Chapter 4 Ordering and Clustering

In this chapter, we discuss the tendency of alloy phases, constrained to grow as epitaxial thin films, to order and cluster. We would like to know whether, during MBE of alloy phases, the individual components will tend on a microscopic scale to attract or repel each other, so that there is shortrange order. We would also like to know whether the individual components will tend on a macroscopic scale to cluster into ordered or disordered phases of particular stoichiometries.

For concreteness, our discussion will center on "pseudobinary" III/V alloys — alloys composed of binary mixtures of two distinct III/V compounds. These alloys are exceedingly useful to device engineers because their lattice constants and electronic properties can be tuned continuously by adjusting the relative fractions of the two III/V compounds. These alloys are also characterized by positive enthalpies of mixing, and hence have a tendency to "unmix."¹ Those enthalpies of mixing originate mainly from microscopic strain caused by the different bond lengths of the two III/V compounds. Therefore, we begin the chapter by describing, in Section 4.1, how to estimate the strain in microscopic clusters using what are known as "valence force field" (VFF) models. If these microscopic clusters are embedded in an epitaxial thin film on a substrate with a different lattice constant, then they will also be "externally" strained. In Section 4.2, we discuss how to estimate that external strain.

In Section 4.3, we introduce a powerful technique, the cluster variation method, for building a macroscopic description of alloy thermodynamics from statistical combinations of such microscopic clusters. In Section 4.4, we apply this method in an approximate way to $In_{1-x}Ga_xAs$, a pseudobi-

¹E.K. Müller and J.L. Richards, "Miscibility of III-V semiconductors studied by flash evaporation," J. Appl. Phys. **35**, 1233 (1964).

nary alloy of current technological interest. We will find that the thermodynamic properties of $In_{1-x}Ga_xAs$ depend greatly on whether the alloy is coherent or incoherent with the substrate, i.e., on whether the interface between the epitaxial film and the substrate is crystallographically perfect or not.² If the alloy is incoherent with the substrate, then it is free to adopt the in-plane lattice constant that minimizes its free energy. If the alloy is coherent with the substrate, then it must adopt the in-plane lattice constant of the substrate; the resulting elastic strain energy can increase its overall free energy significantly.

In fact, such coherency constraints greatly suppress the tendency for alloys to separate into their pure-component "endpoint" phases, and at the same time greatly enhance their tendency to form ordered compounds at certain stoichiometric compositions. These tendencies can be understood quantitatively from the full cluster variation method calculation, but they can also be understood semiquantitatively through simpler semi-empirical models. We end the chapter, therefore, with a simple analytical treatment in Section 4.5 of coherency-constrained clustering and ordering.

4.1 Microscopic Strain

Let us start, in this section, by discussing microscopic strain in pseudobinary III/V alloys. We begin, in Subsection 4.1.1, by introducing a simple bond stretching and bond bending force field model for calculating the equilibrium atomic positions of a small alloy cluster. Then, in Subsection 4.1.2, we use those atomic positions to estimate the strain energy, which is the dominant contribution to the enthalpy of mixing.

4.1.1 Virtual Crystals and Covalent Radii

Let us begin, in this subsection, by calculating the microscopic bond distortions that occur when two III/V compounds are mixed. For concreteness, let us consider GaAs and InAs. Bulk alloys in this system are known to obey Vegard's law quite accurately: their *overall* lattice constants are the averages of the bulk GaAs and InAs lattice constants, weighted by mole fraction. If we imagine the alloy to be a "virtual crystal," in that each atom sits on geometrically perfect zincblende lattice sites,³ then its lattice

²D.M. Wood and A. Zunger, "Epitaxial effects on coherent phase diagrams of alloys," *Phys. Rev.* **B40**, 4062 (1989).

³L. Nordheim, "Electron theory of metals," Ann. Phys. (Leipzig) 9, 607 and 641 (1931).

constant can be expressed as

$$a_{\rm VC} = (1-x)a_{\rm GaAs,o} + xa_{\rm InAs,o},\tag{4.1}$$

where x is the InAs fraction in the alloy, and $a_{\text{GaAs,o}}$ and $a_{\text{InAs,o}}$ are the equilibrium lattice constants of (unstrained) bulk GaAs and InAs, respectively. Indeed, measurements⁴ show that the *second*-nearest-neighbor distances between group III atoms (or between group V atoms) in the lattice are very nearly those — $a_{\text{VC}}/\sqrt{2}$ — expected for such virtual crystals.

In contrast, however, first-nearest-neighbor distances between group III and group V atoms deviate significantly from those — $\sqrt{3}a_{\rm VC}/4$ — expected for such virtual crystals. Instead, Ga–As bonds are shorter, and In–As bonds are longer, than the virtual crystal bonds. That this is so is not unexpected, since the Ga–As bond in bulk GaAs is shorter than the In–As bond in bulk InAs, so in some sense the As "prefers" to be nearer to Ga than to In atoms. Indeed, one might imagine that, instead of occupying virtual crystal lattice sites, the atoms would occupy sites such that the bulk Ga–As and In–As bond lengths, and the associated "covalent radii" of the Ga, As, and In atoms, were preserved.⁵

To see which extreme of behavior is closer to the truth, consider the 5-atom $In_{0.5}Ga_{0.5}As$ tetrahedron at the right of Figure 4.1. This tetrahedron is one of the five distinct tetrahdra shown at the top of Figure 4.2 from which, as discussed in Section 4.4, an $In_{1-x}Ga_xAs$ alloy of arbitrary composition may be constructed. On the one hand, if the central As atom occupies the geometric center of the tetrahedron, then the tetrahedral bond angles associated with the sp^3 hybridized bonds can be preserved, but at the expense of InAs bonds that are too short and GaAs bonds that are too long. On the other hand, if the central As atom moves down slightly, then the InAs and GaAs bonds can approach their bulk equilibrium values, but at the expense of In-As-In bond angles that are too acute and Ga-As-Ga bond angles that are too obtuse.

The "elastic" energies associated with these kinds of distortions are often quantified using what are known as valence force field (VFF) models,⁶ in which the energies of individual bonds and bond angles are considered to be independent of each other. In the most popular representation for diamond-

⁴J.C. Mikkelsen, Jr., and J.B. Boyce, "Extended x-ray-absorption fine-structure study of GaInAs random solid solutions," *Phys. Rev.* **B28**, 7130 (1983).

⁵L. Pauling and M.L. Huggins, "Covalent radii of atoms and interatomic distances in crystals containing electron-pair bonds," Z. Kristallogr. Kristallgeom. Kristallphys. Kristallchem. 87, 205 (1934).

⁶M.J.P. Musgrave and J.A. Pople, "A general valence force field for diamond," *Proc.* Roy. Soc. London A268, 474 (1962).



Figure 4.1: Valence force field potential energies versus As position within a 5atom $In_{0.5}Ga_{0.5}As$ tetrahedron. Left and right panels show contributions due to bond bending and bond stretching forces; center panel shows the sum of the two contributions. The geometric center of the tetrahedron is at z = 0; the actual As position is shifted downward toward the Ga atoms. The predictions of the virtual crystal (VCA) and covalent radius (CRA) approximations discussed in the text are also shown.

structure semiconductors, the "Keating potential,"⁷ the stretching energy associated with bond i is proportional to the squared deviations of the squared actual bond length from the squared equilibrium length,

$$u_{\text{str},i} = \frac{3}{8} \alpha_i \frac{\left(d_i^2 - d_{i,o}^2\right)^2}{d_{i,o}^2},$$
(4.2)

and the bending energy associated with adjacent bonds i and j is proportional to the squared deviations of the dot products of actual adjacent bonds from the dot products of the equilibrium bonds,

$$u_{\mathrm{bnd},ij} = \frac{3}{8} \frac{\beta_i + \beta_j}{2} \frac{\left(\mathbf{d}_i \cdot \mathbf{d}_j - \mathbf{d}_{i,\mathrm{o}} \cdot \mathbf{d}_{j,\mathrm{o}}\right)^2}{d_{i,\mathrm{o}} d_{j,\mathrm{o}}}.$$
(4.3)

The two microscopic stretching and bending force constants, α and β , are assumed sufficient to characterize completely the microscopic elastic behavior of both the pure and mixed III-V compounds. Moreover, they can be used to predict various macroscopic elastic phenomena, and hence can be deduced from bulk elastic constants. The most commonly used values are listed in Table 4.1 for a number of diamond-structure materials.

⁷P.N. Keating, "Effect of invariance requirements on the elastic strain energy of crystals with application to the diamond structure," *Phys. Rev.* **145**, 637 (1966).

Material	α	β	β/α	C ₁₁	C ₁₂	C_{44}
				(10^{10})	(10^{10})	(10^{10})
	(N/m)	(N/m)		$N/m^2)$	$N/m^2)$	$N/m^2)$
С	129.33	84.76	0.655	107.6	12.50	57.68
Si	48.50	13.81	0.285	16.57	6.39	7.96
Ge	38.67	11.35	0.294	12.89	4.83	6.71
AlSb	35.35	6.77	0.192	8.94	4.43	4.16
GaP	47.32	10.44	0.221	14.12	6.25	7.05
GaAs	41.19	8.95	0.217	11.81	5.32	5.92
GaSb	33.16	7.22	0.218	8.84	4.03	4.32
InP	43.04	6.24	0.145	10.22	5.76	4.60
InAs	35.18	5.50	0.156	8.33	4.53	3.96
InSb	29.61	4.77	0.161	6.67	3.65	3.02
ZnS	44.92	4.78	0.107	10.40	6.50	4.62
ZnSe	35.24	4.23	0.120	8.10	4.88	4.41
ZnTe	31.35	4.45	0.142	7.13	4.07	3.12
CdTe	29.02	2.43	0.084	5.35	3.68	1.99
CuCl	12.60	1.00	0.079	2.72	1.87	1.57

Table 4.1: Microscopic bond stretching (α) and bond bending (β) force constants deduced from macroscopic elastic constants (C_{11} , C_{12} , and C_{44}) of various cubic semiconducting materials.^{*a*}

^aR.M. Martin, "Elastic properties of ZnS structure semiconductors," *Phys. Rev.* B1, 4005 (1970).

To calculate the stretching energy of tetrahedra such as that shown at the right of Figure 4.1, we sum Equation 4.2 over the four bonds to the central As atom, divide by two because each bond is shared by two atoms, then multiply by two because there is a pair of atoms per tetrahedron:

$$u_{\rm str} = \sum_{i=1}^{4} u_{{
m str},i}.$$
 (4.4)

To calculate the bending energy of the tetrahedron, we sum Equation 4.3 over each distinct pair of adjacent bonds, and multiply by two because we have only accounted for the bonds centered on the group V atoms, but not those centered on the group III atoms:

$$u_{\text{bnd}} = 2 \sum_{i=1}^{4} \sum_{j < i} u_{\text{bnd}, ij}.$$
 (4.5)

For the particular tetrahedron shown at right in Figure 4.1, the total elastic energy, per atom-pair, is then

$$u_{(\bigstar)} = u_{\text{str}} + u_{\text{bnd}}$$

$$\approx \frac{3}{4} \left[\alpha_{\text{GaAs}} \frac{\left(d_{\text{GaAs}}^2 - d_{\text{GaAs,o}}^2\right)^2}{d_{\text{GaAs,o}}^2} + \alpha_{\text{InAs}} \frac{\left(d_{\text{InAs}}^2 - d_{\text{InAs,o}}^2\right)^2}{d_{\text{InAs,o}}^2} \right]$$

$$+ \frac{3}{4} \left[\beta_{\text{GaAs}} \frac{\left(d_{\text{GaAs}}^2 \cos^2 \theta_{\text{GaAsGa}} - d_{\text{GaAs,o}}^2 \cos^2 \theta_{\text{T}}\right)^2}{d_{\text{GaAs,o}}^2} \right]$$

$$+ \beta_{\text{InAs}} \frac{\left(d_{\text{InAs}}^2 \cos^2 \theta_{\text{InAsIn}} - d_{\text{InAs,o}}^2 \cos^2 \theta_{\text{T}}\right)^2}{d_{\text{InAs,o}}^2} \right]. \quad (4.6)$$

In this equation, the actual and equilibrium GaAs and InAs bond lengths are denoted d_{GaAs} , $d_{\text{GaAs,o}}$, d_{InAs} and $d_{\text{InAs,o}}$; the actual Ga–As–Ga and In–As–In bond angles are denoted θ_{GaAsGa} and θ_{InAsIn} ; and the ideal tetrahedral bond angle is $\theta_{\text{T}} = 2 \tan^{-1}(1/\sqrt{2}) \approx 109.47^{\circ}$. Note that we have used the symmetry of the tetrahedron to set $\theta_{\text{GaAsGa}} = \theta_{\text{AsGaAs}}$, $\theta_{\text{InAsIn}} = \theta_{\text{AsInAs}}$ and $\theta_{\text{GaAsIn}} = \theta_{\text{InAsGa}} \approx \theta_{\text{T}}$.

In terms of the vertical displacement, z, of the As atom from the geometric center of the tetrahedron, the actual GaAs and InAs bond lengths can be written as

$$d_{\text{GaAs}}^{2} = \left(\frac{a_{\text{VC}}}{2\sqrt{2}}\right)^{2} + \left(\frac{a_{\text{VC}}}{4} + z\right)^{2}$$
$$d_{\text{InAs}}^{2} = \left(\frac{a_{\text{VC}}}{2\sqrt{2}}\right)^{2} + \left(\frac{a_{\text{VC}}}{4} - z\right)^{2}, \qquad (4.7)$$

and the actual Ga-As-Ga and In-As-In bond angles can be written as

$$\cos(\theta_{\rm GaAsGa}/2) \approx \frac{(a_{\rm VC}/4) + z}{(\sqrt{3}a_{\rm VC}/4) + z/\sqrt{3}} \\
\cos(\theta_{\rm InAsIn}/2) \approx \frac{(a_{\rm VC}/4) - z}{(\sqrt{3}a_{\rm VC}/4) - z/\sqrt{3}},$$
(4.8)

where $a_{\rm VC}/2$ is the length of an edge of the cube circumscribing the tetrahedron.

Then, substituting back into Equation 4.6, we can calculate, as shown in the left three panels of Figure 4.1, the distortion energies as a function of z. The left and right panels show only the bending and stretching energies, respectively; the center panel shows their total. Those panels illustrate how the actual position of the As atom at the center of the tetrahedron is determined by a competition between bending and stretching forces. Given only bending forces, the virtual crystal approximation (VCA) holds, and bond angles are nearly ideally tetrahedral.⁸ Given only stretching forces, the covalent radius approximation (CRA) holds, and bond lengths are undistorted from the bulk pure component compounds. Given both forces, neither holds exactly, but, as can be seen, the CRA is the better approximation. In this pseudobinary III/V system, bending forces are about 5 times weaker than stretching forces, and bond lengths are very nearly preserved upon mixing. They deviate slightly, however, due to the "steric" constraints provided by bending forces.

To obtain an analytic form for the position of the As atom, we can expand Equations 4.6, 4.7, and 4.8 to second order in z, giving

$$\begin{aligned} u_{(\blacktriangle)} &\approx \frac{3}{2} \alpha_{\rm GaAs} \left[\frac{\sqrt{3}}{4} a_{\rm VC} - d_{\rm GaAs,o} + \frac{z}{\sqrt{3}} \right]^2 \\ &+ \frac{3}{2} \alpha_{\rm InAs} \left[\frac{\sqrt{3}}{4} a_{\rm VC} - d_{\rm InAs,o} - \frac{z}{\sqrt{3}} \right]^2 \\ &+ \frac{3}{8} \beta_{\rm GaAs} \left[\frac{-2}{3} \left(\frac{\sqrt{3}}{4} a_{\rm VC} - d_{\rm GaAs,o} \right) + \frac{2}{\sqrt{3}} z \right]^2 \\ &+ \frac{3}{8} \beta_{\rm InAs} \left[\frac{-2}{3} \left(\frac{\sqrt{3}}{4} a_{\rm VC} - d_{\rm InAs,o} \right) - \frac{2}{\sqrt{3}} z \right]^2. \end{aligned}$$
(4.9)

Then, solving for $\partial u_{(\bigstar)}/\partial z = 0$, the equilibrium position can be deduced to be

$$z_{\rm equ} = \left(\frac{-\sqrt{3}}{2}\right) \frac{\alpha_{\rm GaAs} + \alpha_{\rm InAs} - \beta_{\rm GaAs}/3 + \beta_{\rm InAs}/3}{\alpha_{\rm GaAs} + \alpha_{\rm InAs} - \beta_{\rm GaAs} + \beta_{\rm InAs}} (d_{\rm InAs,o} - d_{\rm GaAs,o}).$$

$$(4.10)$$

In the limit $\beta \to 0$, $z_{\rm equ} \to -(\sqrt{3}/2)(d_{\rm InAs,o} - d_{\rm GaAs,o})$, and the VCA holds; in the limit $\alpha \to 0$, $z_{\rm equ} \to -(d_{\rm InAs,o} - d_{\rm GaAs,o})/(2\sqrt{3})$, and the CRA holds.

To see how the bond lengths in these alloys depend on composition, similar calculations can be performed for 5-atom GaAs, $In_{0.25}Ga_{0.75}As$,

⁸The angles are not exactly tetrahedral because the Keating representation of the "valence forces" does not cleanly separate stretching from bending motions, since Equation 4.3 consists of deviations of dot products (rather than of angles) between adjacent bonds. Other representations do, but at the expense of not appearing to predict distortion energies as accurately [W.A. Harrison, *Electronic Structure and the Properties of Solids* (W.H. Freeman, San Francisco, 1980), pp. 193-197].



Figure 4.2: Measured and calculated Ga-As and In-As bond lengths in $In_{1-x}Ga_x$ As alloys. Experimental data are from (open circles) X-ray-absorption fine structure (EXAFS)^a measurements; calculations are based on valence-force-field potentials using either bond bending forces in the virtual crystal approximation (dashed line), bond stretching forces in the covalent radius approximation (dotted lines), or both (solid lines).^b

In_{0.75}Ga_{0.25}As and InAs tetrahedra, which are the most probable tetrahedra in the corresponding GaAs, In_{0.25}Ga_{0.75}As, In_{0.75}Ga_{0.25}As and InAs alloys. The results are shown in Figure 4.2. The dotted and dashed lines are the stretching-force-only (CRA) and bending-force-only (VCA) bond lengths; the solid lines are the stretching-force plus bending-force bond lengths. The predictions agree extremely well with the measurements shown as open circles. That agreement indicates that, consistent with more complete calculations,⁹ elastic energies dominate chemical energies in this alloy system. Indeed, this dominance appears to hold for most isovalent, though not for heterovalent, mixtures of semiconductors.¹⁰

^aJ.C. Mikkelsen, Jr., and J.B. Boyce, "Extended x-ray-absorption fine-structure study of GaInAs random solid solutions," *Phys. Rev.* **B28**, 7130 (1983).

^bT. Fukui, "Calculation of bond length in InGaAs ternary semiconductors," Jpn. J. Appl. Phys. 23, L208 (1984).

⁹T. Ito, "A pseudopotential approach to mixing enthalpies of III-V ternary semiconductor alloys," Jpn. J. Appl. Phys. 26, 256 (1987).

¹⁰W.A. Harrison and E.A. Kraut, "Energies of substitution and solution in semiconductors," *Phys. Rev.* B37, 8244 (1988).

4.1.2 Mixing Enthalpies

In Subsection 4.1.1, we calculated the microscopic distortions that minimize the sum of the bond stretching and bond bending energies of an As-centered cluster containing both Ga and In. In this subsection, we make use of those distortions to calculate the strain energy associated with the cluster, and then to estimate the mixing enthalpy associated with the alloy as a whole.

To obtain a simplified formula for the energy of the In_{0.5}Ga_{0.5}As tetrahedron, we insert the equilibrium position of the As atom given by Equation 4.10 into Equation 4.9. Then, approximating the individual bond stretching and bending force constants by their averages, $\overline{\alpha} \equiv (\alpha_{\text{GaAs}} + \alpha_{\text{InAs}})/2$ and $\overline{\beta} \equiv (\beta_{\text{GaAs}} + \beta_{\text{InAs}})/2$, we obtain, after some algebra,

$$u_{(\blacktriangle)} \approx \frac{\overline{\alpha}\overline{\beta}}{2(\overline{\alpha} + \overline{\beta})} (\Delta a_o)^2,$$
 (4.11)

where

$$\Delta a_{\rm o} \equiv a_{\rm InAs,o} - a_{\rm GaAs,o} = \frac{4}{\sqrt{3}} (d_{\rm InAs,o} - d_{\rm GaAs,o}). \tag{4.12}$$

The distortion energy of the tetrahedron calculated in this way is listed in Table 4.7 on page 132. The energy is proportional to the square of the difference, Δa_o , between the lattice parameters of the component compounds, precisely what one expects from a model based on linear elasticity. The effective spring constant, $1/[(1/\overline{\alpha}) + (1/\overline{\beta})]$, is the "parallel" sum of the individual stretching and bending force constants. Since, as mentioned above, $\overline{\beta}$ is approximately 5 times weaker than $\overline{\alpha}$, the effective spring constant is dominated by $\overline{\beta}$. In other words, as with all coupled spring systems, most of the energy is stored in the weaker and more deformed spring.

If we now imagine building a lattice solely out of $In_{0.5}Ga_{0.5}As$ tetrahedra, then Equation 4.11 can also be used to estimate the enthalpy of mixing of the $In_{0.5}Ga_{0.5}As$ alloy. On the one hand, it will be an overestimate: our simple calculation did not account for relaxation of the corner group III atoms of the tetrahedron away from their virtual crystal positions, which would decrease the tetrahedron energy. On the other hand, it will be an underestimate: as discussed later in Section 4.4, a real $In_{0.5}Ga_{0.5}As$ alloy at finite temperature would also contain some fraction of more highly deformed tetrahedra of other compositions, which would increase the energy of the alloy as a whole.

To see how well this estimate works, let us approximate the alloy as a strictly regular solution, and identify its interaction enthalpy at x = 1/2 with the VFF elastic energy of the In_{0.5}Ga_{0.5}As tetrahedron: $\Omega_{\rm VFF} \approx$

Material	Crystal	$a_{\rm o} \text{ or } a/c$	$\alpha_T = \partial \ln a_{\rm o} / \partial T$
	Structure	$(\text{\AA or \AA}/\text{\AA})$	$(10^{-6} \mathrm{K}^{-1})$
С	Diamond	3.56683	0.87 + 0.0092(T - 273)
Si	Diamond	5.43095	3.08 + 0.0019(T - 273)
Ge	Diamond	5.64613	6.05 + 0.0036(T - 273)
α -Sn	Diamond	6.48920	
SiC	Wurtzite	3.086/15.117	
BN	Zincblende	3.6150	
BP	Zincblende	4.5380	
AlP	Zincblende	5.4510	
AlAs	Zincblende	5.6605	3.40 + 0.0064(T - 273)
AlSb	Zincblende	6.1355	
GaN	Zincblende	3.189/5.185	
GaP	Zincblende	5.4512	5.81
GaAs	Zincblende	5.6533	5.35 + 0.0080(T - 273)
GaSb	Zincblende	6.0959	6.7
InP	Zincblende	5.8686	
InAs	Zincblende	6.0584	4.33 + 0.0038(T - 273)
InSb	Zincblende	6.4794	
ZnO	Rock Salt	4.580	
ZnS	Zincblende	5.420	6.70 + 0.0128(T - 313)
ZnS	Wurtzite	3.82/6.26	
CdS	Zincblende	5.8320	
CdS	Wurtzite	4.16/6.756	
CdTe	Zincblende	6.482	
CdSe	Zincblende	6.050	
PbS (Galena)	Rock Salt	5.9362	18.81 + 0.0074(T - 273)
PbTe (Altaite)	Rock Salt	6.4620	19.80

Table 4.2: Crystal structures, room-temperature lattice parameters and thermal expansion coefficients of various semiconductors.^a

^aAdapted from S.M. Sze, *Physics of Semiconductor Devices*, 2nd Ed. (John Wiley & Sons, New York, 1981), and R.S. Krishnan, R. Srinivasan and S. Devanarayanan, *Thermal Expansion of Crystals* (Pergamon Press, Oxford, 1979).

 $4u_{(\bigstar)}$. Then

$$\Omega_{\rm VFF} \approx \frac{2\overline{\alpha}\overline{\beta}}{\overline{\alpha} + \overline{\beta}} (\Delta a_o)^2.$$
 (4.13)

This equation can be used to estimate the elastic part of the regular solution interaction parameter for any pseudobinary mixture whose microscopic elastic constants and lattice parameters are known. Its predictions are shown in Figure 4.3 for a number of alloys, using the lattice parameters listed in Table 4.2 and the bond stretching and bending force constants listed in Table 4.1. Within the (fairly large) uncertainty in the values deduced from experimental measurements, the equation predicts the regular solution parameters surprisingly accurately. It represents the physical basis¹¹ for what is known as the Delta-Lattice-Parameter (DLP) model, originally based on the empirical observation that heats of mixing are approximately proportional to the squared mismatches between the lattice parameters of the constituent components.¹²

4.2 Macroscopic Strain

In Section 4.1, we noted that, from a microscopic point of view, pseudobinary III-V alloys can be viewed as a collection of elementary tetrahedra such as those shown in Figure 4.2. Except for the pure-component tetrahedra, none are perfectly tetrahedral: their bond lengths and angles deviate from the CRA lengths and VCA angles, respectively. These *internal* distortions give rise to the elastic strain energies listed in Table 4.7 on page 132 even in tetrahedra embedded in bulk alloys of the same overall composition as the tetrahedra themselves.

Superimposed on these internal distortions, however, are distortions due to *externally* imposed constraints on the dimensions of the tetrahedra. These constraints arise because the tetrahedra, each with an "ideal" dimension or shape, are all embedded in a macroscopic lattice whose unit cells have their own (and possibly different) average dimension or shape. In this section, we discuss these externally imposed distortions. Conceptually, they can be decomposed into two components: one that is mainly volumetric and one that is mainly distortional.

The volumetric component comes about either when alloys are grown in bulk form, or when epitaxial films are grown coherently on a latticematched substrate. Consider such an alloy, whose overall composition is $x_{\rm epi} = 0.5$, and whose mean (or virtual crystal) lattice parameter is given, using Equation 4.1, by

$$a_{\rm epi,o} = 0.5 a_{\rm GaAs,o} + 0.5 a_{\rm InAs,o}.$$
 (4.14)

¹¹P.A. Fedders and M.W. Muller, "Mixing enthalpy and composition fluctuations in ternary III-V semiconductor alloys," *J. Phys. Chem. Solids* **45**, 685 (1984); J.L. Martins and A. Zunger, "Bond lengths around isovalent impurities and in semiconductor solid solutions," *Phys. Rev.* **B30**, 6217 (1984)

¹²G.B. Stringfellow, "Calculation of regular solution interaction parameters in semiconductor solid solutions," J. Phys. Chem. Solids **34**, 1749 (1973).



Figure 4.3: Regular solution parameters for various pseudobinary alloys. Values plotted along the bottom axis were calculated using Equation 4.13; values plotted along the left axis are experimental measurements^a; values plotted along the right axis are the critical temperatures, deduced from Equation 3.38, above which the constituent components are fully miscible.

