Calculations of critical misfit and thickness: An overview

Jan H. van der Merwe and W.A. Jesser

Department of Physics, University of Pretoria, Pretoria 0002, South Africa.

This overview stresses the equilibrium/non-equilibrium nature of the physical principles, as well as the basic properties of the models, used to calculate critical misfit and critical thickness in epitaxy.

Hierdie oorsig beklemtoon die ewewigs/nie-ewewigs karakter van die fisiese beginsels, sowel as die basiese eienskappe van die modelle wat gebruik word om kritieke wanpas en dikte in epitaksie te bereken.

Introduction

The importance of the transition of an epitaxial interface between crystals A and B from coherency to incoherency with misfit dislocations (MD's) needs no motivation. The transition depends primarily on the *misfit* f at the interface. The general specification of misfit has been discussed by Jesser[1] and Braun[2]. In the case of rectangular interfacial symmetry we may write

$$f = (a - b)/c \qquad (1)$$

where $c = \frac{1}{2}(a+b)$ when A and B are both thick and c = a or b when A is thin.[3-5] For crystals with different symmetries, e.g. (111) f.c.c./(110) f.c.c. interfaces, the misfits in two orthogonal directions may be significantly different.[6,7]

The following modes of *misfit accommodation* (MA) will be considered: (a) a misfit vernier (MV) when A-B is rigidlike[7], (b) misfit strain (MS)[4]; crystals homogeneously strained to reduce disregistry, (c) misfit dislocations (MD's)[4,5] (d) and a misfit strain gradient (MSG)[8] from

lattice parameter a of A to b of B. MD's may be efficient (EMD), inefficient (IMD), of screw character (SMD) or split up into partials. Often two or more modes of MA are present.

If *MD's ($\bar{f}$) and MS ($\bar{e}$) coexist*, the portion $\bar{f}$ of f accommodated by MD's satisfies (to first order) for quadratic symmetry the relations[5,9,10]

$$f = P\bar{f} + Q\bar{e}, \quad \bar{e} \equiv |\bar{e}| , \tag{2}$$

where (i) $P = Q = 1$ for a thin layer on a thick substrate (ii) $P = \cos\lambda$, $Q = 1$ when the MD is inefficient; λ being the inclination of the Burgers vector to the interface and (iii) $P = 1$, $Q = 1 + r^{-1}R^{-1}$ when A and B are both thin with $r = h_a/h_b$, $R \approx \mu_a/\mu_b$ and $\bar{e} = |\bar{e}_b|$. For different symmetries MS, MD's and a MV may coexist.[7] Our *main aim* is to report on calculations of critical misfit f_c when a monolayer (ML) is subcritical ($\bar{f} = 0$) and the critical thickness h_c at which such a system becomes critical ($\bar{f} \neq 0$).

We need to define the *physical principles* governing calculations of f_c and h_c, clearly. Consider a growing epilayer which started as a coherent ($\bar{f} = 0$) ML. The acquisition of MD's is subjected to a hierarchy of energy barriers U both in creation and motion (Peierls). The acquisition is driven by the free energy gradients (ΔA), aided by thermal energy (kT_s, $T_s \equiv$ substrate T) and proceeds towards the free energy (A) minimum defining the equilibrium, i.e.

$$\text{min.} = A \approx E, \tag{3}$$

where E is the energy. The observed configuration depends on ΔA, kT_s, U, the deposition rate R_{dep} and any waiting time. Furthermore, the critical parameters will also depend on the *growth mode:* island growth, ML-by-ML growth and Stranski-Krastanov (islands on top of ML-by-ML).[11-13]

Almost all theories *model* the AA and BB interaction in terms of continuum elasticity (harmonic approximation).[3,4,5] Recently ab initio

calculations, using appropriate potentials, have been carried out too.[14] The periodic AB field (potential V(x,y)) has been cast either in the Volterra model[5], which invokes crystal dislocations to accommodate f, or the Frenkel-Kontorowa (FK)[2-4,7] model that expresses V(x,y) in terms of a (truncated) Fourier series. Whereas the Volterra model has the advantage of the mathematical simplicity the FK model allows for adjusting the AB strength as needed. Furthermore, specific epitaxial orientations are linked to specific Fourier terms[7] as matching of A and B wave vectors[2] ("atomic row matching"), whereas the magnitude of the relevant coefficient poses as an energetic measure of the tendency to the realization of the specific epitaxy.

In the FK model the self energy E_D' of a MD between two thick epicrystals is given by[3]

$$E_D' = \frac{\mu_{ab} cp}{4\pi^2}\left[1 + \beta - (1+\beta^2)^{\frac{1}{2}} - \beta \ell n\{2\beta(1+\beta^2)^{\frac{1}{2}} - 2\beta^2\}\right] \tag{4a}$$

$$\beta = \frac{2\pi\lambda_0 c}{\mu_{ab} p}, \quad \frac{1}{\lambda_0} = \frac{1-\nu_a}{\mu_a} + \frac{1-\nu_b}{\mu_b}, \quad c \approx \frac{1}{2}(a+b) \tag{4b}$$

where p is the MD spacing and the other symbols have evident meanings. The Volterra model introduces a "cut off radius R" such that for an epilayer of thickness h_a on a thick substrate[5] the line energy of an EMD becomes

$$E_D' = \frac{Db}{2}\{\ell n(\frac{R}{b}) + 1\}; \quad D = \frac{\lambda_0 b}{\pi}, \quad R = \begin{cases} p/2 & \text{when } h_a > p/2 \\ h_a & \text{when } h_a \leq p/2. \end{cases} \tag{5}$$

This expression must be modified when the MD's are inefficient or partial and the Burgers vector b is essentially the c in Eqs. (4). When p becomes large (4) reduces to (5) but with smaller value of R. When the misfit is partly accommodated by MS $\bar{e}$ the MS energy, which is of the form [4,5]

$$E_{\bar{e}} = Bh_a\bar{e}_a^2 = \frac{2\mu_a(1+\nu_a)}{1-\nu_a} h_a\bar{e}_a^2 \quad \text{per unit area,} \tag{6}$$

for interfaces with quadratic symmetry, must be included.

We need *criteria defining* f_c *and* h_c. The criterion for the equilibrium MS $\bar{e}_m$ (or MD density $\bar{f}_m$) is defined by[4,5]

$$\frac{\partial E}{\partial \bar{e}} = 0 = \frac{\partial E}{\partial \bar{f}} \text{ for } \bar{e} = \bar{e}_m, \quad \bar{f} = \bar{f}_m \tag{7a}$$

using eqs. (2)-(3). Furthermore, from $\partial E/\partial \bar{f} = 0$ we may infer that the work done to introduce a MD vanishes, i.e. that[4,5,7]

$$W = \int F ds = 0, \quad \text{or} \quad F = 0 \text{ when F constant;} \quad \bar{e} = \bar{e}_m, \quad \bar{f} = \bar{f}_m, \tag{7b}$$

where, either F is the net force on a dislocation and ds its displacement, or F is an applied force employed to introduce the dislocation and ds the displacement of its "point" of action. The case F =0 applies when a MD is generated from an existing (threading) dislocation[5]. The conditions for critical misfit f_c ($\bar{f}$ =0) at given h and critical thickness h_c at given f may be written as[5,9]

$$\phi(h, f_c; \bar{f} = 0) = 0 \quad \text{and} \quad \phi(h_c, f; \bar{f} = 0) = 0, \tag{8}$$

where ϕ is one of the functions in eqs. (7). In the FK model the transition to incoherency is sharp[4]. This gives credibility to

$$E_{\bar{e}}(h, f_c; \bar{e} = f) = E_D(h, f_c; \bar{f} = f) \tag{9}$$

as an approximate criterion for f_c. This may however overestimate[9,10] f_c and h_c by nearly 25%.

Equilibrium considerations assume that MD's are freely available. In general energy barriers of the form[4]

$$U = \int_{bar.} F ds; \qquad \text{bar} \equiv \text{barrier} \tag{10}$$

exist, where the integration is from bottom-to-top of barrier. Typical

barriers are the nucleation and Peierls barriers.[5,14] The barrier heights are reduced by MS (Peach-Kohler) driving forces $F_{\bar{e}}$. A special case obtains when the opposing MD line tension F_D is overcome by $F_{\bar{e}}$; i.e.[5]

$$F = F_D - F_{\bar{e}} = 0, \quad \text{for } \bar{e} = f \tag{11}$$

and MD's generate spontaneously without the aid of thermal energy kT_s. When $U/kT_s \gg 1$ the barriers are completely prohibitive and the system will remain almost indefinitely in a metastable configuration. When $U \lesssim kT_s$ the attainment of equilibrium takes time and the observed configuration may differ significantly from equilibrium.

Equilibrium calculations of critical misfit and thickness

Consider the application of the equilibrium criteria (7) and (8) to layerlike growth. Frank and van der Merwe[4] did this for a one-dimensional (1-D) FK model in which the AA (harmonic) and AB (sinusoidal) interaction strengths were respectively embedded in a force of constant μ and an overall amplitude W. They obtained the (ML) critical misfit $f_c^{(1)}$ and critical thickness h_c as (See eqs. (8))

$$f_c^{(1)} = 2/\pi\ell_0 , \quad \ell_0^2 = nb\mu^{(1)}b/2W ; \qquad n=1 \text{ for a ML}, \tag{12}$$

$$h_c = n_c b = 8W/\pi^2\mu^{(1)}bf^2, \tag{13}$$

where of $\mu^{(1)}$ and W in (12) and (13) stresses the importance of field strengths; $f_c^{(1)} \sim 9\%$ for an average case.

This calculation neglects strain in the substrate, the normal strain gradient in the overlayer and the lateral Poisson (ν) contraction. The latter is easily taken care of yielding for quadratic symmetry the results[15]

$$f_c^{(1)} = \frac{2}{\pi\ell(1+\nu)^{\frac{1}{2}}} , \quad \ell^2 = \frac{\mu_a hb^2}{(1-\nu)W} , \quad h_c = \frac{4(1-\nu)W}{\pi^2(1+\nu)\mu_a b^2 f^2} , \tag{14}$$

where μ_a is now the shear modulus of A. This diminishes $f_c^{(1)}$ from 9% to about 7%. With different symmetries, as for (111) f.c.c./(110) b.c.c. interfaces[2,7,12,13] the misfits in two ortogonal directions are different and a variety of MA modes coexists. Also for crystals, other than ML's, 2-D coherency is extremely rare, whereas 1-D coherency follows the same pattern as for quadratic symmetry.

The first calculation to allow properly for relaxation in the substrate and a strain gradient normal to the interface was done using the parabolic (P) model and yielded the exact solution[9]

$$f_c(h) = \frac{1-\nu}{4\pi(1+\nu)} \int_0^\infty \frac{dx}{x[x+\alpha\phi(x,\alpha)]}, \qquad \alpha = \frac{2(1-\nu)h}{a}$$

$$\phi(x,\alpha) = \frac{x(s^2+cs-x-x^2)+\alpha(c+s)^2/(1-2\nu)}{x(s^2-x^2)+\alpha(s^2+cs+x-x^2)/(1-2\nu)}, \qquad \begin{cases} s \equiv \sinh x \\ c \equiv \cosh x \end{cases} \tag{15}$$

for $f_c(h)$ or equivalently $f(h_c)$ in crystals with equal elastic constants. With different elastic constants $f_c(h)$ has the same form but ϕ is vastly more complicated.[10] Matthews[5] obtained using the Volterra model (eqs. (5) and (6))

$$f_c = \frac{D}{2Bh}\left[\ln\left(\frac{h}{b}\right) + 1\right], \qquad h_c = \frac{D}{2Bf}\left[\ln\left(\frac{h_c}{b}\right) + 1\right]. \tag{16}$$

Equation (16) is inaccurate at small thicknesses where the "cut off" approximation is poor and the precise AB field strength, which (16) ignores, becomes important. It may be modified though to account for inefficient and partial MD's.

Most systems exhibit *island growth* (Volmer-Weber or Stranski-Krastanov).[11] This case was first addressed by Cabrera[16] assuming a hemispherical epitaxial island A of radius R on a thick substrate B and equilibrium MS's $\bar{e}_a^m$ and $\bar{e}_b^m$. For the critical misfit $\bar{e}_a^m + \bar{e}_b^m = f_c$,

while also $p = 2R+$. When $\mu_a = \mu_b = \mu$ and $a_x = a_y = a$ it follows, (to first order) using the strain energy (log. term) in Eq. (4a) that

$$f_c(R) = \frac{15(1-\nu)}{16\pi}\beta\ell n[2\beta(1+\beta^2)^{\frac{1}{2}} - 2\beta^2], \quad \beta(R) = \frac{\pi b}{2(1-\nu)R}. \tag{17}$$

For given misfit f, Eq. (17) defines a critical size R_c.

Recently $f_c(h)$ has been calculated for a *superlattice*[9,10] using the exactly solvable parabolic model in conjunction with the assumption that the stack of layers remains plane:

$$f_c(h;r,R) = \frac{(1-\nu_b)(1+r^{-1}R^{-1})\mu}{4\pi(1+\nu)\mu_b}\int_0^\infty \frac{dx}{x[x+\alpha(\eta)\phi(x)]}, \tag{18}$$

where $\alpha(\eta) = 2(1-\nu)h/a$ and $\phi(x)$ is a complicated function of elastic constants and thicknesses, and r and R are defined in eqs. (2). By criterion (8) this also defines a critical thickness $h = h_c$. When the layer thicknesses and elastic constants are respectively equal eq. (18) simplifies greatly. h_c/a varies approximately as f^{-K}; $K = 1.22$ compared to $K = 2$ in (14) and $K = 1$ in (16). When h/a is small $E_D^{sup} \sim E_D/2$ where E_D relates to a single layer on a thick substrate. For thick layers, $E_D^{sup} \sim E_D$. Furthermore $f_c^{sup} \approx 4\ f_c$ and $h_c^{sup} \approx 4h_c$.

Matthews[5] has calculated f_c and h_c for a *thin epilayer on a thick substrate* using the (Volterra) force criterion in (7b) and included the line tension σ_0 of a step formed concurrently at the free surface and the force γ_0 needed for stacking fault formation:

$$h_c = \frac{Db(1-\nu\cos^2\alpha)[\ell n(h_c/b)+1] + \sigma_0 b\sin\alpha}{Bbf\cos\lambda - \gamma_0/\cos\phi}, \tag{19}$$

λ being the angle between the slip direction and the direction in the interface normal to the line of intersection of the slip plane (S) and the interface, α is the angle between the MD and its Burgers vector b

(applicable to partials) and ϕ is the angle between the free surface and the normal to S.

MD's may also be generated by *climb*. Matthews[5] calculated h_c for this case invoking partial MD's and obtained, using the criterion (7b):

$$h_c = \frac{D[\ln(h/b) + 1] \pm \sigma_0}{Bf - \gamma_0/b \cos\phi} \quad , \tag{20}$$

where the $\pm$ signs correspond to surfaces being created or annihilated by climb and σ_0 is the surface free energy. This mechanism is dependent on the availability of point defects.

Calculations of critical parameters (non-equilibrium)

We briefly consider quantities relevant to criticality under non-equilibrium conditions: the activation barriers U_n (eq. (12)) to the formation of MD's and their reduction by MS, the nucleation energy of MD loops, surface (image) and Peierls barriers, and the dynamics of MS relief. Frank and van der Merwe[4] have first shown that in the FK model the barrier to the injection of a MD at the free end of a coherent *monolayer* (ML) is

$$U_n^{(1)} = (4W\ell_0/\pi)[(1-f^2\ell_0^2)^{\frac{1}{2}} - (\pi/2 - \arcsin f\ell_0) f\ell_0] \tag{21}$$

per atom row, (compare eqs. (10). $U_n^{(1)}$ has a maximum value of $4W\ell_0/\pi \sim 9W$ at f=0, diminishes to about 2W at the coherency instability $f_c^{(1)} = 2/\pi\ell_0$ in eq. (12) and vanishes at $f_s^{(1)}$ for a ML or h_s for a multilayer (analogy of eq. (13)):

$$f_s^{(1)} = 1/\ell_0 = \pi f_c^{(1)}/2; \quad h_s = \pi^2 h_c/4 \approx 2h_c. \tag{23}$$

whereafter MD's enter freely without thermal aid. In metals $W \sim 0.2 - 0.5$ eV. The misfit $f_s^{(1)}$ is about 14% which is reduced to about 11% if we use the more appropriate ML relations in eqs. (14).

Continuum theory predicts[17] that the *nucleation energy of a MD loop* in the absence of MS is about 200 eV which completely rules out the creation of such loops by thermal energy alone. The reduction of U_n was also been considered by Dodson[14] in the case of Si. He employed an atomistic approach, using modified Tersoff potentials and obtained an $f_s \approx 11\%$. This is significantly larger than the continuum prediction of 4% and equals exactly the 11% in (23), which is somewhat fortuitous.

More recently Tsao et. al.[17] have considered the overcoming of *Peierls barriers* and measured the MS relief in metastable Si_xGe_{1-x} strained layers grown on (001) Ge substrates. They concluded that strained layer breakdown is most directly correlated with (1) an excess (driving) stress $2(F_{\bar{e}} - F_D)/hb$ given by

$$\sigma_{exc} = \frac{2\bar{e}\mu(1+\nu)}{1-\nu} - \frac{\mu(1-\nu\cos^2\alpha)}{2\pi(1-\nu)} \quad \frac{\ell n(4h/b)}{h/b} \; ; \quad \bar{e} = f \tag{24}$$

and (2) temperature, where $\alpha = 60^\circ$ is the angle between the MD line and its Burgers vector. Whereas $\sigma_{exc} = 0$ is still valid for equilibrium, in the relevant metastable configuration, the energy barriers delay the transition to equilibrium. At a temperature of 494°C MS relief becomes observable when $\sigma_{exc}/\mu \approx 0.024$. Since in the derivation of σ_{exc} the MD's are generated from threading dislocations the relevant barriers are presumably the Peierls' barriers. The excess stress defines an excess force $F_{exc} = \frac{1}{2} hb\sigma_{exc}$ acting on the dislocation threading the overlayer,[17] and when displaced by x reduces the barrier U(x) to $\Delta U(x) = U(x) - xF_{exc}$, or to $\delta U(x) \approx \Delta(x)L/h$ for a critical segment length L jumping the barrier. When $|\delta U|_{max}$ is small enough compared to the thermal energy kT_s the jump frequency increases to the extent that MS relief becomes observable.

An interesting case of critical thickness concerns epilayers on substrates of finite width. It makes use of a significant *misfit strain gradient* (MSG) to accommodate the misfit f. Luryi and Suhir[8] adopted a *rigid* substrate of width 2ℓ along the 1-D misfit direction, a coherent epilayer of the same width and an exponentially decaying strain field with distance normal to the interface. They used a quasi-nucleation criterion and showed that when ℓ is small in comparison to the thickness h, there is an effective thickness h_e smaller than h that characterizes the limit of the strain field, i.e. that $h_e \sim \ell$ when $h \gg \ell$ and $h_e \approx h$ when $\ell \gg h$. These considerations are qualitatively very significant because they point towards a new mechanism of avoiding MD's.

Recently Dodson and Tsao[18] investigated dislocation dynamics based on a phenomenological model and assuming that (i) initially $\bar{f} = 0$, (ii) the biaxial MS is relaxed by a MD network, (iii) the rate process mechanism is the equisition of MD's, (iv) the motion of dislocations is driven by a localized stress $\tau_{\ell oc}$ and opposed by energy barriers and (v) $\tau_{\ell oc}$ can be expressed in terms of the residual mismatch $f - \bar{f}(t) - \bar{e}(h)$, $\bar{f}(t)$ being the (insufficient) density of MD's and $\bar{e}(h)$ the equilibrium MS at thickness h. Starting with a nearly dislocation free coherent layer and introducing a background dislocation density $\bar{f}_o$, the authors obtained the governing equation

$$d\bar{f}(t)/dt = C\mu^2[f_0 - \bar{f}(t) - \bar{e}(h)]^2(\bar{f}(t) + \bar{f}_0), \qquad (25)$$

where C is a constant depending on temperature. This equation is claimed to give a good description of the data for SiGe alloys on (100) Si substrates using as model parameters $C = 30.1$ and $\bar{f}_0 = 10^{-4}$. These considerations constitute significant progress towards a successful dynamical description of strain relief in non-equilibrium cases.

Concluding remarks

In the past, calculations of critical misfit and thickness had mainly been carried out using equilibrium principles because of their simplicity. Recent observations have displayed a significant discrepancy between theoretical and experimental values, particularly in semiconductors. This phenomenon which has been mainly ascribed to non-equilibrium effects, has become topical recently. This paper attempts to put the models and physical principles of the predictions into perspective. Clearly the space allowed is completely inadequate to do justice to this endeavour. A more complete overview is in preparation.

References

1. Jesser W.A. (1973). On the extension of Frank's formula to crystals with different lattice parameters. Phys. Status Solidi (a) 20, 63.
2. Braun M.W.H. (1987). Epitaxy on substrates with hexagonal lattice symmetry. Ph.D. Thesis. University of Pretoria, Pretoria.
3. van der Merwe J.H. (1950). On the stresses and energies associated with intercrystalline boundaries. Proc. Phys. Soc., London A63, 616.
4. Frank F.C. and van der Merwe J.H. (1949). One-dimensional dislocations. Proc. Roy. Soc. 198, 205.
5. Matthews J.W. (1975). In Epitaxial Growth, ed. J.W. Matthews, Chapt. 8 Academic Press, New York; (1979). In Dislocations in Solids, ed. F.R.N. Nabarro, chapt. 7. North Holland, Amsterdam.
6. Bruce L.A. and Jaeger H. (1978). The allignment of (111) f.c.c. - (110) b.c.c. epitaxed metal pairs. Philos. Mag. A38, 223.
7. van der Merwe J.H. (1982). Misfit accommodation in epitaxial monolayers on (111) f.c.c and (110) b.c.c. substrates. Philos. Mag. A45, 145.
8. Luryi S. and Suhir E. (1986). New approach to high quality epitaxial growth of lattice matched materials. Appl. Phys. Lett. 49, 140.
9. van der Merwe J.H. and Jesser W.A. (1987). An exactly solvable model for calculating critical misfit and thickness in epitaxial superlattices. Part I J. Appl. Phys. (in print).
10. Jesser W.A. and van der Merwe J.H. (1987). An exactly solvable model for calculating critical misfit and thickness in epitaxial superlattices. Part II J. Appl. Phys. (in print)
11. Bauer E. (1985) Phänomenologische Theorie der Kristallabscheidung an Oberflächen. Z. Kristallogr. 110, 423.
12. Bauer E. and van der Merwe J.H. (1986). Structure and growth of crystalline superlattices: From monolayer to superlattice. Phys. Rev. B 33, 3657.

13. van der Merwe J.H. and Braun M.W.H. (1987). In MRS Symposia Proceedings, eds. J.D. Dow and I.K. Schuller, Vol. 77, MRS, Pittsburgh.

14. Dodson B.W. (1987). Atomistic study in structural metastability in coherently strained Si-like layers. Phys. Rev. B35, 2795, 5558.

15. van der Merwe J.H. (1970). Equilibrium structure of a thin epitaxial film. J. Appl. Phys. 41, 425.

16. Cabrera N (1964). The equilibrium of crystal surfaces. Surf. Sci. 2, 320.

17. Tsao J.Y., Dodson B.W., Picraux S.T. and Cornelison D.M. (1987). Critical stresses for Si_xGe_{1-x} strained-layer plasticity. Phys. Rev. Lett. (submitted).

18. Dodson, B.W. and Tsao J.Y. (1987). Relaxation of strained-layer semiconductor structures. Appl. Phys. Lett. (submitted).

Descriptions of Low Energy Misfit Dislocation Structures using the Parabolic Interaction Potential

JAN H. VAN DER MERWE
Department of Physics, University of Pretoria,
Pretoria 0002, (South Africa)

W.A. JESSER
Department of Materials Science, University of Virginia,
Charlottesville, VA 22901 (U.S.A.)

