CHAPTER 22

Structural Effects in Coherent Epitaxial Semiconductor Films

ALEX ZUNGER

National Renewable Energy Laboratory, Golden, CO 80401, USA

Contents

HANDBOOK OF CRYSTAL GROWTH, VOL. 3
edited by D.T.J. Hurle

CHAPTER 22

Structural Effects in Coherent Epitaxial Semiconductor Films

ALEX ZUNGER

National Renewable Energy Laboratory, Golden, CO 80401, USA

Contents

HANDBOOK OF CRYSTAL GROWTH, VOL. 3
edited by D.T.J. Hurle

1. Introduction

This chapter reviews the experimental evidence and the theoretical explanations for coherent epitaxy-induced structural changes in films relative to "free-standing" bulk solids. I will not consider growth morphology or growth dynamics. Instead I will focus on the way that coherence with the substrate can lead to

- appearance of *compound* crystal structures that are either unstable or metastable in bulk form,
- stabilization of *alloy solid solutions* that have limited solubility in bulk form,
- pinning of the alloy composition, and
- new electronic properties.

I will focus on physical explanations rather than on detailed mathematical descriptions which are given in the original papers cited. I will further not limit myself to the explanation of experimentally tested systems, but extend the arguments and provide predictions for experimentally unexplored cases. This article explores the thermodynamic consequences of coherent epitaxy rather than kinetic limitations in attaining them. It is aimed at highlighting the fact that modern epitaxial growth techniques provide not only better purity and control of the growing film than "bulk" growth methods, but that epitaxy can also result in *fundamentally different material structures and thus new material properties.* Recent review articles on related subjects include Bruinsma and Zangwill [1], Zunger and Wood [2], and Bauer et al. [3].

Our discussion will consider three forms of solids produced by different growth geometries:

(i) *"Bulk" compounds or alloys.* These are thick, "free-standing" (i.e. without a substrate) solids. They are often grown from the melt. We use such bulk structures as a *reference* against which epitaxial effects will be measured.

(ii) *"Surfaceless" epitaxial compounds or alloys.* These are thin films, grown coherently on a given substrate such that a *free* surface does not exist during growth (e.g. by liquid phase epitaxy).

(iii) *Epitaxial compounds or alloys grown with a free surface.* These are thin films, grown coherently on a substrate such that a *free* surface is exposed during growth, e.g. by VPE, OMVPE, and MBE.

We will show that the expected properties of these three forms of solids can be profoundly different: in (i) the structure that grows reflects largely the consequences of bulk thermodynamics. In (ii) the structure that is produced reflects the thermodynamics of *strained* three-dimensional (3D) solids. In (iii) the structure produced manifests the combined consequences of coherent epitaxy and *surface* effects. Our main results can be summarized as follows:

(a) When a film is epitaxially coherent with a substrate and surface effects during growth are not too significant (class (ii) above), epitaxy can lead to (i) stabilization of new compound crystal structures which are unstable in the bulk (e.g. NaCl in the zinc-blende structure, CdTe in the β-Sn structure), (ii) stabilization of "ordered alloys" (e.g. chalcopyrite-like Ga_2AsSb), (iii) large reductions in the miscibility gap tempera-

ture (e.g. epitaxial solubility of GaP with GaSb, which are bulk insoluble), and (iv) "pinning" of the alloys composition at a value close to lattice matching.

(b) Growth methods that lead to epitaxial coherence with the substrate *and* expose a free surface of the growing film (class (iii) above) can lead, in addition, to spontaneous stabilization of monolayer (1 1 1) superlattices (the "CuPt" structure) even if the alloy is grown homogeneously.

(c) The structural effects indicated in (a) and (b) above can lead to profound changes in the electronic structure of the film, including altering an indirect band gap bulk material to a direct band gap epitaxial film, as well as changing a "type-I" band offset to a "type-II" band offset.

I will consistently use firstly continuum elasticity ideas to explain these phenomena and then expose interesting physical aspects that require a *microscopic theory* for their explanation.

2. *Epitaxial stabilization of bulk-unstable structures: pure compounds*

2.1. *Observations of epitaxial stabilization in pure compounds*

Coherent epitaxy can stabilize crystal structures that are unstable in bulk form under otherwise equivalent conditions. In this section I describe such effects in *pure compounds*, while section 3 describes epitaxial stabilization effects in *alloys*. I start by giving some experimental examples that appear to reflect epitaxial stabilization of novel structures in pure compounds.

(1) Tin (Sn) appears in bulk form either as the *low-temperature* ($T \leqslant 13.2°C$) "grey-tin" cubic semiconductor ($a = 6.489$ Å) or as the *room-temperature* tetragonal "white-tin" metal ($a = 5.831$ Å, $c = 3.181$ Å). When grown epitaxially on the (0 0 1) surface of InSb or CdTe ($a = 6.49$ Å), tin adopts its low-temperature cubic structure *even at room temperature* [4–7], thus avoiding the strain that would have resulted from fitting the tetragonal phase onto this substrate.

(2) Cobalt (Co) is stable at ambient conditions in the bulk hcp structure. When grown epitaxially on a GaAs(1 1 0) substrate, it is stabilized in the bcc structure up to thicknesses of ≈ 1000 Å, after which it reverts to its bulk-stable hcp phase [8]. The measured in-plane lattice parameter of bcc Co ($a = 2.819$ Å) matches to within 0.4% the projected lattice constant of the GaAs(1 1 0) substrate. Similarly, when grown on Ni(0 0 1) or Cu(0 0 1), cobalt crystalizes in the fcc structure [9, 10]. Note that bcc Co does not exist in bulk form, although it is predicted [11] to be a metastable phase, 70 meV higher in energy than hcp cobalt.

(3) Iron (Fe) is stable in bulk form in the bcc structure, yet when grown epitaxially on Cu(1 1 1), Cu(1 0 0), or Cu(1 1 0) it appears in the fcc structure [12–16].

(4) Silver (Ag) is stable in bulk form in the fcc structure, yet when grown on InSb (1 1 0) it crystallizes in the bcc structure [17].

(5) Cu and Ni are stable in bulk in the fcc form, yet when grown on Fe(0 0 1) they appear in the bcc form [18, 19].

(6) MnSe is stable in bulk in the NaCl structure, yet when grown on ZnSe it is

stabilized in the zinc-blende structure [20]. Similarly, MnTe exists in bulk in the NiAs structure, but when grown on ZnTe it is stabilized in the zinc-blende structure [21a]. This stabilization was predicted theoretically [21b].

(7) InSb, which is stable in bulk in the zinc-blende structure, can be grown by sputtering on glass in the NaCl structure [22].

(8) Cesium (Cs) and thallium (Tl) halides, normally stable in the CsCl structure can be grown on mica and alkali halide substrates in the NaCl structure [23].

(9) MgS and MgSe can be grown on alkali halide substrates in the wurtzite form [24]. Total-energy calculations [25] suggest that other structures are more stable in bulk form.

(10) $FeSi_2$ crystallizes epitaxially on an Si substrate in the bulk-unstable CaF_2 structure [26]. At composition 1:1 it appears in the CsCl structure. The bulk-stable phases are β-$FeSi_2$ (orthorombic) and ε-FeSi (the CP8 structure).

(11) "Nb_3Nb" grows in the A 15 structure on an Si substrate despite the fact that this structure is bulk-unstable [27].

(12) An unusual rhombohedral structure of SiGe has been grown on an Si(0 0 1) substrate [28] despite the fact that there is no stable Si–Ge compound in bulk form.

We do not know at present which of the examples cited above represent the thermodynamically stable phase of the *film + substrate system*, as opposed to representing structure at a a local energy minimum surrounded by kinetically insurmountable activation barriers. In fact, some of the examples cited above may not have been stabilized through lattice coherence with the substrate. Future annealing experiments both with the substrate present and after the substrate has been removed (by lift-off or etching) should clarify this important issue. However, in many cases (e.g. the stabilization of bcc metals and α-Sn in particular) one observes a clear correlation between the substrate lattice constant and that of the pseudomorphic epitaxial phase, suggesting true epitaxial stabilization.

I next describe the generic physics of epitaxially strained *pure compounds* (section 2.2) and then (section 2.3) illustrate how these ideas, when combined with first-principles total-energy calculations, predict epitaxial stabilization of bulk-unstable structures.

2.2. *Continuum elasticity description of epitaxial stabilization in pure compounds*

We begin our discussion of the origin of epitaxial effects by considering a continuum elasticity description of bulk and epitaxially constrained single-phased cubic crystals. Our qualitative description in this section illustrates the main effects by considering *cubic* systems and retaining only *harmonic* terms in the energy. Quantitative total-energy calculations using self-consistent first-principles approaches (described in section 2.3) are, however, not restricted in this manner.

Consider a cubic crystal (e.g. zinc-blende) in a free-standing bulk form. Its total energy E (per fcc site), as a function of its cubic lattice parameter a, can be expanded about the equilibrium value a_{eq} as

$$E^{bulk}(a) = E_{eq} + \tfrac{9}{8} B a_{eq} [a - a_{eq}]^2 + \cdots, \tag{1}$$

where E_{eq} is the equilibrium total energy at $a = a_{eq}$ and where

$$B = \tfrac{1}{3}(C_{11} + 2C_{12}) \tag{2}$$

is the cubic bulk modulus. A film of this material grown epitaxially on a thick substrate will, if thinner than the critical thickness h_c for nucleation of misfit dislocations, strain so that its atoms grow in registry with those of the substrate. Thus its dimensions $a_\parallel$ parallel to the substrate are geometrically determined by those of the substrate, labeled a_s in fig. 1a. We may formally denote this elastic constraint by the condition $a_\parallel = a_s$. Within harmonic elasticity theory the strain energy density is

$$U = \tfrac{1}{2}C_{11}(e_{xx}^2 + e_{yy}^2 + e_{zz}^2) + \tfrac{1}{2}C_{44}(e_{yz}^2 + e_{zx}^2 + e_{xy}^2) + C_{12}(e_{yy}e_{zz} + e_{zz}e_{xx} + e_{xx}e_{yy}), \tag{3}$$

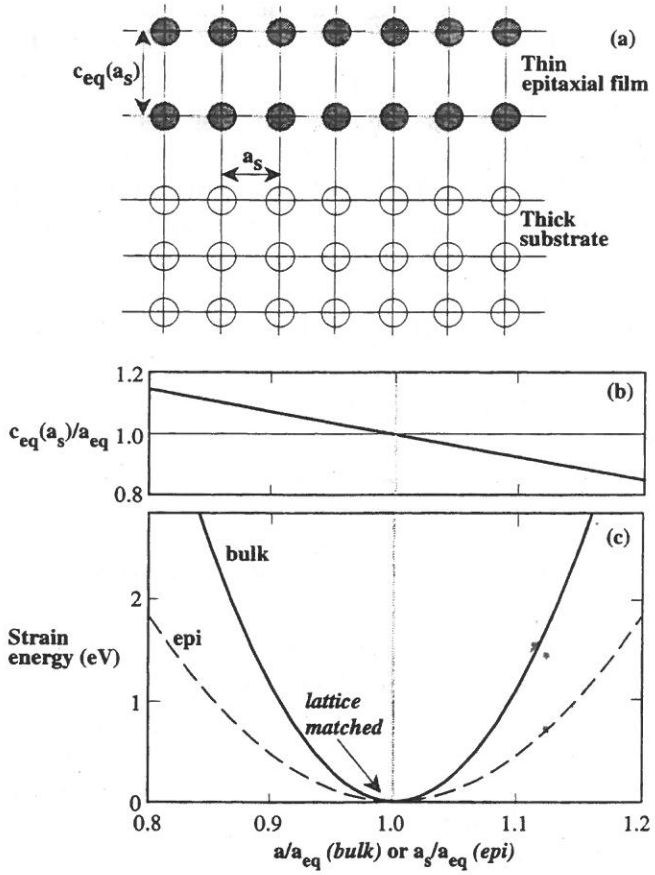


Fig. 1. Schematic illustration of structural and energetic effects of coherent epitaxy. Panel (a) depicts registry of atoms of epitaxial film with those of a substrate having lattice dimensions a_s; (b) shows dependence of induced perpendicular distortion $C_{eq}(a_s, \hat{G})$ of the film on a_s (eq. (7)). Panel (c) contrasts the energy of a hydrostatically compressed bulk (eq. (1)) with that of a coherent epitaxial material $E^{epi}(a_s, \hat{G})$ (eq. (14)). From Wood [33].

where the C_{ij} are the conventional cubic elastic constants. To lowest order, the strain energy per atom may be written

$$U_{\text{strain}} = UV_{\text{eq}}, \tag{4}$$

where V_{eq} is the volume per atom of the unconstrained equilibrium material. For simplicity we consider only situations for which we have no shear strains (e_{ij} with $i \neq j$). Then,

$$e_{\parallel} \equiv e_{xx} = e_{yy} = \frac{a_s - a_{\text{eq}}}{a_{\text{eq}}},$$

$$e_{\perp} \equiv e_{zz} = \frac{c - a_{\text{eq}}}{a_{\text{eq}}}, \tag{5}$$

where the perpendicular dimension of the film is labelled c in fig. 1a. For fixed substrate a_s, the equilibrium value of c is found from the condition

$$\left(\frac{\partial U}{\partial e_{\perp}}\right)_{a_s} = 0, \tag{6}$$

which yields

$$c_{\text{eq}}(a_s, \hat{G}) = a_{\text{eq}} - [2 - 3q(\hat{G})](a_s - a_{\text{eq}}). \tag{7}$$

Here, $\hat{G}$ is the direction of deformation and the "epitaxial strain reduction factor" $q(\hat{G})$ is given by* [2, 29]

$$q(\hat{G}) = 1 - \frac{B}{C_{11} + \gamma(\hat{G})\Delta}, \tag{8}$$

where B is the bulk modulus and

$$\Delta = C_{44} - \tfrac{1}{2}(C_{11} - C_{12}) \tag{9}$$

is the "elastic anisotropy". Here, γ is a purely geometric factor given by [30]

$$\gamma(\hat{G}) \equiv \gamma(\phi, \theta) = \sin^2(2\theta) + \sin^4(\theta)\sin^2(2\phi), \tag{10}$$

where ϕ and θ are spherical polar coordinates defined by

$$\boldsymbol{r} = [r\sin(\theta)\cos(\phi); r\sin(\theta)\sin(\phi); r\cos(\theta)]. \tag{11}$$

For the principal directions we have

$$\gamma[0\,0\,1] = 0, \quad \gamma[0\,1\,1] = 1, \quad \gamma[1\,1\,1] = \tfrac{4}{3}. \tag{12}$$

A parametric plot of $\gamma(\phi, \theta)$ is presented in fig. 2 (from [30]). Note that the minimum value of γ lies along the [0 0 1] direction and the maximum along the [1 1 1] direction.

* I consistently avoid using the Poisson ratio, preferring the *general* form of eqs. (7)–(11), which exposes the analytical dependence on orientation.

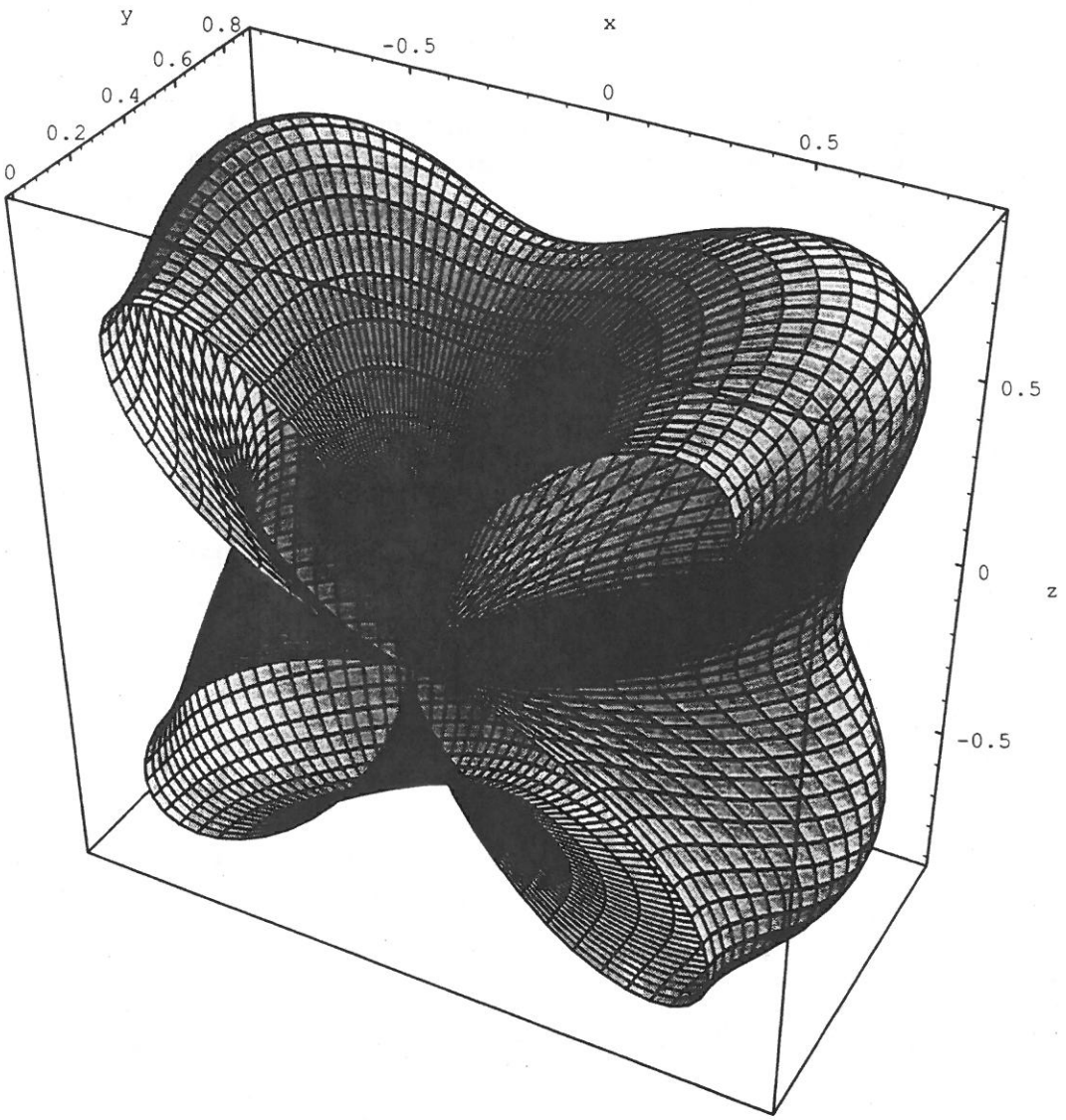


Fig. 2. Parametric plot of $\gamma(\phi, \theta)$ of eq. (10) over half of the unit sphere. The surface shown is defined by the spherical polar coordinates $[\gamma(\phi, \theta), \phi, \theta]$. The function $\gamma(\phi, \theta)$ has a minimum value of 0 along the six $\langle 0\,0\,1 \rangle$ directions and a maximum value of $\frac{4}{3}$ along the eight $\langle 1\,1\,1 \rangle$ directions. From Laks et al. [30].

Explicit expressions for $q(\hat{G})$ along principal directions [2] are

$$q[1\,0\,0] = \frac{2}{3}\left(1 - \frac{C_{12}}{C_{11}}\right),$$

$$q[1\,1\,0] = \frac{1}{3}\frac{C_{11} - C_{12} + 6C_{44}}{C_{11} + C_{12} + 2C_{44}},$$

$$q[1\,1\,1] = \frac{4C_{44}}{C_{11} + 2C_{12} + 4C_{44}}. \tag{13}$$

We see from eq. (7) that on a mismatched substrate the epitaxial film will distort along $\hat{G}$ such that when $a_s < a_{eq}$ then $c/a_{eq} > 1$, while when $a_s > a_{eq}$ then $c/a_{eq} < 1$. This is illustrated in fig. 1b. Note that, in the continuum elasticity description, the film is assumed to be deformed *homogeneously*, so atomic planes near the film/substrate interface deform equally as atomic planes far from the interface. We will see below (section 4.3) that this is not always true *even* if complete coherence with the substrate is maintained. In such cases, we will abandon the conventional *continuum* description in favor of a *microscopic* (i.e. atomistic) model.

