A scheme for the passage from atomic to continuum theory for thin films, nanotubes and nanorods

Gero Friesecke[a], Richard D. James[b],*

[a]*Mathematical Institute, University of Oxford, 24-29 St Giles, Oxford, OX1 3LB, UK*
[b]*Department of Aerospace Engineering and Mechanics, University of Minnesota, Minneapolis, MN 55455, USA*

Abstract

We propose a scheme for the direct passage from atomic level to continuum level. The scheme is applicable to geometries, like films, rods and tubes, in which one or more dimensions are large relative to atomic scale but other dimensions may be of atomic scale. The atomic theory is assumed to be governed by a variational principle resting on the Born–Oppenheimer approximation. The atomic level energy is further assumed to satisfy certain decay properties when evaluated for disjoint sets of atoms. The scheme is based on two hypotheses: (1) distortions are limited, and (2) there are many atoms in certain directions. The scheme produces in a natural way the variables of the continuum theory. In the case of a film, the continuum theory that emerges is a Cosserat membrane theory with $(v - 1)$ Cosserat vectors, v being the number of atomic layers in the film. The arguments presented are not mathematically rigorous. One difficulty is that it is not clear under which circumstances our decay hypothesis on interatomic interactions is consistent with quantum mechanics or density functional theory. © 2000 Elsevier Science Ltd. All rights reserved.

* Corresponding author. Tel.: + 1-612-625-0706; fax: + 1-612-626-1558.
E-mail address: james@aem.umn.edu (R.D. James).

1. Introduction

The ability to reproducibly synthesize structures having an atomic scale dimension, and the subsequent investigation of the unusual properties and possible applications of these nanoscale structures, has given rise to the field of nanotechnology. Properties that are known to be structure-sensitive on the macroscale, such as strength, plasticity, magnetic hysteresis, superconductivity, and properties altered by phase transformation are expected to exhibit unusual behavior on scales that are smaller than the typical microstructural feature that is responsible for the sensitivity.

When all of the dimensions of a nanoscale object are small, it is possible to use purely atomic scale computational methods to understand its behavior. But when at least one dimension of the object is much larger than atomic scale, purely atomistic methods become time consuming, and they also suffer from the drawback of offering only one-configuration one-computation; that is, they sometimes do not give insight into the variety of possible behaviors of the structure.

Lacking a suitable alternative, the current practice is to transport macroscale continuum theories down to the nanoscale. For example, the motion of nanoscale cantilevers is often analyzed using macroscale beam theory (e.g., Wong et al., 1997). The moment–curvature relation in these beam theories is given as the product EI of Young's modulus and moment of inertia. For a nanotube or nanorod, it is not clear what one should use for the moment of inertia I of the cross-section. One could imagine that it represents a moment of inertia of a kind of cross-sectional envelope of the atoms. But cuts at different locations along the rod give very different sets of atoms. Furthermore, the origins of the term EI — the comparison of a solution in Euler beam theory with a similar solution of linear elasticity — indicate that the presence of this product is associated with a linear variation of stress across the cross-section, a situation that is difficult to imagine on the nanoscale. More than that, even the concept of stress (or, in electromagnetism, the concept of the electromagnetic field in a polarizable or magnetizable medium) is essentially macroscopic and may not be a reasonable concept to use at the nanoscale.

The existence of a nanoscale dimension is in the authors' opinion not a reason by itself to reject continuum theory. Roughly, continuum theories should arise when there are one or more dimensions much larger than atomic scale and there is some macroscopic homogeneity. In statics, the problem of deriving continuum theory is essentially a problem of partitioning the configurational variables into two sets: one set is treated by atomic methods and the other set furnishes the variables of the continuum theory. Along the way, a successful method of partitioning achieves simplification. The opportunity for simplification arises from the fact that for an extended film large classes of nanoscale configurations have nearly the same energy, as compared to a typical change of energy of the whole film.

We propose the following method for deriving continuum theory from atomic

theory. Consider, for example, a film having v atomic layers and lateral dimensions $k \times k$, where k is much larger than atomic scale. First, to limit the number of possible states, we assume that the distortion of the film is bounded; that is, we assume that the affine deformation experienced by every triple of atoms has a bounded gradient. This immediately implies that there exist some well-defined macroscale averages. Then the method is described as follows:

1. Calculate all averages of atomic positions that are well-defined in the limit $k \to \infty$. In mathematical terms, for each fixed k rescale the deformation to a fixed domain, and calculate all weak limits of the sequence of rescaled deformations. Use these averages as the variables of continuum theory.
2. Find a shrinking neighborhood $\mathcal{N}_k$ of these averages having the property that, if a sequence $y^{(k)}$ of deformations lies in the neighborhood for each k, then they converge in the sense found above to these averages.
3. Minimize the atomic degrees of freedom in the neighborhood for each k.
4. Calculate the energy per unit area, pass to the limit $k \to \infty$ and extract a continuum theory.

In this paper, we outline this procedure explicitly in the case of thin films. The argument at this stage is formal, but we believe that it can be made rigorous. We concentrate on the physical basis of the scheme in this paper. The scheme has several desirable features. First, it is compatible with the basic physics of interatomic forces, governed by quantum mechanics or simplified versions such as density functional theory. In particular, full many-body interatomic potentials as arising from minimization over electronic states are treated, and the positions of individual atoms are not assigned a priori, just certain averages. Second, it yields a specific procedure for calculating a continuum level energy in terms of the atomic scale relaxation under homogeneous conditions, with certain averages fixed. Essentially, it gives a quantitative interpretation of the notion of "local state" in continuum mechanics. Finally, unlike classical rules for passing from atomic to continuum levels, this method in a natural way produces the variables of the continuum theory.

To develop this procedure, there remains a lot of work to do. We only carry out the argument at a formal level for the thin film case, and indicate briefly how it works for nanotubes and rods. Second, as presented and expected, the scheme only delivers the membrane energy. The bending energy occurs at higher order in $1/k$. One could think of applying a similar procedure to the difference between the original energy and the membrane energy, multiplied by a suitable power of k (k^3 is expected classically), but we have no results in this direction. Third, the method is explicitly static. Lattice vibrations could be added by superimposing the formulas of statistical mechanics on the atomic scale relaxation as discussed by many authors, but full-scale dynamic motions are not treated by this method. Fourth, the continuum theory that emerges needs to be further simplified to be widely applicable. For example, the linearized theory needs to be explored, and any possible separation between the effects of geometry and atomic forces (such as embodied in expressions like EI) should be exposed.
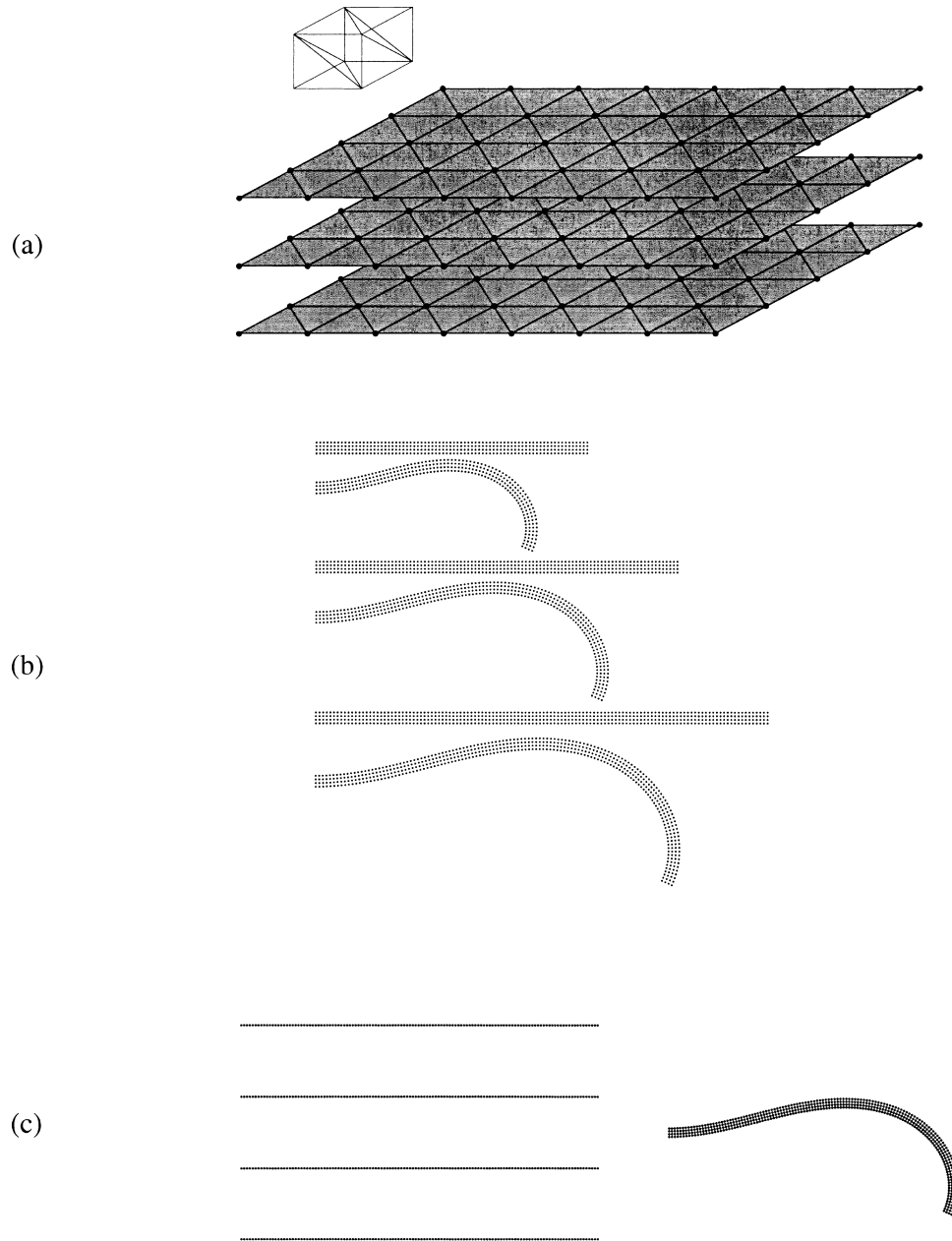
Fig. 1. Kinematics: (a) method of triangulation, (b) picture of a section through the original deformation for several values of k, (c) picture of the rescaled deformation.

