Part One
Fundamentals of Stress and Strain on the Nanoscale

Mechanical Stress on the Nanoscale: Simulation, Material Systems and Characterization Techniques, First Edition.
Edited by Margrit Hanbücken, Pierre Müller, and Ralf B. Wehrspohn.

1
Elastic Strain Relaxation: Thermodynamics and Kinetics

Frank Glas

1.1
Basics of Elastic Strain Relaxation

1.1.1
Introduction

Although frequently used, the phrase *elastic strain relaxation* is difficult to define. It usually designates the modification of the strain fields induced in a solid by a transformation of part or whole of this solid. At variance with *plastic relaxation*, in crystals, elastic relaxation proceeds without the formation of extended defects, thereby preserving *lattice coherency* in the solid.

Elastic strain relaxation is intimately linked with the notion of instability. Indeed, the transformation considered is often induced by the change of a control parameter (temperature, forces applied, flux of matter, etc.). It may imply atomic rearrangements. Usually the realization of the instability is conditioned by kinetic processes (in particular, diffusion), which themselves depend on the stress state of the system. Elastic relaxation may also occur during the formation of part of a system, for instance, by epitaxial growth. The state with respect to which the relaxation is assessed may then exist not actually, but only virtually, as a term of comparison (e.g., the intrinsic state of a mismatched epitaxial layer grown on a substrate). Moreover, it is often only during growth that the kinetic processes are sufficiently active for the system to reach its optimal configuration.

In the present introductory section, we give a general principle for the calculation of strain relaxation and briefly discuss some analytical and numerical methods. In the next sections, we examine important cases where elastic strain relaxation plays a crucial part. Section 1.2 deals with strain relaxation in substitutional alloys with spatially varying compositions and with the thermodynamics and kinetics of the instability of such alloys against composition modulations. Section 1.3 introduces a kinetic process of major importance, namely, diffusion, and summarizes how it is affected by elastic effects. Section 1.4 treats the case of a homogeneous mismatched layer of uniform thickness grown on a substrate. Section 1.5 shows how a system with a planar free surface submitted to a nonhydrostatic stress is unstable with respect to

Mechanical Stress on the Nanoscale: Simulation, Material Systems and Characterization Techniques, First Edition. Edited by Margrit Hanbücken, Pierre Müller, and Ralf B. Wehrspohn.

the development of surface corrugations. Finally, Section 1.6 briefly recalls how the presence of free surfaces in objects of nanometric lateral dimensions, such as quantum dots or nanowires (NWs), permits a much more efficient elastic strain relaxation than in the case of uniformly thick layers.

1.1.2 Principles of Calculation

At given temperature and pressure, any single crystal possesses a reference intrinsic mechanical state $\mathcal{E}_0$ in which the strains and stresses are zero, namely, the state defined by the crystal lattice (and the unit cell) of this solid under bulk form. If the crystal experiences a transformation (change of temperature, phase transformation, change of composition, etc.), this intrinsic mechanical state changes to $\mathcal{E}_1$, where again strains and stresses are zero (Figure 1.1a). The corresponding deformation is the *stress-free strain* (or eigenstrain) ε_{ij}^* with respect to state $\mathcal{E}_0$; for instance, for a change of temperature δT, $\varepsilon_{ij}^* = \delta_{ij}\alpha\delta T$, where α is the thermal dilatation coefficient. If the crystal is mechanically isolated, it simply adopts its new intrinsic state $\mathcal{E}_1$; it is then free of stresses. This is not the case if the transformation affects only part of the system. We then have two extreme cases. The transformation is *incoherent* if it does not preserve any continuity between the crystal lattices of the transformed part and of its environment. If, on the contrary, lattice continuity is preserved at the interfaces, the transformation is *coherent*. This chapter deals with the second case.

Let us call *inclusion* the volume that is transformed and *matrix* the untransformed part of the system (indexed by exponents I and M). Coherency is obviously incompatible with the adoption by the inclusion of its stress-free state $\mathcal{E}_1$, the matrix remaining unchanged. The system will thus relax, that is, suffer additional strains, which in general affect both inclusion and matrix. It is a strain relaxation in the following sense: if one imagines the inclusion having been transformed (for instance, heated) but remaining in its original reference mechanical state $\mathcal{E}_0$ (which restores coherency, since the matrix has not been transformed from state $\mathcal{E}_0$), it is subjected to stresses, since forces must be applied at its boundary to bring it from its new intrinsic state $\mathcal{E}_1$ back to $\mathcal{E}_0$. With these stresses is associated an elastic energy. The coherent deformation of the whole system constitutes the elastic relaxation.

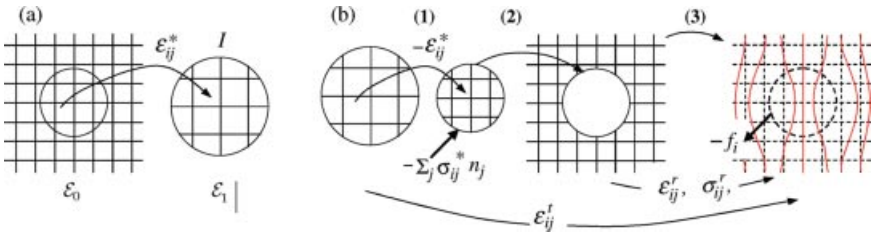

Figure 1.1 (a) Stress-free strain relative to the inclusion. (b) The three stages of an Eshelby's process.

This suggests a way to calculate relaxation, Eshelby's method (Figure 1.1b) [1]:

1) One applies to the transformed inclusion (state $\mathcal{E}_1$) the strain $-\varepsilon_{ij}^{I*}$, which brings it back to state $\mathcal{E}_0$. This implies exerting on its external surface (whose external normal $\mathbf{n}$ has components n_j) the forces $-\sum_{j=1}^{3}\sigma_{ij}^{I*}n_j$ per unit area, where σ_{ij}^{I*} is the stress associated[1)] with the stress-free strain ε_{ij}^{I*}.
2) Having thus restored coherency between inclusion and matrix, one may reinsert the former into the latter. The only change that then occurs is the change of the surface density of forces applied at stage (1) into a body density f_i, since the surface of the inclusion becomes an internal interface.[2)]
3) The resulting state is not a mechanical equilibrium state, since forces f_i must be applied to maintain it. One then lets the system relax by suppressing these forces, that is, by applying forces $-f_i$, *while at the same time maintaining coherency everywhere*. One thus has to compute the strain field, in the inclusion ($\varepsilon_{ij}^{\mathrm{Ir}}$) and in the matrix ($\varepsilon_{ij}^{\mathrm{Mr}}$), solution of the elasticity equations for body forces $-f_i$, under the coherency constraint, which amounts to equal displacements $\mathbf{u}^{\mathrm{Ir}} = \mathbf{u}^{\mathrm{Mr}}$ at the interface.

We may generalize this approach by not differentiating matrix and inclusion. The whole system experiences a transformation producing an inhomogeneous stress-free strain $\varepsilon_{ij}^{*}(\mathbf{r})$ (defined at any point $\mathbf{r}$) with respect to initial uniform state $\mathcal{E}_0$ (perfect crystal). One then applies the body forces producing strain $-\varepsilon_{ij}^{*}$, namely, $f_i = \sum_{j=1}^{3}\partial\sigma_{ij}^{*}/\partial x_j$, where $\sigma_{ij}^{*}(\mathbf{r})$ is the stress associated with strain ε_{ij}^{*}. Finally, one calculates the relaxation field $\varepsilon_{ij}^{\mathrm{r}}$, the solution of the elastic problem with forces $-f_i(\mathbf{r})$ that preserves coherency everywhere (displacements must be continuous). It is important to specify the reference state with respect to which one defines the final state of the system. It is often easier to visualize the relaxed state relative to the uniform state $\mathcal{E}_0$; strain is then simply $\varepsilon_{ij}^{\mathrm{r}}$. If, on the contrary, the elastic energy W stored in the system is to be calculated, we must take as a reference state for each volume element its intrinsic state *after transformation* ($\mathcal{E}_1$), with respect to which the total strain is $\varepsilon_{ij}^{\mathrm{t}} = \varepsilon_{ij}^{\mathrm{r}} - \varepsilon_{ij}^{*}$. Hence, $W = (1/2)\int_V \sum_{i,j=1}^{3}\varepsilon_{ij}^{\mathrm{t}}\sigma_{ij}^{\mathrm{t}}\mathrm{d}V$, where σ_{ij}^{t} is the strain associated with $\varepsilon_{ij}^{\mathrm{t}}$ and where the integral is taken over the whole volume (in the reference state).

In case of an inclusion (Figure 1.1), one may easily show that

$$W = -\frac{1}{2}\int_I \sum_{i,j=1}^{3}\varepsilon_{ij}^{I*}\sigma_{ij}^{\mathrm{It}}\mathrm{d}v \tag{1.1}$$

This is a fundamental result obtained by Eshelby [1]. In particular, the total elastic energy depends only on the stress in the inclusion.

An example of application to an infinite system with a continuously varying transformation will be given in Section 1.2. Eshelby's method may also be adapted to

1) Via the constitutive relations, for instance, Hooke's law in linear elasticity.

2) In the present case (single inclusion), this density is nonzero only in the zero-thickness interface layer, so for a facet $x = x_0$, one has $f_i = -\sigma_{ix}^{I*}\delta(x - x_0)$.

other problems. In particular, if the interface between matrix and inclusion does not entirely surround the latter (which happens if the inclusion has a free surface), it is not necessary to apply strain $-\varepsilon_{ij}^{I*}$ to the inclusion at stage 1. It suffices to apply a strain that restores the coherency in the interface, which may make the solution of the problem simpler. An example is given in Section 1.4.

1.1.3 Methods of Calculation: A Brief Overview

The problem thus consists in determining the fields relative to stage 3 of the process. One has to calculate the elastic relaxation of a medium subjected to a given density $-f_i$ of body forces. In addition to numerical methods, for instance, those based on finite elements, there exist several analytical methods for solving this problem, in particular, the Green's functions method [2] and the Fourier synthesis method.

