on the left also converges by virtue of (3.14) and the Rellich Theorem (this gives strong convergence in L^q of $\nabla_{\mathrm{p}} \mathbf{y}^{(h)}$ and $(1/h)\mathbf{y}^{(h)}_{,3}$) and the bound (2.3). We get

$$\begin{aligned}
\limsup_{h \to 0} \int_{\Omega_1} \left\{ \kappa \left[|\mathbf{E}^{(h)}_{\mathrm{p}}|^2 + |\mathbf{E}^{(h)}_3|^2 + \frac{1}{h^4} |\mathbf{y}^{(h)}_{,33}|^2 \right] \right\} \mathrm{d}\mathbf{z} \leqslant \int_{\Omega_1} & \{ 2\kappa [|\nabla \mathbf{b}_\varepsilon|^2 - |\nabla \bar{\mathbf{b}}|^2] \\
& + \varphi(\bar{\mathbf{y}}_{,1} | \bar{\mathbf{y}}_{,2} | \bar{\mathbf{b}}_\varepsilon) - \varphi(\bar{\mathbf{y}}_{,1} | \bar{\mathbf{y}}_{,2} | \bar{\mathbf{b}}) \} \, \mathrm{d}\mathbf{z}.
\end{aligned} \tag{3.17}$$

Now pass to the limit $\varepsilon \to 0$, again using the bound (2.3). Thus, the "sup" can be dropped in (3.17) and we have improved the convergence in (3.14) and (3.15) to strong. This shows that the limiting energy of $\mathbf{y}^{(h)}$ is given by (3.9) evaluated at $(\bar{\mathbf{y}}, \bar{\mathbf{b}})$.

To establish the minimum principle, choose a test function $\hat{\mathbf{y}}^{(h)}(\mathbf{z}) = \hat{\mathbf{y}}(z_1, z_2) + h\hat{\mathbf{b}}(z_1, z_2)z_3$, with $(\hat{\mathbf{y}}, \hat{\mathbf{b}}) \in (C^\infty(\mathrm{S}, \mathbb{R}^3))^2$ satisfying the boundary conditions (3.10). Copy the argument (3.16 and 3.17). This gives the minimum principle for smooth competitors, and (3.9) follows by approximation. □

Remark 3.2. It is easily seen that Theorem 3.1 extends to boundary conditions of the form

$$\mathbf{y}(\mathbf{z}) = \mathbf{y}_0(z_1, z_2) + \sum_{i=1}^{N} \frac{h^i}{i!} \mathbf{b}_{(i)}(z_1, z_2) z_3^i \tag{3.18}$$

with $\mathbf{y}_0, \mathbf{b}_1, \ldots, \mathbf{b}_N \in W^{2,2}(\mathrm{S}, \mathbb{R}^3)$. For example, (3.18) permits the assignment of an h-independent curvature to vertical lines on the edge of the film.

Theorem 3.1 gives as the limiting theory a Cosserat theory with a single Cosserat

vector $\mathbf{b}: S \to \mathbb{R}^3$. The mapping $\mathbf{y}: S \to \mathbb{R}^3$ describes the average deformation of the film (e.g. the "middle surface") and $\mathbf{b}$ describes the deformation of the cross-section relative to the film. The structure of the test functions used in the proof gives a meaning to $\mathbf{y}$ and $\mathbf{b}$. That is, the minimizer and the typical test functions have the form,

$$\mathbf{y}^{(h)}(\mathbf{z}) \doteq \mathbf{y}(z_1, z_2) + h\mathbf{b}(z_1, z_2)z_3, \tag{3.19}$$

where the error is small in the sense of (3.8). Note that the error in the separate terms of (3.19) is measured differently. In terms of the original variables,

$$\tilde{\mathbf{y}}^{(h)}(\mathbf{x}) \doteq \mathbf{y}(x_1, x_2) + \mathbf{b}(x_1, x_2)x_3. \tag{3.20}$$

Hence, $\mathbf{b}$ describes approximately the shear and expansion of the cross-section, as summarized by Fig. 2. Even though both $\mathbf{y}$ and $\mathbf{b}$ arise from the given sequence of minimizers $\mathbf{y}^{(h)}$, no compatibility restrictions arise from the argument: in the final minimization of (3.9) $\mathbf{y}$ and $\mathbf{b}$ are independent functions.

Though it looks like $\mathbf{b}$ might carry information about bending energy, it does not as we explain in Section 5. The term $\nabla_p\mathbf{b}$ appearing in the integrand of (3.9) is not a bending term, but is rather a contribution of interfacial energy. The calculation shows that, for a sufficiently thin film, the dominant energies are stretching and interfacial. Bending energy is of higher order in h (Section 8). This fact does not seem to be well

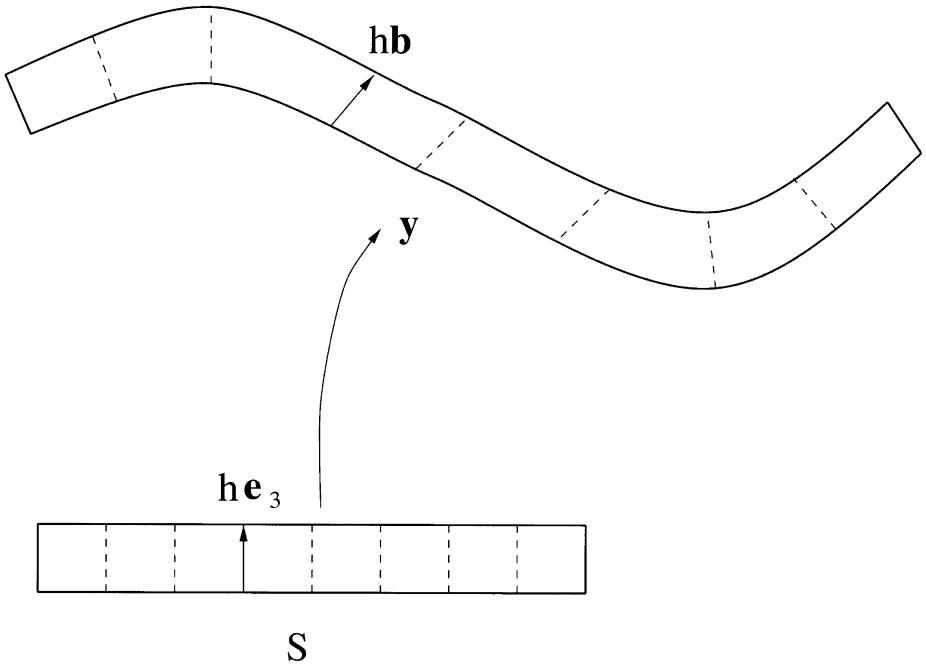

Fig. 2. Interpretation of $\mathbf{y}$ and $\mathbf{b}$.

appreciated in the actuator literature, which has focused on the use of cantilevers to produce actuation.

The limiting theory has inherited the balance between stretching and interfacial energy. Based on bulk measurements, it is expected that κ will be very much smaller than a typical elastic modulus that governs the growth of φ away from an energy well. Therefore, to form an idea of the nature of the minimizers of the limiting problem, it is useful to neglect interfacial energy by putting $\kappa = 0$ (Section 5). Then it is easy to construct various one, two or more phase minimizers of the simplified energy. If these have few interfaces (i.e., jumps of the deformation gradient) then it is expected that these will be rather low energy deformations. In some cases, suitably smoothed versions of these are expected to be minimizers of the limiting problem.

Despite the presence of approximations, the limiting theory is exactly frame-indifferent. Therefore, it is expected to give accurate predictions of the deformation of the film even in cases when the film undergoes very large stretch and rotation.

4. Results for a film bonded to a substrate

We now quote the result of a calculation analogous to Theorem 3.1 in the case that the film is bonded to a substrate. Here, we shall assume that the film is firmly bonded to a substrate governed by a free energy density $\varphi_s(\mathbf{A})$ having same general growth and invariance as $\varphi(\mathbf{A}, \theta)$ but with perhaps a different symmetry group and of course a different orientation; typically $\varphi_s(\mathbf{A})$ would be a more conventional energy having a single energy well. The result depends crucially on the thickness of the substrate. If the substrate is much thicker than the film, say of order $h^{1/2}$, then the energy of the film will disappear altogether from the final theory. Alternatively, if the substrate is much thinner than the film, say of order h^2, then the substrate energy will disappear from the limiting theory. This is well understood by practitioners. Thus, we make the film and substrate have comparable thicknesses by assuming that the following domains are assigned:

$$\begin{aligned} &\text{For the film:} \quad \Omega_h^f = S \times (0, \lambda h), \\ &\text{For the substrate:} \quad \Omega_h^s = S \times ((\lambda - 1)h, 0), \lambda \in (0, 1). \end{aligned} \tag{4.1}$$

Thus, the total thickness is h and the thickness ratio is $\lambda/(1-\lambda)$; see Fig. 3. The total free energy is,

$$e_{fs}^{(h)}[\mathbf{y}] = \int_{\Omega_h^f} \{\kappa_f |\nabla^2 \mathbf{y}|^2 + \varphi_f(\nabla \mathbf{y}, \theta)\} \, d\mathbf{x} + \int_{\Omega_h^s} \{\kappa_s |\nabla^2 \mathbf{y}|^2 + \varphi_s(\nabla \mathbf{y})\} \, d\mathbf{x} \tag{4.2}$$

and we use boundary conditions

$$\mathbf{y}(\mathbf{x}) = \begin{cases} \mathbf{a}_1 x_1 + \mathbf{a}_2 x_2 + \mathbf{a}_3^f x_3, & (x_1, x_2) \in \partial S, \quad 0 < x_3 < \lambda h, \\ \mathbf{a}_1 x_1 + \mathbf{a}_2 x_2 + \mathbf{a}_3^s x_3, & (x_1, x_2) \in \partial S, \quad (\lambda - 1)h < x_3 < 0, \end{cases} \tag{4.3}$$

so as to allow the film and substrate to shear and expand in different ways above and below the $x_3 = 0$ plane. Kinematic compatibility $[|\mathbf{y}|] = 0$ is imposed at $x_3 = 0$. This

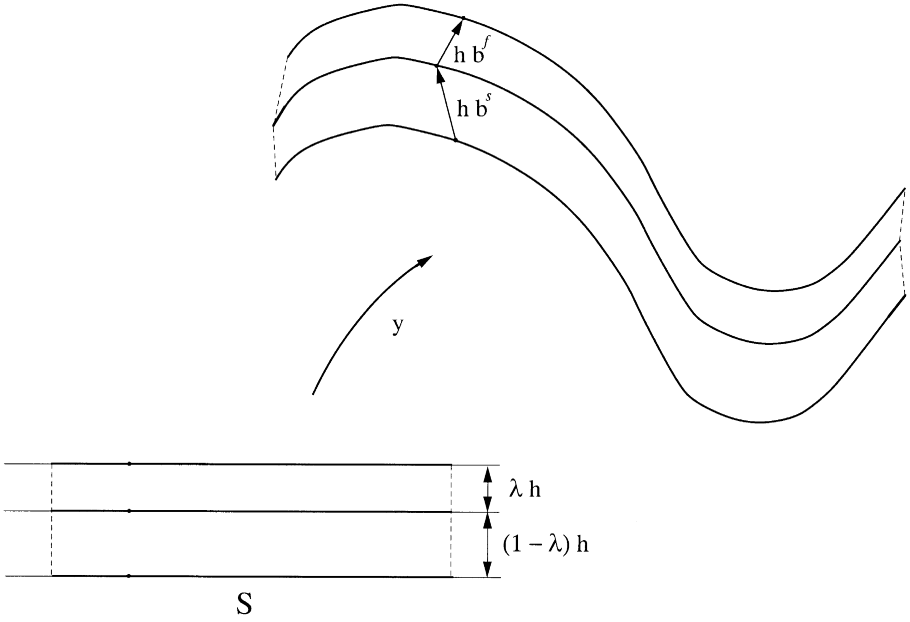

Fig. 3. Interpretation for $\mathbf{y}$, $\mathbf{b}^{\mathrm{f}}$ and $\mathbf{b}^{\mathrm{s}}$ for a film and substrate.

theory does not assign surface energy at the film/substrate interface; in other words, it allows $\nabla\mathbf{y}$ to jump freely across the surface $x_3 = 0$ as long as this does not violate kinematic compatibility. We make this choice for simplicity, but we recognize that for very thin films there may be a surface energy at the film/substrate interface which depends nontrivially on deformation. The existence and limiting theory are obtained as above, after first changing variables as in (3.1). For a film and substrate we get

$$\begin{aligned}
\nabla_{\mathrm{p}}^2\mathbf{y}^{(h)} &\to \nabla_{\mathrm{p}}^2\bar{\mathbf{y}} \quad \text{in } L^2(\mathrm{S}\times(\lambda-1,\lambda)),\\
\frac{1}{h}\nabla_{\mathrm{p}}\mathbf{y}_{,3}^{(h)} &\to \nabla_{\mathrm{p}}\bar{\mathbf{b}}^{\mathrm{s}} \quad \text{in } L^2(\mathrm{S}\times(\lambda-1,0)),\\
\frac{1}{h}\nabla_{\mathrm{p}}\mathbf{y}_{,3}^{(h)} &\to \nabla_{\mathrm{p}}\bar{\mathbf{b}}^{\mathrm{f}} \quad \text{in } L^2(\mathrm{S}\times(0,\lambda)),\\
\frac{1}{h^2}\mathbf{y}_{,33}^{(h)} &\to 0 \quad \text{in } L^2(\mathrm{S}\times(\lambda-1,\lambda)),
\end{aligned} \tag{4.4}$$

where $(\bar{\mathbf{y}}, \bar{\mathbf{b}}^{\mathrm{f}}, \bar{\mathbf{b}}^{\mathrm{s}})$ are independent of z_3 and minimize

$$\int_{\mathrm{S}} \{(\lambda\kappa_{\mathrm{f}}+(1-\lambda)\kappa_{\mathrm{s}})|\nabla_{\mathrm{p}}^2\mathbf{y}|^2 + 2(\lambda\kappa_{\mathrm{f}}|\nabla_{\mathrm{p}}\mathbf{b}^{\mathrm{f}}|^2 + (1-\lambda)\kappa_{\mathrm{s}}|\nabla_{\mathrm{p}}\mathbf{b}^{\mathrm{s}}|^2) + \lambda\varphi_{\mathrm{f}}(\mathbf{y}_{,1}|\mathbf{y}_{,2}|\mathbf{b}^{\mathrm{f}}) + (1-\lambda)\varphi_{\mathrm{s}}(\mathbf{y}_{,1}|\mathbf{y}_{,2}|\mathbf{b}^{\mathrm{s}})\}\,\mathrm{d}z_1\,\mathrm{d}z_2 \tag{4.5}$$

in $W^{2,2}(\mathrm{S},\mathbb{R}^3)\times(W^{1,2}(\mathrm{S},\mathbb{R}^3))^2$, subject to the boundary conditions

$$\left.\begin{aligned}\mathbf{y}(z_1,z_2)&=\mathbf{a}_1z_1+\mathbf{a}_2z_2\\ \mathbf{b}^{\mathrm{f}}(z_1,z_2)&=\mathbf{a}_3^{\mathrm{f}}\\ \mathbf{b}^{\mathrm{s}}(z_1,z_2)&=\mathbf{a}_3^{\mathrm{s}}\end{aligned}\right\}\quad (z_1,z_2)\in\partial\mathrm{S}. \tag{4.6}$$

This theory is also exactly frame-indifferent but involves two Cosserat vectors $\mathbf{b}^{\mathrm{f}}$ and $\mathbf{b}^{\mathrm{s}}$. See Fig. 3.

The limiting theory is interesting with regard to the design of films with unusual transformation properties. As above, the relation between surface and bulk energy is inherited by the thin film theory. Let us discuss the implications of (4.6) by first putting $\kappa_{\mathrm{s}}=\kappa_{\mathrm{f}}=0$. The elastic energy is a convex combination of the elastic energies of the film and the substrate. Therefore, as the thickness ratio $\lambda/(1-\lambda)$ passes from 0–∞ there is a smooth change of the energy from the energy-well structure of the substrate to that of the film. For a given $\mathbf{y}$ in the appropriate space the Cosserat vector fields $\mathbf{b}^{\mathrm{f}}$ and $\mathbf{b}^{\mathrm{s}}$ are found by minimizing pointwise and separately the energy densities $\varphi_{\mathrm{f}}(\mathbf{y}_{,1}|\mathbf{y}_{,2}|\cdot)$ and $\varphi_{\mathrm{s}}(\mathbf{y}_{,1}|\mathbf{y}_{,2}|\cdot)$. After this preliminary minimization that determines the approximate cross-sectional deformations, we minimize the integral with respect to $\mathbf{y}$ to get the shape of the approximate middle surface. To form a qualitative picture of what is expected, it is interesting to plot the energy-well structure of a convex combination of a double-well (representing schematically φ_{f}) and a single-well energy density (representing schematically φ_{s}). This is shown in Fig. 4. First, note that the

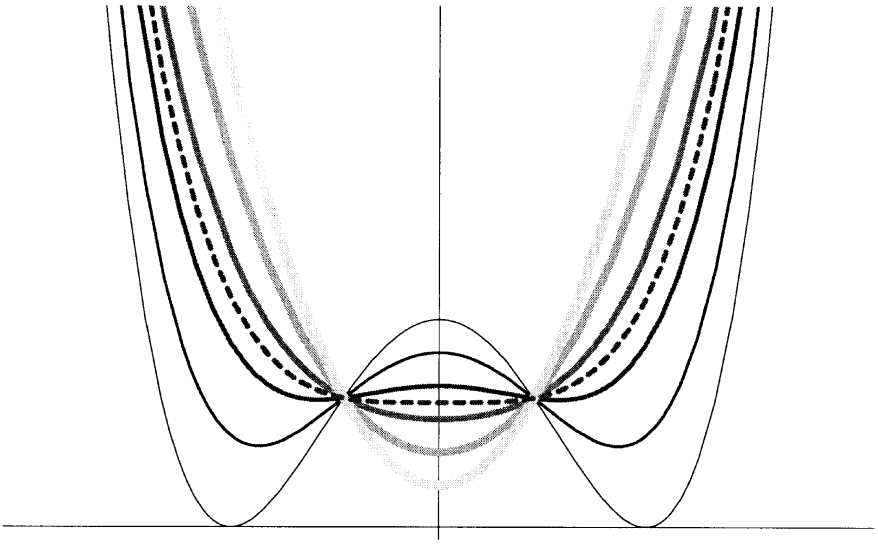

Fig. 4. Convex combination of a double and single well energy. Increasing line thickness corresponds to decreasing values of λ.

total free energy rises substantially, on the order of the barrier height, for intermediate values of λ. In practice, if there is a substantial barrier height, the film might tend to lose its epitaxial relation with the substrate during growth via mechanisms of surface diffusion or dislocation formation; these mechanisms are not included in the present model. Second, notice the second order "phase transition" at a certain value of λ (i.e., dashed in Fig. 4). This relates to Ericksen's (1980) extension of Landau's theory of second order phase transitions to first order phase transformations. The sensitive dependence of the elastic energy on thickness ratio is similar to that observed in the theory of compliant substrates (Freund and Nix, 1996).