2. Kinematics

Consider a single crystal film having ν monolayers. For simplicity we picture the nuclei as occupying a reference configuration on a Bravais lattice

$$\mathscr{L} = \mathscr{L}(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = \left\{ \mu_i \mathbf{e}_i : \mu_i \text{ are integers} \right\} \tag{1}$$

and a film of thickness $h = (\nu - 1)|\mathbf{e}_3|$ and planar section $\mathscr{S}_k = (0, k) \times (0, k)$, so that the nuclei in the reference configuration of the film occupy the positions,

$$\mathscr{L} \cap (\mathscr{S}_k \times [0, h]) \tag{2}$$

The theory is intended to apply to films that have been released from the substrate, or else to films attached to substrates that are also very thin. In other words, we neglect contributions to the energy arising from the substrate or from film–substrate interactions. The reference configuration could have a more complex structure than shown: partially filled layers, different types of atoms, complex crystal structures. It would be reasonable to model the reference configuration after the approximate crystal structure of the undeformed film, but the scheme only assumes that all configurations of interest are deformed from the given reference configuration by a deformation with bounded gradient.

Deformations of the film are described by functions $\mathbf{y} : \mathscr{S}_k \times [0, h] \to \mathbb{R}^3$. Strictly speaking, $\mathbf{y}$ is only defined at the nuclear positions, so in writing the domain of $\mathbf{y}$ as $\mathscr{S}_k \times [0, h]$ we have tacitly assumed a particular interpolation between the nuclei. For definitness we shall assume that each layer is triangulated in the natural way as indicated in Fig. 1, and the ν triangulations are each translates in the direction $\mathbf{e}_3$ of the lowest one. The basic assumption of regularity we make is,

$$|\nabla \mathbf{y}(\mathbf{x})| \leq c_2 \quad \text{and} \quad |\mathbf{y}(\mathbf{x})| \leq c_2 k, \quad \mathbf{x} \in \mathscr{S}_k \times [0, h], \tag{3}$$

where c_2 is a constant. In our view, it bounds on atomic positions as Eq. (3) that are directly responsible for the presence of gradients in continuum theories.

It is also necessary to make a global *minimal strain hypothesis* to prevent the film from looping back and touching itself. In fact, without such an hypothesis the film could, consistent with Eq. (3), repeatedly fold back on itself so as to be effectively a block of bulk material, with some negligible thin film loops at its ends. We do not want such configurations to compete in the energy minimization. We therefore assume,

$$c_1 |\mathbf{x}_2 - \mathbf{x}_1| \leq |\mathbf{y}(\mathbf{x}_2) - \mathbf{y}(\mathbf{x}_1)| \ (\leq c_2 |\mathbf{x}_2 - \mathbf{x}_1|), \quad \mathbf{x}_1, \mathbf{x}_2 \in \mathscr{S}_k \times [0, h], \tag{4}$$

the latter inequality being equivalent to the first inequality in Eq. (3).

As mentioned above, already the conditions (3) imply that existence of some well-defined averages in the limit as $k \to \infty$. By "averages" we mean, in this case, *local spatial averages*. That is, for a sequence of bounded functions $f^{(k)} : \Omega \to \mathbb{R}^3$ we say that the local spatial average of the sequence is f, abbreviated $\prec f^{(k)} \succ = f$, if for a smooth function $\rho : \mathbb{R}^3 \to \mathbb{R}$ with compact support and $\int_{\mathbb{R}^3} \rho = 1$ we have

that

$$\lim_{k\to\infty} \int_\Omega \rho(\mathbf{x} - \mathbf{z}) f^{(k)}(\mathbf{z})\, d\mathbf{z} \to \int_\Omega \rho(\mathbf{x} - \mathbf{z}) f(\mathbf{z})\, d\mathbf{z}. \tag{5}$$

For Eq. (5) to make sense $\mathbf{x} \in \Omega$ has to be a distance greater than the diameter of the support of ρ away from $\partial\Omega$, and f is then guaranteed to be the local spatial average of $f^{(k)}$ only on the subset of Ω where Eq. (5) applies. The definition of local spatial average can be made to apply on all of Ω by simply extending f as zero outside Ω (when this is possible) and then requiring that Eq. (5) hold for all $\mathbf{x} \in \Omega$. The condition (5) clearly means "local spatial average" if ρ is the characteristic function of a ball, for in that case Eq. (5) expresses the idea that the average of $f^{(k)}$ over the ball centered at $\mathbf{x}$ converges to the average of f over the same ball, for every choice of $\mathbf{x}$.

For a bounded region Ω, the definition (5) is equivalent to the standard definition of weak convergence (in L^∞). Extend $f^{(k)}$ to be zero outside Ω and take the Fourier transform of the left-hand side of Eq. (5). The Fourier transform of the convolution is the product of the Fourier transforms $\hat{\rho}\hat{f}^{(k)}$. Since ρ is bounded, has compact support, and is not identically zero, its Fourier transform is analytic and vanishes at most on a set of measure zero. Thus, dividing the Fourier transform of Eq. (5) by $\hat{\rho}$ where possible, we see that Eq. (5) implies that $\hat{f}^{(k)}$ converges pointwise a.e. to $\hat{f}$. Now take the inverse Fourier transform of $\hat{f}^{(k)} \to \hat{f}$ and one gets the standard definition of weak convergence: $\int_\Omega f^{(k)}\varphi\, d\mathbf{x} \to \int_\Omega f\varphi\, d\mathbf{x}$ for all trigonometric test functions φ, or, by expansion in Fourier series, for any smooth φ with compact support. This argument shows that it does not matter what function ρ one uses in Eq. (5), satisfying the given requirements. In particular, one can also use a ρ that has very small support, and one can even let the support of ρ shrink to zero as long as that is done after $k \to \infty$; for example, one can do this to avoid the problem of having to be a certain distance away from $\partial\Omega$. Below, we shall let the support of ρ shrink to zero in a particular way with k in order to select out a certain mesoscale average.

We shall use a standard fact about local spatial averages (weak convergence): if $f^{(k)}$ is uniformly bounded then it necessarily has a subsequence $\{f^{(k_i)}\}$ with a local spatial average, $\prec f^{(k_i)} \succ = f$. A brief word about the interpretation of subsequences, that may cause confusion from a physical point of view. In the above, we can take away the convergent subsequence $\{f^{(k_i)}\}$ from $\{f^{(k)}\}$. Either the whole sequence $\{f^{(k)}\}$ satisfies $\prec f^{(k)} \succ = f$, or we are again left with an infinite sequence that is bounded and we can reapply the result. So a given bounded sequence has families of subsequences, possibly an infinite number, each with a well-defined local spatial average. By selecting one subsequence, we are effectively selecting one macroscopic configuration of the film. It is, therefore, important that any subsequent arguments would apply to any subsequence.

Let $\mathbf{y}^{(k)}$ be any further subsequence of deformations that for each k satisfies Eq. (3). To reveal the averages, it is useful to rescale the deformation so that the scaled deformation is defined on a fixed domain, as in the definition of local

spatial average. A convenient rescaling is given by

$$\tilde{\mathbf{y}}^{(k)}(\mathbf{x}) := \frac{1}{k}\mathbf{y}^{(k)}(kx_1, kx_2, x_3). \tag{6}$$

Here and throughout this paper the superimposed ~ denotes rescaled functions, defined on the fixed domain $\mathscr{S}_1 \times [0, h]$ (see Table 1). By direct appeal to Eq. (3), we have for the rescaled deformation,

$$\left|\nabla_{\mathrm{p}}\tilde{\mathbf{y}}^{(k)}(\mathbf{x})\right| \le c_2, \quad \left|k\tilde{\mathbf{y}}^{(k)}_{,3}(\mathbf{x})\right| \le c_2 \quad \text{and} \quad \left|\tilde{\mathbf{y}}^{(k)}(\mathbf{x})\right| \le c_2, \quad \mathbf{x} \in \mathscr{S}_1 \times [0, h]. \tag{7}$$

Here, the notation ∇_{p} denotes the planar gradient. We consider only those averages that are linear in the positions of the nuclei, i.e., linear in $\tilde{\mathbf{y}}^{(k)}$ or $\mathbf{y}^{(k)}$. From Eq. (7), we have immediately that for a suitable subsequence (not relabeled) there are vector-valued functions $\mathbf{u}$ and $\mathbf{b}$ defined on the fixed domain such that,

$$\prec \nabla\tilde{\mathbf{y}}^{(k)} \succ = \nabla\mathbf{u} \quad \text{and} \quad \prec k\tilde{\mathbf{y}}^{(k)}_{,3} \succ = \mathbf{b} \quad \text{on } \mathscr{S}_1 \times [0, h]. \tag{8}$$

Here, we have written the local spatial average of the deformation gradient as a gradient ($\nabla\mathbf{u}$) because from Eq. (5) it follows that gradient commutes with local spatial average.