In elasticity, Green's function $G_{ij}(\mathbf{r}, \mathbf{r}')$ is defined as the component along axis i of displacement at point $\mathbf{r}$ caused by a unit body force along j applied at point $\mathbf{r}'$. For a solid with homogeneous properties, it is a function $G_{ij}(\mathbf{r}-\mathbf{r}')$ of the vector joining the two points. One easily shows that for an elastically linear solid (with elastic constants C_{jklm}), the displacement field at stage 3 is

$$\mathbf{u}_i(\mathbf{r}) = -\sum_{j,k,l,m=1}^{3} C_{jklm} \int \varepsilon_{lm}^{*}(\mathbf{r}') \frac{\partial G_{ij}}{\partial x_k}(\mathbf{r}-\mathbf{r}')\mathrm{d}\,\mathbf{r}' \tag{1.2}$$

where the integral extends to all points $\mathbf{r}'$ of the volume. Green's functions depend on the elastic characteristics of the medium, but, once determined, any problem relative to this medium is solved by a simple integration. However, if the Green's functions for an infinite and elastically isotropic solid have been known since 1882, only a few cases have been solved exactly. If the medium is not infinite in three dimensions, the Green's functions also depend on its external boundary and on the conditions that are imposed to it. For epitaxy-related problems, the case of the half-space (semi-infinite solid with planar surface) is particularly interesting. These functions have been calculated for the elastically isotropic half-space with a free surface (no external tractions) [3, 4]. Mura's book gives further details [2]. The method also applies to the relaxation of two solids in contact via a planar interface; in this case, this surface is generally not traction-free and the boundary conditions may be on these tractions or on its displacements. Pan has given a general solution in the anisotropic case, valid for all boundary conditions [5].

In the Fourier synthesis method, one decomposes the stress-free strain distribution into its Fourier components: $\varepsilon_{ij}^{*}(\mathbf{r}) = \int \tilde{\varepsilon}_{ij}^{*}(\mathbf{k})\exp(i\mathbf{kr})\mathrm{d}\,\mathbf{k}$, where $\mathbf{k}$ is the running wave vector. In linear elasticity, the solution is simply the sum, weighted by the Fourier coefficients $\tilde{\varepsilon}_{ij}^{*}(\mathbf{k})$, of the solutions relative to each periodic wave of wave vector $\mathbf{k}$, which are themselves periodic with the same wave vector. If the system is infinite, the elementary solution is easily determined (see Section 1.2.2). The only nontrivial point is then the integration. This method allows one to treat elegantly the

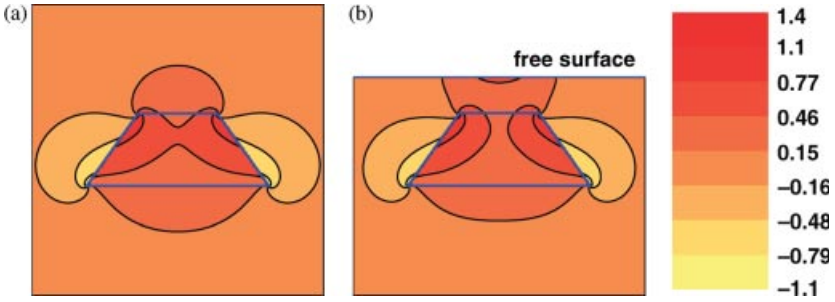

Figure 1.2 Comparison of the elastic relaxation of a truncated square-base pyramidal inclusion in an infinite matrix (a) and in a half-space (b). Maps in the symmetry plane *xz* of strain component ε_{xx} normalized to the intrinsic misfit ε_0 of the inclusion with respect to the matrix. Analytical calculation by the Fourier method (see also Ref. [10]). Thick blue lines mark the inclusion contour and the free surface, and the intensity scale is the same for (a) and (b).

stress-free strain discontinuities, as happening at the interface between a matrix and a misfitting inclusion[3] having the same elastic constants.

The number of problems solved by these methods steadily increases. As for inclusions, let us only mention, in addition to Eshelby's pioneering work on the ellipsoidal inclusion [1], the case of parallelepipedic inclusions in an infinite matrix [6] and in a half-space [7, 8] and that of the truncated pyramidal inclusions in an infinite matrix [9] and in a half-space [10]. These are important for being the shapes commonly adopted by semiconducting quantum dots. For a given inclusion, the elastic relaxation may be deeply modified by a free surface, on which the tractions must vanish (Figure 1.2).

1.2
Elastic Strain Relaxation in Inhomogeneous Substitutional Alloys

As a first example of strain relaxation, we examine the common case of an alloy whose stress-free state (for instance, its lattice parameter) depends on its composition. If composition variations that preserve lattice coherency develop in an initially homogeneous alloy, internal strains and stresses appear. In this case, elucidating strain relaxation amounts to calculating these fields. We shall see that strains may deeply affect the stability of such an alloy. However, the stability is not determined only by elastic effects. To provide a term of comparison, we shall first consider alloys where compositions variations induce no strain (Section 1.2.1), before turning to the calculation of the elastic fields (Section 1.2.2) and to the way in which they alter the alloy thermodynamics (Section 1.2.3). Finally, Section 1.2.4 discusses how the presence of a free surface affects strain relaxation and hence stability.

3) We call misfitting an inclusion (respectively, a layer) having the same crystal structure as the matrix (respectively, substrate) but different lattice parameters.

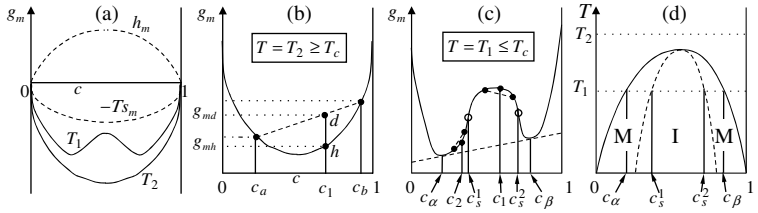

Figure 1.3 (a) Mixing free enthalpy g_m at two temperatures (full lines), with enthalpic and entropic contributions (dashed lines). (b) High temperature case. (c) Low temperature case; tangent construction, equilibrium compositions c_α, c_β and spinodal compositions c_1^s, c_2^s corresponding to inflexion points (circles). (d) Boundaries of miscibility gap (full line) and spinodal gap (dashed lines).

1.2.1
Spinodal Decomposition with No Elastic Effects

Let us consider a bulk binary AB substitutional alloy with atomic concentrations $(1-c)$ and c in species A and B.[4] With each thermodynamic quantity is associated a mixing *quantity*, the difference of this quantity between the alloy and the same numbers of its atomic constituents taken as pure solids. In the regular solution model, the mixing free enthalpy (per atom) of the disordered alloy at temperature T is $g_m = h_m - Ts_m = \Omega c(1-c) + k_B T[c \ln c + (1-c) \ln (1-c)]$, where Ω is an energy, the *interaction parameter* (here taken per atom), and k_B is the Boltzmann's constant (Figure 1.3a). If $\Omega > 0$, the alloy tends to decompose, since alloying its constituents produces an energy h_m. However, at high temperatures, entropy may prevent decomposition. The equilibrium state of the system at given temperature and pressure is found by minimizing g_m at constant numbers of atoms.

Depending on temperature T, g_m adopts two different forms, with either a single or a double minimum (Figure 1.3a).[5] At high T (Figure 1.3b), the mixing free enthalpy g_{mh} of a homogeneous alloy of any composition c_1 is less than that of any mixture of two homogeneous alloys of compositions c_a and c_b into which it might decompose (the latter, g_{md}, is given by the *cord construction*), so the alloy is stable. On the contrary, for $T \leq T_c^0$, with $T_c^0 = \Omega/(2k_B)$, the *tangent construction* gives two stable equilibrium compositions c_α, c_β, into which any alloy with intermediate composition tends to decompose (Figure 1.3c). However, two different behaviors are expected. For an alloy of composition c_1 intermediate between the compositions $c_s^1(T)$ and $c_s^2(T)$ corresponding to the inflexion points of the g_m curve (i.e., such that $g_m''(c_1) = (\partial^2 g_m/\partial c^2)_{c=c_1} < 0$), decomposition into two alloys is always energetically favorable,

4) This also applies to pseudobinary ternary alloys with two sublattices, one homogeneous and the other mixed, such as alloys of compound semiconductors.

5) In the regular solution model, which we retain in the following calculations, the curve is symmetrical about $c = 0.5$. In the figure, we illustrate a slightly more general situation highlighting the salient feature of the curve, its double-well.

even if these two alloys differ infinitesimally in composition (curve g_m is above its cord). Conversely, for an alloy composition such as c_2, between a stable composition and an inflexion point ($g''_m(c_2) > 0$), it is only decomposition into two compositions differing by a finite amount that is favored (curve g_m is below its cord). The homogeneous alloy is unstable in the first case and metastable in the second. In the (c, T) plane (Figure 1.3d), curves $c_\alpha(T)$, $c_\beta(T)$ enclose the miscibility gap, inside which the spinodal curve, locus of points $c_s^1(T)$, $c_s^2(T)$, separates the metastable (M) and unstable (or spinodal, I) domains. The process through which an alloy quenched below T_c^0 starts decomposing via composition variations of arbitrary small amplitude is *spinodal decomposition*.