By playing with different materials as substrate and film, a wide variety of interesting behaviors are possible. One possibility is to adjust λ to be at a second order transition point, thereby designing a film to have certain specific soft moduli. Notice from Fig. 5 that the locations of the energy well minima change significantly with the thickness ratio. Thus, another possibility is to design the energy wells of a multilayer film so as to satisfy certain special conditions of compatibility, e.g., the conditions (5.40) for the formation of a "tent".

5. Energy minimizing deformations

Consider a film occupying the reference domain $\mathrm{S} \subset \mathbb{R}^2$. According to Section 3, the deformations of this film are described by two vector fields, $\mathbf{y}: \mathrm{S} \to \mathbb{R}^3$ and $\mathbf{b}: \mathrm{S} \to \mathbb{R}^3$. According to Theorem 3.1, the behavior of the film is governed by the energy (3.9). We are interested in studying some deformations that minimize the energy (3.9). However, if the film is large enough, the elastic energy is much larger than the interfacial energy. Therefore, we can obtain a very reasonable approximation by setting $\kappa = 0$. We can then write the energy of the film to be

$$\int_{\mathrm{S}} \varphi(\mathbf{y}_{,1} | \mathbf{y}_{,2} | \mathbf{b}) \, \mathrm{d}z_1 \, \mathrm{d}z_2. \tag{5.1}$$

Deformations that minimize the energy (5.1) are exactly the deformations $(\mathbf{y}, \mathbf{b})$ that satisfy

$$(\mathbf{y}_{,1} | \mathbf{y}_{,2} | \mathbf{b}) \in SO(3)\mathbf{U}_0 \bigcup SO(3)\mathbf{U}_1 \bigcup \cdots \bigcup SO(3)\mathbf{U}_n \tag{5.2}$$

according to the structure of φ described in Section 2. Thus, we study deformations that involve the austenite and the different variants of the martensite. As before $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is an orthonormal basis such that $\mathbf{e}_1$ and $\mathbf{e}_2$ lie in the plane of the film while $\mathbf{e}_3$ is perpendicular to the film.

5.1. *Single phase deformations*

We begin with a deformation that involves only a single phase or variant so that

$$(\mathbf{y}_{,1} | \mathbf{y}_{,2} | \mathbf{b}) \in SO(3)\mathbf{U} \tag{5.3}$$

or

$$(\mathbf{y}_{,1}|\mathbf{y}_{,2}|\mathbf{b}) = \mathbf{Q}(z_1, z_2)\mathbf{U},$$

where $\mathbf{U}$ is the (constant) distortion matrix of the particular phase or variant while $\mathbf{Q}: \mathrm{S} \to SO(3)$ is a rotation-valued function. We will now show that even this very simple class of deformations is non-trivial.

Assume momentarily that $\mathbf{U} = \mathbf{I}$. Then, (5.3) becomes

$$(\mathbf{y}_{,1}|\mathbf{y}_{,2}|\mathbf{b}) = \mathbf{Q}. \tag{5.4}$$

The only differential constraints implied by (5.4) are

$$\mathbf{y}_{,1} = \mathbf{Q}\mathbf{e}_1 \quad \text{and} \quad \mathbf{y}_{,2} = \mathbf{Q}\mathbf{e}_2.$$

Since $\mathbf{Q}$ is a rotation,

$$\mathbf{y}_{,i} \cdot \mathbf{y}_{,j} = \delta_{ij}, \quad i,j = 1,2.$$

Therefore, the deformation $\mathbf{y}$ does not stretch the film: $\mathbf{y}$ is an isometric mapping of a subset of the plane into $\mathbb{R}^3$. Physically, $\mathbf{y}$ describes the class of "paper-folding deformations": the deformations that a flat sheet of paper can undergo. Further, for any paper-folding deformation $\mathbf{y}$, the function $\mathbf{b} = \mathbf{y}_{,1} \times \mathbf{y}_{,2}$ is the solution of (5.4). This is interpreted as saying that the (vertical) relative position vector between points on the top and the bottom of the film deforms so as to be approximately perpendicular to the deformed middle surface. If $\mathbf{U} \neq \mathbf{I}$ then the deformations that satisfy (5.3) are characterized by a uniform stretch of the film $\mathbf{U}$ followed by a paper-folding deformation as shown in Fig. 5.

We now report some results from differential geometry (see for example Kreyszig, 1968) that characterize smooth ($\mathbf{y} \in C^2(\mathrm{S})$) paper-folding deformations. First, the image of $\mathbf{y}$ is a developable surface. Second, a developable surface is locally a plane, a cylinder, a cone or a tangent surface (a tangent surface is obtained by sweeping a curve in the direction of its own tangent).

In summary, the deformations of a thin film that involve a single variant are not trivial. Instead, they stretch the film uniformly through the distortion matrix $\mathbf{U}$ and then deform it using a paper folding deformation into a developable surface (see Fig. 5). Thus, the theory captures the floppiness of the thin films.

Let us contrast this with the behavior of bulk specimen which occupies $\Omega \in \mathbb{R}^3$. A deformation $\tilde{\mathbf{y}}: \Omega \to \mathbb{R}^3$ involving a single variant satisfies

$$\nabla \tilde{\mathbf{y}} = \mathbf{Q}\mathbf{U} \tag{5.5}$$

for some rotation-valued function $\mathbf{Q}: \Omega \to SO(3)$. It follows (see for example James and Kinderlehrer, 1989) that $\mathbf{Q}$ = constant. Thus, the only deformation that a bulk specimen can undergo using a single variant is trivial: it involves a uniform stretch followed by a rigid rotation. Mathematically, this difference can be seen by comparing (5.3) and (5.5). In (5.3), the third column of the matrix on the left hand side is not a gradient; in (5.5) it is. This extra constraint forces the rotation to be a constant in (5.5). Physically, bending a bulk specimen requires a lot of energy. In contrast, the bending energy of a thin film is negligible (lower order in thickness than stretching); hence the film can curl up into developable surfaces.

5.2. *Two phase deformations*

We now consider deformations that consist of the austenite and a single variant of martensite or of two variants of martensite. In particular, consider the deformation shown in Fig. 6: we divide the film into two disjoint open regions S_1 and S_2, $S = S_1 \bigcup S_2 \bigcup \mathscr{I}$ and assume that

$$\begin{aligned} (\mathbf{y}_{,1}|\mathbf{y}_{,2}|\mathbf{b}) &= \mathbf{QU} \quad \text{in } S_1 \\ (\mathbf{y}_{,1}|\mathbf{y}_{,2}|\mathbf{b}) &= \mathbf{QV} \quad \text{in } S_2 \end{aligned} \tag{5.6}$$

where $\mathbf{V}$, $\mathbf{U}$ are the positive–definite, symmetric distortion matrices associated with the two phases and $\mathbf{Q}$ is a rotation valued function. We make the following classical assumptions:

(i) $\mathbf{y}$ is continuous with smooth limiting values on the smooth interface $\mathscr{I}$;
(ii) $\mathbf{y}_{,1}$, $\mathbf{y}_{,2}$ and $\mathbf{b}$ may suffer jumps across $\mathscr{I}$;
(iii) $\mathbf{y}$, $\mathbf{b}$ are smooth in S_1, S_2.

Condition (i) assures us that the deformation does not tear apart the film. Note that there is no such condition on $\mathbf{b}$, since it can jump without tearing the film! Condition (ii) allows us to consider interfaces consistent with (5.6).

The restriction that $\mathbf{y}$ be continuous while its derivatives suffer a jump imposes a kinematic compatibility condition or the Hadamard jump condition. This requires that at any point $(z_1, z_2) \in \mathscr{I}$, there exists a vector $\mathbf{a}$ such that

$$[|\mathbf{y}_{,1}|] = n_1\mathbf{a} \quad \text{and} \quad [|\mathbf{y}_{,2}|] = n_2\mathbf{a} \tag{5.7}$$

for some n_1, $n_2 \in \mathbb{R}$. Above, the notation $[|g|]$ denotes jump in any quantity g across $\mathscr{I}$. Therefore,

$$\begin{aligned} [|(\mathbf{y}_{,1}|\mathbf{y}_{,2}|\mathbf{b})|] &= n_1\mathbf{a} \otimes \mathbf{e}_1 + n_2\mathbf{a} \otimes \mathbf{e}_2 + [|\mathbf{b}|] \otimes \mathbf{e}_3, \\ &= \mathbf{a} \otimes \mathbf{n} + \mathbf{c} \otimes \mathbf{e}_3, \end{aligned}$$

where $\mathbf{n} = n_1\mathbf{e}_1 + n_2\mathbf{e}_2$ and $\mathbf{c} = [|\mathbf{b}|]$. Reversing the argument, we conclude that a deformation $(\mathbf{y}, \mathbf{b})$ is consistent with the conditions (i), (ii), (iii) if and only if

$$[|(\mathbf{y}_{,1}|\mathbf{y}_{,2}|\mathbf{b})|] = \mathbf{a} \otimes \mathbf{n} + \mathbf{c} \otimes \mathbf{e}_3 \tag{5.8}$$

for some vectors $\mathbf{a}$, $\mathbf{c}$ and $\mathbf{n}$ that satisfy $\mathbf{n} \cdot \mathbf{e}_3 = 0$. Furthermore, $\mathbf{n}$ is the normal to the interface (in the plane of the film).

We now return to the two-phase deformation (5.6). Substituting this into (5.8) we find that at any point (z_1, z_2) on $\mathscr{I}$,

$$\mathbf{Q}_1\mathbf{U} - \mathbf{Q}_2\mathbf{V} = \mathbf{a} \otimes \mathbf{n} + \mathbf{c} \otimes \mathbf{e}_3 \tag{5.9}$$

for some vectors $\mathbf{a}$, $\mathbf{n}$ and $\mathbf{c}$ where $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are the limiting values of the rotation $\mathbf{Q}$ on the interface $\mathscr{I}$. Pre-multiplying this equation by $\mathbf{Q}_2^T$, denoting $\mathbf{Q}_0 = \mathbf{Q}_2^T\mathbf{Q}_1$ and redefining the vectors $\mathbf{a}$ and $\mathbf{c}$ we can rewrite (5.9) as

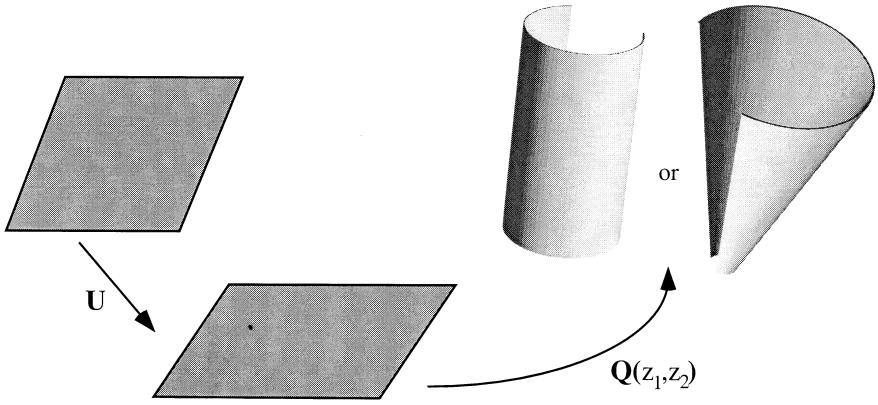

Fig. 5. Single phase deformation.

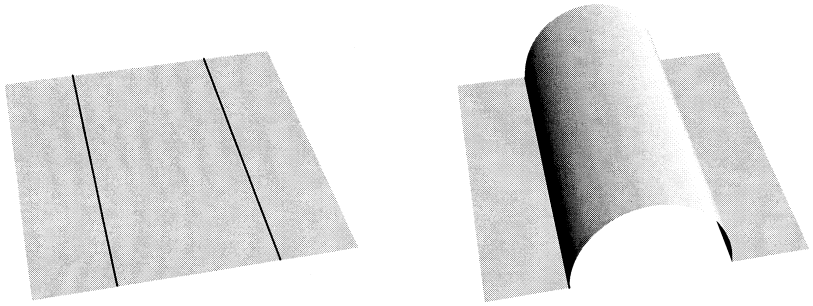

Fig. 7. Tunnel.

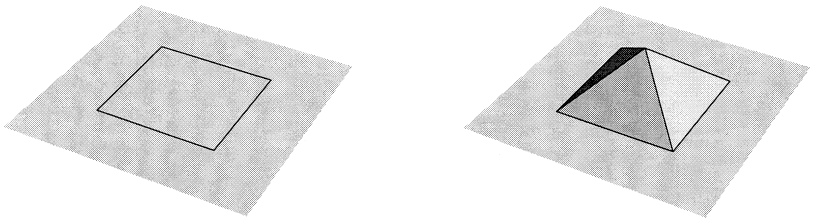

Fig. 8. Tent.

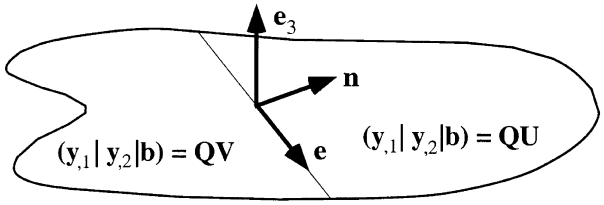

Fig. 6.

$$\mathbf{Q}_0\mathbf{U}-\mathbf{V}=\mathbf{a}\otimes\mathbf{n}+\mathbf{c}\otimes\mathbf{e}_3. \tag{5.10}$$

Equivalently,

$$(\mathbf{Q}_0\mathbf{U}-\mathbf{V})\mathbf{e}=0, \quad \mathbf{e}\cdot\mathbf{e}_3=0 \tag{5.11}$$

for some vector $\mathbf{e}\neq 0$. In what follows we assume without loss of generality that $|\mathbf{n}|=1$ and $|\mathbf{e}|=1$.

Let us now turn the question around: given two phases or variants, can we form an interface between them? Clearly, this is possible if we satisfy (5.11) for some rotation $\mathbf{Q}_0$ and some vector $\mathbf{e}$: construct a planar interface between them with normal $\mathbf{n}$ and set $\mathbf{Q}=\mathbf{Q}_0$ in S_1 and $\mathbf{Q}=\mathbf{I}$ in S_2.

Notice the (5.11) is an invariant line condition: the two deformation gradients $\mathbf{Q}_0\mathbf{U}$ and $\mathbf{V}$ deform a line in the direction of $\mathbf{e}$ in the $\mathbf{e}_1$–$\mathbf{e}_2$ plane identically. Thus, we can form an interface between two variants if and only if we can find an invariant line. The following result allows us to determine when we can satisfy (5.11) for some given matrices $\mathbf{U}$, $\mathbf{V}$ and normal $\mathbf{e}_3$.

Proposition 5.1. Given matrices $\mathbf{V}=\mathbf{V}^T>0$, $\mathbf{U}=\mathbf{U}^T>0$ and a vector $\mathbf{e}_3$, there exist $\mathbf{Q}_0\in SO(3)$ and a vector $\mathbf{e}$ with $|\mathbf{e}|=1$ that satisfy (5.11) if and only if

$$\mathbf{V}^{-1}\mathbf{e}_3\cdot(\operatorname{adj}\mathbf{A})\mathbf{V}^{-1}\mathbf{e}_3\leqslant 0, \tag{5.12}$$

where $\mathbf{A}=\mathbf{V}^{-1}\mathbf{U}^2\mathbf{V}^{-1}-\mathbf{I}$. To find the solutions, let $\mathbf{e}_1$ be any vector on the plane of the film (i.e., $\mathbf{e}_1\cdot\mathbf{e}_3=0$). Set

$$\mathbf{u}=\frac{\mathbf{V}\mathbf{e}_1}{|\mathbf{V}\mathbf{e}_1|}, \quad \mathbf{w}=\frac{\mathbf{V}^{-1}\mathbf{e}_3}{|\mathbf{V}^{-1}\mathbf{e}_3|} \quad \text{and} \quad \mathbf{v}=\mathbf{w}\times\mathbf{u}, \tag{5.13}$$

so that $\{\mathbf{u},\mathbf{v},\mathbf{w}\}$ form an orthonormal basis. Then, all solutions of (5.11) are given by

$$\mathbf{e}=\zeta\mathbf{V}^{-1}(\alpha\mathbf{u}+\beta\mathbf{v}), \tag{5.14}$$

where $\zeta\neq 0$ is chosen to make $|\mathbf{e}|=1$ while α,β are solutions of the quadratic equations

$$\alpha^2+\beta^2=1, \quad \alpha^2\mathbf{u}\cdot\mathbf{A}\mathbf{u}+2\alpha\beta\mathbf{u}\cdot\mathbf{A}\mathbf{v}+\beta^2\mathbf{v}\cdot\mathbf{A}\mathbf{v}=0. \tag{5.15}$$

Further, to classify the solutions $\mathbf{e}$ given by (5.14), let

$$M = \begin{pmatrix} \mathbf{u}\cdot\mathbf{A}\mathbf{u} & \mathbf{u}\cdot\mathbf{A}\mathbf{v} \\ \mathbf{u}\cdot\mathbf{A}\mathbf{v} & \mathbf{v}\cdot\mathbf{A}\mathbf{v} \end{pmatrix}. \tag{5.16}$$

(5.12) is equivalent to $\det M \leqslant 0$ so that the following cases are exhaustive.