^aAdapted from G.B. Stringfellow, "Calculation of ternary and quaternary III-V phase diagrams," J. Cryst. Growth 27, 21 (1974).

The only tetrahedron whose "ideal" dimension is also given by Equation 4.14 is the $In_{0.5}Ga_{0.5}As$ tetrahedron, which will not be externally strained, and whose excess energy will be due solely to internal distortions. All other tetrahedra will have "ideal" dimensions different from that given by Equation 4.14. If embedded in the $x_{epi} = 0.5$ alloy, they will be constrained to occupy volumes different from their ideal volume, and will have additional energies due to the externally imposed volumetric distortions. For example, an $In_{0.75}Ga_{0.25}As$ tetrahedron has an ideal dimension (neglecting relaxations of corner atoms) of $0.25a_{GaAs,o} + 0.75a_{InAs,o}$, and must be compressed before it can fit into a $In_{0.5}Ga_{0.5}As$ lattice.

The distortional component comes about when epitaxial films are grown coherently on a lattice-*mismatched* substrate. Suppose, for example, that the substrate is a single (unstrained) crystal of bulk $In_{1-x_{sub}}Ga_{x_{sub}}As$ itself, whose In composition is x_{sub} and whose mean (or virtual crystal) lattice parameter is a weighted average of the two endpoint lattice parameters,

$$a_{\rm sub} = (1 - x_{\rm sub})a_{\rm GaAs,o} + x_{\rm sub}a_{\rm InAs,o}$$
(4.15)

As illustrated in the right half of Figure 4.4, if the epitaxial film is coherent with the substrate, then its lattice parameter parallel to the interface must be the same as that of the substrate, independent of the composition of the epitaxial film itself:

$$a_{\text{epi},\parallel} = a_{\text{sub}} = (1 - x_{\text{sub}})a_{\text{GaAs,o}} + x_{\text{sub}}a_{\text{InAs,o}}.$$
 (4.16)

In other words, there will be a parallel strain in the film of

$$\epsilon_{\rm epi,\parallel} \equiv 2 \frac{a_{\rm epi,\parallel} - a_{\rm epi,o}}{a_{\rm epi,\parallel} + a_{\rm epi,o}},\tag{4.17}$$

where

$$a_{\rm epi,o} = (1 - x_{\rm epi})a_{\rm GaAs,o} + x_{\rm epi}a_{\rm InAs,o}, \qquad (4.18)$$

is the equilibrium (unstrained) lattice parameter of the epitaxial film.

As illustrated in the left half of Figure 4.4, however, its lattice parameter in a direction perpendicular to the interface will not be the same as the equilibrium lattice parameters of either the substrate or the epitaxial film. If the film is locked to a substrate with a smaller lattice parameter, then the in-plane compressional "squeezing" will force its perpendicular lattice parameter to increase in order to preserve (approximately) its unit cell volume. If the film is locked to a substrate with a larger lattice parameter, then the in-plane tensile "stretching" will force its perpendicular lattice parameter to decrease, again in order to preserve (approximately) its unit cell volume.

To understand both the volumetric and distortional components of the externally imposed strains quantitatively, we write what is known as the generalized Hooke's law for cubic crystals,¹³

$$\begin{pmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \\ \tau_{xy} \\ \tau_{yz} \\ \tau_{zx} \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{12} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{44} \end{pmatrix} \begin{pmatrix} \epsilon_x \\ \epsilon_y \\ \epsilon_z \\ \gamma_{xy} \\ \gamma_{yz} \\ \gamma_{zx} \end{pmatrix}, \quad (4.19)$$

¹³See, e.g., A.J. Durelli, E.A. Phillips, and C.H. Tsao, Introduction to the Theoretical and Experimental Analysis of Stress and Strain (McGraw-Hill, New York, 1958), Chap. 4.



Figure 4.4: Perpendicular (left) and parallel (right) lattice parameters of $In_{1-x}Ga_xAs$ at composition x_{epi} grown coherently on substrates having the lattice parameters of bulk $In_{1-x}Ga_xAs$ at composition x_{sub} . The filled circles represent $In_{1-x}Ga_xAs$ grown on substrates with compositions $x_{sub} = 0, 1/2, 1$. The open circles represent $In_{1-x}Ga_xAs$ grown on "lattice-matched" substrates with compositions $x_{sub} = x_{epi}$, or, alternatively, to incoherent growth.

where the ϵ_i 's and σ_i 's are the normal strains and stresses, respectively, and the γ_{ij} 's and τ_{ij} 's are the shear strains and stresses, respectively.

If the epitaxial film and its substrate are oriented along one of the (100) cubic symmetry directions, then this equation reduces to

$$\begin{pmatrix} \sigma_{\rm epi,\parallel} \\ \sigma_{\rm epi,\perp} \end{pmatrix} = \begin{pmatrix} C_{11} + C_{12} & C_{12} \\ 2C_{12} & C_{11} \end{pmatrix} \begin{pmatrix} \epsilon_{\rm epi,\parallel} \\ \epsilon_{\rm epi,\perp} \end{pmatrix}.$$
(4.20)

If, in addition, the epitaxial film has a free surface, such that perpendicular stresses vanish, then

$$\sigma_{\rm epi,\perp} = 2C_{12}\epsilon_{\rm epi,\parallel} + C_{11}\epsilon_{\rm epi,\perp} = 0, \qquad (4.21)$$

and the perpendicular strain and lattice parameter of the film are

$$\epsilon_{\rm epi,\perp} = \frac{-2C_{12}}{C_{11}} \epsilon_{\rm epi,\parallel}$$

$$a_{\perp}(x_{\rm epi}, x_{\rm sub}) = a_{\rm epi,o} \frac{1 + \epsilon_{\rm epi,\perp}/2}{1 - \epsilon_{\rm epi,\perp}/2}.$$
(4.22)

On average, then, the unit cell of the epitaxial film has parallel and perpendicular dimensions given by Equations 4.16 and 4.22, respectively.

4.3. The Cluster Variation Method

Now consider the microscopic tetrahedra that are embedded within this epitaxial film. On average, they must be constrained to the same dimensions as the unit cell.¹⁴ However, each individual tetrahedron has its own "ideal" size, given approximately by Vegard's law:

$$a_{i,o} = (1 - x_i)a_{GaAs,o} + x_i a_{InAs,o},$$
 (4.23)

where x_i is the composition of the *i*th elementary tetrahedron. If they are all constrained to the average dimension of the unit cell of the epitaxial film, then they will be strained, according to

$$\begin{split} \epsilon_{i,\parallel} &= 2 \frac{a_{\text{epi},\parallel} - a_{i,\text{o}}}{a_{\text{epi},\parallel} + a_{i,\text{o}}} \\ \epsilon_{i,\perp} &= 2 \frac{a_{\text{epi},\perp} - a_{i,\text{o}}}{a_{\text{epi},\perp} + a_{i,\text{o}}}. \end{split}$$
(4.24)

The resulting strain energies of the various tetrahedra (per atom pair) due to these external constraints can then be approximated, through use of Equation 4.20, by^{15}

$$u_{i,\text{ext}} = \frac{1}{2} \left[2\sigma_{i,\parallel}\epsilon_{i,\parallel} + \sigma_{i,\perp}\epsilon_{i,\perp} \right]$$

= $(C_{i,11} + C_{i,12})\epsilon_{i,\parallel}^2 + 2C_{i,12}\epsilon_{i,\perp}\epsilon_{i,\parallel} + \frac{1}{2}C_{i,11}\epsilon_{i,\perp}^2,$ (4.25)

where the elastic constants of the individual tetrahedra can be taken to be Vegard's law averages of the elastic constants of the pure component binary alloys:

$$C_{i,11} = (1 - x_i)C_{\text{GaAs},0,11} + x_iC_{\text{InAs},0,11}$$

$$C_{i,12} = (1 - x_i)C_{\text{GaAs},0,12} + x_iC_{\text{InAs},0,12}$$
(4.26)

4.3 The Cluster Variation Method

In Sections 4.1 and 4.2, we explored the origin of elastic distortion energies in small microscopic tetrahedra such as those shown in Figure 4.2. In

 $^{^{14}}$ Note that the tetrahedra with more In atoms will be somewhat larger than the average, and those with fewer will be somewhat smaller. Nevertheless, we make the simplifying approximation, as we did in Section 4.1.1, that the virtual crystal approximation holds for *second*-nearest-neighbor distances, and that all tetrahedra have the same dimensions.

 $^{^{15}}$ We neglect, in this simple treatment, the strain-induced-splitting of the degeneracies of tetrahedra differing only by permutations of group III atoms, and treat all tetrahedra having the same number of In and Ga atoms to be the same.

this section, we ask: how can we use such microscopic information to deduce macroscopic quantities of interest, such as enthalpies and entropies of mixing, or tendencies toward short- and long-range ordering?

One classic approach to this problem is the cluster variation method (CVM),¹⁶ in which solids are built by statistically combining a finite number of independent, elementary clusters. In principal, the method may be made arbitrarily accurate by choosing arbitrarily large clusters. In practice, actual implementations of the method represent trade-offs between accuracy and speed.¹⁷ The larger the clusters, the less important the intercluster interaction energies are relative to intracluster energies, and the more accurate the assumption of cluster independence becomes. However, the larger the clusters, the more types of elementary clusters (of different composition) there will be, and the more time-consuming the combinatorics become.

In this section, we give a brief introduction to the cluster variation method. The method can be viewed as an increasingly accurate sequence of approximations, and so it is convenient to illustrate it by applying it to successively more complex structures: first alloys on 1D linear (in Subsection 4.3.1), then 2D triangular (in Subsection 4.3.2) and finally 3D zincblende (in Subsection 4.3.3) lattices.

The introduction given in this section is somewhat lengthy, both because the cluster variation method gives insight into so many aspects of alloy thermodynamics and because a comparable introductory treatment does not appear to exist elsewhere. However, it will not be necessary to understand the cluster variation method in detail in order to follow its application in Sections 4.4 and 4.5 to $In_{1-x}Ga_xAs$, a prototypical pseudobinary III/V alloy. The casual reader is advised to begin with Section 4.5.

4.3.1 1D Linear Lattice

We start, in this subsection, by illustrating the cluster variation method using a simple one-dimensional linear lattice. We consider, in turn, two possible ways of constructing this lattice. In the first way, the lattice is constructed from uncorrelated "points" of atoms, as shown in Figure 4.5. In the second way, the lattice is constructed from correlated "pairs" of atoms, as shown in Figure 4.6.

¹⁶R. Kikuchi, "A theory of cooperative phenomena," Phys. Rev. 81, 988 (1951).

¹⁷D.M. Burley, "Closed form approximations for lattice systems," in C. Domb and M.S. Green, Eds., *Phase transitions and critical phenomena* (Academic Press, London, 1972), Vol. 2, Chap. 9.



Figure 4.5: Construction of a 1D linear chain of points by the addition of a new node (open circle) to an existing lattice (filled circles).

Points

Consider first the lowest order "point" approximation, in which the largest clusters are the individual atoms themselves. We imagine building an ensemble of n linear chains, each composed of nodes which are either type A or type B atoms. If the overall fractions of A and B atoms in the ensemble of chains are x_0 and $x_1 = 1 - x_0$, then each node of the ensemble of chains will have x_0n A atoms and x_1n B atoms.

Now suppose we wish to add another node to this ensemble of n chains. Since the nodes are all independent, we are free to add A atoms to x_0n nodes of the ensemble in any order, and then to add B atoms to the rest of the x_1n nodes of the ensemble, again in any order. The number of distinguishable ways the atoms may be added is $W = n!/[(x_0n)!(x_1n)!]$. If we introduce the CVM notation shown in Table 4.3,

() =
$$n!$$
 (4.27)

$$(\bullet) = \prod(x_i n)!, \tag{4.28}$$

then we have the compact expression

$$W = ()/(\bullet).$$
 (4.29)

The entropy per node and per chain in the ensemble can then be calculated, using Stirling's formula, to be

$$s = \frac{k}{n} \ln W = -k \sum x_i \ln x_i. \tag{4.30}$$

As expected, this equation reproduces the entropy of a random mixture of noninteracting components.

Since, by assumption, the nodes do not interact, the energy per node and per chain in the ensemble is just a weighted sum of the energies of the individual A and B atoms:

$$u = \sum x_i u_i. \tag{4.31}$$

Largest	Combinatorial	Uncorrelated
cluster	Factor	cluster Identity
Space	() = n!	
Point	$(\bullet) = \prod (x_i n)!$	
Pair	$(-) = \prod (y_i n)!^{\beta_i}$	$(\neq) = (\bullet)^2/()$
Triangle	$(\Delta) = \prod (z_i n)!^{\gamma_i}$	$(\measuredangle) = (\bullet)^3 / ()^2$
Tetrahedron	$(\mathbf{A}) = \prod (w_i n)!^{\delta_i}$	$(\bigstar) = (\bullet)^4 / ()^3$

Table 4.3: Heirarchy of CVM approximations showing combinatorial factors and uncorrelated cluster identities.

Note, though, that although we have assumed that the nodes do not interact directly, we may still allow them to interact indirectly by allowing the energies u_i to depend on the mean composition. For example, if u_0 is proportional to the average concentration of B, $u_0 = \Omega x_1/2$, and u_1 is proportional to the average concentration of A, $u_1 = \Omega x_0/2$, then the molar energy becomes

$$u = \Omega x_0 x_1, \tag{4.32}$$

which reproduces the strictly regular solution model for alloys.

Finally, the free energy of the system, f = u - Ts, can be seen to be a function of two parameters, x_0 and x_1 . Only one can be chosen freely, however, since, as listed in Table 4.4, they must together obey the constitutive "space" relationship

$$x_0 + x_1 = 1. \tag{4.33}$$

Therefore, given the overall composition, $x \equiv x_1$, the free energy is given directly by Equations 4.30 and 4.31.

Pairs

Consider now the next CVM approximation, in which the largest clusters are pairs of atoms. Again imagine building an ensemble of n linear chains, whose nodes have x_0n A atoms and x_1n B atoms. This time, however, we include only those chains for which the overall fractions of AA, AB, BA, and BB atom pairs (or bonds) assume particular values, say, y_0 , y_1 , y_1 , and y_2 .

Note that we have assumed that y_1 is, by symmetry, the number of both the AB and the BA atom pairs. Then, as listed in Table 4.4, the degeneracies of the configurations are $\beta_1 = 2$ and $\beta_0 = \beta_2 = 1$. Also note that the atom pair fractions y_i are not independent of the atom fractions

Configuration	Fraction	Degeneracy	Constitutive
			Relation
()	1		$1 = x_0 + x_1$
$(\bullet)_A$	x_0		$x_0 = y_0 + y_1$
$(\bullet)_B$	x_1		$x_1 = y_1 + y_2$
$(-)_{A_2}$	y_0	$\beta_0 = 1$	$y_0 = z_0 + z_1$
$(-)_{AB}$	y_1	$eta_1=2$	$y_1 = z_1 + z_2$
$(-)_{B_2}$	y_2	$\beta_2 = 1$	$y_2 = z_2 + z_3$
$(\Delta)_{A_3}$	z_0	$\gamma_0 = 1$	$z_0 = w_0 + w_1$
$(\Delta)_{A_2B}$	z_1	$\gamma_1=3$	$z_1 = w_1 + w_2$
$(\triangle)_{AB_2}$	z_2	$\gamma_2 = 3$	$z_2 = w_2 + w_3$
$(\Delta)_{B_3}$	z_3	$\gamma_3 = 1$	$z_3 = w_3 + w_4$
$(\mathbf{A})_{A_4}$	w_0	$\delta_0 = 1$	
$(\mathbf{A})_{A_3B}$	w_1	$\delta_1 = 4$	
$(\mathbf{A})_{A_2B_2}$	w_2	$\delta_2 = 6$	
$(\mathbf{A})_{AB_3}$	w_3	$\delta_3 = 4$	4
$(\mathbf{A})_{B_4}$	w_4	$\delta_4 = 1$	

Table 4.4: Configurations, fractions, degeneracies and constitutive relations for empty, point, pair, triangular and tetrahedral clusters.

 x_i , but must obey the constitutive "point" relations

$$\begin{aligned} x_0 &= y_0 + y_1 \\ x_1 &= y_1 + y_2. \end{aligned}$$
 (4.34)

These relations arise because all AA and AB pairs are associated on the left with an A atom, and all BA and BB pairs are associated on the left with a B atom.

Now suppose we wish to add another node to this ensemble of n chains. In this case, the nodes are *not* independent, so we are not free to add A atoms to x_0n nodes of the ensemble in any order, nor to add B atoms to the rest of the x_1n nodes of the ensemble in any order. Instead, we must add them in such a way that the fractions of new atom-pairs are also y_0 , y_1 , y_1 , and y_2 .

A convenient way of doing this is illustrated in Figure 4.6. To the x_0n chains in the ensemble having A atoms as their last node we add y_0n A atoms and y_1n B atoms. The number of distinguishable ways these additions can be done is $(x_0n)!/[(y_0n)!(y_1n)!]$. Then, to the remaining x_1n chains in the ensemble having B atoms as their last node we add y_1n A atoms and y_2n B atoms. The number of distinguishable ways these



Figure 4.6: Construction of a 1D linear chain of pairs by the addition of a new node (open circle) to an existing lattice (filled circles).

additions can be done is $(x_1n)!/[(y_1n)!(y_2n)!]$. The total number of ways is the product, or $W = [(x_0n)!(x_1n)!]/[(y_0n)!(y_1n)!^2(y_2n)!]$.

If we introduce the CVM notation

$$(-) = \prod \left[(y_i n)! \right]^{\beta_i}, \qquad (4.35)$$

then we can again write more compactly

$$W = \frac{(\bullet)}{(-)}.\tag{4.36}$$

Equations 4.29 and 4.36 are now seen to take the same form, which by induction can be written

$$W = \frac{\text{The part already filled}}{\text{The whole to be completed}}.$$
 (4.37)

This rule generalizes and simplifies the calculation of combinatoric factors for even the most complicated lattice and cluster topologies.

The entropy per node and per chain in the ensemble can now be deduced, again using Stirling's formula, to be

$$s = \frac{k}{n} \ln W = k \left(\sum x_i \ln x_i - \sum \beta_i y_i \ln y_i \right). \tag{4.38}$$

Note that if the atom pairs were randomly distributed, then $y_0 = x_0^2$, $y_1 = x_0 x_1$, and $y_2 = x_1^2$. Then, we would have $\sum \beta_i y_i \ln y_i = 2 \sum x_i \ln x_i$, and Equation 4.38 would reduce to Equation 4.30, the entropy of mixing in the point approximation. In a more compact notation, we can write

$$(\neq) \equiv (\bullet)^2 / (), \tag{4.39}$$

where (\neq) denotes a pair of "uncorrelated" points. Then,

$$W = \frac{(\bullet)}{(\neq)} = \frac{(\bullet)}{(\bullet)^2/()} = \frac{()}{(\bullet)}, \qquad (4.40)$$

which again is the point approximation result.

Since, by assumption, individual atoms *do* interact in the pair approximation, the energy per node and per chain in the ensemble is written as a weighted sum of the energies of the various kinds of *pairs* of A and B atoms:

$$u = \sum \beta_i y_i u_i. \tag{4.41}$$

The free energy of the system, f = u - Ts, is then seen to be a function of five fractions, x_0 , x_1 , y_0 , y_1 , and y_2 . As before, of the two "point" fractions, at most one can be chosen freely, due to the constitutive "space" relationship of Equation 4.33. In addition, of the three "pair" fractions, only one as well can be chosen freely, due to the constitutive "point" relationships of Equation 4.34.

Now, if the overall composition, $x \equiv x_1$, were free to vary, then the equilibrium value of the free energy would be determined by minimizing f with respect to both x and one of the pair probabilities, say, y_1 . This might be the case, e.g., if the lattice were composed not of atoms whose overall numbers we know, but of spins which are free to flip, as in an Ising model. Then, x would play the role of the overall magnetization.

For problems in alloy thermodynamics, however, $x \equiv x_1$ is usually fixed, and is not free to vary. Then, the equilibrium value of the free energy is determined by minimizing f with respect to one of the pair probabilities, usually taken to be the unlike pair probability, y_1 . In other words, we wish to minimize

$$f = y_0 u_0 + 2y_1 u_1 + y_2 u_2 + kT [y_0 \ln y_0 + 2y_1 \ln y_1 + y_2 \ln y_2 - (1-x) \ln(1-x) - x \ln(x)]$$
(4.42)

with respect to y_1 , where $y_0 = 1 - x - y_1$ and $y_2 = x - y_1$. Taking the derivative and setting it equal to zero then gives

$$\frac{\partial f}{\partial y_1} = 2u_1 - u_0 - u_2 + kT \ln\left(\frac{y_1^2}{y_0 y_2}\right) = 0.$$
(4.43)

This expression can be recast, again using the constitutive point relations of Equation 4.34, into the form

$$\frac{y_1^2}{y_0y_2} = e^{-(2u_1 - u_0 - u_2)/kT}.$$
(4.44)

If each atom pair is considered, in a loose sense, to be a molecule, then the equilibrium ratio between the number of AB or BA molecules and the product of the numbers of AA and BB molecules is seen to be given by a Boltzmann factor. This is exactly the "mass-action" law expected for the chemical reaction

$$AA + BB \rightleftharpoons 2AB, \tag{4.45}$$

which can be derived by equating a forward rate, proportional to the product of the concentrations of the AA and BB species, to a backward rate, proportional to the concentration of the AB or BA species. In this way, the pair approximation is equivalent¹⁸ to what is known as the "quasichemical" treatment¹⁹ of alloy thermodynamics, for which Equation 4.44 is the central assumption.

Equation 4.44 has two limiting behaviors. On the one hand, if $2u_1 \ll u_0 + u_2$, then AB pairs are highly favored over AA and BB pairs, the A and B atoms tend to arrange themselves next to each other, and the pair probability y_1 approaches $(1/2) - \sqrt{(1/4) - x(1-x)}$. On the other hand, if $2u_1 \gg u_0 + u_2$, then AA and BB pairs are highly favored over AB pairs, the A and B atoms tend to segregate away from each other, and the pair probability y_1 approaches 0. In between, if $2u_1 = u_0 + u_2$, then AB pairs are neither favored nor unfavored over AA and BB pairs, the A and B atoms tend to arrange themselves randomly, and the pair probability y_1 approaches x(1-x).

Often, it is useful to characterize these behaviors by a short-range "order parameter,"

$$\sigma^{\text{SRO}} \equiv \frac{y_1 - y_1^{\text{ran}}}{y_1^{\text{ord}} - y_1^{\text{ran}}} = \frac{y_1 - x(1 - x)}{(1/2) - \sqrt{(1/4) - x(1 - x)} - x(1 - x)}, \quad (4.46)$$

which is zero if the atoms are arranged randomly, one if the atoms are ordered, and minus one if the atoms are "anti-ordered." For the special case of x = 1/2, Equations 4.38 and 4.41 can be recast, after some algebra, into the forms

$$s = -k \left[\frac{1-\eta}{2} \ln \left(\frac{1-\eta}{2} \right) + \frac{1+\eta}{2} \ln \left(\frac{1+\eta}{2} \right) \right]$$

$$u = \frac{1}{4} \left(u_0 + 2u_1 + u_2 \right) + \frac{\sigma^{\text{SRO}}}{4} \left(2u_1 - u_0 - u_2 \right). \quad (4.47)$$

If the resulting free energy is minimized with respect to η , then one finds

$$\sigma^{\text{SRO}} = \frac{1 - e^{(2u_1 - u_0 - u_2)/2kT}}{1 + e^{(2u_1 - u_0 - u_2)/2kT}}.$$
(4.48)

¹⁸R. Kikuchi, "Theory of ternary III-V semiconductor phase diagrams," *Physica* **103B**, 41 (1981).

¹⁹E.A. Guggenheim, "The statistical mechanics of regular solutions," *Proc. Roy. Soc.* (London) A148, 304 (1935).

For negative $2u_1 - u_0 - u_2$, $\sigma^{\text{SRO}} > 0$, and A and B atoms order on a microscopic scale; for positive $2u_1 - u_0 - u_2$, $\sigma^{\text{SRO}} < 0$, and A and B atoms anti-order on a microscopic scale.

4.3.2 2D Triangular Lattice

In Subsection 4.3.1, we illustrated the cluster variation method using a simple 1D linear lattice. In this subsection, we illustrate the cluster variation method using the more complicated 2D triangular lattice shown in Figure 4.7. This lattice may be constructed either from points, pairs or triangles.

In the lowest order point approximation, the entropies and energies are the same as those for the 1D linear lattice, and Equations 4.30 and 4.31 for the entropies and energies can be carried over without modification. In the pair and triangle approximations, however, the topology of the lattice must be taken into account, because it imposes correlations *between* the various pairs and triangles of atoms. We consider, in turn, these two possible ways of constructing this lattice.

Pairs

Consider first the pair approximation. As before, we assume that individual atoms interact pairwise, so that the energy per node and per chain in the ensemble can, as in Equation 4.41, still be written as a weighted sum of the energies of the various kinds of pairs of A and B atoms. Also, as before, we imagine building a large ensemble of n lattices, whose nodes have x_0n A atoms and x_1n B atoms, and for which the overall fractions of AA, AB, BA, and BB atom pairs are y_0, y_1, y_1 , and y_2 .

Suppose we wish to add another node to this ensemble of lattices, in such a way that the node contains x_0n A atoms and x_1n B atoms, and each new ensemble of bonds, a-b, a-c and a-d, contains y_0 AA pairs, y_1 AB pairs, y_1 BA pairs, and y_2 BB pairs. This we can do in three steps.

First, add node a with respect to node b without regard to correlations with nodes c and d. The number of ways this can be done is the same as that for the 1D linear lattice, namely,

$$W' = \frac{(\bullet)}{(-)}.\tag{4.49}$$

Second, correct (approximately) for the correlation between a and c by multiplying by the factor

$$W'' = \frac{(\bullet)/(-)}{()/(\bullet)}.$$
(4.50)



Figure 4.7: Construction of a 2D triangular lattice by the addition of a new node (open circle) to an existing lattice (filled circles).

This factor is the ratio between the number of ways atoms *should* have been placed on node a with respect to node c, $(\bullet)/(-)$, and the number of ways atoms *actually were* placed on node a with respect to node c, $()/(\bullet)$. Third, correct (approximately) for the correlation between a and d by multiplying by the same factor

$$W''' = \frac{(\bullet)/(-)}{()/(\bullet)}.$$
(4.51)

Another way to look at these two correction factors is to use Equation 4.39 to rewrite them as

$$W'' = W''' = \frac{(\bullet)/(-)}{()/(\bullet)} = \frac{(\neq)}{(-)},$$
(4.52)

so that, in the spirit of Equation 4.37, they carry the physical meaning that correlated pairs are being built from uncorrelated pairs. Indeed, the first combinatorial factor can itself be rewritten as

$$W' = \frac{(\bullet)}{(-)} = \left[\frac{()}{(\bullet)}\right] \left[\frac{(\neq)}{(-)}\right], \qquad (4.53)$$

which carries the physical meaning that an uncorrelated point is first added, and then a correlated pair is built from an uncorrelated pair.

The overall number of ways of adding atoms to node a then becomes

$$W = W'W''W''' = \frac{()}{(\bullet)} \left[\frac{(\neq)}{(-)}\right]^3 = \frac{(\bullet)^5}{(-)^3()^2}.$$
 (4.54)

In other words, again in the spirit of Equation 4.37, we first add an uncorrelated point, then correlate the resulting three uncorrelated pairs.