The same results are obtained by linear stability analysis [11]. This general method consists in studying the instability of a system not against an arbitrary perturbation but against a Fourier component of the latter. If the problem is linear (a linear combination of solutions is a solution, the boundary conditions being themselves linear), global instability is equivalent to the existence of at least one unstable elementary perturbation. Here, we study the instability of an alloy of average composition c_1 against a sinusoidal composition modulation of amplitude δ_c along the direction x: $c(x) = c_1 + \delta_c \sin kx$. The mixing free enthalpy of the modulated alloy, found by averaging g_m over a modulation wavelength $\lambda = 2\pi/k$, is $\lambda^{-1}\int_0^\lambda g_m[c(x)]dx = g_m(c_1) + g''_m(c_1)\delta_c^2/4$ at order 2 in δ_c. The *excess* of mixing enthalpy due to modulation is $\delta g_{mc} = g''_m(c_1)\delta_c^2/4$. We thus recover the previous result, namely, an alloy of average composition c_1 is unstable ($\delta g_{mc} < 0$) at temperature T against a composition modulation of vanishing amplitude if $g''_m(c_1) < 0$.[6)]

The foregoing analysis assumes that fractioning the alloy into two phases or modulating its composition produces no excess energy.[7)] It is however not generally the case, and this can be taken into account by adding to the energy density of the composition modulation a Landau-type phenomenological gradient term that opposes abrupt composition variations [12]. This produces a critical decomposition wavelength, lower wavelengths being energetically unfavorable. The gradient term may be of purely chemical origin, but, in addition, in alloys with size effects that decompose coherently, stresses appear between regions having different compositions and generate additional elastic energy. The next section describes the modifications of the previous results due to these stresses and to the way in which they relax.

1.2.2
Elastic Strain Relaxation in an Alloy with Modulated Composition

Let us assume, for simplicity, that the alloy is of cubic structure or elastically isotropic and that its intrinsic lattice parameter varies linearly with composition (Vegard's law) $a(c) = a_0(1 + \eta c)$ with $a_0 = a(0)$ and $\eta = [a(1) - a(0)]/a(0)$, the relative *lattice mismatch* between the pure constituents A and B. If the alloy composition lies within

6) In the regular solution model, $g''_m(c_1) = -2\Omega + k_B T/[c_1(1-c_1)]$ and T_c^0 is obtained for $c_1 = 0.5$.

7) In the first case, this corresponds to an interface energy.

the metastable domain M, one may expect the nucleation of finite volumes of phases with distinct compositions, hence with different lattice parameters, and decomposition may be coherent or not, depending on the size and composition contrast. In the unstable domain I, the transformation into the stable phases may however occur via the amplification of composition modulations starting with zero amplitude (Section 1.2.1). One then expects the decomposition to be coherent, at least at the start of the process. This is the case we treat here. Two major questions then arise. What are the elastic strain fields induced by a given composition distribution? How does the associated elastic energy modify the stability of the alloy?

To answer these questions, let us consider again a sinusoidal composition modulation $c(x) = c_1 + \delta_c \sin kx$ in direction x. From Vegard's law, this induces a spatial modulation of *intrinsic* parameter $a(x) = a_0[1 + \eta c_1 + \varepsilon_c \sin kx]$ with amplitude $\varepsilon_c = \eta\delta_c$. As regards elasticity (Section 1.1.2), this corresponds to a stress-free strain modulation $\varepsilon^*_{ij}(x) = \delta_{ij}\varepsilon_c \sin kx$ with respect to the homogeneous state (c_1).[8] From Section 1.1.2, the elastic relaxation with respect to the latter is obtained by imposing a density of body forces $-f_i(x) = -\sum_{j=1}^{3} \partial\sigma^*_{ij}/\partial x_j$. If the alloy is elastically isotropic,[9] $-f_i(x) = -\delta_{ix}[E/(1-2\nu)]\varepsilon_c k \cos kx$. We look for a relaxation displacement field that is continuous (coherency condition) and has the same wavelength as the perturbation: $u^r_i(x) = \delta_{ix}(A \cos kx + B \sin kx)$. The relaxation strain is then $\varepsilon^r_{ij}(x) = \delta_{ix}\delta_{jx}k(-A \sin kx + B \cos kx)$[10] and the associated stress $\sigma^r_{xx} = \{E(1-\nu)/[(1+\nu)(1-2\nu)]\}\varepsilon^r_{xx}$, $\sigma^r_{yy} = \sigma^r_{zz} = [\nu/(1-\nu)]\sigma^r_{xx}$, $\sigma^r_{ij} = 0$, if $i \neq j$. From the equilibrium between stresses and forces $-f_i$, one gets $B = 0$ and $A = -[(1+\nu)/(1-\nu)]k^{-1}\varepsilon_c$. The relaxation field is thus a tetragonal strain modulation along x in phase with the intrinsic modulation but amplified by factor $[(1+\nu)/(1-\nu)]$ [12]. The relaxed lattice parameter remains equal to $a(c_1)$ in the directions normal to x. The elastic energy is that of the total strain fields ($\varepsilon^t_{ij} = \varepsilon^r_{ij} - \varepsilon^*_{ij}$) (Section 1.1.2). Its density is easily found to be $\delta w = (1/2)E'\varepsilon_c^2/(1-\nu)$, where $E' = E/\mathcal{N}$ is the Young's modulus E normalized by Avogadro's number $\mathcal{N}$ in order to obtain a density per atom.

What is the meaning of *relaxation* here? With respect to the intrinsic (reference) state of each volume element of the modulated alloy, in which elastic energy is zero by definition, there is an increase of elastic energy due to the coherency between these elements. On the contrary, the energy is lower than that in the state where all elements (with different compositions) would adopt the average lattice parameter (virtual reference state at the end of stage 2; Figure 1.4). If, at least during the early stages of decomposition, the system prefers continuous composition variations and coherent relaxation, it is because the formation of finite domains with distinct compositions, which would all adopt their intrinsic lattice parameter, would induce interfacial defects whose cost in energy would be even higher (see also Section 1.4).

8) Such a problem, where the stress-free strain is a pure dilatation, is a *thermal stress problem*.

9) Its elastic behavior is then governed by two parameters, Young's modulus E and Poisson's ratio ν, and strains and stresses are linked by relations $\sigma_{ij} = [E/(1+\nu)]\left\{\varepsilon_{ij} + [\nu/(1-2\nu)]\delta_{ij}\sum_{m=1}^{3}\varepsilon_{mm}\right\}$ and $\varepsilon_{ij} = [(1+\nu)/E]\left\{\sigma_{ij} - [\nu/(1+\nu)]\delta_{ij}\sum_{m=1}^{3}\sigma_{mm}\right\}$.

10) The only nonzero element of the strain tensor is ε_{xx}.

1.2.3
Strain Stabilization and the Effect of Elastic Anisotropy

The *total* excess of mixing free enthalpy δg_m due to modulation is found by adding the elastic relaxation energy δw to the term δg_{mc} calculated in Section 1.2.1 without taking stresses into account. One finds $\delta g_m = \left[\eta^{-2} g_m''(c_1) + 2E'/(1-\nu)\right]\varepsilon_c^2/4$. The condition of instability of alloy with composition c_1 against modulations of vanishing amplitude thus changes from $g_m''(c_1) < 0$ to

$$g_m''(c_1) + 2\eta^2 E'/(1-\nu) < 0 \tag{1.3}$$

Since $\nu < 1$, the extra term is positive; so the condition for instability is more restrictive in terms of composition. If the mixing enthalpy at low T is a double-well curve (Figure 1.3c), we see that the domain of unstable compositions is reduced and the critical temperature lowered, an effect often called *stress stabilization*. In the regular solution model, the critical temperature decreases from $T_c^0 = \Omega/(2k_B)$ to $T_c^C = \{1-\eta^2 E'/(\Omega(1-\nu))\}T_c^0$. This reduction may reach several hundreds of kelvins in some metallic alloys [12]. The elastic contribution may even be large enough to render negative the calculated T_c^C, which means that the instability is then totally suppressed (no composition satisfies Eq. (1.3)).

The previous calculations, based on isotropic elasticity, do not specify any favored direction of modulation. To extend the calculations to cubic crystals [13], it suffices to replace term $\eta^2 E'/(1-\nu)$ in Eq. (1.3) by a modulus $Y_{\hat{\mathbf{k}}}$ depending on modulation direction $\hat{\mathbf{k}} = \mathbf{k}/k$. In particular, $Y_{100} < Y_{110} < Y_{111}$ if the elastic constants satisfy to $2C_{44} + C_{12} - C_{11} > 0$. Modulations then tend to form in the *soft* directions of type $\langle 100 \rangle$. These considerations are corroborated by the observation in spinodally decomposed alloys of a characteristic microstructure, manifested in the transmission electron microscopy images by a modulated contrast in the soft directions [11, 14]. More generally, in elastically anisotropic materials, the shape of inclusions and their relative disposition tend to be determined by elastic relaxation [11].

1.2.4
Elastic Relaxation in the Presence of a Free Surface

The presence of a free surface may also deeply affect strain relaxation. In the infinite solid considered so far, the elastic energy of any composition distribution is the sum of those of its Fourier components, which do not interact since $\int_V \mathbf{e}^{i\mathbf{k}\cdot\mathbf{r}}\mathbf{e}^{i\mathbf{k}'\cdot\mathbf{r}}\,d\mathbf{r} = 0$ if $\mathbf{k} \neq \mathbf{k}'$. This does not necessary hold in a solid bounded by a surface (see Section 1.5). However, this is still true for a planar half-space. Moreover, for a modulation with a wave vector parallel to the surface, the elastic energy may be considerably reduced with respect to the infinite solid. This reduction stems from the extra stress relaxation permitted by the free surface. The optimal modulations have an amplitude that is exponentially attenuated in the direction normal to the surface. Strain stabilization is thus less pronounced than in the bulk and the critical temperature increases accordingly [8]. In the regular solution model, one finds a new critical temperature

$T_c^I = \{1-\eta^2 E'/(2\Omega)\}T_c^0$, such that $T_c^C < T_c^I < T_c^0$ [15, 16].[11] Composition modulations in directions parallel to the substrate have indeed been observed in epitaxial layers of semiconducting alloys [17, 18].

More generally, the strain relaxation of misfitting inclusions may be strongly affected if they lie close to a free surface, typically within a distance on the order of their dimensions (Figure 1.2) [10].