There are some compatibility conditions that arise from Eq. (7). The second part of Eq. (7) says that $|\tilde{\mathbf{y}}^{(k)}_{,3}|$ tends uniformly to zero, so (since $\mathscr{S}_1 \times [0, h]$ is convex in the x_3 direction) $\mathbf{u}$ must be independent of x_3:

$$\mathbf{u}(x_1, x_2) \tag{9}$$

There is another condition on $\mathbf{b}$ which is less obvious but becomes clear from Fig. 1. That is, because $\tilde{\mathbf{y}}^{(k)}$ was obtained by interpolation of nuclear positions, $\mathbf{b}$ is completely determined by the values of $\tilde{\mathbf{y}}^{(k)}$ on the ν layers of rescaled atomic positions. This strong restriction is a nuisance because it means that $\mathbf{b}$ cannot be regarded as an independent function in the resulting continuum theory. Rather than carry along the restrictions on $\mathbf{b}$, we adopt an equivalent but simpler format. That is, we note that the quantity,

Table 1
Scales expressed in original and rescables

	Original	Rescaled
Nanoscale	1	$1/k$
Mesoscale	εk	ε
Macroscale	k	1

$$k\left|\tilde{\mathbf{y}}^{(k)}(x_1, x_2, x_3) - \tilde{\mathbf{y}}^{(k)}(x_1, x_2, 0)\right|, \quad x_3 = \left(\frac{h}{(v-1)}\right)i,$$

$$i \in \left\{1, \ldots, (v-1)\right\} \tag{10}$$

is bounded independent of k for each $i \in \{1, \ldots, (v-1)\}$, by the second inequality in Eq. (7). Hence, for a suitable subsequence (again, not relabeled),

$$\prec k\left(\tilde{\mathbf{y}}^{(k)}\left(x_1, x_2, \frac{ih}{(v-1)}\right) - \tilde{\mathbf{y}}^{(k)}(x_1, x_2, 0)\right) \succ = \mathbf{b}_i(x_1, x_2),$$

$$i \in \left\{1, \ldots, (v-1)\right\}. \tag{11}$$

Evidently, $\mathbf{u}$ and the $(v-1)$ functions $\mathbf{b}_i$ are all averages of atomic positions that are linear in the deformation.

One other point is why only consider averages that are linear in the deformation? For example, any nonlinear function of the rescaled deformation gradient $f(\nabla\tilde{\mathbf{y}}^{(k)})$ would also be bounded and therefore (on a suitable subsequence) would have a local spatial average, $F = \prec f(\nabla\tilde{\mathbf{y}}^{(k)}) \succ$. But with nonlinear functions, the gradient does not commute with the nonlinear function. And even for very simple choices of f the possible local spatial averages F that can emerge from this procedure are difficult to characterize. So, the domain of the resulting continuum theory might be strongly restricted but in an unexpected way. In general, this problem is equivalent to the difficult problem of characterizing the set of all Young measures which can arise from a sequence of gradients, a problem which has attracted considerable activity in recent years.

3. Weak neighborhoods

Having identified $\mathbf{u}$ and the $(v-1)$ functions $\{\mathbf{b}_i\}$ as the basic functions of the continuum theory, we now face the task of designing a suitable shrinking neighborhood of these functions, which we denote by $\mathcal{N}_\varepsilon[\mathbf{u}, \mathbf{b}_1, \ldots, \mathbf{b}_{(v-1)}]$. The subscript ε refers to the a mesoscopic length scale $\varepsilon = \varepsilon_k$. All deformations inside the neighborhood should be allowed to compete in the atomic scale relaxation. Thus, the problem of energy minimization is reduced to a problem of, first, minimizing over deformations in the neighborhood using atomic methods and, second, minimizing over $(\mathbf{u}, \mathbf{b}_1, \ldots, \mathbf{b}_{(v-1)})$ using continuum theory. The mesoscale should be much smaller than macroscale and much bigger than atomic scale: $1 \ll \varepsilon k \ll k$, see Table 1.

The following seem to be reasonable physical requirements on weak neighborhoods.

1. If $\tilde{\mathbf{y}}^{(k)} \in \mathcal{N}_{\varepsilon_k}[\mathbf{u}, \mathbf{b}_1, \ldots, \mathbf{b}_{(v-1)}]$ for each k, then it should follow that

$$\nabla\mathbf{u} = \prec \nabla\tilde{\mathbf{y}}^{(k)} \succ,$$

$$\mathbf{b}_i = \prec k\left(\tilde{\mathbf{y}}^{(k)}\left(x_1, x_2, \frac{ih}{(\nu-1)}\right) - \tilde{\mathbf{y}}^{(k)}(x_1, x_2, 0)\right) \succ, \tag{12}$$

i.e., $(\mathbf{u}, \mathbf{b}_1, \ldots, \mathbf{b}_{(\nu-1)})$ are the local spatial averages based on this sequence. In particular, the weak neighborhoods should separate local spatial averages: given $(\mathbf{u}^1, \mathbf{b}_1^1, \ldots, \mathbf{b}_{(\nu-1)}^2) \neq (\mathbf{u}^2, \mathbf{b}_1^2, \ldots, \mathbf{b}_{(\nu-1)}^2)$, then for k sufficiently large $\tilde{\mathcal{N}}_{\varepsilon_k}[\mathbf{u}^1, \mathbf{b}_1^1, \ldots, \mathbf{b}_{(\nu-1)}^1] \cap \mathcal{N}_{\varepsilon_k}[\mathbf{u}^2, \mathbf{b}_1^2, \ldots, \mathbf{b}_{(\nu-1)}^2] = \varnothing$.

2. The replacement of $(\mathbf{u}; \mathbf{b}_1, \ldots, \mathbf{b}_{(\nu-1)})$ by their mesoscale (piecewise linear; piecewise constant) approximations should only slightly perturb the weak neighborhoods. That is, let $\mathcal{T}_\varepsilon$ be a triangulation of $\mathbb{R}^2$ using triangles of side length ε, and let $\mathbf{u}_\varepsilon$ be a piecewise linear approximation of $\mathbf{u}$ and $(\mathbf{b}_{\varepsilon 1}, \ldots, \mathbf{b}_{\varepsilon(\nu-1)})$ be a piecewise constant approximation of $(\mathbf{b}_1, \ldots, \mathbf{b}_{(\nu-1)})$ on $\mathcal{T}_\varepsilon$ restricted to $\mathscr{S}_1$, as shown in Fig. 2. Then, $\tilde{\mathcal{N}}_\varepsilon[\mathbf{u}_\varepsilon, \mathbf{b}_{\varepsilon 1}, \ldots, \mathbf{b}_{\varepsilon(\nu-1)}]$ should approximate $\tilde{\mathcal{N}}_\varepsilon[\mathbf{u}, \mathbf{b}_1, \ldots, \mathbf{b}_{(\nu-1)}]$ in that the minimum energy obtained in the atomic relaxation is close to the sum of the energies of each of the linearly deformed triangles. The precise requirement is that the resulting error divided by the macroscale area k^2 tend to zero as $k \to \infty$. This requirement is important for reducing the atomic level calculation to one which is carried out under homogeneous conditions.

Condition 2 is interesting even when the film is not being deformed. It says that there is sufficient screening so that the energy per atom is affected little by
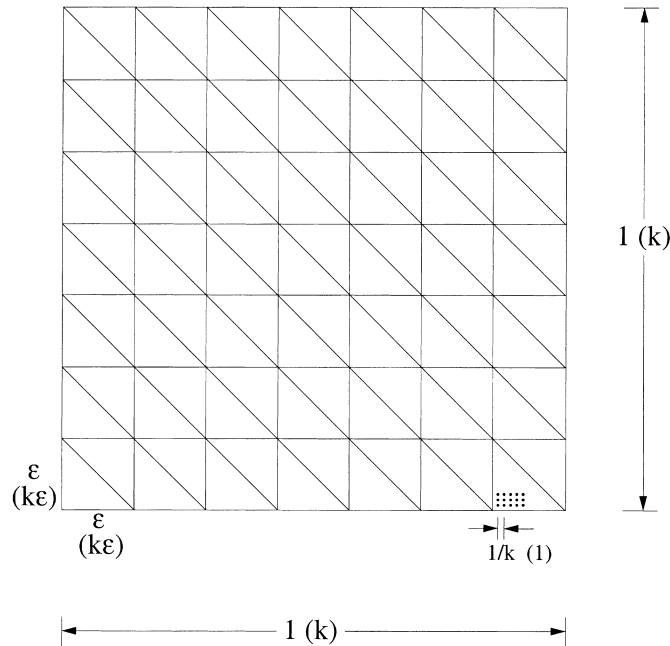

Fig. 2. Mesoscale triangulation of $\mathscr{S}_1$ in rescaled variables; original variables in parentheses.

intertriangle interactions. To understand the ensuing restriction on ε_k consider the case of strong screening, that is to say of short-range interactions. The error in neglecting intertriangle interactions would be of the order of the sum of the sidelengths of all triangles, $O(k/\varepsilon)$. This is of lower order than the macroscale area k^2 if and only if $1 \ll \varepsilon_k k$.