The distortion along $\hat{G}$ lowers the energy of the epitaxial film relative to the energy $E^{bulk}(a = a_s)$ of the bulk deformed hydrostatically to the same a_s (eq. (1)). This can be seen by substituting eq. (7) into eq. (3). The strain energy for coherent epitaxy per fcc cubic site is then

$$E^{epi}(a_s, \hat{G}) = E_{eq} + \tfrac{9}{8} q(\hat{G}) B a_{eq} [a_s - a_{eq}]^2 + \frac{1}{8} \frac{B}{1 - q(\hat{G})} a_{eq} [c - c_{eq}(a_s, \hat{G})]^2. \tag{14}$$

If the film is distorted along $\hat{G}$ to *minimize* its strain energy, then $c = c_{eq}$ and the final term in eq. (14) disappears (in strained-layer systems, however, it might persist). Comparing the epitaxial energy of eq. (14) with the energy of a hydrostatically deformed bulk (eq. (1)) we see that

$$q(\hat{G}) = \frac{E^{epi}(a_s, \hat{G}) - E_{eq}}{E^{bulk}(a_s) - E_{eq}}. \tag{15}$$

Hence, the epitaxial reduction factor of eq. (8) measures the relative energy lowering of an epitaxial film deformed in *two dimensions* to a_s, with respect to a bulk film deformed in *three dimensions* to the same a_s. In the harmonic approximation, $q(\hat{G})$ does not depend on a_s. More general theories [30] show some dependence. Inspection of eqs. (8)–(13) and fig. 2 shows that $q(\hat{G}) \leqslant 1$. In fact, for most semiconductors the elastic anisotropy of eq. (9) obeys $\Delta > 0$ (see [31]) so the epitaxial reduction factor $q(\hat{G})$ of eq. (8) is determined by the geometrical factor $\gamma(\hat{G})$ (eq. (12)). In this case q is the smallest along [0 0 1] (softest deformation direction) while it is largest along [1 1 1] (hardest deformation direction). Using tabulated elasticconstants [31] we have

$$q[0\,0\,1] = 0.21, \quad q[1\,1\,0] = 0.35, \quad q[1\,1\,1] = 0.39 \quad \text{(CdTe)}$$

and

$$q[0\,0\,1] = 0.36, \quad q[1\,1\,0] = 0.48, \quad q[1\,1\,1] = 0.51 \quad \text{(GaAs and GaSb)}.$$

By contrast, in many alkali halides $\Delta < 0$ [31], so q is the largest along [0 0 1].

Figure 1c illustrates eq. (14), showing how the epitaxial energy is reduced relative to the hydrostatic bulk energy. *The central point here is that the epitaxial energy curve is flatter (i.e. "softer") than the bulk curve*, since q, appearing in the second term of eq. (14) but not in eq. (1), is smaller than unity. When $a_{eq} = a_s$ the two curves coincide. Note that while only volumes smaller than the equilibrium value are accessible when pressure is applied to bulk materials, "two-dimensional" negative (epitaxial) pressures are readily applied to a coherent epitaxial film by choice of a substrate with $a_s > a_{eq}$. This could lead to epitaxial stabilization of structures never found in bulk for pressures $P \geqslant 0$.

For $a_s \neq a_{eq}$, the epitaxial film is strained, i.e. $E_{epi}(a_s \neq a_{eq}) > E_{eq}$ (fig. 1c). The energy of such a film could be lowered towards E_{eq} by the nucleation of misfit dislocations. This energy lowering is described within simple continuum elasticity by Matthews [32]. It modifies eq. (14) to finite thickness (for $c = c_{eq}$) giving

$$E^{epi}(a_s, \hat{G}, h) = E_{eq} + \tfrac{9}{8} q(\hat{G})\, B\, a_{eq}\, [a_s - a_{eq}]^2\, G(h), \tag{16}$$

where

$$G(h) = 1, \quad \text{when } h < h_c,$$
$$G(h) = \chi(2 - \chi), \quad \text{when } h \geqslant h_c, \tag{17}$$

and

$$\chi(h) = \frac{h_c}{h}\left[1 + \ln\left(\frac{h}{b}\right)\right] \Big/ \left[1 + \ln\left(\frac{h_c}{b}\right)\right],$$

where b is the Burger's vector of the relevant misfit dislocation. Figure 3 illustrates schematically how $E^{epi}(a_s \neq a_{eq})$ lies *above* E_{eq} but *below* $E^{bulk}(a_s)$ (part (b)), and how $E^{epi}(a_s \neq a_{eq})$ approaches E_{eq} as the film thickness h increases (part (c)). The film is said to be "coherently strained" (e.g. in registry with the substrate) for h below the critical thickness h_c for nucleation of misfit dislocations. In practice, activation barriers against nucleation of such dislocations permit, for semiconductors, coherent growth considerably beyond the thermodynamic value of h_c; such films are said to be "metastably strained".

To see how epitaxy could stabilize a bulk-unstable phase, consider (fig. 4a) a compound whose lowest-energy state in bulk corresponds to a structure β, while structure α, which we wish to stabilize, has a higher energy in bulk. Suppose now that we grow this material on a substrate whose lattice constant is a_s. According to eq. (14)

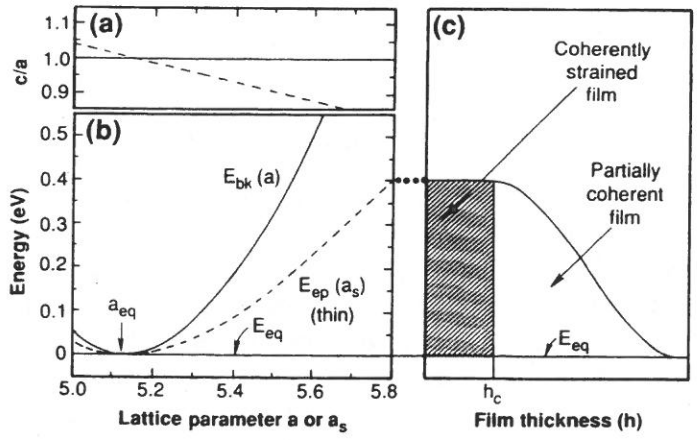


Fig. 3. Schematic plot showing how the epitaxial energy relaxes to the equilibrium bulk value, as a function of film thickness h. Part (a) shows the c/a ratio, part (b) shows the bulk and epitaxial energies for a thin ($h \ll h_c$) film, and part (c) shows the epitaxial energy versus h. From Zunger and Wood [2].

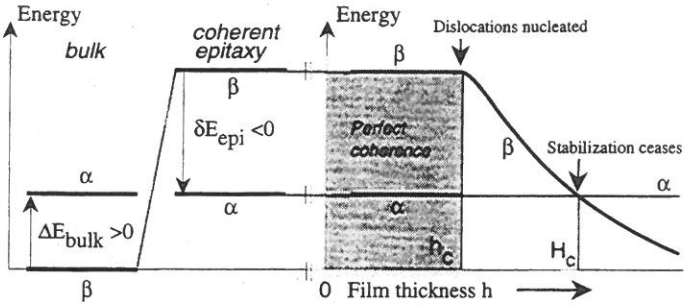


Fig. 4. Schematic relative energies of two competing phases α and β under bulk and coherent epitaxial conditions, (a) on a substrate lattice-matched to α (phase α is the more stable, although β is stable in bulk). Panel (b) shows the reduction in strain energy of phase β as strain-relieving mechanisms occur with increasing film thickness. From Wood [33].

the *excess* epitaxial energies are

$$\Delta E_\alpha^{\text{epi}} = E_\alpha^{\text{epi}}(a_s) - E_{\alpha,\text{eq}} = \tfrac{9}{8} q_\alpha B_\alpha a_\alpha [a_s - a_\alpha]^2,$$
$$\Delta E_\beta^{\text{epi}} = E_\beta^{\text{epi}}(a_s) - E_{\beta,\text{eq}} = \tfrac{9}{8} q_\beta B_\beta a_\beta [a_s - a_\beta]^2. \tag{18}$$

The potential for epitaxial stabilization of the bulk-unstable phase α is apparent by considering the most favorable situation, i.e. when the desired phase (α) is lattice-matched to the substrate ($a_s = a_\alpha$). Then $\Delta E_\alpha^{\text{epi}} = 0$ and the *epitaxial* energy difference between α and β equals

$$\delta E^{\text{epi}}(a_s) \equiv E_\alpha^{\text{epi}}(a_s) - E_\beta^{\text{epi}}(a_s) = \Delta E^{\text{bulk}} - \Delta E_\beta^{\text{epi}}, \tag{19a}$$

where

$$\Delta E^{\text{bulk}} \equiv E_{\alpha,\text{eq}} - E_{\beta,\text{eq}} \tag{19b}$$

is the α–β energy difference in *bulk*. Here, by assumption, β is more stable than α in bulk, so $\Delta E^{\text{bulk}} > 0$. Yet, eq. (18) shows that $\Delta E_\beta^{\text{epi}} > 0$, so $-\Delta E_\beta^{\text{epi}} < 0$. Hence, the relative energy δE^{epi} *could become negative*. This is shown schematically in the center part of fig. 4a.

Equations (18) and (19) clarify which physical conditions could lead to epitaxial stabilization of the bulk-unstable phase α. This is the case if (i) we select a substrate such that a_β is sufficiently different from a_s (so that $-\Delta E_\beta^{\text{epi}}$ is sufficiently negative), (ii) phase β is sufficiently stiff (large B_β) and has a high q_β-value, and if (iii) the initial energy difference ΔE_{bulk} between the bulk phases α and β is sufficiently small. In this case epitaxy could select α over β, while the reverse is true in bulk.

The analysis above assumes that the film is thinner than the critical thickness, labeled h_c in fig. 3 and fig. 4. The manner in which the film's strain energy drops as the thickness increases beyond h_c and misfit dislocations relieve strain is schematically shown in fig. 4b. Beyond a *second* critical thickness H_c [2, 33], where the energy curves of the two thermodynamic competitors α and β cross, the absolute stabilization of α over β ceases.

On a substrate with $a_s = a_\alpha$, lattice-matched to α,

$$h_c \propto \frac{a_\beta^2}{|a_s - a_\beta|}, \tag{20}$$

$$H_c \propto \frac{q_\beta B_\beta a_\beta^3 |a_s - a_\beta|}{E_{\beta,eq} - E_{\alpha,eq}}. \tag{21}$$

We see that a large misfit $|a_s - a_\beta|$ with the substrate *limits* the critical thickness h_c for misfit dislocations in phase β but *increases* the critical thickness H_c for stability of the competing phase α.

Our foregoing discussion illustrates the principles of epitaxial stabilization of bulk-unstable compounds. It clarifies how such effects depend on the "material parameters" $(q_\alpha, B_\alpha, E_{eq,\alpha})$ and $(q_\beta, B_\beta, a_\beta, E_{eq,\beta})$ of the stable (β) and unstable (α) phases. While the material parameters of *bulk-stable phases* can be measured, those of *bulk-unstable phases* are generally unknown. I describe therefore in the next section how these can be predicted from first-principles self-consistent total-energy calculations, and how combining these results with the ideas described in this section leads to quantitative predictions of epitaxially stabilized phases.

2.3. *Microscopic theory of epitaxial stabilization in pure compounds*

In this section I describe the energetic basis for epitaxial selection of a bulk-unstable phase in *pure* compounds [34]. I will use two generic examples (i) a compound that is bulk-stable in the zinc-blende structure (CdTe) and is predicted to be epitaxially stabilized in the NaCl structure, and (ii) compounds that are bulk stable in the NaCl structure (MgS, NaCl) but could be epitaxially stabilized in the zinc-blende structure.

I use here the total energies of *extended periodic solids* to characterize epitaxial phases and transitions. These are calculated from the local density theory [35] in which the Schrödinger equation of a solid (bulk or epitaxial) is solved self-consistently without any empirical input. Unlike continuum elasticity, where the properties of stressed solids are described in terms of the properties of the unstressed solids, our microscopic (local-density) theory treats each structure in its own right. For example, the total energy of a given compound (CdTe) in one crystal structure (NaCl) is not derived from information on the other crystal structure (zinc-blende) but calculated from scratch. Thus, different patterns of charge transfer and other electronic effects that may distinguish various crystal structures of the same compound are treated explicitly by solving the appropriate Schrödinger equation. Details of the method are given in [35]. For small deformations about equilibrium, we expect that such microscopic descriptions will reproduce the predictions of continuum elasticity.

Before describing such results, it may be appropriate to compare the predictions of continuum elasticity and those of the microscopic description for a simple case, e.g. a uniformly stressed pure compound.

Brandt et al. [36a] have recently described a high-resolution electron microscopic analysis of buried (0 0 1) layers of InAs in GaAs that appears to reveal a breakdown of

continuum elasticity theory (eq. (7)) in the limit of monatomic films. In particular, they find that for a *monatomic* film of InAs in GaAs the relative tetragonal deformation $c_{eq}(a_s, \hat{G})/a_{eq}$ was 12.46%. By contrast, eq. (7) with $q[0\,0\,1]$ given by eq. (13) predicts

$$\frac{c_{eq}(a_s, 0\,0\,1)}{a_{eq}} - 1 = -2\frac{C_{12}}{C_{11}}\left(\frac{a_s - a_{eq}}{a_s}\right) = 7.27\%.$$

The numerical value is obtained using the measured C_{12} and C_{11} of GaAs [31], the substrate lattice constant $a_s = a(\text{GaAs}) = 5.65$ Å, and the equilibrium lattice constant $a(\text{InAs}) = 6.06$ Å. Careful self-consistent total-energy calculations by Bernard and Zunger [36b] yield a relative deformation of 7.38%, in excellent agreement with continuum elasticity *but in conflict with experiment*. Bernard and Zunger conclude therefore that continuum elasticity does not break down even in the limit of ultrathin layers. The source of discrepancy with the experiment of Brandt et al. (interdiffusion in their sample, error in data analysis?) is, however, not known. In section 4.3 we will describe another example where continuum elasticity does fail. In almost all cases, however, we find that for semiconductors continuum elasticity produces quantitative agreement with the microscopically calculated c/a ratios.

We now proceed to discuss the microscopic origins of epitaxial structural transformations using first CdTe [34a] as an example. Figure 5a shows the calculated equation of state of CdTe versus applied external hydrostatic pressure; as pressure is increased CdTe undergoes two phase transformations: (i) B3 (= zinc-blende) to B1 (= NaCl) and (ii) B1 to A5 (β-Sn). Volume integration of the pressure data gives the energy as a function of volume for each phase (again under hydrostatic conditions). The offsets between energy minima of different phases (i.e. $E_{\alpha,eq} - E_{\beta,eq}$ of eq. (19b)) are obtained from the transformation pressures through the common tangent constructions shown in fig. 5b. As noted before, when a solid is constrained by coherent growth on a substrate, parallel to the interface the overlayer takes the lattice dimensions of the substrate. The appropriate external parameter for the energy is no longer pressure or volume, but, as eq. (14) indicates, the substrate lattice parameter(s), a_s. This entails two changes: *first*, changing from a volume representation V to an a_s representation involves structural geometric factors relating a_s to the cell volume in different phases. With $a_s = \lambda(V)^{1/3}$, we find

$$\lambda_{B1} = A\,2^{1/6} = \lambda_{B3},$$
$$\lambda_{A5} = A\,(2/\eta)^{1/3}, \tag{22}$$

where $\eta(V)$ is the c/a ratio under hydrostatic conditions and A is a geometric scaling factor. This would be used, e.g., to describe the $A = \sqrt{2}$ scaling when changing from a simple to a face-centered square lattice. *Second*, the unconstrained c-axis, perpendicular to the interface, is free to relax. As indicated by eq. (14) this reduces the energy of the epitaxial phase relative to the hydrostatically compressed phase evaluated at the same substrate lattice constant. These two steps are illustrated for CdTe in fig. 5c. The solid lines show the energy as a function of a_s before c-axis relaxation ($q \equiv 1$) and the dashed lines give the final epitaxial energy (q given by eq. (8) or (13)). The epitaxial energies differ from the hydrostatic energies in fig. 5b by (i) reduced elastic energies, causing a

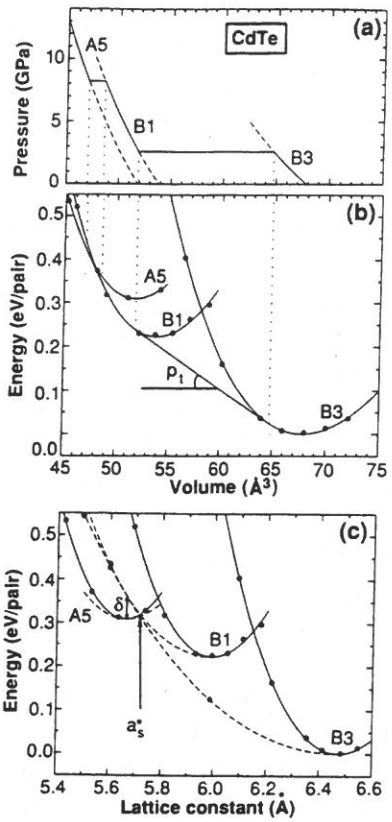


Fig. 5 Calculated structural energies for CdTe in the structures indicated. (a) and (b) are given under hydrostatic-pressure conditions, while (c) shows the energy as a function of (substrate) lattice parameter before (solid lines) and after (dashed line) c-axis relaxation. The epitaxial relaxation energy for A5 is estimated. We used $A = \sqrt{2}$ for B1 and B3 and $A = 1$ for the A5 phase (eq. (22)). From Froyen et al. [34a].

flattening of the curves (fig. 1c); this increases the region of stability for the lowest energy phase, and (ii) horizontal shifts of the energy minima (eq. (22)). This horizontal shift separates the two minima that coincided in the bulk, thus exposing an inaccessible phase, e.g. bcc ($\lambda = 1.26$) and fcc ($\lambda = 1.12$). These effects can result in qualitative changes in phase stability. *Under hydrostatic pressure CdTe has two transitions: B3 to B1 and B1 to A5. Epitaxially, these are predicted to be replaced by a single transition: B3 to A5. Relaxation ($q < 1$) of B3 leads to a situation where B1 is never a stable epitaxial phase.* This shows how epitaxy (i) stabilizes at zero pressure a bulk-unstable phase (A5) and (ii) eliminates a phase (B1) that occurs in bulk at high pressures.

The existence of a competing phase with a lower energy minimum (B3) leads to a situation where nucleation of misfit dislocations can limit the thickness of the pseudomorphic phase (A5) even when the latter is lattice-matched to the substrate (see fig. 4). Selecting a substrate lattice-matched to the A5 phase makes its energy independent of thickness. The competing B3 phase is higher in energy by δ (fig. 5c). However, the B3 film experiences large misfit and, once its critical thickness has been exceeded, it can lower its energy by nucleating misfit dislocations. The energy of the B3 phase will

therefore decrease with increasing thickness (see fig. 4b), at some point H_c dropping below the A5 phase. Then, B3 (with misfit dislocations) becomes the ground state, defining a new critical thickness H_c (fig. 4) for the A5 phase. Note that, contrary to the situation in regular strained-layer epitaxy [32], the critical thickness H_c *increases* with the lattice mismatch (eq. (21)). This is because it is the *competing* phase (B3) which experiences the misfit.