Case 1. $M = 0$

Any unit vector $\mathbf{e}$ in the plane of the film is a solution.

Case 2. $M \neq 0$, $\det M = 0$

There is a unique solution up to change in sign, $\mathbf{e} \to -\mathbf{e}$.

Case 3. $\det M < 0$

There are exactly two solutions up to change in sign, $\mathbf{e} \to -\mathbf{e}$.

Finally, for each $\mathbf{e}$ given above, $\mathbf{Q}_0$ belongs to the one-parameter family of rotations that satisfy

$$\mathbf{Q}_0\mathbf{U}\mathbf{e} = \mathbf{V}\mathbf{e}.$$

Proposition 5.1 has a simple interpretation. Without loss of generality, we can change the reference configuration so that $\mathbf{V} = \mathbf{I}$; then the deformed side of the film can be viewed as having been stretched by amounts $(1+\mu_1)$ and $(1+\mu_2)$ in two perpendicular directions, and there exists an invariant line if and only if

$$\mu_1\mu_2 \leqslant 0. \tag{5.17}$$

The matrix M in (5.16) describes the relative stretch of the deformed film in the following sense: its eigenvectors are the principal axes of the stretch in the film and its eigenvalues are $\lambda_1 = (1+\mu_1)^2 - 1$, $\lambda_2 = (1+\mu_2)^2 - 1$. Hence, the condition $\det M \leqslant 0$ is equivalent to (5.17). If $\mu_1 = \mu_2 = 0$, then the film is undeformed and any line is an invariant line (Case 1); if $\mu_1 = 0$ or $\mu_2 = 0$, but not both, then there is one and only one invariant line along the unstretched principal axis of strain (Case 2) if (5.17) holds with $\mu_1 \neq 0$ and $\mu_2 \neq 0$ then there are two distinct invariant lines (Case 3). The last sentence of Proposition 5.1 says simply that the deformed side of the film admits an arbitrary superposed rotation with axis on the invariant line. In Section 7 the terminology *in-plane principal stretches* refers to the values of $(1+\mu_1)$ and $(1+\mu_2)$.

Proof of Proposition 5.1. Notice that (5.11) is equivalent to $|\mathbf{U}\mathbf{e}|^2 = |\mathbf{V}\mathbf{e}|^2$, $\mathbf{e}\cdot\mathbf{e}_3 = 0$ since $\mathbf{Q}_0$ is a rotation. Set $\mathbf{f} = (\mathbf{V}\mathbf{e}/|\mathbf{V}\mathbf{e}|)$ so that (5.11) is equivalent to

$$|\mathbf{U}\mathbf{V}^{-1}\mathbf{f}|^2 = |\mathbf{f}|^2, \quad \mathbf{f}\cdot\mathbf{V}^{-1}\mathbf{e}_3 = 0, \quad |\mathbf{f}| = 1;$$

or in the notation introduced in (5.12) and (5.13),

$$\mathbf{f}\cdot\mathbf{A}\mathbf{f} = 0, \quad \mathbf{f}\cdot\mathbf{w} = 0, \quad |\mathbf{f}| = 1. \tag{5.18}$$

Set $\mathbf{f} = \alpha\mathbf{u} + \beta\mathbf{v}$ so that $\mathbf{f}\cdot\mathbf{w} = 0$ is automatic. Then, $(5.18)_{1,3}$ reduces to the quadratic eqns (5.15). We can find a solution (α, β) to these if and only if $(\mathbf{u}\cdot\mathbf{A}\mathbf{u})$

$(\mathbf{v}\cdot\mathbf{A}\mathbf{v})-(\mathbf{u}\cdot\mathbf{A}\mathbf{v})^2 \leqslant 0$ or equivalently if $\det M \leqslant 0$. However notice that the matrix $\mathbf{A}$ in the $\{\mathbf{u},\mathbf{v},\mathbf{w}\}$ basis is given by

$$\begin{pmatrix} \mathbf{u}\cdot\mathbf{A}\mathbf{u} & \mathbf{u}\cdot\mathbf{A}\mathbf{v} & \mathbf{u}\cdot\mathbf{A}\mathbf{w} \\ \mathbf{u}\cdot\mathbf{A}\mathbf{v} & \mathbf{v}\cdot\mathbf{A}\mathbf{v} & \mathbf{v}\cdot\mathbf{A}\mathbf{w} \\ \mathbf{u}\cdot\mathbf{A}\mathbf{w} & \mathbf{v}\cdot\mathbf{A}\mathbf{w} & \mathbf{w}\cdot\mathbf{A}\mathbf{w} \end{pmatrix} \tag{5.19}$$

so that the "$\mathbf{w}-\mathbf{w}$" component of $\operatorname{adj}\mathbf{A}$ is equal to $\det M$. Therefore, we conclude that (5.15) has a solution (α,β) if and only if

$$\mathbf{w}\cdot(\operatorname{adj}\mathbf{A})\mathbf{w} \leqslant 0$$

which is exactly (5.12) since $\mathbf{w}=(\mathbf{V}^{-1}\mathbf{e}_3/|\mathbf{V}^{-1}\mathbf{e}_3|)$. From a solution (α,β) or (5.15) we obtain a solution $\mathbf{e}$ to (5.11) from (5.14).

The classification is also easy. If $M=0$, then any α, β satisfying $\alpha^2+\beta^2=1$ solves (5.15) and hence any unit vector $\mathbf{e}$ satisfies (5.11). If $M\neq 0$, but $\det M=0$; then one of the two eigenvalues of M is equal to zero while the other is nonzero; hence (5.15) has only one solution (up to sign) and (5.14) gives a unique direction. Finally, if $\det M<0$, M has two distinct eigenvalues and (5.14) gives two distinct directions.

Finally, the rotation $\mathbf{Q}_0$. We already know that the $\mathbf{e}$ in (5.14) satisfies $|\mathbf{U}\mathbf{e}|=|\mathbf{V}\mathbf{e}|$. Therefore, we can find a rotation $\mathbf{Q}_0$ such that $\mathbf{Q}_0\mathbf{U}\mathbf{e}=\mathbf{V}\mathbf{e}$. For any rotation $\mathbf{R}'$ with axis $\mathbf{V}\mathbf{e}$, $\mathbf{R}'\mathbf{Q}_0\mathbf{U}\mathbf{e}=\mathbf{V}\mathbf{e}$ so that $\mathbf{R}'\mathbf{Q}_0$ is also a solution. We can also obtain an alternative characterization of this one-parameter family as $\mathbf{Q}_0\mathbf{R}''$ where $\mathbf{R}''$ is any rotation about $\mathbf{U}\mathbf{e}$.

Let us now contrast the situation in thin films with that in a bulk specimen. Here, two variants with distortion matrices $\mathbf{U}$, $\mathbf{V}$ can form an interface if and only if they satisfy

$$\mathbf{Q}_0\mathbf{U}-\mathbf{V}=\mathbf{a}\otimes\mathbf{n} \tag{5.20}$$

for some rotation $\mathbf{Q}_0$, and vectors $\mathbf{a}$ and $\mathbf{n}$ (see for example Ball and James, 1987). This is an invariant plane condition in contrast with the thin film which requires only an invariant line condition (5.10). The limiting thin film theory allows for interfaces which may be incompatible in the thickness direction. We understand this as follows: in sufficiently thin films, it is possible to overcome an incompatibility in the thickness direction with an elastic deformation whose energy is small compared to the interfacial and membrane energies. We will explore this idea further with bounds and explicit constructions in Section 6.

Thus, the condition to form an interface in bulk is much more restrictive than in thin films. In particular, if it is possible to form an interface in the bulk, it continues to be possible to form an interface in the film. Further, the interface in the film is the trace of the bulk interface in the plane of the film. In contrast, there are interfaces in the film that are not possible in the bulk.

To understand this further, let us recall from Ball and James (1987) that for a given pair of matrices $\mathbf{U}$, $\mathbf{V}$, we can solve (5.20) if and only if the eigenvalues $(\lambda_1,\lambda_2,\lambda_3)$ of $\mathbf{V}^{-1}\mathbf{U}^2\mathbf{V}^{-1}$ satisfy the condition

$$\lambda_1 \leqslant \lambda_2 = 1 \leqslant \lambda_3. \tag{5.21}$$

In contrast, (5.10) is satisfied for some film orientation if and only if,

$$\lambda_1 \leqslant 1 \leqslant \lambda_3. \tag{5.22}$$

Notice that the stringent requirement that the middle eigenvalue be equal to one is not required in the film.

5.3. *Tunnels*

Figure 7 shows a pair of configurations that could be the basis of a microactuator. A film is deposited on the substrate in the austenite state. Assume that the film is unstressed as deposited. A strip is released as shown. At high temperatures, the film is in the austenite and is undeformed as shown on the left. On cooling, the film transforms to the martensite and bulges up, under perhaps a small back pressure, as shown on the right.

This deformation is possible if and only if the following conditions hold for some rotation $\mathbf{Q}$ and vector $\mathbf{e}$:

$$\begin{aligned} &1.\ (\mathbf{QU}-\mathbf{I})\mathbf{e} = 0, \quad \mathbf{e}\cdot\mathbf{e}_3 = 0, \quad |\mathbf{e}| = 1 \\ &2.\ \mathbf{n}\cdot\mathbf{U}^2\mathbf{e} = 0, \quad \text{where } \mathbf{n} = \mathbf{e}_3 \times \mathbf{e}, \\ &3.\ |\mathbf{Un}| > 1. \end{aligned} \tag{5.23}$$

The first condition is the compatibility condition that must be satisfied in order to form the two interfaces along the edges. Recall from Section 5.2 that for a given direction $\mathbf{e}$, $\mathbf{Q}$ can belong to a one parameter family of rotations. Therefore, the rotation continuously changes within this family as we go from one edge to another giving us a tunnel-shaped deformation. In the proposed deformation, two parts of the film are attached to the substrate. This implies that the film should suffer no shear in the strip when it transforms to martensite. The second condition of (5.23) is exactly a condition of no shear. The final condition says that the central part of the film is stretched in a direction perpendicular to the interface: this ensures that the film will bulge into a tunnel on deformation. We have the following result.

Proposition 5.2. Given a matrix $\mathbf{U} = \mathbf{U}^T > 0$ and a vector $\mathbf{e}_3$, there exist $\mathbf{Q} \in SO(3)$ and a vector $\mathbf{e}$ that satisfy (5.23) if and only if

$$\mathbf{e}_3 \cdot \operatorname{adj} \mathbf{A}\mathbf{e}_3 = 0 \quad \text{and} \quad \operatorname{tr} \mathbf{U}^2 - \mathbf{e}_3 \cdot \mathbf{U}^2\mathbf{e}_3 - 2 > 0 \tag{5.24}$$

where $\mathbf{A} = \mathbf{U}^2 - \mathbf{I}$.

Proof. Recall the notation of Proposition 5.1. Since $\mathbf{V} = \mathbf{I}$, $\mathbf{f} = \mathbf{e}$ and $\mathbf{w} = \mathbf{e}_3$, we have that $\mathbf{n} = -\beta\mathbf{u} + \alpha\mathbf{v}$ so that $(5.23)_{1,2}$ are equivalent to

$$\alpha^2 + \beta^2 = 1; \quad \{\alpha, \beta\} M \begin{Bmatrix} \alpha \\ \beta \end{Bmatrix} = 0; \quad \{-\beta, \alpha\} M \begin{Bmatrix} \alpha \\ \beta \end{Bmatrix} = 0. \tag{5.25}$$

This is possible if and only if 0 is an eigenvalue of M (with eigenvector $\{\alpha, \beta\}$), or equivalently $\det M = 0$. Therefore, $(5.23)_{1,2}$ are equivalent to $(5.24)_1$.

We now turn to $(5.24)_2$. Since $\{\mathbf{e}, \mathbf{n}, \mathbf{e}_3\}$ is an orthonormal basis,

$$\operatorname{tr} \mathbf{U}^2 = \mathbf{e}\cdot\mathbf{U}^2\mathbf{e} + \mathbf{n}\cdot\mathbf{U}^2\mathbf{n} + \mathbf{e}_3\cdot\mathbf{U}^2\mathbf{e}_3. \tag{5.26}$$

However, if $(5.23)_1$ holds then, $|\mathbf{U}\mathbf{e}| = 1$ or $\mathbf{e}\cdot\mathbf{U}^2\mathbf{e} = 1$ so that

$$|\mathbf{U}\mathbf{n}|^2 = \mathbf{n}\cdot\mathbf{U}^2\mathbf{n} = \operatorname{tr} \mathbf{U}^2 - \mathbf{e}_3\cdot\mathbf{U}^2\mathbf{e}_3 - 1 \tag{5.27}$$

and $(5.23)_3$ is equivalent to $(5.24)_2$. □

5.4. *Tents*

Figure 8 shows another possible design of a microactuator. A film is deposited on the substrate in the austenite state. Assume that the film is unstressed as deposited. A polygonal region is released. At high temperatures, the film is in the austenite and is undeformed as shown on the left. On cooling, the film transforms to the martensite and bulges up (under perhaps a small back pressure) like a tent as shown on the right.

Figure 9 shows a schematic of this deformation. This figure shows a four-sided tent; but it will be clear that the results generalize to n-sided tents. We look for a deformation $\mathbf{y}$, $\mathbf{b}$ such that $\mathbf{y}$ is continuous while the matrix $(\mathbf{y}_{,1}|\mathbf{y}_{,2}|\mathbf{b})$ takes the values shown in Fig. 9.

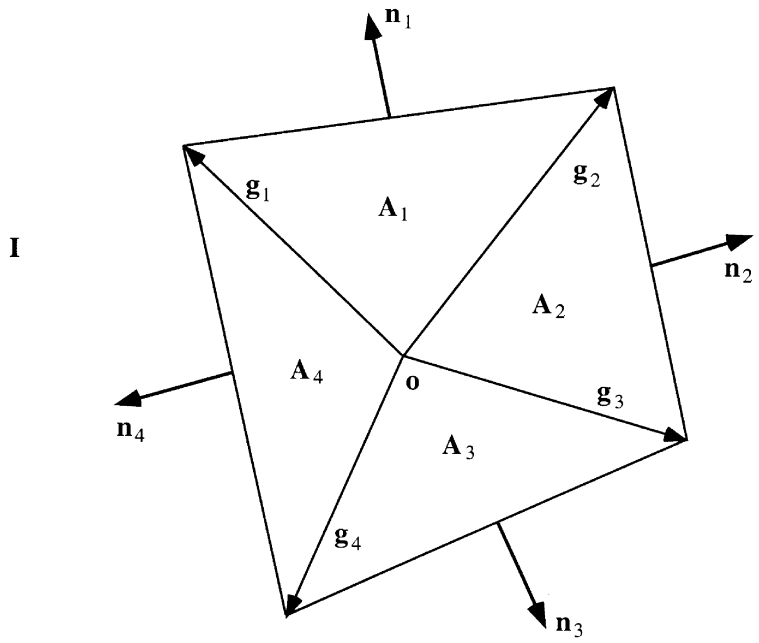

Fig. 9. Tent.

Consider the vectors $\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_n$ and $\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_n$ as shown in Fig. 9. They satisfy

$$\begin{aligned} &1.\ \mathbf{g}_i = \cos\theta_i \mathbf{e}_1 + \sin\theta_i \mathbf{e}_2 \quad \text{for } \theta_1, \theta_2, \ldots, \theta_n \quad \text{that satisfy } 0 < \theta_i - \theta_{i+1} < \pi. \\ &2.\ \mathbf{n}_i \cdot \mathbf{e}_3 = 0, \quad \mathbf{n}_i = R_{\pi/2}\left(\frac{\mathbf{g}_{i+1} - \mathbf{g}_i}{|\mathbf{g}_{i+1} - \mathbf{g}_i|}\right) \end{aligned} \tag{5.28}$$

where $R_{\pi/2}$ is a counter-clockwise rotation by $\pi/2$ about $\mathbf{e}_3$. These conditions say that $\mathbf{g}_i$ are numbered consecutively as we go clockwise around 0, that the angle between $\mathbf{g}_i$ and $\mathbf{g}_{i+1}$ is acute (and hence encloses a positive area), and the $\mathbf{n}_i$ are outward normals as shown.

Lemma 5.3. Let $\{\mathbf{g}_i\}$, $\{\mathbf{n}_i\}$ be consistent with (5.28). A continuous deformation $\mathbf{y}: \mathbb{R}^2 \to \mathbb{R}^3$ has gradient $(\mathbf{G}\mathbf{e}_1, \mathbf{G}\mathbf{e}_2)$ where $\mathbf{G} \in \{\mathbf{I}, \mathbf{A}_1, \ldots, \mathbf{A}_n\}$ as shown in Fig. 9 if and only if there exist vectors $\mathbf{a}, \mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$ and numbers $\alpha_1, \alpha_2, \ldots, \alpha_n$ such that

$$\mathbf{A}_i = \mathbf{I} + \mathbf{c}_i \otimes \mathbf{e}_3 + \alpha_i \mathbf{a} \otimes \mathbf{n}_i \quad i = 1, 2, \ldots, n, \tag{5.29}$$

where $\{\alpha_i\}$ satisfy the equations

$$(\alpha_{i+1}\mathbf{n}_{i+1} - \alpha_i \mathbf{n}_i) \cdot \mathbf{g}_{i+1} = 0 \quad i = 1, 2, \ldots, n. \tag{5.30}$$

(here we use the convention $\alpha_{n+1} = \alpha_1$, $\mathbf{n}_{n+1} = \mathbf{n}_1$, $\mathbf{g}_{n+1} = \mathbf{g}_1$.) Further, for given $\{\mathbf{n}_i\}$ and $\{\mathbf{g}_i\}$ consistent with (5.28), there is a solution $\{\alpha_i\}$ which is unique up to scaling: $\alpha_i \to \delta\alpha_i$, $\delta \in \mathbb{R}$.