ABSTRACT

The need for and development of the parabolic model as a power series approximation of the sinusoidal Frenkel–Kòntorowa model, used to study low energy misfit dislocation (MD) structures in epicrystals, is briefly discussed: Its application to monolayers on thick substrates, thickening epilayers, superlattices and structural ledges is outlined with special emphasis on critical misfit and thickness in coherent to incoherent transitions. The "nucleation" of a MD by climb from the free surface is calculated. The critical misfits obtained from (a) the continuum (approximation) and exact solutions, using the parabolic model and (b) the continuum solutions of the sinusoidal and parabolic models, are compared and the differences of the various approaches assessed.

1. INTRODUCTION

Intercrystalline boundaries play an important role in the physical behavior of crystalline systems, e.g. in plasticity and the electrical and optical properties of epitaxial films. In the present considerations we are primarily interested in the near interface atomic displacements resulting from the competing intralayer and interlayer atomic interactions in an epitaxial bicrystal comprizing two different crystals [1,2,3]. When the interlayer forces are moderately strong the interfacial atoms are in fair registry everywhere except near line defects constituting a sequence of misfit dislocations (MDs) which are said to accommodate the misfit. One or both crystals may also be homogeneously (misfit) strained to eliminate the disregistry completely and accordingly accommodate the misfit by socalled misfit strain (MS). Because of the mathematical complexities involved in a proper quantum mechanical description of the competing forces, either ab initio models [4,5], using appropriate interatomic potentials, or phenomenological descriptions [1,2,3], have been employed in predicting interface and near interface atomic configurations.

Of interest here are the phenomenological models. Except for a few anelastic [5] refinements investigators almost invariably assumed that the relative intralayer atomic displacements are small enough for the harmonic (elastic) approximation to be valid. For simplicity mostly isotropic elasticity is usually adopted [1,2,3,7].

In modelling the interlayer atomic interaction provision is made for the discrete atomic nature of the crystals by expressing the relevant forces as periodic functions of the relative displacement (parallel to the interface) of atoms

opposing each other across the interface. There are two somewhat different approaches: (a) the Volterra model [2] which assumes the atomic relative displacement to be zero on one side of the dislocation and to be one atomic spacing b on the other side and (b) the Frenkel Kontorowa (FK) model [8] (and its generalizations and approximations) that expresses the periodic interaction potential V in terms of a truncated Fourier series V(x). The truncation is justified by the fact that the Fourier coefficients decay rapidly with harmonic order [9].

In the original one–dimensional (1D) model Frenkel and Kontorowa expressed the energy of their system as [8]

$$E = \sum_n \tfrac{1}{2}\lambda(x_{n+1} - x_n - a)^2 + \sum_n V(x_n) \tag{1}$$

with

$$V(x_n) = \tfrac{1}{2}V_0[1-\cos(2\pi x_n/b)], \tag{2a}$$

$$x_n = b(n+\xi_n) = nb + U_n. \tag{2b}$$

Equations (1) and (2) model one halfcrystal, simulating a monolayer (ML), in terms of a linear chain of particles connected by elastic springs (harmonic forces) of force constant λ and natural length a and the other (a thick, effectively "rigid" substrate of atomic spacing b) to be the source of the periodic potential V [1]. Its overall amplitude V_0 is seen as a measuere of the interlayer interaction. The integer n enumerates particles and potential troughs starting from a position of (nearest) registry so that $b\xi_n = U_n$ represents the displacement of particle n from

trough n.

The parabolic reproduction of the FK model replaces the sinusoidal relation (1) by a sequence of parabolic arcs, for example [10,11]

$$V = \pi^2 V_0 \xi^2 = \pi^2 V_0 U^2/b^2, \quad b|\xi| = |U| \leq b/2, \tag{3}$$

which are obtained by truncating the Taylor series of (2) at quadratic terms. Hence the arcs in (2) osculate with the troughs of (1). If we assume that relation (2) is a true representation of the interlayer interaction then (3) is an approximation which is still acceptable [10,12] at $U = \pm b/4$ and poor towards the boundaries $U = \pm b/2$, as may also be inferred from Fig. 1. In semiconductors in which the crest of the potential V is believed to be fairly sharp as compared to the troughs the approximation (2) is rather inadequate and it has been suggested that the parabolic form (3) may be nearer to the truth [13].

The simple representation (2) has subsequently been extended to two–dimensional (2D) potentials V(x,y) appropriate to 2D interfaces characteristic of monolayers (MLs) [1,14], for example, and to linear interfaces between thick epicrystal halves [7]. In the latter case it had been convenient to represent the interaction in terms of an interfacial shear stress $p_{zx}(x)$ (shear force per unit area) which, in the parabolic approximation [15], takes the form

$$p_{zx}(x) = \mu U(x)/d, \ |U| \leq b/2 \tag{4}$$

where d is the separation of the atomic planes on either side of the interface, x and z are respectively parallel and normal to the interface and μ an interfacial

shear modulus of magnitude characteristic of the interlayer interaction strength. One may use either Eq. (2) or (3) to derive the equivalent of (4) and obtain by comparison the relation

$$\mu/d = 2\pi^2 V_0/b^2 s, \tag{5}$$

where s is the interfacial surface area per atom. In any interface with quadratic symmetry $s = b^2$.

When the crystal halves are different in thickness and/or elastic constants the relative normal displacement W of the opposing atoms at the interface varies along the interface. This induces a normal interfacial force which has been modelled [13,15,16], using Hookian relations, as

$$p_{zz}(x) = [2\mu/(1-2\nu)]W(x)/d, \tag{6}$$

where ν is a Poisson's ratio for the interface. In terms of an interaction *potential*, (6) is also parabolic. Whereas in case (4) a nonlinear relation exists at $|U| = \pm b/2$ when a MD is present, W in Eq. (6) will always be small enough for the linear Hookian relation to apply.

Having described the background of the parabolic model it is now our objective to briefly review the use of the parabolic model in describing low energy misfit dislocation structures (LEMDSs). Of specific interest is the transition from a coherent (C) interface, where the misfit f is fully accommodated by MS $\bar{e} = f$, to an incoherent (IC) interface with MD density $\bar{f}$. When MDs and MS coexist [2,13]

$$f \equiv (a-b)/b \approx \bar{f} + |\bar{e}| \ . \qquad (7)$$

The C–IC transition is mathematically characterized by a change from $\bar{f} = 0$ to $\bar{f} \neq 0$.

The behavior of our system is governed by the law that all natural processes carry thermodynamic systems towards equilibrium. The equilibrium configuration is determined by minimization of the free energy which, in the case of MDs; may be approximated by minimization of mechanical energy E [1–3]. The solutions thus obtained for the governing equations within the crystals, which are normally in the form of partial differential equations, must satisfy certain boundary conditions, for example, that the "upper" surface of an epitaxial film on a thick substrate is a *free* surface. This implies that the forces that act on the surface, and [1,3] accordingly the surface values of the corresponding calculated stresses, vanish.

Apart from the suggestion that the parabolic model could be a better representation of the interlayer interaction potential in semiconductors [13], for example, than a low order Fourier truncation, it enters piecewise *linearly* in the governing equations making them exactly solvable. Although the need for exact solutions was most pressing in the case of thickening epilayers, the model has been useful and even indenspensible in other cases.

2. APPLICATIONS

2.1. *Monolayers: One–Dimensional Model*

Originally the 1D FK model was introduced to simulate the plastic behavior of crystals [8]. It was subsequently applied to describe the accommodation of misfit in epicrystals, in which misfit exists in one interfacial direction only [1,7,17]. In certain interfaces the misfit may be so large in one direction e.g. {111} fcc/{110} bcc interfaces [18], that the relevant misfit accommodation approaches a misfit vernier (MV) with little effect on the atomic arrangements in either crystal. The small misfit direction accordingly poses a 1D MD problem. Minimization of (1), using (2b) and (3), yields the piecewise linear governing equations [10,11]

$$\xi_{n+1} - 2\xi_n + \xi_{n-1} = \frac{2\pi^2 V_0}{\lambda b^2}[\xi_n]; \; |[\xi_n]| \leq \tfrac{1}{2} . \tag{8}$$

where $[\xi_n]$ means that ξ_n must be replaced by $\xi_n \bullet 1$ if atom n enters the parabolic domain of neighboring troughs $n \pm 1$. The disadvantage of the parabolic approximation, as may be seen from (8), is that the right hand side has discontinuities at $\xi_n = \pm\frac{1}{2}$, which is a mathematical complication.

Possibly the parabolic model was used for the first time [19] to support results obtained by Frank and van der Merwe [1] using the 1D FK potential in Eq. (2). In that case the governing equations were a series of second order nonlinear difference equations which simplify in the continuum *approximation* to solvable Sine Gordon equations. Frank and van der Merwe obtained critical misfits [1]

$$f_c = 2/\pi\ell_0, \quad f_s = 1/\ell_0, \tag{9a}$$

$$\ell_0^2 = \lambda b^2/2V_0. \tag{9b}$$

Below a misfit $f = f_c$ the equilibrium configuration of the chain is the one in which it is misfit strained to coherency (a→b), but still needs an activation energy for the introduction of MDs. Above $f = f_s$ the energy barrier vanishes and MDs enter spontaneously. The question as to the extent to which the approximation effects the predictions was investigated using a modified parabolic potential [19]. The modification was an improvement on the form (3) in that the lower part of the potential was represented by the trough segment of (3) and towards the crests at $x = \pm b/2$ by segments of inverted parabolas. This eliminated the discontinuities. The modification of the governing equation (8) contains linear pieces of positive and negative slope on the right hand side and connect smoothly. The investigation using the modified parabolic potential showed that the effect on the predicted critical values was not serious.

Of interest is that, instead of the critical misfits f_c and f_s of Eqs. (9), the authors obtained for the parabolic model the results

$$f_c' = \frac{\pi^2}{8(1+\pi^2/4\ell_0^2)^{\frac{1}{2}}}\frac{2}{\pi\ell_0}, \quad f_c'' = \frac{\pi^2}{8}\frac{2}{\pi\ell_0}, \tag{10a}$$

$$f_s' = \frac{\pi}{2}\{(1+\frac{\pi^2}{4\ell_0^2})^{\frac{1}{2}} + \frac{\pi}{2\ell_0}\}\frac{1}{\ell_0}, \quad f_s'' = \frac{\pi}{2}\frac{1}{\ell_0}. \tag{10b}$$

when expressed in terms of ℓ_0. The values f' and f'' correspond to exact and continuum (approximate) solutions, respectively. It is seen that $f_s' > f_s'' = \ell_0^{-1}$ whereas $f_c' < f_c''$. Both f_c' and f_c'' are greater than $f_c' = 2/\pi\ell_0$; f_c only when $\ell_0 > 2$.

Frank and van der Merwe [1] showed that $\ell_0 \approx 7$ in average cases in which intra– and interlayer forces are the same. It may thus be surmised that the regime $\ell_0 \geq 2$ covers most common cases. Also if we take conveniently $\ell_0 = 2\pi$ it follows that f_c' differs from f_c'' by only about 3% whereas the difference between f_s' and f_s'' is about 25%.

2.2. *Adsorbed Monolayers: 2D Models*

According to the Frank–van der Merwe (FVDM) theory an adsorbed (2D) ML grows coherently (incoherently) on a crystalline substrate under quasi–equilibrium conditions when the misfit falls below (exceeds) a critical value which is somewhat below the value f_c in Eq. (9a) because of Poisson phenomenon [1,14]. The coherent to incoherent (C–IC) transition is accomplished by the introduction of MDs which may be of finite length and terminate within the ML (see Fig. 2) [20]. The MD contains one more (less) ML atomic row as compared to substrate potential troughs when $a < b$ ($a > b$). When the MD involves an extra atomic row the MD grow (shrink) by the capture (ejection) of atoms from the surface (row). Shrinkage by the absorption of vacancies from the ML may also be invisaged. Phase transitions in adsorbed MLs had become a subject of intensive theoretical and experimental study. Not only had these studies generalized the theory to account for temperature effects but they had also introduced new nomenclature: solitons, discommensurations, etc. for MDs and (Taylor) dislocations [20,21] for the atomic configuration within the ML where the MD terminates. This configuration differs from that of a

conventional dislocation in that the atomic positions within the ML not only depends on the atomic interaction within the ML but also on the modulating potential V(x,y) emanating from the substrate. Although this situation can be modelled in terms of Fourier truncations the resulting governing equations could not be solved. However when parabolic models are used instead for both the soliton and the dislocation they become solvable [20] for the atomic displacements in terms of Fourier transforms. Expectedly the analysis is moderately complicated in this case. Different sets of difference equations, with a variety of boundary conditions apply to different areas depicted in Fig. 2. Simple solutions had only been obtained in the continuum approximation in which difference equations reduce to differential equations. Of particular interest is the result that the interaction between the two (Taylor) dislocations terminating the soliton (MD) falls off exponentially with separation, whereas it had previously been assumed in theories of two–dimensional melting transitions in adlayers that dislocation dipole unbinding follows a logarithmic dependence [21].

2.3. *Thickening Epilayers*

One of the important problems that "evaded" an exact solution is the one of LEMDSs at the interface between a thickening epilayer and its substrate. By "exact" we mean a solution that satisfies the governing equation even though the equation itself already involves a degree of approximation, as for example the continuum approximation which reduces a set of difference equations to a single

partial differential equation in the 1D FK model [1]. Exact solutions using sinusoidal FK models exist for the extreme cases, namely a ML on a thick (essentially rigid) substrate [1,14] and an epitaxial bicrystal consisting of two "infinitely" thick crystals [7]. When both crystals are thicker than half the MD spacing the "infinite" solution is accurate to within 2% [15]. An extrapolation of the ML solution to multilayers, already involves a nonneglegible error at two and three fold layers. Attempts at interpolations between the ML case and the thick layer case had not been very reliable; the errors have not been calculated.

Matthews has adapted the Volterra model to study LEMDSs in growing epilayers [2]. For a growing epilayer A on a thick substrate B the energy per unit length of a MD is written as

$$E_D = Db(\ell n\frac{R}{b} + 1)\ ;\ D = \frac{\mu_A \mu_B b}{[(1-\nu_A)\mu_B + (1-\nu_B)\mu_A]},$$
$$R = \begin{cases} \frac{1}{2}p \text{ when } h_A > \frac{1}{2}p \\ h_A \text{ when } h_A < \frac{1}{2}p. \end{cases} \qquad (11)$$

R is a "cut off" radius accounting for cancelling of overlapping of strain fields, p and h are respectively the MD spacing and layer thickness and μ and ν respectively the shear modulus and Poisson's ratio. The constant term Db makes provision for the core energy of the dislocation. By a proper choice of R Matthews could handle epilayers of all thicknesses. However, the Volterra model does not provide independently for the interfacial bond strength between A and B [3]. This is evidently important in epilayers of small multiplicity [13] as one may deduce from the fact that V_0 appears in the relation (9) for f_c and that V_0 is not necessarily determined by μ_A and μ_B. In thick epilayers the interfacial bond

strength vanishes asymptotically from the results [3].

Exact solutions of the problem of a growing epilayer on a thick substrate of a different crystal had previously been studied by various authors and more recently in greater detail by van der Merwe and Jesser [13,16], using the parabolic model in the form of Eq. (4). Since the crystals have different thicknesses, provision is also made for a normal interface stress in the form of Eq. (6). With the conventional assumption that the MDs are long and straight the problem is one of plane strain that is most efficiently analysed in terms of an Airy stress function Ψ satisfying a biharmonic equation that can be solved in terms of Fourier transforms [7,13]. The stresses are obtained from Ψ as simple derivatives. The Fourier transforms are selected to satisfy appropriate periodicity and symmetry conditions for the film, a free top surface and surface stresses (4) and (6) at the interface. The strains accordingly follow from Hooke's law and the displacements U and W by integration. The periodicity, symmetry and boundary conditions yield a set of linear equations from which the Fourier coefficients have been obtained. Thus the problem is solved in principle. By energy minimization the equilibrium distribution of misfit between MS $\bar{e}_m$ and MDs $\bar{f}_m$ as well as the critical misfit f_c, or equivalently critical thickness h_c. are obtained. For the critical misfit of an epilayer of thickness h having the same elastic properties as the substrate, we have, for example [13]

$$f_c(h) = \frac{1-\nu}{4\pi(1+\nu)} \int_0^\infty \frac{dx}{x[x+\alpha(h)\phi(x,h)]} \equiv \frac{1-\nu}{4\pi(1+\nu)} I(h) ,$$

$$\phi(x,h) = \frac{1}{2} \frac{x(s^2+cs-x-x^2) + \delta h(c+s)^2}{x(s^2-x^2) + \delta h(s^2+sc+x-x^2)} ,$$

$$\delta = \frac{i-\nu}{(1-2\nu)d}, \quad \alpha = \frac{2(1-\nu)h}{d}, \tag{12}$$

$s = \sinh x$, $c = \cosh x$.

Supposing that a MD enters a coherent epilayer by climb from the free surface, the question is: how does the energy of the epilayer system depends on the distance ℓ of the MD from the free surface and on the layer thickness h? Whereas the energy per unit length of the MD increases as the MD recedes from the free surface energy of coherency (misfit) strain $\Delta E_{\bar{e}}$ (per unit length of MD) will be released. The energy of formation of a MD by climb, i.e. the change in energy of the system per unit length of MD ΔE will thus be given by (see Appendix)

$$\Delta E = -\Delta E_e + \Delta E_D$$

$$\approx -\frac{2\mu a[\ell](1+\nu)f}{(1-\nu)} + \frac{\mu a}{2\pi}\ell\, I(\ell)\,, \tag{13}$$

where

$$[\ell] = \begin{cases} \ell \text{ when } \ell < h \\ h \text{ when } \ell \geq h \end{cases}$$

and $I(\ell) = I(h)$ with $h = \ell$.

The analytical evaluation of Eq. (13) is difficult because of the complexity of the function $\phi(x)$, however numerically it was possible to show that the equation is very closely approximated, for the case of Poisson's ratio equal to 1/3, as

$$\Delta E/(\mu ac/4\pi^2) = 11.4 + 4\,\ell n(\ell/c) - 16\pi^2(\ell/c)f. \tag{14a}$$

The maximum energy ΔE^*, which in this analysis is a "nucleation energy" barrier to be overcome by the MD as it climbs to the interface, is given by the expression

$$\Delta E^*/(\mu ac/4\pi^2) = 7.4 - 4\,\ell n(4\pi^2 f) \tag{14b}$$

and occurs at the thickness ℓ^* where

$$\ell^* = c/4\pi^2 f. \tag{14c}$$

This critical thickness for nucleation is related to the critical thickness for loss of coherency, h_c by

$$h_c = \ell^*[2.85 + \ell n(h_c/c)]. \tag{14d}$$

From the relation (14d) one sees that the critical thickness h_c is greater than three times the critical thickness for nucleation ℓ^*. When the film thickness is less than the critical thickness for nucleation of an MD then the energy per unit

dislocation length always rises as it approaches the interface and no critical thickness for nucleation exists. One could state this result another way. If a MD exists at the interface when the film thickness is less than the critical thickness ℓ^* then the MD is attracted to the surface and should be eliminated from the film.

The dependences of ΔE, $-\Delta E_e$ and ΔE_D on ℓ are illustrated in Fig. 3. Whereas ΔE_D increases monotonically (essentially logarithmically), $-\Delta E_e$ decreases linearly until $\ell = h$; thereafter it remains constant. ΔE rises at first to a maximum value at thickness ℓ^*. Thereafter it decreases until $\ell = h$ at which point the slope changes discontinuously to a positive value characteristic of ΔE_D at $\ell = h$, and ΔE_e assumes a constant value $\Delta E_f(h)$. The value of thickness for which ΔE vanishes is $\ell = h_c(f)$, the equilibrium thickness at misfit f.

2.4. *Superlattices*

The problem of predicting LEMDSs in epitaxial superlattices has previously also been dealt with by Matthews and Blakeslee using the Volterra model as outlined in Eqs. (10) [22]. Naturally this involves the same shortcomings as mentioned in relation to the growing epilayer on a thick substrate. The parabolic model had also been used beneficially in this case [13,16]. The analysis in this case differs from the "growing layer on a thick substrate" in two important aspects: (i) both layers now participate in MS and (ii) the boundary conditions are different. As to MS we have the relations

$$f \approx \bar{f} + \bar{e}(1+r^{-1}R^{-1}) \; ; \; \bar{e} = |\bar{e}_B| \; ,$$

$$r = h_A/h_B\ ,\ R = \frac{\mu_A(1+\nu_A)(1-\nu_B)}{\mu_B(1+\nu_B)(1-\nu_A)}, \tag{15}$$

where the superlattice consists of alternating layers of crystals A and B of thicknesses $2h_A$ and $2h_B$ and the other quantities in (15) have evident meanings. The relations in (14) assume that no bending occurs. When the crystals have (more or less) equal elastic constants, as is often the case, also their thicknesses are the same and $f = \bar{f} + 2\bar{e}$ ($\bar{e}=|\bar{e}_A|=|\bar{e}_B|$) [13].

As to boundary conditions there are no free surfaces except the outer boundaries. The problem of LEMDSs for the interior of the superlattice has been analysed using the parabolic model. Apart from the relations (4) and (6) that hold for the interlayer interaction field periodicity with a wavelength equal to the MD spacing p exists. Additionaly the unit lying between the two midplanes of neighboring layers A and B are constrained so that the normal displacement and shear stresses vanish on these planes. When the elastic constants are equal the condition of vanishing normal displacement may be replaced more simply by zero normal stress on this plane. In this case of superlattices the problem of LEMDSs could be solved for the equilibrium MS $\bar{e}_m$ and MD density $\bar{f}_m$. For example for the critical thickness $\eta_c = h_B/b$ of the layers of type B the authors obtained

$$f(\eta_c;r,R) = \frac{(1-\nu_B)\mu(1+r^{-1}R^{-1})}{4\pi(1+\nu_B)\mu_B}\int_0^\infty \frac{dx}{x[x+\alpha\phi(x)]}, \tag{16}$$

$$\alpha = 2(1-\nu)\eta_c\ ,$$

where the function $\phi(x)$ is somewhat more complicated than $\phi(x)$ in Eq. (12).

Although an analytical expression for the integral could not be found it lends itself to fairly accurate approximations and the role of various parameters could be investigated. For example, certain useful scaling relations could be obtained: thus for $R' = 1 = r$ it was shown that $f_c(\eta_1,R_1) = f_c(\eta_1,R_2)$ where $R' = \mu/2\mu_+$, $\mu_+^{-1} \equiv \mu_A^{-1} + \mu_B^{-1}$ and $\nu_A = \nu_B$. The authors also showed that $\eta_c \sim f^{-1.3}$ whereas $\eta_c \sim f^{-2}$ for the extrapolated Frank–van der Merwe theory [1] and $\eta_c \sim f^{-1}$ (approximately) for the Volterra approach [22].

2.5. *Structural Ledge Interfaces*

An interesting LEMD related mechanism of misfit accommodation (MA) occurs in solid–solid phase transformations where the introduction of a sequence of atomic ledges of the same sign (down or up) effects a trade of the conventional sequence of efficient MDs that accommodate the disregistry parallel to a plane interface for: (i) interface terrace patches of zero average disregistry and (ii) widely spaced MDs with Burgers vector normal to the interface to accommodate the misfit that accumelates (stepwise at the ledges) over a number of patches [23,24]. Each terrace commences with a maximum disregistry on one side which decreases along the terrace to reach an equal but opposite value on the other side. In addition this disregistry is reduced by elastic relaxation due to the interlayer interaction. Primarily a ledge effects a relative displacement of the atomic paterns on either side of the interface (pattern advance), resetting the disregistry to the same value as at the beginning of the terrace. It has been shown using the parabolic model for disregistry as well as for the MDs; that this trading effects an

energetic gain depending on the misfit f and pattern advance δ at a ledge. Energetically stepped interfaces become progressively more favorable than plane interfaces as δ increases and f decreases.