Based on this analysis, we can now identify the physical factors that could lead to epitaxial stabilization (maximum δ) of bulk-unstable compounds:

(i) The difference in energy between the two competing bulk phases (eq. (19b)) should be as small as possible. This is often the case in transition metals (fcc versus bcc or hcp versus bcc) but less so for semiconductors.

(ii) The difference between the equilibrium lattice constants of the competing phases should be as large as possible (eq. (20)), so that H_c is maximized.

(iii) The lowest of the two bulk phases should be as stiff as possible, i.e. have a large bulk modulus and a large q (eqs. (18)–(20)).

Based on these considerations, Froyen et al. [34a] selected three compounds where a pseudomorphic phase might be stabilized through epitaxy: CdTe (normally B3), MgS, and NaCl (normally B1). Other candidates are AgBr, AgCl, CaS, CaSe, and MgSe (normally B1) and AgI, CdSe, CuI, HgS, HgSe, and HgTe (normally B3). Figures 6 and 7 show results for MgS and NaCl, respectively, corresponding to fig. 5c for CdTe. Also shown are the relaxed c/a ratios. Figure 7 shows that for NaCl the substrate lattice parameter for epitaxial B1 to B3 crossover, $a_s^* = 6.5$ Å, lies to the right of the undistorted B3 minimum, which therefore cannot be reached. However, it may be possible to achieve a distorted B3 phase, with $c/a < 1$, by choosing $a_s > 6.5$ Å. Figure 6 shows that the B3 phase of MgS is accessible: the epitaxial energies cross at

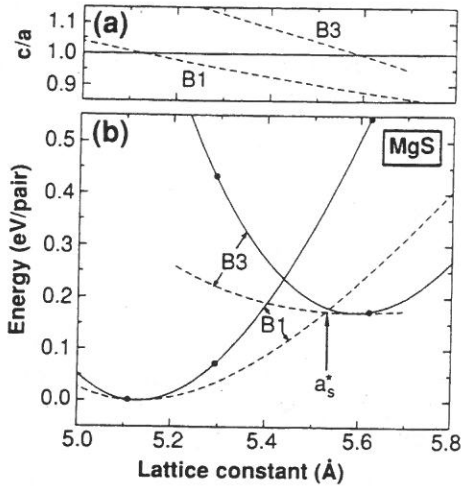


Fig. 6. Calculated structural energy and c/a ratios for B1 and B3 MgS as a function of (substrate) lattice parameter. Solid and dashed lines show the energy before and after c-axis relaxation, respectively. From Froyen et al. [34a].

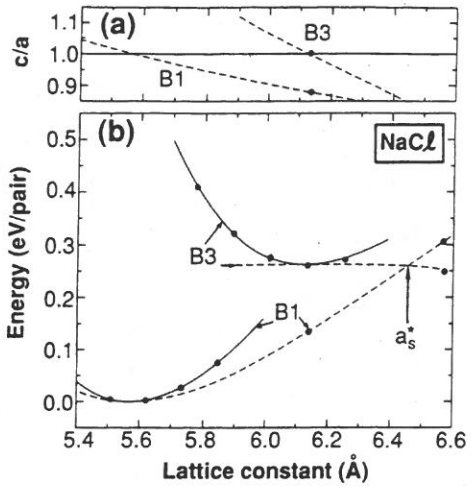


Fig. 7. Calculated structural energy and c/a ratios for B1 and B3 NaCl as a function of (substrate) lattice parameter. Solid and dashed lines show the energy before and after c-axis relaxation, respectively. From Froyen et al. [34a].

$a_s^* = 5.5$ Å. Some substrates which are nearly lattice-matched to B3 MgS are BeTe, MnS, AlAs, and GaAs (all B3). I am unaware of experimental testing of the predictions of figs. 5–7.

3. *Epitaxial stabilization in alloys*

Section 2 discussed epitaxial stabilization effects in *pure, single-phased compounds*. I now extend the discussion to binary alloys $A_{1-x}B_x$ of constituents A and B, or to the pseudobinary case $A_{1-x}B_xC$, where the constituents are AC and BC. I start this discussion by outlining some pertinent experimental observations.

3.1. *Observations of epitaxial stabilization in alloys*

A few effects are pertinent here:

(i) *Epitaxially enhanced solid solubility*: GaAs and GaSb have a rather restricted solid solubility in the bulk: the miscibility gap at 600°C extends from $x = 0.2$ to 0.8. Yet, epitaxial growth (by OMVPE) shows that the $GaAs_{1-x}Sb_x$ alloy can be grown over the entire composition range [37]. Even $GaP_{1-x}Sb_x$, whose bulk miscibility gap extends at 600°C from $x = 0.01$ to 0.99, can be grown epitaxially over a wide composition range [38]. Similarly, Farrow [7] and Farrow et al. [39] illustrated that, despite the limited bulk solubility of group-II fluorides (e.g. BaF_2, CaF_2, and SrF_2), their alloys can be grown epitaxially on suitable semiconductor substrates, e.g. $SrBaF_2$ on InP, $CaBaF_2$ on GaAs, and $CaCdF_2$ on Si. Noreika et al. [40,41] have likewise demonstrated

epitaxial solubility of InSb/BiSb (the maximum bulk solubility is only 2.2%). Finally, Sood et al. [42, 43] demonstrated epitaxial solutions of the bulk-immiscible PbS/CdS system and Asom et al. [44] demonstrated epitaxial solutions of Sn and Ge. In most of these cases detailed diffraction data is absent, so the real microstructure is not known: it could be that some phase separation does exist. However, there is little doubt that the overall solubility is much enhanced in the epitaxial case.

(ii) *Pinning of the epitaxial alloy composition near the value lattice-matched to the substrate*: Stringfellow [45], Beneking et al. [46], Quillec et al. [47], Antypas and Moon [48], and Asai and Oe [49] observed that the composition of epitaxial alloys grown by LPE is close to the lattice-matched value despite the fact that the alloy compositions of bulk samples grown under identical conditions vary widely. This has been discussed by Kume et al. [50], Asai and Oe [51], Oe et al. [52], Suzuki et al. [53], Joncour et al. [54], Larché and Cahn [55], and Wood and Zunger [56].

While the current evidence for epitaxially enhanced solubility and composition pinning is robust, it is not clear to what extent this reflects thermodynamic stabilization or kinetic limitations. Again, careful annealing experiments are much needed. As in section 2 we will first discuss the simple predictions of continuum elasticity (section 3.2) and then (section 3.3) consider a microscopic theory.

3.2. *Continuum elasticity description of epitaxial stabilization in alloys*

While the energy of a *pure* bulk compound is given by eq. (1), the energy of a bulk *alloy* is measured with respect to the equilibrium energies of equivalent amounts of its bulk constituents, i.e.

$$\Delta H^{\text{bulk}}(x) = E_{\text{eq}}[\text{A}_{1-x}\text{B}_x\text{C}] - (1-x)E_{\text{eq}}[\text{AC}] - x\,E_{\text{eq}}[\text{BC}]. \tag{23}$$

The *epitaxial* formation enthalpy is then

$$\Delta H^{\text{epi}}(a_s, \hat{G}, x) = E^{\text{epi}}[\text{A}_{1-x}\text{B}_x\text{C}, a_s, \hat{G}] - E^{\text{cs}}[\text{AC/BC}, a_s, \hat{G}], \tag{24}$$

where the "constituent strain" (cs) is the energy of the epitaxially constrained components AC and BC:

$$E^{\text{cs}}[\text{AC/BC}, a_s, \hat{G})] = (1-x)E^{\text{epi}}[\text{AC}, a_s, \hat{G}] + x\,E^{\text{epi}}[\text{BC}, a_s, \hat{G}]. \tag{25}$$

Here, E^{epi} appearing in eqs. (24), (25) is given by eq. (14). Comparing eqs. (24), (25) with eq. (23) then gives the relative energies of epitaxy versus bulk (eq. (19a)), i.e. the "substrate strain" (ss) energy

$$\delta E_{\text{epi}} \equiv \Delta H^{\text{epi}}(a_s, \hat{G}, x) - \Delta H^{\text{bulk}}(x) = \Delta E_{\text{ss}}(a_s, \hat{G}, x). \tag{26}$$

Abbreviating the coefficients in eq. (14) as

$$\begin{aligned} K(\hat{G}, x) &= \tfrac{9}{8} q(\hat{G}) B(x) a(x), \\ K_{\text{A}}(\hat{G}) &= \tfrac{9}{8} q_{\text{A}}(\hat{G}) B_{\text{A}} a_{\text{A}}, \\ K_{\text{B}}(\hat{G}) &= \tfrac{9}{8} q_{\text{B}}(\hat{G}) B_{\text{B}} a_{\text{B}}, \end{aligned} \tag{27a}$$

the epi versus bulk ("substrate strain") energy becomes

$$\Delta E_{ss}(a_s, \hat{G}, x) = \Delta H^{epi} - \Delta H^{bulk} = K(\hat{G}, x)[a_s - a(x)]^2 - (1-x)K_A(\hat{G})(a_s - a_A)^2 - xK_B(\hat{G})(a_s - a_B)^2. \quad (27b)$$

The last two terms are the *negative* of the "constituent strain" of eq. (25). The constituent strain energy of eq. (25) depends on orientation through $q(\hat{G})$. This is illustrated in fig. 8, which shows the equilibrium constituent strain energy for GaP + InP at $x = 0.5$ as a function of orientation $\hat{G}$, as calculated by Laks et al. [30]. It is small for (0 0 1) deformations and large for (1 1 1) deformations.

All epitaxial stabilization effects discussed here have their origins in the *substrate strain energy* $\Delta E_{ss}(a_s, \hat{G}, x)$. Its dependence on a_s, q, B, and x can be used to manipulate the relative stabilities of bulk and epitaxial systems. A simple illustration (fig. 4) serves to make this point. (Here we reinterpret the identity of the symbols α and β relative to section 2.2 as follows.) Consider an alloy $A_{1-x}B_xC$ denoted in fig. 4 as α and its

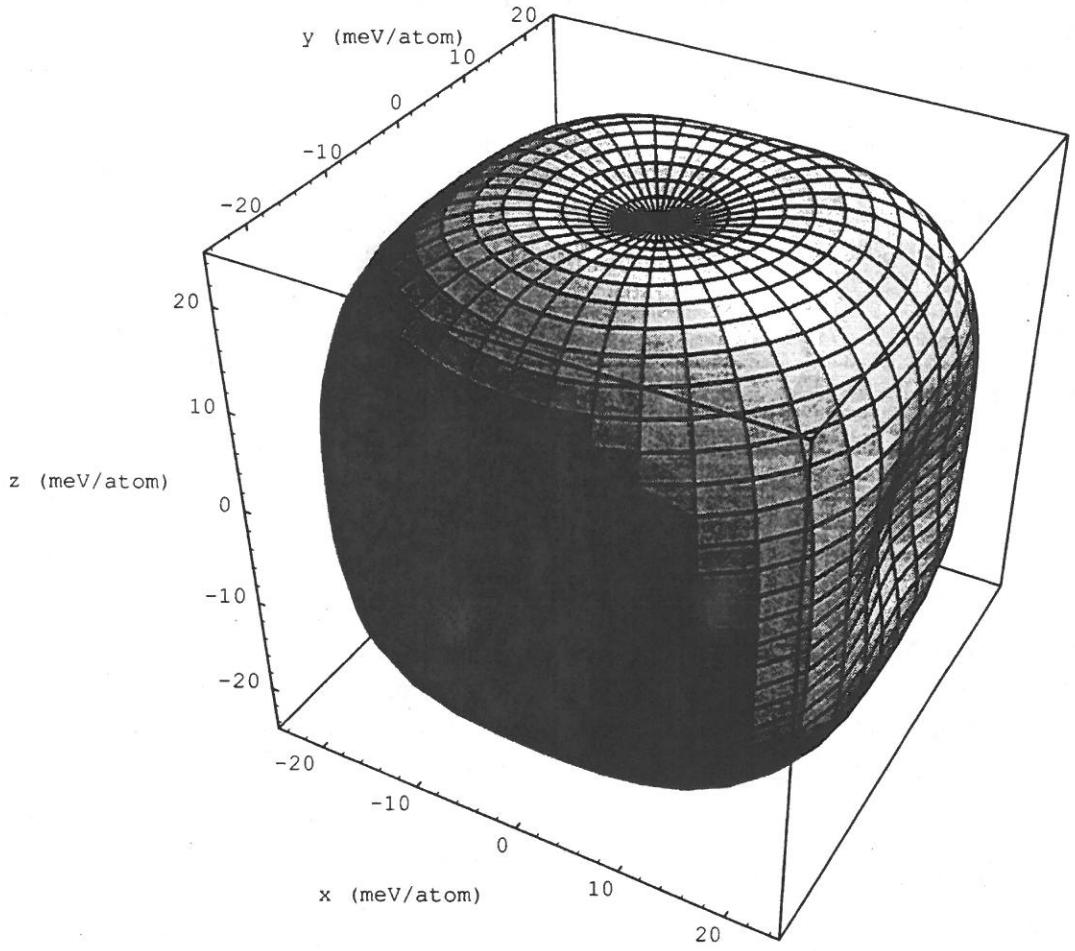


Fig. 8. Parametric plot of the equilibrium constituent strain energy (eq. (25)) for GaP/InP at $x = 0.5$ over the unit sphere. The surface shown is defined by the spherical polar coordinates (E_{cs}, ϕ, θ) in units of MeV/atom. From Laks et al. [30].

decomposition products AC + BC denoted there as β. Since for pseudobinary semiconductor alloys $\Delta H^{\text{bulk}} \geqslant 0$ [2], the homogeneous alloy α has a higher enthalpy than its disproportionation products β, as shown in fig. 4a. If we grow the alloy on a lattice-matched substrate, $a_\alpha = a_s$, then the first term of eq. (27b) vanishes and $\Delta E_{ss} = -\Delta E_{cs}$. Note that ΔE_{cs} is positive (fig. 8) since it represents the energy increase due to straining AC and BC from their equilibrium lattice constants a_A and a_B, respectively, to the substrate value a_s. Hence, $\Delta E_{ss} = -\Delta E_{cs}$ is negative, as shown in fig. 4a (where δE_{epi} corresponds now to ΔE_{ss}). We see that when the alloy is lattice-matched to the substrate

$$\Delta H^{\text{epi}}[a_s = a(x); \hat{G}] - \Delta H^{\text{bulk}}(x) < 0, \tag{28}$$

as shown in fig. 4a. The discussion of section 2.2 is pertinent here: the bulk-unstable alloy phase α would become epitaxially stabilized up to a critical thickness H_c. It is thus clear that the origin of the epitaxial stabilization is *not* the substrate-induced strain in the growing alloy film $A_{1-x}B_x \equiv \alpha$ (since, by assumption, this is lattice-matched to the substrate), but in the epitaxial destabilization of the constituents (AC + BC) $\equiv \beta$. This becomes clearer when we assume that the alloy follows Vegard's rule,

$$a(x) = a_A + x(a_B - a_A), \tag{29}$$

and that the substrate itself is an alloy with composition X_s,

$$a_s \equiv a(X_s) = a_A + X_s(a_B - a_A). \tag{30}$$

Inserting these into eq. (27b) then gives

$$\Delta E_{ss}(X_s, x, \hat{G}) = (a_B - a_A)^2[K(x)(X_s - x)^2 - (1 - X)X_s^2 K_A - x(1 - X_s)^2 K_B]. \tag{31}$$

We distinguish here and in the remainder of the text two forms of (often confused) mismatches: (a) the mismatch between the film $a(x)$ and the substrate $a(X_s)$. This vanishes at the "lattice-matched" composition: $x = X_s$. And (b) the microscopic mismatch $a_A \neq a_B$ between the atomic constituents of the film. Equation (31) shows that (i) if the *constituents* are size-matched ($a_A = a_B$) the substrate strain ΔE_{ss} vanishes, as do epitaxial stabilization effects. If the constituents are size-mismatched, then (ii) at the lattice-matched composition $x = X_s$ the substrate strain energy is non-zero: it is negative and proportional to the weighted stiffnesses of the constituents, i.e.

$$\Delta E_{ss}(X_s = x, \hat{G}) = -(a_B - a_A)^2[K_B X(1 - X)^2 + K_A(1 - X)X^2]. \tag{32}$$

To understand how coherent epitaxy affects solid solubility, note that the spinodal temperature T_c is given by the second compositional variation of the free energy [55]. The epitaxial excess enthalpy, from eq. (26), is

$$\Delta H^{\text{epi}}(a_s, \hat{G}, x) = \Delta H^{\text{bulk}}(x) + \Delta E_{ss}(a_s, \hat{G}, x), \tag{33}$$

while the *free* energy is

$$\Delta F^{\text{epi}}(a_s, \hat{G}, x, T) = [\Delta H^{\text{bulk}}(x) - T\Delta S(x)] + \Delta E_{ss}(a_s, \hat{G}, x). \tag{34}$$

The first term in square brackets in in eq. (34) is the free energy $\Delta F_{\text{bulk}}(x, T)$ of the bulk, so the zero of its second variation with respect to composition gives the *bulk*

spinodal temperature. We see that the epitaxy-induced change in the spinodal temperature is given by the compositional derivatives of the substrate strain energy ΔE_{ss}. Assuming Vegard's rule for a moment (eqs. (29), (30)), we find from eqs. (31) and (34) the epitaxial critical temperature condition

$$\Delta F''_{\text{bulk}} - 2K(x)(a_B - a_A)^2 - (x - X_s)(a_B - a_A)^2[K''(x - X_s) + 4K'] = 0. \tag{35}$$

Here the prime(s) denote compositional derivatives. To evaluate $\Delta F''_{\text{bulk}}$ we will use, for the moment, the regular solution model, whereby the mixing enthalpy of the bulk is

$$\Delta H^{\text{bulk}}(x) = \Omega x(1 - x) \tag{36a}$$

and the mixing entropy is just the ideal entropy

$$\Delta S = RT[x \ln x + (1 - x)\ln(1 - x)]. \tag{36b}$$

Here R is the gas constant and Ω is the bulk "interaction parameter". These assumptions and the assumption of Vegard's rule are rather crude approximations which will be completely avoided in our *quantitative* discussion in section 3.3. Here, we use these approximations only to provide a *qualitative* discussion of epitaxial effects on solubility. Thus, evaluating $\Delta F''_{\text{bulk}}$ under the regular solution model gives the epitaxial transition temperature:

$$\frac{RT_c}{x(1 - x)} = 2\Omega - 2(a_B - a_A)^2 K(x) - (a_B - a_A)^2(x - X_s)[K''(x)(x - X_s) + 4K'(x)]. \tag{37}$$

All but the first term vanish if the constituents have the same lattice constants ($a_A = a_B$); the last term vanishes if the alloy is lattice-matched to the substrate ($x = X_s$). The terms of eq. (37) have the following physical significance:

- The first term, yielding $T_c = \Omega/2R$ at $x = \frac{1}{2}$, is the spinodal temperature of the bulk alloy without microscopic strain (lattice-matched constituents, $a_A = a_B$) and without a substrate strain ($x = X_s$). This term alone was used in standard regular-solution bulk phase-diagram calculations by deCremoux et al. [57], Stringfellow [58], Panish and Illegems [59], and Glas [60]. It leads to high values of T_c, e.g. 910 K for GaInP and 850 K for GaAsSb [60].
- The correction of the second term corresponds to "coherent strains" [61]. It lowers T_c in proportion to the microscopic size-mismatch $(a_B - a_A)^2$ and the elastic stiffness in the direction of phase separation. This lowering of T_c exists even at the lattice-matched composition $x = X_s$. This correction was used by deCremoux [62], Stringfellow [63], and Glas [60], leading to a strong suppression of T_c, e.g. -160 K for GaInP and -250 K for GaAsSb [60]. While this predicts in many cases complete epitaxial solubility, actual experiments show that the solubility is only enhanced relative to bulk, but still some phase separation is evident. Glas [60] corrected the second term of eq. (37) by introducing transverse sinusoidally modulated lattice parameters in the substrate plane and surface effects. This reduces the elastic energy with respect to the bulk, particularly at the film/vacuum and film/substrate interfaces. This adds a multiplicative factor $m \leqslant 1$ to the second term of eq. (37). The predicted T_c is

increased to +400 K for GaInP and +330 K for GaAsSb. The effect of the third term of eq. (37) was neglected [60].