The entropy per node and per chain in the ensemble can now be calculated, using Stirling's formula, to be

$$s = \frac{k}{n} \ln W = k \left(5 \sum x_i \ln x_i - 3 \sum \beta_i y_i \ln y_i \right). \tag{4.55}$$

Aside from the different numerical factors for the point and pair sums in Equations 4.38 and 4.55, all the arguments in Section 4.3.1 hold.

Triangles

Consider now the triangle approximation. In this case, we assume that the energies of atoms can be expressed as sums over triangular triplets of atoms, so that the energy per node and per chain in the ensemble can be written as a weighted sum of the energies of the various kinds of triangles:

$$u = \sum \gamma_i z_i u_i, \tag{4.56}$$

Here, the overall fractions of A_3 , A_2B , AB_2 and B_3 triplets are $\gamma_0 z_0$, $\gamma_1 z_1$, $\gamma_2 z_2$ and $\gamma_3 z_3$, with the degeneracies, γ_i , listed in Table 4.4 on page 111.

Now suppose we wish to add another node to this ensemble of n lattices, in such a way that (1) the node contains x_0n A atoms and x_1n B atoms, (2) each new ensemble of pairs, a-b, a-c and a-d, contains y_0 AA pairs, y_1 AB pairs, y_1 BA pairs and y_2 BB pairs, and (3) each new ensemble of triangles, a-b-c and a-c-d, contains $\gamma_0 z_0$ A₃ triangles, $\gamma_1 z_1$ A₂B triangles, $\gamma_2 z_2$ AB₂ triangles, and $\gamma_3 z_3$ B₃ triangles. Again, we proceed in steps.

First, we add node a with respect to the pair b-c without regard to correlations with node d. In the spirit of Equation 4.37, the number of ways this can be done is

$$W' = \frac{(-)}{(\Delta)},\tag{4.57}$$

where

$$(\Delta) \equiv \prod \left[(z_i n)! \right]^{\gamma_i} . \tag{4.58}$$

Second, correct for the correlation within the triangle a-c-d by multiplying by the factor

$$W'' = \frac{(-)/(\Delta)}{(\bullet)/(-)}.$$
(4.59)

The numerator of this factor is the ratio between the number of ways atoms should have been placed on node a with respect to the pair c-d. The denominator is the number of ways atoms actually were placed on node awith respect to the pair c-d, namely, the number of ways the correlated pair a-c forming one side of the correlated triangle a-b-c could be formed from the point c.

Again, it is useful to rewrite W' and W'' as

$$W' = \frac{(-)}{(\Delta)} = \frac{()}{(\bullet)} \left[\frac{(-)}{(\neq)} \right] \left[\frac{(\not\Delta)}{(\Delta)} \right]$$

Chapter 4. Ordering and Clustering

$$W'' = \frac{(-)/(\Delta)}{(\bullet)/(-)} = \left[\frac{(-)}{(\neq)}\right]^2 \left[\frac{(\measuredangle)}{(\Delta)}\right], \qquad (4.60)$$

where

$$(\measuredangle) \equiv \frac{(\bullet)^3}{()^2} \tag{4.61}$$

generalizes Equation 4.39 to uncorrelated triangles.

The physical meaning of W' can now be seen to be the addition of an uncorrelated point, the *decorrelation* of the previously correlated pair b-c, and the correlation of the now uncorrelated triangle a-b-c. The physical meaning of W'' is seen to be the decorrelation of the pairs a-c and c-d, which had been previously correlated, followed by the correlation of the now uncorrelated triangle a-c-d.

Finally, then, the overall number of ways of adding atoms to node a becomes

$$W = W'W'' = \frac{\binom{1}{(\bullet)} \left[\frac{(-)}{(\neq)} \left(\underline{\measuredangle} \right) \right] \left[\frac{(-)^2}{(\neq)^2} \left(\underline{\measuredangle} \right) \right]}{\left(\frac{(-)^2}{(\neq)^2} \left(\underline{\measuredangle} \right) \right]} = \frac{(-)^3}{(\triangle)^2(\bullet)}.$$
 (4.62)

In other words, we first form an uncorrelated point, then for each of the two triangles that the point belongs to, we uncorrelate all previously correlated pairs in the triangles and then correlate the triangles.

The entropy per node and per chain in the ensemble can now be calculated, using Stirling's formula, to be

$$s = \frac{k}{n} \ln W = k \left(3 \sum \beta_i y_i \ln y_i - 2 \sum \gamma_i z_i \ln z_i - \sum x_i \ln x_i \right). \quad (4.63)$$

The free energy of the system, f = u - Ts, is a function of nine parameters, x_0 , x_1 , y_0 , y_1 , y_2 , z_0 , z_1 , z_2 , and z_3 . As before, however, of the two point parameters, only one can be chosen freely, due to the constitutive space relationship. In addition, none of the three pair parameters can be chosen freely, because they must obey the constitutive pair relationships

$$y_0 = z_0 + z_1$$

$$y_1 = z_1 + z_2$$

$$y_2 = z_2 + z_3.$$
(4.64)

These equations express the fact that each A_3 or A_2B triangle is formed by coupling an atom to an A_2 pair, that each A_2B or AB_2 triangle is formed by coupling an atom to an AB pair, and that each AB_2 or B_3 triangle is formed by coupling an atom to a B_2 pair.

Finally, of the four triangle parameters, only two can be chosen freely, because of the constitutive point relations listed in Table 4.4 on page 111.

Therefore, for a fixed overall composition, $x \equiv x_1$, the equilibrium value of the free energy is determined by minimizing f with respect to two of the triangle probabilities, which can be taken to be the mixed triangle probabilities, z_1 and z_2 .

4.3.3 3D Zincblende Lattice

In Subsections 4.3.1 and 4.3.2, we illustrated the cluster variation method using first a 1D linear lattice and then a 2D triangular lattice. In this subsection, we illustrate the cluster variation method using a 3D zincblende lattice, whose projection onto an (001) plane is shown in Figure 4.8. For a III-V semiconductor, such a lattice would be built from a superposition of two face-centered-cubic sublattices, one containing group III species and the other containing group V species. Since we are ultimately interested in treating pseudobinary III-III-V alloys, we are interested in the entropy of mixing of group III species on the group III sublattice. Note, though, that these group III species do not form nearest-neighbor bonds with each other; instead, they form next-nearest-neighbor bonds mediated by the group V atoms on the group V sublattice. Therefore, two, three, or four group III atoms can be considered to form a pair, triangle, or tetrahedron if and only if they are all bonded to the same group V atom.

Triangles

Consider first the triangle approximation, in which we assume that the energies of atoms can be expressed as sums over triangular triplets of atoms, as given by Equation 4.56. Suppose we wish to add another node to an ensemble of zincblende lattices, in such a way that all the point, pair and triangle probabilities are preserved. To do so, we use the following simplified rules,²⁰ generalized from Section 4.3.2:

- 1. Add an uncorrelated point via the combinatorial factor $()/(\bullet)$.
- 2. Enumerate all the largest clusters created by adding that point, regardless of overlap.
- For each such cluster: (a) uncorrelate all (previously correlated) subclusters via the combinatorial factors (△)/(△), (−)/(≠), etc., starting from large to small; and (b) correlate the cluster itself via the combinatorial factors (▲)/(▲), (△), (≠)/(−), etc.

 $^{^{20}\}mathrm{The}$ rules are not exact, but must be made recursive when clusters overlap in subclusters larger than pairs.

We start, then, by adding an uncorrelated point a, via the combinatorial factor

$$W' = ()/(\bullet).$$
 (4.65)

Then, we note that by adding point a, we have formed three new triangles, a-b-c, a-b-d, and a-e-f, and one new pair, a-g. We do not include the triangles a-d-g and a-f-g, because these clusters of group III atoms are not all bonded to a common group V atom. Within triangle a-e-f, we must uncorrelate the pair e-f and then correlate the triangle via the combinatorial factor

$$W'' = \begin{bmatrix} (-) \\ (\neq) \end{bmatrix} \begin{bmatrix} (\not\Delta) \\ (\triangle) \end{bmatrix}.$$
(4.66)

Similarly, within triangle a-b-c, we must uncorrelate the pair b-c and then correlate the triangle via the combinatorial factor

$$W^{\prime\prime\prime\prime} = \left[\frac{(-)}{(\neq)}\right] \left[\frac{\langle \underline{\mathscr{A}} \rangle}{(\triangle)}\right]. \tag{4.67}$$

Within triangle a-b-d, we must now uncorrelate two pairs, a-b (which we just correlated in correlating the triangle a-b-c) and b-d, before correlating the triangle:

$$W'''' = \left[\frac{(-)}{(\neq)}\right]^2 \frac{\langle \not\Delta \rangle}{\langle \triangle \rangle}.$$
(4.68)

Finally, we must correlate the pair a-g via the combinatorial factor

$$W''''' = (\neq)/(-). \tag{4.69}$$

Altogether, the number of ways of adding an atom at a is

$$W = W'W''W'''W'''W'''' = \frac{(-)^3(\bullet)^2}{(\triangle)^3()^2},$$
(4.70)

and the entropy is

$$s = k \left(3 \sum \beta_i y_i \ln y_i + 2 \sum x_i \ln x_i - 3 \sum \gamma_i z_i \ln z_i \right).$$
(4.71)

As before, the free energy of the system, f = u - Ts, is a function of nine parameters, $x_0, x_1, y_0, y_1, y_2, z_0, z_1, z_2$ and z_3 . However, for a fixed overall composition, $x \equiv x_1$, the constitutive relations eliminate all but two. The equilibrium value of the free energy is then determined by minimizing f with respect to two of the triangle probabilities, e.g., the mixed triangle probabilities z_1 and z_2 .



Figure 4.8: Top view of the construction of a 3D zincblende lattice by the addition of a new node (open circle) to an existing lattice (filled circles). For a III-III-V alloy, the filled circles and squares would correspond to group III and group V atoms, respectively. The atoms in each (001) sheet are represented by symbols of the same size; the smaller the symbol the deeper the sheet. The solid lines represent next-nearest-neighbor bonds between group III atoms mediated by group V atoms.

Tetrahedra

In the tetrahedron approximation we assume that the energies of atoms can be expressed as sums over tetrahedral quadruplets of atoms,

$$u = \sum \delta_i w_i u_i, \tag{4.72}$$

Here, the overall fractions of A_4 , A_3B , A_2B_2 , A_1B_3 , and B_4 quadruplets are δ_0w_0 , δ_1w_1 , δ_2w_2 , δ_3w_3 , and δ_4w_4 , respectively, with the degeneracies, δ_i , listed in Table 4.4 on page 111.

Suppose we wish to add another node to this ensemble of lattices, in such a way that all the point, pair, triangle, *and* tetrahedron probabilities are preserved. To do so, we again use the rules outlined in Section 4.3.3.

We start by adding an uncorrelated point a, via the combinatorial factor

$$W' = ()/(\bullet).$$
 (4.73)

Then, we note that by adding point a, we have formed one new tetrahedron, a-b-c-d, one new triangle, a-e-f, and one new pair, a-g. Again, we include neither the tetrahedron a-d-g-f nor the triangles a-d-g and a-f-g, because these groups of atoms are not all bonded to a common group V atom.

Within tetrahedron a-b-c-d, we must uncorrelate the triangle b-c-d and then correlate the tetrahedron via the combinatorial factor

$$W'' = \left[\frac{(\Delta)}{(\Delta)}\right] \left[\frac{(\Delta)}{(\Delta)}\right], \qquad (4.74)$$

where

$$(\bigstar) \equiv \frac{(\bullet)^4}{()^3} \tag{4.75}$$

generalizes Equation 4.39 to uncorrelated tetrahedra. Within triangle a-e-f, we must uncorrelate the pair e-f before correlating the triangle:

$$W''' = \left[\frac{(-)}{(\neq)}\right] \left[\frac{(\measuredangle)}{(\bigtriangleup)}\right]. \tag{4.76}$$

Finally, we must correlate the pair a-g via the combinatorial factor

$$W'''' = (\neq)/(-). \tag{4.77}$$

Altogether, the number of ways of adding an atom at a is

$$W = W'W''W'''W''' = \frac{(\bullet)^3}{(\blacktriangle)()^2},$$
(4.78)

and the entropy is

$$s = k \left(3 \sum x_i \ln x_i - \sum \delta_i w_i \ln w_i \right).$$
(4.79)

Note that, for the peculiar topology of the zincblende lattice, the free energy of the system,

$$f = \sum \delta_i w_i u_i + kT \left(\sum \delta_i w_i \ln w_i - 3 \sum x_i \ln x_i \right), \qquad (4.80)$$

contains no pair and triangle probabilities. If it had, though, they could have been eliminated either through the constitutive pair relations listed in Table 4.4 on page 111 or the constitutive "triangle" relations

$$z_{0} = w_{0} + w_{1}$$

$$z_{1} = w_{1} + w_{2}$$

$$z_{2} = w_{2} + w_{3}$$

$$z_{3} = w_{2} + w_{3}.$$
(4.81)

The equilibrium value of the free energy is determined by minimizing f with respect to the five tetrahedron probabilities w_i and the two point probabilities x_i , subject to the three constraints embodied in the two constitutive pair relations and the constitutive space relation. In general, this minimization can be performed through standard techniques based on Lagrange multipliers, one of which can be identified with the chemical potential for species B. This leads to a set of seven nonlinear equations that

122

can be solved through a compact procedure called the "natural iteration method."²¹ In essence, that method begins by guessing values for the point probabilities, using those guesses to calculate the tetrahedron probabilities, from which the point probabilities can be recalculated, etc.

For the zincblende lattice, however, it is simpler to eliminate directly two of the five tetrahedra probabilities using the constitutive pair and space relations. Taking these to be the "pure" cluster probabilities, we then have

$$w_0 = 1 - x - (3w_1 + 3w_2 + w_3)$$

$$w_4 = x - (w_1 + 3w_2 + 3w_3)$$
(4.82)

Therefore, for a fixed overall composition, $x \equiv x_1$, the equilibrium value of the free energy is determined by minimizing f with respect to the remaining three "mixed" tetrahedron probabilities w_1 , w_2 and w_3 .

Taking derivatives of Equation 4.80 with respect to w_1 , w_2 , and w_3 gives, after some algebra,

$$\frac{w_1^4}{w_0^3 w_4} = e^{-(4u_1 - 3u_0 - u_4)/kT}$$
$$\frac{w_2^4}{w_0^2 w_4^2} = e^{-(4u_2 - 2u_0 - 2u_4)/kT}$$
$$\frac{w_3^4}{w_0 w_4^3} = e^{-(4u_3 - u_0 - 3u_4)/kT}$$
(4.83)

Note that Equations 4.83 are in exactly the "mass-action" form expected for chemical reactions between "molecular" tetrahedra:

$$\begin{array}{rcl}
4A_3B &\rightleftharpoons& 3A_4 + B_4 \\
4A_2B_2 &\rightleftharpoons& 2A_4 + 2B_4 \\
4AB_3 &\rightleftharpoons& A_4 + 3B_4,
\end{array} \tag{4.84}$$

and are therefore equivalent, as were Equations 4.44, to a "quasi-chemical" treatment, though of tetrahedra rather than of pairs. In general, chemical reactions between pairs, triplets, and quadruplets form the basis for what are known as the first, second and third quasi-chemical approximations.²² The tetrahedron approximation of the CVM, applied to a zincblende lattice, is therefore equivalent to the third quasi-chemical approximation. It should be emphasized, though, that CVM calculations are not always equivalent to

 $^{^{21}}$ R. Kikuchi, "Superposition approximation and natural iteration calculation in cluster-variation method," J. Chem. Phys. **60**, 1071 (1974).

²²E.A. Guggenheim, "Statistical mechanics of regular mixtures," *Proc. Roy. Soc. (London)* **A206**, 335 (1951).

quasi-chemical approximations; in this case the equivalence is a consequence of the peculiar topology of the zincblende lattice, whose combinatorial factor of Equation 4.78 contains no intermediate subclusters such as triangles or pairs. Otherwise, an equivalence can only be established by the additional assumption that those intermediate subclusters are uncorrelated.

Equations 4.83, together with Equations 4.82, form a set of coupled nonlinear equations which can be solved for the tetrahedra probabilities, w_i , in terms of the tetrahedra energies, u_i . To do so, it is convenient to make use of their equivalence to a quasi-chemical treatment by reformulating them as chemical rate equations that can be solved by numerical simulation.

If we rewrite Equations 4.84 in terms of reactions between tetrahedra differing by only the exchange of one atom, then we have

$$\begin{array}{rcl}
 & k_1^+ \\
 & 4A_3B & \rightleftharpoons & 2A_4 + 2A_2B_2 \\
 & k_1^- \\
 & k_2^+ \\
 & 4A_2B_2 & \rightleftharpoons & 2A_3B + 2AB_3 \\
 & k_2^- \\
 & k_3^+ \\
 & 4AB_3 & \rightleftharpoons & 2A_2B_2 + 2B_4. \\
 & k_3^- \\
\end{array}$$

$$(4.85)$$

The forward and backward reaction rates can be conveniently chosen to be

$$k_{1}^{+} = w_{0}w_{2}e^{-(2u_{1}^{*}-u_{0}-u_{2})/kT}$$

$$k_{1}^{-} = w_{1}^{2}e^{-(2u_{1}^{*}-2u_{1})/kT}$$

$$k_{2}^{+} = w_{1}w_{3}e^{-(2u_{2}^{*}-u_{1}-u_{3})/kT}$$

$$k_{2}^{-} = w_{2}^{2}e^{-(2u_{2}^{*}-2u_{2})/kT}$$

$$k_{3}^{+} = w_{2}w_{4}e^{-(2u_{3}^{*}-u_{2}-u_{4})/kT}$$

$$k_{3}^{-} = w_{3}^{2}e^{-(2u_{3}^{*}-2u_{3})/kT},$$
(4.86)

where u_1^* , u_2^* , and u_3^* are activation energies that can be chosen to match the time-step of the numerical simulation. In practice, the choices

$$u_1^* = \max\{(u_0 + u_2)/2, u_1\}$$

$$u_2^* = \max\{(u_1 + u_3)/2, u_2\}$$

$$u_3^* = \max\{(u_2 + u_4)/2, u_3\}$$
(4.87)

give convergence to steady-state in a reasonable number of time-steps. Note also that these choices of rate constants guarantee that in the steady-state, defined by setting $k_1^+ = k_1^-$, $k_2^+ = k_2^-$ and $k_3^+ = k_3^-$, Equations 4.83 will be satisfied.

In terms of these rates, the time evolution of the tetrahedra probabilities can be written as

$$\begin{aligned} \gamma_{0}\dot{w_{0}} &= -(k_{1}^{+}-k_{1}^{-}) \\ \gamma_{1}\dot{w_{1}} &= 2(k_{1}^{+}-k_{1}^{-}) - (k_{2}^{+}-k_{2}^{-}) \\ \gamma_{2}\dot{w_{2}} &= -(k_{1}^{+}-k_{1}^{-}) + 2(k_{2}^{+}-k_{2}^{-}) - (k_{3}^{+}-k_{3}^{-}) \\ \gamma_{3}\dot{w_{3}} &= - (k_{2}^{+}-k_{2}^{-}) + 2(k_{3}^{+}-k_{3}^{-}) \\ \gamma_{4}\dot{w_{4}} &= - (k_{3}^{+}-k_{3}^{-}) \end{aligned}$$
(4.88)

Note that these rate equations are *conservative*, so that an initial probability distribution will remain correctly normalized, and an initial overall composition, $x = w_1 + 3w_2 + 3w_3 + w_4$, will remain constant. In practice, two convenient initial probability distributions are the completely random Bernoullian distribution,

$$\delta_i w_{i,\mathrm{ran}} = \begin{pmatrix} 4\\i \end{pmatrix} x^i (1-x)^{4-i}, \qquad (4.89)$$

and the completely nonrandom linear distribution,

$$\delta_i w_{i,\text{ord}} = \max\{0, 1 - 4 | x - x_i | \}, \tag{4.90}$$

where x_i is the composition of the *i*th cluster.

4.4 A Pseudobinary III-V Alloy: "InGaAs"

In Section 4.3, we described how, given the energies of various elementary tetrahedra, their occupation statistics could be calculated using the cluster variation method, and the free energy of an alloy as a whole could be determined. In this section, we apply this procedure in an approximate way to the pseudobinary alloy $In_{1-x}Ga_xAs$. The treatment is only semiquantitative, but will include all the most interesting and important features that have been observed in alloys of this type, such as short- and long-range ordering.²³ Tables 4.5 and 4.6, e.g., list the ordered alloys that have been observed thus far in III/V compound semiconductors.

²³H. Nakayama and H. Fujita, "Direct observation of an ordered phase in a disordered $In_{1-x}Ga_xAs$ alloy," *Inst. Phys. Conf. Ser.* **79**, 289 (1985); H.R. Jen, M.J. Cherng and G.B. Stringfellow, "Ordered structures in GaAsSb alloys grown by organometallic vapor phase epitaxy," *Appl. Phys. Lett.* **48**, 1603 (1986); T.S. Kuan, W.I. Wang and E.L. Wilkie, "Long-range order in $In_{1-x}Ga_xAs$," *Appl. Phys. Lett.* **51**, 51 (1987); and M.A. Shahid and S. Mahajan, "Long-range atomic order in $Ga_xIn_{1-x}As_yP_{1-y}$ epitaxial layers [(x, y) = (0.47, 1), (0.37, 0.82), (0.34, 0.71) and (0.27, 0.64)]," *Phys. Rev.* **B38**, 1344 (1988).

Alloy GaPAs	Growth Tech- nique MOVPE	Sub- strate	Struc- ture	Reference H.R. Jen. D.S. Cao and G.B.
		()	1	Stringfellow, Appl. Phys. Lett. 54, 1890 (1989).
InPAs	MOVPE	(001)	$L1_1$	D.H. Jaw, G.S. Chen and G.B. Stringfellow, Appl. Phys. Lett. 59, 114 (1991).
GaPSb	MOVPE	(001)	$L1_1$ (weak)	J.R. Pessetto and G.B. Stringfel- low, J. Cryst. Growth 62 , 1 (1983).
GaAsSb	MOVPE	(001) (110) (221) (311)	L1 ₀ E1 ₁	H.R. Jen, M.J. Cherng and G.B. Stringfellow, J. Cryst. Growth 48, 1603 (1986).
GaAsSb	MBE	(001)	L1 ₁	 I.J. Murgatroyd, A.G. Norman and G.R. Booker, <i>J. Appl. Phys.</i> 67, 2310 (1990); and Y.E. Ihm, N. Otsuka, J.F. Klem and H. Morkoç, <i>Appl. Phys. Lett.</i> 51, 2013 (1987).
InPSb	MOVPE	(001)	$L1_1$	J.R. Pessetto and G.B. Stringfel- low, J. Cryst. Growth 62 , 1 (1983).
InAsSb	MOVPE	(001)	L1 ₁	H.R. Jen, K.Y. Ma and G.B. Stringfellow, <i>Appl. Phys. Lett.</i> 54 , 1154 (1989).

Table 4.5: Ordered III/V-V alloys observed to date in layers formed by any epitaxial growth technique.^a The growth techniques referred to are molecular beam epitaxy (MBE), metal-organic vapor phase epitaxy (MOVPE), liquid-phase epitaxy (LPE), and vapor-levitation epitaxy (VLE). The structures referred to are illustrated in Figures 4.9 and 4.10.

^aAdapted from G.B. Stringfellow and G.S. Chen, "Atomic ordering in III/V semiconductor alloys," J. Vac. Sci. Technol. **B9**, 2182 (1991).

	Growth			
	Tech-	Sub-	Struc-	
Alloy	nique	strate	ture	Reference
GaInP	MOVPE	(001)	L1 ₁	J.P. Goral, M.M. Al-Jassim, J.M.
				Olsen and A. Kibbler, Mat. Res.
				Soc. Symp. Proc. 102, 583 (1988);
				T. Suzuki, A. Gomyo, and S.
				Iijima, J. Cryst. Growth 93 ,
				396 (1988); and O. Ueda, M.
				Takikawa, J. Komeno, and I.
				Umebu, Jpn. J. Appl. Phys. 26,
				L1824 (1987).
AlGaInP	MOVPE	(001)	L1 ₁	G.S. Chen, T.Y. Wang, and G.B.
				Stringfellow Appl. Phys. Lett. 56,
				1463 (1990).
AlGaAs	MOVPE	(001)	L1 ₀	T.S. Kuan, T.F. Kuech, W.I.
		(110)		Wang, and E.L. Wilkie, Phys.
				Phys. Lett. 54, 201 (1985).
AlInAs	MOVPE	(001)	L1 ₀	A.G. Norman, R.E. Mallard, I.J.
				Murgatroyd, G.R. Booker, A.H.
				Moore, and M.D. Scott, Inst.
				Phys. Conf. Ser. 87, 77 (1987).
InGaAs	LPE	(001)	$L1_0$	H. Nakayama and H. Fujita, Inst.
			$E1_1$	Phys. Conf. Ser. 79, 289 (1985).
			$DO_{22}?$	
InGaAs	MBE	(110)	$L1_0$	T.S. Kuan, W.I. Wang, and E.L.
				Wilkie, Appl. Phys. Lett. 51, 51
				(1987).
InGaAs(P)	VLE	(001)	$L1_1$	M.A. Shahid and S. Mahajan,
				Phys. Rev. Lett. B38, 1344 (1988).
InGaAs(Sb)	MOVPE	(001)	$L1_0$	H.R. Jen, M.J. Cherng, and G.B.
			$E1_1$	Stringfellow, Inst. Phys. Conf.
				Ser. 83, 159 (1987).

Table 4.6: Ordered III-III/V alloys observed to date in layers formed by any epitaxial growth technique.^a The growth techniques referred to are molecular beam epitaxy (MBE), metal-organic vapor phase epitaxy (MOVPE), liquid-phase epitaxy (LPE) and vapor-levitation epitaxy (VLE). The structures referred to are illustrated in Figures 4.9 and 4.10.

^aAdapted from G.B. Stringfellow and G.S. Chen, "Atomic ordering in III/V semiconductor alloys," J. Vac. Sci. Technol. **B9**, 2182 (1991).

We begin, in Subsection 4.4.1, by estimating the composition-dependent energies of the various elementary tetrahedra. Then, in Subsection 4.4.2, we apply the cluster variation method to estimate the composition and temperature dependent probabilities of the various elementary tetrahedra. Then, in Subsection 4.4.3, we estimate from these tetrahedra energies and probabilities the composition and temperature dependent molar Gibbs free energy of the alloy as a whole. Finally, in Subsection 4.4.4, we discuss the tendency of these alloys to order, i.e., for the tetrahedra probabilities to be peaked at film compositions that match those of the tetrahedra themselves.

4.4.1 Tetrahedra Energies

Let us start, in this subsection, by describing the energetics of the elementary tetrahedra of which such an alloy is composed. Those energies can be thought of as arising from the two kinds of distortions discussed in Sections 4.1 and 4.2. The first kinds are distortions *internal* to the tetrahedra due to the different equilibrium Ga-As and In-As bond lengths. The second kinds are distortions of the tetrahedra as a whole due to *external* constraints imposed by coherency of the epitaxial film with a substrate. Strictly speaking, these two kinds of distortions are not independent, because various externally imposed distortions may be more or less compatible with particular internal distortions.²⁴ In this simplified treatment, however, we neglect interactions between the two.