3. CONCLUDING REMARKS

A basic aspect of the theoretical study of LEMDSs at epitaxial interfaces is the modelling of the periodic interfacial interaction. Frenkel and Kontorowa modelled this by a Fourier series (sinusoidal function), truncated at second order harmonics. The parabolic model is primarily a second degree Taylor series approximation of the Frenkel–Kontorowa model and its extensions. Although it is an acceptable approximation only over about one half of the period, it is believed to constitute a better approximation to the short ranged covalent type interfacial interaction in many semiconductors as compared to a low order Fourier truncation. Also it compares favorably to the Volterra model that does not provide independently for the interfacial bond strength. The main merit of the parabolic model is that it linearizes the equations governing the atomic arrangements, (resulting from the intralayer and interlayer interactions) to render the equations solvable. The model had been used to obtain useful exact solutions for the following problems: (i) in the linear chain, to assess the discrepancies between critical misfits predicted by (a) the continuum FK and parabolic models and (b) the continuum and discrete parabolic models; (ii) the nucleation and energetics of a MD that climbs from the free surface; (iii) the critical thickness (and misfit) in (a) an epilayer on a thick substrate and (b) a

superlattice and (iv) the energetics of structural ledges in solid–solid phase transformations. Clearly the parabolic model had been useful. It is foreseen that it will continue to be useful in analytical studies of interfaces.

ACKNOWLEDGEMENTS

We wish to thank the Foundation for Research Development of the South African Council for Scientific and Industrial Research, and NASA under grant NAG 1–350, for generous financial support.

REFERENCES

1. F.C. Frank and J.H. van der Merwe, Proc. R. Soc. London, Ser. A, 198 (1949) 205, 216.
2. J.W. Matthews (ed.), Epitaxial Growth, Academic Press, New York, 1975, pp 559–609.
3. J.H. van der Merwe, J. Woltersdorf and J.H. van der Merwe, Mater. Sci. Eng., 81 (1986) 1.
4. N.H. Fletcher; J. Appl. Phys. 35 (1964) 234.
5. B. W. Dodson, Phys. Rev. B 35 (1987) 2795, 5558.
6. I. Markov and A. Milchev, Surf. Sci. 136 (1984) 519; 145 (1984) 313.
7. J.H. van der Merwe, Proc. Phys. Soc. London A 63 (1950) 616.
8. J. Frenkel and T. Kontorowa, Phys. Z. Sowjetunion 13 (1938) 151.

9. P.M. Stoop and J.A. Snyman, Thin Solid Films 158 (1988) 1.

10. L.C.A. Stoop and J.H. van der Merwe, Thin Solid Films, 17 (1973) 291.

11. I. Markov and V.D. Karaivanov, Bulg. J. Phys. 5 (1978) 379; Thin Solid Films 61 (1979) 115.

12. J.H. van der Merwe and E. Bauer, Phys. Rev. B (1988) To be published.

13. J.H. van der Merwe, and W.A. Jesser, J. Appl. Phys. 41 (1988) 1509.

14. J.H. van der Merwe, J. Appl. Phys. 41 (1970) 425.

15. J.H. van der Merwe, J. Appl. Phys. 34 (1963) 117, 123.

16. W.A. Jesser and J.H. van der Merwe, J. Appl. Phys., 63 (1988) 1928.

17. V.L. Pokrovskii and A.L. Talanov, Sov. Phys. JETP, 51 (1980) 134.

18. J.H. van der Merwe, Philos. Mag. A 45 (1982) 127, 145.

19. J.H. van der Merwe, Discuss. Faraday Soc., 5 (1949) 201.

20. D.L. Tönsing and J.H. van der Merwe, Phys. Rev. B (1989) To be published.

21. J.M. Kosterlitz and Thoules, J. Phys. C 6 (1973) 1181.

22. J.W. Matthews and A.E. Blakeslee, J. Cryst. Growth, 27 (1974) 118; 29 (1975) 273; 32 (1976) 265.

23. J.M. Rigsbee and H.I. Aaronson, Acta Met. 27 (1979) 365.

24. J.H. van der Merwe, S. Afr. J. Phys. 9 (1985) 55.

APPENDIX: ENERGY OF FORMATION OF A MD

Equation (29) of ref. 13, when written in terms of the MSs $\bar{e}_x$ and $\bar{e}_y$ for an epilayer on a thick substrate, becomes

$$E_{\bar{e}} = \frac{\mu c \ell}{(1-\nu)a}[\bar{e}_x^2 + \bar{e}_y^2 + 2\nu\bar{e}_x\bar{e}_y] ,$$

where c is the reference lattice spacing [7]. The misfit stress $\bar{p}_{xx}$ which may be written as $\partial E_e/\partial(\ell\bar{e}_x)$, accordingly becomes

$$\bar{p}_{xx} = 2\mu(1+\nu)f/(1-\nu)$$

in the coherent $\bar{e} = f$ configuration of an interface with quadratic symmetry $\bar{e}_x = \bar{e}_y = \bar{e}$. The work done per unit length of MD by $\bar{p}_{xx}$ when the MD climbs to a depth ℓ below the free surface is accordingly

$$-\Delta E_e = -\bar{p}_{xx}\ell a = -2\mu a\ell f(1+\nu)/(1-\nu) \qquad (A1)$$

The work done per unit MD length to form a MD at a depth ℓ against the linear stress p_{xx} emanating from the MD (p_{xx} may be obtained from Eqs. (14) and (A8) of ref. 13) is given by

$$\frac{1}{2}\int_0^{\ell} p_{xx}(0,z)a\,dz .$$

This is nothing but the energy per unit length of MD and may be obtained from Eqs. (30) and (A8) of ref. 13 as

$$\Delta E_D = \lim_{p\to\infty} \frac{\mu c}{4\pi^2} p\epsilon_D ,$$

$$\epsilon_D = \sum_{n=1}^{\infty} \frac{1}{n[n+\beta^{-1}\phi(x)]}, \beta = \pi d/(1-\nu)p \; . \tag{A2}$$

where $\phi(x)$ has been defined in Eqs. (13). If we write

$$x = \frac{2\pi a\ell}{p}n \;\; , \;\; dx = \frac{2\pi a}{p}dn \; ; \; dn = 1$$

and let $p \to \infty$ we obtain the result in Eqs. (13).

FIGURE CAPTIONS

Fig. 1. Comparison of the sinusoidal (Eq. (2a)) and parabolic (Eq. 3) potential representations; curves A and B respectively.

Fig. 2. Monolayer (ML) configuration with a MD of finite length (between ML and substrate) terminating with Taylor dislocations (within the ML) at A and B.

Fig. 3. Graph of the energy per unit length of MD (see Eqs. (13) and (14)) in units of $\mu ac/4\pi^2$ plotted against depth in units of ℓ/c. The curves illustrate the energetics of a MD that is formed at the free surface and climbs to a depth ℓ in an epilayer of thickness h on a thick substrate. Curves A, B, C represent respectively the work (ΔE_D) done in forming the MD in a strain free overlayer, B the work (ΔE_e) gained from the MS (Peach Koehler) stresses and C the resultant energy of formation ΔE of the MD. The nucleation parameters ℓ^* and ΔE^* have evident meanings. The figure is drawn for f = 0.001, h = 260c and h_c = 207c.

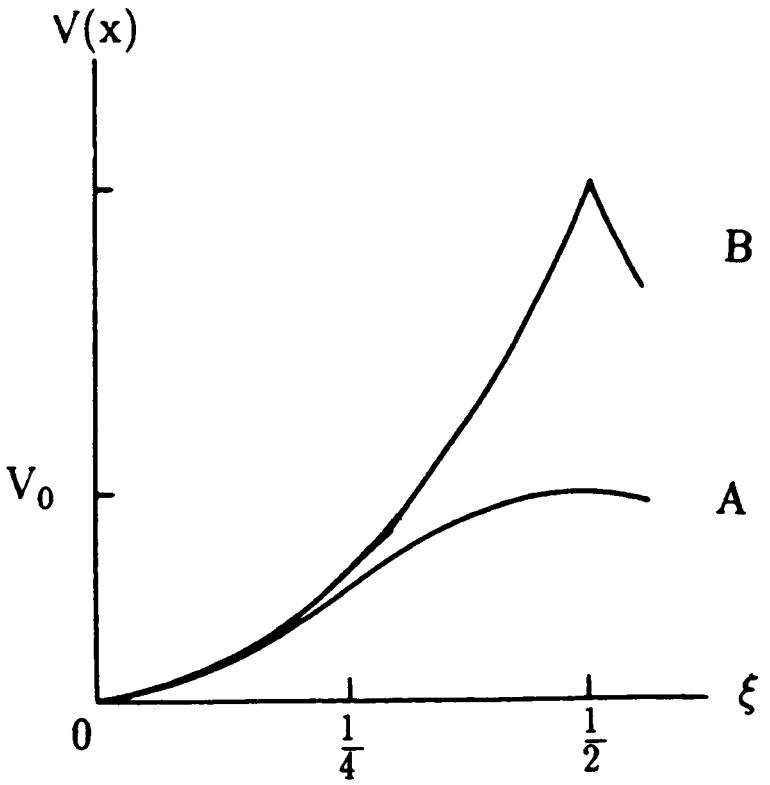

Fig. 1

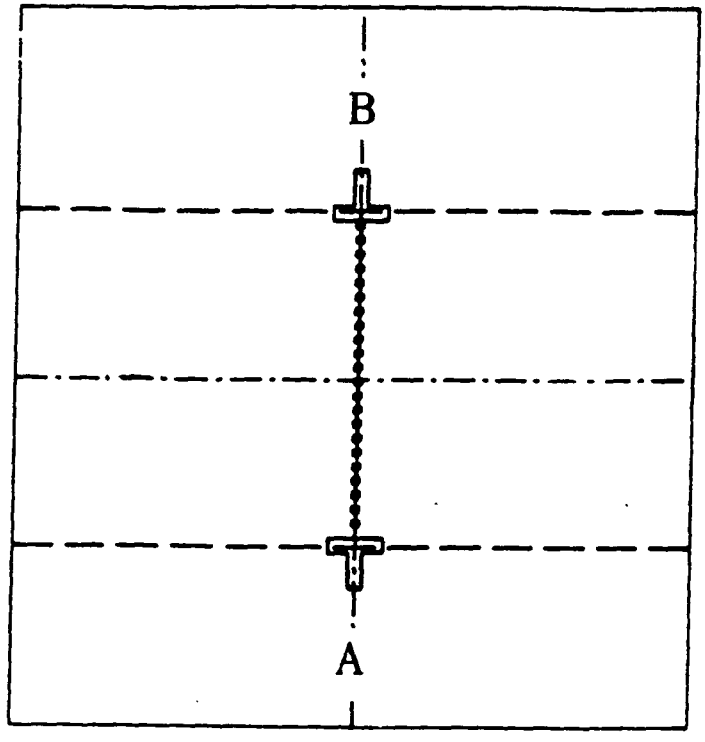

Fig. 2

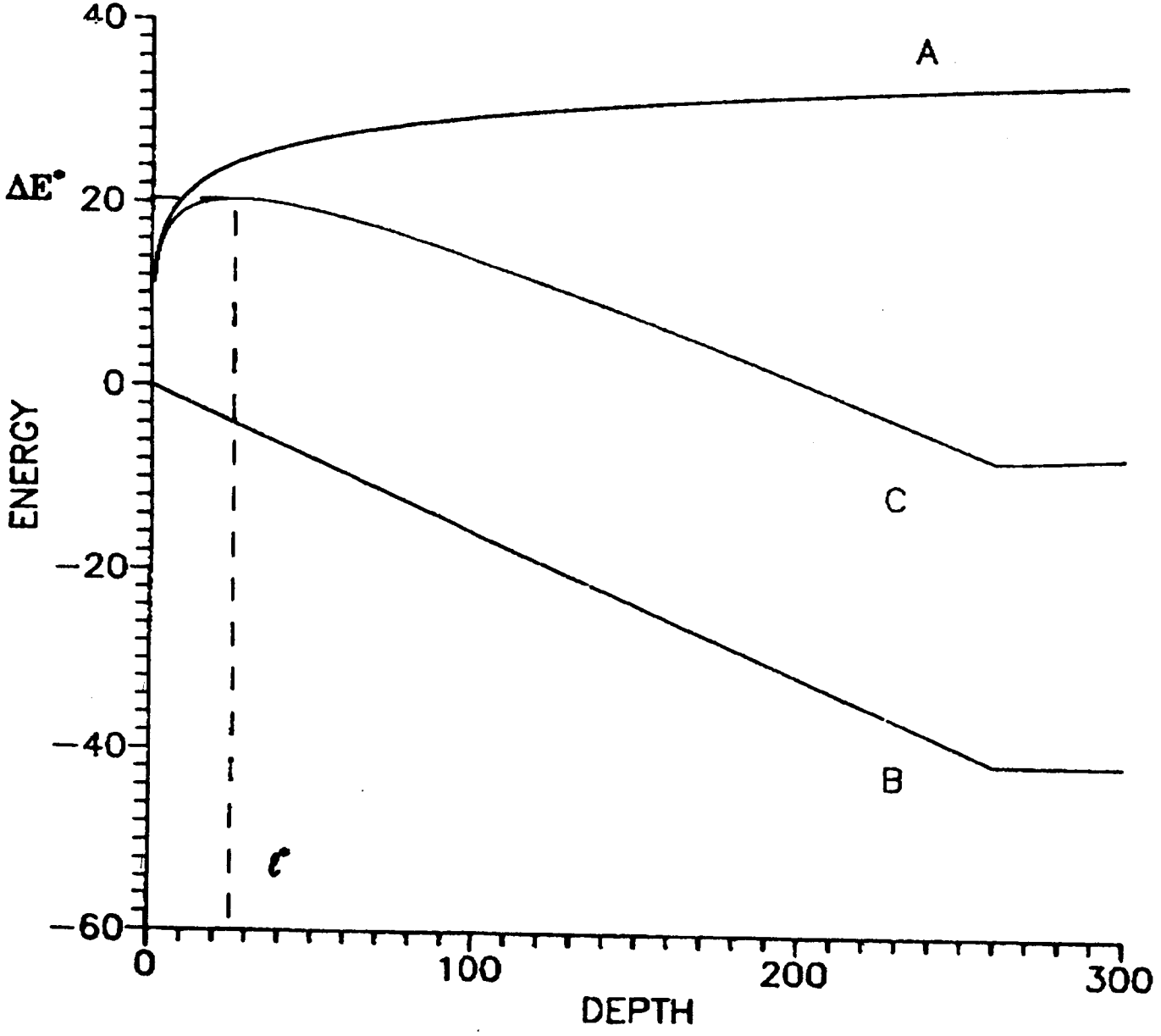

Fig. 3

THE PHYSICAL FOUNDATIONS OF CRITICAL PARAMETER CALCULATIONS IN EPITAXY

Jan H van der Merwe and W A Jesser*

Physics Department, University of Pretoria, Pretoria 0002
Republic of South Africa

The critical misfit f_c and critical thickness h_c at which an epitaxial bicrystal undergoes a coherent to incoherent (C- IC) transition are of great importance. The purpose of this paper is to reconcile the conditions (quasi- equilibrium/ non- equilibrium) under which growth occurs with the physical bases of the theoretical predictions of the critical parameters. Because of the existence of energy barriers to the acquisition of the MDs, which are needed for the transition, it is inevitable that some data will reflect a time lag between growth and equilibration. The speeding up of the equilibration process by the barrier height reducing Peach- Koehler MS forces is highlighted. Criteria for predicting critical parameters are set up and the phenomenological (Frenkel- Kontorowa and Volterra) and ab initio models used in calculating these parameters are explained and appraised. Examples, demonstrating calculations and their correlation with empirical data in each category, are briefly described.

INTRODUCTION

The transition of an epitaxial bicrystal, A on B, from registry (coherency) at the interface to disregistry (incoherency) is of great fundamental and technological importance. This transition, which will be referred to as the C- IC transition, has recently become topical again (1- 4). It is therefore vital that the principles and models used to predict the critical thickness h_c (or misfit f_c) at which the transition occurs, be clearly understood and their reliability be assessed, theoretically and experimentally (5,6).

The misfit between A and B is most simply defined for interfaces with rectangular lattice symmetry as (4)

$$f_i = (a_i - b_i)/d_i \; ; \; i = x,y \, , \qquad (1)$$

where the value of d_i is between a_i and b_i and depends primarily on the relative thicknesses h_A and h_B. When the interfacial symmetries are different, as for {111} fcc/{110} bcc interfaces (8) specification of f needs more care. Recently Braun (9) has introduced a powerful description of the interface geometry (including misfit) in terms of reciprocal lattice vectors.

Of the various modes of misfit accommodation (MA) illustrated in Fig. 1 (5,6) we shall limit ourselves mainly to (i) misfit (homogeneous) strain (MS $\equiv \bar{e}$), (ii) misfit dislocations (MDs of density f) which may be efficient (EMDs) or inefficient (IMDs) when their Burgers vector is inclined to the interface or even split into partials and (iii) a misfit vernier (MV) when the misfit is large and/or the interfacial bonding small so that the periodic strains associated with MDs practically vanish. An interesting atomic configuration of MA, referred to as a *misfit- Taylor dislocation* (MTD), (10) occurs in ultrathin overlayers of one ML or less. The C- IC transition may namely be realized by the nucleation of a finite segment of MD at the coherent interface and its subsequent growth by a zip- like mechanism effected by the consecutive inclusion (or ejection) of adatoms until the MD reaches a boundary edge of the overlayer. In the finite segment stage the configuration is equivalent to an extra (or missing) row of overlayer atoms terminating within the overlayer sheet. This constitutes a "Taylor" dislocation within the overlayer sheet and hence the name MTD for this combination of dislocations. MTDs may be operative at the C- IC transition in MLs. Two more forms of MA, namely, "*ruptures*" and "*distortions*" will conveniently be introduced at a later stage. Misfit curvature (MC) and misfit strain gradient (MSG) modes of MA (5,6) also exist but will be ignored below.

When the misfit in a given direction is jointly accommodated by MDs and MS we have approximately

$$f = \bar{f} + \bar{e}, \quad \bar{e} \equiv |\bar{e}_A|, \text{ A thin (11)} \tag{2a}$$

$$f = \bar{f} + P\bar{e}, \quad \bar{e} \equiv |\bar{e}_B|, \quad P = 1 + \frac{h_B \mu_B (1-\nu_A)}{h_A \mu_A (1-\nu_B)} \tag{2b}$$

where Eq. (2b) applies to layers with arbitrary thicknesses $2h_A$ and $2h_B$, for a system that is constrained not to bend (12). If the temperature at which measurements are made differs from the growth temperature the thermal strains may be different in A and B and must be taken into account. In technology thermal strains are often matched (approximately) in accordance with Vegards Law by selecting suitable alloy compositions (13).

BASIC CONSIDERATIONS

The growth of epitaxial bicrystals involves both the dynamics of adatom surface migration and of the acquisition of MDs. Both processes require the overcoming of characteristic energy barriers of which the migration activation barriers Q are typically between 0.5 and 1.0 eV in metal surfaces with closest

packing (14). The energy barriers U_{ac} opposing the acquisition of MDs, form a hierarchy with a maximum near 200 eV for the nucleation of a MD loop in a perfect 3D crystal, down to a fraction of an electron volt per atom length of MD for overcoming the Peierls barrier to glide (11). Whereas the overcoming of the former by *thermal* energy is completely prohibitive, the latter may be reduced by Peach- Koehler (MS) forces (5,6) to within the range (< 30 kT) of thermal activation for a finite length of MD, and may even be reduced to zero if the MS is large enough. The driving forces (5,6,11) for the abovementioned processes - adatom migration and MD acquisition - are the free energy gradients ΔA. These forces vanish when the equilibrium configuration is realized at minimum (free) energy A :

$$A = \text{min.} \approx E \equiv \text{mechanical energy.} \qquad (3)$$

The approximation $A \approx E$, applies to MDs which form regular arrays in an otherwise perfect crystal. It is exactly valid only at absolute zero. The majority of past calculations have been based on (11) the equilibrium principle in Eq. (3) . Non- equilibrium calculations are of mostly of recent origin and represent attempts at (i) calculating activation energies (15), (ii) at establishing the reduction of such barriers by Peach- Koehler forces (16) and (iii) at describing the dynamics of MS relief by the introduction of MDs (17).

An aspect which is of great significance in the considerations below is the growth mode (18). For the clear exposition of the basics of our considerations we may retrict ourselves to the growth of overlayers of uniform thickness as in ML- by- ML growth. In quasi- equilibrium growth this requires the adatom- substrate (A- S) bonding to be strong in comparison with adatom- adatom (A- A) bonding. Relatively strong A- A bonding induces island growth at quasi- equilibrium, but may be coerced (19,20) into layerlike growth by an appropriately high supersaturation (non- equilibrium) so that single atoms are super critical nuclei and have sufficient time and mobility to move into a local (crystalline) stable position. This is a technologically important case.

MODELS

The minimization procedures, involved in the application of the governing condition (3), can be carried out either through ab initio or phenomenological model calculations (5,6). In the former the minimization proceeds directly by Monte Carlo or other methods using appropriate interatomic potentials (15). In the phenomenological models (11,21,22) the most relevant features of the epitaxial system are at first modelled phenomenologically for use in deriving energy expressions that can be minimized with respect to the relevant variants. Whereas the ab initio models are quantatively more accurate, depending on the reliability of the potentials employed, the phenomenological models have greater generality and wider predictive powers.

The phenomenological models can be broadly classified according to the representation of the interfacial (A- B) interaction in: (a) the Volterra model (11) with an interfacial registry constraint (except at the dislocation line singularity), and (b) the Frenkel- Kontorowa (FK) model (20- 23) in which the A- B interaction is represented by a periodic potential V of wave length d lying

between a and b as in Eq. (1). In the original 1D FK model (21,23) consisting of single (ML) chain of particles on a rigid (thick) substrate (see Fig. 2)

$$V(x) = \tfrac{1}{2} V_0[1-\cos(2\pi x/b)], \quad d = b . \tag{4}$$

V_0 may be linked (8) to the adatom migration barrier Q and the desorption energy E_{des} (A- B bond strength) by

$$Q = gV_0 = \kappa E_{\mathrm{des}} . \tag{5}$$

The factors g and κ depend on the dimensionality, geometry, and bond type and strength. In Eq. (4) $g = 1$.

The form (4) of V constitutes a 1D truncated Fourier series which may be extended to 2D (8,9,20) in the interface between a ML and a thick substrate or to a "linear" interface with 1D misfit between two thick crystals, as in the generalized Peierls- Nabarro model (22). The truncation is justified by the fact that the Fourier coefficients C_i decay rapidly with harmonic order (24). A significant property of the Fourier description is that each Fourier term defines an epitaxial orientation and geometry in which the magnitude of the corresponding coefficient is a measure of the tendency towards realization of that orientation (8,9).

The potential in Eq. (4) degenerates into the *parabolic* potential (25,26) - - a sequence of parabolic arcs - - when Taylor- developed and truncated at lowest (2 degree) powers:

$$V = \tfrac{1}{2} b^{-2}\pi^2 V_0 x^2 , \quad |x| \le \tfrac{1}{2} b . \tag{6}$$

The A- A and B- B interactions in both the Volterra and FK (11,21,23) models have been modelled in terms of the nearest- neighbor isotropic harmonic (linear elasticity) approximation. More recently this has been extended to anisotropic elasticity (9), nearest- neighbor anharmonicy (Fig. 3), (27) and to the case where the bulk structure of the overlayer has a "nearby" metastable phase that can be reached by a homogeneous strain (MS) (28,29). In the latter two cases the point(s) of inflection where the pair potential and free energy of the initial structure becomes unstable, has dramatic consequences. In the former case this leads to additional modes of MA "ruptures" in case of strong A- B bonding in combination with positive misfit or to "*distortions*" at negative misfit. In both cases this occurs because the (nearest- neighbor) bonds become streched beyond the point of inflection. While the "rupture" may be seen as a degenerated MD, the "distortion" represents a stretched bond that separates overlayer segments of close registry with the substrate. The overlayer segments represent dimer, trimer.... groupings, depending on the magnitude of the misfit.