• The third term of eq. (37) represents corrections due to mismatch with the substrate, $x \neq X_s$. The quantity $(a_B - a_A)^2(x - X_s)^2 K''(x)$ is a modulus effect and does not change sign as x crosses the lattice-matched composition X_s, since it depends on $(x - X_s)^2$. When $K''(x) < 0$, the elastic energy is reduced by phase separation, so T_c *increases*. The quantity $4(a_B - a_A)^2(x - X_s)K'(x)$ is a combination of a modulus effect with coherency strains. This term changes sign when the lattice-matched composition is crossed*.

Larché et al. [64] computed T_c for $Ga_xIn_{1-x}P$ assuming (i) the regular solution theory of eq. (36), (ii) Vegard's rule of eq. (29), and (iii) linear variation of the elastic constants with composition ($K'' = 0$). On a lattice-matched substrate they find $T_c \cong 0$, 20, and 220 K for $\hat{G} = (1\,1\,1)$, $(1\,1\,0)$, and $(0\,0\,1)$, respectively. As the substrate lattice constant is reduced, T_c was found to increase.

3.3. Microscopic theory of epitaxial stabilization in alloys

Section 3.2 describes a semi-quantitative theory of epitaxial stabilization of alloys of the sort used by Glas [60], Larché et al. [64], Johnson and Chiang [65], and Chiang and Johnson [66]. While qualitatively correct, this approach makes a number of severe approximations:

(i) The enthalpy of the bulk is described by the regular solution model (eq. (36)), where ΔH^{bulk} depends only on composition but not on the atomic configurations. This model cannot distinguish, therefore, ordered from disordered solids at the same composition x. We will see below that epitaxial effects *in alloy* can stabilize *ordered phases* that are bulk unstable. The approach of Glas, Cahn, and Johnson and Chiang misses this effect.

(ii) The ideal "mean-field" mixing entropy (eq. (36)) is used rather than the correlated entropy. The entropy of a bulk alloy is set equal to that of an epitaxial alloy.

(iii) The formalism used is only appropriate for systems that phase-separate at low temperatures ($\Omega > 0$). Compound-forming alloys ($\Omega < 0$) are not treated.

(iv) The lattice parameter $a(x)$ and epitaxial reduction factor $q(x)$ are assumed to be linear with composition.

(v) The dependence of the equilibrium composition on strain (leading to "composition pinning"; see section 3.3.3) is neglected.

These approximations can be avoided by replacing continuum elasticity by a microscopic statistical treatment, as done by Mbaye et al. [67], Wood and Zunger [68], Mbaye et al. [69], Zunger and Wood [2], Bernard et al. [70], Wood [33], and Osorio et al. [73]. In what follows I describe the essential features of this approach and its application to epitaxial stabilization of alloys.

* An additional term, proportional to the *deviations* from Vegard's rule is absent from eq. (37). It is generally very small in semiconductor alloys.

3.3.1. Bulk versus epitaxial phase diagrams

Disordered pseudobinary $A_{1-x}B_xC$ alloys which result from mixing two III–V or two II–VI semiconductors are crystalline but substitutionally disordered. Each site on the A/B ("mixed") sublattice is occupied by an A or a B atom. The atoms on the common C sublattice do not contribute to alloy configurational entropy, and may be tacitly omitted from the thermodynamic description. Such binary alloys are usually treated via an Ising model (see Wei et al. [72], Lu et al. [29], and Laks et al. [30]) in which one defines a "pseudospin" variable $S_i = -1$ for an A atom and $S_i = 1$ for a B atom at lattice site i. With two different atom types possible at each site, for a lattice of N sites there will be 2^N distinct ways of arranging atoms; a direct calculation of the thermodynamics for macroscopic N is obviously impractical. Instead, we may hope to make progress by identifying the important atom–atom (spin–spin) interactions in the alloy. For a particular arrangement (configuration) σ of the N spins, we can certainly express the excess energy of the alloy as

$$\Delta E[\{S_i\}_\sigma] = \sum_{ij}{}' J_{ij}\langle S_iS_j\rangle_\sigma + \sum_{ijk}{}'' J_{ijk}\langle S_iS_jS_k\rangle_\sigma + \sum_{ijkl}{}''' J_{ijkl}\langle S_iS_jS_kS_l\rangle_\sigma + \cdots. \quad (38)$$

Here the primes indicate omission of site coincidences (e.g. $i = j$) and the $\langle\ \rangle_\sigma$ indicates that the spin products are to be evaluated for the specified configuration σ. No constant (J_0) or one-body (J_1) term appears when evaluating an excess energy such as δE_{epi} or ΔE_{bulk}, since there are as many A and B atoms in the alloy as in the constituents. In eq. (38) we see two-body (pair) interactions J_{ij}, three-body interactions J_{ijk}, etc. between atoms. Since, e.g. sites i and j need not be nearest neighbors, distant-neighbor (as well as many-body) interactions are clearly present. Hand-in-hand with the description thus comes a set of one-, two-, and three-dimensional geometrical "figures" (points, pairs, triangles, tetrahedra, etc.) which show pictorially which sites are coupled to which, and by which J's (see [29, 30, 72]).

Equation (38) is thus very general, but is useful only if it can be truncated to a manageable number of terms while retaining quantitative accuracy. To see how these can be found, notice that if one truncates the expansion of eq. (38) to some maximum number M of J's regarded as physically important, then the excess energy of configuration σ is a linear combination of those M distinct J's, whose coefficients are the spin products $\langle\cdots\rangle_\sigma$. So far the configuration σ is arbitrary. The key to the whole approach is the recognition that for special ordered configurations the spin products $\langle S_iS_jS_kS_l\rangle$, etc. can be easily evaluated. For an ordered A_mB_n compound the A and B atoms occupy known crystal sites, so the relevant spin products are all known exactly by inspection of the unit cell. If we happen to know the ΔE for energy of M ordered A_mB_n compounds, we have M equations in M unknowns, and can solve for all the J's!

The bottom line is that if truncation of eq. (38) to a reasonable number M of physically important atom–atom couplings $J_{ijk\ldots}$ is feasible, each J is a known linear combination of the excess energies ΔE of M distinct ordered compounds. Once we have the J's, we can use them to predict the energies $\Delta E(\sigma')$ of arrangements of atoms $\sigma' \neq \sigma$ not already included in the M structures above; ΔE is itself a linear combination of the J's. Under conditions specified rigorously in [29] and [30], the excess energy of the

substitutionally disordered alloy in phase λ (ordered or disordered) then has the general form

$$\Delta E^{\lambda}(x, T) = \sum_{n=1}^{M} P_n^{\lambda}(x, T)\Delta E_n(\phi_{eq}). \tag{39}$$

Here ϕ_{eq} denotes the external conditions: for bulk we use for ΔE the enthalpy of eq. (23), while for epitaxial solids we use the enthalpy of eq. (24) applied to ordered structures. The interpretation of eq. (39) is simple: as far as thermodynamics is concerned, phase λ is a statistical mixture of ordered structures, with weights which are the "cluster probabilities" $\{P_n^{\lambda}(x, T)\}$. This formalism provides a practical means of calculating phase diagrams entirely from first principles. Modern total-energy methods [35] are designed to predict the ΔE_n above, and are precise and reasonably accurate for crystalline materials, provided there are not too many atoms within the primitive cell. With the energies ΔE_n known, finding alloy properties becomes an exercise in computational statistical mechanics. The cluster variation method of Kikuchi [71] is particularly well-suited for determining the contribution of the geometrical figures mentioned above to the thermodynamics of alloys, and has been used in most recent calculations. In practice [29, 30, 72], the $\{P_n\}$ are found by minimizing the free energy of phase λ with respect to them; distinct phases are distinguished by different symmetry-imposed occupations of the sites making up the figures retained. The free energies for all phases of interest can be compared and regions of phase coexistence (the phase diagram) computed. Systematic examination indicates that first-principles calculations for 8–12 ordered $(AC)_{4-n}(BC)_n$ compounds are generally sufficient to provide a reasonably accurate description of bulk semiconductor alloys. Figure 9 shows a number of such possible structures.

The formalism above is flexible, and makes possible the first-principles study of disordered alloys at a level previously achievable only for crystalline materials. Any elastic constraint which is understood for a pure crystalline material may thus be theoretically applied to an alloy, via the ordered structures used in eq. (39). *Elastic* effects in the alloy phase diagram are automatically directly incorporated. The process of alloying two semiconductor constituents AC and BC with different equilibrium molar volumes, for example, will cost strain energy as the two are dilated or compressed to the common, intermediate volume of the alloy. The structural parameter ϕ which appears in eq. (39) reflects such effects: ϕ is the alloy volume for a bulk alloy, while for a coherent epitaxial alloy on a specified substrate ϕ is the perpendicular dimension c of fig. 1a. In all cases, the equilibrium value $\phi_{eq}(x, T)$ is found by minimizing the energy of the phase with respect to ϕ.

Using this formalism, we can directly compare alloy properties under bulk and coherent epitaxial conditions. $GaAs_xSb_{1-x}$ was selected as a paradigm for semiconductor alloys since elastic effects will be important because of the large mismatch in the lattice constants of GaAs and GaSb, and because a large body of first-principles theoretical and experimental data exists [2, 73]. The ingredients are the quantities $\{\Delta E_{bulk}, a_{eq}, C_{11}, C_{12}\}$ for each of eight ordered $(GaSb)_{4-n}(GaAs)_n$ compounds [7, 33, 72] and the lattice constant a_s of the substrate selected.

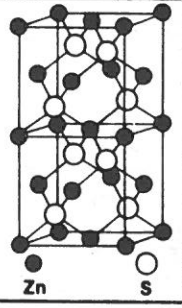
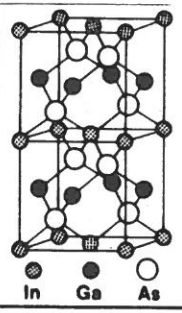
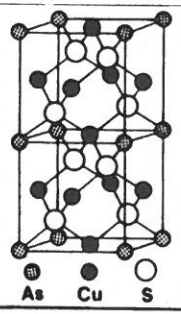
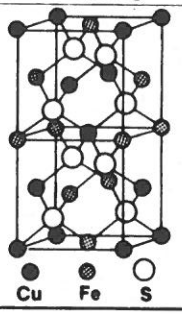
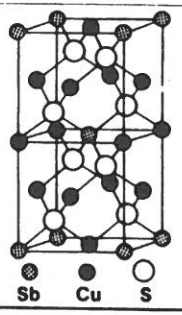
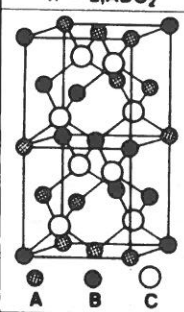
Ordering Vectors	(0,0,0)	(0,0,1)		(2,0,1)		(1,1,1)
Name (ternary)	Zincblende (Sphalerite)	Layered Tetragonal	"Luzonite"	Chalcopyrite	Famatinite	Layered Trigonal
Formula:	n = 0,4;AC	n = 2;ABC_2	n = 1,3; A_3BC_4	n = 2;ABC_2	n = 1,3; A_3BC_4	n = 2;ABC_2
						
Example: (ternary)	ZnS-type	$InGaAs_2$-type	Cu_3AsS_4-type	$CuFeS_2$-type	Cu_3SbS_4-type	$CrCuS_2$-type ($Na\ V\ S_2$)
Bravais Lattice:	Face centered cubic	Simple tetragonal	Simple cubic	Body centered tetragonal	Body centered tetragonal	Rhombohedral
Space Group						
Int. Tables:	$F\bar{4}3m$	$P\bar{4}m2$	$P\bar{4}3m$	$I\bar{4}2d$	$I\bar{4}2m$	R3m
Schoenflies:	T_d^2	D_{2d}^5	T_d^1	D_{2d}^{12}	D_{2d}^{11}	C_{3v}^5
Number:	216	115	215	122	121	160
Struckturbericht	B3	—	$H2_4$	$E1_1$	$H2_8$	
Pearson symbol:	CF8	—	CP8	tI16	tI16	hR4
Atomic positions: (ternary)	Zn: 4 a $\bar{4}3m$ S: 4 c $\bar{4}3m$	1A: 1 a $\bar{4}2m$ 1B: 1 c $\bar{4}2m$ 2C: 2 g mm	3 Cu: 3 c $\bar{4}2m$ 1 As: 1 a $\bar{4}3m$ 4 S: 4 e 3m	2 Cu: 4 a $\bar{4}$ 2 Fe: 4 b $\bar{4}$ 4 S: 8 d 2	1 Cu: 2 b $\bar{4}2m$ 2 Cu: 4 d $\bar{4}$ 1 Sb: 2 a $\bar{4}2m$ 4 S: 8 i m	1 A: 1 a 3m 1 B: 1 a 3m 1 C: 1 a 3m 1 C: 1 a 3m
Equivalent superlattice:	None	(1,1) in [0 0 1] direction	None	(2,2) in [2 0 1] direction	(1,3) in [2 0 1] direction	(1,1) in [1 1 1] direction

Fig. 9. Ordered structures between binary zinc-blende constituents. From Wei and Zunger [92a].

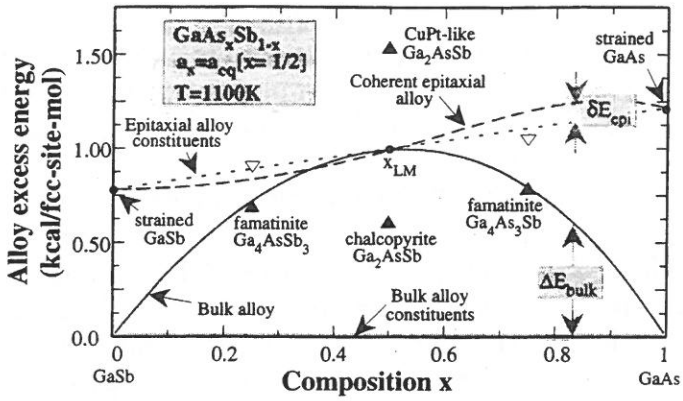


Fig. 10. Calculated enthalpy of $GaAs_xSb_{1-x}$ at 1100 K measured with respect to pure bulk GaAs and GaSb. Solid curve shows result for bulk alloy, dashed curve for coherent epitaxial alloy on a substrate lattice-matched at $x_{LM} = \frac{1}{2}$, and dotted curve the energy, for any x, of equivalent amounts of strained GaAs and GaSb. Triangles show formation energies of likely $(GaSb)_{4-n}(GaAs)_n$ compounds (see fig. 9 for the identification of these structures) under bulk (filled) and coherent epitaxial (open) conditions. Shaded gray labels show how bulk (ΔH_{bulk}) and epitaxial (δE_{epi}) formation energies are defined. Note coincidence of bulk and epitaxial curves at $x = x_{LM}$. From Wood [33].

Figure 10 shows the calculated [33] mixing enthalpy of the homogeneous, disordered alloy in bulk ($\Delta H^{bulk}(x)$ of eq. (23)) and under coherent epitaxial conditions ($\Delta H^{epi}(a_s, \hat{G}, x)$ of eq. (24)) at $T = 1100$ K. The substrate selected is lattice-matched to the homogeneous bulk alloy at $x = X_s = \frac{1}{2}$, i.e. $a_s = a_{eq}[GaAs_{0.5}Sb_{0.5}]$. Also shown are the formation energies of the most likely candidates for $(GaSb)_{4-n}(GaAs)_n$ ordered compounds (see fig. 9 for the identification of these).

Consider first the energetics of *bulk* phases in fig. 10 (solid lines and filled triangles). The curved solid line depicts the mixing enthalpy of the bulk alloy. We see that the CuPt-structure of Ga_2AsSb is high in energy, but the $x = \frac{1}{4}$ and $\frac{3}{4}$ ordered compounds are barely more stable than a disordered alloy of the same composition. Only the chalcopyrite structure is significantly lower in energy than the corresponding bulk disordered alloy. Qualitatively we expect this to lead to the possibility of an ordered compound stable against disordering, but (since the bulk formation energies of the ordered compounds are all positive) unstable with respect to decomposition into GaAs and GaSb.

The *epitaxial* situation is radically different, however. Equations (26), (27) imply that now the quantity determining stability is $\Delta E_{ss}(a_s, \hat{G}, x)$, the energy of the coherent epitaxial alloy measured with respect to equivalent amounts of the *strained* alloy constituents. The dashed line in fig. 10 indicates the enthalpy of the coherent epitaxial disordered alloy, which at $x = 0$ ends at the energy of strained GaSb. Thus the dotted line defines the constituent strain energy of equivalent amounts of the strained alloy constituents and ΔE_{ss} (denoted in fig. 10 as δE_{epi}) is shown via gray arrows for a particular x in the upper right of fig. 10. *We note that the disordered alloy is stable with respect to decomposition (at this temperature) for* $x \lesssim 0.5$. Furthermore, of the three epitaxial, ordered compounds (open triangles), two lie below the dotted line, indicating that they too are epitaxially stable. Moreover, these two also lie below the energy

for the coherent epitaxial, disordered alloy, so they are also stable with respect to disordering. Thus the chalcopyrite and famatinite Ga_4As_3Sb phases (see fig. 9) represent the thermodynamic equilibrium phases at $T = 0$ for the alloy at $x = \frac{1}{2}$ and $\frac{3}{4}$, respectively.

Figure 11 shows the bulk and coherent epitaxial phase diagrams for $GaAs_xSb_{1-x}$ calculated with the enthalpies of fig. 10 and the cluster variation method. The principal feature of the bulk phase diagram (fig. 11a) is a miscibility gap or binodal separating the region where the homogeneous, disordered alloy is thermodynamically stable from the region where a GaAs- and GaSb-rich mixture is preferred. The phase diagram is in good agreement with the experimental data which exist (ref. [72] compares the measured and calculated bulk phase diagram for this alloy). Deep in the miscibility gap, at temperatures below those shown in the figure, are the ordered compounds mentioned above; these are merely theoretical curiosities since, as noted above, they are unstable with respect to decomposition into equivalent amounts of GaAs and GaSb.