1.3 Diffusion

Many transformations (such as considered in Section 1.1.1) require a redistribution of matter in the system to become effective. Hence, the realization of the instability is conditioned by kinetic processes, in particular, diffusion. In this sense, the elastic relaxation that accompanies the transformation is also conditioned by diffusion. However, one usually considers that the timescale for diffusion is much longer than that for the mechanical adjustment (relaxation) of the system to the instantaneous distribution of atoms, so the system is continuously in a state of *mechanical* equilibrium (but of course not in global thermodynamic equilibrium) during the transformation (quasistatic approximation). This is not to say that diffusion and strain relaxation are independent: we shall see in Section 1.3.2 that diffusion is affected by the elastic strain and stress fields. Before this, we briefly consider diffusion without elastic effects.

1.3.1 Diffusion without Elastic Effects

As an example of how diffusion conditions the realization of an instability, consider a bulk alloy AB subject to decomposition (as in Section 1.2), and first ignore the elastic effects. If the homogeneous alloy is quenched below its critical temperature T_c^0, it becomes unstable and tends to decompose (Section 1.2.1). The concentration c of, say, component B becomes nonuniform. According to Fick's law, this induces a *diffusive flux* of B atoms, $\mathbf{J}_B^d = -D_B\nabla(c/\omega)$, where $D_B > 0$ is the diffusion coefficient of B [19] and ω is the atomic volume in the homogeneous reference state of the alloy.[12] Setting $D'_B = D_B/\omega$, we have $\mathbf{J}_B^d = -D'_B\nabla c$. Since $D'_B > 0$, this flux tends to smooth the concentration gradient, which should inhibit the formation of the composition modulation. It is because the B flux is not limited to Fick's diffusive term that the modulation may actually develop.

Indeed, generally, to the diffusive flux must be added a *transport flux* depending on the force $\mathbf{F}_B$ exerted on each B atom [19]. If this force derives from a potential ($\mathbf{F}_B = -\nabla\phi_B$), the Nernst–Einstein equation (demonstrated by canceling the total flux at equilibrium and by using the appropriate statistics) leads to

11) It is obtained for modulations with a z-dependent amplitude, such as will be considered in Section 1.5.5.

12) This appears here since we use, as in Section 1.2.1, the atomic concentration c rather than the number of atoms per unit volume.

$\mathbf{J}_B^t = D'_B/(k_B T)c\mathbf{F}_B = -D'_B/(k_B T)c\nabla\phi_B$. For instance, in the case of charged particles in an electric field, the transport flux is directly related to the electrostatic force exerted on each particle. Now, if the alloy is ideal, the chemical potential of species B at concentration c is $\mu_B = k_B T \ln\left(c/c_{eq}\right)$, where c_{eq} is the equilibrium concentration at temperature T, so the diffusive flux becomes $\mathbf{J}_B^d = -D'_B/(k_B T)c\nabla\mu_B$ and the total flux is $\mathbf{J}_B = \mathbf{J}_B^d + \mathbf{J}_B^t = -D'_B/(k_B T)c\nabla(\mu_B + \phi_B) = -D'_B/(k_B T)c\nabla\tilde{\mu}_B$, where $\tilde{\mu}_B = \mu_B + \phi_B$ is a generalized potential, including external forces.

In the case of the alloy, a similar approach applies, but the definition of the appropriate potential and the generalization of Fick's law require some care. If the alloy is substitutional, A and B share the same crystal lattice and, in the absence of vacancies, any B atom leaving a site must be replaced by an A atom so that the fluxes of A and B are opposite: $\mathbf{J}_B = -\mathbf{J}_A$. The B flux then becomes $\mathbf{J}_B = -\mathcal{D}\nabla(\mu_B - \mu_A) = -\mathcal{D}\nabla M$, where $\mathcal{D}$ is a phenomenological diffusion mobility, μ_A and μ_B are the local and concentration-dependent values of the chemical potentials of A and B, and M is the *diffusion potential* that replaces and generalizes the chemical potential. Since $\mu_B - \mu_A$ is the change of free enthalpy when an A atom is replaced by a B atom at the point considered, we have $\mu_B - \mu_A = \partial g/\partial c$, where g is the free enthalpy (per atom) of the alloy at composition c. Moreover, by definition of the free enthalpy of mixing g_m introduced in Section 1.2, we have $g = c\mu_B^0 + (1-c)\mu_A^0 + g_m$, where μ_A^0, μ_B^0 are the position-independent chemical potentials of the pure elements. Hence, $\mathbf{J}_B = -\mathcal{D}\nabla M = -\mathcal{D}\nabla(\partial g_m/\partial c) = -\mathcal{D}(\partial^2 g_m/\partial c^2)\nabla c$. It now becomes clear that in the spinodal domain, which is precisely defined by $\partial^2 g_m/\partial c^2 < 0$ (Section 1.2.1), the diffusion flux of species B does not oppose the gradient (as follows from the sole consideration of Fick's diffusive flux) but actually tends to amplify it. In other words, as expected in this domain, the alloy is unstable against composition modulations of vanishing amplitude.

This calculation of the diffusion flux is actually much richer than simply confirming the static analysis carried out in Section 1.2.1. It opens the way to a study of the kinetics of decomposition (provided $\mathcal{D}$ is known), since the evolution of the composition profile with time t obeys the usual conservation equation $\mathrm{div}\,\mathbf{J}_B + \partial c/\partial t = 0$. The interested reader is referred to, for example, Cahn's publications [14]. This general approach also permits to take into account any effect that modifies the energy of the system (e.g., electric or magnetic fields) by simply adding its contribution to the free enthalpy g. In the next section, we show how the effect of elastic strain fields on diffusion can be calculated in this way.

1.3.2
Diffusion under Stress in an Alloy

The elastic contribution to the free enthalpy per atom is simply $[1/2]\omega\varepsilon_{ij}^t\sigma_{ij}^t$, where $\varepsilon_{ij}^t = \varepsilon_{ij} - \varepsilon_{ij}^*$ is the strain corresponding to the transformation of a given volume element from its intrinsic state to its final relaxed state (Figure 1.1b and 1.4) and σ_{ij}^t is the associated stress (see Section 1.1.2). Since σ_{ij}^t is defined only formally (ε_{ij}^t being the strain between the independent stress-free volume elements and the final relaxed coherent state), it is preferable to express the density of elastic energy as a function of the elastic fields describing the transformation of the homogeneous alloy with composition c_1 into the modulated and relaxed alloy (the extreme stages of Figure 1.4),

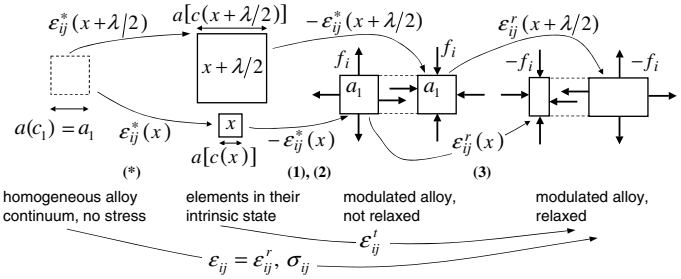

Figure 1.4 Eshelby process for a bulk alloy with modulated composition.

namely, $[1/2]\omega\left(\varepsilon_{ij}-\varepsilon^*_{ij}\right)\sigma_{ij}$. Elasticity introduces a complication, namely, nonlocal effects, since the stress and strain at a given point depend not only on local composition (as was the case with the chemical potentials) but also on the whole concentration distribution. Nevertheless, it is possible to express the effect of elasticity as an additional term in the diffusion potential M [20], so that the diffusion equation becomes $\mathbf{J}_B = -\mathcal{D}\nabla M'$, with

$$M' = \frac{\partial g_m}{\partial c} - \omega \sum_{i,j} \frac{\partial \varepsilon^*_{ij}}{\partial c}\sigma_{ij} - \omega \sum_{i,j,k,l} \frac{\partial S_{ijkl}}{\partial c}\sigma_{ij}\sigma_{kl} \tag{1.4}$$

where S_{ijkl} are the elastic compliances. If the latter do not depend on composition, the third term of Eq. (1.4) disappears. If this is the case and if the material is cubic or elastically isotropic, then $\varepsilon^*_{ij} = \eta(c-c_1)\delta_{ij}$, where η is the relative lattice mismatch between the pure constituents introduced in Section 1.2.2; so $M' = \partial g_m/\partial c - \omega\eta\sum_k \sigma_{kk}$, where $\sum_k \sigma_{kk}$ is simply the local dilatation. Note that only nonhomogeneous stresses affect diffusion in this way, although, in addition, stress may affect the elastic constants and the diffusion coefficients. For more details on diffusion under stress, including surface diffusion, see Refs [21–23].

Spinodal decomposition may be studied in this way, considering a possible gradient energy of chemical origin (Section 1.2.1) plus the elastic energy. This leads to different growth rates for perturbations of different wavelengths, with a rather narrow peak centered around a fastest developing wavelength. Experimentally, spinodally decomposed alloys indeed tend to exhibit a microstructure with a fairly well-defined decomposition periodicity [14].

1.4
Strain Relaxation in Homogeneous Mismatched Epitaxial Layers

1.4.1
Introduction

Knowing which amount of a given material can be deposited coherently on a mismatched substrate is of great importance in the field of epitaxy. For semicon-

ducting materials in particular, the extended defects that form when plastic relaxation occurs often affect deleteriously the electrical and optical properties of heterostructure-based devices, and great care is usually taken to remain in the coherency domain during growth. In a simple equilibrium picture, the transition between elastic and plastic relaxation is governed by the total energy of the system. It is thus important to analyze elastic strain relaxation in the coherent case. In the present section, we consider the standard case of a layer of uniform thickness, before turning to possible thickness variations in Section 1.5.