To summarize, local spatial averages, which are the input into continuum theory, are approximately constant in the mesoscale neighborhood $\tilde{\mathcal{N}}_\varepsilon$ provided $\varepsilon k \ll k$; atomic relaxation of the film decouples, to highest order, into independent relaxation in linearly deformed spatial regions of length scale ε when, at the very least, $1 \ll \varepsilon_k k$.

Some consequences of the bounds (3) and the definition of local spatial average (5) suggest a design of the weak neighborhood. It is a consequence of the Rellich theorem that if $\prec \nabla \tilde{\mathbf{y}}^{(k)} \succ = \nabla \mathbf{u}$, then $\tilde{\mathbf{y}}^{(k)} \to \mathbf{u}$ uniformly. This suggests the use of a uniform neighborhood of $\mathbf{u}$ to define $\tilde{\mathcal{N}}_\varepsilon$, whose size depends on ε. In rescaled variables, that is:

$$|\tilde{\mathbf{y}} - \mathbf{u}| \le \ell_1(\varepsilon) \quad \text{on } \mathscr{S}_1. \tag{13}$$

For $\mathbf{b}_i$ it is advantageous to use more directly the notion of local spatial average, but restricted so that the support of the kernal ρ is on the mesoscale, i.e., a mesoscale spatial average. Recalling that ρ is associated with rescaled variables, let $\varepsilon^2 \rho_\varepsilon(\mathbf{x}_p)$ be the characteristic function of a square of size $\varepsilon \times \varepsilon$:

$$\rho_\varepsilon(\mathbf{x}_p) = \varepsilon^{-2} \chi_{\mathscr{D}_\varepsilon(\mathbf{x}_p)},$$
$$\mathscr{D}_\varepsilon(\mathbf{x}_p) := \left\{ (x_1, x_2) : |x_1 - (\mathbf{x}_p \cdot \mathbf{e}_1)| < \varepsilon/2, |x_2 - (\mathbf{x}_p \cdot \mathbf{e}_2)| < \varepsilon/2 \right\}. \tag{14}$$

With this choice of ρ_ε sub-mesoscale oscillations of $k(\tilde{\mathbf{y}}(x_1, x_2, \frac{ih}{(v-1)}) - \tilde{\mathbf{y}}(x_1, x_2, 0))$ will be averaged during local spatial averaging but oscillations on larger scale will be relegated to the continuum theory. Using rescaled variables, we will complete the definition of the neighborhood by imposing a local spatial average based on the kernel (Eq. (14)):

$$\left| \frac{1}{\varepsilon^2} \int_{\mathscr{D}_\varepsilon(\mathbf{x}_p)} k \left\{ \tilde{\mathbf{y}} \left(x_1, x_2, \frac{ih}{(v-1)} \right) - \tilde{\mathbf{y}}(x_1, x_2, 0) \right\} \, dx_1 \, dx_2 - \right.$$
$$\left. \frac{1}{\varepsilon^2} \int_{\mathscr{D}_\varepsilon(\mathbf{x}_p)} \mathbf{b}_i(x_1, x_2) \, dx_1 \, dx_2 \right| < \ell_2(\varepsilon), \tag{15}$$
$$i \in \left\{ 1, \ldots, (v-1) \right\}.$$

In summary, we propose the neighborhood

$$\tilde{\mathcal{N}}_\varepsilon\big(\mathbf{u}, \mathbf{b}_1, \ldots, \mathbf{b}_{(v-1)}\big):$$

$$= \big\{\tilde{\mathbf{y}} : \mathscr{S}_1 \to \mathbb{R}^3 :$$

Eqs. (4), (7), (13) *and* (15) hold for some $\ell_1(\varepsilon)$ and $\ell_2(\varepsilon)\big\}.$ \hfill (16)

The goal of the following sections is to find restrictions on $\ell_1(\varepsilon)$ and $\ell_2(\varepsilon)$ that imply the conditions 1 and 2 above. Condition 1, which is purely kinematical, is guaranteed by

$$\ell_1(\varepsilon) \to 0, \quad \ell_2(\varepsilon) \to 0 \quad \text{as } \varepsilon \to 0. \hfill (17)$$

Condition 2 enforces more subtle restrictions on ε, k and the $\ell_i(\varepsilon)$ which account for the nature of the atomic level energy, so these restrictions are postponed until Section 5, where we set up the atomic relaxation. Condition 2 relies on approximation. Consider a triangulation $\mathscr{T}_\varepsilon$ of $\mathbb{R}^2$ as shown in Fig. 2, using triangles of height ε and base length ε. Let $\mathbf{u}_\varepsilon$ be a standard piecewise linear approximation of $\mathbf{u}$ and $(\mathbf{b}_{\varepsilon 1}, \ldots, \mathbf{b}_{\varepsilon(v-1)})$ be a piecewise constant approximation of $(\mathbf{b}_1, \ldots, \mathbf{b}_{(v-1)})$ on $\mathscr{T}_\varepsilon$ restricted to $\mathscr{S}_1$. For simplicity, let us assume that $(\mathbf{b}_{\varepsilon 1}, \ldots, \mathbf{b}_{\varepsilon(v-1)})$ is constant on each of the $\varepsilon \times \varepsilon$ rectangles. For a moment assume that $\mathbf{u} \in C^1(\mathscr{S}_1)$ and $(\mathbf{b}_1, \ldots, \mathbf{b}_{(v-1)}) \in C^0(\mathscr{S}_1)$. Then,

$$|\mathbf{u}_\varepsilon - \mathbf{u}| < \varepsilon g(\varepsilon),$$

$$|\mathbf{b}_{\varepsilon i} - \mathbf{b}_i| < g(\varepsilon) \quad \text{on } \mathscr{S}_1, \hfill (18)$$

where $g(\varepsilon) \to 0$ as $\varepsilon \to 0$. Without loss of generality, we have used the same measure of error in both parts of Eq. (18). The bounds (18) suggest that the replacement of $(\mathbf{u}; \mathbf{b}_1, \ldots, \mathbf{b}_{(v-1)})$ by their mesoscale (piecewise linear; piecewise constant) approximations only slightly perturb the weak neighborhoods if, for example,

$$\ell_1(\varepsilon) \geq \varepsilon g(\varepsilon)^\gamma, \quad \text{and} \quad \ell_2(\varepsilon) \geq \varepsilon^2 g(\varepsilon)^\gamma, \hfill (19)$$

where $0 < \gamma < 1$. We note that, to make the argument below work, it is necessary to frame the definition of the weak neighborhood by using information about the regularity of $(\mathbf{u}; \mathbf{b}_1, \ldots, \mathbf{b}_{(v-1)})$, e.g., it would not be sufficient to put $g = 1$ in Eq. (19).

In general, the functions $(\mathbf{u}, \mathbf{b}_1, \ldots, \mathbf{b}_{(v-1)})$ are only assured of being in Sobolev space $W^{1,\infty}(\mathscr{S}_1) \times L^\infty(\mathscr{S}_1)^{(v-1)}$, that is, to satisfy the bounds $|\mathbf{u}| + |\nabla \mathbf{u}| \leq c_2$, $|\mathbf{b}_i| \leq \frac{ihc_2}{(v-1)}$ on $\mathscr{S}_1$. Thus, the conditions (18) are not guaranteed to hold everywhere on $\mathscr{S}_1$ but only after removing a set whose area tends to zero as $\varepsilon \to 0$. This is the realm of approximation theory, and is one reason why the argument we present is a formal one. In a more careful treatment, one has to estimate precisely the energy of this small mesoscale set of atoms that has to be discarded.

4. Interpretation of the weak neighborhood

Since $\mathbf{y}$ and therefore $\tilde{\mathbf{y}}$ are entirely determined by the nuclear positions, any condition on these deformations is interpretable in terms of nuclear positions. In particular, it is interesting to interpret the weak neighborhood (17) in this way. We begin with Eq. (13), and rewrite this condition in terms of $\mathbf{y}$ by using Eq. (6):

$$\left| \mathbf{y}(x_1, x_2, x_3) - k\mathbf{u}\left(k^{-1}x_1, k^{-1}x_2\right) \right| \leq k\ell_1(\varepsilon). \tag{20}$$