The predicted epitaxial phase diagram (fig. 11b) is rich in structure, however. *Not only is the disordered alloy stable to temperatures at least 680°C lower than in bulk, but, as advertised, the chalcopyrite phase is now thermodynamically stable.* The positive δE_{epi} evident in fig. 10 for $x \gtrsim 0.5$ manifests itself as a miscibility gap for $300 \leqslant T \leqslant 360$ K. Although the phase diagrams of other semiconductor $A_{1-x}B_xC$ alloys will differ

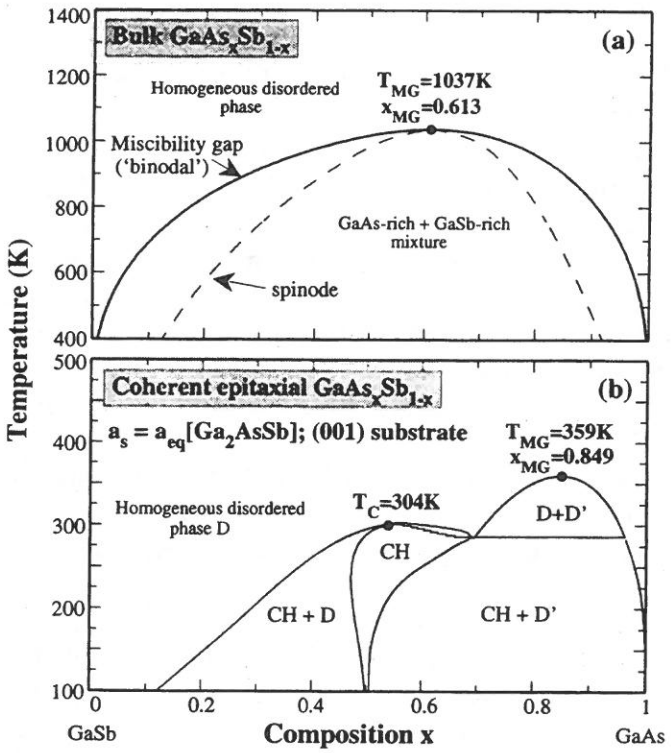


Fig. 11. Calculated (x, T) phase diagrams for $GaAs_xSb_{1-x}$ under bulk (a) and coherent epitaxial conditions on a (0 0 1) substrate lattice-matched to $GaAs_{0.5}Sb_{0.5}$ (b). D and D′ indicate disordered phases of distinct compositions. "Spinode" in (a) indicates limits of metastability of homogeneous, disordered phase. From Wood [33].

quantitatively, for constituents AC and BC with significant lattice mismatch the strain effects of coherent epitaxy will (i) permit growth of such alloys at temperatures far below what the bulk alloy phase diagram would suggest and (ii) cause stable crystalline $(AC)_{4-n}(BC)_n$ compounds to be a generic feature. Theoretically calculated epitaxial phase diagrams depend surprisingly weakly on the value of a_s selected [68]. Thus the principal effect of the loss of coherence with the substrate (as the film thickness increases, or, for fixed film thickness, as the lattice mismatch between the alloy and the substrate increases) is to restrict the validity of the description to a composition "window" centered on the lattice-matched composition X_s.

3.3.2. The surface phase diagram

The previous section described how epitaxial coherence alone, without surface effects, modifies the phase diagram of a bulk alloy. I next introduce the effect of a free surface during growth.

During vapor phase epitaxial growth one introduces (i) coherence with the substrate at the film/substrate interface and (ii) a free surface at the film/vacuum interface. Item (i) was treated in section 3.3.1. Here I treat cases for which item (ii) has a dominant effect.

A free surface can exhibit surface reconstructions [74, 75] which can be affected by coherence effects, thus leading to new stable structures absent both from bulk growth and from epitaxial growth without a film/vacuum interface (e.g. LPE). The relevant experimental background here is the observation of spontaneous formation of $(GaP)_1/(InP)_1$ (1 1 1) superlattices (the "CuPt" structure, see fig. 9) during *homogeneous* growth of $Ga_{0.5}In_{0.5}P$ alloys on a GaAs(0 0 1) substrate in vapor growth (OMVPE, MBE). By contrast, LPE growth does not produce this type of ordering. This rapidly advancing area of research has recently been reviewed by Zunger and Mahajan [76].

There are two noteworthy observations regarding the ordering process in vapor phase growth. *First*, the ordering is spontaneous, i.e. it is not introduced deliberately by, e.g. a shutter-controlled growth sequence. *Second*, the ordering cannot be of bulk or 3D epitaxial origin. This is inferred from the facts that (i) only two out of the four bulk-equivalent ⟨1 1 1⟩ directions are observed (the observed variants are termed $CuPt_B$), and (ii) surface preparation, and in particular the surface orientation, strongly influences the ordering and the domain distribution [76], and that (iii) epitaxial phase diagram calculations (e.g. fig. 10) predict that, in the absence of a free surface, chalcopyrite rather than CuPt structure will be stabilized. In fact these calculations show that if bulk energetics were driving the ordering, the observed CuPt structure would be among the least favorable structures.

It has been known for some time [74, 75] that the (0 0 1) surface of pure III–V *compounds* exhibits dimerization. If one considers the surface of an $A_{1-x}B_x$ alloy, one could envision either homopolar dimers (A–A and B–B) or heteropolar dimers (A–B). It is interesting to investigate how these topological possibilities interact with epitaxial coherence to stabilize new phases.

To understand quantitatively the role of surface reconstructions on ordering on a (0 0 1) substrate, Froyen and Zunger [77] and Bernard et al. [70] have considered the

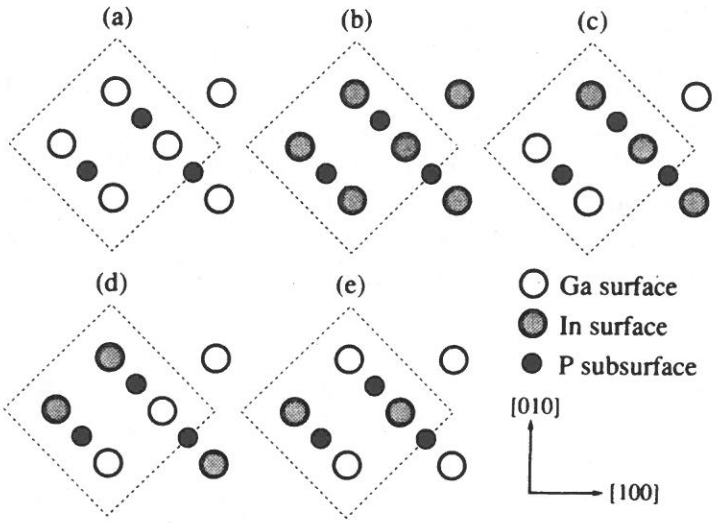


Fig. 12. Two-dimensional layer structures used in calculations of surface energies. See table 2. From Froyen and Zunger [77].

following hierarchy of questions: (a) Why is atomic ordering preferred at the surface to phase separation, given that the reverse trend exists in the bulk (fig. 11a)? (b) Why is there a preference for the (1 1 1) ordered CuPt structure rather than for the (2 0 1) ordered chalcopyrite structure, which is stable in 3D epitaxy (fig. 11b)? (c) Given the (1 1 1) ordered structures, why is the $CuPt_B$ variant preferred over other ($CuPt_A$) variants and (d) why is not CuPt ordering observed in the lattice-matched $Al_{0.5}Ga_{0.5}As$ alloy [78]?

The basic methodology used by Froyen and Zunger to address these questions was to perform accurate first-principles calculations [35] of *surface thermodynamics* and invoke "kinetic arguments" only if the former explanation fails. These questions can be phrased in terms of the relative stability of the five prototype *two-dimensional* (0 0 1) *alloy layers* shown in fig. 12. By stacking these prototype a, b, c, d, and e layers in different ways, many of the commonly observed 3D structures can be obtained (see table 1). These include phase separation, (0 0 1) ordering (= CA), chalcopyrite ordering (= CH), $CuPt_B$ ordering ($= CP_B$), or the unobserved $CuPt_A$ ordering ($= CP_A$). Table 2 gives the calculated total energies of the various surface topologies, as obtained by Froyen and Zunger [77], using the first-principles local-density pseudopotential method [35]. These energies correspond to eq. (24), but now the energy E is calculated for a finite solid with a *relaxed and reconstructed free surface* rather than for a three-dimensional periodic system. These calculations address questions (a)–(d) above as follows:

(i) Table 2 shows that many ordered structures have lower energies than surface phase-separation ("a + b" in Table 2), hence $\Delta H^{epi} < 0$ *at the surface* despite $\Delta H^{epi} > 0$ in the 3D solid (fig. 10). In other words, phase separation, which is the stablest arrangement in the bulk is predicted to be partially suppressed at the surface if the latter orders. This reflects the combined energy lowering due to coherent epitaxy and surface reconstruction: dimerization of pure a = GaP and pure b = InP lowers the energy *less*

Table 1

3D structures characterized by stacking of the (0 0 1) bilayers shown in fig. 12. The structure is identified both by a superlattice notation (the direction G and the repeat period $2n$) and by the notation used in the text. Each layer is shifted laterally as indicated in parentheses (in units of the zinc-blende lattice constant). Several other structures are degenerate with those tabulated: [1 1 0] $n = 1$ is identical to [0 0 1] $n = 1$ CA, [2 0 1] $n = 2$ CH is degenerate with [0 2 1] $n = 2$ CH, and [0 1 0] CA is degenerate with [1 0 0] CA. The [1 0 2] $n = 2$ and [0 1 2] $n = 2$ CH structures have (0 0 1) layers that cannot be represented by the 2 × 2 patterns in fig. 12. Notice that the d layer occurs only in the observed CP_B phase and in the Y2 superlattice. The two are distinguished only by the third- and fourth-layer stacking. From Froyen and Zunger [77].

Structure		Layer number			
G, n	Name	1	2	3	4
(Binary)	A	a(0, 0)	a($\frac{1}{2}$, 0)	a(0, 0)	a($\frac{1}{2}$, 0)
(Binary)	B	b(0, 0)	b($\frac{1}{2}$, 0)	b(0, 0)	b($\frac{1}{2}$, 0)
[0 2 1], 2	CH	e(0, 0)	e($\frac{1}{2}$, 0)	e($\frac{1}{2}$, $\frac{1}{2}$)	e(0, $\frac{1}{2}$)
[1 0 0], 1	CA	e(0, 0)	e($\frac{1}{2}$, 0)	e(0, 0)	e($\frac{1}{2}$, 0)
[0 0 1], 1	CA	a(0, 0)	b($\frac{1}{2}$, 0)	a(0, 0)	b($\frac{1}{2}$, 0)
[1 1 1], 1	CP_A	c(0, 0)	c($\frac{1}{2}$, 0)	c($\frac{1}{2}$, $\frac{1}{2}$)	c(0, $-\frac{1}{2}$)
[1 1 0], 2	Y 2	c(0, 0)	c($\frac{1}{2}$, 0)	c(0, 0)	c($\frac{1}{2}$, 0)
[$\bar{1}$ 1 1], 1	CP_B	d(0, 0)	d($\frac{1}{2}$, 0)	d($\frac{1}{2}$, $-\frac{1}{2}$)	d(0, $\frac{1}{2}$)
[$\bar{1}$ 1 0], 2	–	d(0, 0)	d($\frac{1}{2}$, 0)	d(0, 0)	d($\frac{1}{2}$, 0)

Table 2

Surface energies for the various $GaInP_2$/GaAs (0 0 1) reconstruction modes (fig. 12) discussed in the text. The energies are in meV per surface atom relative to the unreconstructed a + b (phase-separated) surface (surfaces a and b have their own separate zero of energy). We show separately the results of the dimerized + buckled + tilted (DBT) surface with Ga up and with In up. Blank entries in the "dimerized" line refer to cases where a pure dimerization without buckling or tilting cannot be defined. From Froyen and Zunger [77].

Surface geometry	Surface type					
	a	b	a + b	c	d	e
Unreconstructed	0	0	0	14	−9	2
Dimerized	−785	−366	−575	–	–	−620
Dimerized + buckled	−732	−448	−590	–	−692	−602
DBT–Ga up	−836	–	−701	−684	−623	−715
DBT–In up	–	−564	−701	−684	−799	−705

than dimerization in mixed structures, e.g. e = CH. This answers question (a) above.

(ii) For relaxed but *unreconstructed* surfaces the energy difference between the various surface arrangements are considerably smaller than kT_g, where $T_g \approx 900$ K is the growth temperature. Hence, simple relaxations derived by atomic size-mismatch cannot explain preferential ordering.

(iii) The situation changes significantly when reconstructions are permitted (figs. 13 and 14). Surface dimerization results in *heteropolar* (Ga–In) dimers on surfaces c = CP_A and e = CH (fig. 14) but in *homopolar* (Ga–Ga or In–In) dimers on surfaces

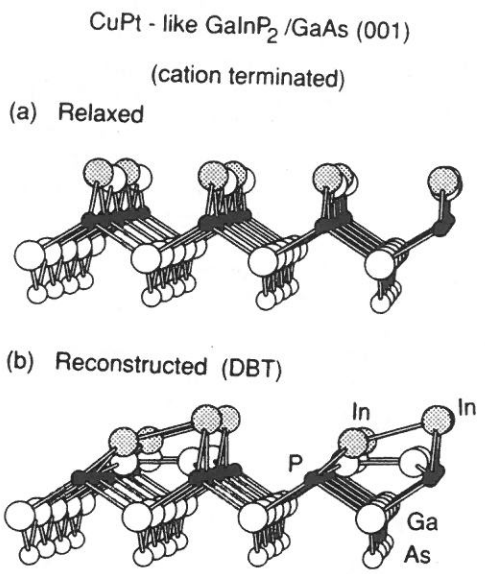


Fig. 13. Calculated equilibrium surface geometrics of CuPt-like cation terminated $GaInP_2/GaAs$ (0 0 1). (a) Relaxed but unreconstructed, (b) reconstructed. From Bernard et al. [70].

a = GaP, b = InP, and d = CP_B (fig. 13). The significant topological difference between CP_A and CP_B is hence the occurrence of unlike atom versus like atom surface dimers, respectively. Relative to the undimerized surfaces, dimerization lowers the energy by an average of 600 meV per surface atom (table 2, row 2). In addition to dimerization there are two other energy-lowering reconstructions within the 2 × 2 surface unit cell: first, dimers relax perpendicularly to the surface creating [$\bar{1}$ 1 0] dimer rows of alternating high and low dimers (e.g. see fig. 14c). We will refer to this as "buckling". Relative to the dimerized, unbuckled surface, buckling lowers the energy of the b surface (table 2, row 3), but raises the energy for surfaces a and e (so these surfaces will not buckle). Hence, buckling prefers the d = CP_B structure over the e = CH structure. This answers question (b) above, i.e. CP surface ordering is preferred to CH surface ordering mainly due to buckling. Second, the high dimer tilts in the [1 1 0] direction becoming non-horizontal, whereas the low dimer remains virtually horizontal (e.g. see figs. 13b, 14d). Since the four surface sites are inequivalent in the final geometry, there are two different ways of distributing the two Ga and the two In atoms in each of the topologies, c, d, and e. We will characterize this by the type of atom (Ga or In) occupying the site on the high dimer that tilts upwards. The last two rows of table 2 show that the d = CP_B surface strongly prefers having the larger In atom on the high dimer, whereas the e = CH surface shows a slight preference for the smaller (but more electronegative) Ga atom to be tilted up. Tilting leads to a uniform energy lowering of 100 meV but does not affect the relative stability of the surface (table 2).

As shown in table 2, the full reconstructions (dimerization, buckling, and tilting) considerably lower the energy of all the surfaces and, most significantly, make the surface corresponding to the observed CP_B ordering (d) the lowest in energy by 84 meV per surface atom. *Thus, surface reconstruction not only favors atomic arrangements that*

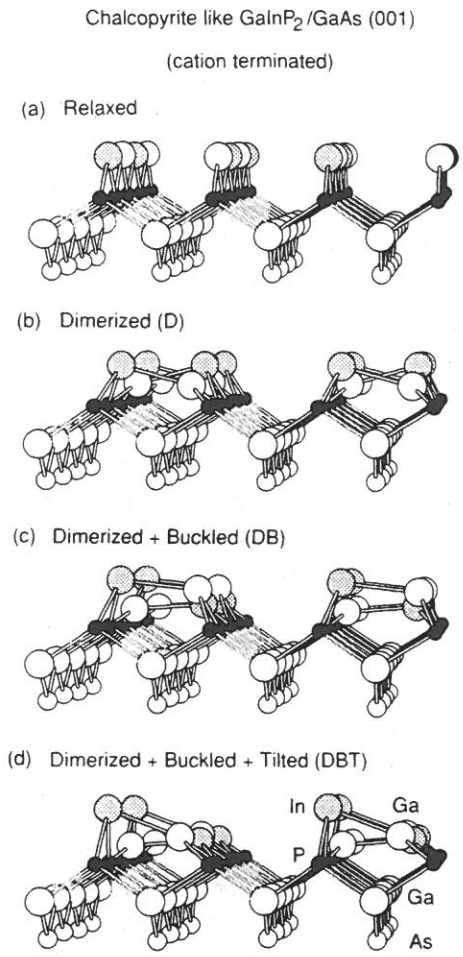


Fig. 14. Same as fig. 13 but for the chalcopyrite surface structure. Parts (a)–(d) show the changes in structure as additional energy-lowering displacements are allowed. Note that the dimers here are heteropolar, while those of fig. 13 are homopolar. Ga (white), In (grey), and P (black) are on top of GaAs (white). From Bernard et al. [70].

are unstable in the bulk, but it also results in energy differences large enough to produce order at T_g. This answers question (c) above.

(iv) Since $Al_{0.5}Ga_{0.5}As$ *does not order* when grown on (0 0 1) substrates but $Ga_{0.5}In_{0.5}P$ does, it is interesting to examine whether the above theory will reproduce this different behavior. Froyen and Zunger [77] have compared the energies of the $GaInP_2$ surfaces to those of $AlGaAs_2$. They separately considered Ga and Al as the highest atom on the high dimer. They find that the energy of surface d = CP_B is now only 9 meV per surface atom below that of the next lowest configuration (the e surface), so the energy difference is negligible compared to kT_g. Hence, the resulting surface topology in AlGaAs/GaAs(0 0 1) is likely to be disordered. This is consistent with the experiment of Kuan et al. [78], where $Al_{0.5}Ga_{0.5}As$/GaAs (0 0 1) was found to be disordered while ordering was seen only for growth on the (1 1 0) surface. This answers question (d) above.

The previous discussion showed how $T = 0$ surface energetics could drive selective ordering in the subsurface layers. This does not imply that ordering will persist at finite (i.e. growth) temperatures. To find out what happens at finite temperatures, one needs to incorporate entropy effects [73] and consider "non-ideal" structures, such as the *imperfectly ordered* CuPt and *imperfectly disordered* alloy. In addition to questions (a)–(d) posed and addressed above, we now add (e): would the CP_B ordering survive finite (growth) temperatures? (f) Would ordering exist also in non-stoichiometric compositions, e.g. in $Ga_{1-x}In_xP$ with $x \neq 0.5$ and (g) would the CP_B arrangement be stable with respect to partial Ga–In atomic swaps? Such a study was undertaken by Osorio et al. [73]. They used the Ising Hamiltonian of eq. (38) fitted to many total-energy calculations of *reconstructed surface structures*. This fit produced the *surface* interaction energies (J) of eq. (38). Then, this generalized Ising equation was used to find (i) the $T = 0$ ground-state structures of this Hamiltonian and (ii) its $T > 0$ free energies $E - TS$. This was done by solving the configurational Hamiltonian using the cluster variation method of Kikuchi [71], thus including entropy effects and site correlations. Both ideal CuPt as well as partially ordered structures were considered.

Figure 15 shows the $CuPt_B$ long-range order parameter η for the $Ga_{0.5}In_{0.5}P$ cation-terminated surface, while fig. 16 shows the results for the $Ga_{0.7}In_{0.3}P$ surface. The major findings of this study were:

(i) As a consequence of reconstruction, the cation-terminated surface orders in the $CuPt_B$ structure and shows a gradual decrease in the degree of order with temperature, *with no phase transition* (fig. 15). We see that a *thermodynamic* model can account for the existence of ordering at $x = 0.5$ and that *a significant $CuPt_B$ order parameter persists up to* ≈ 1500 K. This answers question (e). Note that in the *absence* of reconstruction, ordering disappears already at ≈ 150 K (fig. 15). Furthermore, fig. 16 shows that $CuPt_B$ ordering persists at growth temperatures even at other compositions, e.g. $Ga_{0.7}In_{0.3}P$. This is consistent with the hitherto surprising experimental findings by Kondow et al. [79] showing $CuPt_B$ ordering even for $x \neq \frac{1}{2}$. This answers question (f).

(ii) A study of the energies of intra-layer atomic swaps shows a remarkable resilience of the two-dimensional $CuPt_B$ structures at the surface of cation-terminated alloys and at the second cation subsurface layer of anion-terminated alloys. This answers question (g).