It is proper at this point, to record some expressions for the energies as calculated for the simplest case of isotropic modelling. Since the MS energy $E_{\bar{e}}$ and the energy E_{p}' per unit length of MD in a sequence of regularly spaced (p) MDs are additive, we may specify them separately:

$$E_{\bar{e}}^{(A)} = B_A h_A \bar{e}_A^2 \text{ per unit area for } A, \tag{7a}$$

$$B_A = 2\mu_A(1+\nu_A)/(1-\nu_A), \tag{7b}$$

as introduced by Matthews (11) and by Jesser and Kuhlmann- Wilsdorf (30),

$$E_D' = (\mu_{AB} cp/4\pi^2)[1+\beta-(1+\beta^2)^{\frac{1}{2}}-\beta \ell n\{2\beta(1+\beta^2)^{\frac{1}{2}}-2\beta^2\}], \tag{8a}$$

$$\beta = \frac{2\pi\lambda_0 d}{\mu_{AB} p}, \quad \frac{1}{\lambda_0} = \frac{1-\nu_A}{\mu_A} + \frac{1-\nu_B}{\mu_B} \tag{8b}$$

in the FK model (22,25) (for $h_A, h_B \geq \frac{1}{2}p$) and

$$E_D' = Db\{\ell n(R/b) + 1\}, \quad D = \lambda_0 b/2\pi, \quad d = b \tag{9a}$$

$$R = \begin{cases} \frac{1}{2}p & \text{when } h_A > \frac{1}{2}p \\ h_A & \text{when } h_A < \frac{1}{2}p \end{cases} \tag{9b}$$

in the Volterra model (11). In the foregoing μ and ν are respectively the shear modulus and Poisson's ratio, the subscripts A and B have evident meanings, and R is referred to as the "cut off" radius. Note that the interfacial shear modulus μ_{AB}, measuring the strength of interfacial interaction, is absent in the Volterra model. It is of interest that the expression (8a) reduces to the form (9a) in the limit of large p, with $R \cong p/10$ rather than $\frac{1}{2}p$. Furthermore, the relations (8) and (9) (relating to zero MS $\bar{e}$) may be adapted for nonzero MS simply by replacing d and p by $\bar{d}$ and $\bar{p}$ with $\bar{f} = \bar{d}/\bar{p}$, where the bar designates quantities in a configuration with MS. For purposes of energy minimization we need expressions for the total energy per unit area of interface:

$$E = E_D + E_{\bar{e}} \equiv E(h_A, h_B, f, \bar{f} \text{ or } \bar{e}) \tag{10}$$

$$E_D = (\bar{p}_y E_D'^{(x)} + \bar{p}_x E_D'^{(y)})/\bar{p}_x\bar{p}_y \text{ per unit area} \tag{11a}$$

$$= 2E_D'/\bar{p} \text{ for quadratic symmetry } (\bar{p}_x = \bar{p}_y \equiv \bar{p}). \tag{11b}$$

Equations (11) relate to cross grids of MDs in which the crossing energy of MDs may be neglected (22). $E_D^{(x)}$ is the energy per unit length of a MD in a sequence of MDs spaced at distances $\bar{p}_x$ along the x- axis.

CRITERIA

Because of its greater analytical simplicity we adopt a *criterion* for locating criticality (5,6,11), the limiting operation

$$\bar{f} \to 0 \quad \text{(equivalently } \bar{e} \to f\text{)} \tag{12}$$

defining the inverse (IC- C) of the C- IC transition. Using relations (3) and (10) we accordingly obtain the minimization equations

$$0 = \frac{\partial E}{\partial \bar{e}} = \frac{\partial E}{\partial \bar{f}} \text{ for } (\bar{e},\bar{f}) = (\bar{e}_m,\bar{f}_m) \,. \tag{13a}$$

The relation $\partial E/\partial \bar{f} = 0$ in (13a) is equivalent to requiring that the work W needed to introduce an additional MD vanishes:

$$W = \int F ds = 0 \text{ at } (\bar{e},\bar{f}) = (\bar{e}_m,\bar{f}_m). \tag{13b}$$

In the relation (13b) F is either an external force on the "free" surface that is used to introduce the dislocation and ds an element of displacement of the surface (21) or F is the resulting integrated force on an existing TD (threading dislocation) (11) that glides to lay the MD in the interface and ds in the element of displacement of the TD. When F is constant Eq. (13b) is satisfied by

$$0 = F = F_{\bar{e}} - F_D, \tag{13c}$$

where $F_{\bar{e}}$ is the MS (Peach- Koehler) force on the TD and F_D the line tension of the MD being laid in the interface.

By combining the criticality criterion (12) with one of the equilibrium conditions (13) we obtain the equations (5,6)

$$\Phi(h,f_c;\bar{f} = 0) = 0 \text{ and } \Phi(h_c,f,\bar{f} = 0) = 0 \tag{14}$$

defining respectively the critical misfit f_c and the critical thickness h_c. Φ is the corresponding function in Eqs. (13).

The fact that in the 1D FK model the density $\bar{f}$ of MD rises abruptly and closes rapidly in on f at the C- IC transition was taken as ground to introduce the approximation (25)

$$E_{\bar{e}}(h,f_c;\bar{e} = f) \approx E_D(h,f_c;\bar{f} = f) \tag{15}$$

However, this equation overestimates (12) f_c by up to 30%.

The adoption of equilibrium criteria implies that MD are freely available,

which they are not because of the existence of energy barriers (5,6,11). In analogy to (13b) the energy barriers to nucleation and motion may be obtained by integration as

$$U_{ac} = \int_{bar} F\, ds. \tag{16}$$

When reduced by Peach- Koehler forces to about 30 kT and less, the barriers can be overcome by thermal fluctuations, given enough time. When reduced to zero

$$U = 0 \tag{17}$$

the nucleation or motion is spontaneous. In the sections below we briefly describe calculations of f_c and h_c with the view of illustrating the foregoing considerations. The illustrations are not exhaustive. Equilibrium predictions using Eq. (13c) will be referred to as force criteria, and energy criteria otherwise (5,6). When applicable force criteria are superior in simplicity.

EQUILIBRIUM CALCULATIONS

Although the Frank and van der Merwe theory (21) based on the 1D FK model, lacks accuracy when applied to real systems, it has the merits of simplicity and capability of making predictions about most of the features of present interest: a critical misfit

$$f_c^{(1)} = \frac{2}{\pi \ell_0}, \quad \ell_0^2 = \frac{n\mu^{(1)} b^2}{2V_0}; \quad n = 1 \text{ for a ML}, \tag{18}$$

obtained using either (13a) or (13b) and a critical thickness (see Eq. (14))

$$h_c^{(1)} = n_c a = 8 V_0 a / \pi^2 \mu^{(1)} b^2 f^2 \tag{19}$$

In these equations the superscript (1) designates 1D FK model. $\mu^{(1)}$ is the force constant of the springs modelling the harmonic forces (Fig. 2). Accordingly the ratio ℓ_0 is a measure of the relative strength of the competing interactions; as expressed by $\mu^{(1)}$ and V_0. The authors estimated that $f_c^{(1)} \approx 9\%$ and that $f_c^{(1)}$ decreases in going to a 2D quadratic model to $f_c^{(2)} = 7\%$, due to the Poisson effect (31) in a model with a 2D quadratic interface.

In 2D interfaces of the kind {110} fcc/{110} bcc (8,32) the misfits f_x and f_y differ greatly and are most suitably expressed in terms of the ratio

$$r = a_{nn}/b_{nn} \tag{20}$$

of nearest- neighbor distances a_{nn} and b_{nn}. Aspects of great significance in the Fourier description for these cases are (8,9): (i) that minima of interfacial energy exist for atomic (ideal- epitaxial) configurations in which sequences of atomic

rows on either side of the interface come into 1D registry (coherency, i.e. wave vectors on either side of the interface match), (ii) that the matching and depth are determined by the wavelength of the Fourier term and the magnitude of its coefficient respectively, (iii) that the depth of the minimum is a measure of the epitaxial tendency of the related epitaxial configuration (dimensionally and orientationally) and accordingly provides the driving force for MS towards coherency, and (iv) the depths are additive when 2D registry is attained. "Phase diagrams" defining the critical values r_{xc}, r_{yc} and r_c have been constructed (8,9). In these, in which r_{xc} and r_{yc} mark (19,20,21) 1D coherency breakdown the critical thickness is also defined approximately as in Eq. (19).

The main consequence of including anharmonicity (27) is that the critical parameter values branch into two: the lower branch for positive natural misfit (coherent overlayer compressed) and upper one for negative natural misfit. The magnitudes of the gaps depend on the details of anharmonicity and the misfit in relation to the point of inflection of the nearest- neighbor (nn) pair interaction. The "rupture" and "distortion modes of MA obtain when the nn bond is stretched beyond its theoretical strength (see Eqs. 1(f) and 3(a)).

An important generalization of the foregoing is the one in which there exists a structural transition of the crystal to a nearby metastable phase (28,29) that can be reached by homogeneous (misfit) strain $\bar{e}$, for example, the tetragonal distortion of a bcc phase into a metastable fcc phase. There exist critical parameters for both the stable and metastable configurations as well as two points of inflection of the free energy vs strain curve. If the critical misfit in the stable ML configuration exceeds the first point of inflection and the natural misfit closely matches the tetragonal strain corresponding to the metastable configuration, the "critical" thickness (thickness to which the metastable configuration persists) will be very much higher than would be expected on basis of the harmonic (approximation) calculation, assuming that the existing misfit be accommodated by MS (referred to the stable phase) alone.

The model on which Eq. (19) (and its 2D extension) for the critical thickness is based, suffers from the deficiencies as listed above. Equally important deficiencies are that the strain gradients in the overlayer normal to the interface, have been neglected (8,9,19,20). Previous attempts at including these have all been approximate. *Exact solutions* (12,26) had only been obtained by using a parabolic (approximate) potential of the form in Eq. (6). The resulting expressions are complicated. For the case of a single multilayer on a thick substrate (both isotropic and with equal elastic moduli μ and ν, including the interface) the critical misfit is given by (26)

$$f_c(h) = \frac{1-\nu}{4\pi(1+\nu)} \int_0^\infty \frac{dx}{x[x+\alpha\Phi(x,\alpha)]}, \quad \alpha(h) = \frac{2(1-\nu)h}{a}, \tag{21}$$

where $h \equiv h_A$ and Φ is a complicated function of x and α, that increases in complication when the moduli are different. The Volterra approach instead yields (11) the results (see Eqs. (7), (9), (13))

$$f_c(h) = \frac{D}{2Bh}[\ell n(h/b)+1],\ h_c = \frac{D}{2Bf}[\ell n(h_c/b) + 1]. \qquad (22)$$

These are evidently simpler than Eq. (21) to analyse, but are rather inaccurate at small thicknesses because of the approximations involved in Eqs. (9). Note that in Eq. (22) (and also in (24) below) $h \equiv h_A$ falls in the class $h_A < \frac{1}{2}p$ of Eq. (9b).

The parabolic model has also been used to obtain (exact solutions) critical parameters (12) e.g. for a superlattice obtained by stacking A and B layers in alternating succession, giving

$$f_c(h_B) = \frac{(1-\nu_B)P\mu_{AB}}{4\pi(1+\nu_B)\mu_B}\int_0^\infty \frac{dx}{x[x+\alpha(h_B)\Phi(x)]}, \qquad (23)$$

where P is defined in Eq. (2b), $\alpha(h_B) = 2(1-\nu_{AB})h_B/d$ and Φ is a complicated function of elastic constants and thicknesses, which simplifies greatly when these quantities are respectively equal. Of interest is that h_c/d varies approximately as $f^{-\kappa}$ with $\kappa \approx 1.22$ as compared to $\kappa = 2$ in Eq. (19) and $\kappa \approx 1$ in Eq. (22). Energy criteria were also used in conjunction with the relations (7) and (8) to calculate approximate values for (i) the critical thickness in an epitaxial bicrystal of two equally thick halfcrystals (33) and (ii) the critical size R_c of a hemispherical island on an extended 3D substrate crystal with similar elastic constants (34). The results in (i) correlated excellently with the observed critical thickness in PbS/PbSe epitaxial bilayers (35). The results in (ii) were qualitatively correct (34).

The power of the force criterion in Eq. (13) (combined with the Volterra model) to handle rather complicated cases is demonstrated by the derivation of the critical layer thickness (11)

$$h_c = \frac{Db(1-\nu\cos^2\alpha)[\ell n\ (h_c/b)+1]\ +\ \sigma_0 b\ \sin\ \alpha}{Bbf\cos\ \lambda\ -\ \gamma_0/\cos\ \Phi} \qquad (24)$$

in which the C- IC transition was accomplished by imperfect partial dislocations. In relation (24) λ is the angle between the slip direction and the interfacial direction which is normal to the line of intersection of the slip plane (S) and the interface, α is the angle between the line of the mixed MD and its Burgers vector, and Φ is the angle between the unit vector $\hat{n}$, normal to S and the line of intersection of the free surface and a plane which is normal to the free surface and contains $\hat{n}$. Also, γ_0 and σ_0 are respectively the stacking fault energy and the surface free energy of the overlayer surface where the gliding TD traces a step of height $b\cdot\sin\ \alpha$. A relation analogous to (24) can be written down for a MD that is generated by climb (11). The force criterion was also applied to superlattices by Hirth and Evans (36) and to multilayers, (allowing for bending) by Bokii and Kuznetsov (37) with satisfactory results.

NON- EQUILIBRIUM CALCULATIONS

We have classified as *non- equilibrium* (5,6) all calculations (i) of energy barriers to the acquisition (nucleation as well as motion) of MDs, (ii) of barrier reduction and elimination by Peach- Koehler forces and (iii) of the dynamics of the acquisition processes, e.g. dynamics of MS relief.

The Frank- van der Merwe (21) analysis of the 1D FK model is also very powerful in illustrating non- equilibrium considerations, particularly features (i) and (ii) listed above. The authors considered the introduction (nucleation) of a MD from the free end of of the chain by displacing the endpoint by an amount $a\xi$, obtaining for a coherent layer $U(\xi,f)$ and $U_{ac}(f)$

$$U(\xi,f) = (2V_0\ell_0/\pi)[(\cos\pi\xi_0-\delta_1\cos\pi\xi)+f\ell_0(\pi\xi_0-\pi\xi)+2\delta_2], \tag{25a}$$

$$U_{ac}(f) = (4V_0\ell_0/\pi)[(1-f^2\ell_0^2)^{\frac{1}{2}} - f\ell_0\arccos(f\ell_0)], \tag{25b}$$

where $\delta_{1,2} = 1,0$ and $\delta_{1,2} = -1,1$ when $\xi < 1$ and $\xi > 1$ respectively. The displacement $a\xi_0$ of the *free* end is defined by

$$\sin\pi\xi_0 = f\ell_0. \tag{26}$$

A MD is created when the endpoint is displaced from the stable equilibrium position $\xi = \xi_0$ to the nearest stable equilibrium position $\xi = 1+\xi_0$. When $f = 0$ it follows from Eq. (26) that $\xi_0 = 0$, defining (see Eq. (25a)) the nucleation barrier at zero misfit as

$$U(\xi,0) = (2V_0\ell_0/\pi)(1-\cos\pi\xi)\ . \tag{27}$$

The residual part of $U(\xi,f)$ constitutes the reduction of the barrier by the Peach- Koehler MS forces. These considerations are illustrated in Fig. 4. Though Fig. 4 corresponds to a misfit $f = f_c$ ($\equiv$ equilibrium critical misfit $2/\pi\ell_0$) the acquisition of MDs is still confronted with a nucleation energy barrier. Curves A, B and C display respectively the natural barrier $U(\xi,0)$, the details of the reduction of the barrier by the Peach- Koehler forces and the net barrier $U(\xi,f_c)$. The barrier height attains the values $4W\ell_0/\pi$, $1.6V_0\ell_0/\pi$ (approx.) and 0 when $f = 0$, f_c and f_s respectively. At

$$f_s = \ell_0^{-1} = \tfrac{1}{4}\pi f_c^{(1)} \quad \text{or} \quad h_s = \tfrac{1}{4}\pi^2 h_c\ , \tag{28}$$

the barrier vanishes and MDs enter spontaneously (without thermal aid). For metals $V_0\ell_0/\pi$ is of the order of 1.0 eV so that even at $f = f_c$ the activation energy barrier is still about 1.0 eV per atom length of MD. The nucleation barrier for a finite (critical) length of MD may thus still be above the common thermal threshold of about 30kT. It thus seems likely that "easy" sources of MDs are normally needed for the C- IC transition, for example threading dislocations. The energy barriers to the nucleation of MDs at the crystal surface using continuum theory and in the interior of a crystal with MS, using ab initio (16) calculations had also been carried out with equivalent conclusions. The foregoing considerations are helpful in interpreting recent results obtained by

Tsao et al (16) regarding MS relief in Si_xGe_{1-x} strained layers grown on (001) Ge substrates. The authors have compared the components $F_{\bar{e},par}$ of the Peach- Koehler MS and the dislocation (presumably TDs) line tension $F_{D,par}$ forces with the observed MS relief and concluded that "metastable strained layer breakdown is most directly correlated with (i) an excess stress $\tau_{exc} = 2(F_{\bar{e},par} - F_{D,par})/hb$ and (ii) absolute temperature T". When written as an excess force

$$F_{exc} = F_{\bar{e},par} - F_{D,par} \tag{29}$$

the result represents the non- equilibrium $(F \neq 0)$ equivalent of relation (13c) and may be correlated with the reduction of the activation barrier as described by Eqs. (25a) and (25b), and displayed in Fig. 4. Although Eqs. (29) and (25), refer to different cases, the former to the motion of a TD and the latter to the nucleation of a MD, the principles involved are the same. The relation $F_{exc} = 0$ still describes equilibrium as in Eq. (13c). When $F_{exc} \neq 0$, the related barriers delay MS relief which only begins to occur with an observable rate at an excess stress $\tau_{exc}/\mu \approx 0.024$ that has reduced the Peierls barrier of a critical length of TD to within the thermal threshold at 494°C.

A very significant mechanism of reducing the energy barrier to the acquisition of MDs, and to tailoring them, has been discovered by Fischer et al (39) for the growth of *GaAs* on (100) *Si* substrates. The authors had cut the (100) *Si* surface off axis (tilted optimally by about 4° towards [001]) so that the exposed *Si* surface forms a 2D staircase with ledges along [011] and [0$\bar{1}$1]. The ledges apparently facilitate the recruitment of substrate dislocations that terminate at the *Si* surface, drawing them into the ledges where they become fairly tightly bound EMDs that are prevented from threading the *GaAs* overlayer.

Dodson and Tsao (17) have also analysed the dynamics of MD acquisition using a phenomenological model and making various assumptions of which the most prominent one is that the dislocations are driven by a local stress τ_{loc} that can be expressed in terms of a residual mismatch $f - \bar{f}(t) - \bar{e}(h)$, in which $\bar{e}(h)$ is the equilibrium MS at thickness h and the MD density $\bar{f}(t)$ at time t falls short in accommodating the residual mismatch. The authors derived the equation

$$d\bar{f}/dt = C\mu^2[\bar{f} - \bar{f}(t) - \bar{e}(h)]^2(\bar{f} - \bar{f}_0) \tag{30}$$

for the dynamics of MS relief; C being a constant and $\bar{f}_0$ a background dislocation density. Equation (30) gave a satisfactory description for the observed MS relief of SiGe alloys on (100) Si substrates with values of C and $\bar{f}_0$ given by C = 30.1 and $\bar{f}_0 = 10^{-4}$.

SUMMARY AND CONCLUSIONS

The main objectives of this paper have been (i) to correlate the principles, specifically equilibrium/non-equilibrium principles on which critical misfit/critical thickness calculations are based, with the experimental conditions, under which they are measured, (ii) to describe the models used in the calculations, (iii) to highlight the deficiencies, advantages and disadvantages of the various approaches, and (iv) to briefly describe sample calculations, making no attempt to be exhaustive. We have given prominence to the prediction of critical parameters in layer growth, an approach which we justified on the grounds that either an overlayer has a natural tendency to grow in the ML-by-ML mode because of strong adatom-substrate bonding and subsequent homoepitaxial growth, or the overlayer can be coerced into growing layerlike by a suitable supersaturation. In a C-IC transition the coherency breakdown is accomplished by the acquisition of MDs, a process which is opposed and retarded by a hierarchy of energy barriers, some of which naturally fall within the thermal threshold and others which need be reduced, mainly by Peach-Koehler forces to come within this threshold. It is therefore not suprising that measured critical values are scattered over regimes covering the entire range from non-equilibrium to equilibrium values. In metals where the Peierls barriers to dislocation glide are believed to be relatively small many observations correlate with equilibrium predictions whereas in semiconductors with strong localized covalent bonding the barriers are high and many observations cannot be correlated with equilibrium calculations (5,6,11). Because of their relative simplicity most calculations of the past have been equilibriumwise. However, recently there had been a growing interest and activity in the non-equilibrium regime. This is very timely particularly because of the technological importance of semiconducting materials.

As to models, the phenomenological models of Volterra and of Frenkel and Kontorowa have been the main goals of the past and will still play a significant role in the future because of their simplicity and greater analytical predictive power. Ab initio calculations, rightly aimed at greater accuracy, will be carried out with increased frequency.

The Volterra model has the advantage of greater simplicity, particularly in allowing for varying layer thickness, whereas the FK-model is more accurate at small (1-2 MLs) thicknesses and in addition is very powerful in selecting epitaxial orientations for which the coherency lock-in power is maximal.

Acknowledgements

The authors gratefully acknowledge financial support from the Foundation for Research Development of the South African Council for Scientific and Industrial Research and also from NASA under grant NAC-1-350.

*Permanent address: Department of Materials Science, University of Virginia,

Charlottesville, Virginia, USA.

REFERENCES

1. R.People and J.C. Bean, Appl. Phys. Lett. 47, 322 (1985) [Errata; 49, 229 (1986)].
2. B.W. Dodson and P.A. Taylor, Appl. Phys. Lett. 49, 642 (1986).
3. Y. Kohama, Y. Watanabe and Y. Fukuda, Jap. J Appl. Phys. 26, 11944 (1987).
4. P.L. Gourley, I.J. Fritz and L.R. Dawson, Appl. Phys. Lett. 52, 377 (1988).
5. W.A. Jesser and Jan H. van der Merwe, in Dislocations in Solids edited by F.R.N. Nabarro (North Holland, Amsterdam, 1989) to be published.
6. Jan H. van der Merwe and W.A. Jesser, J. Appl. Phys. (1988) to be published.
7. W.A. Jesser, Phys. Status Solidi (a) 20, 63 (1973).
8. J.H. van der Merwe, Philos Mag. A45, 145 (1982).
9. M.W.H. Braun, Ph.D. Thesis (Pretoria University, 1987).
10. D.L. Tõnsing, S.A. J. Sci. (1988) to be published.
11. J.W. Matthews, in Epitaxial Growth, edited by J.W. Matthews (Academic Press, New York 1975), chapter 8; in Dislocations in Solids, edited by F.R.N. Nabarro, (North Holland, Amsterdam 1979) chapt. 7.
12. W.A. Jesser and J.H. van der Merwe, J. Appl. Phys. 63, 1509 (1988).
13. C.A.B. Ball and J.H. van der Merwe, in Dislocations in Solids, edited by F.R.N. Nabarro (North Holland, Amsterdam 1983) chap. 27.
14. T.T.Tsong, Rep. Prog. Phys. 51, 759 (1988).
15. B.W. Dodson, Phys. Rev. B 35, 2795, 5558 (1987).
16. J.Y. Tsao, B.W. Dodson, S.T. Picraux and D.M. Cornelison, Phys. Rev. Lett. 59, 2455 (1987).
17. B.W. Dodson and J.Y. Tsao, Appl. Phys. Lett. 51, 1325 (1987).
18. E. Bauer, Z. Kristallogr. 110, 423 (1958).
19. E. Bauer and J.H. van der Merwe, Phys. Rev. B 33, 3657 (1986).
20. J.H. van der Merwe and M.W.H. Braun, in MRS Symposium Proceedings Vol. 77, edited by J.D. Dow and I.K. Schuller (MRS, Pittsburgh 1987) p. 133.
21. F.C. Frank and J.H. van der Merwe, Proc. Roy. Soc. Ser. A 198, 205 (1949).
22. J.H. van der Merwe, Proc. Phys. Soc. London A 63, 616 (1950).
23. J. Frenkel and T. Kontorowa, Phys. Z. Sowjetunion 13, 1 (1938).
24. P.M. Stoop and J.A. Snyman, Thin Solid Films 158, 151 (1988).
25. J.H. van der Merwe, J. Appl. Phys. 34, 123 (1963).
26. J.H. van der Merwe and W.A. Jesser, J. Appl. Phys. 63, 1509 (1988).
27. A. Milchev and I. Markov, Surface Sci. 136, 503, 519 (1984); 145, 313 (1984).
28. R. Bruinsma and A Zangwill, J. Physique 47, 2055 (1986).
29. K. Yang, H. Homma and I.K. Schuller, J. Appl. Phys. 63, 4066 (1988).
30. W.A. Jesser and D. Kuhlmann-Wilsdorf, Phys. Stat. Sol. 19, 95 (1967).