Coherency and External Distortions

First, consider the energies of tetrahedra due to *external* distortions. These distortions arise, as discussed in Section 4.2, because of macroscopic strains imposed by coherency with a substrate. The additional energy due to these distortions is given by Equation 4.25.

Ordering and Internal Distortions

Second, consider the energies of tetrahedra due to *internal* distortions. Those energies were estimated in Section 4.1, in a calculation which assumed that the corner group III atoms were pinned at their virtual crystal positions. In fact, those corner group III atoms will have a tendency to relax away from their virtual crystal positions, thereby decreasing the cluster energy.

 $^{^{24}}A.A.$ Mbaye, D.M. Wood and A. Zunger, "Stability of bulk and pseudomorphic epitaxial semiconductors and their alloys," *Phys. Rev.* B37, 3008 (1988).
4.4. A Pseudobinary III-V Alloy: "InGaAs"

Two extremes of behavior can be imagined. On the one hand, if the various tetrahedra were distributed randomly, as in a disordered alloy, then the relaxations of the various corner group III atoms will themselves tend to be random. Then, since each group III atom belongs to four tetrahedra, relaxations that decrease the energy of one tetrahedron will just as likely as not increase the energy of the other three. For this reason, the incoherent superposition of relaxations of group III atoms characteristic of a disordered alloy is not expected to greatly reduce the internal distortional energy from those estimated in Section 4.1 and listed in the first row of Table 4.7.

On the other hand, if the various tetrahedra were distributed in an ordered arrangement, then the relaxations of the various corner group III atoms will themselves tend to be ordered. Relaxations that decrease the energy of one tetrahedron may be exactly the relaxations required to reduce the internal distortion of the adjacent tetrahedra, and so on. For this reason, the coherent superposition of relaxations of group III atoms characteristic of an ordered alloy *is* expected to reduce the internal distortional energy from those estimated in Section 4.1. For example, for the GaAsSb alloy, calculations indicate that the chalcopyrite and famatinite structures illustrated in Figure 4.10 may be the least distorted,²⁵ although the layered tetragonal and layered trigonal ordered compounds are experimentally more commonly observed (see Tables 4.5 and 4.6). Note also that surface thermodynamics and kinetics effects not taken into account here may influence which of the ordered structure actually appears.²⁶

For the disordered alloy, then, we would like to use the cluster energies calculated in Section 4.1 and listed in the first row of Table 4.7; for the ordered alloys, we would like to use the reduced values listed in the second row of Table 4.7; and for partially ordered alloys, we would like to use values somewhere in between. To incorporate these ideas in a semiquantitative way, we assume that the energies of the various tetrahedra depend on the occupation probability of the tetrahedra themselves:

$$u_{i,\text{int}} = u_{i,\text{int,dis}} + (u_{i,\text{int,ord}} - u_{i,\text{int,dis}})(\delta_i w_i)^{\lambda}.$$
(4.91)

In other words, as the probability of particular clusters increases, their tendency to interact coherently and lower their energy also increases. At low enough temperatures, this kind of cooperative interaction ultimately leads to long-range ordering into stoichiometric structures. Note, though, that only a few of the "wrong" kind of tetrahedra might be expected to destroy

²⁵A.A. Mbaye, D.M. Wood and A. Zunger, "Stability of bulk and pseudomorphic epitaxial semiconductors and their alloys," *Phys. Rev.* B37, 3008 (1988).

²⁶See, e.g., S. Froyen, and A. Zunger, "Surface-induced ordering in GaInP," *Phys. Rev. Lett.* **66**, 2132 (1991).

Ordering Vectors	(0,0,0)	(0,0	,1)	
Name (ternary)	Zincbiende (Sphalarito)	Layered Tetragonal	"Luzonito"	
Fermula:	n = 0,4;AC	n = 2;ABC2	n = 1,3; AjBC4	
Example: (ternary)	ZnS-type	InGaAs ₂ -type	Cu ₃ AsS ₄ -type	
Bravels Lattice:	Face centered cubic	Simple tetragenal	Single cubic	
Space Group				
Int, Tables:	F43m	Pim2	Pi3m	
Schoonilies:	T _d ²	D 5	T,	
Number:	216	115	215	
Struckturbericht	83	—	H24	
Pearson symbol:	CF8		CPO	
Atomic positions: (ternary)	Zn: 4 в 73m S: 4 с 73m	1A: 1 a E2m 18: 1 c 72m 2C: 2 9 mm	3Cu: 3 c 72m 1Aa: 1 a 73m 4S: 4 e 3m	
Equivalent superlattice:	None	(1,1) in {0 0 1} direction	None	
Space group and type (binary)				
tvi. Tables	Fm3m	P4/mmm	Pm3m	
Scheenilles:	U.,	D m	0	
Number	225	123	221	
Pearson symbol	CF,	1	C _P A	
Examples (binary)	СилдАн	CuANË	Cu 3 Au Au 3 Cu	

Figure 4.9: Examples of ordered fcc (or pseudobinary zincblende) structures and their space groups.^a

^aReprinted from L.G. Ferreira, S-H Wei and A. Zunger, "First-principles calculation of alloy phase diagrams: the renormalized-interaction approach," *Phys. Rev.* **B40**, 3197 (1989).

(2,0,1)		(1,1,1)		
Chaicopyrite	Famatinite	Layered Trigonal	"X"	
n = 2;ABC;	n = 1,3; AjBC4	n = 2;ABC2	n = 1,3;A ,BC ,	
CuFeS ₂ -type	Cu ₃ SbS ₄ -type	CrCuS ₂ -type (Ne V S ₂)		
Body centered tetragonal	Body centered tetragonal	Rhembohedral	Ortherheaddic	
NZ26 D 20 12 22 E1, 122 E1, 116	i42m D 31 121 121 142 a 116	R3m C 3 190 194		
2 Cu: 4 a 4 2 Fe: 4 b 4 4 S: 8 d 2	1 Cu: 2 b 42m 2 Cu: 4 4 4 1 Sb: 2 s 42m 4 S: 8 j m	1A: 1 a 3m 18: 1 a 3m 1C: 1 a 3m 1C: 1 a 3m		
(2,2) in [2 0 1] direction	(1,3) in [2 0 1] direction	(1,1) in [1 1 1] direction	Mone	
14, /amd D ¹⁰ 66, 141 141	H/wmmm D ¹⁷ 139 DO 22 88	₹3m D 2d 146 L1 , MR32	Crime D 19 es	
Nb-P(7)	AI ₃ TI, NI ₃ V	Cu Pl, SiGe	None	

Figure 4.10: Examples of ordered fcc (or pseudobinary zincblende) structures and their space groups.^a

^aReprinted from L.G. Ferreira, S-H Wei, and A. Zunger, "First-principles calculation of alloy phase diagrams: the renormalized-interaction approach," *Phys. Rev.* **B40**, 3197 (1989).

	$u_{0,\mathrm{int}}$	$u_{1,\mathrm{int}}$	$u_{2,\mathrm{int}}$	$u_{3,\mathrm{int}}$	$u_{4,\mathrm{int}}$
Disordered	Õ	0.023	0.031	0.023	0
Ordered	0	0.017	0.016	0.017	0

Table 4.7: Estimated internal distortion energies (in eV per atom pair) of the elementary InGaAs tetrahedra shown in Figure 4.2. The energies listed in the first row were estimated for a disordered arrangment of tetrahedra, whose corner atoms, on average, are bound to virtual crystal sites.^{*a*} The energies listed in the second row are those (very roughly) estimated for an ordered arrangement of tetrahedra, whose corner atoms can relax "in-phase" with the corner atoms of adjacent tetrahedra.^{*b*}

the coherency of the tetrahedron relaxations. Therefore, we expect the ordering energy to be a highly nonlinear function of the cluster probability itself. In this treatment, we take λ , the nonlinearity parameter, to be eight.

In a sense, we have augmented the tetrahedron approximation of the CVM, which allows different tetrahedra to have different energies, with a point, or mean-field approximation of the CVM to allow each tetrahedron's energy to depend also on the average tetrahedra populations. We must emphasize, though, that this simple, mean-field treatment of long-range order is only a semiquantitative one. To treat long-range order quantitatively within the CVM, it is necessary to distinguish between the (up to) four group III sublattices in the ordered structures, and to account explicitly for the occupation statistics on each sublattice of the (up to) 16 kinds of tetrahedra.²⁷

Total Energies

The internal and external strain energies can now be summed to give

$$u_i = u_{i,\text{int}} + u_{i,\text{ext}}.\tag{4.92}$$

^aM. Ichimura and A. Sasaki, "Short-range order in III-V ternary alloy semiconductors," J. Appl. Phys. **60**, 3850 (1986); A. Sher, M. van Schilfgaarde, A.-B. Chen and W. Chen, "Quasi-chemical approximation in binary alloys," *Phys. Rev.* **B36**, 4279 (1987).

^bEstimated very roughly by scaling the results of calculations in the GaAsSb system by L.G. Ferreira, S-H Wei, and A. Zunger, "First-principles calculation of alloy phase diagrams: the renormalized-interaction approach," *Phys. Rev.* **B40**, 3197 (1989).

²⁷W.L. Bragg and E.J. Williams, "The effect of thermal agitation on atomic arrangement in alloys" *Proc. Roy. Soc. (London)* A145, 699 (1934); H.A. Bethe, "Statistical theory of superlattices," *Proc. Roy. Soc. (London)* A150, 552 (1935); C.M. van Baal, "Order-disorder transformations in a generalized Ising alloy," *Physica* 64, 571 (1973); and D. de Fontaine, "Configurational thermodynamics of solid solutions," *Solid State Physics* 34, 73 (1979).

These energies cannot be evaluated directly, since, through Equation 4.91, they depend on the tetrahedron probabilities, which in turn depend (selfconsistently) on the energies themselves. However, we can get an approximate idea of how the tetrahedra energies depend on the compositions, x_{epi} and x_{sub} , of the epitaxial film and the substrate by calculating the energies of a completely disordered alloy, so that the tetrahedra probabilities contribute negligibly to Equation 4.91, and $u_{i,int} \rightarrow u_{i,int,dis}$. Those energies are plotted in the right column of Figure 4.11 as functions of x_{epi} and x_{sub} .

Consider first the type 0 tetrahedron at the bottom of that column. All of its group III atoms are Ga, and so its externally imposed strain energy is zero when it is embedded in a film of pure GaAs grown undistorted and lattice-matched to a substrate of pure GaAs. Moreover, its internal distortional strain energy is also zero, since the central As atom is symmetrically situated within a tetrahedron of equivalent Ga atoms. Therefore, its total strain energy is zero at $x_{epi} = x_{sub} = 0$.

If now we increase x_{epi} , then the average lattice parameter of the epitaxial film increases, and the unit cell of the epitaxial film grows. At the same time, the size of the tetrahedra embedded in the film are tied to those of the unit cell. Therefore, the type 0 tetrahedra themselves must grow, even if they would "prefer" not to, and their strain energies must increase.

Note that even as x_{epi} increases, we can choose either to increase the substrate lattice parameter at the same rate $(x_{sub} = x_{epi})$ or to keep it fixed $(x_{sub} = 0)$. If we increase it at the same rate (open circles in Figure 4.11), then the unit cell of the epitaxial film remains an undistorted, albeit larger, cube. The energy of the type 0 tetrahedra increases due to that volume mismatch. If, however, we keep it fixed (filled near circles in Figure 4.11), then the unit cell of the epitaxial film is not only larger, but distorted as well. The energy of the type 0 tetrahedra is therefore also quite high when $x_{epi} = 1$ and $x_{sub} = 0$.

Suppose, now, that we fix x_{epi} at zero, but increase x_{sub} . Then, the unit cell of the epitaxial film remains approximately the same size, but it distorts, as its parallel lattice parameter increases and its perpendicular lattice parameter decreases. Therefore, its energy increases, reaching a maximum at $x_{sub} = 1$. If x_{epi} is now increased, then the volume of the unit cell increases, but the distortion in the unit cell decreases. Initially, the strain energy in the type 0 tetrahedra decreases as the unit cell distortion decreases, but eventually it increases as the volume mismatch between the type 0 tetrahedra and the film unit cell increases.

Similar arguments can be used to understand the dependences of the energies of the other types of tetrahedra on x_{epi} and x_{sub} . In general, the energy minima for the various tetrahedra occur when both x_{epi} and x_{sub} are equal to the composition of the cluster itself. The reasons are that when



Figure 4.11: Energies of various tetrahedra embedded in disordered $In_{1-x}Ga_xAs$ of composition x_{epi} grown coherently on substrates having the lattice parameters of bulk $In_{1-x}Ga_xAs$ at composition x_{sub} .

 $x_i = x_{epi}$, the volume of cluster *i* is best matched to the volume of the unit cell of the epitaxial film, and when $x_{epi} = x_{sub}$ the unit cell of the film is least distorted. Deviations from $x_{epi} = x_{sub} = x_i$ along the $x_{epi} = x_{sub}$ diagonal lead to volume mismatches and relatively large increases in energy. Deviations from $x_{epi} = x_{sub} = x_i$ through changes in x_{sub} lead to distortions and somewhat smaller increases in energy. Deviations from $x_{epi} = x_{sub} = x_i$ through changes in x_{epi} lead to some of both, and intermediate increases in energy.

Note that Equation 4.92 includes only the elastic potential energy contribution to the energies of each cluster. In principle, the temperature dependences of the molar energies and entropies could also be determined by heat capacity functions for each cluster, via Equations 2.9 and 2.8. However, in this simple treatment, we make the approximation that the various tetrahedra all have the same heat capacities. Then, the temperature dependences to their molar energies and entropies are all the same. Since tetrahedra probabilities depend only on the *relative* energies, we can neglect those temperature dependences.

4.4.2 Tetrahedra Probabilities

In Subsection 4.4.1, we estimated the energetics of the various elementary tetrahedra from which the InGaAs alloy may be constructed. In this subsection, we use these elementary tetrahedra energetics to calculate the tetrahedra probabilities using the rate equation method outlined in Section 4.3.3. These are shown in Figure 4.12 as functions of $x_{\rm epi}$ and $x_{\rm sub}$ at fixed temperatures of 100, 600, and 1100 K. Two opposing tendencies determine the probability distributions.

The first tendency is energy minimization. For a given composition of the epitaxial film, the two tetrahedra whose compositions just straddle x_{epi} will be the least volume mismatched, and will usually have the lowest energies. The film energy will then be minimized if it is composed of a weighted combination of only those two tetrahedra. For example, if $x_{epi} =$ 3/8, then the type 1 ($x_1 = 1/4$) and type 2 ($x_2 = 1/2$) tetrahedra will have the lowest energies, and the lowest energy film will be that composed of half type 1 and half type 2 tetrahedra. Therefore, at 100 K (left column of Figure 4.12), where energy minimization is most important, only two kinds of tetrahedra are ever significantly populated, and the probability distribution approaches the linear ramp given by Equation 4.90.

The second, opposing, tendency is entropy maximization. As can be seen in Figure 4.12, as temperature increases and entropy becomes an increasingly important component of the molar Gibbs free energy, the probabilities "diffuse" away from the tetrahedra whose compositions straddle that of the epitaxial film. The probabilities cannot diffuse too far away, however, since the overall composition of the film is still constrained to be $\sum p_i x_i = x_{epi}$. Ultimately, at 1100 K, the probability distribution approaches the Bernoullian distribution given by Equation 4.89.

4.4.3 Free Energies

In Subsections 4.4.1 and 4.4.2, we estimated the energetics and probabilities of the various elementary tetrahedra from which the InGaAs alloy may be constructed. In this subsection, we use these energetics and probabilities to calculate the molar Gibbs free energy of the film as a whole using Equation 4.80. These free energies are shown in Figure 4.13 as functions of x_{epi} and x_{sub} , again for three fixed temperatures: 100, 600, and 1100 K. These temperatures are representative of three distinct regimes of behavior.

At the highest temperature, 1100 K, the molar Gibbs free energy is everywhere and in every direction concave up. Therefore, films cannot lower their molar Gibbs free energies by decomposing spatially into local regions, some having higher x_{epi} and others having lower x_{epi} . Epitaxial films at this temperature are *stable* against such macroscopic compositional clustering.

At the intermediate temperature, 600 K, the molar Gibbs free energy is concave up with respect to horizontal fluctuations in x_{epi} (at fixed x_{sub}), but concave down with respect to diagonal fluctuations in x_{epi} (mimicked by identical flucations in x_{sub}). Therefore, films cannot lower their molar Gibbs free energies by composition fluctuations that preserve x_{sub} , but can by fluctuations that do not preserve x_{sub} . In other words, fluctuations in which the local regions remain coherent with the substrate are suppressed, while fluctuations in which the local regions are incoherent (and hence free to adopt their equilibrium lattice parameter) are not. Epitaxial films at this temperature are stable against *coherent* macrosopic clustering, but unstable against *incoherent* macrosopic clustering.

At the lowest temperature, 100 K, the molar Gibbs free energy is, for some combinations of x_{epi} and x_{sub} , concave down with respect to *both* horizontal and diagonal fluctuations in x_{epi} . Therefore, these films can lower their molar Gibbs free energies both by composition fluctuations that preserve x_{sub} , as well as by fluctuations that do not preserve x_{sub} . These films at this temperature are *not* stable against either coherent or incoherent macroscopic compositional clustering.

Note that the downward concavity of the molar Gibbs free energy at 100 K is most exaggerated at those special compositions (1/4, 1/2, 3/4) for which we have assumed ordering may take place. The sharpness of those cusps is a consequence of the cooperative nature of the ordering process.



Figure 4.12: 100, 600, or 1100 K probabilities of various tetrahedra embedded in $In_{1-x}Ga_xAs$ of composition x_{epi} grown coherently on substrates having the lattice parameters of bulk $In_{1-x}Ga_xAs$ at composition x_{sub} .

In the $In_{0.25}Ga_{0.75}As$ alloy, for example, the more $In_{0.25}Ga_{0.75}As$ tetrahedra there are, the closer w_1 approaches unity, the lower the energy $u_{1,int}$ becomes, and the more favored the $In_{0.25}Ga_{0.75}As$ tetrahedra become. At higher temperatures or at compositions slightly off $In_{0.25}Ga_{0.75}As$, there are never enough $In_{0.25}Ga_{0.75}As$ to "get the process going," and the tetrahedra energies are dominated by their disordered values.

Also note that with respect to horizontal fluctuations, the resulting cusps are even, for intermediate substrate compositions, *global* minima in the molar Gibbs free energies. Therefore, films that are constrained to be coherent with a substrate are unstable against clustering into ordered alloys. With respect to diagonal fluctuations, however, the cusps are only *local* minima. Therefore, films not constrained to be coherent with a substrate are unstable against clustering into ordered compounds, but those ordered compounds are themselves unstable against further clustering into (nearly) pure GaAs and (nearly) pure InAs.

To understand these three temperature regimes more concretely, consider an epitaxial film at $x_{epi} = 0.6$ grown on a substrate also at $x_{sub} = 0.6$. Because the film is lattice-matched to the substrate, it is free from macroscopic elastic strain. It is, however, also composed preferentially of type 2 and type 3 tetrahedra. Those tetrahedra are internally distorted, and hence, on a microscopic scale, contain a significant amount of internal distortional elastic energy.

Suppose we force the film to decompose into macroscopic clusters, of which in some $x_{epi} = 0$ and in others $x_{epi} = 1$. These clusters are composed preferentially of type 0 and type 4 tetrahedra, respectively. Neither type of tetrahedron is internally distorted, and hence both are free of internal distortional elastic energy. However, they may or may not be externally distorted, and hence may or may not be free of external distortional elastic energy.

On the one hand, if the $x_{epi} = 0$ and $x_{epi} = 1$ clusters were each free to change their average lattice parameters (i.e., free to change x_{sub}), then the type 0 and type 4 tetrahedra would be free from external distortional elastic energy. Hence, the decomposition of regions having mainly type 2 and 3 tetrahedra into macroscopic clusters having mainly type 0 and type 4 tetrahedra decreases the overall strain energy and will tend to occur.

Note, though, that the number of ways different tetrahedra can be combined to form a macroscopically uniform alloy at $x_{epi} = 0.6$ is larger than the number of ways they can be combined to form alloys at $x_{epi} = 0$ and $x_{epi} = 1$. Since at high enough temperatures, entropic contributions to the molar Gibbs free energies ultimately dominate, there will then be a critical temperature above which mixing will be favored over decomposition.

On the other hand, if the $x_{epi} = 0$ and $x_{epi} = 1$ clusters were not free to



Figure 4.13: 100, 600, or 1100 K molar Gibbs free energies of $In_{1-x}Ga_xAs$ at composition x_{epi} grown coherently on substrates having the lattice parameters of bulk $In_{1-x}Ga_xAs$ at composition x_{sub} . The filled circles represent $In_{1-x}Ga_xAs$ grown on substrates with compositions $x_{sub} = 0, 1/2, 1$. The open circles represent $In_{1-x}Ga_xAs$ grown on "lattice-matched" substrates with compositions $x_{sub} = x_{epi}$, or, alternatively, to incoherent growth.

change their average lattice parameters (i.e., not free to change x_{sub}), then the type 0 and type 4 tetrahedra would not be free from external distortional elastic energy. If that energy is higher than the internal distortional elastic energy of the original type 2 and 3 tetrahedra, then the decomposition is suppressed. Instead, the film will decompose into macroscopic ordered clusters, in some of which $x_{epi} = 0.5$ and in others $x_{epi} = 0.75$. These clusters are composed preferentially of type 2 and 3 tetrahedra, respectively, which fit together in such a way as to minimize their internal distortional elastic energy.

Note, though, that just as before, the number of ways different tetrahedra can be combined to form a macroscopically uniform alloy at $x_{epi} = 0.6$ is larger than the number of ways they can be combined to form ordered alloys at $x_{\rm epi} = 0.5$ and $x_{\rm epi} = 0.75$. Since at high enough temperatures, entropic contributions to the molar Gibbs free energies ultimately dominate, there will then be a critical temperature above which mixing will be favored over ordering.

4.4.4 Short-Range Ordering

In Subsection 4.4.2, we estimated the probabilities of the various elementary tetrahedra from which an InGaAs alloy may be constructed. From Figure 4.12, it can be seen that as the temperature is lowered, the tetrahedra probabilities become less random and more peaked at the film compositions that match those of the tetrahedra themselves. This is a consequence of the fact that the tetrahedra energies are not the same, but are minimum for film compositions that match those of the tetrahedra themselves.

In this subsection, we discuss in more detail this deviation from randomness. Now, first suppose the tetrahedra *were* distributed randomly, according to Equation 4.89. Then, from the constitutive pair and triangle relations listed in Table 4.4 on page 111, the "unlike" pair probability would be

$$y_1 = w_1 + 2w_2 + w_3$$

= $x(1-x)^3 + 2x^2(1-x)^2 + x^3(1-x)$
= $x(1-x),$ (4.93)

as expected. Since the clusters are *not* distributed randomly, we expect deviations from this purely random mixed pair probability.²⁸

To quantify these deviations from randomness, we define a short-range order parameter associated with pairs of unlike (next-nearest-neighbor) group III atoms, analogous to that of Equation 4.46,

$$\sigma^{\text{SRO}} \equiv \frac{w_1 + 2w_2 + w_3 - x(1-x)}{w_1^{\text{ord}} + 2w_2^{\text{ord}} + w_3^{\text{ord}} - x(1-x)},\tag{4.94}$$

where the w_i^{ord} are the completely ordered cluster probabilities given by Equation 4.90. σ^{SRO} is unity if every In atom is surrounded by as many Ga atoms as possible, zero if every In atom is surrounded by a random number of Ga atoms, and negative if every In atom is surrounded by as many In atoms as possible.

²⁸M.T. Czyżyk, M. Podgórny, A. Balzarotti, P. Letardi, N. Motta, A. Kisiel and M. Kimnal-Starnawska, "Thermodynamic properties of ternary semiconducting alloys," *Z. Phys.* B62, 153 (1986).



Figure 4.14: 100, 600, or 1100 K short-range order parameters in $In_{1-x}Ga_xAs$ at composition x_{epi} grown coherently on substrates having the lattice parameters of bulk $In_{1-x}Ga_xAs$ at composition x_{sub} . The filled circles represent $In_{1-x}Ga_xAs$ grown on substrates with compositions $x_{sub} = 0, 1/2, 1$. The open circles represent $In_{1-x}Ga_xAs$ grown on "lattice-matched" substrates with compositions $x_{sub} = x_{epi}$, or, alternatively, to incoherent growth.

This short-range order parameter is plotted in Figure 4.14 for the three temperatures 100, 600, and 1100 K. Note that even at the highest temperature, there is a preference toward bonding between unlike group III atoms, although the preference becomes more pronounced at the lower temperatures. Note also that the short-range ordering becomes somewhat less pronounced (cusped downward) at compositions corresponding to the various elementary tetrahedra. This is so even though the unlike pair probability itself decreases smoothly on both sides of $x_{epi} = 1/2$.

The reason is that exactly at those stoichiometric compositions, it is more difficult to suppress the occupation of composition-straddling tetrahedra. For example, at $x_{\rm epi} = 1/2$, the film will be dominated by type 2 tetrahedra, but some type 1 and 3 tetrahedra, differing in composition from $x_{\rm epi}$ by only 1/4, will also be present. At $x_{\rm epi} = 5/8$, the film will be dominated by a mix of type 2 and 3 tetrahedra. The type 1 and 4 tetrahedra, however, differ in composition from $x_{\rm epi}$ by 3/8. Since the elastic distortion energies of the tetrahedra vary with the square of the composition mismatch, the type 1 and 4 tetrahedra will be suppressed more effectively for $x_{\rm epi} = 5/8$ than the type 1 and 3 tetrahedra were for $x_{\rm epi} = 1/2$.

Finally, we are in a position to understand the microscopic origins of clustering and ordering. In epitaxial films at all temperatures, different tetrahedra have different energies. Usually, tetrahedra that are most nearly volume-matched to the average unit cell volume have the lowest energies, and hence are most favored. Therefore, films of intermediate composition will be preferentially composed of tetrahedra of intermediate composition, and will be short-range ordered in the sense of having an excess of pairs between unlike next-nearest neighbors.

At high temperatures, a homogeneous film of intermediate composition will always be favored over macroscopic clusters at endpoint compositions, because of the increased entropy associated with an increased number of ways of combining tetrahedra of different compositions. Moreover, because of the relatively wide distribution of tetrahedra of different compositions, the tetrahedra will tend to be arranged randomly with respect to one another.

At low temperatures, homogeneous films are no longer favored. As the occupation probabilities become more and more concentrated among those tetrahedra whose compositions straddle the composition of the film, it becomes possible for the tetrahedra to order in such a way that their internal distortions are minimized. Then, films will have a tendency to form macroscopic, ordered clusters composed nearly exclusively of tetrahedra of a certain kind, arranged in a certain way. In *coherent* films, these ordered clusters are the stable state of the system, because tetrahedra in disordered clusters at intermediate compositions have too much internal distortional energy, and tetrahedra in clusters at endpoint compositions have too much external distortional energy. In *incoherent* films, however, the tetrahedra in clusters at the endpoint compositions have no external distortional energy, and hence will ultimately form at the expense of both a homogeneous film or a film composed of ordered clusters.