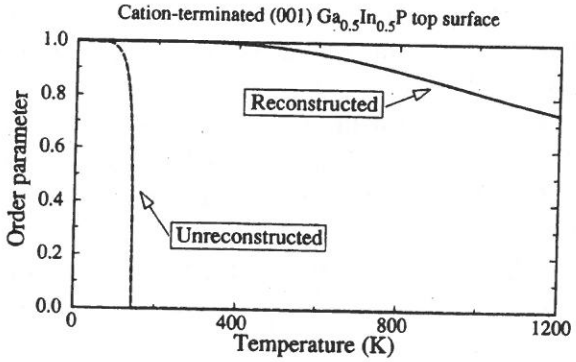


Fig. 15. Calculated temperature dependence of the CuPt long-range order parameter for a cation-terminated $Ga_{0.5}In_{0.5}P$ surface. From Osorio et al. [73].

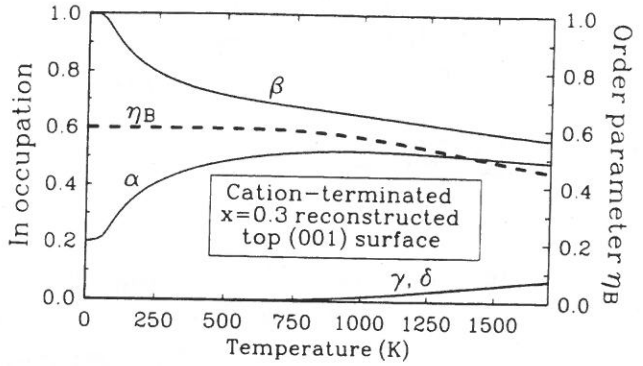


Fig. 16. Calculated temperature dependence of the CuPt order parameter η_B (right axis) and the occupations by In atoms of the four cation sublattices α, β, γ, and δ (left axis) for off-stochiometric $Ga_{0.7}In_{0.3}P$. The value $\eta_B = 1$ corresponds to $C_\alpha = C_\beta = 1$ and $C_\gamma = C_\delta = 0$. From Osorio et al. [73].

In summary, this section showed how the combination of coherent epitaxy with surface reconstruction during vapor phase growth can stabilize new "ordered-alloy" structures which are unstable in liquid phase epitaxial growth. The recent review article on this subject [76] discusses other mechanisms for this ordering and highlights the fact that CuPt ordering is seen virtually in all vapor-grown III–V alloys.

3.3.3. Composition pinning in epitaxial alloys

It has been observed [45–48] that the measured composition x_{epi} of the epitaxial alloy tends to be "pinned" at a value near x_{LM}, where the alloy is lattice-matched (LM) to the substrate ($X = X_s$), even though the composition x_{bk} of the corresponding (unsupported) bulk alloy varies widely. This pronounced deviation from behavior expected from the bulk phase diagram ("latching" or "pulling" effect) was first observed in an elegant experiment of Stringfellow [45], who found that large changes of the liquid composition in liquid phase epitaxy growth of $Ga_xIn_{1-x}P$ on a GaAs substrate produced analogously large composition variations in bulk-like platelets (as expected from the bulk phase diagram), while the composition of epitaxial layers x was pinned to a narrow region near x_{LM}. For large $|x_{epi} - x_{LM}|$ misfit dislocations – which destroy coherency between the epitaxial alloy and the substrate – nucleate and remove the effect. This effect has since been seen in LPE for a variety of systems on diverse substrates of various orientations [45–48].

For any phase there is a monotonic relation between the chemical potential $\mu = d\Delta F/dx$ and the composition x. We may Taylor expand about μ_{LM} (where $a(x_{LM}) = a_s$ and bulk and epitaxial alloys are indistinguishable according to eq. (27b)) to find

$$\delta x_{bk}(\mu) \equiv x_{bk}(\mu) - x_{LM} = \left.\frac{\mu - \mu_{LM}}{d^2\Delta F^{bk}/dx^2}\right|_{x_{LM}}, \tag{40}$$

$$\delta x_{epi}(\mu) \equiv x_{epi}(\mu) - x_{LM} = \left.\frac{\mu - \mu_{LM}}{d^2\Delta F^{epi}/dx^2}\right|_{x_{LM}} + \cdots, \tag{41}$$

using the identity $dx/d\mu = (d\mu/dx)^{-1} = (d^2\Delta F/dx^2)^{-1}$. Under identical growth conditions μ is common to bulk and epitaxial alloys growing near equilibrium. We may thus

conveniently measure the degree of composition pinning *at the lattice-matched composition* by the slope $Q(x_{LM}) \equiv dx_{epi}/dx_{bk}$ of the curve $x_{epi}(\mu)$ versus $x_{bk}(\mu)$ at x_{LM}, or

$$Q = \frac{d^2\Delta F^{bk}(x_{LM})/dx^2}{d^2\Delta F^{epi}[x_{LM}, a_s = a(x_{LM})]/dx^2} = \frac{T - T_{bk}(x_{LM})}{T - T_{epi}(x_{LM})}. \tag{42}$$

As described above, a homogeneous disordered phase of a phase-separating alloy such as $Ga_{1-x}In_xAs$ ceases to be stable for temperatures below the spinodal temperature $T(x)$, whose maximum coincides with the maximum miscibility gap temperature (fig. 11a). The spinodal temperature $T(x)$ is the locus of points satisfying $d^2\Delta F/dx^2 = 0$; Wood and Zunger [68] demonstrated that the epitaxial constraint strongly modifies $\Delta H(x)$ while leaving $\Delta S_{epi}(x) \approx \Delta S_{bk}(x)$. Thus, from eq. (37), we see that at the lattice-matched composition $x = X_s = X_{LM}$ we have

$$T_{epi}(X_{LM}) = T_{bk}(x) - \frac{9}{4R} q(x) B(x) a(x) [a_B - a_A]^2, \tag{43}$$

where the last term is denoted $\Delta T(X_{LM})$.

Equations (42) and (43) connect composition pinning and epitaxial stabilization of alloys. Since $\Delta T \leqslant 0$, for any T above the bulk miscibility gap temperature $Q < 1$, i.e. composition pinning is a *universal* feature of coherent epitaxial growth, even on a lattice-matched substrate. Perfect composition pinning ($Q = 0$) occurs at the bulk spinodal temperature $T = T_{bk}(x)$, while $Q \to 1$ as $T \to \infty$. (While composition pinning may persist below the bulk miscibility temperature, it must do so in some guise other than an x_{epi} versus x_{bk} plot, since the full range of bulk alloy composition is inaccessible.) Larché and Cahn [55] explained composition pinning in LPE in precisely the same terms, but failed to note the quantitative connection with epitaxial stabilization of alloys ($\Delta T < 0$).

Equations (42) and (43) show that composition pinning and epitaxial stabilization both scale as $(da/dx)^2 \propto (a_B - a_A)^2$, i.e. the lattice mismatch between the *alloy constituents* (not between substrate and the film material); nearly lattice-matched systems (e.g. $Ga_{1-x}Al_xAs$) will show no epitaxial effects. Both effects are larger for elastically stiff alloys, via $q(x)B(x)$, and depend on orientation through q. The physical origin of both is not epitaxial strain per se (since both occur even on lattice-matched substrates), but rather the composition dependence of the alloy parameter $a(x, T)$. This dependence obviously may be traced to the size mismatch $|a_B - a_A|$ between alloy constituents, reflecting different atomic sizes or distinct bond lengths and their imperfect accommodation in an alloy environment.

Figure 17 gives results for $GaAs_xSb_{1-x}$ grown on substrates lattice-matched to the alloy at $x = 0$, $x \approx 0.5$, and $x = 1$ at 1300 K [68]. The epitaxy-induced increase in $d\mu/dx = d^2\Delta F/dx^2$ (i.e. $Q < 1$) in fig. 17a is most pronounced at $x_{LM} \approx 0.5$; a given μ (fixed growth conditions) corresponds to different values of x_{epi} and x_{bk}. Figure 17b shows strong deviations from $x_{epi} = x_{bk}$ (i.e. $Q = 1$) even well away from x_{LM} and 50°C above the calculated maximum bulk miscibility gap temperature (≈ 1245 K). These

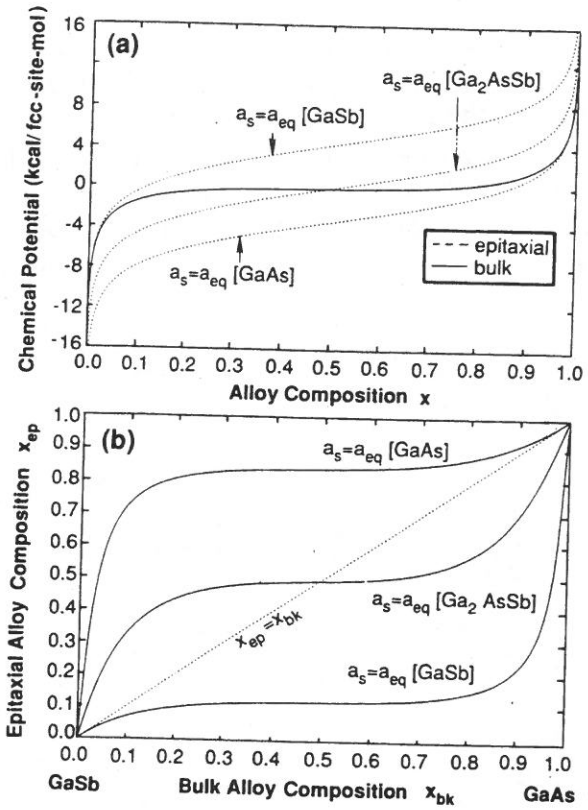


Fig. 17. Predictions for bulk and epitaxial $GaAs_xSb_{1-x}$ phase diagrams calculated with the cluster variation method: (a) dependence of alloy composition on chemical potential μ; (b) x_{epi} (μ) versus x_{bk} (μ), exhibiting composition pinning. From Wood and Zunger [68].

results are directly relevant to liquid phase epitaxy and strongly resemble the original curve of Stringfellow [45].

4. *Electronic consequences of coherent epitaxy*

Our foregoing discussion showed that coherent epitaxy leads to (i) deformation of the unit cell parameter(s) in the direction $\hat{G}$ perpendicular to the substrate (eq. (7)) and (ii) stabilization of new structures of pure compounds (section 2) or alloys (section 3). Both effects lead to modifications in the electronic properties of the film material. These are reviewed next.

4.1. *Effects of coherent epitaxy on the band structure of cubic compounds or alloys: strain splitting*

The valence-band maximum (VBM) of a cubic zinc-blende compound or alloy (in the absence of spin–orbit splitting) is the triply degenerate Γ_{15v} state. High-symmetry conduction-band states include the singly degenerate Γ_{1c} state, as well as the X_{1c}, X_{3c}, .

and L_{1c} valleys. When grown coherently on a substrate whose lattice constant a_s differs from the equilibrium value a_{eq} of the cubic film material, the films lattice spacing will distort in the direction $\hat{G}$ perpendicular to the substrate plane, as indicated by eq. (7). Consider, for example, the cubic $GaAs_{0.5}P_{0.5}$ alloy, whose bulk equilibrium lattice constant $a(x = 0.5)$ is denoted by $\bar{a}$. If this alloy is grown coherently on a GaAs (0 0 1) substrate, the film will contract in the substrate plane to match the (smaller) dimension $a_s = a$(GaAs); according to eq. (7) and fig. 1b the c-axis of the film will elongate to

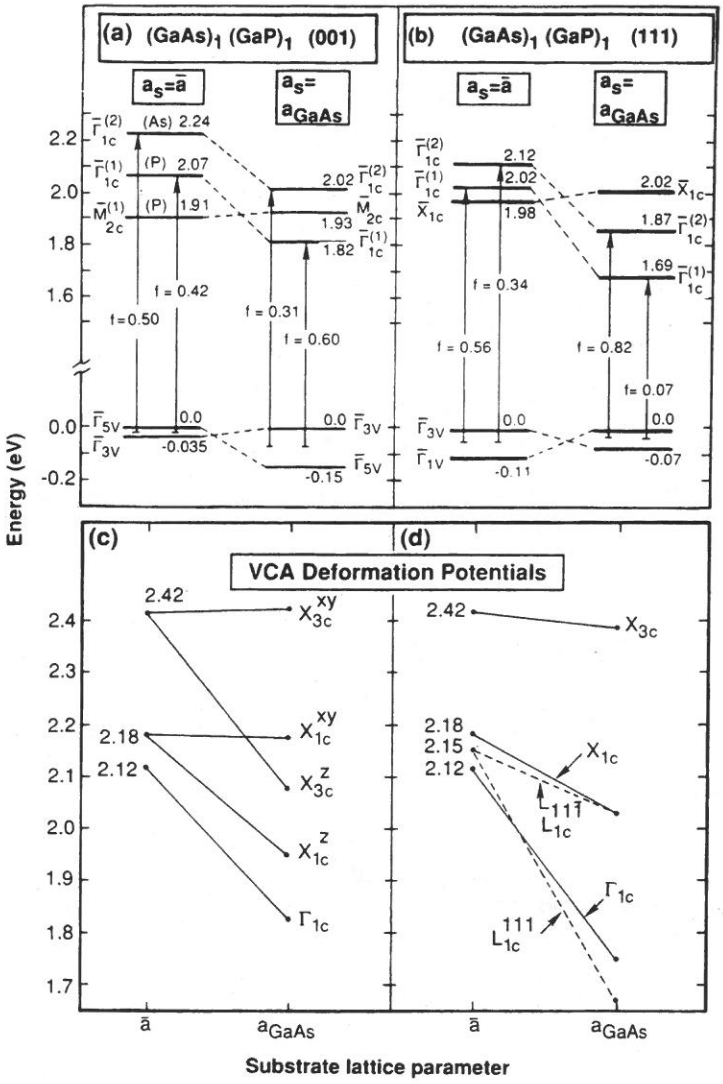


Fig. 18. Calculated energy levels for the monolayer $(GaAs)_1(GaP)_1$ SLs in: (a) (0 0 1) orientation and (b) (1 1 1) orientation on two substrates with different a_s. The by LDA (local density approximation) calculated conduction bands were rigidly shifted upwards by 0.75 eV. Valence energies here are without spin–orbit splitting, consideration of which (within a quasi-cubic model) decreases all gaps 0.04–0.07 eV. Here f is the square of the momentum matrix element (summed over the three valence maxima at $\bar{\Gamma}$), in Ry. In (c) and (d) we show the respective biaxial deformation potentials for VCA Ga⟨AsP⟩. Note how increasing a_s from $\bar{a}$ to a_{GaAs} leads to an indirect-to-direct crossover.

$c_{eq}/a_{eq} > 1$. This splits the Γ_{15v} state into $\Gamma_{5v} + \Gamma_{3v}$. If the substrate is GaAs (1 1 1), the VBM will split into $\Gamma_{3v} + \Gamma_{1v}$. The bottom part of fig. 18 shows the changes in the conduction-band states of $GaAs_{0.5}P_{0.5}$ for (i) a GaAs (0 0 1) substrate and (ii) a GaAs (1 1 1) substrate, as calculated from the local density theory by Dandrea and Zunger [80]. We see that for a GaAs (0 0 1) substrate the X_{1c} alloy state splits into $X^z_{1c} + X^{xy}_{1c}$ and the X_{3c} state splits into $X^z_{3c} + X^{xy}_{3c}$, where $z = (0\,0\,1)$ and xy are the orthogonal directions. Note the larger slope of the z-component energies relative to the xy-component energies. For a GaAs (1 1 1) substrate the film L_{1c} conduction band splits into $L_{1c}(1\,1\,\bar{1}) + L_{1c}(1\,1\,1)$, where the latter component has a larger slope. These splittings are readily calculated by performing band-structure calculations [35] using the (distorted) lattice geometry prescribed by eq. (7). They naturally lead to changes in the optical spectra. Numerous examples exist illustrating the optical consequences of such strain-induced splittings and shifts both in epitaxial compounds and epitaxial alloys [81–83]. In such simple cases the essential physics is embodied by the response of the otherwise cubic band structure (fig. 18) to non-cubic unit-cell deformations (eq. (7)). In what follows I will focus on somewhat more subtle consequences of coherent epitaxial strain, namely on the way such strains affect (i) valence-band offsets and (ii) the nature (direct versus indirect) of the optical transitions in coherently strained compounds.

4.2. *Effects of coherent epitaxy on heterojunction band offsets*

I start this discussion by explaining the factors determining the band offset between materials A and B and the way that epitaxial strain affects these quantities.

The valence-band offset ΔE_v across an A/B interface can be written as

$$\Delta E_v = [E_v - \langle V \rangle]_B - [E_v - \langle V \rangle]_A + [\langle V \rangle_B - \langle V \rangle_A]_{IF}. \tag{44}$$

The first term on the right-hand side is the energy of the valence-band edge in pure material B with respect to the average potential $\langle V \rangle_B$ in that material. The second term is the similar quantity in material A. Both these band-structure terms are properties of the bulk materials. These two terms have nothing to do with the A/B heterointerface, except for the fact that the strain configurations of pure materials A and B must be identical to the configurations of A and B on either side of the A/B interface. The third term in eq. (44) is the interfacial (IF) dipole: the difference in the average potential on either side (but far from) the A/B interface. In the case of doped materials leading to space-charge effects, the interface dipole must be evaluated on a length scale across which negligible band bending occurs. We denote the interface dipole as ΔV_{IF} below. Since it depends only on the charge density of the heterojunction, the interface dipole is a ground-state property, and not a band-structure property like the first two terms of eq. (44). To simplify the discussion I first illustrate the physical origins of the terms in eq. (44) in a lattice-matched case (no strain), and introduce the effects of strain in the second step.

Figure 19 illustrates the physical significance of the terms in eq. (44) for the lattice-matched AlAs/GaAs (0 0 1) case. The value of ΔV_{IF} shown in fig. 19 is calculated

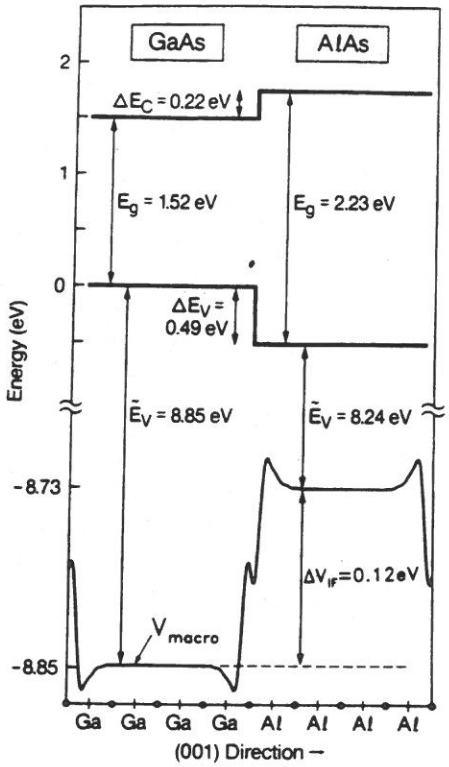


Fig. 19. Calculated terms appearing in eq. (44) for the valence-band offset across a (0 0 1) GaAs/AlAs interface. The first two terms in eq. (44) of the text are labeled $\tilde{E}_v = E_v - \langle V \rangle$ here. From Dandrea et al. [84].