Proof. For the deformation shown in Fig. 9, we have two sets of compatibility conditions: one set for the boundary of the polygon and another for the internal boundaries between the triangles:

$$\left.\begin{aligned} \mathbf{A}_i - \mathbf{I} &= \mathbf{c}_i \otimes \mathbf{e}_3 + \mathbf{a}_i \otimes \mathbf{n}_i \\ (\mathbf{A}_{i+1} - \mathbf{A}_i)\mathbf{g}_{i+1} &= 0 \end{aligned}\right\} \quad i = 1, 2, \ldots, n, \tag{5.31}$$

for some vectors $\mathbf{c}_i$ and $\mathbf{a}_i$ (Here, we use the notation $\mathbf{A}_{n+1} = \mathbf{A}_1$). Notice that we have used two different but equivalent forms of the compatibility conditions [cf (5.10), (5.11)]. Substituting $(5.31)_1$ into $(5.31)_2$, we obtain

$$\mathbf{a}_{i+1}(\mathbf{n}_{i+1} \cdot \mathbf{g}_{i+1}) = \mathbf{a}_i(\mathbf{n}_i \cdot \mathbf{g}_{i+1}) \quad i = 1, 2, \ldots, n \tag{5.32}$$

from which we conclude [with (5.28)] that

$$\mathbf{a}_i = \alpha_i \mathbf{a} \quad i = 1, 2, \ldots, n \tag{5.33}$$

for some vector $\mathbf{a}$ and for $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R}$. Substituting (5.33) in $(5.32)_2$ gives (5.30). Conversely, (5.29) and (5.30) are sufficient for (5.31).

Notice that the eqns (5.30) are cyclic; hence their sum is zero. Also, (5.28) imply that none of the coefficients of α_i vanish. Therefore, (5.30) has a unique solution up to scaling. □

Generally, the conditions (5.29) and (5.30) are highly restrictive. But if $\mathbf{e}_3$ is an n-fold symmetry axis ($n > 2$) for the austenite: $\mathbf{R}\mathbf{e}_3 = \mathbf{e}_3$, $\mathbf{R}^n = \mathbf{I}$ for some $\mathbf{R} \in P$, they become considerably less restrictive. For such films, it may be possible to form a

symmetric n-sided tent. We now examine the conditions that we need to satisfy in order to do so. Let $\mathbf{U}_1 = \mathbf{U}_1^T > 0$ and define,

$$\{\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_n\} = \{\mathbf{R}^k \mathbf{U}_1 (\mathbf{R}^k)^T : k = 0, \ldots, n-1\}. \tag{5.34}$$

Note that some of the matrices $\mathbf{U}_1 \mathbf{U}_2, \ldots, \mathbf{U}_n$ may be repeated. Assume that $\mathbf{U}_1$ is compatible with $\mathbf{I}$, so that there are vectors $\mathbf{a}$, $\mathbf{n}_1$, $\mathbf{c}_1$ and a rotation $\mathbf{Q}_1$ such that

$$\mathbf{Q}_1 \mathbf{U}_1 = \mathbf{I} + \mathbf{c}_1 \otimes \mathbf{e}_3 + \mathbf{a} \otimes \mathbf{n}_1, \quad \mathbf{n}_1 \cdot \mathbf{e}_3 = 0, \quad |\mathbf{n}_1| = 1. \tag{5.35}$$

Set $\mathbf{A}_1 = \mathbf{Q}_1 \mathbf{U}_1$. Notice that it satisfies $(5.29)_{i=1}$ with $\alpha_1 = 1$. The goal is to satisfy the remaining equations in (5.29) and (5.30) by symmetry. From (5.34) and (5.35),

$$\mathbf{Q}_k \mathbf{U}_k = \mathbf{I} + \mathbf{c}_k \otimes \mathbf{e}_3 + \mathbf{a}_k \otimes \mathbf{n}_k, \quad \mathbf{n}_k \cdot \mathbf{e}_3 = 0, \quad |\mathbf{n}_k| = 1 \tag{5.36}$$

where $\mathbf{Q}_{k+1} = \mathbf{R}^k \mathbf{Q}_1 (\mathbf{R}^k)^T$, $\mathbf{U}_{k+1} = \mathbf{R}^k \mathbf{U}_1 (\mathbf{R}^k)^T$, $\mathbf{c}_{k+1} = \mathbf{R}^k \mathbf{c}_1$, $\mathbf{n}_{k+1} = \mathbf{R}^k \mathbf{n}_1$, $\mathbf{a}_{k+1} = \mathbf{R}^k \mathbf{a}$, $k = 1, \ldots, n-1$. Setting $\mathbf{A}_k = \mathbf{Q}_k \mathbf{U}_k$, we see that $(5.29)_{i=2,\ldots,n}$ are satisfied if and only if $\mathbf{R}\mathbf{a} = \mathbf{a}$ or equivalently

$$\mathbf{a} = \alpha \mathbf{e}_3 \tag{5.37}$$

in which case we can choose $\alpha_2, \ldots, \alpha_n = 1$. Now consider $(5.30)_{i=1}$: $(\mathbf{n}_2 - \mathbf{n}_1) \cdot \mathbf{g}_2 = 0$. Notice that this is satisfied by choosing $\mathbf{g}_2 = \gamma(\mathbf{n}_2 + \mathbf{n}_1)$, $\gamma > 0$. Further, defining $\mathbf{g}_1 = \mathbf{R}^{n-1} \mathbf{g}_2$, $\mathbf{g}_{k+1} = \mathbf{R}^{k-1} \mathbf{g}_2$, we find that all of the equations in (5.30) are satisfied. Finally, $(5.28)_1$ holds with $\theta_i = \theta_0 - (2\pi/n)i$ for suitable θ_0 and so does $(5.28)_2$.

In summary, (5.29) and (5.30) are satisfied if (5.35) and (5.37) hold and $\mathbf{e}_3$ is an n-fold symmetry axis ($n > 2$) of austenite. Recall from Proposition 5.1 that there is nonuniqueness of the rotation $\mathbf{Q}_1$ in (5.35). Let (5.35) hold and let $\mathbf{e} = \mathbf{n}_1 \times \mathbf{e}_3$. From Proposition 5.1, given $\mathbf{e}_3$, $\mathbf{n}_1$, then all solutions of (5.35) are characterized by

$$\begin{aligned} &\mathbf{Q}_1 \to \mathbf{Q}' \mathbf{Q}_1, \quad \mathbf{c}_1 = (\mathbf{Q}_1 \mathbf{U}_1 - \mathbf{I})\mathbf{e}_3 \to (\mathbf{Q}' \mathbf{Q}_1 \mathbf{U}_1 - \mathbf{I})\mathbf{e}_3 \\ &\mathbf{a} = (\mathbf{Q}_1 \mathbf{U}_1 - \mathbf{I})\mathbf{n}_1 \to (\mathbf{Q}' \mathbf{Q}_1 \mathbf{U}_1 - \mathbf{I})\mathbf{n}_1 \end{aligned} \tag{5.38}$$

where the rotation $\mathbf{Q}'$ satisfies $\mathbf{Q}'\mathbf{e} = \mathbf{e}$. We can use the freedom in (5.38) to try and satisfy (5.37) which reduces to

$$\mathbf{Q}' \mathbf{Q}_1 \mathbf{U}_1 \mathbf{n}_1 - \mathbf{n}_1 = \alpha \mathbf{e}_3. \tag{5.39}$$

The following result tells us when we can satisfy this equation.

Lemma 5.4. Assume that (5.35) and (5.37) hold and let $\mathbf{e} = \mathbf{n}_1 \times \mathbf{e}_3$. There exists $\alpha \in \mathbb{R}$ and rotation $\mathbf{Q}'$, $\mathbf{Q}'\mathbf{e} = \mathbf{e}$ satisfying (5.39) if and only if

$$\mathbf{n}_1 \cdot \mathbf{U}_1^2 \mathbf{e} = 0 \quad \text{and} \quad |\mathbf{U}_1 \mathbf{n}_1| \geqslant 1. \tag{5.40}$$

Proof. Suppose (5.39) holds. Then, taking the dot product with $\mathbf{e}$ and using $\mathbf{Q}'\mathbf{e} = \mathbf{e}$ and $(\mathbf{Q}_1 \mathbf{U}_1 - \mathbf{I})\mathbf{e} = 0$ we obtain $(5.40)_1$. The condition $(5.40)_2$ follows easily by a simple application of the Pythagorean theorem: $|\mathbf{U}_1 \mathbf{n}_1|^2 = |\mathbf{n}_1 + \alpha \mathbf{e}_3|^2 = 1 + |\alpha \mathbf{e}_3|^2 \geqslant 1$. Conversely, assume (5.40) and let $\alpha = \sqrt{|\mathbf{U}_1 \mathbf{n}_1|^2 - 1}$. Define

$$\mathbf{Q}' = \mathbf{e} \otimes \mathbf{e} + \left(\frac{\mathbf{n}_1 \pm \alpha \mathbf{e}_3}{\sqrt{1+\alpha^2}} \right) \otimes \left(\frac{\mathbf{Q}_1 \mathbf{U}_1 \mathbf{n}_1}{|\mathbf{U}_1 \mathbf{n}_1|} \right) + \left(\mathbf{e} \times \left(\frac{\mathbf{n}_1 \pm \alpha \mathbf{e}_3}{\sqrt{1+\alpha^2}} \right) \right) \otimes \left(\mathbf{e} \times \frac{\mathbf{Q}_1 \mathbf{U}_1 \mathbf{n}_1}{|\mathbf{U}_1 \mathbf{n}_1|} \right). \tag{5.41}$$

Since $\sqrt{1+\alpha^2} = |\mathbf{U}_1\mathbf{n}_1|$, it follows from (5.40) that $\mathbf{Q}'$ is a rotation and satisfies (5.39). □

Remark 5.5. For a nontrivial tent, i.e. one that actually rises, $\alpha > 0$. The reason for the two $\pm$ signs in (5.41) is that the tent can go either up or down.

Thus we can form an n-sided tent if we satisfy the eqns (5.35) and (5.40) and $\mathbf{e}_3$ is an n-fold axis of symmetry for the austenite. However, notice that (5.35) and (5.40) are exactly the conditions required for forming a tunnel (5.23), so Proposition 5.2 applies. Therefore, we conclude that we can form a symmetric n-sided tent if $\mathbf{e}_3$ is an axis of n-fold symmetry of the austenite and $\mathbf{U}_1$ and $\mathbf{e}_3$ satisfy eqns (5.24).

6. Transition layers

The calculations of Section 5 give minimum energy states for the thin film theory in the case $\kappa = 0$. These consist of two (or more) variants of martensite, or austenite plus martensite, meeting at certain interfaces on the film. It is expected that some of these solutions will be close to smooth energy minimizing deformations in both the thin film theory based on (3.9) with $\kappa > 0$, and also in the original theory based on (2.12); the sharp interfaces are replaced with transition layers in these theories. From the form of (3.9) it is clear that any smooth deformation which is close to one of these two-phase deformations contributes positive energy to the thin film theory. Hence, it is expected that, with respect to the original theory (2.12), the corresponding transition layers will have energy $O(h)$.

In the original theory for bulk material it is well-known that transition layers between compatible deformation gradients have energy that scales as their cross-sectional area. In a thin film it is therefore expected that a transition layer between two deformation gradients which are compatible in bulk will have energy of $O(h)$. But the $\kappa = 0$ thin film theory has compatible interfaces that have no bulk counterpart. These are compatible along a line in the film but not along a plane, so they necessarily generate elastic energy in the original theory. To understand the structure and energy of both types of interface, it is of interest to go back to the original theory and examine the energy of the corresponding transition layers. A complete study of these layers would need to be done numerically, but some light is shed on the issue by examining upper and lower bounds for the energy of a transition layer between two regions having constant deformation gradients.

In this section we work in the original coordinates (x_1, x_2, x_3). We consider a film Ω_h of thickness h and width L, bisected by a transition layer of width $s^{(h)}$ as shown in Fig. 10. The dependence of $s^{(h)}$ on h is to be determined by energy minimization. Let $\mathbf{A}$ and $\mathbf{B}$ be two 3×3 matrices that minimize the elastic energy density: $\varphi(\mathbf{G}) \geqslant \varphi(\mathbf{A}) = \varphi(\mathbf{B}) = 0$ for all $\mathbf{G} \in \mathbf{M}^{3\times 3}$. Assign $\mathbf{y}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ on $\{\mathbf{x}\cdot\mathbf{n} < 0\} \bigcap \Omega_h$ and $\mathbf{y}(\mathbf{x}) = \mathbf{B}\mathbf{x} + \mathbf{c}_1$ on $\{\mathbf{x}\cdot\mathbf{n} > s^{(h)}\} \bigcap \Omega_h$. Assume that $\mathbf{A}$ and $\mathbf{B}$ satisfy conditions of compatibility that permit a thin film interface in the sense of Section 5, i.e.,

$$\mathbf{A} - \mathbf{B} = \mathbf{a} \otimes \mathbf{n} + \mathbf{c} \otimes \mathbf{e}_3, \quad \mathbf{n}\cdot\mathbf{e}_3 = 0, \quad |\mathbf{n}| = 1. \tag{6.1}$$

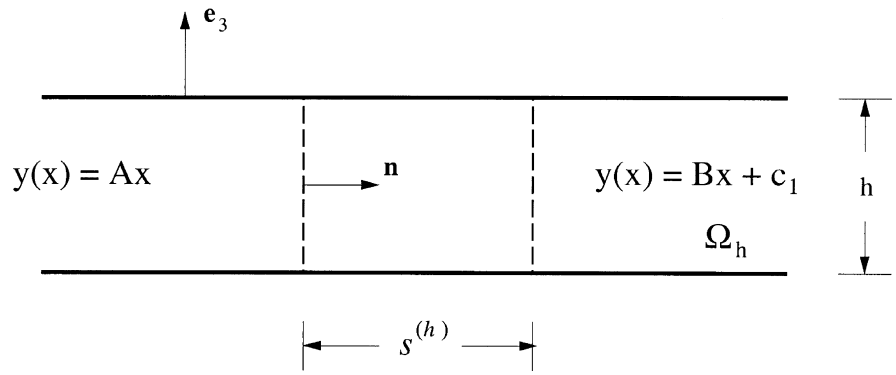

Fig. 10. Notation for transition layers.

The distinction between conditions of compatibility for bulk vs thin film concerns whether **a** is parallel to **c** or not. If $\mathbf{a} \parallel \mathbf{c}$ then **A** and **B** differ by a rank-one matrix and there is a continuous (but not smooth) interpolation of the deformation which makes the elastic energy identically zero: this is accomplished by extending the deformations $\mathbf{y}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ and $\mathbf{y}(\mathbf{x}) = \mathbf{B}\mathbf{x}+\mathbf{c}_1$ into the transition layer up to a suitable inclined plane in the layer, and using a suitable choice of $\mathbf{c}_1$. If $\mathbf{a} \parallel \mathbf{c}$, and the only zeros of φ are at **A** and **B**, then every interpolation of the layer (in, say, $W^{1,q}$) has positive elastic energy.

We begin with a naive interpolation based on weighted convex combination:

$$\mathbf{y}(\mathbf{x}) = \lambda(\mathbf{x}\cdot\mathbf{n})\mathbf{A}\mathbf{x}+(1-\lambda(\mathbf{x}\cdot\mathbf{n}))(\mathbf{B}\mathbf{x}+\mathbf{c}_1). \tag{6.2}$$

Here, $\lambda(s)$ is a smooth transition function,

$$\lambda(s) = \begin{cases} 1, & s \leqslant 0, \\ 0, & s \geqslant s^{(h)}, \end{cases} \tag{6.3}$$

satisfying the inequalities

$$\begin{aligned} &|\lambda'| \leqslant 2/s^{(h)}, \\ &|\lambda''| \leqslant 4/s^{(h)2}. \end{aligned} \tag{6.4}$$

The gradient and second gradient of (6.2) are

$$\begin{aligned} \nabla\mathbf{y}(\mathbf{x}) &= \lambda'[\mathbf{a}(\mathbf{n}\cdot\mathbf{x})+\mathbf{c}(\mathbf{e}_3\cdot\mathbf{x})-\mathbf{c}_1]\otimes\mathbf{n} \\ &\quad +\lambda\mathbf{A}+(1-\lambda)\mathbf{B}, \\ \nabla^2\mathbf{y}(\mathbf{x}) &= [\lambda''(\mathbf{x}\cdot\mathbf{n})+2\lambda']\mathbf{a}\otimes\mathbf{n}\otimes\mathbf{n} \\ &\quad +\lambda''(\mathbf{e}_3\cdot\mathbf{x})\mathbf{c}\otimes\mathbf{n}\otimes\mathbf{n} \\ &\quad -\lambda''\mathbf{c}_1\otimes\mathbf{n}\otimes\mathbf{n} \\ &\quad +\lambda'\mathbf{c}\otimes(\mathbf{n}\otimes\mathbf{e}_3+\mathbf{e}_3\otimes\mathbf{n}). \end{aligned} \tag{6.5}$$

Hence using (6.4) we have the bound on the interfacial energy,

$$\int_{\Omega_h} \kappa|\nabla^2\mathbf{y}|^2\,\mathrm{d}\mathbf{x} \leqslant c_1 L\kappa\left(\frac{h}{s^{(h)}}+\left(\frac{h}{s^{(h)}}\right)^2+\left(\frac{h}{s^{(h)}}\right)^3\right) \tag{6.6}$$

and the elastic energy bound,

$$\begin{aligned}\int_{\Omega_h} \varphi(\nabla\mathbf{y})\,\mathrm{d}\mathbf{x} &\leqslant c_2\int_{\Omega_h}(1+|\nabla\mathbf{y}|^q)\,\mathrm{d}\mathbf{x},\\ &\leqslant \int_{\Omega_h} c_2+c_3\left(\frac{h}{s^{(h)}}\right)^q\mathrm{d}\mathbf{x},\end{aligned} \tag{6.7}$$

where we have used the convexity of $|\mathbf{G}|^q$, $q>1$. Using (6.6) and (6.7) and the suggestive constant $E>0$ to indicate "elastic modulus", we get the following bound on the total energy,

$$e_1^{(h)}[\mathbf{y}^{(h)}] \leqslant c_4\kappa L\left(\frac{h}{s^{(h)}}\right)+c_5 EL(s^{(h)}h). \tag{6.8}$$

We optimize the right hand side of (6.8) over $s^{(h)}$ to get the following upper bound on the energy of the transition layer

$$Lh\sqrt{c_4c_5\kappa E}. \tag{6.9}$$

As expected based on the derivation of the limiting theory, transition layers have energy at most of order h, and the constants also enter as anticipated.