4.5 Semi-empirical Models

In Sections 4.1–4.4, we have been concerned with developing a microscopic description of the thermodyamics of coherent and incoherent pseudobinary III-V alloys. There is of course no substitute for the physical insight that such a microscopic description gives. However, many of the overall *results* of such a description, such as the molar Gibbs free energy, can be understood using simpler, macroscopic, semi-empirical models. Such models have the advantage, as discussed in Chapter 3, of being described by analytic equations that can be more easily used to calculate phase diagrams and other thermodynamic quantities of interest. In this section, we develop such a semi-empirical model.

We will begin, in Subsection 4.5.1, by describing semi-empirical, physically motivated expressions for the molar Gibbs free energies of disordered and ordered pseudobinary alloys. Then, in Subsection 4.5.2, we use these molar Gibbs free energies to calculate equilibrium alloy phase diagrams.

4.5.1 Free Energies

Let us start, in this subsection, by describing and justifying a semi-empirical expression for the molar Gibbs free energy of an epitaxial pseudobinary alloy grown coherently on a thick substrate. The three main components of the molar Gibbs free energy that we need to account for are (1) the enthalpy of mixing, due to the internal distortional energies of the various tetrahedra, (2) the entropy of mixing, and (3) the coherency energy, due to the external distortional energies of the various tetrahedra.

We describe the internal distortions, as discussed in Section 4.1, by an enthalpy of mixing of the regular solution form

$$h_{\rm int} = \Omega x_{\rm epi} (1 - x_{\rm epi}), \qquad (4.95)$$

where the interaction parameter, Ω , is identified with that calculated in Equation 4.13. The description could easily be improved further through the use of a sub-regular solution form, in order to account for composition-dependent elastic constants. The description could also easily be improved by allowing the mixing enthalpy to depend on temperature through a composition- *and* temperature-dependent heat capacity.

We describe the entropy of mixing by the ideal solution form:

$$s = -k \left[x_{\rm epi} \ln(x_{\rm epi}) + (1 - x_{\rm epi}) \ln(1 - x_{\rm epi}) \right]. \tag{4.96}$$

Finally, we describe the external distortions, following the treatment of Section 4.2, with a coherency energy of the form²⁹

$$h_{\rm ext} = C_{\rm eff} \eta^2 (x_{\rm epi} - x_{\rm sub})^2 a_{\rm avg}^3 / 4, \qquad (4.97)$$

where

$$C_{\text{eff}} = (1 - x_{\text{epi}}) \left(C_{\text{GaAs},11} + C_{\text{GaAs},12} - \frac{2C_{\text{GaAs},12}^2}{C_{\text{GaAs},11}} \right) + x_{\text{epi}} \left(C_{\text{InAs},11} + C_{\text{InAs},12} - \frac{2C_{\text{InAs},12}^2}{C_{\text{InAs},11}} \right)$$
(4.98)

is an effective elastic coefficient that varies linearly between that of GaAs and that of $\rm InAs^{30}$ and

$$\eta = 2 \frac{a_{\text{InAs},o} - a_{\text{GaAs},o}}{a_{\text{InAs},o} + a_{\text{GaAs},o}}$$
(4.99)

²⁹J.W. Cahn, "On spinodal decomposition," Acta Metall. 9, 795 (1961).

³⁰F.C. Larché, W.C. Johnson, C.S. Chiang, and G. Martin, "Influence of substrateinduced misfit stresses on the miscibility gap in epitaxial layers: application to III-V alloys," J. Appl. Phys. 64, 5251 (1988).

is the coefficient of linear expansion per unit composition change.

We note in passing that this elastic energy term, present only for coherent epitaxy, can be an important determinant of the overall driving force for epitaxy. Coherent alloys whose lattice parameters differ from that of the substrate will have higher molar Gibbs free energies than those that are lattice matched. Condensation of lattice-matched alloys will therefore be favored over condensation of lattice-mismatched alloys, as has been observed for both liquid phase epitaxy³¹ as well as MBE.³²

We emphasize here that Equation 4.97 only applies under special circumstances. In particular, it only applies to the geometry we are considering — namely, a thin epitaxial film whose parallel lattice parameter is constrained to be that of its infinitely thick substrate but whose perpendicular lattice parameter is free to adjust — if all phases present have some physical dimension that is large compared to the film thickness. More general treatments of coherent phase equilibria are complicated immensely by the possibility that the elastic coherency energies depend on the details of the phase morphology, which in turn depend on the relative amounts of the different phases present.³³

Finally, the total molar Gibbs free energy of the disordered alloy is

$$g^{dis}(x_{epi}, x_{sub}) = h_{int}(x_{epi}) - Ts(x_{epi}) + h_{ext}(x_{epi}, x_{sub}),$$
 (4.100)

and depends on the compositions of both the epitaxial film and the substrate. As can be seen from the top panels of Figure 4.15, the semi-empirical expression of Equation 4.100 reproduces surprisingly well the molar Gibbs free energies deduced from the CVM calculation shown in Figure 4.13.

Now, as discussed in Section 4.4, ordered and disordered phases should really be treated on a single footing. Doing so requires, however, a microscopic treatment that is difficult to incorporate into a semi-empirical model. Instead, we treat ordered alloys as if they were distinct "compound" phases which exist only within a narrow range of special compositions, as illustrated in the bottom panels of Figure 4.15. In other words, we write

³¹G.B. Stringfellow, "The importance of lattice mismatch in the growth of GaInP epitaxial crystals," J. Appl. Phys. **43**, 3455 (1972); and R.E. Nahory, M.A. Pollack, E.D. Beebe, J.C. DeWinter, and M. Ilegems, "The liquid phase epitaxy of AlGaAsSb and the importance of strain effects near the miscibility gap," J. Electrochem. Soc. **125**, 1053 (1978).

 $^{^{32}}$ M. Allovon, J. Primot, Y. Gao, and M. Quillec, "Auto lattice matching effect for AlInAs grown by MBE at high substrate temperature," *J. Electron. Mater.* **18**, 505 (1989).

³³J.W. Cahn and F.C. Larché, "A simple model for coherent equilibrium," Acta Metall. 11, 1915 (1984); W.C. Johnson and C.S. Chiang, "Phase equilibrium and stability of elastically stressed heteroepitaxial thin films," J. Appl. Phys. 64, 1155 (1988).



Figure 4.15: Semi-empirical molar Gibbs free energies of disordered (top) and ordered (bottom) $In_{1-x}Ga_xAs$ at composition x_{epi} grown coherently on substrates having the lattice parameters of bulk $In_{1-x}Ga_xAs$ at composition x_{sub} .

their free energies as

$$g^{\alpha}(x_{\rm epi}, x_{\rm sub}) = h^{\alpha} + A(x_{\rm epi} - 1/4)^2 + h_{\rm ext}(x_{\rm epi}, x_{\rm sub})$$

$$g^{\beta}(x_{\rm epi}, x_{\rm sub}) = h^{\beta} + A(x_{\rm epi} - 1/2)^2 + h_{\rm ext}(x_{\rm epi}, x_{\rm sub})$$

$$g^{\gamma}(x_{\rm epi}, x_{\rm sub}) = h^{\gamma} + A(x_{\rm epi} - 3/4)^2 + h_{\rm ext}(x_{\rm epi}, x_{\rm sub}).$$
(4.101)

The first terms in these equations are the enthalpies of the ordered compounds, which for InGaAs we identify with those listed in Table 4.7. The second terms are phenomenological terms reflecting expected sharp dependences of the ordering enthalpies on composition near the special compositions, with A a large constant. The third terms are the energies given by Equation 4.97.

4.5.2 Phase Diagrams

In Subsection 4.5.1, we described a semi-empirical expression for the molar Gibbs free energy of a pseudobinary alloy. In this subsection, we use these free energies and the common tangent prescription described in Chapter 3 to calculate two-dimensional x_{epi} -T cuts through the full three-dimensional



Figure 4.16: x_{epi} -T phase diagrams for $In_{1-x}Ga_xAs$ during (right) coherent epitaxy on a substrate of composition x_{sub} and during (left) incoherent epitaxy. Above and below each phase diagram are also shown the molar Gibbs free energies of the various phases at 600 K and 100 K, their common tangents, and the critical compositions (open circles) determined by those common tangents.

 $x_{\text{epi}}-x_{\text{sub}}-T$ phase diagram, as illustrated in Figure 4.16, or calculate the full $x_{\text{epi}}-x_{\text{sub}}-T$ phase diagram itself, as illustrated in Figure 4.17.

In both cases the common tangents may be drawn in two ways. On the one hand, if we constrain the epitaxial film to be coherent, then x_{sub} must be preserved, and so, as illustrated in the right halves of Figures 4.16 and 4.17, we must take horizontal tangents at constant x_{sub} . On the other hand, if we do not constrain the epitaxial film to be coherent, then x_{sub} is free to accommodate x_{epi} , and so, as illustrated in the left half of Figures 4.16 and 4.17, we must take diagonal tangents for which $x_{sub} = x_{epi}$.

Incoherent Phase Equilibria

Consider first the incoherent case for which $x_{sub} = x_{epi}$. Then, Equation 4.100 reduces to

$$g^{\text{dis,inc}}(x_{\text{epi}}) = h_{\text{int}}(x_{\text{epi}}) - Ts(x_{\text{epi}}), \qquad (4.102)$$

since h_{ext} vanishes for $x_{\text{epi}} = x_{\text{sub}}$.

At high temperatures, the mixing entropy term in Equation 4.102 causes the molar Gibbs free energy of the disordered phase to be concave up, and to lie below the molar Gibbs free energies of all of the ordered phases. A disordered InGaAs alloy cannot, for any composition, lower its molar Gibbs free energy either by phase-separating into disordered InAs and GaAs rich clusters or into ordered stoichiometric phases.

At low temperatures, the mixing entropy term becomes small, and is no longer sufficient to bring the molar Gibbs free energy of the disordered phase below those of the ordered phases. Therefore, a uniform disordered alloy can decrease its molar Gibbs free energy by phase-separating into either a combination of two ordered phases or a combination of an ordered phase and an InAs or GaAs rich disordered phase.

In addition, however, the (positive) mixing enthalpy term in Equation 4.102 causes the molar Gibbs free energy of the disordered phase to now be concave enough down that it becomes, near its endpoint compositions, lower than those of all of the ordered phases. Therefore, the ordered phases are themselves unstable with respect to phase separation into pure InAs and GaAs disordered phases. A miscibility gap opens up that destabilizes the ordering.

Coherent Phase Equilibria

Consider now the coherent case for which $x_{sub} = \text{constant}$. Then, the molar Gibbs free energy does not simplify to Equation 4.102, and the full Equation 4.100 must be used.

In this case, at high temperatures, *both* the mixing entropy term and the elastic energy term cause the molar Gibbs free energy of the disordered phase to be concave up. Because of the contribution from both terms, the molar Gibbs free energy remains concave up to lower temperatures, and of course the miscibility gap shifts to lower temperatures.

At low temperatures, the mixing entropy term becomes small, and again is no longer sufficient to bring the molar Gibbs free energy of the disordered phase below those of the ordered phases. Therefore, a uniform disordered alloy can again decrease its molar Gibbs free energy by phase-separating



Figure 4.17: Pseudobinary x_{epi} - x_{sub} -T phase diagrams of $In_{1-x}Ga_xAs$. In the usual, "incoherent" diagram on the left, variations in x_{epi} are accompanied by identical variations in the effective composition of the substrate, x_{sub} . In the "coherent" diagram on the right, variations in x_{epi} occur at constant x_{sub} .

into either a combination of two ordered phases or a combination of an ordered phase and an InAs or GaAs rich disordered phase.

In this case, however, the (positive) mixing enthalpy term is countered by the elastic energy term, and the molar Gibbs free energy of the disordered phase remains concave up. Therefore, the ordered phases remain lower in energy than the disordered phase, even near its endpoint compositions. The ordered phases are therefore stable with respect to phase separation into (nearly) pure InAs and (nearly) pure GaAs disordered phases. Coherency suppresses the miscibility gap. Then, if ordered phases are present, as in this example, coherency stabilizes them.³⁴ If, however, ordered phases are not present, then a uniform disordered alloy will persist to lower temperatures (perhaps even to 0 K) than in the incoherent case.³⁵

Suggested Reading

- 1. W.A. Harrison, *Electronic Structure and the Properties of Solids* (W.H. Freeman, San Francisco, 1980).
- 2. E.A. Guggenheim, *Thermodynamics* (North-Holland, Amsterdam, 1959)

³⁴C.P. Flynn, "Strain-assisted epitaxial growth of new ordered compounds," *Phys. Rev. Lett.* 57, 599 (1986).

³⁵G.B. Stringfellow, "Spinodal decomposition and clustering in III/V alloys," *J. Electron. Mater.* **11**, 903 (1982); and M. Quillec, C. Daguet, J.L. Benchimol, and H. Launois, "InGaAsP alloy stabilization by the InP substrate inside an unstable region in liquid phase epitaxy," *Appl. Phys. Lett.* **40**, 325 (1982).



Figure 4.18: Intermediate stage of building of a 2D lattice of face-centered triangles.

- 3. T. Hill, Introduction to Statistical Thermodynamics (Addison Wesley, Reading, MA, 1960).
- 4. G.B. Stringfellow, Organometallic Vapor-Phase Epitaxy: Theory and Practice (Academic Press, Boston, 1989).
- 5. A. Zunger and D.M. Wood, "Structural phenomena in coherent epitaxial solids," J. Cryst. Growth 98, 1 (1989).

Exercises

- 1. Calculate the distortion energies of the cluster shown in Figure 4.1 for (a) As atoms forced onto VCA and CRA positions and (b) for an As atom at its actual equilibrium position.
- 2. From Figure 4.13, it can be seen that there is a greater tendency toward phase decomposition for strained but coherent epitaxy of In-GaAs on a GaAs substrate than on an InAs substrate. Why?
- 3. What is the limiting value of the order parameter in the pair approximation of the CVM [Equation 4.46] when $2u_1 \gg u_0 + u_2$, i.e., when AA and BB pairs are greatly preferred over AB pairs?
- 4. Derive Equation 4.48 for the temperature dependence of the order parameter in the pair approximation of the CVM.
- 5. Calculate the entropy of the two-dimensional face-centered triangular lattice shown in Figure 4.18 in the point, pair, and triangle CVM approximations.
- 6. Construct a ball-and-stick 3D zincblende lattice and identify the tetrahedra, triangles, and pairs associated with adding point a in Figure 4.8.

- 7. Derive Equation 4.83.
- 8. Draw envelopes of minimum molar Gibbs free energies for the various 2D cuts shown at the bottom of Figure 4.13, and identify the composition ranges within which various phases or phase mixtures are stable.
- 9. Construct phase decomposition scenarios for which coherency energies either depend on, or are independent of, the amounts of the different phases present.
- Starting from Equation 4.100, derive expressions for the chemical potentials of InAs and GaAs in coherent and incoherent epitaxial In-GaAs.
- 11. Using Equations 3.59 and 4.100, derive an expression for how much the vapor pressure of Ga over coherent $In_{1-x}Ga_xAs$ lattice-matched to InP differs from that over incoherent $In_{1-x}Ga_xAs$.

Chapter 5

Coherency and Semi-coherency

In Chapter 4, we described how the thermodynamics of epitaxial alloy films depend on whether those films are coherent or not with their underlying substrate. Films that are coherent often tend to form ordered compounds at certain stoichiometric compositions, while films that are not often tend to separate into their pure-component "endpoint" phases. Coherency with an underlying substrate is thus a crucial determinant of the compositional integrity of alloy films.

Coherency is also a crucial determinant of other properties of alloy films. Consider, e.g., an epitaxial layer whose bulk lattice parameter differs from that of its substrate. On the one hand, if the layer is coherent with its substrate, it will be mechanically strained, and its electronic and optoelectronic properties will be modified through strain-induced changes in electronic band structure.¹ On the other hand, if the layer is not coherent with its substrate, then structural defects must be present, some of which degrade significantly the performance of semiconductor devices.

In this chapter, we discuss the conditions under which coherency between film and substrate can be maintained. In particular, we will focus on the transition from coherency to "semi-coherency." A coherent interface is one that is crystallographically perfect, and that separates epitaxial and substrate atoms in perfect "registry" with each other. If the bulk lattice parameters of the epitaxial layer and the substrate differ, then the epitaxial layer accommodates by developing in-plane strain. A semi-coherent

¹G.C. Osbourn, "Strained-layer superlattices from lattice mismatched materials," J. Appl. Phys. 53, 1586 (1982).

interface, in contrast, is one for which the registry between epitaxial and substrate atoms is punctuated by occasional localized regions of disregistry, i.e., by dislocations. These localized regions of disregistry compensate for lattice parameter misfit between the epitaxial layer and the substrate, allowing the in-plane strain of the epitaxial layer to relax.

We begin, in Section 5.1, by discussing the energies associated with those two kinds of interfaces. The energy associated with a coherent interface is due solely to "coherency strain" in the epitaxial film, and increases linearly with film thickness. The energy associated with a semi-coherent interface is due partly to coherency strain and partly to "misfit" dislocations at the interface. Much of the energy of the misfit dislocations is due to the disregistered atoms at the dislocation core, and is independent of film thickness. Therefore, thin coherent films will tend to have lower energies than thin semi-coherent films, but thick coherent films will tend to have higher energies than thick semi-coherent films.²

As a consequence, in the early stages of film growth, an epitaxial film will usually be coherent with its substrate. Only when the film becomes thick enough will it tend to become semi-coherent with the substrate, and even then, it may not actually become semi-coherent. To become semi-coherent, misfit dislocations must be created at the film/substrate interface, but that creation may be impeded by kinetic barriers. Therefore, in Section 5.2 we discuss the forces, or "excess stresses," acting to create misfit dislocations, and in Section 5.3 we describe how an understanding of those forces can be used to develop semi-empirical macroscopic descriptions of the overall kinetics of strain relaxation.

Note that this chapter deals only with the most common form of heteroepitaxy, in which the film has the same crystal structure as the substrate. Then, provided the lattice parameters of the film and substrate are not too mismatched, epitaxy will occur, and the orientation of the film will mimic that of the substrate. From a practical point of view, we need only be concerned with predicting the conditions under which the film will be coherent or semi-coherent. This chapter also deals only with the simplest form of heteroepitaxy, in which the film grows as layers, rather than as islands (see, e.g., Exercise 2 and Chapter 6).

We emphasize, though, that a deposited film need not have the same crystal structure as the substrate.³ In such cases, it is not always easy to predict (1) whether epitaxy will even occur at all and (2) even if it does, what the orientation relationship will be between the film and the substrate.

²F.C. Frank and J.H. van der Merwe, "One-dimensional dislocations. II. Misfitting monolayers and oriented overgrowth," *Proc. R. Soc. London* **A198**, 216 (1949).

³E. Grünbaum, "List of epitaxial systems," in *Epitaxial Growth*, J.W. Matthews, Ed. (Academic Press, New York, 1975), pp. 611-673.

These two questions are among the most basic in the science of epitaxy, and have been studied for nearly a century, beginning with the work of Barker⁴ and Royer.⁵ However, they are also exceedingly difficult questions that are far from being fully answered.

From a purely crystallographic point of view, one anticipates that those orientation relationships will be favored for which the three dimensional film and substrate lattices coincide most closely at the two-dimensional interface.⁶ For example, such purely crystallographic considerations are evidently responsible⁷ for what are known as the Nishiyama-Wasserman⁸ and Kurdjumov-Sachs⁹ orientation relationships between fcc and bcc crystals found both in solid-phase precipitation reactions¹⁰ as well as in vaporphase epitaxy.¹¹

However, it will not always be sufficient to consider the crystallography of the known equilibrium bulk phases. Occasionally, it will be possible to epitaxially stabilize crystal phases which are not normally stable in bulk form.¹² Elemental tin, e.g., adopts a metastable diamond structure when deposited epitaxially on the (001) surfaces of InSb and CdTe.¹³

Moreover, the epitaxial film may also be chemically different from the substrate, and so the orientation relationship will depend not just on crystallography, but on bond chemistry as well. For these reasons, an under-

⁴T.V. Barker, "Contributions of the theory of isomorphism based on experiments on the regular growths of crystals of one substance on those of another," *J. Chem. Soc. Trans.* **89**, 1120 (1906).

⁵L. Royer, "Recherches expérimentales sur l'épitaxie ou orientation mutuelle de cristaux d'espèces différentes," *Bull. Soc. Franc. Mineral* **51**, 7 (1928).

⁶R.W. Balluffi, A. Brokman, and A.H. King, "CSL/DSC lattice model for general crystal-crystal boundaries and their line defects," *Acta. metall.* **30**, 1453 (1982); and A. Zur and T.C. McGill, "Lattice match: an application to heteroepitaxy," *J. Appl. Phys.* **55**, 378 (1984).

⁷R. Ramirez, A. Rahman, and I.K. Schuller, "Epitaxy and superlattice growth," *Phys. Rev.* B30, 6208 (1984).

⁸Z. Nishiyama, "X-ray investigation of the mechanism of the transformation from face-centred cubic lattice to body-centred cubic," *Sci. Rep. Tohoku Univ.* **23**, 638 (1934); and G. Wasserman, *Arch. Eisenhuettenwes.* **126**, 647 (1933).

 $^{^9 {\}rm G.}$ Kurdjumov and G. Sachs, "Über den Mechanismus der Stahlhärtung," Z. Phys. 64, 325 (1930).

¹⁰U. Dahmen, "Orientation relationships in precipitation systems," Acta Metall. **30**, 63 (1982).

¹¹L.A. Bruce and H. Jaeger, "Geometric factors in f.c.c. and b.c.c. metal-on-metal epitaxy III. The alignments of (111) f.c.c.-(110) b.c.c. epitaxed metal pairs," *Phil. Mag.* A38, 223 (1978).

 $^{^{12}\}mathrm{R.}$ Bruinsma and A. Zangwill, "Structural transitions in epitaxial overlayers," J. Physique 47, 2055 (1986).

¹³R.F.C. Farrow, D.S. Robertson, G.M. Williams, A.G. Cullis, G.R. Jones, I.M. Young, and P.N.J. Dennis, "The growth of metastable, heteroepitaxial films of α -Sn by metal beam epitaxy," *J. Cryst. Growth* **54**, 507 (1981).

standing of orientation relationships in epitaxy is an enormously complicated ongoing area of research that will not be treated here.

5.1 Energetics of Misfit Accommodation

Let us begin, in this section, by discussing the *energetics* of epitaxial films attached through coherent or semi-coherent interfaces to substrates with different lattice parameters. We discuss first, in Subsection 5.1.1, the coherency strains and energies associated with the epitaxial films. Then, we discuss, in Subsection 5.1.2, the strain fields and energies associated with misfit dislocations at the interfaces between the epitaxial films and their substrates, with particular emphasis on face-centered cubic (fcc) and diamond lattices. Finally, we discuss, in Subsection 5.1.3, the dependence of both kinds of energies on misfit dislocation density. Minimizing the sum of the two energies with respect to misfit dislocation density determines how the overall misfit is partitioned, in equilibrium, between coherency strain and misfit dislocation density. We will find that for thin, low misfit films, energy is minimized when the misfit dislocation density is zero. For thick, high misfit films, however, energy is minimized when the misfit dislocation density is nonzero.¹⁴

5.1.1 Coherency Strain

Let us start, in this subsection, by discussing the strain energy associated with epitaxial films that are coherent with their substrates. In particular, consider the simplest strained heterostructure: a single, thin, planar layer of one material and a thick substrate of a different material. As illustrated in Figure 5.1, in the absence of a connection between the two materials, each is unstrained and will adopt its own bulk lattice parameter — either $a_{\rm epi,o}$ or $a_{\rm sub}$. Note that we neglect changes in the lattice parameter of a free-standing film due to surface stresses, changes that may be important for very thin films.¹⁵

Suppose we exert a compressive in-plane force on the epitaxial layer and an equal but opposing tensile in-plane force on the substrate. Then, the in-plane lattice parameter of the epitaxial film will shrink and that of the substrate will grow. If the bulk lattice parameter of the epitaxial layer were larger than that of the substrate, as is the case in Figure 5.1, then the

¹⁴J.H. van der Merwe, "Crystal interfaces. Part II. Finite overgrowths," J. Appl. Phys. 34, 123 (1963).

¹⁵R.C. Cammarata and K. Sieradzki, "Surface stress effects on the critical film thickness for epitaxy," Appl. Phys. Lett. 55, 1197 (1989).



Figure 5.1: Hypothetical simple cubic epitaxial layer and substrate with bulk lattice parameters 5.0 and 5.5 Å, respectively. The epitaxial layer is imagined to be disconnected from the substrate, and so is free to adopt its bulk lattice parameter. As a consequence, it is both unstrained and unstressed.

two in-plane lattice parameters can eventually be made to match, and the epitaxial layer can be joined coherently to the substrate.

Note that if the substrate is much thicker than the epitaxial layer, then it will experience a much lower average in-plane stress than will the epitaxial layer, and its lattice parameter will change much less. Therefore, we make the usual approximation that *all* of the lattice parameter misfit is accommodated by strain in the epitaxial layer, rather than in the substrate. In the general case, though, partitioning of lattice parameter misfits between film and substrate, and even between layers within a multilayered film, must be taken into account.¹⁶

Note also that the Hooke's law energies associated with straining the epitaxial layer and substrate are each proportional to thickness and to the square of the change in lattice parameter. Since the changes in lattice parameter are proportional to applied stress, which is inversely proportional to thickness, the Hooke's law energies are themselves inversely proportional to thickness. Therefore, we can also make the approximation that all of the strain energy associated with coherently joining the epitaxial layer to the substrate is in the epitaxial layer, rather than in the substrate. In other words, just as we saw in Section 4.1.2, most of the energy associated with

 $^{^{16}}$ Z.C. Feng and H.D. Liu, "Generalized formula for curvature radius and layer stresses caused by thermal strain in semiconductor multilayer structures," J. Appl. Phys. 54, 83 (1983).

coupled spring systems is stored in the weaker and more deformed spring.

To calculate the actual strain energy in a fully coherent epitaxial layer, we follow the discussion in Section 4.2, in which a generalized Hooke's law was written in terms of the elastic coefficients C_{ij} . That law is also commonly written, for cubic materials, in terms of Poisson's ratio, ν (defined as the negative of the ratio between lateral and longitudinal strains under uniaxial longitudinal stress), and the shear modulus, μ (defined as the ratio between applied shear stress and shear strain under pure shear):

$$\begin{pmatrix} \epsilon_x \\ \epsilon_y \\ \epsilon_z \end{pmatrix} = \frac{1}{2\mu(1+\nu)} \begin{pmatrix} 1 & -\nu & -\nu \\ -\nu & 1 & -\nu \\ -\nu & -\nu & 1 \end{pmatrix} \begin{pmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \end{pmatrix}.$$
 (5.1)

The relationships between the C_{ij} , ν , and μ are

$$C_{11} = 2\mu \left(\frac{1-\nu}{1-2\nu}\right)$$

$$C_{12} = 2\mu \left(\frac{\nu}{1-2\nu}\right).$$
(5.2)

The shear modulus itself is related to the modulus of elasticity, E, by $2\mu = E/(1+\nu)$.