31. J.H. van der Merwe, J. Appl. Phys. 41, 425 (1970).
32. L.A. Bruce and H. Jaeger, Philos. Mag. A 38, 223 (1978).
33. J.H. van der Merwe, Philos. Mag. 7, 1433 (1962).
34. N. Cabrera, Surface Sci. 2, 320 (1964); Mem. Scient. Revue Metall. 62, 205 (1965).
35. J.W. Matthews, Philos. Mag. 6, 1347 (1961).
36. J.P. Hirth and A.G. Evans, J. Appl. Phys. 60, 2372 (1986).
37. G.B. Bokii and G.F. Kuznetsov, Sov. Phys. Dokl. 29, 1048 (1984).
38. S. Sharan, K. Jagannadham and J. Narayan, in MRS Symposium Proceedings, Vol. 91, Heteroepitaxy on Si, edited by J.C.C. Fan, J.M. Phillips and B.Y. Tsaur (MRS, Pittsburgh, 1987) p. 311.
39. R. Fischer, H. Morkoc, D.A. Neumann, H. Zabel, C. Choi, N. Otsuka, M. Longerbone and L.P. Erickson, J. Appl. Phys. 60, 1640 (1986).

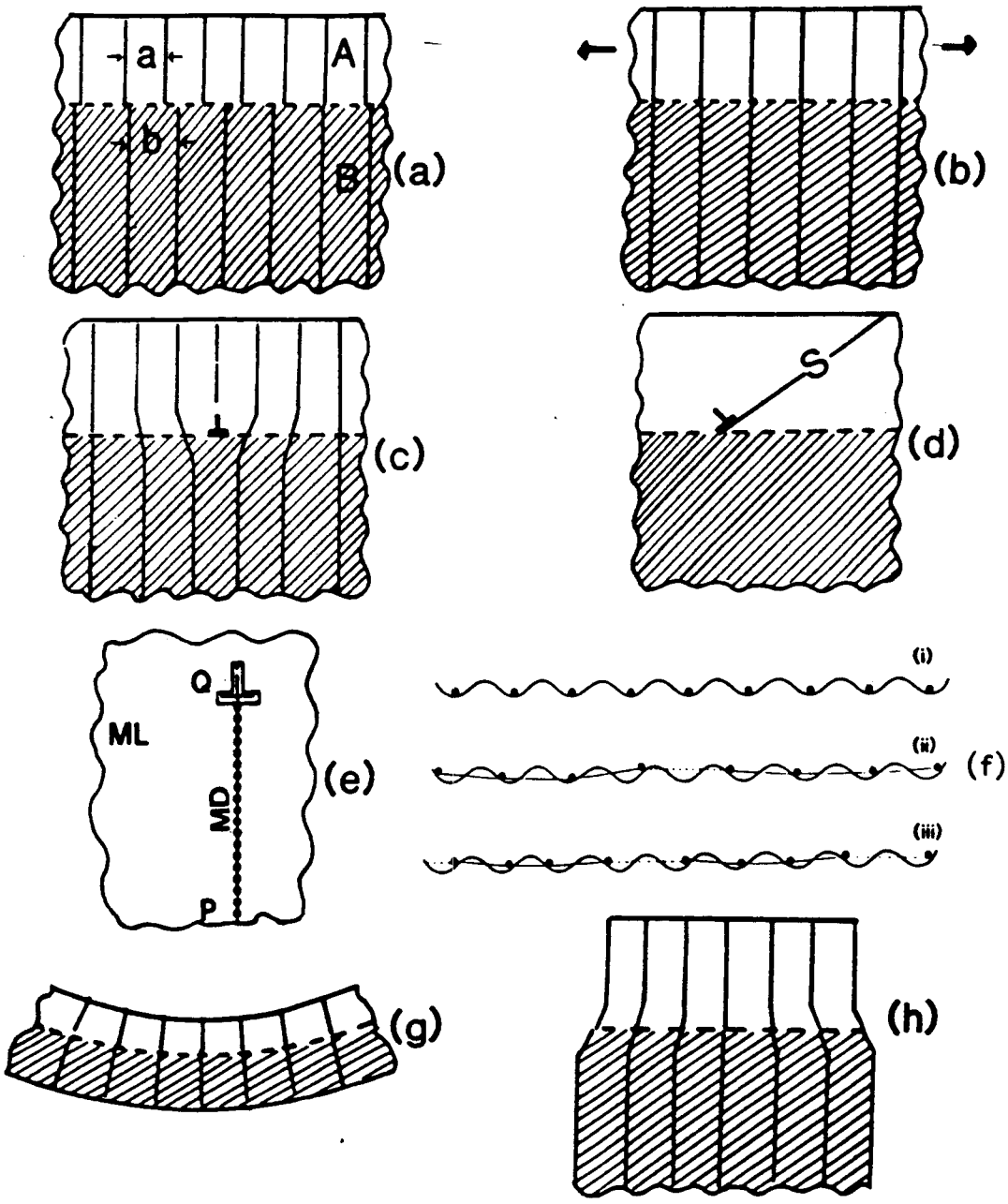

Fig. 1. Misfit accommodation modes: (a) misfit vernier, (b) misfit strain, (c) efficient MD, (d) inefficient MD, (e) misfit- Taylor dislocation i.e. MD PQ (between a monolayer ML and a substrate) ending in a Taylor dislocation at Q within the ML, (f): (i) a coherent chain of particles, (ii) a chain with a>b containing "ruptures" (broken segments), (iii) a chain with a<b containing "distortions", (g) misfit curvature and (h) a misfit strain gradient.

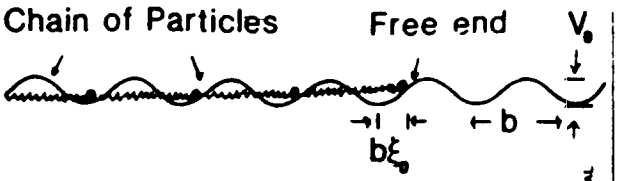

Fig. 2. 1D Frenkel- Kontorowa (FK) model: A chain of particles (adatoms) connected with elastic springs (harmonic forces) of natural length a and force constant μ in adatom- substrate periodic potential $V(\xi)$ of overall amplitude V_0.

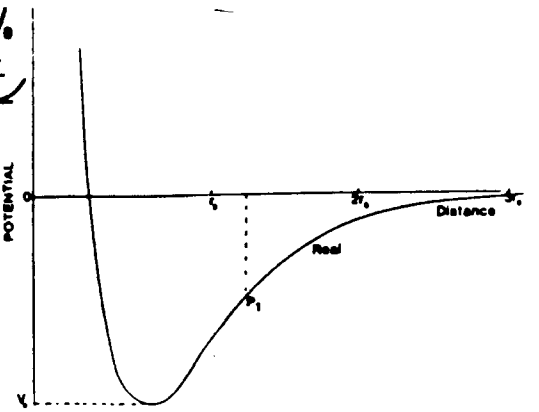

Fig. 3(a)

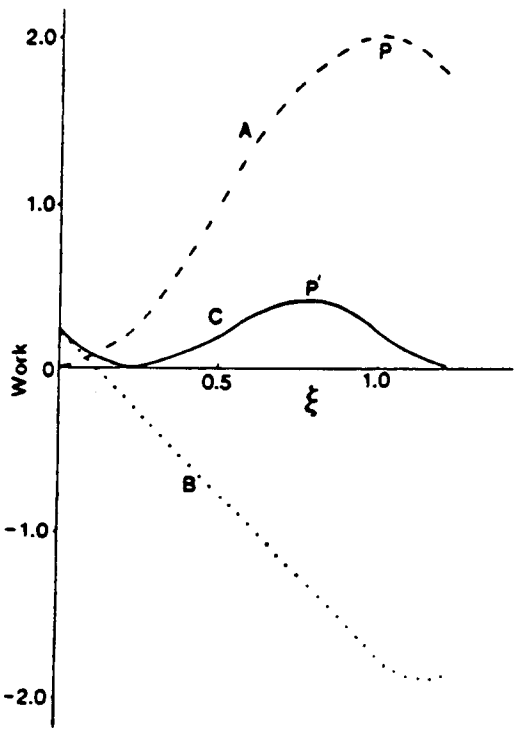

Fig. 4. Curves displaying the work (units $4V_0\ell_0/\pi$) done in generating a MD by displacing the endpoint of the chain, starting at the displacement $a\xi_0$ of stable equilibrium. Curve A represents the primary barrier (defined in (Eq. (27)), B the reduction by MS and C the nett work $U(\xi,f_c)$ with the activation barrier $U_{ac}(f_c) = U(1-\xi_0,f_c)$ when the displacement $a(1-\xi_0)$ of unstable equilibrium is reached.

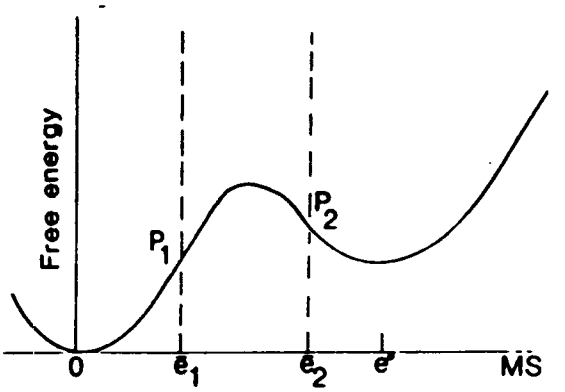

Fig. 3(b)

Fig. 3. (a) Curve of atomic pair potential (in arbitrary units) vs distance r between members of the pair, displaying the equilibrium distance r_0 and point of inflection at P_1.
(b) Curve of the free energy (in arbitrary units) vs MS $\bar{e}$ showing stable configuration at $\bar{e}=0$, metastable configuration at $\bar{e}^*$ and points of inflection at P_1 and P_2.

KEY WORDS

THE PHYSICAL FOUNDATIONS OF CRITICAL PARAMETER CALCULATIONS IN EPITAXY

Jan H. van der Merwe and W.A. Jesser

p.1	epitaxy, coherency- incoherency (C- IC) transition
p.2	misfit accommodation (MA), misfit dislocation (MD), misfit strain (MS)
p.3	activation barrier, growth mode
p.4	Frenkel- Kontorowa (FK), Fourier truncation, Harmonic approximation
p.5	Volterra, energy of MS and MD
p.6	criteria for C- IC transition, equilibrium, Peach- Koehler
p.7	activation barrier, critical misfit
p.8	one- dimensional and two- dimensional registry, rupture
p.9	parabolic approximation, superlattice
p.10	force criterion, non- equilibrium
p.11	nucleation barrier, excess stress (force)
p.12	misfit strain relief, tilted substrate
p.13	semiconductors, ab initio

CHAPTER 41

The Prediction of Critical Misfit and Thickness in Epitaxy

W. A. JESSER* and J. H. van der MERWE

Department of Physics
University of Pretoria
South Africa

*Permanent address: Department of Materials Science, University of Virginia, Charlottesville, VA, USA.

Elsevier Science Publishers B.V., 1989

Dislocations in Solids
Edited by F. R. N. Nabarro

Contents

1. Introduction
   - 1.1. Misfit
   - 1.2. Modes of misfit accommodation
   - 1.3. Goals
   - 1.4. Governing physical principles
   - 1.5. Growth modes
   - 1.6. Models
   - 1.7. Criteria
2. Calculations: Equilibrium
   - 2.1. Energy criteria
   - 2.2. Stress and force criteria
   - 2.3. Geometrical criteria
3. Calculations: Non-equilibrium
   - 3.1. Monolayer approach
   - 3.2. Nucleation of misfit dislocation loops
   - 3.3. Overcoming the Peierls barrier
   - 3.4. Overcoming the surface barrier
   - 3.5. Quasi-nucleation approach
   - 3.6. Misfit strain gradient
   - 3.7. Strain relief by crack formation
   - 3.8. Dynamics of strain relief
4. Correlation with experiment
5. Discussion and conclusions

Abbreviations

AAI adatom–adatom interaction
ASI adatom–substrate interaction
C-IC commensurate–incommensurate, coherent–incoherent
FK Frenkel–Kontorova
IMD inefficient (imperfect) misfit dislocation
KS Kurdjumov–Sachs
NW Nishiyama–Wassermann
MA, MC, MD, MF, MS, MSG and MV *Misfit* accommodation, curvature, dislocation, fracture, strain, strain gradient and vernier
ML monolayer
TD threading dislocation
VM Volterra model

1. Introduction

1.1. Misfit

The epitaxial growth of one crystal on another has been studied as a fundamental problem for over six decades. Only in the past two decades, however, has there existed a practical interest in understanding the phenomena more completely. This recent practical interest has arisen out of the great need of the semiconductor industry for crystals free from dislocations and other defects. As the use of electronic devices became more demanding and sophisticated, the need for perfection of the crystals comprising the devices has increased. This is particularly true for superlattice structures.

Central to the growth of nearly perfect epitaxial overlayers is the concept of a coherent or dislocation free interface. The thickness above which dislocations appear in the interface during the growth of an epitaxial layer is known as the critical thickness. Because of the importance of this parameter to the fabrication of acceptable semiconductor devices, classification of the various approaches to its calculation is desirable. Before discussing the calculations of critical thickness, it is necessary to define the nature of the misfit associated with the interface between two different crystals.

The general case of fitting together two crystals of different lattices in an oriented way across an arbitrary but flat interface, was dealt with theoretically by Pond [1] and by Bilby et al. [2]. We follow the latter authors who introduced a geometrical closure failure $\boldsymbol{B}$ to a Burgers circuit enclosing a section of the

interface. The vector $\boldsymbol{B}$ varies with direction of the interface and may be written in terms of a reference vector V lying in the interface as

$$\boldsymbol{B} = [S_1^{-1} - S_2^{-1}]V , \tag{1}$$

where S_1 and S_2 are the "deformation" matrices that generate respectively the lattices of crystal 1 and crystal 2 from a reference lattice in which $\boldsymbol{B}$ and $\boldsymbol{V}$ are defined. The general case that includes crystal symmetry and coincidence lattices can be incorporated in the use of eq. (1) to express the deviations from coincidence and symmetry positions. In this chapter we are concerned with cases in which the crystals are simply related to one another through the matrix S of small deformations. In this case the closure failure is represented by a set of dislocations of sufficient separation to be considered individually, When $\boldsymbol{V}$ cuts m_i dislocation axes of Burgers vectors $\boldsymbol{b}_i$ then

$$\boldsymbol{B} = \sum_i m_i \boldsymbol{b}_i , \tag{2}$$

and the average spacing between like dislocations of given type i measured along $\boldsymbol{V}$ is Vm_i^{-1}. It is convenient to express the closure failure in terms of three physically distinguishable components [3]: B_t is the component normal to the interface and represents the component of $\boldsymbol{B}$ produced by tilt around $\hat{V}$, where $\hat{V}$ is the unit vector parallel to V; B_s is in the plane of the interface and normal to $\boldsymbol{V}$ and represents the twist or shear component; and lastly B_f is parallel to $\boldsymbol{V}$. The component B_f is the misfit component which is of interest to the present chapter, and it is this component that provides a definition of misfit f as measured along the direction $\hat{V}$ of the interface as

$$f = \frac{B_f}{V} = \frac{\boldsymbol{B} \cdot \hat{V}}{V} . \tag{3}$$

When the deformations S are small and contain no rotation terms a pure misfit boundary between two parallel orthorhombic crystals results and typically suggests one of two choices for the reference lattice. When the two crystals across the interface are of comparable thickness, or are both "thick", then the choice of a reference lattice that has lattice constants between those of either crystal is appropriate. According to a simplified treatment of the closure failure [3] one of the reference lattice constants, say A is chosen to be

$$c = a_1 a_2 / \bar{a}_2 , \tag{4a}$$

where a_1 and a_2 are the lattice constants in a chosen principal direction $\hat{a}$ of the respective crystals 1 and 2 and

$$\bar{a} = \tfrac{1}{2}(a_1 + a_2) \tag{4b}$$

is the average lattice constant. This choice of reference lattice constant is also the one adopted by van der Merwe in his model for thick crystals [4]. Equation (3) in conjunction with eqs. (1) and (4a) and (4b) yields for this case

$$f \equiv f_a = (a_2 - a_1)/c: \qquad \text{similar thickness} , \tag{4c}$$

as the misfit in the $\hat{a}$-direction of the lattices.

The second choice of reference lattice is appropriate to a thin crystal on a thick substrate. In this case it is natural to take the substrate as the reference crystal. If we call the reference crystal 1 then $S_1^{-1} = I$, the identity matrix, and hence eq. (1) becomes

$$\boldsymbol{B} = [I - S_2^{-1}]V \,, \tag{5a}$$

which is equivalent to the approach of Bollmann [5]. The misfit now, is by eqs. (3) and (5a),

$$f_a = (a_2 - a_1)/a_2 : \qquad \text{thin overlayer} \,. \tag{5b}$$

The advantages of definitions (4c) and (5b) have been discussed by Matthews [6].

Epitaxy of crystals with different symmetries is also of interest. A case that had been considered extensively is the epitaxy of metals at (111) fcc/(110) bcc interfaces [7, 8] in which two dominant epitaxial orientations (the Kurdjumov–Sachs and the Nishiyama–Wassermann orientations) differing by about 5° occur depending on the ratio of nearest-neighbor distances. The two misfits f_1 and f_2, of the form (4c) or (5b), are usually defined for two suitable orthogonal interfacial directions, f_1 and f_2 being significantly different.

degree symbol round 0

1.2. Modes of misfit accommodation

The above considerations are purely geometrical. The special response of the near interface atoms to the fields AA and BB within each of two misfitting epitaxial crystals A and B and the coupling field AB between them is normally referred to as "misfit accommodation" (MA). In the foregoing we have tacitly assumed that the misfit f is accommodated by dislocations, so-called misfit dislocations (MD's). However, this is not the only mode of MA. The resulting mode in any given case depends largely on the magnitude and nature of f as well as those of the fields. The possibilities considered are illustrated in fig. 1.

Consider a pure misfit boundary, i.e. S_1 and S_2 are diagonal matrices that represent small dimensional changes, not rotations. There are numerous modes by which the misfit of this interface can be accommodated. These mechanisms can operate individually as a single mode or in concert with several modes participating. Homogeneous elastic strains that reduce the effective misfit from f to zero (fig. 1b), i.e. $\boldsymbol{B} \rightarrow 0$, have first been observed in metals [9–13] as reviewed by Matthews [14] and are now commonly observed in strained layer superlattices of semiconductor crystals [15]. This strain is referred to as the misfit strain (MS) [16–17], and interfaces for which the misfit is accommodated entirely by MS are said to be pseudomorphic or coherent [18]. Similarly, free-standing crystals may bend with neutral planes that are positioned so that the misfit at the interface is accommodated by the lattice parameter gradient introduced through lattice curvature [19–22]. We can refer to this mechanism (fig. 1c) as misfit curvature (MC). A third strain mechanism for accommodating the misfit also exists. It is a simple gradient of strain (misfit strain gradient: MSG) that takes a coherent interface from its constrained lattice parameter at the interface to its bulk lattice

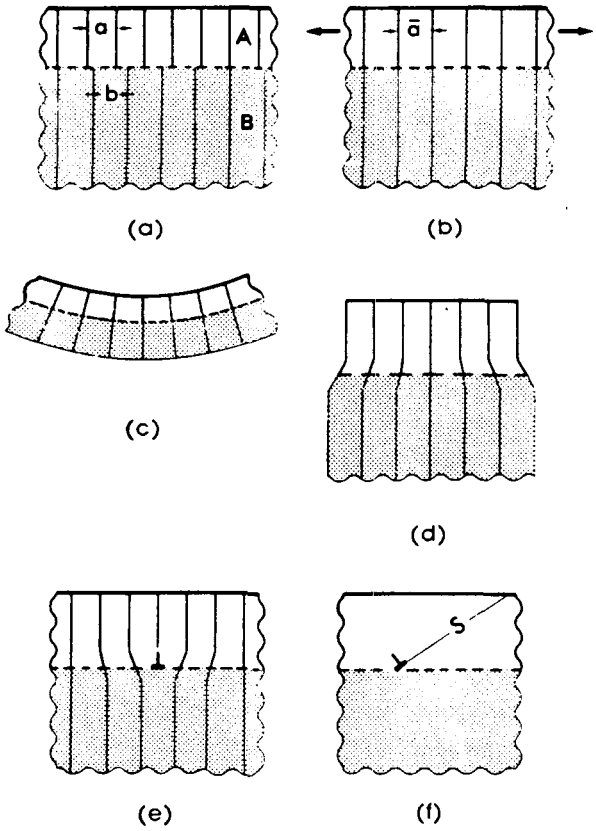

Fig. 1. Diagrammatic illustration of modes of Misfit Accommodation (MA) in a bicrystal of crystals A and B with natural lattice parameters a and b. The full lines designate atomic planes and the dashed lines interfaces, all viewed edge on: (a) Crystals as though rigid, or with vanishing interfacial interaction, or both; Misfit Vernier (MV) mode of MA. (b) Crystal A homogeneously strained ($a \rightarrow \bar{a}$) into registry ($\bar{a} = b$) with B; Misfit Strain (MS) mode of MA. (c) Crystals bent into registry at interface; Misfit Curvature (MC) mode of MA. (d) Misfit accommodated by a Misfit Strain Gradient (MSG) within a finite thickness. (e) Crystals elastically relaxed so that both the residual interfacial misfit and the strain are very much localized around regularly spaced (dislocation) lines; Misfit Dislocation (MD) model of MA. (f) A MD formed from a glide dislocation on a slip plane S has its Burgers vector inclined to the interface and constitutes an Inefficient Misfit Dislocation (IMD).

parameter at the free surface [23] as in fig. 1d. This can occur when the height to width ratio for a deposit is comparable to the mutual misfit. More precisely, since the MSG is not expected to be linear but to decay rather faster than linearly with distance normal to the interface, there is an effective thickness above which the lattice parameter is essentially that of the bulk. It is this feature which means in practice that there is no critical thickness (or the critical thickness is infinite), but there is a critical misfit which decreases with increasing width. Heterostructures involved in ultralarge scale integration can take advantage of this misfit accommodation mode. Recent evidence of misfit dislocations between the reconstructed

topmost surface of a small gold platelet and its bulk underlying layers [24, 25] suggests that the MSG mode of accommodating misfit may not always be energetically favored.

When the interfacial bond is weak and the misfit large [8], the structure of the interface (fig. 1a) resembles the rigid misfit vernier (MV). This mode of misfit accommodation or one approximating it has been observed in some metal–non metal epitaxial systems [26].

An uncommon mode of misfit accommodation is fracture (misfit fracture: MF). If the interfacial bond is strong and the epitaxial layer brittle, then a tensile misfit strain can be relieved by a crack on the nearest cleavage plane that is normal to the interface. Epitaxial garnet films show this mode of misfit accommodation [27].