We consider mismatched heterostructures formed by depositing a homogeneous layer L *of uniform thickness t* ($0 < z < t$) onto a semi-infinite planar substrate S ($z < 0$), both being single crystals. In the spirit of Section 1.1.2, one may consider that in its intrinsic stress-free state, L results from an elastic distortion of S, described by the stress-free strain that transforms the intrinsic lattice of S into that of L [8, 24]. The methods of Section 1.1 may then be applied. We further assume that S and L have the same structure and differ only by the magnitude of their lattice parameters, so the stress-free strain ε_{ij}^{*} is a pure isotropic dilatation, equal to the relative difference between the lattice parameters a_{L} of the layer and a_{S} of the substrate (or lattice mismatch), $\varepsilon_0 = (a_{\mathrm{L}} - a_{\mathrm{S}})/a_{\mathrm{S}}$ (which is positive if the layer is compressed by the substrate).[13)]

If the deposit L extends infinitely parallel to the planar interface (plane xy), the lattice mismatch makes it impossible for S and L to retain their intrinsic bulk stress-free states if the interface is coherent, simply because the spacings of the lattice planes that cross the interface are different for S and L. Hence the necessity of an *accommodation* of the lattice mismatch, which can take two extreme forms. If the coherency at the interface is preserved, thanks to a deformation of one or both materials, the accommodation is purely elastic. Conversely, accommodation may be realized plastically, via the formation of a network of misfit dislocations at the interface, which thereby becomes incoherent (section 1.1.1). Leaving aside plastic relaxation, the detailed discussion of which falls outside the scope of the present chapter, we shall briefly show how the methods of Section 1.1 give the solution of the *elastic* problem.

1.4.2
Elastic Strain Relaxation

If the deposit is much thinner than the substrate, one can safely consider (ignoring possible curvature effects) that only the former is strained while the latter retains its bulk lattice. Moreover, the layer has a free surface $z = h$, so the substrate/layer interface $z = 0$ does not entirely surround the layer. As mentioned in Section 1.1.2, it is then not necessary to apply strain $-\varepsilon_{ij}^{*} = -\varepsilon_0$ at stage 1 of the Eshelby process. It suffices to apply a strain that restores lattice continuity across the S/L interface, that is, such that $\varepsilon_{xx}^{(1)} = \varepsilon_{yy}^{(1)} = -\varepsilon_0$. This may be achieved by applying forces along x and y on the elementary cubes that compose the layer, but not along z, so at this (modified)

13) The problem is then equivalent to heating or cooling a layer of thickness h of a stress-free half-space of material S.

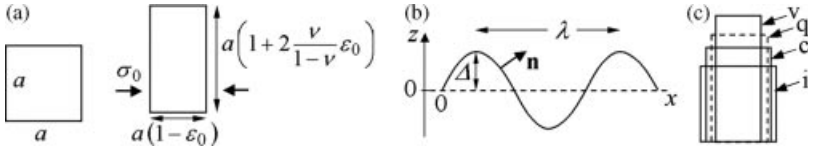

Figure 1.5 (a) Tetragonal strain of a unit cell of parameter a subjected to biaxial stress σ_0 (here taken negative, corresponding to compression); $\varepsilon_0 = -(1-\nu)\sigma_0/E$. (b) Planar (dashed line) and corrugated (full line) surfaces. (c) Schematics of strain at crests (c) and valleys (v) with respect to intrinsic (i) and tetragonally strained (q) unit cells.

stage 1, $\sigma_{zz}^{(1)} = 0$ (Figure 1.5a). Considering, to simplify, an elastically isotropic medium, the equations of elasticity (see footnote 9) indicate that one then has $\sigma_{xx}^{(1)} = \sigma_{yy}^{(1)} = -E\varepsilon_0/(1-\nu) = \sigma_0$ and $\varepsilon_{zz}^{(1)} = 2\nu\varepsilon_0/(1-\nu)$ (all nondiagonal strains and stresses are zero). Because the forces corresponding to $\sigma_{xx}^{(1)}$, $\sigma_{yy}^{(1)}$ are uniform, they cancel mutually when the layer elements are reassembled at stage 2, so the corresponding density of body forces is zero. Moreover, since there are no forces along z, the layer is entirely free of applied forces at the end of stage (2), so that considering stage (3) becomes irrelevant. This adaptation of Eshelby's method thus offers a particularly simple solution of the problem.

To summarize, elastic relaxation affects only the thin layer, which adapts its parameter to that of the substrate in the interface plane and strains tetragonally (extending or contracting, depending on the sign of ε_0) in the normal (z) direction, with an amplification factor $(1+\nu)/(1-\nu)$ with respect to intrinsic strain ε_0, since the dilatation along z with respect to the substrate is $\varepsilon_{zz}^{(1)} + \varepsilon_0 = [(1+\nu)/(1-\nu)]\varepsilon_0$. The corresponding elastic energy *per unit volume* of layer is readily found to be $(E/(1-\nu))\varepsilon_0^2$.

1.4.3
Critical Thickness

One of the major issues regarding strain accommodation in heterostructures is to find out which factors determine the mode of relaxation (elastic or plastic). It is usually observed that dislocations do not form until the growing layer reaches some *critical thickness* h_c. Basically, such a critical thickness exists because the energies stored in the system per unit area scale differently with layer thickness h. In the coherent state (elastic relaxation), we have just seen that the energy is uniformly distributed in the layer, so the energy per unit area scales with h. On the contrary, in the plastically relaxed state, the density of dislocations of a given type that accommodates a given mismatch (ensuring that the material passes from the stress-free lattice parameter of the substrate to that of the layer across their interface) is fixed and inversely proportional to the relative mismatch [25]. In this case, as a first approximation, the layer is strained in the vicinity of the interface (because of the nonuniform strain fields of the dislocations) but quickly recover its stress-free parameter away from it, so the elastic energy per unit area does not depend on h.

Actually, since the dislocations have a long-range strain field, the elastic energy increases with h, but only logarithmically, much more slowly than in the elastic case (it even saturates when the layer thickness becomes larger the dislocation spacing). In addition, the dislocation cores contribute a constant term to the energy. Hence, the energy stored is larger in the plastic case at low layer thicknesses and in the elastic case at high thicknesses.

These considerations are at the origin of the most widely used criterion for calculating h_c for a given couple of materials, which consists in comparing the total energies of a given heterostructure in the coherent and plastically relaxed states as a function of thickness and in defining h_c as that thickness at which the energy in the former becomes larger than in the latter. This is an *equilibrium* criterion, equivalent to finding the thickness at which the misfit-induced force acting on a preexisting dislocation tends to pull and extend it into the S/L interface [26–28]. Other criteria are of a kinetic nature and deal with the nucleation of the misfit dislocations or with their motion toward the interface. The critical thickness decreases rapidly when the S/L mismatch increases [26]. In practice, for typical semiconductor materials, $h_c \sim 10$ nm for $\varepsilon_0 = 1\%$ and $h_c \sim 1$ nm for $\varepsilon_0 = 4\%$.

1.5
Morphological Relaxation of a Solid under Nonhydrostatic Stress

1.5.1
Introduction

Consider a *homogeneous* half-space $z \leq 0$, subjected to a uniform biaxial stress (exponent q) $\sigma^q_{xx} = \sigma^q_{yy} = \sigma_0$ in plane xy and with a *planar* traction-free surface $z = 0$ ($\sigma^q_{iz} = 0,\ i = x, y, z$). Assuming, for the sake of simplicity, that the medium is elastically isotropic, its response (relaxation) to this stress has been calculated in another context in Section 1.4.2:[14] it is a uniform tetragonal strain $\varepsilon^q_{xx} = \varepsilon^q_{yy} = -\varepsilon_0$, $\varepsilon^q_{zz} = 2\nu\varepsilon_0/(1-\nu)$, with $\varepsilon_0 = -(1-\nu)\sigma_0/E$; the nondiagonal terms are zero (Figure 1.5a).

It has been known for a few decades that if one abandons the arbitrary constraint that the free surface be planar (Figure 1.5b), the system can relax even more and adopt a different state, with lower elastic energy [29–31].[15),16)] Why such a *morphological* (planar → nonplanar) transformation might reduce the elastic energy is easily understood: at the crests of the surface, the system may deform not only in direction z but also laterally Figure 1.5c), since there is no matter to prevent it from doing so [31]. To confirm the decrease of the total energy, we must however also examine the

14) Whether the stress is applied by a rigid substrate or externally is irrelevant.

15) In Section 1.5.1, we assume that the system remains chemically homogeneous. The coupling between morphological and compositional instabilities is treated in Section 1.5.5.

16) Note that we consider here a true *redistribution of matter* along the surface with respect to the planar state and not a simple elastic deformation of the planar surface.

strain in the valleys and allow for the fact that the area of the corrugated surface is larger than that of the planar one.