For concreteness, let us assume, as is common, that the epitaxial film and its substrate are not only cubic, but are oriented along one of the $\langle 100 \rangle$ cubic symmetry directions.¹⁷ Then, the in-plane strains are symmetric and can be taken to be along the x and y axes. If we denote in-plane quantities as "parallel," and out-of-plane quantities as "perpendicular," then we can write

$$\begin{pmatrix} \epsilon_{\parallel} \\ \epsilon_{\perp} \end{pmatrix} = \frac{1}{2\mu(1+\nu)} \begin{pmatrix} 1-\nu & -\nu \\ -2\nu & 1 \end{pmatrix} \begin{pmatrix} \sigma_{\parallel} \\ \sigma_{\perp} \end{pmatrix},$$
(5.3)

which is just the inverse of Equation 4.20.

Equation 5.3 contains two known and two unknown quantities. The first known quantity is the parallel strain, ϵ_{\parallel} , which is determined by the lattice mismatch. The second known quantity is the perpendicular stress, σ_{\perp} , which, since the epitaxial layer is free to expand vertically, vanishes. Therefore, Equation 5.3 determines the two unknown quantities — the parallel stress, σ_{\parallel} , and perpendicular strain, ϵ_{\perp} — in terms of ϵ_{\parallel} only:

$$\sigma_{\parallel} = 2\mu \left(\frac{1+\nu}{1-\nu}\right) \epsilon_{\parallel} \tag{5.4}$$

 $^{^{17}}$ Otherwise, more general expressions are required. See, e.g., J.P. Hirth, "On dislocation injection into coherently strained multilayer structures," *S. Afr. J. Phys.* **9**, 72 (1986).



Figure 5.2: Hypothetical simple cubic epitaxial layer and substrate with bulk lattice parameters 5.5 and 5.0 Å, respectively. The epitaxial layer is still imagined disconnected from the substrate, but has been strained in a direction parallel to the interface so that its parallel lattice parameter matches that of the substrate. As a consequence, it develops both a parallel (in-plane) stress and a perpendicular (out-of-plane) strain.

$$\epsilon_{\perp} = \frac{-2\nu}{1-\nu}\epsilon_{\parallel}. \tag{5.5}$$

As illustrated in Figure 5.2, if the epitaxial layer is strained in a direction parallel to the interface so that its parallel lattice parameter matches that of the substrate, then it must develop a parallel stress. It also develops a perpendicular strain, in the same direction as that which would preserve unit-cell volume. In fact, if ϵ_{\perp} were exactly $-2\epsilon_{\parallel}$, or if $2\nu/(1-\nu)$ were exactly 2, then unit-cell volume would be exactly preserved. Poisson's ratio, however, lies in the range 0.25–0.35 for most materials, so that $2\nu/(1-\nu)$ is actually approximately 1, and unit-cell volume is only approximately conserved.

The "coherency" energy associated with strain in the epitaxial layer can now be calculated, per unit area, to be

$$u_{\rm coh} = \frac{1}{2}h\left(2\sigma_{\parallel}\epsilon_{\parallel} + \sigma_{\perp}\epsilon_{\perp}\right) = 2\mu\left(\frac{1+\nu}{1-\nu}\right)h\epsilon_{\parallel}^2,\tag{5.6}$$

where h is the thickness of the film.

In an epitaxial film composed of multilayers each with a different lattice parameter, the multilayer coherency energy will just be a sum of (or integral



Figure 5.3: Pure edge (left) and pure screw (right) dislocations lying in an interface separating an epitaxial film from its substrate.

over) expressions such as Equation 5.6 for each layer:

$$u_{\rm coh} = 2\mu \left(\frac{1+\nu}{1-\nu}\right) \sum_{i} h_i \epsilon_{i,\parallel}^2, \qquad (5.7)$$

where h_i and $\epsilon_{i,\parallel}$ are the thicknesses and parallel strains of the *i*th layer.

5.1.2 Misfit Dislocations

In Subsection 5.1.1, we discussed the strain energy associated with epitaxial films that are coherent with their substrates. In this subsection, we discuss the energy associated with epitaxial films that are semi-coherent with their substrates. In particular, consider a single-layer heterostructure in which the perfect coherent registry between the epitaxial film and substrate is broken by a localized region of "disregistry." In the simplest case, as illustrated in the left half of Figure 5.3, the disregistry might consist of a half plane missing from the epitaxial film. Physically, we might imagine that the half plane had been "squeezed" upward out of the epitaxial film by a compressive coherency stress, thereby relieving some (or all) of that coherency stress.

Geometrically, the disregistry can be thought of as formed by making what is known as a "Volterra" cut in the epitaxial film perpendicular to the interface, removing a plane of atoms, and then rejoining the remaining crystal by inwardly collapsing atoms in the adjacent planes. The disregistry can then be seen to be equivalent to a negative edge dislocation along the line labeled \bar{l} in the left half of Figure 5.3, with Burgers vector along the line labeled \bar{b} . In a sense, such interface dislocations act to relieve the coherency strain in the epitaxial layer by concentrating the lattice misfit into localized regions of disregistry confined mainly to the interface.

Note that pure edge dislocations such as this, whose Burgers vectors are both perpendicular to the dislocation line and in the plane of the interface, are the most efficient means for relieving coherency strain. Screw dislocations such as that illustrated in the right half of Figure 5.3, whose Burgers vectors are parallel to the dislocation line, do not relieve coherency strain. Likewise, edge dislocations whose Burgers vectors are perpendicular both to the dislocation line and to the plane of the interface, do not relieve coherency strain. Therefore, in the general case of "mixed" dislocations, having both edge and screw character, only that component of the Burgers vector that is both "edgelike" and in the plane of the interface acts to relieve coherency strain. In particular, if, as illustrated in Figure 5.23 on page 195, λ is the angle between (a) the Burgers vector and (b) the direction that is both normal to the dislocation line and that lies within the plane of the interface, then only the component,

$$b_{\text{edg},\parallel} \equiv b \cos \lambda,$$
 (5.8)

acts to relieve lattice misfit.

Also note that dislocations with partial or full edge character move most easily by gliding *within* the plane containing both the dislocation line and its Burgers vector. The pure edge dislocation illustrated in Figure 5.3, e.g., will move most easily within the interface between the epilayer and the substrate. Therefore, such a dislocation, if created at the free surface, would be unable to glide to the interface between the epilayer and the substrate. Instead, it would be constrained to glide parallel to that interface.

To be *practically* effective at relieving misfit strain, then, dislocations must usually have some component of their Burgers vector out of the interface. Otherwise, they must move by "climbing" *out* of the plane containing both the dislocation line and its Burgers vector. Such motion requires the creation or annihilation of vacancies at the dislocation core, and hence a diffusive flux of vacancies either away from or toward the dislocation core. For example, to move the edge dislocation illustrated in the left half of Figure 5.3 down from the interface by one lattice spacing, a row of vacancies must be removed from the dislocation core. Such vacancy removal ultimately requires diffusion away from the core, which usually only becomes significant at fairly high temperatures.¹⁸

¹⁸E.A. Fitzgerald, P.D. Kirchner, R.E. Proano, G.D. Pettit, J.M. Woodall and D.G. Ast, "Totally relaxed $Ge_x Si_{1-x}$ layers with low threading dislocation densities grown on Si substrates," *Appl. Phys. Lett.* **59**, 811 (1991).

While acting to relieve misfit strain, interface dislocations also cost energy, due to the disruption in bonding associated with the disregistered atoms at their core and to the long-range elastic stress and strain fields away from their core. For pure edge and screw dislocations, the energies per unit length associated with the elastic stresses and strains in a cylindrical ring surrounding a long straight dislocation core can be shown, in a continuum model, to be approximately

$$U_{\text{edg}} = \frac{\mu b_{\text{edg}}^2}{4\pi (1-\nu)} \ln(R/r_o)$$
$$U_{\text{scr}} = \frac{\mu b_{\text{scr}}^2}{4\pi} \ln(R/r_o), \qquad (5.9)$$

where r_o and R are the inner and outer radii of the cylinder. The energies associated with the disregistered core atoms, however, are difficult to determine. Instead, it is common to adjust the inner "cutoff" radius r_o so that the core energies are included in Equations 5.9. In practice, a value of $r_o = b/4$ for covalent semiconductors is often used.

For dislocations having mixed character, the energy is the sum of Equations 5.9, with the edge and screw components of the Burgers vectors used accordingly. If, as illustrated in Figure 5.23 on page 195, β is the angle between the Burgers vector and the dislocation line, then the edge component is $b \sin \beta$ and the screw component is $b \cos \beta$. Therefore,

$$U_{\rm dis} = \frac{\mu b^2}{4\pi} \left(\frac{\sin^2 \beta}{1-\nu} + \cos^2 \beta \right) \ln(4R/b)$$
$$= \frac{\mu b^2}{4\pi} \left(\frac{1-\nu \cos^2 \beta}{1-\nu} \right) \ln(4R/b). \tag{5.10}$$

Note that, because of the long-range nature of the elastic stresses and strains, the dislocation energy diverges logarithmically with the radius of the outer radius of the cylinder. Therefore, a dislocation embedded in an infinite crystal has infinite energy. In fact, the long-range elastic stresses and strains are always disrupted (and bounded) either by free surfaces or by the stress and strain fields of neighboring dislocations.

For example, if a free surface at z = 0 is placed a distance h away from a dislocation at z = h, the normal and shear stress components acting on the surface must vanish, because the surface is free to expand outward or contract inward. The effect of the surface can be accounted for approximately by placing an "imaginary" dislocation of the opposite sign at z = -h, thereby largely cancelling the long-range stress field at distances much greater than h from the dislocation core. The energy associated with



Figure 5.4: Dislocation lines lying along [110] (left) and $[1\bar{1}0]$ (right) directions. The edges of the tetrahedra inscribed within each unit cube represent the possible directions of the Burgers vectors for each of those dislocations.

a dislocation a distance h from a free surface is therefore approximately

$$U_{\rm dis} = \frac{\mu b^2}{4\pi} \left(\frac{1 - \nu \cos^2 \beta}{1 - \nu} \right) \ln(4h/b).$$
 (5.11)

Note that the energy is proportional to b^2 , because the strains around the dislocations are proportional to b, and the energy is proportional to the square of the strains. Therefore, dislocations with shorter Burgers vectors will be more common than those with longer Burgers vectors. For this reason, the most common Burgers vectors in fcc-based diamond and zincblende lattices are of the $\frac{1}{2}\langle 110 \rangle$ type, since these are the shortest lattice vectors in these crystals.¹⁹ Since there are six possible $\langle 110 \rangle$ directions, there are six possible directions for the Burgers vectors. These six directions are the edges of the tetrahedra shown in Figure 5.4.²⁰

Consider, for example, misfit dislocations lying along either [110] or $[1\overline{10}]$ directions, as shown in the left and right halves, respectively, of Figure 5.4. For (001) oriented fcc-lattice-based epitaxial films, these two dislocation line directions are the most common, as they lie both in the (001) interface plane as well as in one of the close-packed {111} slip planes within which dislocations move most readily. Dislocations having these line directions

 $^{^{19} {\}rm Dissociation}$ into "partial dislocations" having shorter Burgers vectors separated by stacking faults is also possible.

²⁰N. Thompson, "Dislocation nodes in face-centred cubic lattices," *Proc. Phys. Soc.* B66, 481 (1955).



Figure 5.5: Left: $[\bar{1}01]$ threading screw dislocation segments and $[110] 60^{\circ}$ misfit dislocation segment with Burgers vector along $[\bar{1}01]$ direction. Right: crossed grid of two arrays of misfit dislocations along the $[\bar{1}10]$ and [110] directions.

can be one of three types, depending on the directions of their Burgers vectors. If the Burgers vector of the $\bar{l} = [110]$ dislocation illustrated in the left half of Figure 5.4 lies along the line A-B, parallel to \bar{l} , then it is screw in character. If its Burgers vector lies along the line C-D, perpendicular to \bar{l} , then it is edge in character. If its Burgers vector lies along any of the four other lines, A-C, A-D, B-C or B-D, at 60° to \bar{l} , then it is a "mixed" 60° dislocation. Likewise, the $\bar{l} = [1\bar{1}0]$ dislocation illustrated in the right half of Figure 5.4 will either be screw, edge, or 60° mixed, depending on the direction of its Burgers vector.

An example of a commonly observed dislocation configuration is illustrated in the left half of Figure 5.5. An $\bar{l} = [\bar{1}01]$ screw dislocation with $\bar{b} = [\bar{1}01]/2$ is shown threading up diagonally from the substrate into the epilayer. Just at the epilayer/substrate interface, the dislocation has bent over to form a misfit dislocation segment with $\bar{l} = [110]$. Since Burgers vectors must be preserved along the length of any particular dislocation, the misfit dislocation segment is a 60° dislocation with $\cos \beta = \hat{l} \cdot \hat{b} = \cos 60^\circ = 1/2$. A plan-view transmission electron micrograph of a crossed grid of such misfit dislocations is shown in Figure 5.6.



Figure 5.6: Bright-field plan-view transmission electron micrograph of the interface between a 200-nm $Si_{0.9}Ge_{0.1}$ layer grown on a Si (001) substrate.^{*a*} The misfit dislocations are arranged in a crossed grid running along the $\langle 110 \rangle$ directions within the (001) interface.

^aY. Fukuda, Y. Kohama, M. Seki, and Y. Ohmachi, "Misfit dislocation structures at MBE-grown Si_{1-x}Ge_x/Si interfaces, Jpn. J. Appl. Phys. **27**, 1593 (1988).

5.1.3 Equilibrium Strains and Dislocation Densities

In Subsection 5.1.1 we discussed the strain energy cost associated with a perfectly coherent interface, and in Subsection 5.1.2 we discussed the dislocation energy cost associated with a semi-coherent interface. In this subsection, we ask which of the two interfaces costs the least energy, and hence will be thermodynamically preferred. To answer this question, let us calculate how the two kinds of energies depend on misfit dislocation density. For concreteness, we assume that the semi-coherent interface is composed, as illustrated in the right half of Figure 5.5, of a crossed-grid of two identical arrays of dislocations, each having a linear density of $\rho_{\rm md}$.

For a fully coherent interface, for which $\rho_{\rm md} = 0$, there is only the coherency strain energy, which we have already calculated to be $2\mu[(1 + \nu)/(1-\nu)]hf^2$, where f is the lattice parameter misfit between the epitaxial layer and the substrate. For a semi-coherent interface, for which $\rho_{\rm md} > 0$, the misfit will be partially taken up by localized regions of disregistry,

thereby decreasing the coherency strain energy, but increasing the misfit dislocation energy.

To see how the coherency strain energy will decrease with misfit dislocation density, we note that in one dimension, the misfit taken up by dislocations, $f_{\rm dis}$, is the lattice displacement parallel to the interface per dislocation, $b_{\rm edg,\parallel}$, divided by the spacing between dislocations, $1/\rho_{\rm md}$. In other words,

$$f_{\rm dis} = \rho_{\rm md} b_{\rm edg,\parallel}. \tag{5.12}$$

Therefore, the dislocation density that would relieve all the misfit in one dimension would be $\rho_{\rm md} = f/b_{\rm edg,\parallel}$. If, on average, the strain in the epitaxial film decreases linearly with the dislocation density, then

$$\epsilon_{\parallel} \approx f - f_{\rm dis} = f - \rho_{\rm md} b_{\rm edg,\parallel}. \tag{5.13}$$

The dependence of the coherency strain energy on misfit dislocation density can then be written as

$$u_{\rm coh} = 2\mu \left(\frac{1+\nu}{1-\nu}\right) h \left(f - \rho_{\rm md} b_{\rm edg,\parallel}\right)^2.$$
(5.14)

As indicated by the dashed lines in the left and right halves of Figure 5.7, the coherency strain energy depends parabolically on dislocation density, and vanishes when $\rho_{\rm md} = f/b_{\rm edg,\parallel}$.

At the same time, the energy associated with the dislocations themselves will increase as the misfit dislocation density increases. For most applications, it is sufficient to approximate the energy associated with each of the two dislocation arrays to be the dislocation density times the energy of an isolated dislocation, or

$$u_{\rm dis} \approx \rho_{\rm md} \frac{\mu b^2}{4\pi} \left(\frac{1 - \nu \cos^2 \beta}{1 - \nu} \right) \ln \left(\frac{4h}{b} \right). \tag{5.15}$$

This linear dependence of the dislocation array energy on dislocation density is shown as the dotted lines in the left and right halves of Figure 5.7.

For more precise calculations, however, we note that interactions between dislocation should be taken into account. The reason is that when the dislocation spacing is less than the film thickness, the stress fields of individual dislocations are not fully screened from each other by the free surface, and mediate an "interaction" between them.²¹

 $^{^{21}}$ See, e.g., J.P. Hirth and X. Feng, "Critical layer thickness for misfit dislocation stability in multilayer structures," *J. Appl Phys.* **67**, 3343 (1990); and J.R. Willis, S.C. Jain, and R. Bullough, "The energy of an array of dislocations: implications for strain relaxation in semiconductor heterostructures," *Philos. Mag.* **A62**, 115 (1990).


Figure 5.7: Areal energy densities, normalized by the product of the shear modulus and the Burgers vector, as a function of the misfit taken up by dislocations, $\rho_{\rm md} b_{\rm edg,\parallel}$. The dotted lines represent the energies of two dislocation arrays; the dashed lines represent the coherency strain energies; the full lines represent the sum. The film on the left is thin enough that it is stable when it is coherent with the substrate; the film on the right is thicker and is stable when it is semi-coherent with the substrate.

The total areal energy density is the sum of the areal energy densities associated with the coherency strain and *both* of the dislocation arrays, or

$$u_{\text{tot}} = u_{\text{coh}} + 2u_{\text{dis}}$$

$$= 2\mu \left(\frac{1+\nu}{1-\nu}\right) h \left(f - \rho_{\text{md}} b_{\text{edg},\parallel}\right)^{2}$$

$$+ \rho_{\text{md}} \frac{\mu b^{2}}{2\pi} \left(\frac{1-\nu \cos^{2}\beta}{1-\nu}\right) \ln \left(\frac{4h}{b}\right). \quad (5.16)$$

This dependence of u_{tot} on ρ_{md} displays two distinct kinds of behavior, as illustrated in Figure 5.7. For thin, low-misfit films, the total energy is minimum at $\rho_{md} = 0$. Misfit dislocations cost more energy than is regained by release of coherency strain. For thick, high-misfit films, however, the total energy is minimum at $\rho_{md} > 0$. The introduction of *some* misfit dislocations costs less energy than is regained by release of coherency strain.

Mathematically, these two kinds of behaviors arise according to whether the energy associated with either of the dislocation arrays, $u_{tot}/2$, increases or decreases for an incremental increase in ρ_{md} from $\rho_{md} = 0$. In other words, according to whether

$$\left[\frac{\partial u_{\text{tot}}}{2\partial\rho_{\text{md}}}\right]_{\rho_{\text{md}}=0} = -2\mu\left(\frac{1+\nu}{1-\nu}\right)hb_{\text{edg},\parallel}f + \frac{\mu b^2}{4\pi}\left(\frac{1-\nu\cos^2\beta}{1-\nu}\right)\ln(4h/b)$$
(5.17)

is greater than or less than zero. On the one hand, if it is greater than zero, then the change in energy upon introducing the first few misfit dislocations is positive. Misfit dislocations will *not* tend to form, and the fully strained, coherent epilayer will be thermodynamically stable. On the other hand, if it is less than zero, then the change in energy upon introducing the first few misfit dislocations is negative. Misfit dislocations *will* tend to form, and the strain in the epilayer will tend to "relax."

The critical misfit for a given thickness and the critical thickness for a given misfit, beyond which misfit dislocations will tend to form, are determined by the condition $(1/2)[\partial u_{tot}/\partial \rho_{md}]_{\rho_{md}=0} = 0$, or

$$f_{\rm c} = \frac{b}{8\pi h \cos \lambda} \left(\frac{1 - \nu \cos^2 \beta}{1 + \nu} \right) \ln (4h/b)$$

$$h_{\rm c} = \frac{b}{8\pi f \cos \lambda} \left(\frac{1 - \nu \cos^2 \beta}{1 + \nu} \right) \ln (4h_{\rm c}/b)$$
(5.18)

where we have used Equation 5.8, $b_{\text{edg},\parallel} = b \cos \lambda$. These expressions reproduce exactly those derived originally by Matthews and Blakeslee²² and more recently by Ball and van der Merwe.²³ They are illustrated in Figure 5.8 for $\lambda = \beta = 60^{\circ}$, which is often the case for fcc-lattice-based diamond and zincblende crystals. Films having thickness/misfit combinations below the curves are stable against the introduction of misfit dislocations; films having thickness/misfit combinations above the curves are not.

Also shown in Figure 5.8 are experimental data points corresponding to $In_{1-x}Ga_xAs$ films grown on GaAs substrates and $Si_{1-x}Ge_x$ films grown on Si substrates. As can be seen, the boundary separating the coherent from the semi-coherent films is given very closely by Equation 5.18. That equation is also believed to describe the thermodynamic boundary dividing coherent from semi-coherent epitaxy of metal films.²⁴

Above the critical layer thickness, the energy decreases at first upon the introduction of the first few misfit dislocations, but eventually increases

²²J.W. Matthews and A.E. Blakeslee, "Defects in epitaxial multilayers I. Misfit dislocations," J. Cryst. Growth 27, 118 (1974).

²³C.A.B. Ball and J.H. van der Merwe, "The growth of dislocation-free layers," in *Dislocations in Solids*, F.R.N. Nabarro, Ed. (North-Holland, Amsterdam, 1983), Chap. 27.

 $^{^{24}}$ Y. Kuk, L.C. Feldman, and P.J. Silvermann, "Transition from the pseudomorphic state to the nonregistered state in epitaxial growth of Au on Pd (111)," *Phys. Rev. Lett.* **50**, 511 (1983).



Figure 5.8: Logarithmic (left) and linear (right) plots of critical layer thicknesses for a given misfit (or, alternatively, critical layer misfits for a given thickness), as deduced from Equation 5.18. The triangles represent experiments in which $In_{1-x}Ga_xAs$ layers were grown on GaAs substrates and then annealed.^a The circles represent experiments in which $Si_{1-x}Ge_x$ layers were grown on Si substrates and then annealed.^b The filled data points correspond to structures that maintained coherency; the open data points correspond to structures that became semi-coherent.

again. The dislocation density that minimizes u_{tot} can be found by solving for that ρ_{md} for which the derivative

$$\frac{\partial u_{\text{tot}}}{2\partial \rho_{\text{md}}} = -2\mu \left(\frac{1+\nu}{1-\nu}\right) h b_{\text{edg},\parallel} \left(f - \rho_{\text{md}} b_{\text{edg},\parallel}\right) \\
+ \frac{\mu b^2}{4\pi} \left(\frac{1-\nu \cos^2 \beta}{1-\nu}\right) \ln \left(\frac{4h}{b}\right)$$
(5.19)

vanishes. In other words, the equilibrium dislocation density is given by

$$\rho_{\rm md,equ} = \frac{f}{b\cos\lambda} - \frac{b}{8\pi h\cos^2\lambda} \left(\frac{1-\nu\cos^2\beta}{1+\nu}\right) \ln\left(\frac{4h}{b}\right)$$

^aP.S. Peercy, B.W. Dodson, J.Y. Tsao, E.D. Jones, D.R. Myers, T.E. Zipperian, L.R. Dawson, R.M. Biefeld, J.F. Klem and C.R. Hills, "Stability of strained quantum-well field effect transistors," *IEEE Electron Dev. Lett.* **9**, 621 (1988).

^bD.C. Houghton, C.J. Gibbings, C.G. Tuppen, M.H. Lyons, and M.A.G. Halliwell, "Equilibrium critical thickness for $Si_{1-x}Ge_x$ strained layers on (100) Si," Appl. Phys. Lett. 56, 460 (1990).



Figure 5.9: Equilibrium misfit strain (f_{dis}) taken up by dislocations as a function of thickness at constant misfit (left) or as a function of misfit at constant thickness (right). Below either the critical thickness (h_c) or critical misfit (f_c) the strain taken up by dislocations is zero; above the critical thickness or critical misfit it becomes an increasingly large fraction of the total misfit (f).

$$= \frac{f}{b\cos\lambda} \left[1 - \frac{h_{\rm c}}{h} \right]. \tag{5.20}$$

This equation determines $\rho_{\rm md}$ self-consistently in terms of $\rho_{\rm md}$ itself, and can be solved iteratively; a reasonable initial guess is $\rho_{\rm md} = f/(2b\cos\lambda)$. Once determined, the equilibrium $\rho_{\rm md}$ then determines that portion of the total misfit, $f_{\rm dis}$, that is taken up by dislocations through Equation 5.12.

For a given misfit, $f_{\rm dis}$ is zero for thicknesses less than the critical layer thickness, but increases sharply for thicknesses greater than the critical layer thickness. This dependence is shown in the left half of Figure 5.9. Note that even for $h > h_c$, the equilibrium dislocation density is less than that which would eliminate all coherency strain. In other words, even above the critical layer thickness the coherency strain in the epilayer is only partially, and not fully, relaxed.

Likewise, for a given thickness, $\rho_{\rm md}$ is zero for misfits less than the critical layer misfit, but increases logarithmically for misfits greater than the critical layer misfit. This dependence is shown in the right half of Figure 5.9. Again, even for $f > f_c$, the equilibrium dislocation density is always less than that which would eliminate all coherency strain.

5.2 Forces on Dislocations

In Section 5.1, we found that the transition between coherency and semicoherency corresponded to the thicknesses and misfits at which introduction of the first few misfit dislocations became *energetically* favorable. Ultimately, though, the transition between coherency and semi-coherency also requires the *motion* of dislocations to (or near to) the epilayer/substrate interface. There must then be forces acting on the dislocations to cause them to move. In this section, we describe the forces acting on dislocations in strained heterostructures. We will find that the thermodynamic transition between coherency and semi-coherency also corresponds exactly to the thicknesses and misfits at which the forces acting to elongate existing misfit dislocations are positive or negative. We will also find that the average force acting to elongate existing misfit dislocations, the "excess stress" of the structure, is a natural measure of the driving force for strain relaxation by misfit dislocation creation.

We will begin, in Subsection 5.2.1, by describing excess stress in a simple structure: a single strained surface layer grown on a very thick substrate. Then, in Subsection 5.2.2, we describe excess stress in a more complicated structure: a single strained layer buried within a very thick substrate. Finally, in Subsection 5.2.3, we generalize the concept of excess stress to even very complicated heterostructures, for which the excess stress depends on depth within the structure.

5.2.1 Strained Surface Layers

We start, in this subsection, by describing excess stress in a simple structure: a single strained surface layer grown on a very thick substrate. Consider a dislocation "threading" upward through the epilayer/substrate interface and into the epilayer itself, as illustrated in the left half of Figure 5.10. If the dislocation bends over, then new length of misfit dislocation will be created at the epilayer/substrate interface. If, in steady state, the shape of the threading segment as it moves from A–C to B–D does not change, then the net change in energy is due solely to the new misfit segment C–D, which we may imagine has moved downward from AB.