The most common mode of accommodating misfit is by misfit dislocations (MD's; fig. 1e). A MD is defined as an interfacial dislocation that generates the transition between the two lattices and is said to accommodate the misfit f. Essentially it replaces the long-range stresses due to the MS by the oscillatory strains of the MD's to reduce the overall energy. There are several choices of MD arrays that will accommodate the same misfit. Minimum energy considerations coupled with barriers to MD generation determine which of the available choices will be realized. The most common MD's have an edge component of the Burgers vector $\boldsymbol{b}$ that projects along the direction of interfacial misfit. The most efficient MD has its Burgers vector in the interface and is therefore one of pure edge character. It is referred to as an efficient or perfect MD, and was first observed by TEM in PbS/PbSe bicrystals [28]. It is geometrically possible to accommodate a pure misfit boundary of (001) orientations between orthorhombic crystals with a crossed grid of screw dislocations [29] if the crystals are rotated so that the a-axis of the one is parallel to the b-axis of the other and the misfits in the two directions are therefore equal but of opposite sign. This same misfit boundary can also be accommodated by a crossed grid of perfect MD's. While the above considerations show that MD's are not restricted to any particular type of dislocation one usually expects MD's to have edge character.

The first experimental evidence for misfit accommodation by MD's was obtained from etch-pit analyses of germanium crystals with small gradients in chemical composition [30]. Examples of misfit accommodation by only inefficient (or imperfect) misfit dislocations (IMD) (Burgers vectors lying out of the interface as shown in (fig. 1f) can also be found [31]. It is also possible to have both inefficient and pure misfit dislocations present concurrently in the same interface [32]. Indeed, accommodation of misfit by dislocation half-loops and inefficient dislocations that have glided into the interface has been documented [33–35]. Clearly then, the mechanisms available for the introduction of the various types of MD's is of significant practical importance. It is now commonly observed that MD's occur in most epitaxial systems, as has been reviewed by Matthews [6, 14].

The last mechanism for accommodating misfit to be discussed here is that of ledges. Often substrates are oriented off major crystallographic planes so that steps in the surface are available for nucleation sites. Further, the appearance of

growth steps and ledges at interphase boundaries is common [36]. The occurrence of stepped interfaces to improve the matching of interphase boundaries with misfit has been known for some time [37–40]. The suggestion that steps associated with terraces may accommodate misfit at an epitaxial boundary has also been made [41]. It is only recently that the energy of a misfit boundary with ledges has been calculated [42]. A planar misfit boundary may improve the overall interfacial matching if a sequence of steps, that effects a relative displacement of the atomic patterns on either side of the interface so as to insure that disregistry is never greater than about one quarter of an atomic spacing, is introduced [42, 43]. A detailed treatment of this problem is in progress [44].

1.3. Goals

The main goal of the present paper concerns the accommodation of misfit by MD's for the case of a pure, flat misfit boundary. It is the MD's that are the center of attention because of their importance to electronic device performance and reliability. If MD's and MS($\bar{e}$) coexist, the portion of the misfit f accommodated by MD's being $\bar{f}$, one may write for each interfacial direction of a thin epilayer A on a thick substrate B (to first order) [14]

$$f = \bar{f} + \bar{e}\,; \quad \bar{e} \equiv |\bar{e}| \tag{6a}$$

the corresponding quantities in the two directions being equal when the interfacial symmetries are quadratic, but otherwise unequal. In the Frank and van der Merwe formalism [46]

$$f = \frac{a-b}{b}\,, \qquad \bar{f} = \frac{\bar{a}-b}{b}\,, \qquad \bar{e} = \frac{\bar{a}-a}{a}\,, \tag{6b}$$

where a and b are the normal atomic spacings in A and B and $\bar{a}$ is the misfit strained spacing in A; B, being thick, is unstrained. When both crystals are thin provision must be made for MS in both crystals. Furthermore, coherence may exist in one direction and disregistry with MD's or a MV in a perpendicular direction [8]. An implicit assumption, also in eq. (6a), is that the temperature remains fixed at the growth temperature, or the half crystals A and B have the same thermal expansion coefficients. Otherwise the equilibrium MS $\bar{e}$ may decrease, or increase or even change sign. Technologically this is undesirable and provision must be made to fit both crystallographical misfit and thermal expansion coefficient [45].

The goal has thus been reduced to describing existing considerations on the conditions that determine the distribution of f between $\bar{f}$ and $\bar{e}$. Here it suffices to report [46] that under certain conditions there exists a critical misfit $f_c^{(1)}$ so that a monolayer (ML) with $f < f_c^{(1)}$ grows coherently ($\bar{f} = 0$) while MD's are present when $f > f_c^{(1)}$. If for $f < f_c^{(1)}$ more ML's are added a critical thickness h_c is reached above which there is a transition to MA with MS and MD's jointly. Conversely for given $h = h_c$ the misfit f is critical, i.e. for $f > f(h_c) \equiv f_c(h)$ there will be a

transition from coherency (registry, commensurate configuration) to incoherency (disregistry, incommensurate configuration). The goals of this chapter are to review and appraise existing calculations of f_c and h_c with special emphasis on their dependence on the physical properties of the bicrystals concerned and to their correlation with experimental data – particularly the influence of the experimental conditions on this correlation.

1.4. Governing physical principles

The systems (vapors, crystals) and processes (adsorption, desorption, adatom diffusion, MD generation and motion) involved in epitaxy fall respectively in the categories of thermodynamic systems and rate processes. Therefore at an intermediate stage any internal change under constant external constraints brings the system nearer to the equilibrium (stable) minimum free energy (A) configuration. The process is driven by free energy gradients $-\Delta A$ and progress is facilitated by thermal (temperature T) fluctuations to overcome a hierarchy of energy barriers U (e.g. adatom migration activation energy, and energy barriers towards nucleation and motion (Peierls) of dislocations) separating consecutive metastable configurations. Thus the closeness of an observed configuration to equilibrium or the degree to which it had advanced towards equilibrium, will otherwise depend on the deposition rate (supply rate of adsorbate).

Since the analysis of equilibrium configurations on the basis of minimum free energy, i.e.

$$A = \text{minimum} \qquad \text{(equilibrium)}, \quad (7)$$

is so much easier than the dynamics of non-equilibrium processes, most theoretical predictions of the past were equilibrium based, the assumption being that in many practical cases sufficient time had been available for the system to come close to equilibrium. The perception is that knowledge of equilibrium configurations will also provide useful guidelines of tendencies. Additional simplification had been introduced by noting that, because of the regularity of the crystals and of the MD arrangements adopted, the energy (E) contribution to A is dominant and hence that the equilibrium configurations may be adequately characterized by the condition that [6, 8, 46]

$$E = \text{minimum} \qquad \text{(equilibrium)}, \quad (8)$$

Clearly the applicability of the theory requires the barriers to be small enough as compared to thermal energies.

In the problem of epitaxy the scale of E, which has often been referred to as the interfacial energy, is primarily determined by the strength of the AB interfacial interaction and secondarily by the misfit. The primary goal has therefore been to find the dependence of E on these parameters. The basic assumption is then that the structure (mode of MA) is determined by the minima of E with respect to these parameters.

1.5. Growth modes

Of great importance to the consideration of critical size, e.g. critical thickness, are the epitaxial growth modes that have been named by Bauer [47]: Frank and van der Merwe (ML-by-ML, 2D) growth, Volmer–Weber (island, 3D) growth and Stranski–Krastanov (island on top of a few ML's in 2D mode). Also the growth mode is partly controlled by the prevailing degree of non-equilibrium as expressed in terms of supersaturation. The supersaturation increases with decreasing substrate temperature T_s and increasing deposition rate. The equilibrium criterion for the growth mode can be expressed in terms of the specific surface (γ_A, γ_B) and interface (γ_{AB}) free energies [47].or bondings [48] per unit area E_{AA} and E_{AB} as

$$\begin{aligned} \Delta\gamma_{AB} \equiv \gamma_A + \gamma_{AB} - \gamma_B \leqslant 0 \quad \text{or} \quad E_{AA} - E_{AB} \leqslant 0 \quad &\text{for 2D growth} \\ > 0 \qquad\qquad\qquad\quad > 0 \quad &\text{for 3D growth}, \end{aligned} \tag{9}$$

where B is the substrate and A the adsorbate. For growth of the $(n+1)$-st ML on the n-th ML, $\Delta\gamma_{AB}$ must be replaced by $\Delta\gamma_{AA} \sim 0$. However, $\Delta\gamma_{AA}$ may even become effectively positive due to influence of the nearby substrate. This is an important reason for Stranski–Krastanov growth [49]. Equations (9) are simply saying that when AB bonding is strong compared to AA bonding the tendency will be to form AB rather than AA bonds, i.e. lateral rather than vertical growth.

The formation of smooth and planar interfaces and of uniform epilayers are highly desirable, particularly in superlattices consisting of thin alternating A and B layers, and is facilitated by 2D growth [48, 49]. In this case one half of the growth cycle will inevitably be 3D-wise at equilibrium. 3D growth can be modified to occur 2D-like by having the appropriate supersaturation. The smoothness of the interfaces will furthermore be facilitated by a choice of materials for which the degree of three-dimensionality is low [48] as may approximately be expressed in terms of a surface free energy "mismatch" [49]. Since most practical applications require smooth planar interfaces and uniform thicknesses, the fabrication processes are tailored to achieve these goals. This lends more credibility towards the usual practice of predicting critical misfits assuming layerlike growth.

1.6. Models

Almost all the earlier theories modeled the interaction (AA and BB) within the crystals in terms of continuum elasticity (harmonic approximation) theory [4, 6, 46]. Subsequently anharmonicity has been introduced [50, 51], and more recently ab initio atomic calculations, using appropriate potentials, have also been carried out [52].

The earlier theories adopted essentially two approaches for modeling the interfacial (AB) interaction: the Volterra continuum model in which A and B interfacial atoms undergo relative displacements parallel to the interface that are limited to integer values of the (reference) lattice parameter, and the Frenkel–

Kontorova (FK) model [53] that views the interfacial atoms of A to be in a (sinusoidal) periodic field of crystal B and vice versa. The Volterra model (VM) which has been extensively and profitably employed by Matthews [6, 14] has the advantage of versatility and mathematical simplicity over all regimes of thickness. Its main disadvantages are (i) its inaccuracy at small thicknesses and (ii) that it does not allow for the widely varying interfacial (AB) bond strengths which are important for thin epicrystals. The FK-model on the other hand makes appropriate allowance for AB interaction and has fairly simple analytical solutions for the extreme cases: (a) a ML on a thick substrate for which the governing equations reduce to sine–Gordon types [46, 53–55] and (b) two thick crystals [4, 56], but has to resort to approximations for intermediate cases except when the periodic AB interaction is represented by parabolic arcs [56–57].

The model introduced by Frenkel and Kontorova is one-dimensional and was first applied to epitaxy by Frank and van der Merwe [46] in their classic paper concerning misfitting monolayers on thick (essentially rigid) substrates. It was subsequently extended to two interfacial directions [8, 46, 54, 55]. As in the FK-model the complexity of the analysis was limited by truncating the corresponding two-dimensional Fourier series of the periodic interfacial (AB) potential at low harmonic order, a practice which has now been properly justified [58]. This extended model has been used to show [8] (i) that minima in the (free) energy of an epitaxial bicrystal occur when the interfacial structure and dimensions, and the relative crystal orientations, are such that parallel interfacial atomic rows of the two crystals are equally spaced, irrespective of whether in the rows the atoms are in registry, (ii) that the depth of the free-energy minima may be directly correlated with the values of specific Fourier coefficients and (iii) that the depths are additive should matching also occur in a perpendicular direction to achieve complete 2D registry. Thus, while the values of the Fourier coefficients are indicative of the maximal energies that can be gained (the tendency to epitaxy) from row matching in one or two dimensions, the matching itself predicts precisely for given crystal structures the orientations for which epitaxy is observed. Naturally the net gain will be reduced by strain energy if the matching is achieved by MS.

The difficult part of the relevant analysis is to find expressions for the self energies (line tension) E_D per unit length of a MD. In the FK model and its generalizations the analytical techniques have been different for a ML on a thick substrate and for two thick crystals and approximate otherwise. Because of its frequent use it is convenient for later reference to give the expression of E_D for two thick epicrystals [4, 56]:

$$E_D = \frac{\mu_{ab} c p}{4\pi^2} \left[1 + \beta - (1+\beta^2)^{1/2} - \beta \ln\{2\beta(1+\beta^2)^{1/2} - 2\beta^2\}\right], \tag{10a}$$

$$\beta = \frac{2\pi\lambda_0 c}{\mu_{ab} p}, \qquad \frac{1}{\lambda_0} = \frac{1-\nu_a}{\mu_a} + \frac{1-\nu_b}{\mu_b}, \tag{10b}$$

where μ_{ab}, μ_a and μ_b are respectively shear moduli for the interface and crystals

A and B, ν_a and ν_b Poisson's ratios, $c \approx \frac{1}{2}(a + b)$ is a reference lattice parameter and p the MD spacing. When p is large ($c/p \ll 1$) eq. (10a) becomes

$$E_D \approx \frac{\lambda_0 c^2}{2\pi}\left\{\ln\left(\frac{\mu_{ab} p}{4\pi\lambda_0 c}\right) + 1\right\}. \tag{10c}$$

In the Volterra model the self energy per unit length of a MD between an epilayer of thickness h_a and a thick substrate is given by [6, 14]

$$E_D = Db\left\{\ln\left(\frac{R}{b}\right) + 1\right\}, \qquad D = \lambda_0 b/2\pi, \tag{11a}$$

where R is the "cut-off radius" usually taken to be

$$R = \begin{cases} p/2 & \text{when } h_a > p/2, \\ h_a & \text{when } h_a \leqslant p/2, \end{cases} \tag{11b}$$

and the Burgers vector b is essentially the c in eqs. (10). We note the close similarity between (10c) and (11a) with R replaced by $\mu_{ab} p/4\pi\lambda_0 \sim p/10$ for crystals with similar atomic interactions instead of $p/2$ as in (11b).

This expression for E_D can be adopted for the line energy of a MD or used to calculate the energy per unit area,

$$\begin{aligned} E_D &= p_x^{-1} E_D^x + p_y^{-1} E_D^y \\ &= 2E_D/p \text{ for quadratic symmetry}, \end{aligned} \tag{12a}$$

of an interface for a network of MD's assuming that the crossing energy is negligible [6, 14]. When the misfit is partly accommodated by MS($\bar{e}$) the MS energy, which is of the form [6, 54, 59]

$$E_{\bar{e}} = Bh_a\bar{e}_a^2 = \frac{2\mu_a(1+\nu_a)}{1-\nu_a} h_a\bar{e}_a^2 \quad \text{per unit area}, \tag{12b}$$

for interfaces with quadratic symmetry, must be included. For this symmetry the principal strains $\bar{e}$ are equal and the shear strain vanishes. When both crystals are thin each will be misfit strained and contains MS energy of the form (12b) [6, 57, 60]. In any case the total energy is

$$E = E_D + E_{\bar{e}} \tag{12c}$$

where (12a) and (12b) are special cases of E_D and $E_{\bar{e}}$ respectively.

1.7. Criteria

Various criteria had been used in the past to calculate the critical misfit f_c and critical thickness h_c. These have recently been analysed in some detail [61]. It is therefore necessary that we understand the basis of each and put them into perspective. We begin by considering equilibrium criteria. For this purpose we may display the functional dependence of the total energy of a thin epilayer A of

uniform thickness h on a thick substrate as $E(h, f; \bar{f}$ or $\bar{e})$. We may further minimize E by varying the MS($\bar{e}$) (or equivalently the MD density $\bar{f}$ in eqs. (6)) at constant h and f, obtaining from eq. (8) the governing equations [6, 46]

$$0 = \frac{\partial E}{\partial \bar{e}} = \frac{\partial E}{\partial \bar{f}}, \qquad \text{for } \bar{e} = \bar{e}_m \quad \text{and} \quad f = \bar{f}_m . \tag{13a}$$

This defines the equilibrium values $\bar{e}_m$ and $\bar{f}_m$ of $\bar{e}$ and $\bar{f}$.

The form (13a) of the equilibrium condition clearly implies extensive interfaces with many MD's so that it is meaningful to speak of varying $\bar{f}$ continuously rather than discretely. The second part of eq. (13a) lends itself to an interpretation that not only overcomes this difficulty but in some cases also leads to a drastic simplification of the analysis: if E does not change with $\bar{f}$, it is saying that the work W done to create the MD vanishes, i.e.

$$W = \int F \, ds = 0 \qquad \text{for } \bar{e} = \bar{e}_m \quad \text{and} \quad \bar{f} = \bar{f}_m ; \tag{13b}$$

for $\bar{e} \lesseqgtr \bar{e}_m$ the work $W \lesseqgtr 0$ respectively. In eq. (13b) F is an external force needed to generate the dislocation and ds is the surface displacement, *or* F is the integrated force on an existing dislocation (extending from the interface on the free surface or other interface) from which the interfacial MD is created (drawn out along the interface) and ds is an infinitesimal displacement of the (existing) dislocation line. Both F and ds are measured in the direction of increased length of MD line laid along the intersection of the interface and the slip plane of the "threading" segment of the crystal dislocation in the strained epilayer. It is the glide of this threading dislocation (TD) that creates increased length of MD line; the glide direction and direction of "increased" MD length are the same. When F is constant (independent of the displacements) the equilibrium equation (13b) reduces to [6, 14]

$$0 = F = F_{\bar{e}} - F_D , \tag{13c}$$

where $F_{\bar{e}}$ is the MS-induced Peach–Koehler force on the TD and F_D the resistance due to the MD line tension. F may also be an external force needed to inject a MD and ds correspondingly a displacement of the point(s) of application of F as in refs. [46, 54, 55].

The problems associated with the forces between MD's in an interface of finite width between two cylindrical half crystals, their interplay with misfit and MS, and the mechanics of consecutive incoming MD's, have been analysed in some detail by Nabarro [62]. These provide useful insights into the behavior of MD's at the onset of coherency breakdown.

Equations (13) are three equivalent criteria (except that (13c) is a special case) for the equilibrium distribution of given misfit f between MD's ($\bar{f}$) and MS ($\bar{e}$) at given thickness h of epilayer. As h decreases, the MS energy for fixed $\bar{e}$ decreases; alternatively the MS $\bar{e}$ for fixed MS energy increases so that the density $\bar{f}$ of MD's, needed to satisfy eq. (2), diminishes until it vanishes ($\bar{f} = 0$) at a critical thickness h_c [6, 14]. Conversely $f_c = f$ is the critical misfit at given thickness $h = h_c$. The

conditions for the critical misfit f_c at given h and the critical thickenss h_c at given f, are respectively [6, 57]

$$\phi(h, f_c; \bar{f} = 0) = 0 \quad \text{and} \quad \phi(f, h_c; \bar{f} = 0) = 0 , \tag{14}$$

where ϕ is one of the functions of eqs. (13).

Consider the introduction of MD's into an initially coherent epilayer. The line tension increases the energy according to E_D in eqs. (10) or (11) and the Peach–Koehler force decreases the MS energy $E_{\bar{e}}$ in eq. (12b). Furthermore, in the FK-model, at equilibrium there are no MD's ($\bar{f} = 0$) below the critical misfit f_c and many – almost enough to accommodate all ($\bar{f} = f$) the misfit – above f_c [46]. On the basis of these considerations it has previously [6] been assumed that the condition for f_c is approximately equivalent to the condition that $f = f_c$ when *all* ($\bar{e} = f$) the MS energy E in eqs. (12b) can be traded for MD energy in eq. (12a) with $\bar{f} = f$ [56]:

$$E_{\bar{e}}(h, f_c; \bar{e} = f) \approx E_D(h, f_c; \bar{f} = f) , \tag{15}$$

instead of a configuration where $\bar{e}$ and $\bar{f}$ coexist. This approximation has subsequently been used by other authors, as dealt with below. Recent calculations [57] have however shown that this approximation is rather inaccurate and may overestimate the critical misfit by as much as 25% and the critical thickness by as much as 35%.

The foregoing considerations apply to the case where the availability of MD's is not an obstacle [6, 46]. Normally the introduction of MD's is subjected to the overcoming of energy barriers [46]

$$U = \int_{\text{bar}} F \, ds . \tag{16}$$

The most prominent barriers are associated with the nucleation of MD and the Peierls stresses opposing their motion [6, 14]. The barrier heights are reduced by the misfit stress (Peach–Koehler forces). When U vanishes, i.e.

$$U = 0 , \tag{17}$$

MD's generate spontaneously without the aid of temperature fluctuations [46].

When $U > 0$ the acquisition of MD's depends upon the thermal energy kT. When $U \gtrsim 30kT$ the barriers are completely prohibitive and the system will remain almost indefinitely in a metastable configuration. When $U < 30kT$ the attainment of equilibrium takes time and the observed configuration may differ significantly from equilibrium.

2. *Calculations: equilibrium*

The criteria employed for calculating critical parameters as defined in eqs. (14)–(17) may be broadly separated into equilibrium and non-equilibrium criteria. The equilibrium category can be subdivided into (i) energy criteria when

the energy values are compared as in eq. (15) or the energy is properly minimized as in eqs. (13), (ii) stress or force criteria as given in eq. (13c) and (iii) geometrical criteria where good "matching" is implied to be equivalent to low energy. The non-equilibrium category, which recently became the subject of intensive study, can be subdivided into calculations concerning (i) nucleation energies of MD's (ii) other energy barriers and (iii) the reduction of the activation energy by MS until it ultimately vanishes as in eq. (17) and MD's generate spontaneously, and (iv) the solution of the rate equation to establish the degree of progress at any time t towards achieving equilibrium. Often diverse approaches are encountered within each of these categories.

2.1. Energy criteria

2.1.1. Precipitates

The first calculation of a crtical size was made by Nabarro for the case of a misfitting sphere in a crystalline matrix [63]. Although this is not really an example of "epitaxy" it is directly related and, being the first, warrants some consideration. Nabarro essentially employed the criterion (15) and equated the elastic strain energy associated with the coherency strains, MS, to the rise in interfacial energy that occurs when coherency is destroyed. The interfacial energy was estimated by the energy required to melt a monolayer at the interface. The same model was employed by Jesser [64] but with a more realistic expression for the interfacial energy. The sum of the misfit strain energy and interfacial energy was minimized as in eq. (13a) to find the critical radius above which coherency loss is energetically favored. In this case the interfacial energy expression of van der Merwe was used. A similar model in which the interfacial energy is replaced by the energy of a single Volterra type prismatic misfit dislocation loop of radius r equal to that of the misfitting spherical precipitate (r_p) provides yet another expression for the critical radius r_c [65]. The authors considered the nucleation of a prismatic dislocation loop within the precipitate, its growth by climb and its ultimate positioning on an equatorial plane in a circle of radius infinitesimally smaller than r_p. Its energy of formation $E_f = E_\ell - \Delta E_{\bar{e}}$ is approximately given by

$$E_f = \frac{\mu b^2 r_p}{2(1-\nu)}\left(\ln\frac{8r_p}{e^2 b} + 1\right) - \frac{\pi r_p^2 4\mu b f}{1 + 4\mu/3K}\,, \tag{18a}$$

where E_ℓ and $\Delta E_{\bar{e}}$ ($\bar{e} = f$) are respectively the (Volterra) self energy of the loop and the MS energy release in the precipitate and, μ, K, ν, b and f respectively the shear modulus, the bulk modulus, Poisson's ratio and the Burgers vector (all of the precipitate) and f the misfit. From (18a) two alternative critical radii may be deduced:

$$r_c^s = \frac{b(1 + 4\mu/3K)}{8\pi f(1-\nu)}\left[\frac{1}{s}\ln\frac{8r_c^s}{e^2 b} + s\right]; \quad s = 1, 2\,, \tag{18b}$$

the one ($s = 1$) when $E_f = 0$, i.e. the formation becomes energetically favored (see

eq. 15), and the other ($s = 2$) when $\partial E_f/\partial r = 0$, i.e. the precipitate can support an already existing loop at its surface. Note that eqs. (18b) are implicit equations in the unknowns $r_c^{s} = r_p$.