1.5.2
Calculation of the Elastic Relaxation Fields

In the present case, *elastic relaxation* stands for the modification of the strain fields accompanying the planar → nonplanar transformation. Since the latter implies an actual change of shape following a redistribution of matter, Eshelby's method (which deals with the change of intrinsic state of a given volume) is not adapted. Instead, we directly solve the elastic problem, the boundary condition being that the corrugated surface remains traction-free. Since this problem has no exact solution for an arbitrary surface profile, we study the elastic response of the system to an elementary perturbation (in the spirit of the linear stability analysis, Section 1.2.1), namely, a sinusoidal modulation (hereafter, *undulation*) along x of the position h of the surface along z, $h(x) = \Delta \sin kx$, measured with respect to the planar state $h = 0$ (Figure 1.5b).[17),18)] Let us look for the fields ε_{ij}^{s}, σ_{ij}^{s} that have to be added to the "q" fields to obtain the total (equilibrium) field. Given the symmetry of the problem, no quantity depends on y, hence $\varepsilon_{iy}^{s} = (1/2)\partial u_{y}^{s}/\partial x_i$. For the same reason, $u_{y}^{s} = 0$, so $\varepsilon_{iy}^{s} = 0$ for $i = x, y, z$. The solution of such a *plane strain* problem is known to derive from an *Airy function* $\chi(x, z)$, solution of differential equation $\partial^4\chi/\partial x^4 + 2\partial^4\chi/\partial x^2\partial z^2 + \partial^4\chi/\partial z^4 = 0$, via relations $\sigma_{xx}^{s} = \partial^2\chi/\partial z^2$, $\sigma_{xz}^{s} = -\partial^2\chi/\partial x\partial z$, and $\sigma_{zz}^{s} = \partial^2\chi/\partial x^2$ [33]. Setting $\chi(x, z) = \varphi(z)\sin kx$, one finds that φ must satisfy differential equation $\mathrm{d}^4\varphi/\mathrm{d}z^4 - 2k^2\mathrm{d}^2\varphi/\mathrm{d}z^2 + k^4\varphi = 0$, the general solution of which is $\varphi(z) = (A + Bz)\mathrm{e}^{kz} + (C + Dz)\mathrm{e}^{-kz}$, with A, B, C, and D constants. Since the fields must remain finite for $z \to -\infty$, one has $C = D = 0$. Finally,

$$\begin{aligned} \sigma_{xx}^{s} &= k(2B + kA + kBz)\mathrm{e}^{kz}\sin kx, \quad \sigma_{zz}^{s} = -k^2(A + Bz)\mathrm{e}^{kz}\sin kx \\ \sigma_{xz}^{s} &= -k(B + kA + kBz)\mathrm{e}^{kz}\cos kx, \quad \sigma_{yy}^{s} = 2k\nu B\mathrm{e}^{kz}\sin kx \end{aligned} \tag{1.5}$$

since $\varepsilon_{yy}^{s} = 0$ and hence $\sigma_{yy}^{s} = \nu\left(\sigma_{xx}^{s} + \sigma_{zz}^{s}\right)$. A shear strain thus appears in plane xz. The undulated surface with normal $(n_x, 0, n_z)$ remains traction-free under *total* field $\sigma_{ij} = \sigma_{ij}^{q} + \sigma_{ij}^{s}$. Hence,

$$\left[\sigma_{xx}^{s} - \frac{E}{1-\nu}\varepsilon_0\right]n_x + \sigma_{xz}^{s}n_z = 0, \qquad \sigma_{xz}^{s}n_x + \sigma_{zz}^{s}n_z = 0 \tag{1.6}$$

the stress being calculated in $z = 0$. We now assume that the amplitude Δ of the undulation is small compared to its wavelength $2\pi/k$ (i.e., $k\Delta \ll 1$), and compute the fields at first order in $k\Delta$ (even in the case of a sinusoidal perturbation, there is no exact solution at finite amplitude). Then, $n_x = -k\Delta\cos kx$, $n_z = 1$. From Eqs (1.6)

17) Even in this case, the problem is not linear, due to the boundary conditions. Indeed, the cancellation of the tractions applies to a different surface for each (Δ, k) couple. The solution for an arbitrary h profile is not found by summing the solutions for each Fourier component of h.

18) One may also consider localized wavelet surface perturbations [32].

and (1.5), one gets $A = 0,\ B = E(1-\nu)\varepsilon_0\Delta$. By using Eq. (1.5), one obtains the stresses, from which the relaxation strains derive via the appropriate relations (see footnote 9). The nonzero components of this strain field are as the following:

$$\varepsilon^{\mathrm{s}}_{xx} = \frac{1+\nu}{E} kB[2(1-\nu)+kz]\mathrm{e}^{kz}\sin kx \tag{1.7}$$

$$\varepsilon^{\mathrm{s}}_{zz} = -\frac{1+\nu}{E} kB[2\nu+kz]\mathrm{e}^{kz}\sin kx \tag{1.8}$$

$$\varepsilon^{\mathrm{s}}_{xz} = -\frac{1+\nu}{E} kB[1+kz]\mathrm{e}^{kz}\cos kx \tag{1.9}$$

1.5.3 ATG Instability

The elastic energy W of the solid with undulated surface, *per unit area* of planar surface (reference state), is easily calculated from the total fields: $W = (1/2)\lambda^{-1} \int_0^\lambda \mathrm{d}x \int_{-\infty}^{h(x)} \left(\sigma^{\mathrm{q}}_{ij} + \sigma^{\mathrm{s}}_{ij}\right)\left(\varepsilon^{\mathrm{q}}_{ij} + \varepsilon^{\mathrm{s}}_{ij}\right) \mathrm{d}z$. It follows that the *variation* of elastic energy per unit area when the system transforms from planar to undulated state is

$$\begin{aligned} \delta W &= \frac{1}{2\lambda}\int_0^\lambda \mathrm{d}x \int_{-\infty}^{h(x)} \left(\sigma^{\mathrm{q}}_{ij}\varepsilon^{\mathrm{s}}_{ij} + \sigma^{\mathrm{s}}_{ij}\varepsilon^{\mathrm{q}}_{ij} + \sigma^{\mathrm{s}}_{ij}\varepsilon^{\mathrm{s}}_{ij}\right)\mathrm{d}z \\ &\approx \frac{1}{2\lambda}\left\{\int_0^\lambda \mathrm{d}x \int_{-\infty}^{0} \sigma^{\mathrm{s}}_{ij}\varepsilon^{\mathrm{s}}_{ij}\mathrm{d}z + \int_0^\lambda \mathrm{d}x \int_0^{h(x)} \left(\sigma^{\mathrm{q}}_{ij}\varepsilon^{\mathrm{s}}_{ij} + \sigma^{\mathrm{s}}_{ij}\varepsilon^{\mathrm{q}}_{ij}\right)\mathrm{d}z\right\} \end{aligned} \tag{1.10}$$

the second equality being valid at lowest (second) order in $k\Delta$. Replacing in Eq. (1.10) the "s" fields by their expressions calculated from Eqs (1.7)–(1.9) via the linear elasticity formulas, we get

$$\delta W = -\frac{1}{2}\frac{1+\nu}{1-\nu}\frac{1}{k} E\varepsilon_0^2 (k\Delta)^2 \tag{1.11}$$

Since $\delta W < 0$, *any surface undulation reduces the elastic energy of the system.*[19] This fundamental result specifies the driving force for the instability.

However, an undulation also increases the effective area of the free surface and induces a surface strain. Here, we only consider the first effect, which translates into an excess energy per unit reference area, equal to $\gamma\delta A/A$, where γ is the surface free energy (assumed to be independent of the slight orientation changes of the surface) and $\delta A/A$ is the relative variation of area. The latter equals $\lambda^{-1}\int_0^\lambda \sqrt{1+(\mathrm{d}h/\mathrm{d}x)^2}\mathrm{d}x$,

19) This is not obvious. If indeed, with respect to the planar state, the elastic energy is reduced at the crests, it increases in the valleys (Figure 1.5c). It is only because the former contribute more than the latter that the global balance is favorable.

that is, $k^2\Delta^2/4$ at order 2 in $k\Delta$. The *total* energy variation per unit area due to the undulation is thus

$$\delta W' = \left(\frac{\gamma}{4} - \frac{1+\nu}{1-\nu}\frac{E}{2}\frac{1}{k}\varepsilon_0^2\right)(k\Delta)^2 \tag{1.12}$$

The planar surface is unstable with respect to an undulation if $\delta W' < 0$. This is equivalent to the undulation wavenumber $k = |\mathbf{k}|$ being less than the critical value $k_c = 2(1+\nu)(1-\nu)\sigma_0^2/(E\gamma)$. In other words, there exists a critical wavelength $\lambda_c = 2\pi/k_c$, such that the planar free surface of the biaxially stressed half-space is unstable with respect to any undulation with wavelength larger than λ_c, with

$$\lambda_c = \pi\frac{1}{1-\nu^2}\frac{\gamma E}{\sigma_0^2} \tag{1.13}$$

The existence of a critical wavelength is due to the fact that the undulation-induced fields (which reduce elastic energy) penetrate the solid over a depth of the order of λ (Eqs (1.5) and (1.7)–(1.9)), whereas the excess energy due to the increased area is independent of λ. Equation (1.12) indicates that an undulation with a given wavelength is all the more easy to create that σ_0 is high and all the more difficult that surface energy is high.

This analysis can be generalized. The point is that the stress must be nonhydrostatic. The instability of a solid subjected to such stresses with respect to morphological perturbations of its surface is often called ATG (Asaro–Tiller–Grinfel'd) [29, 30]. It has been observed (with millimetric wavelengths) at the surface of ^{4}He crystals under uniaxial stress [34]. However, the ATG instability is particularly important for epitaxy. One indeed attributes to it the often-observed formation of undulations at the surface or at the interfaces of semiconducting layers mismatched with respect to their substrates [35, 36]. Indeed, as seen in Section 1.4, if the mismatch is ε_0, the semi-infinite substrate exerts biaxial strain $\sigma_0 = -[E/(1-\nu)]\varepsilon_0$.[20] Hence, the free surface of the layer is unstable against surface undulations with wavelengths larger than the critical value given by Eq. (1.13). For typical strains on the order of a percent, this wavelength is only on the order of a few tens to a few hundreds of nanometers. This is a mode of strain relaxation that differs from the usual tetragonal distortion of uniformly thick layers (Section 1.4) by its morphology and its elastic fields and also from their plastic relaxation, since relaxation remains elastic (no extended defect appears). However, at the atomic scale, the surface undulation of a low-index surface corresponds to the modulation of the spacing of preexisting or newly created surface steps. One may indeed recover the instability by considering, for instance, a vicinal surface, slightly misoriented with respect to a high-symmetry orientation, and hence composed of facets separated by steps. It has been shown that under nonhydrostatic stress, the steps interact attractively and thus tend to accumulate in bunches [37].

20) In the half-space case, the relaxation fields are attenuated over a distance of the order of λ perpendicularly to the interface. Hence, the two problems become similar if $h \gtrsim \lambda$, so that the top surface and the interface are elastically decoupled.