The energy gained by moving unit length of that segment a distance h downward (or, equivalently, by bending the threading dislocation unit length to the right) is $h\hat{z}$ dotted into what is known as the Peach-Koehler force, $d\bar{F} = (\bar{b} \cdot \bar{\sigma}) \times \hat{l}$, which describes the force acting on unit length of dislocation in an external stress field. In particular, the bending force due



Figure 5.10: Strain relaxation by single (left) and double (right) kinking of a threading dislocation to form misfit dislocations.

to the coherency stress in the film is

$$\bar{F}_{\rm coh} = h\hat{z} \cdot \{(\bar{b} \cdot \bar{\sigma}) \times \hat{l}\},\tag{5.21}$$

where

$$\bar{\bar{\sigma}} = \begin{pmatrix} \sigma_{\rm coh} & 0 & 0\\ 0 & \sigma_{\rm coh} & 0\\ 0 & 0 & 0 \end{pmatrix}$$
(5.22)

is the stress tensor in the epilayer film, and, using Equation 5.4,

$$\sigma_{\rm coh} = \sigma_{\parallel} = 2\mu \left(\frac{1+\nu}{1-\nu}\right) \epsilon_{\parallel} \tag{5.23}$$

is the in-plane stress acting in the epilayer.

For the 60° misfit dislocation segment illustrated in Figure 5.5, whose Burgers vector is $\bar{b} = [\bar{1}01]b/\sqrt{2}$ and whose line direction within the interface is $\hat{l} = [110]/\sqrt{2}$, the force is

$$\bar{F}_{\rm coh} = \begin{bmatrix} 0 & 0 & h \end{bmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ -b\sigma_{\rm coh}/2 \end{pmatrix}, \qquad (5.24)$$

with a magnitude of

$$F_{\rm coh} = b\sigma_{\rm coh}h/2. \tag{5.25}$$

Since for this geometry

$$b_{\text{edg},\parallel} = b \cos \lambda = b \cos 60^\circ = b/2, \qquad (5.26)$$

the coherency force can also be written as

$$F_{\rm coh} = b_{\rm edg, \parallel} \sigma_{\rm coh} h, \tag{5.27}$$

5.2. Forces on Dislocations

which can be shown to be generally true for arbitrary geometries (see Exercise 4 at the end of this chapter).

Opposing this force is a line tension associated with the energy required to create the new misfit dislocation segment C–D. From Equation 5.11, this force, or the energy per unit length, is

$$F_{\rm dis} = \frac{\mu b^2}{4\pi} \left(\frac{1 - \nu \cos^2 \beta}{1 - \nu} \right) \ln(4h/b), \tag{5.28}$$

where we have assumed noninteracting dislocations. If we recast this equation into the form $F_{\text{dis}} = b_{\text{edg},\parallel} \sigma_{\text{dis}} h$, then we can write

$$\sigma_{\rm dis} = \frac{\mu b}{4\pi h \cos \lambda} \left(\frac{1 - \nu \cos^2 \beta}{1 - \nu} \right) \ln(4h/b). \tag{5.29}$$

which is an effective stress, $\sigma_{\rm dis}$, associated with the dislocation line tension that opposes the coherency stress, $\sigma_{\rm coh}$.

The net, or "excess" stress driving the bending of threading dislocations to form single-kink misfit segments is therefore

$$\sigma_{\text{exc}}^{\text{SK}} = \sigma_{\text{coh}} - \sigma_{\text{dis}}$$

$$= 2\mu \left(\frac{1+\nu}{1-\nu}\right) \left(f - \rho_{\text{md}} b_{\text{edg},\parallel}\right)$$

$$- \frac{\mu b}{4\pi h \cos \lambda} \left(\frac{1-\nu \cos^2 \beta}{1-\nu}\right) \ln(4h/b). \quad (5.30)$$

When $\sigma_{\rm exc}^{\rm SK} > 0$, threading dislocations will tend to bend over to form strainrelaxing misfit segments. When $\sigma_{\rm exc}^{\rm SK} < 0$, threading dislocations that have bent over to form strain-relaxing misfit segments will tend to straighten. When $\sigma_{\rm exc}^{\rm SK} = 0$, threading dislocations will have neither tendency.²⁵

Note that this excess stress reproduces exactly the variation of energy with dislocation density found in Equation 5.19, assuming noninteracting dislocations. Therefore, the condition $[\sigma_{\text{exc}}^{\text{SK}}]_{\rho_{\text{md}}=0} = 0$ is equivalent to the condition $(1/2)[\partial u_{\text{tot}}/\partial \rho_{\text{md}}]_{\rho_{\text{md}}=0} = 0$ for the thickness/misfit boundary between coherent and semi-coherent films, and the condition $\sigma_{\text{exc}}^{\text{SK}} = 0$ is equivalent to the condition $(1/2)\partial u_{\text{tot}}/\partial \rho_{\text{md}} = 0$ for the equilibrium misfit dislocation density beyond the critical layer thickness. Physically, the force required to form new misfit dislocations by bending of existing threading dislocations is equivalent to that required to increase the density of a dislocation array by "squeezing" laterally on the dislocation array.

 $^{^{25}}$ L.B. Freund, "The driving force for glide of a threading dislocation in a strained epitaxial layer on a substrate," J. Mech. Phys. Solids **38**, 657 (1990).



Figure 5.11: Contours of constant single-kink (left) and double-kink (right) excess stresses on a thickness/equivalent-strain diagram.

As we will discuss later, it is often possible, for kinetic reasons, to grow coherent epilayers to thicknesses well beyond those for which they should become semi-coherent. In these cases, the excess stress evaluated at $\rho_{\rm md} = 0$ is a useful measure for the degree of metastability of the structure. The critical thickness/misfit relationship for a given degree of metastability, or a given value of $\sigma_{\rm exc}^{\rm SK}$, is then found, from Equation 5.30, to be

$$f_{\rm c}(\sigma_{\rm exc}^{\rm SK}, h) = \frac{1}{2} \left(\frac{1-\nu}{1+\nu}\right) \frac{\sigma_{\rm exc}^{\rm SK}}{\mu} + \frac{b}{8\pi h \cos \lambda} \left(\frac{1-\nu \cos^2 \beta}{1+\nu}\right) \ln\left(4h/b\right).$$
(5.31)

These metastable critical thicknesses and misfits are illustrated in the left half of Figure 5.11 for various values of $\sigma_{\rm exc}^{\rm SK}/\mu$. For thin (≈ 150 Å) "quantum-well" type structures, which are capped immediately by unstrained material, it is often possible to grow coherent metastable structures up to values of $\sigma_{\rm exc}^{\rm SK}/\mu = 0.04$. For thicker structures, the maximum $\sigma_{\rm exc}^{\rm SK}/\mu$ values decrease considerably. In the limit of very thick structures, grown for very long times, the maximum $\sigma_{\rm exc}^{\rm SK}/\mu$ values are zero, and the equilibrium critical layer thickness boundary holds.

5.2.2 Strained Buried Layers

In Subsection 5.2.1, we described excess stress in a simple structure consisting of a strained surface layer. In this subsection, we describe excess stress in a more complicated structure consisting of a strained *buried* layer. Consider a dislocation threading upward through this buried strained layer, as illustrated in the right half of Figure 5.10. The buried strained layer has thickness $h_{\rm str}$, and is capped by an unstrained layer of thickness $h_{\rm cap}$. For this structure, it is possible for the dislocation to bend twice, thereby relieving strain only within the buried layer.²⁶ We can apply all the arguments of the previous subsection to calculate the net bending force, except that the line tension force must now be taken twice. The excess stress driving misfit dislocation formation by this macroscopic "double-kinking" of a threading dislocation is therefore

$$\sigma_{\text{exc}}^{\text{DK}} = \sigma_{\text{coh}} - \sigma_{\text{dis},1} - \sigma_{\text{dis},2}$$

$$= 2\mu \left(\frac{1+\nu}{1-\nu}\right) \left(f - \rho_{\text{md}} b_{\text{edg},\parallel}\right)$$

$$- \frac{\mu b}{4\pi R_1 \cos \lambda} \left(\frac{1-\nu \cos^2 \beta}{1-\nu}\right) \ln(4R_1/b)$$

$$- \frac{\mu b}{4\pi R_2 \cos \lambda} \left(\frac{1-\nu \cos^2 \beta}{1-\nu}\right) \ln(4R_2/b), \quad (5.32)$$

where ρ_{md} is now the density of dislocations at *each* interface surrounding the buried strained layer.

Note that the stresses associated with the dislocation line tensions are different for the two dislocations, because they may have different cutoff "screening" distances for their elastic energies. The cutoff distance for the dislocation farthest from the free surface will be approximately the distance to the dislocation closest to the free surface, or h. However, the cutoff distance for the dislocation closest to the free surface will be the smaller of the distances to the free surface, $h_{\rm cap}$, or to the adjacent dislocation, or approximately $h_{\rm eff} = h_{\rm str} h_{\rm cap}/(h_{\rm str} + h_{\rm cap})$. Therefore,

$$\sigma_{\rm exc}^{\rm DK} = 2\mu \left(\frac{1+\nu}{1-\nu}\right) \left(f - \rho_{\rm md} b_{\rm edg,\parallel}\right) - \frac{\mu b}{4\pi h_{\rm str} \cos \lambda} \left(\frac{1-\nu \cos^2 \beta}{1-\nu}\right) \ln \left(\frac{4h_{\rm str}}{b}\right) - \frac{\mu b}{4\pi h_{\rm eff} \cos \lambda} \left(\frac{1-\nu \cos^2 \beta}{1-\nu}\right) \ln \left(\frac{4h_{\rm eff}}{b}\right)$$
(5.33)

In the limit $h_{\rm cap} \rightarrow 0$ (or, to avoid singularities, $h_{\rm cap} \rightarrow b/4$), the

²⁶W.D. Nix, D.B. Noble, and J.F. Turlo, "Mechanisms and kinetics of misfit dislocation formation in heteroepitaxial thin films," *Mat. Res. Soc. Symp. Proc.* **188**, 315 (1990).

double-kink excess stress becomes

$$\sigma_{\text{exc}}^{\text{DK}} = 2\mu \left(\frac{1+\nu}{1-\nu}\right) \left(f - \rho_{\text{md}} b_{\text{edg},\parallel}\right) - \frac{\mu b}{4\pi h_{\text{str}} \cos \lambda} \left(\frac{1-\nu \cos^2 \beta}{1-\nu}\right) \ln \left(\frac{4h_{\text{str}}}{b}\right), \quad (5.34)$$

and is equivalent to the single-kink excess stress. Effectively, there is no cap, and the energy of the dislocation elongating along the surface is zero.

In the opposite limit $h_{\rm cap} \to \infty$, the double-kink excess stress becomes

$$\sigma_{\rm exc}^{\rm DK} = 2\mu \left(\frac{1+\nu}{1-\nu}\right) \left(f - \rho_{\rm md} b_{\rm edg,\parallel}\right) - \frac{\mu b}{2\pi h_{\rm str} \cos \lambda} \left(\frac{1-\nu \cos^2 \beta}{1-\nu}\right) \ln \left(\frac{4h_{\rm str}}{b}\right). \quad (5.35)$$

In this limit, the critical thickness/misfit relationship for a given degree of metastability, or a given value of $\sigma_{\text{exc}}^{\text{DK}}$, can be calculated to be

$$f_{\rm c}(\sigma_{\rm exc}^{\rm DK},h) = \frac{1}{2} \left(\frac{1-\nu}{1+\nu}\right) \frac{\sigma_{\rm exc}^{\rm DK}}{\mu} + \frac{b}{4\pi h \cos \lambda} \left(\frac{1-\nu \cos^2 \beta}{1+\nu}\right) \ln\left(\frac{4h_{\rm str}}{b}\right).$$
(5.36)

These metastable critical layer thicknesses are illustrated in the right half of Figure 5.11. The equilibrium critical layer thickness boundary is determined by $\sigma_{\rm exc}^{\rm DK} = 0$, and is seen to be shifted to the right from the single-kink curves. Because the line tension enters in twice, for strained layers having the same thickness and misfit, this double-kink mechanism for strain relaxation is usually less likely than the single-kink mechanism discussed earlier.

5.2.3 Generalized Excess Stress

In Subsections 5.2.1 and 5.2.2, we described the excess stresses associated with two strained layer structures, one in which the layer is at the surface, another in which the layer is buried. In this subsection, we describe excess stress in general structures composed of multilayers of different misfits and thicknesses. Such structures are susceptible to either single-kink or double-kink relaxation at different depths within the structure. In other words, dislocations may bend anywhere within a given structure. Then, it is useful to generalize the driving force for that bending to include a dependence on depth.²⁷

²⁷J.Y. Tsao and B.W. Dodson, "Excess stress and the stability of strained heterostructures," *Appl. Phys. Lett.* **53**, 848 (1988).

5.2. Forces on Dislocations

In doing so, we note that for many applications it is only necessary to calculate the excess stress in unrelaxed, fully coherent structures. Therefore, we restrict ourselves to the simplest case of unrelaxed ($\rho_{\rm md} = 0$) structures. Structures that are partially relaxed ($\rho_{\rm md} \neq 0$) are considerably more difficult to treat.

If we make the approximation that the elastic moduli of the different layers are equal, then for single-kink relaxation, the depth-dependent excess stress can be written as

$$\sigma_{\rm exc}^{\rm SK}(z) = 2\mu \left(\frac{1+\nu}{1-\nu}\right) \epsilon_{\rm equ}^{\rm SK}(z) - \frac{\mu b}{4\pi z \cos \lambda} \left(\frac{1-\nu \cos^2 \beta}{1-\nu}\right) \ln \left(\frac{4z}{b}\right), \quad (5.37)$$

where z is the depth from the free surface, and

$$\epsilon_{\rm equ}^{\rm SK}(z) = \int_0^z e(z') \frac{dz'}{z},\tag{5.38}$$

the equivalent strain, is the average parallel strain associated with the structure from the surface to that depth.

Physically, the coherency stress acting to bend a dislocation at a depth z is an integral of the strains over the length of the dislocation above that depth. The dislocation line tension stress acting to straighten a dislocation at a depth z is the energy associated with creating a dislocation at that depth. If $\sigma_{\text{exc}}^{\text{SK}}(z) < 0$ at a particular depth z, misfit dislocation formation at that depth leads to an increase in energy. If $\sigma_{\text{exc}}^{\text{SK}}(z) > 0$ at a particular depth z, misfit dislocation formation at that depth leads to a decrease in energy.

Note that even if $\sigma_{\text{exc}}^{\text{SK}}(z) > 0$ at a particular depth, a threading dislocation will not necessarily bend there. Kinetic limitations may prevent such bending, and there may be other depths in the structure at which the excess stress is even higher, and which will be even more favored for misfit dislocation creation.

To illustrate this concept of a depth-dependent excess stress, in Figures 5.12 and 5.13 we show $\sigma_{\text{exc}}^{\text{SK}}(z)$ for two double quantum-well heterostructures. In the unstrained caps, $\sigma_{\text{exc}}^{\text{SK}}(z)$ increases gradually from the surface, as the line tension stress associated with misfit dislocation creation decreases. In the strained layers themselves, $\sigma_{\text{exc}}^{\text{SK}}(z)$ increases more quickly, as the coherency stresses increase. In the unstrained buffers beneath the strained layers $\sigma_{\text{exc}}^{\text{SK}}(z)$ decreases, as the coherency stress associated with the strained layer is "diluted," so that the average coherency stress above a depth z decreases.

In these examples, the single-kink excess stress is maximum at the rear of the deepest buried strained layers. The rear of that strained layer is



Figure 5.12: Strains, equivalent strains, and excess stresses in a double buried quantum-well heterostructure. Both the single- and double-kink excess stresses maximize at a depth of 800 Å, but only the single-kink excess stress exceeds zero, and even then just barely. Therefore, this structure is stable with respect to double-kink strain relaxation, but slightly unstable with respect to single-kink strain relaxation.

therefore the weakest point in the structure, where misfit dislocations are most likely to form. It is important to emphasize, though, that the rear of the shallowest buried strained layer is also a weak point, at which misfit dislocations may form.

The double-kink excess stress can be generalized in a similar way:

$$\sigma_{\rm exc}^{\rm DK}(z) = 2\mu \left(\frac{1+\nu}{1-\nu}\right) \epsilon_{\rm equ}^{\rm DK}(z) - \frac{\mu b}{2\pi h \cos \lambda} \left(\frac{1-\nu \cos^2 \beta}{1-\nu}\right) \ln \left(\frac{4h}{b}\right),\tag{5.39}$$

where z is the depth of the lower kink from the free surface, h is the thickness of material between the kinks, and

$$\epsilon^{\rm DK}(z) = \int_{z}^{z+h} \epsilon(z') \frac{dz'}{h}$$
(5.40)

is the equivalent strain associated with the material between the kinks.

Note that the double-kink excess stress depends not just on depth, but on the thickness of material between the kinks. It will be maximum when it is matched to the thicknesses of the buried strained layers. For example,



Figure 5.13: Strains, equivalent strains, and excess stresses in a double buried quantum well heterostructure. Both the single- and double-kink excess stresses maximize at a depth of 800 Å, but only the single-kink excess stress exceeds zero. Therefore, this structure is stable with respect to double-kink strain relaxation, but unstable with respect to single-kink strain relaxation.

Figures 5.12 and 5.13 show the depth-dependent double-kink excess stress for the double buried quantum-well heterostructures evaluated at h equal to the actual thicknesses, $h_{\rm str}$, of the buried quantum wells. The excess stresses can be seen to be small everywhere except in the quantum wells themselves.

Also note that even at their maxima, at the rear of the buried strained layers, the double-kink excess stresses are less than the single-kink excess stresses. These particular buried structures will therefore be more likely to relax by generation of single-kink rather than double-kink misfit dislocations.

If we require a structure to be absolutely stable, both with respect to single-kink and double-kink relaxation, then we require $\sigma_{\text{exc}}^{\text{SK}}(z) < 0$ and $\sigma_{\text{exc}}^{\text{DK}}(z) < 0$ for all z. For the single buried strained layer structure shown in Figure 5.14, whose weakest point is at a depth $z = h_{\text{cap}} + h_{\text{str}}$, where h_{cap} is the thickness of the unstrained capping layer, we must then satisfy both

$$f_{\rm equ}^{\rm SK} \leq \frac{b}{8\pi h_{\rm equ}^{\rm SK} \cos \lambda} \left(\frac{1-\nu \cos^2 \beta}{1+\nu}\right) \ln \left(\frac{4h_{\rm equ}^{\rm SK}}{b}\right)$$
(5.41)



Figure 5.14: Stability curves for structure consisting of a strained layer of thickness $h_{\rm str}$ and misfit $f_{\rm str}$, buried underneath a capping layer of thickness $h_{\rm cap}$ and misfit $f_{\rm cap}$.

$$f_{\rm equ}^{\rm DK} \leq \frac{b}{4\pi h_{\rm equ}^{\rm DK} \cos \lambda} \left(\frac{1 - \nu \cos^2 \beta}{1 + \nu} \right) \ln \left(\frac{4h_{\rm equ}^{\rm DK}}{b} \right), \qquad (5.42)$$

where

$$\epsilon_{equ}^{SK} = \frac{f_{str}h_{str}}{h_{cap} + h_{str}}$$

$$h_{equ}^{SK} = h_{cap} + h_{str}$$

$$\epsilon_{equ}^{DK} = f_{str}$$
(5.43)

$$h_{\rm equ}^{\rm DK} = h_{\rm str}$$
 (5.44)

are the single- and double-kink equivalent strains and thicknesses. Note that we have made the approximation $h_{\rm equ}^{\rm DK} = h_{\rm str} h_{\rm cap} / (h_{\rm str} + h_{\rm cap}) = h_{\rm str}$ for the double-kink equivalent thicknesses.

Substituting Equations 5.43 into Equation 5.41, we find that for a given thickness cap, $h_{\rm cap}$, and a given thickness buried strained layer, $h_{\rm str}$, the critical strained layer misfit below which the structure will be stable with respect to single-kink relaxation is

$$f_{\rm str} = \frac{h_{\rm cap} + h_{\rm str}}{h_{\rm str}} \frac{b}{8\pi (h_{\rm cap} + h_{\rm str}) \cos \lambda} \left(\frac{1 - \nu \cos^2 \beta}{1 + \nu}\right) \ln \left[\frac{4(h_{\rm cap} + h_{\rm str})}{b}\right].$$
(5.45)

The resulting critical layer misfits for a given thickness (or, equivalently, the critical layer thicknesses for a given misfit) are shown as the dashed curves in Figure 5.14 for various capping layer thicknesses. Buried strained layer thickness and misfit combinations that lie to the left of the curves are stable with respect to single-kink relaxation.

If the cap thickness is zero, then the critical thickness curve (leftmost solid curve) for simple surface strained layers is obtained. As the cap becomes thicker, the critical thickness curves shift to the right, as the unstrained cap "dilutes" the strain of the buried layer and stabilizes it. Ultimately, for infinitely thick caps, the structure becomes more unstable with respect to double-kink relaxation than to single-kink relaxation. In this limit, the critical strained layer misfit below which the structure will be stable is found by substituting Equations 5.44 into Equation 5.42, or

$$f_{\rm str} = \frac{b}{4\pi h_{\rm str} \cos \lambda} \left(\frac{1 - \nu \cos^2 \beta}{1 + \nu} \right) \ln \left(\frac{4h_{\rm str}}{b} \right). \tag{5.46}$$

The resulting critical layer misfit for a given thickness (or, equivalently, the critical layer thickness for a given misfit) is shown as the rightmost solid curve in Figure 5.14. Buried strained layer thickness and misfit combinations that lie to the right of the curve are not stable with respect to double-kink relaxation.

To the left of the leftmost solid curve, then, structures are absolutely stable with respect to both single- and double-kink relaxation, regardless of cap thickness. Within the window between the solid curves, structures are absolutely stable with respect to double-kink relaxation, but require stabilization with respect to single-kink relaxation by a finite-thickness cap layer. To the right of the rightmost solid curve, structures are not stable with respect to double-kink relaxation, even if they have been stabilized against single-kink relaxation by an infinitely thick cap layer.

5.3 Relaxation of Strain

In Sections 5.1 and 5.2, we described the *thermodynamics* of the creation of misfit dislocations. In particular, we described the driving force, or "excess stress," acting to bend vertical dislocation segments into misfit dislocations lying in the interface between the strained layer and the substrate. Basically, the sign of $\sigma_{\rm exc}$ determines whether or not misfit dislocations will have a tendency to form, while the magnitude of $\sigma_{\rm exc}$ determines the driving force for them to form. Even if $\sigma_{\rm exc} > 0$, however, misfit dislocations will not form instantly during growth.²⁸ Instead, they will form at a finite

²⁸A.T. Fiory, J.C. Bean, R. Hull, and S. Nakahara, "Thermal relaxation of metastable strained-layer $Ge_x Si_{1-x}/Si$ epitaxy," *Phys. Rev.* B31, 4063 (1984); and E. Kasper, "Growth and properties of Si/SiGe superlattices," *Surface Sci.* 174, 630 (1986).

rate determined, to first order, by the magnitude of $\sigma_{\rm exc}$.

In this section, we discuss the *kinetics* of the creation of misfit dislocations. We begin, in Subsection 5.3.1, with a brief introduction to the dynamics of dislocations in bulk materials, as summarized in what are known as deformation-mechanism maps. Then, in Subsection 5.3.2, we describe a simple qualitative model for the dynamics of dislocations in epitaxial thin films, and use it to simulate, in an approximate way, the evolution of strain and misfit dislocation density during actual growth and processing of strained heterostructures. Finally, in Subsection 5.3.3, we discuss the construction of stability diagrams, which describe the stress-temperaturetime regimes within which strained heterostructures will be metastable to various amounts of relaxation.

5.3.1 Deformation Mechanism Maps

Let us begin, in this subsection, with a brief introduction to the dynamics of dislocations in bulk materials. At the outset, we note that the mechanisms underlying the introduction, motion, and multiplication of dislocations in bulk materials are exceedingly complex. The mechanisms are many, and each may be important only under certain conditions. To illustrate this, let us first consider some of the ways in which *bulk* materials deform plastically under the application of externally imposed stresses. The classic way of representing the plastic deformation of bulk materials is through the use of deformation-mechanism maps.²⁹ These maps are stress-temperature diagrams on which are indicated regimes within which various mechanisms for plastic deformation.

Consider, for example, the deformation-mechanism maps illustrated in Figures 5.15 and 5.16 for Si and Ge, respectively. At relatively low temperatures and high stresses, deformation is dominated by "low-temperature plasticity," in which dislocations move mainly by conservative motion, or glide, within the plane containing both the dislocation line and its Burgers vector. At relatively high temperatures and moderate stresses, deformation is dominated by "power-law creep," in which dislocations are increasingly able to move by the nonconservative motion, or climb, of dislocations out of the plane containing both the dislocation line and its Burgers vector. At the lowest stresses, deformation of polycrystalline materials is dominated by "diffusional flow," in which, even in the absence of dislocations, grain boundaries move and change shape via diffusion of matter through the grains or along the grain boundaries themselves.

 $^{^{29}\}mathrm{H.J.}$ Frost and M.F. Ashby, Deformation-Mechanism~Maps (Pergamon, Oxford, 1982).



Figure 5.15: Stress-temperature deformation mechanism map for silicon of grain size 100 μm .^{*a*} Iso-strain rate contours are drawn from 1/s to 10⁻¹⁰/s.

^aReprinted from H.J. Frost and M.F. Ashby, *Deformation-Mechanism Maps* (Pergamon, Oxford, 1982), p. 71.



Figure 5.16: Stress-temperature deformation mechanism map for germanium of grain size 100 μm .^{*a*} Iso-strain rate contours are drawn from 1/s to 10⁻¹⁰/s.

^aReprinted from H.J. Frost and M.F. Ashby, *Deformation-Mechanism Maps* (Pergamon, Oxford, 1982), p. 73.

Superimposed on these diagrams are iso-strain rate contours, which indicate the temperature-dependent stresses required to cause a given strain rate. Generally, lower stresses are required to cause a given strain rate at higher temperatures, due to increased dislocation mobilities. This increased dislocation mobility may occur for a number of reasons. For example, the nucleation rate of microscopic double kinks, by which dislocations glide laterally on an atomic scale, may increase. The rate at which dislocations pass through obstacles may increase. The rate at which vacancies diffuse to and from dislocations may also increase, thereby increasing the rate at which dislocations climb.

We emphasize that deformation-mechanism maps represent an enormous simplification of a number of complex mechanisms, and can only be a rough guide to deformation behavior. In particular, their construction requires the assumption of a particular microstructure, e.g., dislocation density and, in polycrystalline materials, grain size. As materials deform, however, their microstructure will change; if the change is severe, the corresponding change in the deformation-mechanism map may also be quite severe. In other words, a complete picture of plastic deformation must include the time evolution of dislocation densities and other aspects of microstructure, and how that evolving microstructure in turn influences the further evolution of dislocation densities.

5.3.2 A Simple Phenomenological Model

In Subsection 5.3.1, we discussed briefly plastic deformation in bulk materials. The geometry of thin film single-crystal heterostructures is much simpler than that of a bulk polycrystalline material, and so in principle should be correspondingly easier to treat. However, this has not yet proven so. A general treatment of the plastic deformation of thin film strained heterostructures must itself include a number of complex microscopic mechanisms. In this subsection, we briefly discuss these microscopic mechanisms, and then discuss a simple phenomenological model based on these mechanisms.