Hence, the actual atomic positions are uniformly close to a scaled family of larger and larger surfaces of the form $k\mathbf{u}(k^{-1}x_1, k^{-1}x_2)$, related by "elasticity scaling". Members of this family of surfaces are geometrically similar to each other (see Fig. 3). If we take the gradient of $k\mathbf{u}(k^{-1}x_1, k^{-1}x_2)$, we see that the local gradient of the surface (which, of course, encodes the normal and the local stretching) is the same at corresponding points of the family. However, the second gradient tends to zero as $k \to \infty$, consistent with geometric similarity. Since $\varepsilon = \varepsilon_k \to 0$, then on the mesoscale $k\varepsilon$, the surfaces are more and more nearly linearly deformed as $k \to \infty$. Being in the weak neighborhood, it is necessary for each nuclear position to be close to its corresponding point on this surface. The maximum distance of
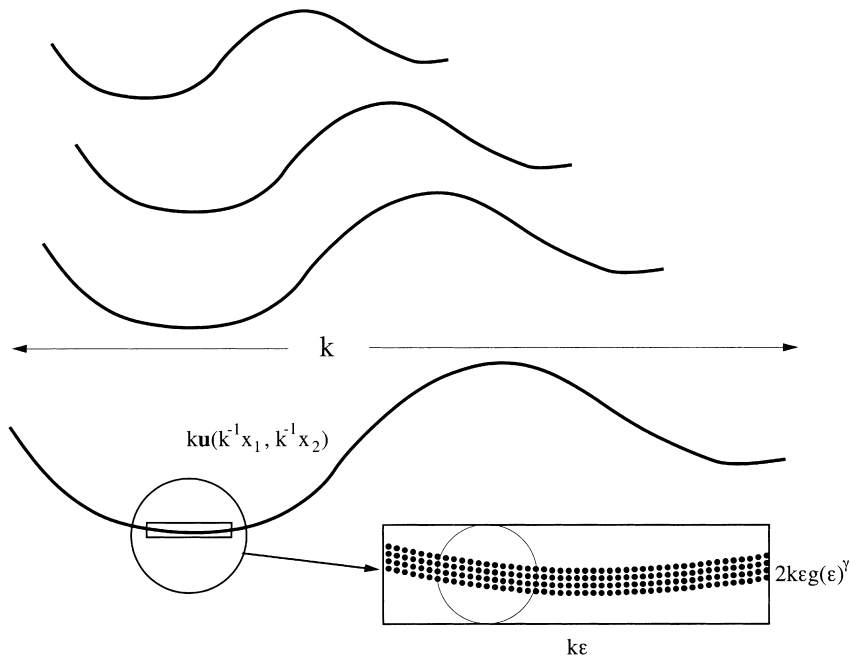

Fig. 3. Cross-section through the film in original variables showing the weak neighborhood for $\mathbf{y}$. The film is approximately linearly deformed on the mesoscale and each nucleus is in a ball of radius $k\varepsilon \circ (1)$.

the nucleus from this point is $k\ell_1(\varepsilon)$. If we take equality in Eq. (19), this distance is $k\varepsilon g(\varepsilon)^\gamma$. It then follows that the deviation of the nuclear position from its corresponding point on the surface is much less than $k\varepsilon$. Geometrically, we can picture a film-like neighborhood, much wider than its thickness, enclosing a nearly linear surface, in which each nucleus is allowed to roam within a ball having the diameter equal to the thickness of the neighborhood (Fig. 3).

For **b** we can also rescale to original variables to interpret Eq. (15). Changing variables $(x_1, x_2) \rightarrow (k^{-1}x_1, k^{-1}x_2)$ in Eq. (15), we get

$$\left| \frac{1}{(k\varepsilon)^2} \int_{\mathscr{D}_{k\varepsilon}(\mathbf{x}_p)} \left\{ \mathbf{y}\left(x_1, x_2, \frac{ih}{(\nu-1)}\right) - \mathbf{y}(x_1, x_2, 0) \right\} \, dx_1 \, dx_2 - \right. \tag{21}$$

$$\left. \frac{1}{\varepsilon^2} \int_{\mathscr{D}_\varepsilon(\mathbf{x}_p)} \mathbf{b}_i(x_1, x_2) \, dx_1 \, dx_2 \right| < \ell_2(\varepsilon).$$

Since $\mathscr{D}_{k\varepsilon}$ has area $(k\varepsilon)^2$, the left-hand side of Eq. (21) is an area average. The first integral on the left-hand side of Eq. (21) is the average over the mesoscale neighborhood of the average layer position of the ith layer relative to the first layer. As in the case above, and consistent with Eq. (19), the right-hand side of Eq. (21) can be chosen ≤ 1. This interpretation translates immediately into a restriction on nuclear positions. It says that the average deformed position of all nuclei in the ith layer on a mesoscale set, minus a similar quantity for the first layer, weighted by the area (in the reference configuration) per nucleus, is close to the mesoscale area average of $\mathbf{b}_i$.

5. Passage to continuum theory

The atomic level theory shall rest on the assumption that the energy is expressible as a function of the nuclear positions. This is the fundamental setting delivered by quantum mechanics in the Born–Oppenheimer approximation after minimization over electronic states. (More precisely, the energy depends on the nuclear positions and their charges, i.e. the atomic species involved, but for the sake of simplicity we focus on a single species here.) In the present framework, in original and rescaled variables, let the energy be denoted by,

$$E\left[\mathbf{y}(\mathbf{x}); \mathbf{x} \in \mathscr{L} \cap (k\mathscr{S}_1 \times (0, h))\right] = \tilde{E}\left[\tilde{\mathbf{y}}(\mathbf{x}); \mathbf{x} \in \frac{1}{k}\mathscr{L} \cap (\mathscr{S}_1 \times (0, h))\right]. \tag{22}$$

We shall also use a shorthand notation $E[\mathscr{M}]$ to denote this energy, where $\mathscr{M}$ is a set of position vectors. We make the following decay hypothesis: if $\mathscr{M}$ and $\mathscr{N}$ are disjoint sets of atomic positions, then

$$|E[\mathcal{M} \cup \mathcal{N}] - (E[\mathcal{M}] + E[\mathcal{N}])| \le \sum_{\mathbf{x} \in \mathcal{M}, \, \mathbf{z} \in \mathcal{N}} \psi(|\mathbf{x} - \mathbf{z}|) \tag{23}$$

where $|\psi| < M$ on $[c_1, \delta]$, and $|\psi(r)| < cr^{-q}$ on (δ, ∞),

for constants $c > 0$, $M > 0$, $\delta > c_1$ and $q > 4$. This formula can be iterated: if $\{\mathcal{M}_1, \ldots, \mathcal{M}_n\}$ denotes a set of disjoint sets of atomic positions, then from Eq. (23),

$$|E[\mathcal{M}_1 \cup \cdots \cup \mathcal{M}_n] - (E[\mathcal{M}_1]$$

$$+ \cdots + E[\mathcal{M}_n])| \le \sum_{\substack{\mathbf{x} \in \mathcal{M}_1 \\ \mathbf{z} \in \mathcal{M}_2 \cup \cdots \cup \mathcal{M}_n}} \psi(|\mathbf{x} - \mathbf{z}|) + \sum_{\substack{\mathbf{x} \in \mathcal{M}_2 \\ \mathbf{z} \in \mathcal{M}_3 \cup \cdots \cup \mathcal{M}_n}} \psi(|\mathbf{x} - \mathbf{z}|) \tag{24}$$

$$+ \cdots + \sum_{\substack{\mathbf{x} \in \mathcal{M}_{(n-1)} \\ \mathbf{z} \in \mathcal{M}_n}} \psi(|\mathbf{x} - \mathbf{z}|).$$

The assumption (23) is reminiscent of classical phenomenological expressions for atomic forces. As an upper bound it is plausible that in certain situations an assumption of this form will be valid for potential surfaces obtained from quantum mechanics or density functional theory. Clarification of the circumstances under which such screening properties hold would require a deeper understanding of quantum many-body mechanics than currently available (see Fefferman, 1985; Friesecke, 2000; Lieb and Lebowitz, 1972 for further discussion).

We shall explore the possibility here of deriving continuum theory under the hypotheses (19) but with equality:

$$\ell_1(\varepsilon) = \varepsilon g(\varepsilon)^\gamma, \quad \text{and} \quad \ell_2(\varepsilon) = \varepsilon^2 g(\varepsilon)^\gamma, \quad 0 < \gamma < 1. \tag{25}$$

The main point about the bounds (19) is that they ensure that if $(\mathbf{u}, \mathbf{b}_1, \ldots, \mathbf{b}_{(v-1)})$ is replaced by $(\mathbf{u}_\varepsilon, \mathbf{b}_{\varepsilon 1}, \ldots, \mathbf{b}_{\varepsilon(v-1)})$ then the weak neighborhood is negligibly perturbed. In particular, with the original weak neighborhood of $(\mathbf{u}, \mathbf{b}_1, \ldots, \mathbf{b}_{(v-1)})$ defined by $\ell_1(\varepsilon) = \varepsilon g(\varepsilon)^\gamma$, $\ell_2(\varepsilon) = \varepsilon^2 g(\varepsilon)^\gamma$, then, by the triangle inequality, the slightly larger neighborhood defined using $\bar{\ell}_1(\varepsilon) = \varepsilon g(\varepsilon)^\gamma (1 + g(\varepsilon)^{1-\gamma})$, $\bar{\ell}_2(\varepsilon) = \varepsilon^2 g(\varepsilon)^\gamma (1 + g(\varepsilon)^{1-\gamma})$ contains $\tilde{\mathcal{N}}_\varepsilon[\mathbf{u}_\varepsilon, \mathbf{b}_{\varepsilon 1}, \ldots, \mathbf{b}_{\varepsilon(v-1)}]$. We shall assume that the one atom derivative of the energy is uniformly bounded on the slightly enlarged neighborhood in which $\ell_1(\varepsilon)$ and $\ell_2(\varepsilon)$ are replaced by $\bar{\ell}_1(\varepsilon)$ and $\bar{\ell}_2(\varepsilon)$ in the definition of $\tilde{\mathcal{N}}_\varepsilon(\mathbf{u}, \mathbf{b}_1, \ldots, \mathbf{b}_{(v-1)})$.