To confirm that transition layers typically have energy no less than order h, it is useful to have a lower bound. For this purpose let us adopt a particular energy-well structure appropriate to the cubic-to-tetragonal transformation at the transformation temperature:

$$\begin{aligned}&\varphi \geqslant 0,\\ &\quad = 0 \quad \text{on } K = SO(3)\mathbf{U}_1 \bigcup SO(3)\mathbf{U}_2 \bigcup SO(3)\mathbf{U}_3 \bigcup SO(3)\mathbf{I}.\end{aligned} \tag{6.10}$$

For definiteness consider an austenite/martensite interface with $\mathbf{A}=\mathbf{I}$ and $\mathbf{B}=\mathbf{U}_1$ and the film normal $\mathbf{e}_3$ coincident with the third vector of the cubic basis. Then the condition $\mathbf{e}_3\cdot\mathrm{adj}(\mathbf{U}_1^2-\mathbf{I})\mathbf{e}_3 \leqslant 0$ becomes $(\eta_1^2-1)(\eta_2^2-1)\leqslant 0$ and is satisfied, e.g., if $0<\eta_1<1<\eta_2$, which we assume. According to the results of Section 5, there are two possible solutions for two phase deformations in the limiting theory; we choose one with in-plane normal $\mathbf{n}$. Without loss of generality, we assume that $1<|\mathbf{U}_1\mathbf{n}|<|\mathbf{U}_2\mathbf{n}|$ and we let $\{\mathbf{e}_1=\mathbf{n},\mathbf{e}_2,\mathbf{e}_3\}$ be an orthonormal basis. If we consider the energy-well structure (6.10), we have that

$$A\in K \Rightarrow |\mathbf{A}\mathbf{e}_1| = |\mathbf{A}\mathbf{n}| \in K_1 = \{\eta_1, 1, |\mathbf{U}_1\mathbf{n}|, |\mathbf{U}_2\mathbf{n}|\}. \tag{6.11}$$

Therefore, it is reasonable to assume that $\varphi(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ is minorized by a smooth nonnegative function of $|\mathbf{a}_1|$ that vanishes exactly on K_1:

$$\varphi(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3) \geqslant (\psi(|\mathbf{a}_1|))^2 \geqslant 0, \quad (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3) \in (\mathbb{R}^3)^3; \quad \psi(\delta) = 0 \Rightarrow \delta \in K_1. \tag{6.12}$$

Assume without loss of generality that $\mathbf{y}: \Omega_h \to \mathbb{R}^3$ is smooth and bound the energy below using a variant of the trick of Modica and Mortola (1977):

$$\begin{aligned}
&\int_{\Omega_h} \{\kappa |\nabla^2 \mathbf{y}|^2 + \varphi(\mathbf{y}_{,1} | \mathbf{y}_{,2} | \mathbf{y}_{,3})\} \, \mathrm{d}\mathbf{x}, \\
&\geqslant \int_{\Omega_h} \{\kappa |\mathbf{y}_{,11}|^2 + \psi(|\mathbf{y}_{,1}|)^2\} \, \mathrm{d}\mathbf{x}, \\
&\geqslant 2\sqrt{\kappa} \int_{\Omega_h} \psi(|\mathbf{y}_{,1}|) \frac{\mathbf{y}_{,1}}{|\mathbf{y}_{,1}|} \cdot \mathbf{y}_{,11} \, \mathrm{d}\mathbf{x}.
\end{aligned} \tag{6.13}$$

Here, we have used the simple inequalities $|\mathbf{a}|^2 + |\mathbf{b}|^2 \geqslant 2|\mathbf{a}||\mathbf{b}|$ and $|\mathbf{a}| \geqslant \mathbf{a} \cdot \mathbf{e}$ for $|\mathbf{e}| = 1$, with $\mathbf{e} = \mathbf{y}_{,1}/|\mathbf{y}_{,1}|$. Here, readers uncomfortable with the transition between $(6.13)_2$ and $(6.13)_3$ can smooth slightly the function $|\mathbf{z}|$ and replace $\mathbf{e} = \mathbf{y}_{,1}/|\mathbf{y}_{,1}|$ by the gradient of the smoothed version. Now let Ψ be the antiderivative of ψ. The right hand side of (6.13) can be therefore further bounded below by,

$$2\sqrt{\kappa} \int_{\Omega_h} \frac{\mathrm{d}}{\mathrm{d}x_1} [\Psi(|\mathbf{y}_{,11}|)] \, \mathrm{d}\mathbf{x} = 2\sqrt{\kappa} Lhs^{(h)} [\Psi(|\mathbf{U}_1 \mathbf{n}|) - \Psi(1)]. \tag{6.14}$$

Choosing $s^{(h)} > 0$ independent of h as in the derivation of (6.9), we get a similar bound as in (6.9).

We therefore reach the following conclusions. Generally transition layers between austenite and martensite or between different variants of martensite have energy of order h. The derivation of the lower bound would not have been changed in any essential way if $\mathbf{A}$ and $\mathbf{B}$ were to differ by a rank-one matrix, although the constants in front of h could possibly be quite different in the two cases. This indifference can be attributed to the dominating effect of interfacial energy at small scales.

7. Applications to special materials

The formulas obtained in Section 5 can be evaluated for the distortion matrices (2.6)–(2.10) corresponding to various martensitic transformations. In this section we organize these results and point out applications that appear to be the most interesting from a practical viewpoint. Throughout this section we adhere to the notation and assumptions on lattice parameters given in (2.6)–(2.10).

The results in this paper are appropriate to single crystal films. The preferred method for growing such films is epitaxial, e.g., molecular beam epitaxy. In the cases

studied here with cubic austenite, this method is likely to lead to films with low index normals such as in the families $\{100\}$, $\{110\}$ or $\{111\}$, so we concentrate on those here. Even sputtered films often exhibit a strong texture with an excess of grains having a normal from one of these families.

First, we consider straight-line interfaces between austenite and a single variant of martensite. By crystal symmetry we can choose without loss of generality only the first distortion matrix in each of (2.6)–(2.10), and this is done in all the calculations below. We first evaluate and simplify the condition (5.12) with $\mathbf{V} = \mathbf{I}$ and $\mathbf{U} =$ [the first matrix in each of (2.6)–(2.10)], to check the existence of austenite/martensite interfaces in films with various orientations. The results are shown in the Appendix. In general, these conditions are seen to be rather mild.

For example, we consider $Ni_{50}Ti_{50}$ films of various orientation with distortion matrices of the form (2.10). Table 1 gives results for exact interfaces between austenite and variant 1 of martensite for films with orientation in the families $\{100\}$, $\{110\}$ and

Table 1
Exact interfaces between austenite and variant 1 of martensite for films of various orientation in $Ni_{50}Ti_{50}$

Film normal	Austenite/ martensite interface?	Interface lines [solutions **e** of (5.11)]	In-plane principal stretches
100	yes	(0, −0.9639, 0.2664) or (0, 0.3841, 0.9233)	(0.9358, 1.0473)
010	yes	(−0.9639, 0, 0.2664) or (0.3841, 0, 0.9233)	(0.9358, 1.0473)
001	yes	(−0.9728, 0.2317, 0) or (0.2317, −0.9728, 0)	(1.0840, 0.9663)
110	no		(0.9582, 0.9663)
1-10	yes	(0.1892, 0.1892, 0.9636) or (0.6080, 0.6080, −0.5105)	(1.1066, 0.9320)
101	yes	(−0.3339, 0.8815, 0.3339) or (0.5018, 0.7046, −0.5018)	(0.9464, 1.0286)
10-1	yes	(0.1351, −0.9816, 0.1351) or (0.6840, −0.2538, 0.6840)	(1.1005, 0.9574)
011	yes	(0.8815, −0.3339, 0.3339) or (0.7046, 0.5018, −0.5018)	(0.9464, 1.0286)
01-1	yes	(−0.9816, 0.1351, 0.1351) or (−0.2538, 0.6840, 0.6840)	(1.1005, 0.9574)
111	no		(0.9422, 0.9663)
-111	yes	(0.3505, 0.8139, −0.4634) or (0.5952, −0.1864, 0.7816)	(1.1001, 0.9424)
1-11	yes	(0.8139, 0.3505, −0.4634) or (−0.1864, 0.5952, 0.7816)	(1.1001, 0.9424)
11-1	yes	(−0.3238, 0.8110, 0.4872) or (0.8110, −0.3238, 0.4872)	(1.0582, 0.9663)

Interface lines are lines on the film at which austenite and martensite meet. The in-plane principal stretches refer to the martensite. See the Appendix for a remark about crystallographic equivalence.

{111}. The in-plane principal stretches have been computed using the procedure described after (5.17). In only two of the thirteen cases is there no such interface. Notice that in some cases with {110} films, the martensite undergoes over 10% strain while remaining exactly compatible with austenite in the thin film sense.

Another interesting case is alloys of Ni–Mn–Ga near the Heusler composition Ni_2MnGa. This alloy undergoes a cubic-to-tetragonal martensitic transformation as described in (2.6). This alloy is interesting because it belongs to the emerging class ferromagnetic shape-memory alloys (see James and Wuttig, 1996, 1997; Tickle et al., 1997). In this alloy both the austenite and martensite are ferromagnetic: this leads to the possibility of rearranging martensite variants by applying a magnetic field. In thin films this gives the added flexibility of inducing the formation of a tunnel or tent by applying a field, as discussed by DeSimone et al. (1997). Recently, Bensaoula et al. (1997) have grown single crystal films of Ni_2MnGa on $Ga_{(1-x)}In_xAs$ by MBE. The possible interfaces in this alloy up to crystallographic symmetry are listed in Table 2.

Finally, we consider austenite/martensite interfaces $Ni_{64}Al_{36}$. This alloy is interesting because of its premartensitic behavior and its high transformation temperature. It has been studied intensively by Schryvers, Tanner and collaborators (1990, 1991, 1995). As a polycrystalline bulk material, it is unlikely to exhibit a reversible shape-memory effect (Bhattacharya and Kohn, 1996), but this is overcome in single crystal films. It undergoes a cubic tetragonal transformation, $(2.7)_{\text{ff}}$. The results are given in Table 3.

Now we turn to a discussion of tents and tunnels, governed by (5.23), Proposition 5.2 and Lemmas 5.3 and 5.4. We first note that $\mathbf{U}_1$ for NiTi has an eigenvalue nearly equal to one. This fact can be exploited to design an alloy of NiTiCu with low hysteresis, by arranging the concentration of Cu to make an eigenvalue of $\mathbf{U}_1$ equal to one (see Ball et al. (1995) for the connection between hysteresis and this condition). In addition, the presence of more than 10 at.% Cu collapses the symmetry to orthor-

Table 2
Exact interfaces between austenite and variant 1 of martensite for films of various orientation in Ni_2MnGa

Film normal	Austenite/ martensite interface?	Interface lines [solutions **e** of (5.11)]	In-plane principal stretches
100	no		(0.9512, 0.9512)
010	yes	(0.5058, 0, 0.8626) or (0.5058, 0, −0.8626)	(1.130, 0.9512)
110	yes	(0.5058, −0.5058, −0.6987) or (0.5058, −0.5058, 0.6987)	(0.9512, 1.044)
011	yes	(0.5058, −0.610, 0.610) or (0.5058, 0.610, −0.610)	(1.130, 0.9512)
111	yes	(0.5058, 0.3021, −0.808) or (0.5058, −0.808, 0.3021)	(1.0737, 0.9512)

Interface lines are lines on the film at which austenite and martensite meet. The in-plane principal stretches refer to the (thermal) martensite. The cases in the {100}, {110} and {111} families not shown can be obtained by crystallographic equivalence as described in the Appendix.

Table 3
Exact interfaces between austenite and variant 1 of martensite for films of various orientation in $Ni_{64}Al_{36}$

Film normal	Austenite/martensite interface?	Interface lines [solutions **e** of (5.11)]	In-plane principal stretches
100	no		(0.9392, 0.9392)
010	yes	(0.5462, 0, 0.8377) or (0.5462, 0, −0.8377)	(1.1302, 0.9392)
110	yes	(0.5462, −0.5462, −0.6351) or (0.5462, −0.5462, 0.6351)	(0.9392, 1.0391)
011	yes	(0.5462, −0.5923, 0.5923) or (0.5462, 0.5923, −0.5923)	(1.1302, 0.9392)
111	yes	(0.5462, 0.2525, −0.7987) or (0.5462, −0.7987, 0.2525)	(1.0703, 0.9392)

See the footnote of Table 2.

hombic (i.e. $\varepsilon = 0$). It can be seen from (5.24) that in the case that $\mathbf{U}_1$ has an eigenvalue equal to one the conditions for tunnel and tent formation become easier to satisfy. We have constructed Table 4 using the lattice parameters for $Ni_{30.5}Ti_{49.5}Cu_{20.0}$, $(2.8)_{ff}$. The conditions (5.23) are equivalent to the existence of an austenite/martensite interface with corresponding in-plane principal stretches of the form $(1, 1+\mu)$ for $\mu > 0$. It is seen from Table 4 that this alloy supports a tunnel or a four-fold tent on a (001) film. Another interesting case in Cu–Zn–Al and related alloys has been found by Hane (1997).

8. Higher order theories

We showed in Section 3 that the behavior of very thin films is governed by the "limiting energy" e^0 given in (3.9). This energy contains interfacial and membrane terms. For thicker films, we expect other terms to become important. In this section we examine the corrections to (3.9) at higher orders in the thickness. We will show that the first correction to e^0 is second order in thickness h and consists of additional interfacial energy and a bending energy.

Recall the energies $e_1^{(h)}$ and e^0 defined in (3.4) and (3.9), respectively. Throughout this section, let $\mathbf{y}^{(h)} : \Omega_1 \to \mathbb{R}^3$ minimize $e_1^{(h)}$ amongst all deformations $\tilde{\mathbf{y}} : \Omega_1 \to \mathbb{R}^3$ that satisfy the boundary condition

$$\tilde{\mathbf{y}}(z_1, z_2, z_3) = \mathbf{y}_0(z_1, z_2) + h\mathbf{b}_1(z_1, z_2)z_3 \quad \mathbf{z} \in \partial \mathrm{S} \times (-\tfrac{1}{2}, \tfrac{1}{2}) \tag{8.1}$$

and let $\bar{\mathbf{y}} : \mathrm{S} \to \mathbb{R}^3$, $\bar{\mathbf{b}} : \mathrm{S} \to \mathbb{R}^3$ minimize e^0 amongst all functions $\mathbf{y}, \mathbf{b} : \mathrm{S} \to \mathbb{R}^3$ that satisfy the boundary conditions

$$\mathbf{y}(z_1, z_2) = \mathbf{y}_0(z_1, z_2), \quad \mathbf{b}(z_1, z_2) = \mathbf{b}_1(z_1, z_2) \quad (z_1, z_2) \in \partial \mathrm{S} \tag{8.2}$$

Table 4
Exact interfaces between austenite and variant 1 of martensite for films of various orientation in $Ni_{30.5}Ti_{49.5}Cu_{20.0}$

Film normal	Austenite/ martensite interface?	Interface lines [solutions **e** of (5.11)]	In-plane principal stretches
100	yes	(0, −0.7607, −0.649) (0, −0.7607, 0.649)	(0.9579, 1.03)
010	yes	(−0.7607, 0, −0.649) (−0.7607, 0, 0.649)	(0.9579, 1.03)
001	yes	(0.7071, 0.7071, 0) (0.7071, 0.7071, 0)	(1, 1.058)
110	yes	(−0.4512, 0.4512, −0.7699) (−0.4512, 0.4512, 0.7699)	(1.058, 0.9579)
1-10	yes	(0.7071, 0.7271, 0) (0.7071, 0.7271, 0)	(1, 0.9579)
101	yes	(0.7019, −0.1208, −0.7019) (0.3858, 0.838, −0.3858)	(1.039, 0.9844)
10-1	yes	(−0.7019, 0.1208, −0.7019) (−0.3858, −0.838, −0.3858)	(1.039, 0.9844)
011	yes	(−0.1208, 0.7029, −0.7019) (0.838, 0.3858, −0.3858)	(1.039, 0.9844)
01-1	yes	(0.1208, −0.7019, −0.7019) (−0.838, −0.3858, −0.3858)	(1.039, 0.9844)
111	yes	(0.7344, −0.05819, −0.6762) (−0.05819, 0.7344, −0.6752)	(1.058, 0.9721)
-111	yes	(0.7071, 0.7071, 0) (0.7071, 0.7071, 0)	(1, 0.9925)
1-11	yes	(−0.7071, −0.7071, 0) (−0.7071, −0.7071, 0)	(1, 0.9925)
11-1	yes	(−0.7344, 0.5819, −0.6762) (0.05819, −0.7344, −0.6762)	(1.058, 0.9721)

See the footnote of Table 2 and the text.

for given smooth $\mathbf{y}_0$, $\mathbf{b}_1 : \mathrm{S} \to \mathbb{R}^3$. We assume that $\mathbf{y}^{(h)}$ converges to $(\bar{\mathbf{y}}, \bar{\mathbf{b}})$ as described by (3.8).

We assume that the functions $\mathbf{y}^{(h)}$, $\bar{\mathbf{y}}$, $\bar{\mathbf{b}}$ are sufficiently smooth throughout this section and we invoke growth conditions on φ and its derivatives as needed. We also assume a strong second variation condition below. To this extent the calculations presented below are formal and therefore are called "results". We discuss these assumptions at the end.

Our first result shows that the first order correction to e^0 is zero.