2.1.2. Layer-like growth

The first calculation of critical misfit and thickness in epitaxy was made by Frank and van der Merwe [46] in their analysis of a one-dimensional Frenkel–Kontorova model [53]. In this model the competing forces positioning the adatoms were modeled as follows: the adatom–adatom interaction (AAI) that favors the crystalline structure of the overlayer by harmonic (approximation) forces of force constant $\mu^{(1)}$ and the periodic (periodicity b) adatom–substrate interaction (ASI) by a one-dimensional Fourier series (truncated at first order) of overall amplitude W_0:

$$V = \tfrac{1}{2}W_0[1 - \cos(2\pi x/b)] .$$

Equilibrium considerations (eqs. 13a, b) yield a critical misfit $f_c^{(1)}$ of

$$f_c^{(1)} = 2/\pi\ell_0 , \tag{19a}$$

$$\ell_0 = \mu^{(1)}b^2/2W_0 . \tag{19b}$$

The appearance of $\mu^{(1)}$ and W_0, as measures of AAI and ASI respectively, stresses the importance of interaction strengths in determining f_c.

These results can be extended to a thickening overlayer (crudely though) by assigning to a layer of multiplicity n the force constant $n\mu^{(1)}$. Equations (19) in conjunction with (14) accordingly predict a critical thickness [46]

$$h_c = 8W_0/\pi^2\mu^{(1)}bf^2 . \tag{20}$$

The authors estimated $f_c^{(1)}$ to be about 9% when the AAI and the ASI's are the same. The extension of these considerations to non-equilibrium phenomena will be discussed in section 3.

Although the foregoing model is one-dimensional it has proved to be quite successful and has been the basis and stimulus for numerous extensions to more general cases. The most obvious extension was to two-dimensional model monolayers (ML's) [8, 46, 54]. This extension introduced Poisson's ratio ν_a of the ML, whereby eqs. (19a) and (19b) became

$$f_c^{(2)} = 2/\pi\ell\sqrt{1 + \nu_a} , \tag{21a}$$

$$\ell^2 = \mu_a h b^2/(1 - \nu_a)W_0 \tag{21b}$$

for cases with quadratic interfacial symmetry, μ_a now being the shear modulus and h the thickness; $h = a$ for a ML. Critical thicknesses as in eq. (20) may also be defined for this two-dimensional extension. The effect of Poisson's phenomenon is essentially to reduce the critical misfit; in the average case (AAI and ASI the same) from $f_c^{(1)} = 9\%$ to about 7%, with an analogous effect on critical thickness.

A relevant case that had been very topical recently is the epitaxy of metals at (111) fcc/(110) bcc interfaces [7, 8, 49, 55] where the misfit differs greatly in two mutually perpendicular directions and the occurrence of coherency in one direction only is a natural consequence. Loss of coherency for this case has been studied extensively on the ML or quasi-ML level as for eqs. (21). Also in this case a "configurational" parameter ℓ of the form in eq. (21b), containing the bond ratio $\mu b^2/(1-\nu)W_0$ and thickness h, together with the nearest-neighbor distance ratio $r = a[\text{fcc (111)}]/b[\text{bcc (110)}]$ rather than the misfits explicitly, enter naturally into the theory. When ℓ is small enough a ML grows with 2D coherency for a wide range of r values. As h, and accordingly ℓ, increases there is a critical thickness h_c at which there is a transition to 1D coherency in the Kurdjumov–Sachs (KS) orientation at a critical thickness h^c_{KS} when r is near 1.09 (small misfit f^c_{KS}) or to 1D coherency in the Nishiyama–Wassermann (NW) orientation at h^c_{NW} when r is either near 0.94 or 1.15 (small misfits f^c_{NW}). The transition to the KS orientation requires a rigidlike rotation of about 5°, and its realization may thus be impeded. Furthermore, higher critical thicknesses for loss of the residual 1D coherencies exist. These critical thicknesses are only known semi-quantitatively, like h_c in eq. (20). In all these cases the Poisson effect is important, particularly in the MS energy. The role of Poisson's ratio on critical thickness in a variety of cases has also been studied by Jesser and Kuhlmann-Wilsdorf [59].

degree symbol round o

The most useful extension to the considerations of Frank and van der Merwe was to thickening epilayers, including combinations with different elastic constants [56, 57, 61, 66]. Additional improvements to the model have been made in the form of more accurate representations of the MD energy, as has recently been reviewed [67]. While the extensions of the model to very thin epilayers (less than about two ML's) and to very thick ones are acceptable, it is poor in the regime exceeding two ML's; particularly when the misfit is small. It is in this regime that the approximations employed in the calculations introduce significant effects because the energy of the MD's in the interface is given by an expression that relies on the epilayer thickness being greater than half the spacing between MD's. This limitation led Matthews to adopt a Volterra dislocation approach to calculate the energy of an array of MD's [6, 14]. The Volterra model employs a "cut-off" radius R to the dislocation strain field defined with regard to eqs. (11). Since near $f = f_c$ the MD spacing becomes large the finite thickness h of the epilayer A must be taken as R. Minimization of the total energy in eq. (12c) yields for f_c at given h (using the relations (14)) the result [6, 14]

$$f_c = \frac{D}{2Bh}\left[\ln\left(\frac{h}{b}\right) + 1\right] \tag{22a}$$

or, alternatively for the critical thickness h_c at given misfit f the implicit equation

$$h_c = \frac{D}{2Bf}\left[\ln\frac{h_c}{b} + 1\right] \tag{22b}$$

for a bicrystal AB with quadratic interfacial symmetry. Note that relations (22) imply that $\nu_a = \nu_b = \nu$. Also no dependence on interfacial interaction is incorpo-

rated. This approach is good for thick epilayers but is inaccurate when the epilayer is thin because the role of the ASI is significant here, and the strain field of the MD can extend along the plane of the interface to greater distances than the cut-off radius, which is equal to the film thickness. However, it has the great advantage of simplicity in its representation.

The calculation of critical thickness by energy minimization has also been carried out for partial MD's. Such MD's will have stacking faults associated with them. The stacking fault will extend through one of the crystals and will terminate at the MD in the interface. This case has been treated in the model of van der Merwe using eq. (10) modified for the incomplete Burgers vector [68], as well as from the Volterra dislocation-energy point of view using force-balance criteria [6, 14]. The model of van der Merwe makes provision for a variable interaction. The effect of generating incomplete MD's associated with stacking faults is to reduce the critical thickness. This can be understood as the result of the shorter length of Burgers' vector reducing the energy of the MD's to a greater extent than the stacking fault raises the energy. Negative stacking-fault energy and lattice parameter changes normal to the plane of the fault will also affect the critical thickness, but are usually not considered in the analysis.

The question of critical thickness h_c has also addressed more recently by Bruinsma Zangwill [69] using the Volterra approach to the energy of a MD but a different analytical technique. They included anisotropy effects, but made the simplfying approximation that there exists an "easy" direction for loss of coherency by the introduction of a single sequence of MD's. One would expect this to occur for epitaxy at (111) fcc/(110) bcc interfaces characterized by widely different misfits in two perpendicular interfacial directions. The authors obtained, like Matthews, an implicit equation of the form (22b) but with somewhat different entries and give their (approximate) result as

$$h_c = \kappa_4 / f \,, \tag{23}$$

where κ_4 is a "constant" depending on elastic constants as in eq. (22b).

Only one model exists, i.e. the one in which the periodic adatom–substrate interaction potential is represented by a succession of parabolic arcs [56, 57], rather than a truncated Fourier series, that has been solved exactly for all thicknesses (h) and strengths of atomic interaction. In this model the effect of unrealistic potential maxima (high cusped peaks) on the prediction of critical thickness is drastically reduced because at the critical thickness the MD's are widely spaced and the contribution of the peaks to the overall energy is presumably small. The critical misfit for given h and materials with identical atomic interactions is now given by

$$f_c = \frac{1-\nu}{4\pi(1+\nu)} \int_0^{\infty} \frac{\mathrm{d}X}{X[X + \alpha\phi(X,\alpha)]} \,, \tag{24a}$$

$$\phi(X,\alpha) = \frac{1}{2} \frac{X(s^2 + cs - X - X^2) + \delta(c+s)^2/2(1-2\nu)}{X(s^2 - X^2) + \delta(s^2 + sc + X - X^2)/2(1-2\nu)} \,, \tag{24b}$$

$$\delta = 2(1-\nu)h/a \,, \qquad s = \sinh X \,, \qquad c = \cosh X \,.$$

For different elastic constants and thicknesses the critical misfit f_c has the same form as (24a), but now the function ϕ is vastly more complicated [70]. Alternatively, for given misfit ($f_c = f$) this is an implicit equation for the critical thickness h_c.

2.1.3. Island growth

The foregoing case of uniform epilayers covering a substrate is of great practical importance. In the initial stages of epitaxial growth, however, most systems exhibit island growth, which may be Volmer–Weber or Stranski–Krastanov growth [47]. If the misfit is not small or if the thickness at which the growing deposit becomes nearly continuous is large then loss of coherency will tend to occur when the deposit is in the form of individual islands. The coherency strain energy for the case of island deposits was first calculated by Cabrera [71]. This model included elastic strains in both the island and the thick substrate. It was used to calculate the critical radius above which it is energetically favorable to introduce MD's [59]. The calculation of critical misfit and thickness in this case differs somewhat from that for layer growth. The authors have assumed that at the interface between a non-coherent hemispherical epitaxial island A of radius R and a thick substrate B MS's of magnitude $\bar{e}_a$ and $\bar{e}_b$ exist with equilibrium values $\bar{e}_a^{\mathrm{m}}$ and $\bar{e}_b^{\mathrm{m}}$. The condition for the critical misfit was taken as

$$\bar{e}_a^{\mathrm{m}} + \bar{e}_b^{\mathrm{m}} = f_c\,, \tag{25a}$$

i.e. that *almost* all the misfit is accommodated by MS. For the case in which the atomic interactions are the same ($\mu_a = \mu_b = \mu$) and the interfaces have quadratic symmetry ($a_x = a_y = a$) it follows, using eq. (10a), that for given R (approximately)

$$f_c = -\frac{3b(4a_0 + a_s)}{R(a_0 + a_s)}\,\beta \ln[2\beta(1+\beta^2)^{1/2} - 2\beta^2]\,, \tag{25b}$$

$$\beta = \frac{\pi f_c}{(1-\nu)R}\,. \tag{25c}$$

where a_0 and a_s are the nearest-neighbor distances in overlayer and substrate respectively. This is an implicit equation in f_c. For given misfit f, eq. (25b) defines a critical size R_c.

Because of the finite size of an island the introduction of discrete MD's causes an abrupt decrease in MS which was observed experimentally in tin deposits on tin telluride by Vincent [72]. For this case of discrete MD generation the calculation of critical thickness and subsequent MD introduction with island growth laterally has been considered from the point of view of energy minimization using Volterra MD's [6], and using the Frank–van der Merwe model [73]. Vincent used the criterion that MD's would be generated at an island size for which the MD would reduce the MS to zero. The main significance of the above calculations for the present paper is that the island morphology is the ideal one for availability of MD's because an efficient MD can be generated at the edge of the island. This mechanism for the generation of efficient MD's has been

confirmed experimentally by Yagi et al. [74]. Therefore one expects the barriers to MD generation to play only a small role in epitaxial systems exhibiting island formation, particularly at elevated temperatures where dislocation mobility is high. In such cases good agreement between experimentally observed values of critical size and calculated values is found.

2.1.4. Superlattices

In view of the increasing importance of superlattices over the past decade it became similarly important to develop an energetic approach to calculating the critical thickness for this case. A first approximation was made by applying the van der Merwe expression for the energy of an array of MD's to the stack of interfaces [67]. This approximation did not, however, include the reduction of MD energy caused by the cancellation of strain fields across the layer between interfaces. This cancellation is strongest when the layer thickness is small. Further, the expression loses accuracy when the misfit is small. Both of these effects, which are relevant to the case of superlattices of layers A and B, have been recently accounted for more properly by reworking the interfacial misfit boundary from first principles including the boundary conditions applicable to superlattices [57].

In this approach [57] the following boundary conditions were adopted for the interior of an extended superlattice: (i) the normal and tangential forces at the interface are almost everywhere linear functions of the corresponding relative displacements of atoms on either side of the interface [56]. (ii) the stresses acting through the interface are continuous there, (iii) the midplanes between interfaces remain plane and contain no shear parallel to the interface, (iv) the superlattice faces normal to the interface are free of forces and (v) the crystals may otherwise be approximated by isotropic elastic continua. The governing equations of this model are exactly solvable. The critical misfit f_c, as determined by eq. (14), depends on the ratios of elastic constants (M) and thicknesses (r). i.e. on

$$M = \frac{(1+\nu_a)(1-\nu_b)\mu_a}{(1+\nu_b)(1-\nu_a)\mu_b}, \qquad H = \frac{h_a}{h_b}, \qquad h_B \equiv \eta c \tag{26a}$$

and on the thickness of any one layer (taken as $2h_B$), through an equation of the form

$$f_c(\eta; H, M) = \frac{(1-\nu_b)(1+H^{-1}M^{-1})\mu_{ab}}{4\pi(1+\nu)\mu_b} \int_0^\infty \frac{dX}{X[X+\delta(\eta)\phi(X)]}, \tag{26b}$$

where $2h_a$ is the thickness of layer A, c is a reference lattice spacing, $\delta(\eta) = 2(1-\nu)\eta$ and $\phi(X)$ is a complicated function of the variable X, elastic constants and thicknesses. In accordance with the criterion (14) this equation defines for given misfit $f_c = f$, and ratios H and M, a critical thickness $\eta = \eta_c$.

Various simple and asymptotic cases of this prediction have been considered of which the simplest is the one in which the layer thicknesses and elastic constants are respectively equal. In this case

$$\phi = \frac{c^2}{sc + X} \equiv \frac{\cosh^2 X}{\sinh X \cosh X + X}\,, \tag{26c}$$

i.e. it is no longer dependant on thickness and elastic constants. Also, approximately

$$f_c(\eta) = \frac{1-\nu}{2\pi(1+\nu)}\left[\frac{2}{\delta B}\arctan 2B + \frac{1}{\delta}\ln\left(1+\frac{\delta}{2}\right)\right] \tag{27a}$$

$$B(\eta) = [(9+4\eta)/6\eta]^{1/2}\,, \qquad \delta = 2(1-\nu)\eta\,. \tag{27b}$$

Two conclusions from these results warrant special mention. Firstly it follows from a log–log plot that η_c varies approximately as $f^{-\chi}$; $\chi = 1.22$ as compared to $\chi \approx 1$ for previous estimates and $\chi = 2$ for eq. (20) and the approach of People and Bean [75], which apparently agreed with observations of Bean et al. [76]. Also this calculation shows that when the layer thickness is small the energy of the MD's decreases by a factor of two over that of a single layer on a thick substrate. For thick layers, these energies approach one another. Further, since the misfit in a superlattice is accommodated equally by homogeneous elastic strains in each of the layers when elastic constants are equal, the critical misfit and the critical thickness both increase by a factor of up to five over their respective values for the case of a single layer on a thick substrate.

2.1.5. Graded epilayer

A critical thickness calculation for a bicrystal of a different kind was performed by Ball and Laird [77]. The authors considered coherency loss by the introduction of 60° imperfect MD's in a graded epilayer in III–V compounds, expressing the variation of lattice parameters by Vegard's law for a uniform composition gradient. They used the Volterra approach in expressing the stresses and calculated the work W needed for MD's to glide in from the free surface; a mechanism which also involves the overcoming of surface image barriers. Defining h_c as the value of h which satisfies the relations (13b) and (14) they obtained (approximately)

$$h_c = \frac{d}{1-\rho_m}\exp\frac{1+\rho_m}{2\rho_m}\,, \tag{28}$$

where d is the average spacing of atomic planes aligned parallel to the interface, ρ_m is the value of g/h at which W rises to a maximum as t increases and g is the distance of the MD from the *free surface*. The dependence of h_c on grading s is calculated using the parametric equation (28) $h_c = h_c(\rho_m)$ and $s = s(\rho_m)$ (the latter from the equilibrium equation), rather than calculating ρ_m directly from an extremum condition. In comparing these predictions with experiment the authors concluded that even in epilayers without threading dislocations, dislocation multiplication, which will inevitably take place, may introduce serious discrepancies between theory and experiment.

2.2. Stress and force criteria

We now consider in some detail the application of force or stress balance in calculating equilibrium critical parameters. We have discussed, in relation to eqs. (13) the conditions under which force balance and energy minimization (equilibrium) procedures are equivalent. Neither explicitly involves the nucleation energy barriers associated with the acquisition of MD's. The minimization procedure may ignore the barriers because only initial and final energies are relevant. The force balance procedure, on the other hand, presupposes the existence of threading dislocations (TD's), so that nucleation is in any case irrelevant. The mechanism of gliding a threading dislocation that threads the substrate, interface and epilayer on a single slip plane into the interface to make an increased line length of MD was first proposed and observed by Matthews [10]. An inevitable consequence of this mechanism is that the resulting MD is of imperfect type (fig. 1); as for example, when the generating dislocation is a 60° glide dislocation in the semiconductor. The Peach–Koehler driving force $F_{\bar{e}}$ on the dislocation, due to MS, is proportional to the thickness of the epilayer. The force F_D opposing this glide force is the line tension. Matthews has used the Volterra approach (analogous to eq. (11a)) to calculate the line tension. When a critical thickness is reached for which the resultant vanishes so that the condition (13c) is satisfied, then any further increase in thickness will sweep the threading dislocation along the glide plane while it draws out a MD along the interface [13].

round o

2.2.1. *Thin epilayer on a thick substrate*

This model has been improved by adding the line tension σ_0 (force F_σ) of the step generated at the free surface when sweeping the threading dislocation [6, 14]. Furthermore, if the threading dislocation splits into partials separated by a stacking fault (energy γ_0 per unit area) in the epilayer then the above model can be extended to include the relevant force F_γ needed for stacking fault formation. The condition in (13c) may accordingly be used to derive an expression for the critical thickness above which imperfect MD's will be introduced [6, 13, 14, 35]. The force balance equation (13c) now takes the form

$$0 = F_{\bar{e}} + F_D + F_\sigma + F_\gamma$$
$$= Bbh\bar{e}\cos\lambda + Db(1 - \nu\cos^2\alpha)[\ln(h/b) + 1] + \sigma_0 b\sin\alpha + \gamma_0 h/\cos\phi \tag{29a}$$

and yields for the critical thickness h_c ($\bar{e} = f$) the implicit equation

$$h_c = \frac{Db(1 - \nu\cos^2\alpha)[\ln(h_c/b) + 1] + \sigma_0 b\sin\alpha}{Bbf\cos\lambda - \gamma_0/\cos\phi}, \tag{29b}$$

where λ is the angle between the slip direction and the interfacial direction normal to the line of intersection of the slip plane (S) and the interface, α is the angle between line of the mixed MD and its Burgers vector $\boldsymbol{b}$ and ϕ is the angle between the unit vector $\hat{\boldsymbol{n}}$, normal to S, and the line of intersection of the free

the

surface and a plane which is normal to the free surface and contains $\hat{n}$. Note (i) that in the term $\ln(h/b)$ h is the cut off radius assuming that h is less than half of the MD spacing and (ii) that for a perfect MD the γ_0 term is absent and $\boldsymbol{b}$ is the complete Burgers vector. This example illustrates the power of the force balance approach. A disadvantage of the approach is that it handles less well the interaction between MD's.

3.2.2. Superlattices

Hirth and Evans developed a criterion for critical thickness by considering injection of a dislocation dipole or a crack from the edge of a stack of coherent layers [78], into a superlattice in which consecutive layers have equal thicknesses and elastic constants. They calculated the Peach–Koehler force on the mobile segment of a dislocation dipole with one leg in one interface and the other leg in an adjacent interface connected by a segment threading a layer on its slip plane. This force was equated to the line tension of the mobile segment as was done in previous models [15]. The thickness for which this force balance occurs is the critical thickness h_c that was calculated for isotropic elasticity as well as anisotropic elasticity. The results of this approach essentially agree with those of the earlier similar model [15], and for isotropic elasticity give

$$h_c = \frac{\boldsymbol{b}(1 - \nu \cos^2 \alpha)}{2\pi(1 + \nu)} \ln\left(\frac{\sqrt{6}h_c}{2\boldsymbol{b}r_0}\right). \tag{30}$$

This result is very much the same as the ones already dealt with except that the $\sqrt{6}$ corresponds specifically to the $GaAs/GaAs_{0.5}P_{0.5}$ system considered and r_0 is a core cut-off parameter.

2.2.3. Multilayers

Bokii and Kuznetsov have also calculated the critical layer thickness for multilayered double heterostructures of $Ga_xIn_{1-x}P_{1-y}As_y$ alloys from the point of view of balancing the MD line tension and surface step resistance to dislocation motion with the Peach–Koehler force on the MD due to the MS [79]. They started with the expression for critical thickness by Matthews [80] and allowed for elastic bending (a non-vanishing couple) by employing the elasticity calculation of Olsen and Ettenberg for the elastic bending of multilayer heterostructures [81] to develop a modified expression for the critical thickness of the topmost layer in the stack. They obtained for the critical thickness h_{jc} in the topmost layer of a multilayer of multiplicity j the expression

$$h_{jc} = \frac{\boldsymbol{b}(1 - \nu \cos^2 \alpha)[\ln(h_j/\boldsymbol{b}) + 1] + 0.4\pi a_j(1 - \nu^2) \sin \alpha}{8\pi(1 + \nu) \cos \theta \sin \psi(\varepsilon_j + N\boldsymbol{b})}, \tag{31}$$

where $\boldsymbol{b}$, ν, α, h_j have evident meanings (see also eqs. (29)), a_j is the lattice parameter and ε_j the elastic deformation into the topmost layer, N the number of MD's generated, ψ the angle between the {111} slip planes in the alloy and the free (001) surface and θ the angle between the MD Burgers vector $\boldsymbol{b}$ and the $[1\bar{1}0]$

direction. Note that the line energy of the surface step has been expressed in terms of shear modulus and Poisson's ratio; the last term in the numerator. Their experimental results are consistent with the introduction of MD's by the Matthews mechanism of activating glide of the threading dislocations on their slip planes and also with this mechanism occurring in the topmost layer being grown at any given time. Comparison of their calculated values of critical thickness with their measured values shows reasonable agreement.

2.2.4. MD formation by climb

The above arguments rely on the glide of dislocations into the interface. It has also been observed that MD's are introduced into the interface by climb [74, 82]. Even at relatively low temperatures climb can be the dominant mechanism for relieving MS through the introduction of MD's. Matthews equated the climb force on a dislocation due to MS to those forces that oppose climb, and calculated the critical thickness for the generation of MD's by climb [6, 14]. He included the possibility of faulted dislocation loops as well as the extra energy associated with creating or annihilating new surface during climb. Since MD's generated by climb would naturally be in edge orientation the climb force is simply $B\boldsymbol{b}h\bar{e}\cos\psi_0 + D\boldsymbol{b}\ln(h/\boldsymbol{b})$, where ψ_0 is the angle between the glide plane and the free surface. Since ψ_0 is naturally small we may approximate the width of the surface strip created or annihilated by $\boldsymbol{b}$ and the relevant force due to line tension by $\pm\sigma_0\boldsymbol{b}$, where σ_0 is the surface free energy. It follows as for glide that

$$h_c = \frac{D[\ln(h_c/\boldsymbol{b}) + 1] \pm \sigma_0}{Bf - \gamma_0/\boldsymbol{b}\cos\psi_0} . \tag{32}$$

Whether or not this mechanism operates will be controlled by the availability of point defects provided thermally or otherwise.