It then follows that with a suitable (delicate!) relation between ε and k the atomic level minimum energy is negligibly perturbed by this enlargement of the weak neighborhood. First, by enlarging the neighborhood, we decrease the minimum energy, so we only need to worry about a lower bound. With k fixed the atomic level energy is a finite-dimensional function, whose derivative with respect to a single position vector is bounded by, say, c. Then, the maximum

decrease of the energy caused by enlarging the neighborhood is,

$$c \frac{hk^2}{|(\mathbf{e}_1 \times \mathbf{e}_2) \cdot \mathbf{e}_3|} \big(k\varepsilon g(\varepsilon)\big). \tag{26}$$

We shall be interested in the energy per unit area, so we divide Eq. (26) by k^2. Thus, we want $k\varepsilon g(\varepsilon) \to 0$. This is just possible. A reasonable physical choice seems to be:

$$k\varepsilon g(\varepsilon)^\gamma = \text{const.}, \tag{27}$$

so that,

$$k\varepsilon = \frac{\text{const.}}{g(\varepsilon)^\gamma} \to \infty \quad \text{and} \quad k\varepsilon g(\varepsilon) = \text{const.} g(\varepsilon)^{1-\gamma} \to 0. \tag{28}$$

In original variables this choice makes the thickness of the neighborhood of the surface $k\mathbf{u}(k^{-1}x_1, k^{-1}x_2)$, shown in Fig. 3, fixed relative to atomic dimensions. These considerations imply that

$$\inf_{\tilde{\mathbf{y}} \in \tilde{\mathcal{N}}_\varepsilon [\mathbf{u}, \, \mathbf{b}_1, \dots, \mathbf{b}_{(\nu-1)}]} \frac{1}{k^2} \tilde{E}\left[\tilde{\mathbf{y}}(\mathbf{x}); \mathbf{x} \in \frac{1}{k}\mathscr{L} \cap (\mathscr{S}_1 \times (0, h)) \right]$$

$$= \inf_{\tilde{\mathbf{y}} \in \tilde{\mathcal{N}}_\varepsilon [\mathbf{u}_\varepsilon, \, \mathbf{b}_{\varepsilon 1}, \dots, \mathbf{b}_{\varepsilon(\nu-1)}]} \frac{1}{k^2} \tilde{E}\left[\tilde{\mathbf{y}}(\mathbf{x}); \mathbf{x} \in \frac{1}{k}\mathscr{L} \cap (\mathscr{S}_1 \times (0, h)) \right] + \mathcal{O}\big(k\varepsilon g(\varepsilon)\big). \tag{29}$$

The next step is to break down the expression for the energy into a sum on the mesoscale triangulation $\mathscr{T}_\varepsilon$ shown in Fig. 2. Let the triangles in $\mathscr{T}_\varepsilon$ be enumerated $T_1, \dots, T_{2/\varepsilon^2}$, and apply the formula (24) to the sets,

$$\mathcal{M}_1 = \left\{ k\tilde{\mathbf{y}}(\mathbf{x}) : \mathbf{x} \in \frac{1}{k}\mathscr{L} \cap (T_1 \times (0, h)) \right\}, \dots,$$

$$\mathcal{M}_{2/\varepsilon^2} = \left\{ k\tilde{\mathbf{y}}(\mathbf{x}) : \mathbf{x} \in \frac{1}{k}\mathscr{L} \cap \big(T_{2/\varepsilon^2} \times (0, h)\big) \right\}. \tag{30}$$

From Eq. (24) the error in replacing the energy by its sum over the triangles is,

$$\sum_{\substack{\mathbf{v} \in \mathcal{M}_1 \\ \mathbf{w} \in \mathcal{M}_2 \cup \cdots \cup \mathcal{M}_n}} \psi(|\mathbf{v} - \mathbf{w}|) + \sum_{\substack{\mathbf{v} \in \mathcal{M}_2 \\ \mathbf{w} \in \mathcal{M}_3 \cup \cdots \cup \mathcal{M}_n}} \psi(|\mathbf{v} - \mathbf{w}|) + \cdots + \sum_{\substack{\mathbf{v} \in \mathcal{M}_{n-1} \\ \mathbf{w} \in \mathcal{M}_n}} \psi(|\mathbf{v} - \mathbf{w}|)$$

$$= \frac{1}{2} \sum_{i \neq j} \sum_{\substack{\mathbf{v} \in \mathcal{M}_i \\ \mathbf{w} \in \mathcal{M}_j}} \psi(|\mathbf{v} - \mathbf{w}|) \tag{31}$$

Because of the hypotheses (23), the terms in Eq. (31) can be broken down into local and long-range parts, according to whether the argument of ψ is in $[c_1, \delta]$ or

(δ, ∞). If $k\varepsilon > \delta/c_1$, the local terms contain contributions from only nearest neighbor triangles, because by (4)

$$c_1|\mathbf{x} - \mathbf{z}| < |\mathbf{y}(\mathbf{x}) - \mathbf{y}(\mathbf{z})| < \delta \Longrightarrow |\mathbf{x} - \mathbf{z}| < \delta/c_1. \tag{32}$$

The local contribution can be bounded by a straightforward but tedious argument. The long-range contribution can also be bounded by using the hypotheses (23). The sum of the local and long-range bounds is found to be:

$$\text{const.} h^2 k^2 \Bigg\{ \frac{1}{|(\mathbf{e}_1 \times \mathbf{e}_2) \cdot \mathbf{e}_3|^3} \Bigg[\left(\frac{\delta}{c_1}\right)^5 \frac{1}{k\varepsilon} + \left(\frac{\delta}{c_1}\right)^6 \frac{1}{(k\varepsilon)^2} \Bigg]$$

$$+ \left(\frac{1}{(q-2)(q-3)} \right) \left(\frac{\delta}{c_2}\right)^{(3-q)} \left(\frac{1}{k\varepsilon}\right) \Bigg\}. \tag{33}$$

After dividing by the area of the film k^2, it is seen from Eq. (33) that this bound vanishes in the limit $k\varepsilon \to \infty$, as expected. Combining Eqs. (33) and (24), we have for the energy per unit area,

$$\frac{1}{k^2} E\big[\mathbf{y}(\mathbf{x}); \mathbf{x} \in \mathscr{L} \cap (k\mathscr{S}_1 \times (0, h))\big] = \frac{1}{k^2} \sum_{i=1}^{2/\varepsilon^2} E\big[\mathbf{y}(\mathbf{x}); \mathbf{x}$$

$$\in \mathscr{L} \cap (kT_i \times (0, h))\big] + \mathcal{O}\left(\frac{1}{k\varepsilon}\right). \tag{34}$$

Taking the inf of Eq. (34), using Eq. (29) and switching to rescaled variables, we have,

$$\inf_{\tilde{\mathbf{y}} \in \tilde{\mathscr{N}}_\varepsilon[\mathbf{u}, \, \mathbf{b}_1, \ldots, \mathbf{b}_{(\nu-1)}]} \frac{1}{k^2} \tilde{E}\Bigg[\tilde{\mathbf{y}}(\mathbf{x}); \mathbf{x} \in \frac{1}{k} \mathscr{L} \cap (\mathscr{S}_1 \times (0, h)) \Bigg]$$

$$= \inf_{\tilde{\mathbf{y}} \in \tilde{\mathscr{N}}_\varepsilon[\mathbf{u}_\varepsilon, \, \mathbf{b}_{\varepsilon 1}, \ldots, \mathbf{b}_{\varepsilon(\nu-1)}]} \frac{1}{k^2} \sum_{i=1}^{2/\varepsilon^2} \tilde{E}\Bigg[\tilde{\mathbf{y}}(\mathbf{x}); \mathbf{x} \in \frac{1}{k} \mathscr{L} \cap (T_i \times (0, h)) \Bigg] \tag{35}$$

$$+ \mathcal{O}\left(\frac{1}{k\varepsilon}\right) + \mathcal{O}\big(k\varepsilon g(\varepsilon)\big).$$

Finally, we would like to bring the inf through the summation on the right-hand side of Eq. (35). This is an atomic version of issues of quasiconvexity in macroscopic elasticity. The problem with moving the inf through the summation is compatibility: the infima of the separate terms in the summation (35) may be achieved at functions $\tilde{\mathbf{y}}$ which, when joined together at the boundaries of the triangles, do not satisfy the basic bounds (4) and therefore do not belong to the weak neighborhood. But it is extremely plausible that a modification of the minimizing deformations in a layer near the boundary of each triangle will restore

these bounds, while only affecting the energy a little. The plausibility arises from two facts. First, the neighborhood in, say, original variables is of dimensions $\varepsilon k \times$ const. (assuming Eq. (27)) and it surrounds the continuous mapping $k\mathbf{u}_\varepsilon(k^{-1}x_1, k^{-1}x_2)$. Second, arguments such as those summarized by Eq. (33) strongly suggest that modification of a layer of width $k\varepsilon \circ (1)$ (1) will affect only slightly the total energy per unit area. The main difficulty in carrying out this argument is the complications involved in simultaneously respecting all of the conditions (4), (7), (13) and (15). While standard methods of linear interpolation do not work, it appears likely from exploratory calculations that use of the exact piecewise (linear; constant) fields in a small neighborhood of the boundary, together with an interpolation layer that connects the piecewise (linear; constant) fields to the interior (minimizing) field will work. For the purposes of our formal argument we assume that this is possible, and therefore we rewrite Eq. (35) in the suggestive form,

$$
\inf_{\tilde{\mathbf{y}}\in\tilde{\mathscr{N}}_\varepsilon[\mathbf{u},\,\mathbf{b}_1,\ldots,\mathbf{b}_{(v-1)}]}\frac{1}{k^2}\tilde{E}\left[\tilde{\mathbf{y}}(\mathbf{x});\,\mathbf{x}\in\frac{1}{k}\mathscr{L}\cap(\mathscr{S}_1\times(0,h))\right]
$$

$$
=\sum_{i=1}^{2/\varepsilon^2}\left\{\frac{2}{(\varepsilon k)^2}\inf_{\tilde{\mathbf{y}}\in\tilde{\mathscr{N}}_\varepsilon[\mathbf{u}_\varepsilon,\,\mathbf{b}_{\varepsilon 1},\ldots,\mathbf{b}_{\varepsilon(v-1)}]}\tilde{E}\left[\tilde{\mathbf{y}}(\mathbf{x});\,\mathbf{x}\right.\right.
$$

$$
\left.\left.\in\frac{1}{k}\mathscr{L}\cap(T_i\times(0,h))\right]\right\}\frac{\varepsilon^2}{2}+\mathscr{O}\left(\frac{1}{k\varepsilon}\right)+\mathscr{O}\big(k\varepsilon g(\varepsilon)\big). \tag{36}
$$