Result 8.1. $\lim_{h\to 0} (1/h)(e_1^{(h)}[\mathbf{y}^{(h)}] - e^0[\bar{\mathbf{y}}, \bar{\mathbf{b}}]) = 0$

Proof. We will first show that

$$\liminf_{h\to 0} \frac{1}{h}(e_1^{(h)}[\mathbf{y}^{(h)}] - e^0[\bar{\mathbf{y}}, \bar{\mathbf{b}}]) \geqslant 0. \tag{8.3}$$

Given any $z_3 \in (-\frac{1}{2}, \frac{1}{2})$, define $\hat{\mathbf{y}} : \mathrm{S} \to \mathbb{R}^3$, $\hat{\mathbf{b}} : \mathrm{S} \to \mathbb{R}^3$ by

$$\hat{\mathbf{y}}(z_1,z_2)=\mathbf{y}^{(h)}(z_1,z_2,z_3)-hz_3\mathbf{b}_1,\quad \hat{\mathbf{b}}(z_1,z_2)=\frac{1}{h}\mathbf{y}^{(h)}_{,3}(z_1,z_2,z_3). \tag{8.4}$$

Notice that $\hat{\mathbf{y}}$, $\hat{\mathbf{b}}$ satisfy the boundary conditions (8.2). Therefore,

$$e^0[\hat{\mathbf{y}},\hat{\mathbf{b}}]-e^0[\bar{\mathbf{y}},\bar{\mathbf{b}}]\geqslant 0 \tag{8.5}$$

for each $z_3\in(-\frac{1}{2},\frac{1}{2})$. Integrating (8.5) with respect to z_3 and dividing by h, we obtain

$$\frac{1}{h}\left(\int_{-(1/2)}^{1/2} e^0[\hat{\mathbf{y}},\hat{\mathbf{b}}]\,\mathrm{d}z_3-e^0[\bar{\mathbf{y}},\bar{\mathbf{b}}]\,\mathrm{d}z_3\right)\geqslant 0. \tag{8.6}$$

Add $1/h^5\|\mathbf{y}^{(h)}_{,33}\|^2_{L^2}$ to both sides, substitute for $\hat{\mathbf{y}}$, $\hat{\mathbf{b}}$ and expand the first term to obtain

$$\frac{1}{h}(e^{(h)}_1[\mathbf{y}^{(h)}]-e^0[\bar{\mathbf{y}},\bar{\mathbf{b}}])+T^{(h)}+O(h)\geqslant\frac{1}{h^5}\|\mathbf{y}^{(h)}_{,33}\|^2_{L^2}\geqslant 0, \tag{8.7}$$

where

$$T^{(h)}=-\int_{\Omega_1}\left\{2\kappa(\nabla^2_{\mathrm{p}}\mathbf{y}^{(h)})\cdot(\nabla^2_{\mathrm{p}}\mathbf{b}_1)+\left(\frac{\partial\varphi}{\partial\mathbf{F}}\bigg|_{(\mathbf{y}^{(h)}_{,1}|\mathbf{y}^{(h)}_{,2}|(1/h)\mathbf{y}^{(h)}_{,3})}\right)\cdot(\mathbf{b}_{1,1}|\mathbf{b}_{1,2}|0)\right\}z_3\,\mathrm{d}\mathbf{z}. \tag{8.8}$$

However,

$$\begin{aligned}\lim_{h\to 0}T^{(h)}&=-\int_{\Omega_1}\left\{2\kappa(\nabla^2_{\mathrm{p}}\bar{\mathbf{y}})\cdot(\nabla^2_{\mathrm{p}}\mathbf{b}_1)+\left(\frac{\partial\varphi}{\partial\mathbf{F}}\bigg|_{(\bar{\mathbf{y}}_{,1}|\bar{\mathbf{y}}_{,2}|\bar{\mathbf{b}})}\right)\cdot(\mathbf{b}_{1,1}|\mathbf{b}_{1,2}|0)\right\}z_3\,\mathrm{d}\mathbf{z}\\&=\int_S\int_{-(1/2)}^{1/2}f(z_1,z_2)z_3\,\mathrm{d}z_3\,\mathrm{d}z_1\,\mathrm{d}z_2=0.\end{aligned} \tag{8.9}$$

We obtain (8.3) by taking the "lim inf" of (8.7) and using (8.9).

Therefore, the desired result follows if we can show that

$$\limsup_{h\to 0}\frac{1}{h}(e^{(h)}[\mathbf{y}^{(h)}]-e^0[\bar{\mathbf{y}},\bar{\mathbf{b}}])\leqslant 0. \tag{8.10}$$

Define $\hat{\mathbf{y}}^{(h)}:\Omega_1\to\mathbb{R}^3$ by

$$\hat{\mathbf{y}}^{(h)}(z_1,z_2,z_3)=\bar{\mathbf{y}}(z_1,z_2)+hz_3\bar{\mathbf{b}}(z_1,z_2). \tag{8.11}$$

Notice that it satisfies the boundary condition (8.1). Therefore,

$$\frac{1}{h}(e^{(h)}_1[\mathbf{y}^{(h)}]-e^0[\bar{\mathbf{y}},\bar{\mathbf{b}}])\leqslant\frac{1}{h}(e^{(h)}_1[\hat{\mathbf{y}}^{(h)}]-e^0[\bar{\mathbf{y}},\bar{\mathbf{b}}]) \tag{8.12}$$

since $\mathbf{y}^{(h)}$ is a minimizer of $e^{(h)}_1$. However, it is possible to show that

$$\frac{1}{h}\int_{-(1/2)}^{1/2}e^0[\bar{\mathbf{y}}+hz_3(\bar{\mathbf{b}}-\mathbf{b}_1),\bar{\mathbf{b}}]\,\mathrm{d}z_3=\frac{1}{h}e^{(h)}_1[\hat{\mathbf{y}}^{(h)}]+T^{(h)}+O(h) \tag{8.13}$$

for some $T^{(h)}$ which satisfies $\lim_{h\to 0} T^{(h)} = 0$, by expanding the left-hand-side and using an argument similar to (8.9). Substitute (8.13) in (8.12) and note that $\int_{-(1/2)}^{1/2} e^0[\bar{\mathbf{y}}, \bar{\mathbf{b}}]\, \mathrm{d}z_3 = e^0[\bar{\mathbf{y}}, \bar{\mathbf{b}}]$. Therefore,

$$\limsup_{h\to 0} \frac{1}{h}(e_1^{(h)}[\mathbf{y}^{(h)}] - e^0[\bar{\mathbf{y}}, \bar{\mathbf{b}}]) \leqslant \limsup_{h\to 0} \int_{-(1/2)}^{1/2} \left\{ \frac{1}{h}(e^0[\bar{\mathbf{y}} + hz_3(\bar{\mathbf{b}} - \mathbf{b}_1), \bar{\mathbf{b}}] - e^0[\bar{\mathbf{y}}, \bar{\mathbf{b}}]) \right\} \mathrm{d}z_3. \quad (8.14)$$

Since $\bar{\mathbf{y}}$, $\bar{\mathbf{b}}$ is a minimizer,

$$\lim_{h\to 0} \frac{1}{h}(e^0[\bar{\mathbf{y}} + h\mathbf{f}, \bar{\mathbf{b}} + h\mathbf{g}] - e^0[\bar{\mathbf{y}}, \bar{\mathbf{b}}]) = 0 \quad (8.15)$$

for any smooth $\mathbf{f}$, $\mathbf{g}$ that vanish on the boundary ∂S. For any z_3, set $\mathbf{f} = z_3(\bar{\mathbf{b}} - \mathbf{b}_1)$, $\mathbf{g} = \mathbf{0}$ in (8.15). Then we can use it to conclude that the limit exists on the right hand side of (8.14) and that we can switch the limit and the integral to obtain the desired inequality (8.10).

We now turn to the second order correction. It is useful to introduce the notation

$$P_2^{(h)} = \frac{1}{h^2}(e_1^{(h)}[\mathbf{y}^{(h)}] - e^0[\bar{\mathbf{y}}, \bar{\mathbf{b}}]). \quad (8.16)$$

Result 8.2. Suppose the minimizers $(\bar{\mathbf{y}}, \bar{\mathbf{b}})$ of e^0 satisfy the strong second variation condition: there exists $\varepsilon > 0$ such that

$$\frac{1}{h^2}\{e^0[\bar{\mathbf{y}} + h\mathbf{f}, \bar{\mathbf{b}} + h\mathbf{g}] - e^0[\bar{\mathbf{y}}, \bar{\mathbf{b}}]\} \geqslant \varepsilon \int_{\mathrm{S}} \{|\nabla_{\mathrm{p}}^2 \mathbf{f}|^2 + 2|\nabla_{\mathrm{p}} \mathbf{g}|^2\}\, \mathrm{d}z_1\, \mathrm{d}z_2 \quad (8.17)$$

for any smooth $\mathbf{f}$, $\mathbf{g}$ that vanish on the boundary ∂S and for every sufficiently small $h > 0$. Then,

$$\left.\begin{aligned} &\frac{1}{h}\nabla_{\mathrm{p}}^2(\mathbf{y}^{(h)} - \bar{\mathbf{y}} - hz_3\bar{\mathbf{b}}) \to \nabla_{\mathrm{p}}^2 \bar{\mathbf{c}} \\ &\frac{1}{h}\nabla_{\mathrm{p}}\left(\frac{1}{h}\mathbf{y}_{,3}^{(h)} - \bar{\mathbf{b}}\right) \to \nabla_{\mathrm{p}} \bar{\mathbf{d}} \\ &\frac{1}{h^3}\mathbf{y}_{,33}^{(h)} \to 0 \end{aligned}\right\} \quad \text{in } L^2(\Omega_1), \quad (8.18)$$

where $(\bar{\mathbf{c}}, \bar{\mathbf{d}})$ are independent of z_3 and $(\bar{\mathbf{c}}, \bar{\mathbf{d}})$ minimize the energy

$$b[\mathbf{c}, \mathbf{d}] = \int_{\Omega_1} \{\kappa(|\nabla_{\mathrm{p}}^2(\mathbf{c} + \bar{\mathbf{b}}z_3)|^2 + 2|\nabla_{\mathrm{p}}\mathbf{d}|^2) + \mathscr{Q}(\mathbf{c}_{,1} + \bar{\mathbf{b}}_{,1}z_3 | \mathbf{c}_{,2} + \bar{\mathbf{b}}_{,2}z_3 | \mathbf{d})\}\, \mathrm{d}\mathbf{z}, \quad (8.19)$$

among $(\mathbf{c}, \mathbf{d}) \in W_0^{2,2}(\mathrm{S}, \mathbb{R}^3) \times W_0^{1,2}(\mathrm{S}, \mathbb{R}^3)$. Here $\mathscr{Q}$ denotes the quadratic form

$$\mathcal{Q}(\mathbf{F}) = \frac{1}{2}\mathbf{F}\cdot\left(\left.\frac{\partial^2\varphi}{\partial\mathbf{F}^2}\right|_{(\bar{\mathbf{y}}_{,1}|\bar{\mathbf{y}}_{,2}|\bar{\mathbf{b}})}\right)\mathbf{F}. \tag{8.20}$$

Further,

$$\lim_{h\to 0} P_2^{(h)} = b[\bar{\mathbf{c}}, \bar{\mathbf{d}}]. \tag{8.21}$$

Proof. This proof uses an argument very similar to that of Proposition 3.1 to calculate the limit of $P_2^{(h)}$. In order to do so, we need to bound $P_2^{(h)}$ from above and below.

Our first task is to find an upper bound for $P_2^{(h)}$ which is independent of h. Define $\hat{\mathbf{y}}^{(h)}:\Omega_1\to\mathbb{R}^3$ by

$$\hat{\mathbf{y}}^{(h)}(z_1,z_2,z_3) = \bar{\mathbf{y}}(z_1,z_2) + hz_3\bar{\mathbf{b}}(z_1,z_2). \tag{8.22}$$

Notice that this function satisfies the boundary condition (8.1). Since $\mathbf{y}^{(h)}$ is a minimizer of $\mathbf{e}_1^{(h)}$, and $\int_{-(1/2)}^{1/2} e^0[\bar{\mathbf{y}},\bar{\mathbf{b}}]\,\mathrm{d}z_3 = e^0[\bar{\mathbf{y}},\bar{\mathbf{b}}]$,

$$P_2^{(h)} \leqslant \frac{1}{h^2}\left(e_1^{(h)}[\hat{\mathbf{y}}^{(h)}] - \int_{-(1/2)}^{1/2} e^0[\bar{\mathbf{y}},\bar{\mathbf{b}}]\,\mathrm{d}z_3\right). \tag{8.23}$$

Expanding $e^{(h)}[\hat{\mathbf{y}}^{(h)}]$ and cancelling some terms, this inequality reduces to

$$P_2^{(h)} \leqslant \frac{1}{h^2}\int_{\Omega_1}\{\kappa(2\nabla_{\mathrm{p}}^2\bar{\mathbf{y}}\cdot\nabla_{\mathrm{p}}^2(hz_3\bar{\mathbf{b}}) + |\nabla_{\mathrm{p}}^2(hz_3\bar{\mathbf{b}})|^2) + \varphi(\bar{\mathbf{y}}_{,1}+hz_3\bar{\mathbf{b}}_{,1}|\bar{\mathbf{y}}_{,2}+hz_3\bar{\mathbf{b}}_{,2}|\bar{\mathbf{b}}) - \varphi(\bar{\mathbf{y}}_{,1}|\bar{\mathbf{y}}_{,2}|\bar{\mathbf{b}})\}\,\mathrm{d}\mathbf{z}. \tag{8.24}$$

Since $(\bar{\mathbf{y}},\bar{\mathbf{b}})$ is a minimizer, it satisfies the first variation condition (8.15), or equivalently:

$$\int_{\mathrm{S}}\left\{2\kappa(\nabla_{\mathrm{p}}^2\bar{\mathbf{y}}\cdot\nabla_{\mathrm{p}}^2\mathbf{f} + 2\nabla_{\mathrm{p}}\bar{\mathbf{b}}\cdot\nabla_{\mathrm{p}}\mathbf{g}) + \left(\left.\frac{\partial\varphi}{\partial\mathbf{F}}\right|_{(\bar{\mathbf{y}}_{,1}|\bar{\mathbf{y}}_{,2}|\bar{\mathbf{b}})}\right)\cdot(\mathbf{f}_{,1}|\mathbf{f}_{,2}|\mathbf{g})\right\}\mathrm{d}z_1\,\mathrm{d}z_2 = 0 \tag{8.25}$$

for any smooth $\mathbf{f}$, $\mathbf{g}$ that vanish on the boundary $\partial\mathrm{S}$. For any given z_3, set $\mathbf{f} = hz_3(\bar{\mathbf{b}}-\mathbf{b}_1)$, $\mathbf{g}=0$ in the first variation condition (8.25), integrate (8.25) over z_3 from $-\frac{1}{2}$ to $\frac{1}{2}$ and subtract from (8.24):

$$\begin{aligned} P_2^{(h)} &\leqslant \frac{1}{h^2}\int_{\Omega_1}\Big\{\kappa(|\nabla_{\mathrm{p}}^2(hz_3\bar{\mathbf{b}})|^2 + 2\nabla_{\mathrm{p}}^2\bar{\mathbf{y}}\cdot\nabla_{\mathrm{p}}^2(hz_3\mathbf{b}_1)) \\ &\quad + \varphi(\bar{\mathbf{y}}_{,1}+hz_3\bar{\mathbf{b}}_{,1}|\bar{\mathbf{y}}_{,2}+hz_3\bar{\mathbf{b}}_{,2}|\bar{\mathbf{b}}) - \varphi(\bar{\mathbf{y}}_{,1}|\bar{\mathbf{y}}_{,2}|\bar{\mathbf{b}}) - \left(\frac{\partial\varphi}{\partial\mathbf{F}}\right)\cdot(hz_3\bar{\mathbf{b}}_{,1}|hz_3\bar{\mathbf{b}}_{,2}|0) \\ &\quad + hz_3\left(\frac{\partial\varphi}{\partial\mathbf{F}}\right)\cdot(\mathbf{b}_{1,1}|\mathbf{b}_{1,2}|0)\Big\}\mathrm{d}\mathbf{z} \\ &= \frac{1}{h^2}\int_{\Omega_1}\{\kappa h^2 z_3^2|\nabla_{\mathrm{p}}^2\bar{\mathbf{b}}|^2 + h^2 z_3^2\mathcal{Q}(\bar{\mathbf{b}}_{,1}|\bar{\mathbf{b}}_{,2}|0) + O(h^3)\}\,\mathrm{d}\mathbf{z} \end{aligned}$$

$$+\frac{1}{h}\int_{\Omega_1}\left\{2\kappa(\nabla_p^2\bar{\mathbf{y}})\cdot(\nabla_p^2\mathbf{b}_1)+\left(\frac{\partial\varphi}{\partial\mathbf{F}}\right)\cdot(\mathbf{b}_{1,1}|\mathbf{b}_{1,2}|0)\right\}z_3\,\mathrm{d}\mathbf{z}$$

$$=\int_{\Omega_1}\{\kappa z_3^2|\nabla_p^2\bar{\mathbf{b}}|^2+z_3^2\mathscr{Q}(\bar{\mathbf{b}}_{,1}|\bar{\mathbf{b}}_{,2}|0)\}\,\mathrm{d}\mathbf{z}+\int_{\Omega_1}f(z_1,z_2)z_3\,\mathrm{d}\mathbf{z}+O(h)$$

$$=\int_{\Omega_1}\{\kappa z_3^2|\nabla_p^2\bar{\mathbf{b}}|^2+z_3^2\mathscr{Q}(\bar{\mathbf{b}}_{,1}|\bar{\mathbf{b}}_{,2}|0)\}\,\mathrm{d}\mathbf{z}+O(h). \tag{8.26}$$

Above, $(\partial\varphi/\partial\mathbf{F})$ is always evaluated at $(\bar{\mathbf{y}}_{,1}|\bar{\mathbf{y}}_{,2}|\bar{\mathbf{b}})$. We have used the Taylor expansion of φ and the definition (8.20) to obtain the first equality. Notice that the integral in the final line is independent of h. Therefore, we can conclude that there exists a constant C independent of h such that

$$P_2^{(h)}\leqslant C \tag{8.27}$$

for h small enough.