(following the scheme of Baroni et al. [86]) as follows: a macroscopic average $V_{\text{macro}}(z)$ of the potential $V(x, y, z)$ throughout the heterojunction is obtained from the relations

$$V_{\text{macro}}(z) = \frac{1}{2L} \int_{-L}^{L} dz' \bar{V}(z + z'), \tag{45}$$

$$\bar{V}(z) = \frac{1}{A_{xy}} \int_{A_{xy}} dx\, dy\, V(x, y, z). \tag{46}$$

$\bar{V}(z)$ is a planar-averaged potential, i.e. it is the full potential $V(x, y, z)$ averaged (over a two-dimensional unit cell A_{xy}) in a plane parallel to the interface. Far from the interface, this planar-averaged potential $\bar{V}(z)$ is periodic over a two-monolayer distance $2L$. This is a reflection of the fact that the charge density far from the interface is identical to that of the pure binary compound (GaAs or AlAs in the example of fig. 19), and therefore that the potential in this region must be identical to the potential of the pure binary compound to within a constant. The macroscopic-averaged potential as defined in eq. (45) is thus constant far from the interface, and the difference in the constant values on either side of the interface is simply the interfacial dipole ΔV_{IF}. Figure 19 shows that $E_v - \langle V \rangle$ is 8.85 eV on the GaAs side and 8.24 eV on the AlAs side, while $\Delta V_{\text{IF}} = 0.12$ eV. Thus, the band offset is $\Delta E_v = 8.85 - 8.24 - 0.12 = 0.49$ eV. This agrees well with the recently measured values [85].

Since the only contribution to the band offset due to the interface itself is the interfacial dipole term ΔV_{IF}, the change in the band offset with some change in interfacial atomic geometry is completely and exactly the change in the interfacial dipole:

$$\Delta(\Delta E_v) = \Delta(\Delta V_{IF}). \tag{47}$$

On the other hand, if we consider a fixed interfacial structure and change only the lattice mismatch between A and B, the band-offset changes are confined entirely to the first two terms of eq. (44): the energy separation $E_v - \langle V \rangle \equiv \tilde{E}_v$ (fig. 19) between the average potential and the VBM of material A or B shifts as these materials are strained.

With these general clarifications we now demonstrate how epitaxial strain alters the band offset. In the case of a coherent strained-layer interface between two lattice-mismatched materials, the epitaxial constraint implies the constancy of the in-plane repeat area A_{xy} across the junction, but the existence of two different tetragonal deformations means that the macroscopic average in eq. (45) must now be done with two different interplanar spacings:

$$V_{\text{macro}}(z) = \frac{1}{4l_1 l_2} \int_{-l_1}^{l_1} dz' \int_{l_2}^{l_1} dz'' \bar{V}(z + z' + z''). \tag{48}$$

As in the lattice-matched case, the interfacial dipole ΔV_{IF} is simply the change in $V_{\text{macro}}(z)$ as the interface is crossed. Figure 20 shows the calculated (Dandrea and Zunger [80]) band offsets between GaP and GaAs in different strain conditions (for the present discussion, focus first on the left-hand sides of panels (a), (b), and (c) of fig. 20). We first consider the case where both GaP and GaAs remain *cubic*, with the lattice parameters $a_z = a_{xy} = a(\text{GaAs}_{0.5}\text{P}_{0.5})$ (fig. 20a). Here each of these components deforms *hydrostatically* from the equilibrium values $a_{eq}(\text{GaAs})$ and $a_{eq}(\text{GaP})$ to $\bar{a} = a(\text{GaAs}_{0.5}\text{P}_{0.5})$, so band states shift but do not split. Relative to equilibrium bulk the valence-band maximum on the GaAs side shifts upwards by 0.11 eV, that on the GaP side shifts downwards by 0.14 eV, giving an overall "cubic offset" of 0.25 eV. The conduction-band offsets are shown in the left-hand side of fig. 20a.

Next, require that both GaP and GaAs be coherent with each other, with an in-plane lattice constant $a_{xy} = a(\text{GaAs}_{0.5}\text{P}_{0.5}) \equiv \bar{a}$ and a (0 0 1)-oriented substrate. The GaP half thus *expands* in the xy-plane, resulting, according to eq. (7), in $c/a = 0.966$, while the GaAs half *contracts* in the xy-plane, resulting in $c/a = 1.034$. The VBM splits into $\Gamma_{5v} + \Gamma_{3v}$ (fig. 20b), and the conduction-band states split according to fig. 18. We see that for an (0 0 1) substrate (fig. 20b) the Γ_{5v} valence-band offset is now $0.21 + 0.31 = 0.52$ eV, while the Γ_{3v} offset is $0.01 + 0.13 = 0.14$ eV. Recall that the *cubic* case (fig. 20a) gave a single offset of 0.25 eV. Furthermore, while in the cubic case the lowest direct interband transition $\Gamma_{15v} \to \Gamma_{1c}$ was at $2.00 - 0.11 = 1.89$ eV and the lowest indirect transition $\Gamma_{15v} \to X_{1c}$ was at $2.11 - 0.11 = 2.00$ eV, in the tetragonally deformed case (fig. 20b) the lowest direct and indirect transitions shift to $\Gamma_{5v} \to \Gamma_{4c}$ ($1.76 - 0.21 = 1.55$ eV) and $\Gamma_{5v} \to X^z_{1c}$ ($1.94 - 0.21 = 1.73$ eV). Note also that in the tetragonally deformed case the X_{1c} well is a minimum on the GaP half, while in the cubic case the minimum was on the GaAs side.

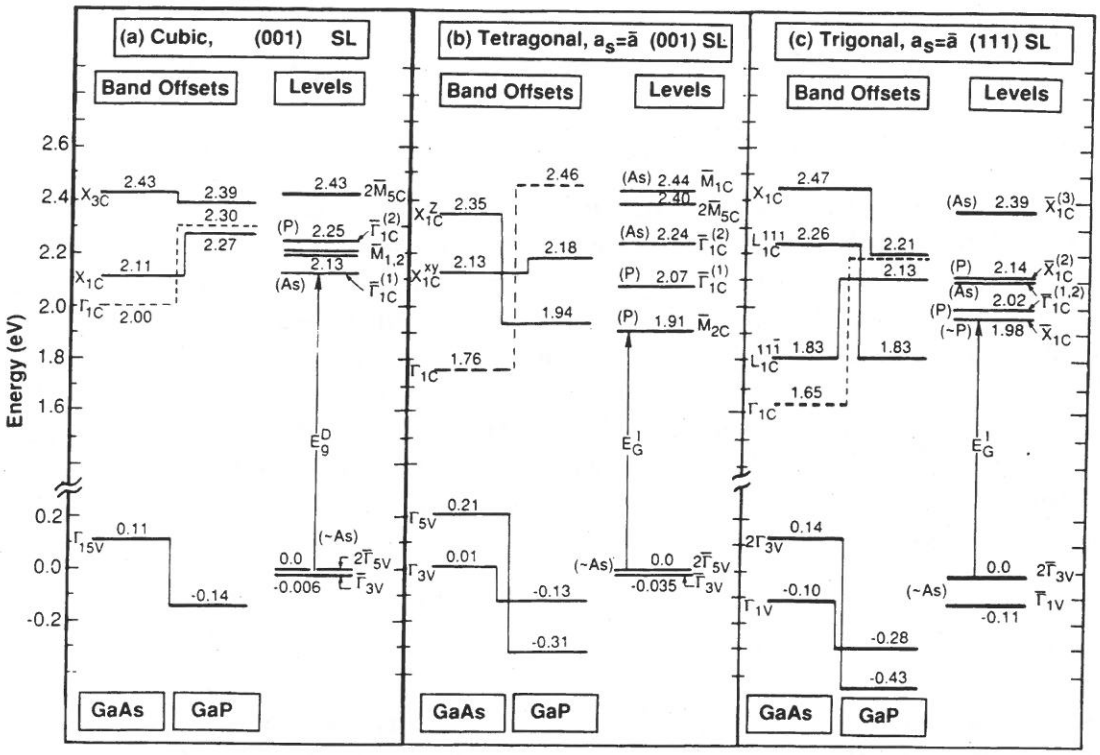


Fig. 20. Calculated band offsets (for thick GaP/GaAs) and SL energy levels (for $(GaAs)_1(GaP)_1$) on $a_s = \bar{a}$: (a) cubic (0 0 1) SL, (b) (0 0 1) SL with cell internal tetragonal relaxations, and (c) (1 1 1) SL with cell internal trigonal relaxations. The labels P and As denote the atom on which localization is largest. Conduction-band levels were shifted upwards by 0.75 eV. From Dandrea and Zunger [80].

Next, require that GaAs and GaP be coherent with each other, but on a (1 1 1)-oriented substrate (fig. 20c). Equation (7) predicts different distortions c_{eq} since $q(\hat{G})$ appearing in this equation depends on the orientation $\hat{G}$ (viz. eqs. (8)–(13)). This leads to different band offsets: the Γ_{15} VBM now splits *trigonally* in $\Gamma_{3v} + \Gamma_{1v}$, with a Γ_{3v} valence-band offset of $0.14 + 0.43 = 0.57$ eV and a Γ_{1v} offset of $0.10 - 0.28 = -0.18$ eV. The lowest direct transition is now $\Gamma_{3v} \rightarrow \Gamma_{1c}$ at $1.65 - 0.14 = 1.51$ eV and the indirect transition to X_{1c} shifts from 2.00 eV in the cubic case (fig. 20a), or 1.73 eV in the tetragonal case (fig. 20b), to $2.47 - 0.14 = 2.33$ eV in the trigonal case (fig. 20c). Thus, epitaxial coherence with a lattice-mismatched substrate can affect profoundly the band offsets as well as the nature of the lowest optical transitions.

4.3. *Effects of microscopic interfacial relaxations on heterojunction band offsets*

Our foregoing discussion focused on *homogeneous* deformations of the heterojunction partners, as predicted by continuum elasticity theory (eq. (7)) after the band offsets and interband transition energies. This neglects possible *microscopic* deformations near the interface. Indeed, the positions of nuclei near a coherent interface between two materials are not simply those positions obtained by layering strained assemblies of the

two constituents on top of each other, as predicted by eq. (7). Rather, nuclei near the interface are shifted by the charge redistribution caused by the interface itself. This is not described by continuum elasticity. In the case of coherent lattice-mismatched interfaces (i.e. strained-layer interfaces) the major cause of these interface-localized relaxations is strain relief: the decreased symmetry of atoms near the interface allows such atoms to shift position very slightly to bring their bond lengths and bond angles closer to their unstrained values. These interface-localized strain relieving relaxations were studied by Dandrea et al. for a large number of systems [87]. There we found that these relaxations are very large and can extend far from the interface in higher-index growth directions such as (1 1 0) and (2 0 1), but are considerably smaller and much more interface-localized in the (0 0 1) and (1 1 1) directions. The purpose of this section is to demonstrate the significance of these small relaxations within (0 0 1) interfaces to the band offset.

To this end we consider a (0 0 1) interface between AlP and InP. The calculated zinc-blende lattice constants of these materials are $a(\mathrm{AlP}) = 5.439$ Å and $a(\mathrm{InP}) = 5.831$ Å, resulting in a large 7% lattice mismatch. We consider an epitaxially constrained system where $a_{xy} = a_{\mathrm{avg}} = 5.635$ Å. In such a system, elasticity theory (eq. (7)) predicts that the tetragonal deformations result in the following c/a ratios: $\eta_{\mathrm{elast}}(\mathrm{AlP}) = 0.2325$ and $\eta_{\mathrm{elast}}(\mathrm{InP}) = 0.2685$. We thus begin our calculation of a sixteen layer (0 0 1) $(\mathrm{AlP})_4(\mathrm{InP})_4$ superlattice by layering eight AlP planes separated by $0.2325a_{xy}$ on top of eight InP planes separated by $0.2685a_{xy}$. This is the geometry predicted by continuum elasticity. To see if this geometry is stable, we now relax the quantum-mechanically calculated [35] Hellman–Feynman forces on this geometry, resulting in a zero-force equilibrium geometry. We find that the central six of the eight AlP layers are separated by $\eta_0(\mathrm{AlP}) = 0.2323$ and the central six of the eight InP layers are separated by $\eta_0(\mathrm{InP}) = 0.2690$, demonstrating the accuracy of the elasticity estimate for the *interior* of the films. However, the AlP and InP layers immediately next to the interface have layer spacings about 1.5% less than and greater than, respectively, their values far from the surface. These are the small interface-localized relaxations referred to above. They are plotted throughout the sixteen layer supercell in fig. 21a, which shows the normalized z-axis displacement for layer i $\Delta\eta_i = (\eta_i - \eta_0)/\eta_0$, where η_i is c/a for layer i and η_0 is the value at the film's center. There it can be seen how rapidly the layer spacings approach their far-from-the-interface values: after the 1.5% variation immediately next to the interface, subsequent layer spacings vary by less than 0.1%. The major relaxation is simply a shift of the interface P atoms (those in the configuration In–P–Al) away from In and toward Al. The small relaxations seen in fig. 21a are actually brought about by the quest for the atoms near the interface for *strain-relief*: the relaxations result in the In–P and Al–P bond lengths at the interface to be nearer their unstrained zinc-blende values.

In order to determine the effects of these small interface-localized relaxations on the band offset, we perform a third calculation, where eight AlP layers separated by $\eta_0(\mathrm{AlP})$ are stacked on top of eight InP layers separated by $\eta_0(\mathrm{InP})$. The atomic coordinates of this geometry differ from those of the equilibrium zero-force geometry only with respect to atomic planes at the interface. In order to find the change in band offset induced by the interface-localized relaxations, we need only subtract the potential of the unrelaxed

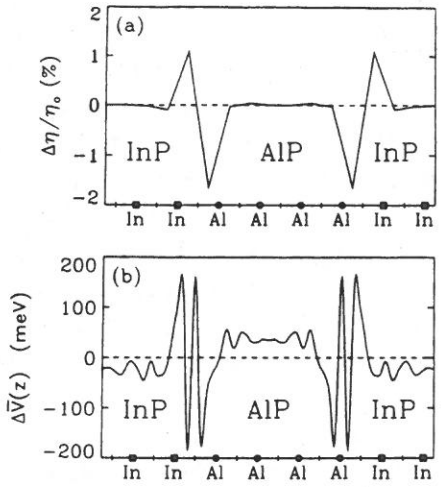


Fig. 21. (a) Calculated variations in interlayer spacings throughout the (0 0 1) AlP/InP interface. η_0 is the layer spacing (in units of a_{xy}) far from the interface: $\eta_0(\text{InP}) = 0.2690$ and $\eta_0(\text{AlP}) = 0.2323$. (b) Change in the planar-averaged potential $\bar{V}(z)$ between the equilibrium (zero-force) geometry shown in (a) and the constant layer spacing geometry. Note that the interface-localized relaxations induce a 58 meV dipole that raises the AlP part of the junction. From Dandrea et al. [84].

geometry from that of the relaxed geometry. This is done in fig. 21b, where it is seen that the interface-localized relaxations create a dipole that raises the AlP half of the junction by 58 meV. In order to demonstrate that the major cause of these interface-localized relaxations is interfacial strain due to lattice mismatch, a similar calculation as above was also performed for the nearly lattice-matched cases of AlAs/GaAs and $Ga_{0.47}In_{0.53}As$/InP. In both cases the relaxations were much smaller than in the AlP/InP case, and their effect on the band offset was found to be less than 10 meV. Our conclusion, therefore, is that obtaining the correct zero-force equilibrium geometry in strained-layer heterojunctions is crucial for calculating accurate offsets and accurate changes in offsets with changes in interfacial composition.

4.4. *Epitaxy-induced indirect-to-direct band gap crossover*

The epitaxially induced changes in band structures (section 4.1) and band offsets (sections 4.2 and 4.3) can lead to dramatic changes in the electronic structure. A particularly striking prediction [80] is the crossover from an *indirect* band gap for $(GaP)_1/(GaAs)_1$ superlattices on a $GaAs_{0.5}P_{0.5}$ substrate to a *direct* band gap for the film grown on a GaAs substrate. I next describe the general underlying physical principles that could lead to such an effect. I use the above-mentioned superlattice film as a particular example illustrating a rather general principle.

In these short-period superlattices (SLs), it is most natural to think of SL states as evolving from the states of a virtual crystal approximation (VCA) binary $Ga\langle PAs\rangle$ compound whose unit cell is commensurate with that of the SL. Figures 18c and 18d show the pseudopotential-calculated VCA band-edge energies as a function of the

lattice parameter a_s parallel to the substrate. In forming an SL on a given substrate, these VCA states are subjected to (i) folding into the SL Brillouin zone as well as (ii) interactions between the folded states through the perturbing potential $\delta V = V_{SL} - V_{VCA}$, which can split and shift them. Since GaP and GaAs have a 3.7% lattice mismatch, the perturbation changing the VCA into the SL can be broken into a chemical piece δV_{chem} (where the virtual $\langle AsP \rangle$ anion is changed to As and P in the respective halves of the SL) and a structural piece δV_{struc} (where the two halves of the SL relax to their equilibrium epitaxial-layer spacings). As there are exactly two VCA $Ga\langle AsP \rangle$ unit cells in each SL cell, the SL *translational* symmetry requires that each of the SL states at wavevector $\bar{K}$ be composed of (generally many) VCA bands coming from just *two* k-points in the VCA Brillouin zone. Denoting SL states by an overbar, these folding relationships for the principal band extrema in the (0 0 1) SL are

$$\bar{\Gamma} = \frac{2\pi}{a}(0,0,0) \longleftrightarrow \Gamma + X^z,$$

$$\bar{M} = \frac{2\pi}{a}(1,0,0) \longleftrightarrow X^x + X^y,$$

$$\bar{R} = \frac{2\pi}{a}(\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2}) \longleftrightarrow K^{11\bar{1}} + L^{111}, \tag{49}$$

while for the (1 1 1) SL they are

$$\bar{X} = \frac{2\pi}{a}(0,0,0) \longleftrightarrow X^z + L^{11\bar{1}},$$

$$\bar{\Gamma} \longleftrightarrow \Gamma + L^{111}. \tag{50}$$

Note that (even without strain) the three-fold degenerate X valley of zinc-blende splits in the (0 0 1) SL into X^z (which folds into $\bar{\Gamma}$) and $X^x + X^y$ (which fold into $\bar{M}$), while in the (1 1 1) SL the four L valleys split into L^{111} ($\bar{\Gamma}$-folding) and $L^{1\,1\,\bar{1}}$ (folds to $\bar{X}$). Clearly the SL will be direct at $\bar{\Gamma}$ only if the $\bar{\Gamma}$-folding states are lower in energy than non-$\bar{\Gamma}$-folding states. Consider the case $a_s = a(GaAs_{0.5}P_{0.5}) \equiv \bar{a}$ where the pertinent VCA states (denoted by angular brackets) correspond to cubic materials with unsplit states (figs. 18c and 18d). Since the $\langle X_{3c} \rangle$ state is higher in energy (fig. 18c), only the $\bar{\Gamma}_{1c}$ and $\bar{M}_{1,2}$ states occur near the CBM of a (0 0 1) SL. Turning on first the chemical perturbation $\delta V_{chem}(r)$ leaves the system cubic (fig. 20a). The non-degenerate $\langle \Gamma_c \rangle$ and $\langle X^z_{1c} \rangle$ VCA states interact (in second order) producing the SL states $\bar{\Gamma}^{(1)}_{1c}$ and $\bar{\Gamma}^{(2)}_{1c}$ (As-like and P-like, respectively) while the VCA degenerate $\langle X^{xy}_{1c} \rangle$ states split (in first order) into the SL states $\bar{M}_{1c}$ and $\bar{M}_{2c}$. The CBM is the $\bar{\Gamma}^{(1)}_{1c}$ state (fig. 20a), hence the gap is *direct*. Although not shown, this occurs also in the cubic (1 1 1) SLs. Note that $(AlAs)_p(GaAs)_p$, being lattice-matched ($\delta V_{str} = 0$), was also predicted to have a direct gap in the (1 1 1) orientation.