1.5.4
Kinetics of the ATG Instability

Here also kinetics matter, since they condition the actual formation of the undulation. Srolovitz has treated in a simple fashion two mechanisms whereby a planar surface may undulate, namely, surface diffusion and evaporation/condensation [31]. In both cases, one obtains a mode that develops more rapidly than the others, with a wavelength on the order of λ_c (as in the case of spinodal decomposition; see Section 1.3.2). This explains simply why if Eq. (1.12) leads to a semi-infinite band $[\lambda_c, +\infty]$ of unstable wavelengths, the experiments show a rather well-defined wavelength. In the case of epitaxy, one must also take into account the influx of matter from the fluid phase (molecular beams, gas or liquid) in addition to the transport of matter along the surface. This has been done by Spencer *et al.* [38] who, in addition to substrate rigidity, identify two kinetic factors that tend to inhibit the formation of the undulation, namely, a low temperature (which reduces surface diffusion) and a high growth rate (which buries the undulations before they can develop).

1.5.5
Coupling between the Morphological and Compositional Instabilities

Let us consider a half-space $z \leq 0$ of a regular solution alloy (A, B) with $\Omega > 0$ (Section 1.2.1). We know that if its free surface remains planar, it is unstable for $T \leq T_c^I$ against composition modulations with arbitrary wavelengths, if gradient energy is ignored (Section 1.2.4). We also know that if it remains homogeneous, its planar surface is unstable against undulations with wave vectors $k \leq k_c$, at any T (Section 1.5.3). In both cases, elastic relaxation is of primary importance: it determines T_c^I and it is the driving force for undulation. In this section, we consider briefly how a possible coupling between the two instabilities affects their respective domains of existence.[21] To answer this question, we calculate the elastic relaxation of a layer $z \leq 0$ of average mismatch ε_0 with a composition modulation along a direction parallel to the substrate/layer interface, but allowing a z-dependent amplitude; without loss of generality, we then write the modulation $\eta^{-1}\varepsilon_c\psi(kz)e^{kz}$ with $\psi(0) = 1$, so that its stress free strain with respect to the substrate is $\varepsilon_{ij}^*(x,z) = \delta_{ij}\left[\varepsilon_0 + \varepsilon_c\psi(kz)e^{kz}\sin kx\right]$. Assume that its free surface is undulated: $z = \Delta \sin kx$. As in the case of the purely morphological perturbation (Section 1.5.2), the elastic relaxation fields cannot be computed exactly; we limit ourselves to the first order in both $k\Delta$ at ε_c [24]. The excess of elastic energy (per unit area) of the undulated/modulated state with respect to the planar/homogeneous state is then quadratic in these two variables:

$$\delta W = \frac{1+\nu}{1-\nu}\frac{E}{2}\frac{1}{k}\left[-\varepsilon_0^2(k\Delta)^2 - 2J_1\varepsilon_0(k\Delta)\varepsilon_c + \left(\frac{J_2}{1+\nu} - J_1^2\right)\varepsilon_c^2\right] \tag{1.14}$$

21) Such a coupling has been observed in mismatched epitaxial alloy layers.

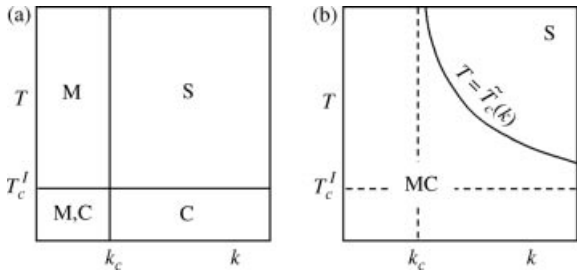

Figure 1.6 Half-space under biaxial stress. (a) Instability domains with respect to separate lateral composition modulation (C) or surface undulation (M), in plane (wave vector, temperature) (S: stability). (b) Domain of joint morphocompositional instability (MC).

where $J_n = \int_{-\infty}^{0} [\psi(u)]^n e^{2u} du$. In Eq. (1.14), the terms in $(k\Delta)^2$ and ε_c^2 correspond respectively to a pure undulation (Eq. (1.12)) and to a composition modulation in a planar half-space [24]. Whatever the sign of product $J_1\varepsilon_0$, the cross-product becomes negative with an appropriate choice of the sign of Δ (the phase of the modulation). *The elastic energy of the mixed perturbation is then less than the sum of the two perturbations taken separately.* If we assume that $\psi(kz) > 0$ and that ε_0 and ε_c have the same sign, then for $x \equiv \lambda/4 (\mathrm{mod}\, \lambda)$, the layer is more mismatched than on average, and it is interesting to have a larger relaxation there, hence also a crest ($\Delta > 0$). The coupling vanishes if the layer is on average lattice matched to its substrate [15]. To obtain the total excess of free energy, we add to δW the excess free enthalpy of mixing and the excess surface energy, as done respectively in Sections 1.2.3 and 1.5.3 [24]. The conclusion is that not only the instability domains but also the very nature of the instability are modified. We must consider together the two parameters that in the case of uncoupled disturbances (Figure 1.6a) have critical values separating stable and unstable domains, namely, T and k for the compositional (C) and morphological (M) instabilities, respectively. With coupling (Figure 1.6b), one finds in the (k, T) plane an extended domain of *morphocompositional* (MC) instability. To each wavenumber k corresponds a critical temperature

$$\tilde{T}_c(k) = T_c^C + \frac{1}{4}\frac{1+\nu}{1-\nu}\frac{E'}{k_B}\eta^2\frac{k}{k-k_c} \tag{1.15}$$

where T_c^C is the bulk critical temperature (Section 1.2.3).

As usual, this thermodynamical analysis must be completed by a kinetic analysis that will decide if the instability will actually develop, depending on the matter transport mechanisms available [39–41].

1.6
Elastic Relaxation of 0D and 1D Epitaxial Nanostructures

We have seen in Section 1.4 that one way to prevent the formation of dislocations during the growth of a mismatched epitaxial layer on a substrate is to keep the

layer thickness below its critical value for plastic relaxation. This becomes impractical at high lattice mismatch ε_0, since the critical thickness decreases rapidly when ε_0 increases (Section 1.4.3). In such cases, one may play on the dimensionality and dimensions of the deposit (and sometimes of the substrate) to prevent or hinder dislocation formation. Indeed, when the constraint of infinite lateral extension is lifted, the deposit may recover its intrinsic (stress-free) state even if the interface remains coherent. The lattice planes may then deform continuously from the spacing of the substrate toward the intrinsic spacing of the deposit over some distance from the interface. This is realized in quantum dots and nanowires.

1.6.1
Quantum Dots

Section 1.5 indicates that a mismatched epitaxial layer may reduce its total energy by developing a surface undulation. The same driving force leads to the nucleation of coherent islands (instead of a uniformly thick 2D layer) of a strongly mismatched epitaxial deposit in the Volmer–Weber (VW) or Stranski–Krastanov (SK) growth modes.[22] Observed for more than 25 years [42], the SK growth of semiconductors spurred the spectacular development of quantum dot nanostructures.[23] Although their origin is the same, the descriptions of the ATG and SK (or VW) instabilities differ in that the ATG instability refers to the development of a smoothly varying surface corrugation (typically realized by step bunching on a singular surface), whereas islands are bound by facets with orientations different from that of the substrate [43]. It is precisely these additional lateral free surfaces that allow an efficient strain relaxation. A similar but weaker effect occurs for quantum wires (1D nanostructures) parallel to the substrate.

Elastic relaxation, which affects both island and substrate, may be calculated by the general methods introduced in Section 1.1, the transformed part of the system being the island. If analytical solutions exist for the island buried in a half-space (which is indeed relevant since, in practice, most quantum dots are capped after growth) (see Section 1.1.3 and Figure 1.2), this is not the case for uncapped islands. One then has to resort to numerical methods, such as the finite elements method [44, 45] or the atomistic valence force field method [46]. In addition, approximate analytical solutions have been proposed for the strains or the energy [43, 47–51]. Elastic relaxation of the substrate is also important. Indeed, for an island under compression, the substrate is dilated under the island and compressed in its vicinity. Hence, when a first island has nucleated, it is unfavorable to nucleate another one close to it. This is at the origin of the tendency of the islands to *self-organize*.

22) These modes differ by the absence (VW) or presence (SK) of a thin wetting layer on the substrate.

23) All dimensions of the islands are nanometric, hence the often-used name *0D nanostructures*.

1.6.2 Nanowires

The same kind of effect occurs in nanowires. Epitaxial NWs growing in a direction perpendicular to their substrate are nowadays fabricated from a large range of elemental and compound semiconductor materials [52]. Similarly to quantum dots, NWs have lateral dimensions ranging from a few nanometers to a few tens of nanometers. At variance with quantum dots, which have inclined facets and nanometric heights, freestanding NWs may extend perpendicularly to the substrate with a uniform diameter and over potentially unlimited lengths (in practice, several microns are easily reached). These 1D nanostructures have remarkable physical properties and many potential applications. For a misfitting layer at the top of a NW, lateral relaxation should be even easier than for a quantum dot, since the effective substrate has the same finite diameter instead of being laterally infinite. The strain relaxation for such relaxation has indeed been calculated [53]. As expected, it is very efficient. For instance, for the same misfitting S/L couple, in the NW, the elastic energy is already only a quarter of its 2D value (for the same volume of layer) when the layer thickness is only 10% of the NW diameter. As a consequence, the critical layer thickness, which is now radius dependent, may be much higher than the 2D value for small, but accessible, NW diameters. Moreover, there exists a critical NW radius under which the critical thickness becomes infinite. Both quantities have been calculated [53]. Finally, a similar effect, albeit somewhat less efficient, operates when a NW is grown on a mismatched substrate. Hence, NWs are nanostructures that allow one to associate with strongly mismatched materials in a way that would not be feasible in the 2D (or even the 0D) geometry.