For each fixed $i\in\{1,\ldots,(v-1)\}$ the term in braces is an atomic level relaxation under homogeneous conditions, since $\mathbf{u}_\varepsilon=\mathbf{A}_\varepsilon\mathbf{x}+\mathbf{a}_\varepsilon$ and $\mathbf{b}_{\varepsilon i}=$ const. on T_i. In fact, we assume that the atomic level energy is translation and rotation invariant, so that this expression is independent of $\mathbf{a}_\varepsilon$. The term $2/(\varepsilon k)^2$ is 1/(the area of a triangle) in original variables. We assume that the following limit exists and is continuous for each rank-two 3×2 matrix $\mathbf{A}$ and distinct vectors $(\mathbf{b}_1,\ldots,\mathbf{b}_{(v-1)})$:

$$
\varphi\big(\mathbf{A},\mathbf{b}_1,\ldots,\mathbf{b}_{(v-1)}\big)=\lim_{\substack{\varepsilon\to 0,\,k\to\infty\\k\varepsilon g(\varepsilon)^\gamma=\text{const.}}}\left\{\frac{2}{(\varepsilon k)^2}\inf_{\tilde{\mathbf{y}}\in\tilde{\mathscr{N}}_\varepsilon[\mathbf{A}(\mathbf{x}),\,\mathbf{b}_1,\ldots,\mathbf{b}_{(v-1)}]}E\big[\mathbf{y}(\mathbf{x});\,\mathbf{x}\right.
$$

$$
\left.\in\mathscr{L}\cap(kT_\varepsilon\times(0,h))\big]\right\}. \tag{37}
$$

In Eq. (37), T_ε represents, for definiteness, the lower left triangle shown in Fig. 2 (In fact, by the translational and rotational invariance of E, any of the triangles in Fig. 2 will do.) The definition (37) is naturally interpreted as the atomic energy per unit reference area. For suitable $(\mathbf{u},\mathbf{b}_1,\ldots,\mathbf{b}_{(v-1)})$, the sum in Eq. (36) is a Riemann sum for $\varphi(\nabla\mathbf{u}(\mathbf{x}),\mathbf{b}_1(\mathbf{x}),\ldots,\mathbf{b}_{(v-1)}(\mathbf{x}))$, $\mathbf{x}\in\mathscr{S}_1$. Therefore, passing to the limit $k\to\infty$ in Eq. (36) under the restriction $k\varepsilon g(\varepsilon)^\gamma=$ const., we get the

continuum level energy,

$$\int_{S_1} \varphi\big(\nabla\mathbf{u}(\mathbf{x}), \mathbf{b}_1(\mathbf{x}), \ldots, \mathbf{b}_{(\nu-1)}(\mathbf{x})\big) \, d\mathbf{x}. \tag{38}$$

6. Comments on the implied atomic scale calculation

Now we discuss briefly the atomic scale calculation implied by the right-hand side of Eq. (37). We work in original variables, this being the natural setting for atomic scale. Technically this calculation takes place near a deformed triangle but the shape of the set likely does not matter with suitable decay properties of E. First we note that if E is rotation and translation invariant, i.e.,

$$E\big[\mathbf{R}\mathbf{y}(\mathbf{x}) + \mathbf{c}; \mathbf{x} \in \mathcal{L} \cap (\mathcal{M} \times (0, h))\big] = E\big[\mathbf{y}(\mathbf{x}); \mathbf{x} \in \mathcal{L} \cap (\mathcal{M} \times (0, h))\big]$$

$$\text{for all } \mathbf{R} \in SO(3), \mathbf{c} \in \mathbb{R}^3, \tag{39}$$

then φ is similarly invariant:

$$\varphi\big(\mathbf{R}\mathbf{A}, \mathbf{R}\mathbf{b}_1, \ldots, \mathbf{R}\mathbf{b}_{(\nu-1)}\big) = \varphi\big(\mathbf{A}, \mathbf{b}_1, \ldots, \mathbf{b}_{(\nu-1)}\big), \quad \mathbf{R} \in SO(3). \tag{40}$$

This implies a certain reduced form of φ, but it seems better to ignore this and simply choose $\mathbf{R}$ to rotate $\mathbf{A}$ into a 2×2 matrix, by, say, making its last row $\{0, 0\}$. Then $\mathbf{A}$ represents any plane distortion. Applying again the invariance (40), it is only necessary to consider 2×2 matrices $\mathbf{A}$ that are symmetric:

$$\mathbf{A} = \begin{pmatrix} \alpha & \beta \\ \beta & \delta \\ 0 & 0 \end{pmatrix}, \tag{41}$$

but, of course, then the $\mathbf{b}_1, \ldots, \mathbf{b}_{(\nu-1)}$ must be considered to be general distinct vectors.

The scheme says to choose a subset of the plane $\mathcal{D}$ that is much larger than atomic dimensions (but, for practical purposes, not too large), and distort the reference lattice $\mathcal{L} \cap (\mathcal{D} \times [0, h])$ using $\mathbf{A}$. For each nucleus $\mathbf{x} \in \mathcal{L} \cap (\mathcal{D} \times [0, h])$, the scheme says to impose two conditions. The first is that the deformed position $\mathbf{y}(\mathbf{x})$ of $\mathbf{x}$ lies in a ball of radius r centered at $\mathbf{A}\mathbf{x}$. According to Eq. (27), r should be chosen comparable to atomic dimensions. Second, the average relative deformed position of the ith layer should be given by the vector $\mathbf{b}_i$. From Eq. (21), this condition is reasonably interpreted as:

$$\frac{|\mathbf{e}_1 \times \mathbf{e}_2|}{\text{area } \mathcal{D}} \sum_{\mathbf{x} \in \mathcal{L} \cap (\mathcal{D} \times [0, h])} \left\{ \mathbf{y}\left(x_1, x_2, \frac{ih}{(\nu-1)}\right) - \mathbf{y}(x_1, x_2, 0) \right\} = \mathbf{b}_i. \tag{42}$$

Note that for $\mathscr{D}$ large the term $((|\mathbf{e}_1 \times \mathbf{e}_2|)/\text{area } \mathscr{D})$ is approximately 1/(the number of atoms in $\mathscr{L} \cap \mathscr{D}$), so the left-hand side is essentially a position average.

7. Nanotubes

We make some brief remarks about the application of the scheme to other geometries. It can be seen from the above that the exact nature of the independent variables of the continuum theory will depend on the atomic structure of the object from which the condition of limited distortion is imposed. For definiteness we focus on a particular geometry. The most well-studied nanotube is certainly the carbon nanotube (Iijima, 1991), whose structure has been carefully elucidated by Amelinckx et al. (Amelinckx et al., 1994; Bernaerts et al., 1995). It consists of layers forming concentric tubes. Each layer is a single layer of carbon atoms lying on an approximately hexagonal grid, as if the usual hexagonal graphite structure was wrapped into the shape of a cylindrical shell. The tubule consists of anywhere from one to several tens of layers. Another similar example is Bacteriophage T4, whose cylindrical tail is one atomic layer thick with a pseudo-triclinic structure (Olson and Hartman, 1982); under certain conditions the tail undergoes a spontaneous martensitic phase transformation resulting in a sudden lengthening of the tube, a feature it employs to invade its host.

Our purpose here is to write down the variables that arise from the assumption of limited distortion and give the expected form of the continuum theory. For simplicity we shall treat the case of one layer. In the reference configuration, the nuclei are assumed to lie on a "long" set $[0, k] \times \{x_1^2 + x_2^2 = r\}$ parameterized in the usual way by $(x_3, \theta) \in [0, k] \times [0, 2\pi)$, and deformations are given by functions $\mathbf{y}^{(k)}(x_3, \theta)$. Distortions are assumed bounded,

$$\left|\mathbf{y}^{(k)}_{,3}\right| + \left|\mathbf{y}^{(k)}_{,\theta}\right| < c_2 \quad \text{on } [0, k] \times [0, 2\pi], \tag{43}$$

and the minimal strain hypothesis has a natural analog that we do not write down. We rescale $\mathbf{y}^{(k)}$ to a cylinder of height 1 by defining, by analogy with Eq. (6),

$$\tilde{\mathbf{y}}^{(k)}(x_3, \theta) := \frac{1}{k}\mathbf{y}^{(k)}(kx_3, \theta) \quad \text{for } (x_3, \theta) \in [0, 1] \times [0, 2\pi]. \tag{44}$$

From Eq. (43) we have that $|\tilde{\mathbf{y}}^{(k)}_{,3}| < c_2$ and $k|\tilde{\mathbf{y}}^{(k)}_{,\theta}| < c_2$ so, for an appropriate subsequence, we have the following local spatial averages:

$$\mathbf{u} = \prec \tilde{\mathbf{y}}^{(k)}_{,3} \succ, \quad \mathbf{b} = \prec k\tilde{\mathbf{y}}^{(k)}_{,\theta} \succ, \tag{45}$$

and it follows immediately that $\mathbf{u}$ is independent of θ, $\mathbf{u}(x_3)$, and represents the deformed shape of the tube from a macroscale perspective. If the tube has several layers, then one must have a deformation defined on a thick cylinder $(x_3, r,$

$\theta) \in [0, k] \times [r_1, r_2] \times [0, 2\pi)$. In that case, the natural kinematics features a vector function $\mathbf{u}(x_3)$ and a 3×2 matrix $\mathbf{B}(x_3, r, \theta)$, which satisfies various restrictions.