This situation changes when we turn on the structural perturbation $\delta V_{str}(r)$, which causes tetragonal distortions of $c/a = 0.966$ and 1.034 in the GaP and GaAs halves,

respectively. As seen in fig. 20b, these tetragonal deformations cause $X^z - X^{xy}$ splittings of 0.22 and 0.24 eV, causing the $\bar{\Gamma}_{1c}^{(1)}$ state to change from 85% As localized to 85% P localized (due to the X_{1c}^z offset now being lower in the P half). Furthermore, the stronger perturbation causes a much larger splitting of $X_{1c}^x + X_{1c}^y$ into $\bar{M}_{1c}$ and $\bar{M}_{2c}$; the relaxed SL is now *indirect* at $\bar{M}_{2c}$ because of the large first-order matrix element $\langle X_{1c}^x | \delta V | X_{1c}^y \rangle$. A similar situation occurs in the (1 1 1) SL on $a_s = \bar{a}$ (fig. 20c), where the $L_{1c}^{1\bar{1}\bar{1}}$-derived SL $\bar{X}_{1c}$ state forms the CBM. Hence, both (0 0 1) and (1 1 1) SLs on $a_s = \bar{a}$ are indirect because of δV_{str}.

This analysis reveals the way to produce a direct band gap SL. Since the structural perturbation alone converted the SLs on $a_s = \bar{a}$ to indirect (compare fig. 20a with 20b), we need to structurally perturb those VCA states which form the indirect band edges (X_{1c}^{xy} for (0 0 1), $X^z + L^{1\bar{1}\bar{1}}$ for (1 1 1)) so that they are displaced to higher energies. The VCA deformation potentials of figs. 18c and 18d show the way: increase the substrate lattice parameter a_s away from $\bar{a}$. The *negative* biaxial deformation potentials for these states increase their energies, whereas the $\bar{\Gamma}$-folding $\langle \bar{\Gamma}_{1c} \rangle$, $\langle X_{1c}^z \rangle$, and $\langle L^{111} \rangle$ states, having positive biaxial deformation potentials, are lowered. *Indeed, self-consistent SL calculations for* $a_s = a_{GaAs}$ (figs. 18a, 18b) *show that they are direct.* Examination of the calculated dipole matrix elements squared (denoted as f in figs. 18a and 18b) reveals that the (1 1 1) SL is "pseudodirect" (greater $\langle L_{1c} \rangle$ folded character than $\langle \Gamma_{1c} \rangle$ character, hence small f) while the (0 0 1) SL is strongly direct (large f). The predicted band gaps (correcting for spin–orbit effects) are 1.78 eV (700 nm) for $(GaP)_1/(GaAs)_1$ on GaAs (0 0 1), while for the (1 1 1) SL the predicted gap was 1.64 eV (760 nm). Recently, Takanohashi and Ozaki [88] measured the band gap of $(GaP)_n/(GaAs)_n$ on GaAs (0 0 1), finding a *direct* gap of 1.77 eV for $n = 1$. This is particularly interesting, given that the equivalent random alloy $GaAs_{0.5}P_{0.5}$ has an *indirect* band gap.

The phenomenon described above is rather general. Another manifestation of it is the prediction [89] of indirect band gap of short-period Si_nGe_n(0 0 1) superlattices on an Si substrate, but a quasi-direct gap on a Ge substrate. Figure 22 illustrates this case. It shows that epitaxial strain splits the conduction band at X into $X_c^\perp$ and $X_c^\parallel$; the valence-band maximum at Γ splits into $\Gamma_c^\perp$ and $\Gamma_c^\parallel$ (here $\perp$ and $\parallel$ denote directions perpendicular and parallel to the substrate interface, respectively). For compressive strain (an Si substrate, figs. 22a and 22b), $X_c^\parallel$ is below $X_c^\perp$ and $\Gamma_c^\perp$ is below $\Gamma_c^\parallel$, but for tensile strain (a Ge substrate, figs. 22e and 22f), $X_c^\perp$ is below $X_c^\parallel$ and $\Gamma_c^\parallel$ is below $\Gamma_c^\perp$. By symmetry, only $X^\perp$ folds into $\bar{\Gamma}$ for (0 0 1)-oriented superlattices. The central point here is that, to make the SL direct, one must invert the order of $X^\perp$ and $X^\parallel$: if the non-folding $X^\parallel$ is below the folding $X^\perp$ (Si substrate), the resulting SL is indirect, while if $X^\perp$ is below $X^\parallel$, then the SL is direct. Consequently, *on an Si substrate* we have the $X^\perp$-derived direct state $\bar{\Gamma}(X_{c2})$ at 1.26 eV (transition A in fig. 22c observed [90, 91] at 1.24 eV), while the $\Delta_c^\parallel$-derived transition is at a lower energy of 0.9 eV (transition I in fig. 22c, observed [90, 91] at 0.8 eV). By contrast, *on a Ge substrate*, the $X^\perp$-derived direct state $\bar{\Gamma}(X_{c2})$ at 0.97 eV (transition A in fig. 22d) is below the $\Delta_c^\parallel$-derived indirect state (transition I in fig. 22d) at a higher energy of 1.33 eV. This shows that changing c/a from > 1 (Si substrate) to $c/a < 1$ (Ge substrate) inverts the order of $X^\perp$ and $X^\parallel$ (much like the situation illustrated in fig. 18c for X^z versus X^{xy}), thus altering the band gap from indirect to direct. This general principle could be applied to many other systems.

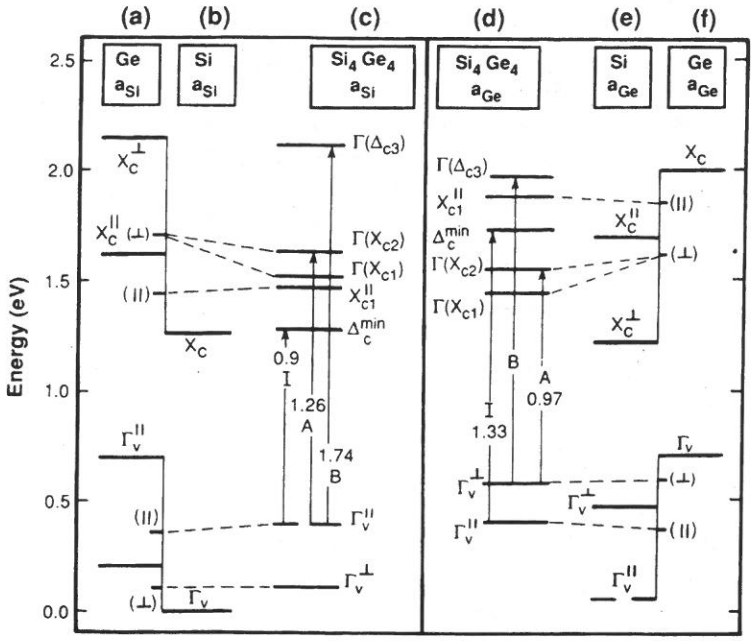


Fig. 22. Calculated electronic energy levels of the Si_4Ge_4 (0 0 1) superlattice on an Si substrate (a, b, c) and on a Ge substrate (d, e, f). Parts (a, b) and (e, f) give the band offsets. From Froyen, Wood and Zunger, Phys. Rev. Lett. 62 (1989) 975. See also ref. [89].

4.5. *Electronic properties of epitaxially stabilized novel compounds*

We have described in sections 2 and 3 how coherent epitaxy (with or without surface effects) can stabilize crystal structures that are unstable in bulk form. Naturally, hand in hand with a new structural form of a compound we expect new materials properties. Froyen et al. [34a] described the electronic properties of epitaxially stabilized zinc-blende MgS (fig. 6) and NaCl (fig. 7), whereas Wei and Zunger [21b] described the properties of epitaxially stabilized zinc-blende MnTe. These properties differ from those of the bulk-stable structures. This highlights the fact that one could use epitaxy to design material phases having qualitatively different properties than those grown in bulk form.

In section 3.3.1 we showed how coherent epitaxial effects can stabilize the bulk-unstable ordered chalcopyrite structure of III–V alloys (see fig. 9), while section 3.3.2 showed how the combination of coherent epitaxy and surface effects stabilize the CuPt structure in such alloys. This modifies significantly the band structure. As we did before, we first clarify the *physical principles* involved and then give calculated results.

Figure 23 (Wei and Zunger [92a]) shows how the band gaps change if one forms an epitaxially ordered (1 1 1) compound (CuPt, left) or a (0 0 1)-ordered compound (CuAu, right). The center part of fig. 23 shows the calculated cubic band offsets of GaAs/GaSb. When GaAs and GaSb are layered to form the CuPt compound, the Γ_{1c} and L_{1c} states fold into $\bar{\Gamma}_{1c}$, giving two states: $\bar{\Gamma}_{1c}(L_{1c})$ and $\bar{\Gamma}(\Gamma_{1c})$ (the quantity in parenthesis denotes the zinc-blende state folded in). The interaction between these folded states

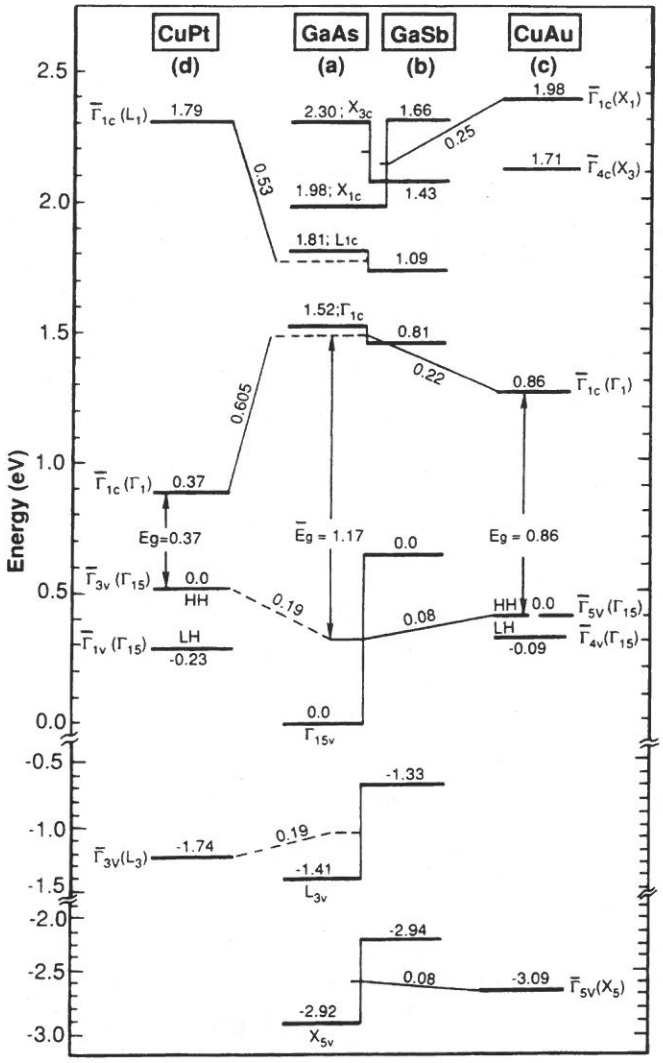


Fig. 23. Parts (a) and (b) show the energy levels of GaAs and GaSb, put on a common energy scale (obtained from the calculated valence-band offset). Parts (c) and (d) show the energy levels of Ga_2AsSb in the CuAuI and CuPt structure, respectively. Note how level repulsion displaces in (c) the $\{\bar{\Gamma}_{5v}(\Gamma_{15v}), \bar{\Gamma}_{5v}(X_{5v})\}$ and the $\{\bar{\Gamma}_{1c}(\Gamma_{1c}), \bar{\Gamma}_{1c}(X_{1c})\}$ pairs relative to their well centers. A stronger effect is seen in (d) for the CuPt structure, where $\{\bar{\Gamma}_{3v}(L_{3v}), \bar{\Gamma}_{3v}(\Gamma_{15v})\}$ and $\{\bar{\Gamma}_{1c}(\Gamma_1), \bar{\Gamma}_{1c}(L_1)\}$ are strongly shifted. From Wei and Zunger [92a].

repels the lower one, $\bar{\Gamma}_{1c}(\Gamma_{1c})$, downwards, thus contributing to a *reduction* in band gap upon ordering. The valence-band maximum Γ_{15v} splits upon ordering into $\bar{\Gamma}_{3v}(\Gamma_{15v}) + \bar{\Gamma}_{1v}(\Gamma_{15v})$. The $\bar{\Gamma}_{3v}(\Gamma_{15v})$ state interacts with the folded-in valence-band $\bar{\Gamma}_{3v}(\Gamma_{3v})$ state. This leads to an upwards displacement of the new VBM state $\bar{\Gamma}_{3v}(\Gamma_{15v})$. This also contributes to a reduction in band gap upon ordering. Thus, the average band gap $\bar{E}_g = 1.17$ eV of the GaAs + GaSb constituents is reduced to 0.37 eV upon ordering into the CuPt structure.

When (0 0 1) ordering is considered, the X_1 point rather than the L_1 point folds into $\bar{\Gamma}$. This leads to two conduction bands $\bar{\Gamma}_{1c}(\Gamma_{1c})$ and $\bar{\Gamma}_{1c}(X_{1c})$. Again, these repel each

other, leading to a lowering of the CBM $\bar{\Gamma}_{1c}(\Gamma_{1c})$. Here, however, the lowering (0.22 eV) is less than in CuPt ordering (0.605 eV), since the *initial* $\Gamma_{1c}-L_{1c}$ energy difference is smaller than the $\Gamma_{1c}-X_{1c}$ energy difference. Thus, (0 0 1) ordering reduces the band gap from the (binary-averaged) value of 1.17 to 0.86 eV, considerably less than for (1 1 1) ordering.

This discussion shows that the reduction in band gaps upon epitaxial ordering depends sensitively on the layer orientation. Wei and Zunger [92b] calculated the band structures of many 50%–50% III–V or II–VI alloys in various ordered structures (shown in fig. 9), including the CuAu, chalcopyrite, and CuPt. Figure 24 depicts their results for the direct $\Gamma_{15v}\rightarrow\Gamma_{1c}$ gaps of such systems. For comparison, they give the results for the random alloy. In all cases a lattice-matched substrate is assumed. We see that surface + epitaxy-induced ordering can lead to a rich repertoire of electronic properties even at a fixed composition. The *order* of direct band gaps is generally

$$E(\text{CuPt}) < E(\text{CuAu}) < E(\text{random}) \leqslant E(\text{chalcopyrite}). \tag{51}$$

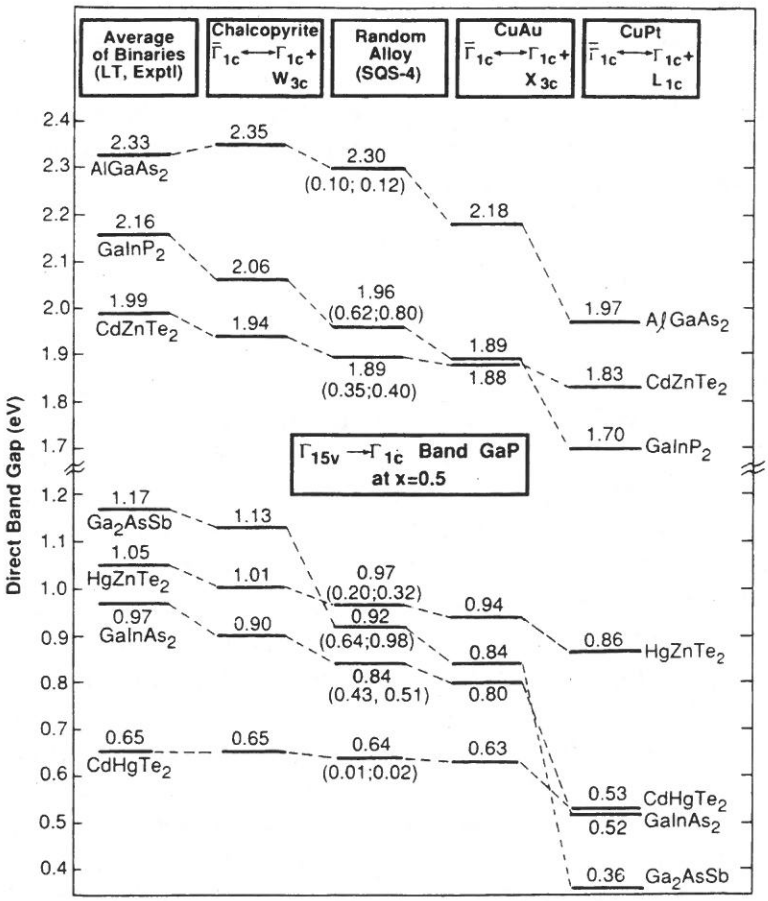


Fig. 24. Calculated direct band gaps of a series of 50%–50% alloys in different structures (depicted in fig. 9) and for the random alloy. From Wei and Zunger [92b].

We see therefore that ordering in the (1 1 1) structure (CuPt) or in the (0 0 1) structure (CuAu) generally leads to a *reduction* in band gap relative to the random alloy. Such effects were noted experimentally in $Ga_{0.5}In_{0.5}P$ ordered in the CuPt structure, e.g., by Kondow et al. [93], Nishino et al. [94], Alonso et al. [95], and Kanata et al. [96]. Note that the ordering attained in these samples is not perfect: successive planes along $\langle 1\,1\,1 \rangle$ are *not* pure GaP and pure InP. Wei et al. [97] have recently shown how imperfect CuPt ordering (measured by the long-range order parameter of fig. 15) affects the band gap reduction. They find that the band gap of an imperfectly ordered sample is related to that of the ideally random alloy (long-rang order parameter $\eta = 0$) and to a perfectly ordered CuPt alloy ($\eta = 1$) by

$$E_g(\eta) = E_g(\eta = 0) + \eta^2 [E_g(\eta = 1) - E_g(\eta = 0)]. \tag{52}$$

Hence, if $\eta < 1$, eq. (52) and the results of fig. 23 can be used to deduce $E_g(\eta)$.

The band-gap reduction upon ordering of the alloy into the CuPt structure can have interesting technological implications. For example, Wei and Zunger [98] predicted that CuPt ordering in InAsSb or GaInSb will reduce their band gaps significantly so as to produce a potentially useful far-infrared III–V material. Recently, the Sandia group (Kurtz et al. [99]) have reported successful MBE growth of spontaneously ordered InAsSb layers exhibiting emission near 9 μm, confirming the above expectations. Superlattices and quantum-well structures consisting of ordered and disordered layers were fabricated. The interfaces in these structures should be extremely sharp since the ordered and disordered layers have the same composition.

5. *Summary*

Coherent epitaxy leads to four simple changes in a film: (i) the interlayer spacings perpendicular to the substrate are deformed (eq. (7)), (ii) the E versus a_s energy curve flattens (eq. (14)), (iii) the alloy reference energy changes from that of *equilibrium* constituents to *strained* constituents (eq. (25)), and (iv) new, structure-dependent geometric factors relating the cell volume to the substrate lattice constant are introduced (eq. (22)). These nearly trivial changes affect profoundly the thermodynamics of coherent epitaxial films, leading to the possibility of stabilization of new *compounds* (section 2), enhanced *alloy* solubility (sections 3.1, 3.2), formation of chalcopyrite-type "ordered alloys" (section 3.3.1), composition pinning (section 3.3.3), and, in combination with surface relaxation effects, to spontaneous CuPt ordering (section 3.3.2). We have discussed the physical basis of these changes starting from ideas of simple continuum elasticity and introducing a microscopic statistical-mechanics viewpoint when the latter is insufficient. All of the above-mentioned *structural* modifications lead to important *electronic* consequences, including the formation of novel materials with new band gaps, the conversion of indirect-gap to direct-gap materials, and tailoring of the heterojunction band offsets. Many of the fascinating structural and crystallographic consequences of epitaxy predicted here are as yet unexplored. These constitute potentially interesting research opportunities.