We now use the strong second variation condition (8.17) to find a lower bound for $P_2^{(h)}$. Set

$$\mathbf{c}^{(h)}=\frac{1}{h}(\mathbf{y}^{(h)}-\bar{\mathbf{y}}-hz_3\bar{\mathbf{b}}),\quad \mathbf{d}^{(h)}=\frac{1}{h}\left(\frac{1}{h}\mathbf{y}_{,3}^{(h)}-\bar{\mathbf{b}}\right). \tag{8.28}$$

For any z_3, set $\mathbf{f}=\mathbf{c}^{(h)}$ and $\mathbf{g}=\mathbf{d}^{(h)}$ in (8.17) and integrate over z_3 from $-\frac{1}{2}$ to $\frac{1}{2}$:

$$\frac{1}{h^2}\left\{\int_{-(1/2)}^{1/2}e^0\left[\mathbf{y}^{(h)}-hz_3\bar{\mathbf{b}},\frac{1}{h}\mathbf{y}_{,3}^{(h)}\right]\mathrm{d}z_3-e^0[\bar{\mathbf{y}},\bar{\mathbf{b}}]\right\}\geqslant\varepsilon\int_{\Omega_1}\{|\nabla_p^2\mathbf{c}^{(h)}|^2+2|\nabla_p\mathbf{d}^{(h)}|^2\}\,\mathrm{d}\mathbf{z}. \tag{8.29}$$

By expansion,

$$\int_{-(1/2)}^{1/2}e^0\left[\mathbf{y}^{(h)}-hz_3\bar{\mathbf{b}},\frac{1}{h}\mathbf{y}_{,3}^{(h)}\right]\mathrm{d}z_3=e_1^{(h)}[\mathbf{y}^{(h)}]-\frac{1}{h^4}\|\mathbf{y}_{,33}^{(h)}\|_{L^2}^2-T_1^{(h)}+T_2^{(h)}+O(h^3), \tag{8.30}$$

where

$$T_1^{(h)}=h\int_{\Omega_1}\left\{2\kappa(\nabla_p^2\mathbf{y}^{(h)})\cdot(\nabla_p^2\bar{\mathbf{b}})+\left(\frac{\partial\varphi}{\partial\mathbf{F}}\right)\cdot(\bar{\mathbf{b}}_{,1}|\bar{\mathbf{b}}_{,2}|\mathbf{0})\right\}z_3\,\mathrm{d}\mathbf{z},$$

$$T_2^{(h)}=h^2\int_{\Omega_1}\left\{\kappa|\nabla_p\bar{\mathbf{b}}|^2+(\bar{\mathbf{b}}_{,1}|\bar{\mathbf{b}}_{,2}|\mathbf{0})\cdot\left(\frac{\partial^2\varphi}{\partial\mathbf{F}^2}\right)(\bar{\mathbf{b}}_{,1}|\bar{\mathbf{b}}_{,2}|\mathbf{0})\right\}z_3^2\,\mathrm{d}\mathbf{z}. \tag{8.31}$$

Above, both $(\partial\varphi/\partial\mathbf{F})$ and $(\partial^2/\partial\mathbf{F}^2)$ are evaluated at $(\mathbf{y}_{,1}^{(h)}|\mathbf{y}_{,2}^{(h)}|(1/h)\mathbf{y}_{,3}^{(h)})$. Substituting (8.30) in (8.29), we can conclude that

$$P_2^{(h)} - \frac{1}{h^2} T_1^{(h)} + \frac{1}{h^2} T_2^{(h)} + O(h) \geqslant \varepsilon' \left(\|\nabla_{\mathrm{p}}^2 \mathbf{c}^{(h)}\|_{L^2}^2 + \|\nabla_{\mathrm{p}} \mathbf{d}^{(h)}\|_{L^2}^2 + \left\| \frac{1}{h^3} \mathbf{y}_{,33}^{(h)} \right\|_{L^2}^2 \right). \tag{8.32}$$

We now estimate the terms on the left hand side. First notice from (8.31) that $\lim_{h\to 0}(1/h^2)T_2^{(h)}$ is finite so that there exists $c > 0$ such that

$$\frac{1}{h^2} T_2^{(h)} < c \tag{8.33}$$

for all h small enough. We now turn to $T_1^{(h)}$ in (8.31). Note that $\mathbf{y}^{(h)} = \bar{\mathbf{y}} + h\mathbf{c}^{(h)} + hz_3\bar{\mathbf{b}}$ and $(1/h)\mathbf{y}_{,3}^{(h)} = h\mathbf{d}^{(h)} + \bar{\mathbf{b}}$. Use the Taylor expansion of $(\partial\varphi/\partial\mathbf{F})$ and some rearrangement to conclude that

$$\begin{aligned}
\frac{1}{h^2} T_1^{(h)} = {} & \frac{1}{h} \int_{\Omega_1} \left\{ 2\kappa(\nabla_{\mathrm{p}}^2 \bar{\mathbf{y}}) \cdot (\nabla_{\mathrm{p}}^2 \bar{\mathbf{b}}) + \left(\left. \frac{\partial \varphi}{\partial \mathbf{F}} \right|_{(\bar{\mathbf{y}}_{,1} | \bar{\mathbf{y}}_{,2} | \bar{\mathbf{b}})} \right) \cdot (\bar{\mathbf{b}}_{,1} | \bar{\mathbf{b}}_{,2} | \mathbf{0}) \right\} z_3 \, \mathrm{d}\mathbf{z} \\
& + \int_{\Omega_1} 2\kappa(\nabla_{\mathrm{p}}^2 \mathbf{c}^{(h)}) \cdot (\nabla_{\mathrm{p}}^2 \bar{\mathbf{b}}) z_3 \, \mathrm{d}\mathbf{z} \\
& + \frac{1}{2} \int_{\Omega_1} (\mathbf{c}_{,1}^{(h)} | \mathbf{c}_{,2}^{(h)} | \mathbf{d}^{(h)}) \cdot \left(\left. \frac{\partial^2 \varphi}{\partial \mathbf{F}^2} \right|_{\mathbf{G}_\lambda} \right) (\bar{\mathbf{b}}_{,1} | \bar{\mathbf{b}}_{,2} | \mathbf{0}) z_3 \, \mathrm{d}\mathbf{z} \\
& + \int_{\Omega_1} \left\{ 2\kappa |\nabla_{\mathrm{p}}^2 \bar{\mathbf{b}}|^2 + \frac{1}{2} (\bar{\mathbf{b}}_{,1} | \bar{\mathbf{b}}_{,2} | \mathbf{0}) \cdot \left(\left. \frac{\partial^2 \varphi}{\partial \mathbf{F}^2} \right|_{\mathbf{G}_\lambda} \right) (\bar{\mathbf{b}}_{,1} | \bar{\mathbf{b}}_{,2} | \mathbf{0}) \right\} z_3^2 \, \mathrm{d}\mathbf{z},
\end{aligned} \tag{8.34}$$

where $\mathbf{G}_\lambda = \lambda^{(h)}(\mathbf{y}_{,1}^{(h)} | \mathbf{y}_{,2}^{(h)} | (1/h)\mathbf{y}_{,3}^{(h)}) + (1-\lambda^{(h)})(\bar{\mathbf{y}}_{,1} | \bar{\mathbf{y}}_{,2} | \bar{\mathbf{b}})$ for some $\lambda^{(h)} \in (0, 1)$. Notice that the first integral above is zero since the integrand is of the form $f(z_1, z_2)z_3$, and the fourth has a finite limit as $h \to 0$. Using Hölder's inequality on the second and the third integral we can conclude that there exists $c > 0$ independent of h such that

$$\left| \frac{1}{h^2} T_1^{(h)} \right| \leqslant c(1 + \|\nabla_{\mathrm{p}}^2 \mathbf{c}^{(h)}\|_{L^2} + \|\nabla_{\mathrm{p}} \mathbf{c}^{(h)}\|_{L^2} + \|\mathbf{d}^{(h)}\|_{L^2}) \tag{8.35}$$

for any h small enough. We now need the following Poincaré inequality which can be proved using methods described for example in Nečas (1983). Given $\mathbf{u}_0 : \Omega_1 \to \mathbb{R}^3$ there exist constants, $\tilde{c}_1$, $\tilde{c}_2$ such that

$$\int_{\Omega_1} \{|\mathbf{u}|^2 + |\nabla \mathbf{u}|^2\} \, \mathrm{d}\mathbf{z} \leqslant \tilde{c}_1 \int_{\Omega_1} |\nabla^2 \mathbf{u}|^2 \, \mathrm{d}\mathbf{z} + \tilde{c}_2$$

$$\text{for every } \mathbf{u} \in W^{2,2}(\Omega_1, \mathbb{R}^3) \text{ with } \mathbf{u}|_{\partial \mathrm{S} \times (-(1/2), 1/2)} = \mathbf{u}_0. \tag{8.36}$$

Using (8.36) and (3.7) in (8.35), we can conclude that there exists $c > 0$ independent of h such that

$$\left|\frac{1}{h^2}T_1^{(h)}\right| \leqslant c(1+\|\nabla_{\mathrm{p}}^2\mathbf{c}^{(h)}\|_{L^2}+\|\nabla_{\mathrm{p}}\mathbf{d}^{(h)}\|_{L^2}) \tag{8.37}$$

for any h small enough.

Substitute (8.27), (8.33) and (8.37) in (8.32): for any h small enough,

$$C'(1+\|\nabla_{\mathrm{p}}^2\mathbf{c}^{(h)}\|_{L^2}+\|\nabla_{\mathrm{p}}\mathbf{d}^{(h)}\|_{L^2}) \geqslant \|\nabla_{\mathrm{p}}^2\mathbf{c}^{(h)}\|_{L^2}^2+\|\nabla_{\mathrm{p}}\mathbf{d}^{(h)}\|_{L^2}^2+\left\|\frac{1}{h^3}\mathbf{y}_{,33}^{(h)}\right\|_{L^2}^2, \tag{8.38}$$

where $C'>0$ is independent of h. Notice that we have first powers on the left and squares on the right. Therefore, we conclude that

$$\|\nabla_{\mathrm{p}}^2\mathbf{c}^{(h)}\|_{L^2}^2 \leqslant C'', \quad \|\nabla_{\mathrm{p}}\mathbf{d}^{(h)}\|_{L^2}^2 \leqslant C'', \quad \left\|\frac{1}{h^3}\mathbf{y}_{,33}^{(h)}\right\|_{L^2}^2 \leqslant C'', \tag{8.39}$$

where $C''>0$ is independent of h. We can add the following to the list.

$$\|\nabla^2\mathbf{c}^{(h)}\|_{L^2}^2 \leqslant \hat{C}, \quad \|\nabla\mathbf{d}^{(h)}\|_{L^2}^2 \leqslant \hat{C}, \quad \|\mathbf{c}^{(h)}\|_{L^2}^2+\|\nabla\mathbf{c}^{(h)}\|_{L^2}^2 \leqslant \hat{C}, \quad \|\mathbf{d}^{(h)}\|_{L^2}^2 \leqslant \hat{C} \tag{8.40}$$

for some $\hat{C}$ independent of h. The first two $(8.40)_{1,2}$ follow by combining the three equations in (8.39). The third $(8.40)_3$ is a consequence of $(8.40)_1$ and the Poincaré inequality (8.36). Finally, $(8.40)_4$ follows from $(8.40)_1$ and the Poincaré inequality (3.7).

Therefore, there is a subsequence (not relabeled) such that

$$\begin{aligned}
&\mathbf{c}^{(h)}=\frac{1}{h}(\mathbf{y}^{(h)}-\bar{\mathbf{y}}-hz_3\bar{\mathbf{b}})\rightharpoonup\bar{\mathbf{c}} \quad \text{in } W^{2,2},\\
&\mathbf{d}^{(h)}=\frac{1}{h}\left(\frac{1}{h}\mathbf{y}_{,3}^{(h)}-\bar{\mathbf{b}}\right)\rightharpoonup\bar{\mathbf{d}} \quad \text{in } W^{1,2}.
\end{aligned} \tag{8.41}$$

Using (8.39) and the fact that Ω_1 is convex in the z_3-direction, it follows that $\bar{\mathbf{c}}$, $\bar{\mathbf{d}}$ are independent of z_3. It is also clear that $\bar{\mathbf{c}}$, $\bar{\mathbf{d}}$ satisfy zero boundary conditions on ∂S. Write

$$\begin{aligned}
&\frac{1}{h}\nabla_{\mathrm{p}}^2(\mathbf{y}^{(h)}-\bar{\mathbf{y}}-hz_3\bar{\mathbf{b}})=\nabla_{\mathrm{p}}^2\bar{\mathbf{c}}+\mathbf{E}_{\mathrm{p}}^{(h)}, \quad \mathbf{E}_{\mathrm{p}}^{(h)}\rightharpoonup 0 \quad \text{in } L^2,\\
&\frac{1}{h}\nabla\left(\frac{1}{h}\mathbf{y}_{,3}^{(h)}-\bar{\mathbf{b}}\right)=\nabla\bar{\mathbf{d}}+\mathbf{E}_3^{(h)}, \quad \mathbf{E}_3^{(h)}\rightharpoonup 0 \quad \text{in } L^2.
\end{aligned} \tag{8.42}$$

We will now show that the convergence in (8.41) or, equivalently, (8.42) is strong and we can pass to the limit in the energy. Set

$$\hat{\mathbf{y}}^{(h)}=(\bar{\mathbf{y}}+h\bar{\mathbf{c}})+(h\bar{\mathbf{b}}+h^2\bar{\mathbf{d}})z_3, \tag{8.43}$$

and notice that it satisfies the boundary conditions (8.1). Now, since $\mathbf{y}^{(h)}$ is a minimizer of $e_1^{(h)}$,

$$\frac{1}{h^2}\left(e_1^{(h)}[\mathbf{y}^{(h)}]-\int_{-(1/2)}^{1/2} e^0[\bar{\mathbf{y}},\bar{\mathbf{b}}]\,\mathrm{d}z_3\right)\leqslant\frac{1}{h^2}\left(e_1^{(h)}[\hat{\mathbf{y}}^{(h)}]-\int_{-(1/2)}^{1/2} e^0[\bar{\mathbf{y}},\bar{\mathbf{b}}]\,\mathrm{d}z_3\right). \tag{8.44}$$

Now, expand both sides, use the first variation condition (8.25), use the Taylor expansion of φ to obtain

$$\int_{\Omega_1}\kappa\left[|\mathbf{E}_{\mathrm{p}}^{(h)}|^2+|\mathbf{E}_3^{(h)}|^2+\frac{1}{h^6}|\mathbf{y}_{,33}^{(h)}|^2\right]\mathrm{d}\mathbf{z}\leqslant\frac{1}{h^2}\int_{\Omega_1}\left\{\mathscr{Q}(h(\bar{\mathbf{c}}+z_3\bar{\mathbf{b}})_{,1}|h(\bar{\mathbf{c}}+z_3\bar{\mathbf{b}})_{,2}|h\bar{\mathbf{d}})\right.$$
$$\left.-\mathscr{Q}\left((\mathbf{y}^{(h)}-\bar{\mathbf{y}})_{,1}|(\mathbf{y}^{(h)}-\bar{\mathbf{y}})_{,2}|\left(\frac{1}{h}\mathbf{y}_{,3}^{(h)}-\bar{\mathbf{b}}\right)\right)+O(h^3)\right\}\mathrm{d}\mathbf{z}. \tag{8.45}$$

Recall (8.41) and notice that the right-hand-side goes to zero as $h\to 0$. Therefore, the convergence in (8.41) and (8.42) have been improved to strong. This establishes the result (8.21). We can obtain the minimum principle by using a test function $\hat{\mathbf{y}}^{(h)}=(\bar{\mathbf{y}}+h\mathbf{c})+(h\bar{\mathbf{b}}+h^2\mathbf{d})z_3$ in (8.44) for any $(\mathbf{c},\mathbf{d})\in W^{2,2}(\mathrm{S},\mathbb{R}^3)\times W^{1,2}(\mathrm{S},\mathbb{R}^3)$ that vanish on $\partial\mathrm{S}$ and then by letting $h\to 0$.

Result 8.2 introduces two additional vector fields $\mathbf{c}:\mathrm{S}\to\mathbb{R}^3$, $\mathbf{d}:\mathrm{S}\to\mathbb{R}^3$. The form of the test function gives meaning to these vectors:

$$\mathbf{y}^{(h)}(\mathbf{z})\doteq(\mathbf{y}(z_1,z_2)+h\mathbf{c}(z_1,z_2))+h(\mathbf{b}(z_1,z_2)+h\mathbf{d}(z_1,z_2))z_3, \tag{8.46}$$

where the error is small in the sense of (3.8) and (8.18). Alternately, in terms of the original variables

$$\tilde{\mathbf{y}}^{(h)}(\mathbf{x})\doteq(\mathbf{y}(x_1,x_2)+h\mathbf{c}(z_1,z_2))+(\mathbf{b}(x_1,x_2)+h\mathbf{d}(x_1,x_2))x_3. \tag{8.47}$$

Therefore, $\mathbf{c}$ is the $O(h)$ correction to $\mathbf{y}$, the deformation of the middle surface of the film, while $\mathbf{d}$ is the $O(h)$ correction to the Cosserat vector $\mathbf{b}$.

Further, Result 8.2 tells us that

$$e_1^{(h)}[\mathbf{y}^{(h)}]\doteq e^0[\mathbf{y},\mathbf{b}]+h^2b[\mathbf{c},\mathbf{d}] \tag{8.48}$$

or in terms of the original energy

$$e^{(h)}[\tilde{\mathbf{y}}^{(h)}]\doteq he^0[\mathbf{y},\mathbf{b}]+h^3b[\mathbf{c},\mathbf{d}]. \tag{8.49}$$

Thus, the correction to e^0 is two orders higher and is given by b in (8.19). The energy b consists of some interfacial terms and the bending energy. Notice that the bending energy depends implicitly on the minimizers $(\bar{\mathbf{y}},\bar{\mathbf{b}})$ of e^0. According to our results, one should not minimize the sum e^0+h^2b with respect to $(\mathbf{y},\mathbf{b},\mathbf{c},\mathbf{d})$. Instead, one should carry out a two-step minimization: first minimize e^0 with respect to $(\mathbf{y},\mathbf{b})$ to obtain the minimizers $(\bar{\mathbf{y}},\bar{\mathbf{b}})$. Then, holding $(\bar{\mathbf{y}},\bar{\mathbf{b}})$ fixed minimize b with respect to $(\mathbf{c},\mathbf{d})$.

Finally, we remark on the additional assumptions in this section. First, we believe that we do not need all the smoothness conditions: since $\mathbf{y}^{(h)}$ as well as $(\bar{\mathbf{y}},\bar{\mathbf{b}})$ are

minimizers, they possess additional regularity. Second, it appears that the strong second variation condition is essential.