Consider the microscopic mechanisms illustrated in Figure 5.17. First, because the initial threading dislocation densities in electronic-grade semiconductor substrates are exceedingly low and cannot by themselves account for the amounts of strain relaxation commonly observed, nucleation of new dislocation loops must be included.³⁰ These loops are most likely "half-

³⁰P.M.J. Marée, J.C. Barbour, J.F. van der Veen, K.L. Kavanagh, C.W.T. Bulle-Lieuwma, and M.P.A. Viegers, "Generation of misfit dislocations in semiconductors," *J. Appl. Phys.* **62**, 4413 (1987); and R. People and J.C. Bean, "Calculation of critical layer thickness versus lattice mismatch for $\text{Ge}_x \text{Si}_{1-x}/\text{Si}$ strained-layer heterostructures,"



Figure 5.17: Schematic illustrations of possible microscopic deformation mechanisms operative during strain relaxation.

loops" nucleated at the free surface, perhaps catalyzed by defects or other stress concentrators.³¹ In compound semiconductors, the situation is even more complicated, due to possible dependences of the nucleation rate on surface conditions and chemistry.

Second, the outward "bowing" motion of these dislocation half-loops to form misfit dislocation segments at the epilayer/substrate interface must be included.³² This motion will be dominated by glide at low temperatures, but will increasingly have a climb component at higher temperatures. It may also be mediated by nucleation of microscopic single kinks at the free surface in very thin films, or by nucleation of microscopic double kinks in thicker films.³³

Third, the slowing and occasional pinning of these dislocation halfloops as they move and encounter other dislocation segments must be included.³⁴ Such pinning has been observed during *in situ* transmission

Appl. Phys. Lett. 47, 322 (1985) and 49, 229 (1986).

³¹B.W. Dodson, "Nature of misfit dislocation sources in strained-layer semiconductor structures," *Appl. Phys. Lett.* **53**, 394 (1988); C.J. Gibbings, C.G. Tuppen, and M. Hockly, "Dislocation nucleation and propagation in Si_{0.95}Ge_{0.05} layers on silicon," *Appl. Phys. Lett.* **54**, 148 (1989); and D.J. Eaglesham, E.P. Kvam, D.M. Maher, C.J. Humphreys, and J.C. Bean, "Dislocation nucleation near the critical thickness in GeSi/Si strained layers" *Phil. Mag.* **A59**, 1059 (1989).

 $^{^{32}}$ J.W. Matthews, S. Mader, and T.B. Light, "Accommodation of misfit across the interface between crystals of semiconducting elements or compounds," *J. Appl. Phys.* **41**, 3800 (1970).

³³R. Hull, J.C. Bean, D. Bahnck, L.J. Peticolas, Jr., K.T. Short, and F.C. Unterwald, "Interpretation of dislocation propagation velocities in strained $\text{Ge}_x \text{Si}_{1-x}/\text{Si}(100)$ heterostructures by the diffusive kink pair model," J. Appl. Phys, to be published.

 $^{^{34}}$ L.B. Freund, "A criterion for arrest of a threading dislocation in a strained epitaxial layer due to an interface misfit dislocation in its path" J. Appl. Phys. **68**, 2073 (1990).

electron microscopy,³⁵ and is likely to be extremely important in the later stages of strain relaxation,³⁶ when both crossed grids of dislocations have become quite dense.

Fourth, the unpinning and possible multiplication³⁷ of dislocation halfloops as they "bow through" obstacles such as other dislocations must be included. These processes have not been directly observed, but may be important in the later stages of strain relaxation.

Finally, the way in which all of these microscopic processes depend on depth within the structure must be included. Treating depth-dependent strain relaxation would represent a nontrivial extension of existing theories, but would be particularly important for compositional graded strained heterostructures, or for strained heterostructures composed of multiple layers, each having its own misfit.

As a consequence, all current models treat only some of these processes, and even then only in simplified ways. For concreteness, let us consider here one model,³⁸ based on the phenomenology of deformation in bulk diamond-structure materials.³⁹ The model is not the most complete,⁴⁰ but is simple and predicts at least qualitatively much of what is known about strain relaxation.

The model assumes that dislocations multiply at a rate proportional to (a) the velocity at which they move, (b) the number of dislocations present, and (c) the excess stress. If the number of dislocations is itself proportional to the amount of strain relaxation, γ , and if the dislocation glide and climb velocities are both thermally activated and proportional to the excess stress, then we can write

$$\frac{d\gamma}{dt} = \frac{\sigma_{\rm exc}^2(\gamma)}{\mu^2} \left(\Gamma_{\rm g} e^{-Q_{\rm g}/kT} + \Gamma_{\rm c} e^{-Q_{\rm c}/kT} \right) (\gamma + \gamma_o) , \qquad (5.47)$$

 40 See, e.g., D.C. Houghton, "Strain relaxation kinetics in $Si_{1-x}Ge_x/Si$ heterostructures," J. Appl. Phys. **70**, 2136 (1991), and Exercise 6 at the end of this chapter.

³⁵R. Hull and J.C. Bean, "Variation in misfit dislocation behavior as a function of strain in the GeSi/Si system" Appl. Phys. Lett. 54, 925 (1989).

³⁶B.W. Dodson, "Work hardening and strain relaxation in strained-layer buffers," *Appl. Phys. Let.* **53**, 37 (1988).

³⁷W. Hagen and H. Strunk, "A new type of source generating misfit dislocations," *Appl. Phys.* **17**, 85 (1978).

³⁸B.W. Dodson and J.Y. Tsao, "Relaxation of strained-layer semiconductor structures via plastic flow," *Appl. Phys. Lett.* **51**, 1325-1327 (1987); B.W. Dodson and J.Y. Tsao, "Erratum: Relaxation of strained-layer semiconductor structures via plastic flow," *Appl. Phys. Lett.* **52**, 852 (1988); and R. People, "Comment on 'Relaxation of strained-layer semiconductor structures via plastic flow," *Appl. Phys. Lett.* **53**, 1127 (1988).

³⁹H. Alexander and P. Haasen, "Dislocations and plastic flow in the diamond structure," in *Solid State Physics* Vol. 22, F. Seitz and D. Turnbull, Eds. (Academic Press, New York, 1968), pp. 27-158.

or

$$\frac{d\ln(\gamma+\gamma_o)}{dt} = \frac{\sigma_{\rm exc}^2(\gamma)}{\mu^2} \left(\Gamma_{\rm g} e^{-Q_{\rm g}/kT} + \Gamma_{\rm c} e^{-Q_{\rm c}/kT} \right).$$
(5.48)

In these equations, $\Gamma_{\rm g}$ and $\Gamma_{\rm c}$ are glide and climb rate prefactors, $Q_{\rm g}$ and $Q_{\rm c}$ are glide and climb activation energies, and γ_o represents a constant "source" term.

Note that the form of Equation 5.48 is general, but the actual values of the kinetic parameters depend on the orientation of the slip planes with respect to the epilayer/substrate interface, and on the direction of slip within those planes. More general treatments can be formulated by replacing the excess in-plane stress with the excess stress resolved on the slip plane and acting in the direction of slip within that plane.⁴¹ For (001) oriented films in the Si_{1-x}Ge_x system, approximate fits to relaxation data give⁴² rate prefactors of $\Gamma_{\rm g} = 2 \times 10^{10} \, {\rm s}^{-1}$ and $\Gamma_{\rm c} = 3 \times 10^{21} \, {\rm s}^{-1}$, a stress-dependent glide activation energy of $Q_{\rm g} = Q_{{\rm g},o}[1 - \sigma_{\rm exc}(\gamma)/\sigma_o]$, where $Q_{{\rm g},o} = 16kT_{\rm m}$ and $\sigma_o \approx 0.1\mu$, a stress-independent climb activation energy of $Q_{\rm c} = 30kT_{\rm m}$, and a "source" term of magnitude $\gamma_o \approx 10^{-4}$. Here, $T_{\rm m}$ is the melting temperature of the Si_{1-x}Ge_x alloy.

Note that the excess stress in Equation 5.48 depends nonlinearly on the actual equivalent strain, ϵ_{equ} which in turn depends on the degree of relaxation, γ :

$$\epsilon_{\rm equ} = f_{\rm equ} - \gamma. \tag{5.49}$$

Therefore, Equation 5.48 is a highly nonlinear differential equation whose full solution requires numerical techniques. However, for practical device heterostructures which are adversely affected by dislocations, small amounts of relaxation ($\gamma < 10^{-3}$) are often of greatest interest. Since these relaxations are less than the unrelaxed equivalent strains in typical structures, the excess stresss may be considered independent of the amount of relaxation. Then, it is straightforward to integrate Equation 5.48 numerically to deduce the time-dependent strain relaxation, γ , and by differentiation to deduce the time-dependent strain relaxation rate, $\dot{\gamma}$.

5.3.3 Time, Temperature and Excess Stress

In Subsection 5.3.2, we described a simple phenomenological model for the relaxation of excess stress and strain. In this subsection, we illustrate the

186

 $^{^{41}}$ L.B. Freund, "The driving force for glide of a threading dislocation in a strained epitaxial layer on a substrate," J. Mech. Phys. Solids **38**, 657 (1990).

 $^{^{42}}$ R. Hull, J.C. Bean, D.J. Werder, and R.E. Leibenguth, "In situ observations of misfit dislocation propagation in Ge_xSi_{1-x}/Si(100) heterostructures," Appl. Phys. Lett. **52**, 1605 (1988); and B.W. Dodson and J.Y. Tsao, "Non-Newtonian strain relaxation in highly strained SiGe heterostructures," Appl. Phys. Lett. **53**, 2498 (1988).

time evolution of excess stress and strain⁴³ in the two simple structures shown in Figures 5.12 and 5.13. We imagine that, as the two structures are grown, they are subjected to the hypothetical (but realistic) temperature cycles shown in Figures 5.18 and 5.19. In each case, growth of a 600-Å Si buffer layer at 750°C commences at t = 0. Toward the end of growth of this buffer, the temperature is ramped down to 550°C for growth of the two 100-Å-thick buried Si_{1-x}Ge_x strained layers. During growth of the strained layers, the single-kink excess stresses (evaluated at the rear of the deepest strained layer) increase, but during growth of the unstrained spacer and capping layers they decrease. At the end of growth, the temperature is ramped down to room temperature (25°C). Finally, we have included the possibility of a 45-s, 900°C post-growth rapid thermal anneal for dopant activation or oxide growth.

For the weakly strained structure in Figure 5.18, the final structure has a single-kink excess stress that is barely positive, and so is fairly resistant to strain relief by plastic flow. Even the 45-s rapid thermal anneal at 900°C causes a strain relaxation less than 10^{-7} . Note that this amount of relaxation may be considered nearly unobservable, even by dislocation counting, since it corresponds to on the order of one dislocation per centimeter. Note also that just after growth of the final strained layer the structure passes through an intermediate structure for which the single-kink excess stress is greatest. However, because the growth temperature is low, negligible strain relaxation occurs.

For the moderately strained structure shown in Figure 5.19, the final structure has a single-kink excess stress that is larger, and hence is less resistant to strain relief by plastic flow. Indeed, during the 45-s rapid thermal anneal at 900°C, strain relaxation is significant. Again, note that just after growth of the final strained layer the structure passes through an intermediate structure for which the single-kink excess stress is greatest. However, because the growth temperature is low, negligible strain relaxation occurs.

5.3.4 Stability Diagrams

From the discussions in Subsections 5.3.1-5.3.3, it is clear that the major determinant of the stability of coherent strained heterostructures is its excess stress, convolved with the time-temperature cycle that it experiences during growth and processing. If the excess stress everywhere in the structure is at all times less than zero, then the coherent structure is absolutely stable. If, during some time interval, the excess stress anywhere in the

⁴³J.Y. Tsao and B.W. Dodson, "Time, temperature and excess stress: relaxation in strained heterostructures," *Surf. Sci.* **228**, 260 (1990).



Figure 5.18: Thickness, excess stress (evaluated at the rear of the deepest strained layer), temperature, strain rate, and strain relaxation in a slightly metastable double buried quantum-well heterostructure.

structure rises above zero, then some strain relaxation will occur. However, the *amount* of strain relaxation may be small if the temperature during that time interval is low, or if the duration of the time interval is short. In other words, it is the *time at temperature while the excess stress is highest* that determines whether significant strain relaxation will occur.

For a given time duration, then, the two parameters that most directly determine the amount of strain relaxation that will occur are the excess



Figure 5.19: Thickness, excess stress (evaluated at the rear of the deepest strained layer), temperature, strain rate, and strain relaxation in a highly metastable double buried quantum-well heterostructure.

stress and temperature.⁴⁴ In this subsection, we describe the use of excess stress versus temperature stability diagrams for depicting various regimes of strain relaxation. To illustrate, we show, in Figure 5.20, contours of constant strain relaxation plotted on a stress-temperature diagram. The contours were calculated according to the simple phenomenological model described earlier by Equation 5.48, and so should only be taken as qual-

 $^{^{44}}$ J.Y. Tsao, B.W. Dodson, S.T. Picraux, and D.M. Cornelison, "Critical stresses for $Si_{1-x}Ge_x$ strained-layer plasticity," *Phys. Rev. Lett.* **59**, 2455 (1987).



Figure 5.20: Stress-temperature stability diagrams for strained $\text{Si}_{1-x}\text{Ge}_x$ heterostructures, assuming times-at-stress of 10 s (left) and 10 min (right). Structures lying below the solid lines are absolutely stable. Structures lying below the dashed lines will have relaxed by less than $\approx 10^{-7}$. Structures lying below the dot-dashed lines will have relaxed by less than $\approx 10^{-3}$.

itative guides. Nevertheless, they illustrate how such kinetic models can be used to construct these "stability" diagrams. Such diagrams are *practical* guides to the degree of relaxation that can be expected for a given structure.

The diagram on the left in Figure 5.20 was calculated assuming a "timeat-stress" of 10 s. Such a diagram would be appropriate for the growth of a buried strained quantum well, in which the excess stress of the structure reaches its maximum just after the buried strained layer has been grown, but diminishes quickly thereafter upon initiation of growth of the unstrained capping layer.

The diagram on the right in Figure 5.20 was calculated assuming a timeat-stress of 10 min. Such a diagram would be appropriate for the growth of a thick surface strained layer, in which the excess stress of the structure reaches its maximum gradually during growth, and persists during the cooldown after growth has terminated.

In both diagrams, structures lying below the solid lines, whose excess stresses are less than zero, are absolutely stable. Structures lying below the dashed lines will have relaxed by less than $\approx 10^{-7}$. This amount of strain relaxation is essentially negligible, because it corresponds to on the order of one misfit dislocation per centimeter. Structures lying below the

dot-dashed lines will have relaxed by less than $\approx 10^{-3}$. This amount of strain relaxation is not negligible, because it corresponds to on the order of one misfit dislocation per micrometer.

Note that the definition of the stress-temperature boundary at which strain relaxation just becomes observable depends on the sensitivity of the technique used to measure the relaxation.⁴⁵ On the one hand, if the measurement technique is sensitive to isolated dislocations in a large field of view, as x-ray topography or etch-pit delineation might be, then at high temperatures the critical stresses approach zero. On the other hand, if the measurement technique is less sensitive (e.g., x-ray diffraction or ion-beam channeling), then the critical stresses may differ significantly from zero, and various degrees of metastability will be observed.

Suggested Reading

- D. Hull and D.J. Bacon, Introduction to Dislocations, 3rd Ed., International Series on Materials Science and Technology, Vol. 37 (Pergamon Press, Oxford, 1984).
- 2. J.P. Hirth and J. Lothe, *Theory of Dislocations*, 2nd Ed. (John Wiley and Sons, New York, 1982).
- T.P. Pearsall, Volume Ed., Strained-Layer Superlattices: Materials Science and Technology, Semiconductors and Semimetals Vol. 33, R.K. Willardson and A.C. Beer, Series Eds. (Academic Press, Boston, 1991).
- H.J. Frost and M.F. Ashby, Deformation-Mechanism Maps (Pergamon, Oxford, 1982).

Exercises

1. An alternative route to misfit accommodation involves tilting of the epitaxial layer with respect to the substrate.⁴⁶ The interface con-

⁴⁵I.J. Fritz, "Role of experimental resolution in measurements of critical layer thickness for strained-layer epitaxy," *Appl. Phys. Lett.* **51**, 1080 (1987).

⁴⁶See, e.g., H. Brooks, "Theory of internal boundaries," in *Metal Interfaces* (American Society of Metals, 1952), pp. 20-64; W.A. Jesser, "On the extension of Frank's formula to crystals with different lattice parameters," *Phys. Stat. Sol.* **A20**, 63 (1973); G.H. Olsen and R.T. Smith, "Misorientation and tetragonal distortion in heteroepitaxial vapor-grown III-V structures," *Phys. Stat. Sol.* **A31**, 739 (1975); R. Du and C.P. Flynn, "Asymmetric coherent tilt boundaries formed by molecular beam epitaxy," *J. Phys.* **C2**, 1335 (1990); and J.E. Ayers, S.K. Ghandhi, and L.J. Schowalter, "Crystallographic



Figure 5.21: Untilted (left) and tilted (right) epitaxy.

tains, instead of "misfit" dislocations with Burgers vectors parallel to the interface, "tilting" dislocations with Burgers vectors *perpendicular* to the interface, and forms what is known as an asymmetric tilt boundary.⁴⁷ Consider the one-dimensional boundary shown in Figure 5.21, containing a linear array of such tilting dislocations. Show that the parallel strain in the epitaxial layer decreases with tilt angle, θ , according to

$$\epsilon = 1 - \frac{1 - f}{\cos \theta},\tag{5.50}$$

where f is the misfit between the epitaxial layer and the substrate. Then show that the dislocation density increases with tilt angle according to

$$\rho \approx \theta / b_{\text{edg},\perp},$$
(5.51)

where $b_{\text{edg},\perp}$ is the magnitude of the edge component of the Burgers vector perpendicular to the interface.

Finally, calculate and compare the dislocation density dependence of the coherency strain and dislocation array energies, and deduce the "critical layer thickness" associated with strain relaxation by tilting dislocations. Is the critical layer thickness greater than or less than that associated with strain relaxation by misfit dislocations? Does the total energy increase or decrease at first for small tilts? All other things equal, which is more likely — strain relaxation by tilting or misfit dislocations?

2. A second alternative route to misfit accommodation is through the introduction of islanding or surface roughness.⁴⁸ Consider the two

tilting of heteroepitaxial layers," submitted to J. Cryst. Growth.

⁴⁷S. Amelinckx and W. Dekeyser, "The structure and properties of grain boundaries," in *Solid State Physics*, Vol. 8, F. Seitz and D. Turnbull, Eds. (Academic Press, New York, 1959), pp. 325–499.

⁴⁸D.J. Eaglesham and M. Cerullo, "Dislocation-free Stranski-Krastanow growth of Ge



Figure 5.22: Uncorrugated (left) and corrugated (right) epitaxial strained layers.

structures in Figure 5.22, the left composed of a planar strained layer of thickness h, the right composed of a strained layer which has developed a one-dimensional sinusoidal corrugation $A \sin(2\pi x/\lambda)$.

Suppose, due to lateral relaxation, that the strain in the corrugated (dotted) part of the strained layer is reduced to zero.⁴⁹ What is the total coherency energy, $u_{\rm coh}$, associated with the entire strained layer? Suppose that the surface energy per unit area of the strained layer is γ . What is the surface energy, $u_{\rm surf}$, due to the corrugation? How does the total energy, $u_{\rm coh}+u_{\rm surf}$, depend on A and λ ? For what values of λ does it decrease with increasing A, and hence for what wavelengths is the surface unstable to roughening.⁵⁰ Are corrugations more or less likely in high or low surface energy systems? How might a surface that lowers the surface energy make corrugations *less* likely?⁵¹

3. Even in relaxed films there may be a thermodynamic driving force for surface roughening. Qualitatively, how does the *equilibrium* energy

on Si (100)," Phys. Rev. Lett. **64**, 1943 (1990); S. Guha, A. Madhukar, and K.C. Rajkumar, "Onset of incoherency and defect introduction in the initial stages of molecular beam epitaxical growth of highly strained $In_{1-x}Ga_xAs$ on GaAs (100)," Appl. Phys. Lett. **57**, 2110 (1990); and K. Sakamoto, T. Sakamoto, S. Nagao, G. Hashiguchi, K. Kuniyoshi, and Y. Bando, "Reflection high-energy electron diffraction intensity oscillations during Ge_xSi_{1-x} MBE growth on Si (001) substrates," Jpn. J. Appl. Phys. **26**, 666 (1987).

 $^{^{49}}$ This is a very crude assumption; for better assumptions see, e.g., S. Luryi and E. Suhir, "New approach to the high quality epitaxial growth of lattice-mismatched materials," *Appl. Phys. Lett.* **49**, 140 (1986).

⁵⁰D.J. Srolovitz, "On the stability of surfaces of stressed solids," Acta Metall. **37**, 621 (1989); and C.W. Snyder, B.G. Orr, D. Kessler, and L.M. Sander, "Effect of strain on surface morphology in highly strained InGaAs films," Phys. Rev. Lett. **66**, 3032 (1991).

⁵¹M. Coppel, M.C. Reuter, E. Kaxiras, and R.M. Tromp, "Surfactants in epitaxial growth," *Phys. Rev. Lett.* **63**, 632 (1989).

per unit volume, u_{tot}/h , of a misfitting layer depend on the height of the film, both below and above the critical layer thickness? Just at the critical layer thickness, can the film reduce its energy by decomposing into some regions infinitesimally thicker, and other regions infinitesimally thinner?

4. Consider a misfit dislocation lying along the *y*-axis, as illustrated in Figure 5.23. Its Burgers vector \bar{b} can be defined either by the pair of angles λ and δ , or by the pair of angles α and β . Show that the Burgers vector of the dislocation is

$$\bar{b} = b \begin{pmatrix} \cos \lambda \\ \sin \lambda \cos \delta \\ \sin \lambda \sin \delta \end{pmatrix} = b \begin{pmatrix} \sin \alpha \sin \beta \\ \cos \beta \\ \cos \alpha \sin \beta \end{pmatrix}, \quad (5.52)$$

and that the Peach-Koehler coherency force acting to create unit length of the dislocation is

$$F_{\rm coh} = h\hat{z} \cdot \left[(\bar{b} \cdot \bar{\sigma}) \times \hat{l} \right]$$

= $b\sigma_{\rm coh}h\cos\lambda = b\sigma_{\rm coh}h\sin\alpha\sin\beta$
= $b_{\rm edg,\parallel}\sigma_{\rm coh}h.$ (5.53)

For a $\beta = 60^{\circ}$ dislocation with, as illustrated in Figures 5.4 and 5.5, $\bar{l} = [110]$ and $\bar{b} = [\bar{1}0\bar{1}]/2$, what are the angles α , λ , γ , and δ ?

5. Consider the double quantum well structure shown in Figures 5.24, in which two strained quantum wells of thicknesses $h_{\rm str}$ and strains $f_{\rm str}$ are spaced apart by an unstrained layer of thickness $h_{\rm spa}$, and capped by an unstrained layer of thickness $h_{\rm cap}$. Given $h_{\rm str}$ and $f_{\rm str}$, what must $h_{\rm cap}$ be in order for the structure to be stable with respect to misfit dislocation formation at a depth $z = z_1$? What must $h_{\rm spa}$ be in order for the structure to be stable with respect to misfit dislocation formation at a depth $z = z_2$? Derive expressions for the time-evolution of the single-kink excess stresses at z_1 and z_2 during growth of the structure. 6. Suppose one considers two types of dislocations, misfit dislocations lying in the epilayer/substrate interface, with density $\rho_{\rm md}$, and dislocation segments threading upward to the epilayer surface, with density $n_{\rm td}$. The units of $\rho_{\rm md}$ and $n_{\rm td}$ are cm⁻¹ and cm⁻², respectively. Suppose that misfit dislocations are created exclusively by lateral bending of threading segments at velocity v; that threading segments are created exclusively by half-loop nucleation at the free surface at a rate



Figure 5.23: Angles commonly used to define misfit dislocations and their Burgers vectors. For convenience, the dislocation line, \bar{l} , is taken to be oriented along the y-axis. Left: λ is the angle between (a) the Burgers vector and (b) the direction that is both normal to the dislocation line and within the plane of the interface; δ is the angle between (a) the dislocation line and (b) the projection of the Burgers vector onto the plane containing the dislocation that is perpendicular to the plane of the interface. Right: β is the angle between (a) the Burgers vector and (b) the dislocation line; α is the angle between (a) the slip plane containing both \bar{b} and \bar{l} and (b) the perpendicular to the plane of the interface; and γ is the angle between (a) the slip plane containing both \bar{b} and \bar{l} and (b) the slip plane containing both \bar{b} and \bar{l} and (b) the slip plane containing both \bar{b} and \bar{l} and (b) the slip plane containing both \bar{b} and \bar{l} and (b) the slip plane containing both \bar{b} and \bar{l} and (b) the slip plane containing both \bar{b} and \bar{l} and (b) the slip plane containing both \bar{b} and \bar{l} and (b) the slip plane containing both \bar{b} and \bar{l} and (b) the slip plane containing both \bar{b} and \bar{l} and (b) the slip plane containing both \bar{b} and \bar{l} and (b) the slip plane containing both \bar{b} and \bar{l} and (b) the slip plane containing both \bar{b} and \bar{l} and (b) the slip plane containing both \bar{b} and \bar{l} and (b) the slip plane containing both \bar{b} and \bar{l} and (b) the slip plane containing both \bar{b} and \bar{l} and (b) the slip plane containing both \bar{b} and \bar{l} and (b) the slip plane containing both \bar{b} and \bar{l} and (b) the plane of the interface.



Figure 5.24: A generic strained double quantum-well heterostructure.

j; that threading segments are pinned with probability η by interactions with misfit dislocations; and that multiplication of threading segments by interactions with misfit dislocations is negligible. Show that

$$\dot{\rho}_{\rm md} = v n_{\rm td} \dot{n}_{\rm td} = j - \eta v n_{\rm td} \rho_{\rm md}$$

$$(5.54)$$

are a set of coupled first-order differential equations for the time evolution of the two kinds of dislocation densities. 52

7. Suppose again that the nucleation rate of dislocation half-loops is j

 $^{^{52}}$ R. Hull, J.C. Bean, and C. Buescher, "A phenomenological description of strain relaxation in Ge_xSi_{1-x}/Si(100) heterostructures," J. Appl. Phys. **66**, 5837 (1989).

and that the velocity at which they propagate to form misfit dislocations is v. Suppose the maximum dislocation propagation length is l, due to pinning by lithographically fabricated boundaries.⁵³ Show that the misfit dislocation creation rate in the low velocity limit is $\dot{\rho_{md}} = vjt$. What is the misfit dislocation creation rate in the high velocity limit? Does this rate increase or decrease with a decrease in the spatial scale of the lithographic patterning?

⁵³E.A. Fitzgerald, G.P. Watson, R.E. Proano, D.G. Ast, P.D. Kirchner, G.D. Pettit, and J.M. Woodall, "Nucleation mechanisms and the elimination of misfit dislocations at mismatched interfaces by reduction in growth area," *J. Appl. Phys.* **65**, 2220 (1989).