The construction of the weak neighborhood and passage to continuum theory follows the pattern given above. It would be interesting to do this in detail and to see if the many observed configurations of nanotubes discussed in Amelinckx et al. (1994), Bernaerts et al. (1995) and Iijima (1991) emerge in a natural way.

8. Discussion

We have given a scheme for the direct passage from atomic to continuum theory applicable to cases in which one or more dimensions of the body are large. Of course, the scheme could be applied to the case in which all dimensions are large, but in this paper we discussed only the lower dimensional cases which have more interesting kinematics. The scheme is very much influenced by geometry, and the small parameter that leads to simplification is taken to be the ratio of length scales.

One could reasonably argue that in passing from atomic to continuum theory one should use not only simplifications guided by geometry, but also by physics. That is, one can look at a family of atomic level computations and try to understand what are the least stable local modes of deformation. In the language of structural mechanics, this involves discovering (say, by linearized stability analysis) the least stable buckling modes of the structure, and then parameterizing them by a small number of parameters, which become the variables of continuum theory. This is related to the methods of Rabe and Waghmare (1995). The present approach has the advantage of being material independent, but the disadvantage of not capitalizing on the possible presence of a simple parameterization of the main atomic modes of deformation. However, the atomic level computation suggested in Section 6 could be explored in this light.

Our derivation is related to other approaches in the literature that rely more directly on computational methods. The approach of Brandt (1992, 1997) is modeled after the successful multigrid method, and involves cycling between the fine scale and a series of coarser scale computations; the fine scale computation takes place on a small domain with periodic boundary conditions, and the coarse scale computations have increasingly larger periods performed with a series of coarse level Hamiltonians. The quasicontinuum method of Tadmor et al. (1996) is a systematic method of introducing a set of kinematic constraints that automatically allow a rather unconstrained atomic relaxation near defects and a more highly constrained (but simpler) computation far from defects; the kinematic constraints are intended to impose large-scale piecewise linearity far from defects as in continuum elasticity calculations. Philosophically, both methods relate to ours, although our method does not treat the defects explicitly. Our scheme features a definite method of identifying the variables of continuum theory as being the well-defined averages, and it provides at a definite form of the

continuum theory, with an exactly invariant energy function which then can be studied by continuum methods.

It is useful to compare our results directly with continuum membrane theories in the thin film case. In recent years, it has become possible to derive theories for thin bodies directly by methods of Γ-convergence (see e.g. Bhattacharya and James, 1999). These methods are rigorous derivations from 3D nonlinear elasticity: there is no ad hoc ansatz or arbitrary choice of asymptotic expansion (see Bhattacharya and James, 1999, and the references therein). Begin with an continuum elastic and interfacial energy per unit reference volume of the form

$$\frac{1}{h(\text{area } \mathscr{S}_1)} \int_{\mathscr{S}_1 \times (0,\, h)} \left\{ \kappa |\nabla^2 \mathbf{y}|^2 + \varphi(\nabla \mathbf{y}) \right\} d\mathbf{x}. \tag{46}$$

Under suitable hypotheses on φ the Γ-limit of Eq. (38) as $h \to 0$ is

$$\int_{\mathscr{S}_1} \left\{ \kappa \left(|\nabla^2 \mathbf{u}|^2 + 2|\nabla \mathbf{b}|^2 \right) + \varphi(\mathbf{u}_{,1}|\mathbf{u}_{,2}|\mathbf{b}) \right\} d\mathbf{x}, \tag{47}$$

Here, $\mathbf{b}$ is a vector field on $\mathscr{S}_1$ that emerges from the calculation and $\mathbf{u}: \mathscr{S}_1 \to \mathbb{R}^3$ is closely related to its counterpart in this paper. The notation $(\mathbf{a}_1|\mathbf{a}_2|\mathbf{a}_3)$ stands for the matrix with columns $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$. In the Γ-convergence argument, $\mathbf{b}$ is the pointwise limit of $(1/h)\tilde{\mathbf{y}}_{,3}$ where $\tilde{\mathbf{y}}$ is a rescaled deformation defined by $\tilde{\mathbf{y}}(x_1, x_2, x_3) = \mathbf{y}(x_1, x_2, hx_3)$, and therefore, it is closely related to the vector $\mathbf{b}_{(\nu-1)}$. Comparing Eq. (47) with (38), there are two striking differences: (1) the energy (38) has no interfacial contribution analogous to the term in Eq. (47), and (2) the energy (47) has one vector $\mathbf{b}$, while Eq. (38) has $(\nu - 1)$. The former is evidently a feature of the scaling: interfacial terms come out at higher order in $(1/k)$ in the present format. This is not unexpected in that κ is usually regarded as much smaller than a typical modulus that describes the growth of φ away from its energy well(s). However, the form of the interfacial energy terms in Eq. (47) may well not be consistent with the atomic level approach, and this is a good topic for further research. The presence of the many vectors $\mathbf{b}_i$ seems to be really an atomic level effect that has no counterpart in usual bulk elasticity.

Without having solved an actual problem with the theory (38) derived in this paper, we might speculate on its main implications. It says that substantial energy can be carried by the overall shape of the film, encoded by $\mathbf{u}$, and by relative shifts among the atomic layers, related to changes of energy with changes of $\mathbf{b}_1, \ldots, \mathbf{b}_{(\nu-1)}$. The latter suggests instabilities involving shifts of atomic layers, perhaps induced by deformation.

If the film is quite thick, then carrying along all of the vectors $\mathbf{b}_1, \ldots, \mathbf{b}_{(\nu-1)}$ would be impractical. And for thick films with periodic crystal structure and periodic composition through the thickness, the dependence of the energy on vectors sufficiently far from the free surfaces is expected to be similar. In this case, it would be reasonable to retain only the vectors that describe the deformation near the free surfaces. For very thin films, there are interesting effects that relate to shifts in atomic structure, that seem to be related to $\mathbf{b}_1, \ldots, \mathbf{b}_{(\nu-1)}$. For example,

the observations of Sander et al. (1996) on Fe/W films show that the residual stress changes sign at between two and three monolayers of Fe, one of the many true "thin film effects" appearing in the literature.

Acknowledgements

References

Amelinckx, S., Zhang, X.B., Bernaerts, D., Zhang, X.F., Ivanov, V., Nagy, J.B., 1994. A formation mechanism for catalytically grown helix-shaped graphite nanotubes. Science 265, 635–639.

Bernaerts, D., Zhang, X.B., Zhang, X.F., Amelinckx, S., Van Tendeloo, G., Van Landuyt, J., Ivanov, V., Nagy, J.B., 1995. Electron microscopy study of coiled carbon nanotubules. Phil. Mag. A 71, 605–630.

Bhattacharya, K., James, R.D., 1999. A theory of thin films with applications to microactuators. J. Mech. Phys. Solids 47, 465–502.

Brandt, A., 1992. Multigrid methods in lattice field computations. Nuclear Physics B, Proc. Suppl. 26, 137–180.

Brandt, A., 1997. Gauss Center research in multiscale scientific computation. Electronic Trans. Num. Anal. (ETNA) 6, 1–34.

Fefferman, C., 1985. The thermodynamic limit for a crystal. Comm. Math. Phys. 98, 289–311.

Friesecke, G., 2000. From Quantum Mechanics to Density Functional Theory. Springer-Verlag, Berlin, to appear.

Lieb, E.H., Lebowitz, J.L., 1972. The constitution of matter: existence of thermodynamics for systems composed of electrons and nuclei. Advances in Mathematics 9, 316–398.

Olson, G.B., Hartman, H., 1982. Martensite and life: displacive transformations as biological processes. J. de Physique 43, 855–865.

Tadmor, E.B., Ortiz, M., Phillips, R., 1996. Quasicontinuum analysis of defects in solids. Phil. Mag. A 73, 1529–1563.

Iijima, S., 1991. Helical microtubules of graphitic carbon. Nature 354, 56–58.

Rabe, K.M., Waghmare, U.V., 1995. Localized basis for effective Hamiltonians: lattice Wannier functions. Phys. Rev. B52, 13236–13246.

Sander, D., Skomski, R., Schmidthals, C., Enders, A., Kirschner, J., 1996. Film stress and domain wall pinning in sesquilayer iron films on W(110). Phys. Rev. Lett. 77, 2566–2569.

Wong, E.W., Sheehan, P.E., Lieber, C.M., 1997. Nanobeam mechanics: elasticity, strength and toughness of nanorods and nanotubes. Science 277, 1971–1975.