The predictions of this theory are a little nonstandard, though evidently appropriate for many thin film problems. That is, given the form of the minimizer (8.46), the use of the higher order theory can at most give a small perturbation to the membrane solution. Consider, for example, a typical problem of a horizontal beam or strip, built-in at one end and loaded by a vertical force $h\mathbf{f}$ at the other end. The membrane solution is expected to cause the beam to hang nearly straight down, the sharp corner at the root of the beam smoothed slightly by the small interfacial energy. The higher order theory can at most perturb this slightly. This looks little like the classical picture of a bent beam. To get a more typical picture, one might consider forces of the form $h^3\mathbf{f}$. In this case the membrane solution is expected to ignore the force altogether, and therefore, because of the positive-definite gradient terms, yield a straight horizontal beam as the unique minimizer. Again, this can at most be perturbed slightly by the bending terms, again deviating from the classical picture (at least for moderate forces). This feature of our theory can be attributed to the dominating effect of gradient-type surface energy at small scales.

To obtain a more classical macroscopic picture, it seems that one should also let $\kappa \to 0$ along with h. We conjecture that there is some regime of the type $(\kappa, h) \to 0$ for which one obtains exactly our forms of the membrane and bending energies e^0 and b, and the sequential minimization property, but with $\kappa = 0$. In the beam problem discussed above with the force $h^3\mathbf{f}$, this would lead to a minimization of the bending energy among all the zero energy deformations of the $\kappa = 0$ membrane theory, i.e., the "floppy modes" discussed in Section 5. This seems reasonable physically, for macroscopic problems. The verification of this conjecture certainly would involve deep analytical issues, if approached by our methods, as the existence of floppy modes are expected to nullify any kind of strong second variation condition, upon which our argument is essentially based.

Acknowledgements

We are grateful to Irene Fonseca for her useful suggestions concerning the treatment of boundary conditions in Section 8. This work was supported by AFOSR (K.B.: F49620-95-1-0109 and R.D.J.: F49620-97-1-0187), ONR/DARPA (R.D.J.: N/N00014-92-J-4034 and N00014-95-1-1145) and NSF (K.B.: CMS-9457573 and R.D.J.: DMS-9505077).

Appendix: conditions on lattice parameters for austenite/martensite interfaces in thin films

The following evaluation refers to the third paragraph of Section 7, in which we discuss conditions on lattice parameters that permit an exact austenite/martensite interface in the thin film theory with $\kappa = 0$. These conditions are embodied in (5.12)

evaluated for the distortion matrices (2.6)–(2.10) and various film normals. Because each set of distortion matrices in (2.6)–(2.10) is invariant under conjugation by the cubic group, it is only necessary to evaluate (5.12) for the first distortion matrix in each case, as long as we use all crystallographically equivalent film normals (under the cubic group). In summary, the following is obtained by evaluating and simplifying (5.12) with $\mathbf{V} = \mathbf{I}$ [$\mathbf{U}$ = the first matrix in each of (2.6)–(2.10)] and $\mathbf{e}_3$ taking on values in the families $\{100\}$, $\{110\}$ and $\{111\}$. In each case there are two interfaces with strict inequality, which degenerate to one or an infinity of interfaces (any line in the plane) with equality; see Proposition 5.1.

A. Cubic to tetragonal transformations as in (2.6). There exist austenite/martensite interfaces if the following conditions are satisfied:

For (100) films:	$\eta_1 = 1$,
For (010) or (001) films:	$(\eta_1^2-1)(\eta_2^2-1) \leqslant 0$,
For (110), (1-10), (101), (10-1) films:	$(\eta_1^2-1)(\eta_1^2+\eta_2^2-2) \leqslant 0$,
For (011), (01-1) films:	$(\eta_1^2-1)(\eta_2^2-1) \leqslant 0$,
For all $\{111\}$ films:	$(\eta_1^2-1)(2\eta_2^2+\eta_1^2-3) \leqslant 0$.

B. Cubic to orthorhombic transformations as in (2.7). There exist austenite/martensite interfaces if the following conditions are satisfied:

For (100) or (010) films:	$(\beta^2-1)(\alpha^2+\gamma^2-2) \leqslant 0$,
For (001) films:	$(\alpha^2-1)(\gamma^2-1) \leqslant 0$,
For (110) films:	$(\beta^2-1)(\gamma^2-1) \leqslant 0$,
For (1-10) films:	$(\beta^2-1)(\alpha^2-1) \leqslant 0$,
For (101), (10-1), (011), (01-1) films:	$(\beta^2-1)(\alpha^2+\gamma^2-2)+2(\gamma^2-1)(\alpha^2-1) \leqslant 0$,
For (111) or (11-1) films:	$(\alpha^2+2\beta^2-3)(\gamma^2-1) \leqslant 0$,
For (-111) or (1-11) films:	$(\gamma^2+2\beta^2-3)(\alpha^2-1) \leqslant 0$.

C1. Cubic to monoclinic transformations as in (2.9). There exist austenite/martensite interfaces if the following conditions are satisfied:

For (100) films:	$1-2\delta^2+\delta^4-2\alpha\delta^2\gamma-\gamma^2+\alpha^2(\gamma^2-1) \leqslant 0$,
For (010) films:	$(\beta^2-1)(\delta^2+\gamma^2-1) \leqslant 0$,
For (001) films:	$(\beta^2-1)(\delta^2+\alpha^2-1) \leqslant 0$,
For (110) or (1-10) films:	$2-3\delta^2+\delta^4-2\alpha\delta^2\gamma-2\gamma^2+\alpha^2(\gamma^2-1)$ $+\beta^2(\delta^2+\gamma^2-1) \leqslant 0$,
For (101) or (10-1) films:	$2-3\delta^2+\delta^4+\beta^2(\delta^2-1)-2\alpha\delta^2\gamma-\gamma^2$ $+\alpha^2(\gamma^2+\beta^2-2) \leqslant 0$,
For (011) films:	$(\beta^2-1)(\alpha^2-2\alpha\delta+2\delta^2-2\delta\gamma+\gamma^2-2) \leqslant 0$,
For (01-1) films:	$(\beta^2-1)(\alpha^2+2\alpha\delta+2\delta^2+2\delta\gamma+\gamma^2-2) \leqslant 0$,
For (111) or (-111) films:	$3-4\delta^2+\delta^4+2\delta\gamma-2\gamma^2-2\alpha\delta(\beta^2+\delta\gamma-1)$ $+\alpha^2(\beta^2+\gamma^2-2)+\beta^2(2\delta^2-2\delta\gamma+\gamma^2-2) \leqslant 0$,
For (1-11) or (11-1) films:	$3-4\delta^2+\delta^4-2\delta\gamma-2\gamma^2-2\alpha\delta(-\beta^2+\delta\gamma+1)$ $+\alpha^2(\beta^2+\gamma^2-2)+\beta^2(2\delta^2+2\delta\gamma+\gamma^2-2) \leqslant 0$,

C2. Cubic to monoclinic transformations as in (2.10). There exist austenite/martensite interfaces if the following conditions are satisfied:

For (100) or (010) films: $\beta^2(\delta^2-1)-2\beta\delta\varepsilon^2-2\alpha(\beta+\delta)\varepsilon^2 + \alpha^2(\beta^2+\varepsilon^2-1)+(\varepsilon^2-1)(\delta^2+2\varepsilon^2-1)\leqslant 0,$

For (001) films: $(\delta^2+\alpha^2-2\alpha\delta-1)(\delta^2+\alpha^2+2\varepsilon^2+2\alpha\delta-1)\leqslant 0,$

For (110) films: $(\delta^2+\alpha^2-2\alpha\delta-1)(\beta^2+2\varepsilon^2-1)\leqslant 0,$

For (1-10) films: $1+\alpha^2(\beta^2-1)-\delta^2+\beta^2(\delta^2-1)-4\varepsilon^2-4\beta\delta\varepsilon^2 + 4\varepsilon^4+2\alpha((\beta^2-1)\delta-2\beta\varepsilon^2)\leqslant 0,$

For (101) or (011) films: $2+\alpha^4-3\delta^2+\delta^4+\beta^2(\delta^2-1)-2\alpha^3\varepsilon+2\delta\varepsilon-2\delta^3\varepsilon - 5\varepsilon^2+3\delta^2\varepsilon^2+2\varepsilon^4+2\alpha\varepsilon(1+2\beta\delta+\delta^2-\beta\varepsilon - 3\delta\varepsilon)-2\beta\varepsilon(\delta^2+\delta\varepsilon-1)+\alpha^2(\beta^2-2\delta^2-2\beta\varepsilon + 2\delta\varepsilon+3\varepsilon^2-3)\leqslant 0,$

For (10-1) or (01-1) films: $2+\alpha^4-3\delta^2+\delta^4+\beta^2(\delta^2-1)+2\alpha^3\varepsilon-2\delta\varepsilon+2\delta^3\varepsilon - 5\varepsilon^2+3\delta^2\varepsilon^2+2\varepsilon^4-2\alpha\varepsilon(1+2\beta\delta+\delta^2+\beta\varepsilon + 3\delta\varepsilon)-2\beta\varepsilon(-\delta^2+\delta\varepsilon+1)+\alpha^2(\beta^2-2\delta^2+2\beta\varepsilon - 2\delta\varepsilon+3\varepsilon^2-3)\leqslant 0,$

For (111) films: $(\delta^2+\alpha^2-2\alpha\delta-1)(\alpha^2+2\beta^2+\delta^2+2\alpha(\delta-2\varepsilon) - 4\beta\varepsilon-4\delta\varepsilon+6\varepsilon^2-3)\leqslant 0,$

For (-111) or (1-11) films: $3+\alpha^4-4\delta^2+\delta^4+2\beta^2(\delta^2-1)-10\varepsilon^2-8\beta\delta\varepsilon^2 + 2\delta^2\varepsilon^2+8\varepsilon^4+2\alpha^2(\beta^2-\delta^2+\varepsilon^2-2) + 4\alpha(-2\beta\varepsilon^2+\delta(\beta^2-\varepsilon^2-1))\leqslant 0,$

For (11-1) films: $(\delta^2+\alpha^2-2\alpha\delta-1)(\alpha^2+2\beta^2+\delta^2+2\alpha(\delta+2\varepsilon) + 4\beta\varepsilon+4\delta\varepsilon+6\varepsilon^2-3)\leqslant 0.$

References

Acerbi, E., Buttazzo, G., Percivale, D., 1991. A variational definition for the strain energy of an elastic string. Journal of Elasticity 25, 137–148.

Anzellotti, G., Baldo, S., Percivale, D., 1994. Dimension reduction in variational problems, asymptotic development in Γ-convergence and thin structures in elasticity. Asympt. Anal. 9, 61–100.

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

Ball, J.M., James, R.D., 1992. Proposed experimental tests of a theory of fine microstructure and the two-well problem. Phil. Trans. Royal Soc. London A338, 389–450.

Ball, J.M., Chu, C., James, R.D., 1995. Hysteresis during stress-induced variant rearrangement. Proceedings of the International Conference on Martensitic Transformations 1995, Journal de Physique IV, Colloque C8 5, 245–251.

Barsch, G.R., Krumhansl, J.A., 1992. Nonlinear physics in martensitic transformations. In Martensite, ed. G.B. Olson, W.S. Owen, pp. 125–147. ASM International.

Barsch, G.R., Krumhansl, J.A., 1988. Nonlinear and nonlocal continuum model of transformation precursors in martensites. Met. Trans. 19A, 761–775.

Bensaoula, A.H., Chen, L.-C., Caldwell, D.A., Palmstrm, C.J., James, R.D., 1997. Molecular beam epitaxy growth of the ferromagnetic compound Ni_2MnGa on $Ga_{(1-x)}In_xAs$, in preparation.

Bhattacharya, K., 1993. A comparison of geometrically nonlinear and linear theories of martensitic transformation. Continuum Mech. Thermodyn. 5, 205–242.

Bhattacharya, K., James, R.D., 1996. A theory of shape-memory thin films with applications. In: Materials for Smart Systems II, ed. E.P. George et al., MRS Proceedings 459, 311–316.
Bhattacharya, K., Kohn, R.V., 1996. Symmetry, texture and the recoverable strain of shape-memory polycrystals. Acta Materialia 44, 529–542.
Chakravorty, S., Wayman, C.M., 1977. Electron microscopy of internally faulted Cu–Zn–Al martensite. Acta Metallurgica 25, 989–1000.
DeSimone, A., James, R.D., Palmstrm, C., Bensaoula, A.H., 1997. Thin films of ferromagnetic shape memory single crystals, with applications, in preparation.
Ericksen, J.L., 1992. Bifurcation and martensitic transformations in Bravais lattices. Journal of Elasticity 28, 55–78.
Evans, L.C., Gariepy, R.F., 1992. Measure Theory and Fine Properties of Functions. CRC Press.
Fonseca, I., Francfort, G., 1998. 3D-2D asymptotic analysis of an optimal design problem for thin films. Max Planck Institute for MIS preprint No. 8. J. Reine Angew. Math., to appear.
Freund, L.B., Nix, W. D., 1996. Critical thickness condition for a strained compliant substrate/epitaxial film system. Applied Physics Letters 69, 173–175.
Fox, D., Raoult, A., Simo, J.C., 1992. Modèles asymptotic invariants pour des structures minces élastiques. C. R. Acad. Sci. Paris 315, Series I, 235–240.
Fox, D., Raoult, A., Simo, J.C., 1993. A justification of nonlinear properly invariant plate theories. Archive for Rational Mechanics and Analysis 124, 157–199.
Grummon, D.S., Hou, L., Zhao, Z., Pence, T.J., 1995. Progress on sputter-deposited thermotractive titanium-nickel films. Procedures of the International Conference on Martensitic Transformations 1995, Journal de Physique IV, Colloque C8 5, 665–670.
Hane, K., 1997. Bulk and thin film microstructures in untwinned martensites, preprint.
Hashinaga, T., Miyazaki, S., Ueki, T., Horikawa, H., 1995. Transformation and deformation behavior in sputter-deposited Ti–Ni–Cu thin films. Proceedings of the International Conference on Martensitic Transformations 1995, Journal de Physique IV, Colloque C8 5, 689–694.
Hou, L., Pence, T.J., Grummon, D.S., 1995. Structure and thermal stability in titanium–nickel thin films sputtered at elevated-temperature on inorganic and polymeric substrates. Materials Research Society Symposium 360, 369–374.
James, R.D., Kinderlehrer, D., 1990. Theory of diffusionless phase transformations. In Proceedings of Équations à Derivées Partielles et Modèles Continues de Transitions de Phases. Lecture Notes in Physics 344, 51–84.
James, R.D., Wuttig, M., 1996. Alternative smart materials. Proceedings of the SPIE Symposium on Smart Structures and Materials, ed. V.V. Varadan, J. Chandra, 2715, 420–426.
James, R.D., Wuttig, M., 1998. Magnetostriction of martensite. Philosophical Magazine A 77, 1273–1299.
Knowles, K.M., Smith, D.A., 1981. The crystallography of the martensitic transformation in equiatomic nickel–titanium. Acta Metallurgica 29, 101–110.
Kreyszig, I., 1968. Introduction to Differential Geometry and Riemannian Geometry. University of Toronto Press.
Krulevitch, P., Ramsey, P.B., Makowiecki, D.M., Lee, A.P., Northrup, M.A., Johnson, G.C., 1996. Mixed sputter deposition of Ni–Ti–Cu shape memory films. Thin Solid Films 274, 101–105.
Krulevitch, P., Lee, A.P., Ramsey, P B., Trevino, J., Hamilton, J., Northrup, M.A., 1996. Thin film shape memory microactuators. Journal of MEMS 5, 270–282.
Le Dret, H., Raoult, A., 1993. Le modèle de membrane non linèaire comme limite variationnelle de l'élasticité non linèaire tridimensionelle. C. R. Acad. Sci. Paris 317, Series I, 221–226.
Le Dret, H., Raoult, A., 1995. The nonlinear membrane model as variational limit of nonlinear three-dimensional elasticity. J. Math. Pures Appl. 73, 549–578.
Le Dret, H., Raoult, A., 1996. The membrane shell model in nonlinear elasticity : a variational asymptotic derivation. Journal of Nonlinear Science 6, 59–84.
Mathews, S.A., Wuttig, M., Su, Q.M., 1996. The effect of substrate constraint on the martensitic transformation in NiTi thin films. Metall. Mat. Trans A 27, 2859–2861.
Miyazaki, S., Nomura, K., Ishida, A., 1995. Shape memory effects associated with the martensitic and R-phase transformations in sputter-deposited Ti–Ni thin films. Proceedings of the International Conference on Martensitic Transformations 1995, Journal de Physique IV, Colloque C8 5, 677–682.

Modica, L., Mortola, S., 1977. Il limite nella Γ-convergenza di una famiglia di funzionali ellittichi. Boll. Un. Math. It. (3) A14, 526–529.

Morrey, C.B. Jr, 1966. Multiple Integrals in the Calculus of Variations. Springer-Verlag.

Nam, T.H., Saburi, T., Nakata, Y., Shimizu, K., 1990. Shape memory characteristics and lattice deformation in Ti–Ni–Cu alloys. Materials Trans. JIM 31, 1050–1056.

Nečas, J., 1983. Introduction to the Theory of Nonlinear Elliptic Equations. Teubner : Leipzig.

Pitteri, M., Zanzotto, G., 1996. Symmetry-breaking and transformation twinning. Archive for Rational Mechanics and Analysis, to appear.

Saburi, T., Watanabe, Y., Nenno, S., 1989. Morphological characteristics of the orthorhombic martensite in a shape memory Ti–Ni–Cu alloy. ISIJ International 29, 405–411.

Shapiro, S.M., Yang, B.X., Noda, Y., Tanner, L.E., Schryvers, D., 1991. Neutron-scattering and electron-microscopy studies of the premartensitic phenomena in Ni_xAl_{100-x} alloys. Physics Review B 44, 9301–9313.

Shu, Y.C., Bhattacharya, K., 1998. The influence of texture on the shape-memory effect in polycrystals. Acta Materialia, to appear.

Su, Q., Hua, S.Z., Wuttig, M., 1994. Martensitic transformation in $Ni_{50}Ti_{50}$ films. Trans. Mat. Res. Soc. Japan 18B, 1057.

Tanner, L.E., Schryvers, D., Shapiro, S.M., 1990. Electron microscopy and neutron scattering studies of premartensitic behavior in ordered Ni–Al β_2 phase. Materials Science and Engineering A127, 205–213.

Tickle, R., James, R.D., Wuttig, M., Kokorin, V.V., Shield, T., Schumacher, P., 1997. Ferromagnetic shape memory in the NiMnGa system, preprint.