References

1 Eshelby, J.D. (1957) The determination of the elastic field of an ellipsoidal inclusion, and related problems. *Proc. R. Soc. Lond. A*, **241**, 376–396.
2 Mura, T. (1991) *Micromechanics of Defects in Solids*, Kluwer, Dordrecht.
3 Mindlin, R.D. (1936) Force at a point in the interior of a semi-infinite solid. *Physics*, **7**, 195–202.
4 Mindlin, R.D. (1950) Nuclei of strain in the semi-infinite solid. *J. Appl. Phys.*, **21**, 926–930.
5 Pan, E. (2003) Three-dimensional Green's functions in an anisotropic half-space with general boundary conditions. *J. Appl. Mech.*, **70**, 101–110.
6 Faivre, G. (1964) Déformations de cohérence d'un précipité quadratique. *Phys. Status Solidi*, **35**, 249–259.
7 Chiu, Y.P. (1978) On the stress filed and surface deformation in a half space with a cuboidal zone in which initial strains are uniform. *J. Appl. Mech.*, **45**, 302–306.
8 Glas, F. (1987) Coherent stress relaxation in a half space: modulated layers, inclusions, steps, and a general solution. *J. Appl. Phys.*, **70**, 3556–3571.
9 Pearson, G.S. and Faux, D.A. (2000) Analytical solutions for strain in pyramidal quantum dots. *J. Appl. Phys.*, **88**, 730–736.
10 Glas, F. (2001) Elastic relaxation of truncated pyramidal quantum dots and quantum wires in a half space: an analytical calculation. *J. Appl. Phys.*, **90**, 3232–3241.

11 Khachaturyan, A.G. (1983) *Theory of Structural Transformations in Solids*, John Wiley & Sons, Inc., New York.

12 Cahn, J.W. (1961) On spinodal decomposition. *Acta Metall.*, **9**, 795–801.

13 Cahn, J.W. (1962) On spinodal decomposition in cubic crystals. *Acta Metall.*, **10**, 179–183.

14 Cahn, J.W. (1968) Spinodal decomposition. *Trans Metall. Soc. AIME*, **242**, 166–180.

15 Glas, F. (2000) Thermodynamic and kinetic instabilities of lattice-matched alloy layers: compositional and morphological perturbations. *Phys. Rev. B*, **62**, 7393–7401.

16 Ipatova, I.P., Malyshkin, V.G., and Shchukin, V.A. (1993) On spinodal decomposition in elastically anisotropic epitaxial films of III–V semiconductor alloys. *J. Appl. Phys.*, **74**, 7198–7210.

17 Cheng, K.Y., Hsieh, K.C., and Baillargeon, J.N. (1992) Formation of lateral quantum wells in vertical short-period superlattices by strain-induced lateral-layer ordering process. *Appl. Phys. Lett.*, **60**, 2892–2894.

18 Glas, F., Treacy, M.M.J., Quillec, M., and Launois, H. (1982) Interface spinodal decomposition in LPE $In_xGa_{1-x}As_xP_{1-y}$ lattice matched to InP. *J. Phys. (Paris)*, **43**, C5-11–C5-16.

19 Philibert, J. (1991) *Atom Movements Diffusion and Mass Transport in Solids*, EDP Sciences, Les Ulis, France.

20 Larché, F.C. and Cahn, J.W. (1982) The effect of self-stress on diffusion in solids. *Acta Metall.*, **30**, 1835–1845.

21 Larché, F.C. and Voorhees, P.W. (1996) Diffusion and stress: basic thermodynamics. *Defect Diffus. Forum*, **129–130**, 31–36.

22 Philibert, J. (1996) Diffusion and stresses. *Defect Diffus. Forum*, **129–130**, 3–8.

23 Wu, C.H. (1996) The chemical potential for stress-driven surface diffusion. *J. Mech. Phys. Solids*, **44**, 2059–2077.

24 Glas, F. (1997) Thermodynamics of a stressed alloy with a free surface: coupling between the morphological and compositional instabilities. *Phys. Rev. B*, **55**, 11277–11286.

25 Hirth, J.P. and Lothe, J. (1982) *Theory of Dislocations*, John Wiley & Sons, Inc., New York.

26 Fitzgerald, E.A. (1991) Dislocations in strained-layer epitaxy: theory, experiment, and applications. *Mater. Sci. Rep.*, **7**, 87–142.

27 Matthews, J.W. and Blakeslee, A.E. (1974) Defects in epitaxial multilayers. 1. Misfit dislocations. *J. Cryst. Growth*, **27**, 118–125.

28 People, R. and Bean, J.C. (1985) Calculation of critical layer thickness versus lattice mismatch for Ge_xSi_{1-x}/Si strained-layer heterostructures. *Appl. Phys. Lett.*, **47**, 322–324.

29 Asaro, R.J. and Tiller, W.A. (1972) Interface morphology development during stress corrosion cracking: Part I. Via surface diffusion. *Metall. Trans.*, **3**, 1789–1796.

30 Grinfel'd, M.A. (1986) Instability of the separation boundary between a non-hydrostatically stressed elastic body and a melt. *Sov. Phys. Dokl.*, **31**, 831–834.

31 Srolovitz, D.J. (1989) On the stability of surfaces of stressed solids. *Acta Metall.*, **37**, 621–625.

32 Colin, J., Grilhé, J., and Junqua, N. (1997) Localized surface instability of a non-homogeneously stressed solid. *Europhys. Lett.*, **38**, 307–312.

33 Love, A.E.H. (1944) *A Treatise on the Mathematical Theory of Elasticity*, 4th edn, Dover, New York.

34 Thiel, M., Willibald, A., Evers, P., Levchenko, A., Leiderer, P., and Balibar, S. (1992) Stress-induced melting and surface instability of ^{4}He crystals. *Europhys. Lett.*, **20**, 707–713.

35 Cullis, A.G., Robbins, D.J., Pidduck, A.J., and Smith, P.W. (1992) The characteristics of strain-modulated surface undulations formed upon epitaxial $Si_{1-x}Ge_x$ alloy layers on Si. *J. Cryst. Growth*, **123**, 333–343.

36 Ponchet, A., Rocher, A., Emery, J.-Y., Starck, G., and Goldstein, L. (1993) Lateral modulations in zero-net strained GaInAsP multilayers grown by gas source molecular-beam epitaxy. *J. Appl. Phys.*, **74**, 3778–3782.

37 Tersoff, J. (1995) Step bunching instability of vicinal surfaces under stress. *Phys. Rev. Lett.*, **75**, 2730–2733.

38 Spencer, B.J., Voorhees, P.W., and Davis, S.H. (1992) Morphological instability in epitaxially strained dislocation-free solid films: linear stability theory. *J. Appl. Phys.*, **73**, 4955–4970.

39 Huang, Z.-F. and Desai, R.C. (2001) Instability and decomposition on the surface of strained alloy films. *Phys. Rev. B*, **65**, 195421.

40 Huang, Z.-F. and Desai, R.C. (2002) Epitaxial growth in dislocation-free strained alloy films: morphological and compositional instabilities. *Phys. Rev. B*, **65**, 205419.

41 Spencer, B.J., Voorhees, P.W., and Tersoff, J. (2001) Morphological instability theory for strained alloy film growth: the effect of compositional stresses and species-dependent surface mobilities on ripple formation during epitaxial film deposition. *Phys. Rev. B*, **64**, 235318.

42 Goldstein, L., Glas, F., Marzin, J.-Y., Charasse, M.N., and Le Roux, G. (1985) Growth by molecular beam epitaxy and characterization of InAs/GaAs strained-layer superlattices. *Appl. Phys. Lett.*, **47**, 1099–1101.

43 Müller, P. and Saúl, A. (2004) Elastic effects on surface physics. *Surf. Sci. Rep.*, **54**, 157–258.

44 Christiansen, S., Albrecht, M., Strunk, H.P., and Maier, H.J. (1994) Strained state of Ge(Si) islands on Si: finite element calculations and comparison to convergent beam electron-diffraction measurements. *Appl. Phys. Lett.*, **64**, 3617–3619.

45 Ponchet, A., Lacombe, D., Durand, L., Alquier, D., and Cardonna, J.-M. (1998) Elastic energy of strained islands: contribution of the substrate as a function of the island aspect ratio and inter-island distance. *Appl. Phys. Lett.*, **72**, 2984–2986.

46 Lin, Y.-Y. and Singh, J. (2002) Self-assembled quantum dots: a study of strain energy and intersubband transitions. *J. Appl. Phys.*, **92**, 6205–6210.

47 Daruka, I. and Barabási, A.-L. (1997) Dislocation-free island formation in heteroepitaxial growth: a study at equilibrium. *Phys. Rev. Lett.*, **79**, 3708–3711.

48 Gippius, N.A. and Tikhodeev, S.G. (1999) Inhomogeneous strains in semiconducting nanostructures. *J. Exp. Theor. Phys.*, **88**, 1045–1049.

49 Tersoff, J. and Tromp, R.M. (1993) Shape transition in growth of strained islands: spontaneous formation of quantum wires. *Phys. Rev. Lett.*, **70**, 2782–2785.

50 Wang, Y.W. (2000) Self-organization, shape transition, and stability of epitaxially strained islands. *Phys. Rev. B*, **61**, 10388–10392.

51 Zinovyev, V.A., Vastola, G., Montalenti, F., and Miglio, L. (2006) Accurate and analytical strain mapping at the surface of Ge/Si(001) islands by an improved flat-island approximation. *Surf. Sci.*, **600**, 4777–4784.

52 Dick, K.A. (2008) A review of nanowire growth promoted by alloys and non-alloying elements with emphasis on Au-assisted III–V nanowires. *Prog. Cryst. Growth Ch.*, **54**, 138–173.

53 Glas, F. (2006) Critical dimensions for the plastic relaxation of strained axial heterostructures in free-standing nanowires. *Phys. Rev. B*, **74**, 121302.