The climb of MD's out of the interface during interdiffusion of two crystal halves of a bicrystal has also been considered [83–85]. These considerations of the climb of MD's have in common that the MD's leave the interface by climbing in both directions i.e. toward both surfaces. This can lead to threading dislocations or to a distribution of MD's throughout a zone of increasing thickness with increasing interdiffusion or to polygonization walls. The additional influence of an excess of point defects generated by the Kirkendall effect has been discussed [6, 86]. The Kirkendall effect can either aid or hinder the climb process of the MD's, depending on the relative diffusivities of the atoms.

2.3. Geometrical criteria

Geometrical criteria are not true equilibrium [24] criteria. Most of the geometrical models for accommodating misfit are of the type introduced earlier [3, 5] which in effect relate a closure failure to a dislocation density. This model is usually applied to planar interfaces that are narrow in thickness. There exist rather thick interfaces over which MD's accommodate the misfit in a thick transition region,

areal areal

referred to as a continuously graded junction. The ~~annealed~~ density n_A of misfit dislocations in a cross-section of the graded interface can be related to the linear density n_L of misfit dislocations in an abrupt interface of the same misfit as that of the graded interface when a transitional interface thickness w_0 is obtained [87]. When the graded interface is sufficiently thin that $w_0 = 1/n_L$ then $n_A \approx n_L^2$. For interface thicknesses below the transitional value w_0 the misfit dislocation spacing is fixed at the value $1/n_L$. For interface thicknesses above w_0 the ~~annealed~~ areal density of misfit dislocations is proportional to the lattice constant gradient. These considerations are useful for relating dislocation densities in the substrate, abrupt interface and graded interface in a unified way.

areal

The physical basis for these models beyond geometrical considerations is the relief of long-range stresses. In effect the geometrical models do not minimize long-range stresses by considerations that balance energies or forces but instead require that the long-range stresses be zero. Therefore they should be reasonable models when the misfit is not small, i.e. in a regime where energy and force balance considerations show that the residual long-range stresses should be near zero once the critical size is exceeded. In fact it has been shown that the mode of MA becomes vernierlike as the misfit increases, with vanishing MS (i.e. longer-range stresses), as for rigid crystals. Furthermore, deep minima in the AB interaction occurs for one-dimensional matching and maximum depths for complete registry. This provides energetic justification for the geometrical models [8]. The half-widths of the minima are zero for infinite interfaces and increase with decreasing interface area.

The usual task of the geometrical model is to find the required MD density that will accommodate the misfit rather than to predict a critical size. It is also useful for selecting the various types of MD's that accommodate the misfit [3, 29, 88]. There is, however, one geometrical criterion for critical size that is both simple and useful for estimating values quickly. Brooks [89] used the classical concepts of dislocations to arrive at essentially a geometrical criterion for critical size of a misfitting sphere by simply equating the misfit f multiplied by the sphere diameter $2R$ to the Burgers vector $\boldsymbol{b}$ of the MD. Physically this criterion is equivalent to requiring the introduction of a MD into the interface when the accumulated displacement from the misfit vernier causes an atomic shift to the next available potential energy trough, as was noted with regard to eq. (25c). The Brooks criterion is a good rule of thumb that gives critical-size values that are not very different from those provided by most sophisticated criteria [64].

3. *Calculations: non-equilibrium*

In the foregoing we have discussed existing calculations of critical misfit and thickness based on equilibrium criteria. Empirical data, particularly on semiconductors, exist that simply cannot be understood in terms of equilibrium principles and can only be interpreted on the basis of non-equilibrium considerations and on the supposition that the path towards equilibrium goes through one or more, or

perhaps a distribution of, metastable configurations separated from each other and from the final stable configuration by energy barriers of the kind discussed in relation to eq. (16). For example, when the bicrystals are initially free of dislocations, MD's need be nucleated, possibly in the form of critical sized loops that would grow spontaneously in the presence of MS stresses. Misfit-accommodating dislocation loops have indeed been observed in metals [34, 35] as well as semiconducting epicrystals [90, 91]. The processes of approach towards equilibrium states clearly involve thermal fluctuations, and knowledge of the height of the energy barriers, e.g. the *nucleation energy* U of a MD loop that would continue to grow in the presence of the MS induced stresses, and of the rate or degree of advance towards equilibrium, are of vital interest. The latter, in particular, would assist in distinguishing the equilibrium or non-equilibrium nature of experimental data.

3.1. Monolayer approach

Both these questions have been addressed previously. Continuum theory predicts that $U \sim 200$ eV [52, 78], which completely rules out the thermal acquisition of MD's in an otherwise perfect crystal. Clearly U will be reduced by the MS (Peach–Koehler) stresses in a coherent epilayer. In fact, Frank and van der Merwe [46] have shown, using the one-dimensional model that the activation energy for the introduction of a MD in a coherent ML is

$$U_1 = (4W_0\ell_0/\pi)[(1 - f^2\ell_0^2)^{1/2} - f\ell_0 \arccos f\ell_0] \tag{33}$$

per atom row of the ML, where the quantities in this relation have been defined in eqs. (19). The energy U_1 has a maximum value of $4W_0\ell_0/\pi \sim 9W_0$ at $f = 0$, diminishes to about $2W_0$ at the misfit $f_c^{(1)} = 2/\pi\ell_0$ in eq. (19) where a coherent ML becomes unstable and vanishes at

$$f_s^{(1)} = 1/\ell_0 = \pi f_c^{(1)}/2 \qquad (\approx 14\%)\,. \tag{34a}$$

In metals $W_0 \sim 0.2$–0.5 eV. Thus above $f = f_s$ MD's enter spontaneously without the assistance of thermal energy. In accordance with eq. (14) this defines (crudely) a critical thickness

$$h_s = \pi^2 h_c/4 \sim 2h_c \tag{34b}$$

above which MD's will enter spontaneously for given $f < f_c^{(1)}$.

3.2. Nucleation of misfit dislocation loops

The spontaneous generation of MD's in practical cases was also considered by Dodson [52]. He realized that the smallness of the critical MD loop radius makes a continuum approach invalid and that therefore an atomistic approach is needed. He carried out calculations on Si considering a model system consisting of a

tetrahedral slab with (111) faces and periodic boundary conditions in the plane of the interface using a super unit cell of 7×7 atoms extending to a depth of 6 monolayers. Dodson [92] employed modified Tersoff potentials using Monte Carlo techniques. He concluded that a climb MD loop of 90° partial having a radius of 2 Å becomes critical (forms spontaneously) at a compressive misfit of 11.2%. This is greater than the 4% estimated by equilibrium theories and very much greater than those which had been observed. Dodson suggests that perhaps the most significant limitation of this calculation is that it neglects the influence of existing surface steps. A feeling for the magnitude of this critical misfit may also be acquired if one compares this value of 11.2% with the critical misfit $f_s^{(1)} = 1/\ell_0$ (in eq. (34a)) of about 14% for spontaneous generation of edge-type glide MD's in the one-dimensional ML model. In any case, these calculations clearly suggest that considerations on the nucleation of MD's in an otherwise perfect crystal will not generate satisfactory understanding of observed non-equilibrium phenomena.

round o

3.3. Overcoming the Peierls barrier

More recently Tsao et al. [93] have extended the foregoing work on non-equilibrium systems and measured the strain relief in metastable Si_xGe_{1-x} strained layers grown on (001) Ge substrates and concluded from their data and the comparison of the Peach–Koehler MS force component $F_{\bar{e},par}$ parallel to the interface and the corresponding opposing force $F_{D,par}$ due to the line tension (self energy) of the dislocation (respectively $F_{\bar{e}}$ and F_D of our notation in eq. (13c)) "that metastable strained layer breakdown is most directly correlated, not with thickness h and lattice mismatch f, but rather with (1) an excess stress $\tau_{exc} = 2(F_{\bar{e},par} - F_{D,par})/hb$ and (2) temperature".

Using the expression for the Peach–Koehler force and the Volterra form of line tension the authors obtained

$$\tau_{exc} = \frac{2\bar{e}\mu(1+\nu)}{1-\nu} - \frac{\mu(1-\nu\cos^2\alpha)}{2\pi(1-\nu)}\,\frac{\ln(4h/b)}{h/b}\;; \quad \bar{e} = f \tag{35}$$

in the foregoing notation, where the angle α between the MD line and its Burgers vector is 60° in the present case.

round o

Whereas the condition $\tau_{exc} = 0$ is equivalent to $F = 0$ in eq. (13c) and is still valid for equilibrium, the energy barriers temporarily delay the transition to equilibrium. At a temperature of 494°C strain relief became observable when $\tau_{exc}/\mu \approx 0.024$.

Scaling excess stress with the shear modulus μ and temperature in Kelvin with the melting temperature T_m allowed the authors to construct experimentally confirmed stress/temperature stability diagrams which provide a simple unified description of the stability–metastability regimes for Si_xGe_{1-x} strained layers.

Since the derivation of τ_{exc} assumes the presence of threading dislocations one may speculate that the relevant barriers are those due to the Peierls stress opposing the glide motion. The net driving force $(F_{\bar{e}} - F_D)$ which consists of the

Peach–Koehler MS force diminished by the line tension is proportional to τ_{exc}. It reduces the height of the Peierls barrier [61, 94]. When the barrier height is sufficiently reduced by increased layer thickness the frequency with which the barrier is overcome increases to the extent that strain relief becomes observable. Thus the finding of the authors can be understood in terms of fluctuation principles.

3.4. Overcoming the surface barrier

Sharan and co-workers [95] also recently dealt with the problem of the critical thickness-misfit dependence of GeSi alloy epilayers on (001) Si surfaces. They considered the inward glide on a (111) slip plane of a long straight 60° dislocation that had been nucleated "at" the surface, and calculated, using "surface dislocation" techniques to satisfy boundary conditions, the (surface) energy barrier U_s per unit area of interface (comprising self energy as well as work done by the MS stress) which it has to overcome in its passage from the free surface to the interface. The barrier U_s has furthermore been calculated on the assumption that there is a sequence of MD's of sufficient density to accommodate all the misfit (as expressed in terms of MD spacing), while not affecting the MS stress. The authors adopted the criterion that strain relief commences (the critical misfit is reached) when $E_{\tilde{e}} \geq U_s$, where $E_{\tilde{e}}$ is the misfit strain energy per unit area of interface when f is entirely accommodated by MS. The fact that this approach yields consistently higher values of critical misfit than the approaches of Matthews [6, 14] and others reflects that the present approach takes into account the opposing image force on dislocations when they recede from the free surface and the fact that the criterion $E_s \geq U_s$ is a non-equilibrium extension of the approximate equilibrium criterion (15) which overestimates the critical parameters. Although the theory is not a true nucleation theory it involves the concept of an energy barrier characteristic of a transition from a metastable to a stable configuration. The approach also does not incorporate the Peierls barriers [14, 93] thought to be important in the semiconducting materials concerned.

3.5. Quasi-nucleation approach

Another critical misfit calculation that is significant and had also been controversial is the one of People and Bean [75], who developed an analytical expression for h_c to account for the large discrepancy between observed values and those calculated from the expressions of Matthews [6, 14] and van der Merwe [56]. The authors essentially assumed that the energy E_D of a MD, which they simply took to be of *screw* type [29], is very much localized within a vertical wall spanning an interfacial area $L \times w$ ($L \equiv$ MD line length and $w =$ MD "width" taken as five [110] spacings for Ge_xSi_{-x} alloys). They furthermore assumed that h_c is reached when the portion of $E_{\tilde{e}}$ contained in the volume of the wall exceeds E_D. This approach, which the authors presented as a nucleation approach but clearly is not

a true one, also contains elements of the criterion in eq. (15). The approach was subsequently also employed by Luryi and Suhir for finite substrates [23]. In equilibrium approaches one finds that the critical thickness depends on the misfit to a power between -1 and -1.5 [57]. The effect of the fixed width of the (screw) dislocation is to yield a misfit dependence of the critical thickness near a power of -2 as in the crude approach to the result in eq. (20), and apparently yielded agreement with their data. In effect this approach is trading all of the MS energy for an almost fixed energy barrier whose value was empirically adjusted through the dislocation width to fit the experimental data on silicon–germanium alloys [76]. To understand the physical details of this approach is difficult but its success seems to lie in picking a nearly fixed energy barrier so that the explicit square dependence of MS energy on misfit is transmitted to the expression for h_c. It is a quasi-nucleation approach in which energies in finite volumes are compared rather than total energies as for equilibrium. Furthermore, it has recently been argued that the authors made serious flaws in the interpretation of the results [96]. Because of its controversiality this paper had the merit of stirring up great activity.

3.6. Misfit strain gradient (MSG)

An interesting case of critical thickness concerns epilayers on substrates of finite width. It makes use of a significant misfit strain gradient (MSG) to accommodate the misfit f. Luryi and Suhir adopted a *rigid* substrate width of 2ℓ along the one-dimensional misfit direction and an epilayer of the same width to calculate the MS energy $E_{\tilde{e}}$ per unit area of the interface for the case of a coherent interface with a strain field decaying exponentially with distance normal to the interface [23]. They showed that when the width ℓ is small in comparison to the thickness h, there is an effective thickness h_e smaller than h that characterizes the limit of the strain field. In other words $E_{\tilde{e}}$ remains finite even though h becomes infinite. The authors defined an effective finite epilayer thickness h_e by equating the MSG energy $E_{\tilde{e}}$ per unit interfacial area, calculated on the basis of the stress distribution on the mid plane, to $[Ef^2/(1-\nu)]h_e$, i.e. as though there exists a uniform energy "density" up to a distance h_e from the interface, obtaining

$$h_e = h\left(1 - \operatorname{sech}\frac{\zeta\ell}{h_e}\right)(1 - e^{-\pi h/\ell})\,\frac{\ell}{\pi h} \equiv h\left[f\psi\left(\frac{\ell}{h}\right)\right]^2, \tag{36a}$$

where the function ψ is such that

$$h_e \approx \begin{cases} h & \text{when } \ell \gg h\,, \quad (36b) \\ \xi^2\ell/\pi & \text{when } \ell \ll h\,, \quad (36c) \end{cases}$$

and $\zeta^2 = 3(1-\nu)/2(1+\nu)$, ν being Poisson's ratio. The quantity ξ, defined for $\ell/h \ll 1$, is obtained numerically from (36a) and (36c) and is about 0.89 for Ge_xSi_{1-x} alloys, so that $h_e/\ell \approx 1/4$ in this case ($\ell \ll h$), whereas $h_e \approx h$ when $\ell \gg h$.

The authors furthermore calculated a critical thickness h_c^{ℓ} following People and

Bean [75] by assuming that coherency breaks down with the introduction of a MD when E_{ℓ} exceeds E_D; E_{ℓ} and E_D having the same meaning as in ref. [69]. This yields for given f and ℓ the equation

$$\frac{b}{t}\ln\frac{t}{b} = A\left[f\psi\left(\frac{\ell}{t}\right)\right]^2; \qquad t = h_c', \qquad A = \frac{20\sqrt{2}\pi a(1+\nu)}{(1-\nu)b} \tag{37a}$$

for h_c', where b ($\approx$4 Å for Ge_xSi_{1-x}) is the Burgers vector of the MD and a is the lattice constant. The solution of (37a) is seen to be a function of $f\psi(\ell/h_c')$ and may accordingly be written in the form

$$h_c'(f) = h_c\{f\psi(\ell/h_c')\} \tag{37b}$$

with the limiting value $h_c\{f\}$ when $\ell/h_c' \gg 1$. The factor $\ln(h_c'/b)$ in (37a) results from the concept of a cut-off radius (see eqs. (11)) $R = h$, applicable when $\ell \gg h$. When $\ell < h$ the authors adopted as cut-off radius $R = \ell$ (= distance to lateral free surface) and obtained for ℓ_c the equation in (37a) with $t = \ell_c$. In the limit $\ell \ll h$ they accordingly obtained (using (36c)) for (37b) the result

$$\ell_{min}(f) = h_c\{f\xi/\sqrt{\pi}\}. \tag{38}$$

The authors showed that the curve of $h_c'(f)$ vs ℓ has two asymptotes h_c and ℓ_{min} and that, at $\ell/b = \ell_{min}/h$, $h_c' \to \infty$, i.e. when $\ell \leqslant \ell_{min}$ the critical thickness is infinite. Equivalently this value of width, ℓ_{min}, represents the size limit below which the misfit is entirely accommodated by a MSG. It then is better described as a situation for which there exists a critical misfit f_c that is a function of $\ell < \ell_{min} \ll h$. These calculations are qualitatively very significant because they point to a new mechanism for avoiding MD's. They could be refined by allowing relaxation of the substrate and by adopting a more physically based energy balance analysis. For example the assumption of a uniform strain f at the interface when ℓ is small may not be realistic [73].

3.7. Strain relief by crack formation

The acquisition of MD's by the nucleation of critical sized MD loops at typical temperatures has been ruled out. It is interesting though that a calculation of the stress intensity factor for a crack gives a critical thickness above which crack growth will occur as being near h_c for the nucleation of a dislocation loop. The experimental value for critical thickness compared to that given by the above calculation gave a value near h_c obtained from a balance of forces. This result points to the growth of pre-existing dislocations as the operative mechanism of coherency loss rather than the nucleation of new MD's. There are, however, several epitaxial systems for which insufficient dislocation slip is available to accommodate the misfit and cracking occurs as an alternative [27]. To model the critical thickness for cracking a Griffith criterion was adopted. The critical thickness for cracking is equal to the critical Griffith crack length:

$$h_{crack} = 2\gamma(1-\nu)^2/\pi E f^2 \tag{39}$$

in which the surface free energy γ of the crack has been approximated by $\gamma = Ea_0/10$, E being Young's modulus [59] and a_0 the interatomic distance. The third and somewhat extreme case, that the misfit stresses are large enough for cracks to form spontaneously, appears to be of less importance. Reasonable agreement between experiment and the relation (39) was found [27] for an epilayer with a cleavage normal to the interface. Plane

Crack formation of a different kind was investigated by Hirth and Evans [78]; crack formation starting at the free lateral faces of a stack and propagating along the interface. Accepting as postulate for h_{crack} that crack propagation proceeds when the driving stress due to MS exceeds the resistance K^{int} to crack growth they obtained the value

$$h_{\text{crack}} \approx \left[\frac{(1-\nu)K^{\text{int}}}{2(1+\nu)\mu}\right]^2. \tag{40}$$

Using reasonable estimates for K^{int} this yields values of the same order as h_c for the introduction for MD's.

3.8. Dynamics of strain relief

Recently Dodson and Tsao [97] used a phenomenological model of dislocation dynamics and plastic flow developed by Haasen and co-workers [98] for diamond structure materials to investigate the role of strain relaxation *rate* in experimental data concerning critical thickness in semiconductors. The main features of the model are: (i) the epilayer is initially strained to coherence (MS $\bar{e}$ = misfit f) in a metastable configuration, (ii) the MS is generated as a constant biaxial deformation relaxed by the introduction of MD network at the interface, (iii) the rate process mechanism is the acquisition of MD's; elongation of existing dislocations and the generation of new ones, (iv) a localized stress τ_{loc} exists that acts as a driving force for the motion of dislocations, while the motion is retarded by energy barriers needing thermal energy for overcoming them and (v) τ_{loc} can be expressed in terms of the residual mismatch $f - \bar{f}(t) - \bar{e}(h)$; $\bar{f}(t)$ being the misfit accommodated at time t by the insufficient density of MD's (the plastic part) and $\bar{e}(h)$ the equilibrium MS at thickness h. The authors consider a nearly dislocation-free coherent layer and make provision for MD sources in terms of a background dislocation density $\bar{f}_0$, obtaining the result

$$\mathrm{d}\bar{f}(t)/\mathrm{d}t = C\mu^2[f - \bar{f}(t) - \bar{e}(h)]^2(\bar{f}(t) - \bar{f}_0), \tag{41}$$

where μ is the shear modulus and C a constant depending on temperature.

The authors concluded that this equation gives an excellent description of the data [76] for SiGe alloys on (100) Si substrates using as model parameters $C = 30.1$ and $\bar{f} = 10^{-4}$. The results are very insensitive to the unknown value for $\bar{f}_0$. For example, while the critical misfit for a 500 Å epilayer is 0.004, both theoretical and experimental values only deviate observably from coherency at $f = 0.1$. These considerations constitute significant progress towards a successful dynamic description of strain relief, i.e. non-equilibrium critical phenomena.

4. Correlation with experiment

There are numerous techniques by which loss of coherency may be observed. Typically the experimental measurements of the critical thickness for coherency loss rely on observations of MD's or such indications of their presence as moiré fringes in electron micrographs. The roles of the MV, MSG and MC mechanisms are usually insignificant and ignored. Further, the measurement of critical thickness and misfit is not a precise one. Nevertheless a sufficient number of measurements and estimates have been made to permit the disclosure of some trends. We will not consider all of the measurements reported in the literature, but rather deal with them in several categories. Metal islands, metal films, semiconductor systems and superlattices will be covered, but precipitates will be omitted. For a collection of data on the critical size of precipitates one may consult refs. [6] and [14]

There are two measurements of the critical size for loss of coherency between island overgrowths and their continuous substrates. For cobalt islands on copper substrates and γ-iron on copper the respective sizes of the largest islands that remained coherent at 38 nm and 75 nm [13]. The corresponding calculated values are 36 nm and 87 nm [59]. This closeness of agreement between measured and calculated values is not typical and therefore may be somewhat fortuitous.

A number of measurements of h_c have been made on thin films of nickel on copper substrates and Cu on Ni. Gradmann measured the critical thickness in a Ni/Cu vapor deposit as about 1.0 nm [9]. Similar results were obtained for electrodeposits of Ni on Cu [11, 99]. Vapor deposits of Cu on Ni showed an h_c of 0.8 nm [100]. The calculated values using both the energy balance and force balance viewpoints yield nearly equal values of near 1.2 nm, which are somewhat high [14, 59]. McWhan reported on h_c measurements of 4.0 nm for the case of a Cu/Ni superlattice [101]. This value for the Cu/Ni superlattice is about four times that for the single layer, a result in agreement with calculation [15, 57]. We note here that the measured values are less than the calculated ones. The misfit in the Cu/Ni system is about 2.6%.

Systems with larger and smaller misfits have also been studied. Thin films of gold on silver which have a misfit of 0.2% have been found to lose coherency when the overgrowth is around 28 nm [10]. This value agrees closely with the calculated value of 24 nm [59]. Two large-misfit systems have been studied. The misfit between films of γ-iron and a (111) gold substrate [102] is about 14% and the measured value of h_c is near 0.1 nm, which is too low by a factor of three. However, at these thicknesses the model is known to become somewhat inaccurate [59]. The Ag/Cu system with a misfit of abot 12% was studied, but instead of a measurement of h_c, the elastic strains were measured as a function of film thickness down to a value of about 0.4 nm [103]. The deposit was not continuous at this thickness but the islands were over 1 μm wide and hence approximated a two-dimensional overgrowth very well. Coherent growth was not observed. The functional dependence fit well the model of Ball [104] except for very small thicknesses where the van der Merwe calculation fits better [56]. The model of

Ball is based on a parabolic interaction potential while that of van der Merwe is based on a sinusoidal interaction potential. The agreement between the measured and calculated elastic strains may be taken as agreement between the critical thickness approach and experiment. Another way of comparing data to theory is to look at the calculated value of critical misfit for a continuous deposit of monolayer thickness. The calculation by van der Merwe [54, 56] gives the equilibrium value of this critical misfit as about 7% for average atomic bonding. At this misfit the introduction of MD's still requires an activation energy which only vanishes at about 10% misfit depending on relative bond strengths [54]. From this point of view the Fe/Au results that show coherency at 14% misfit suggest that the adatom–substrate bonding is strong in comparison with adatom–adatom bonding as is implied by eqs. (19). For Ag/Cu the agreement likewise suggests bonding of about equal strength. There is another observation related to this critical misfit value. The Pd/Cu and Cu/Pd systems were found to be coherent in the initial stages of growth [105]. The misfit for these deposits is about 7%, a value near the critical misfit for a monolayer and hence in agreement with calculation. Only fair agreement with this calculated value of critical misfit is obtained from antimony deposits on tungsten [106]. Here the misfit is just near 8% and the critical thickness for loss of coherency was measured to be three monolayers for (100) and (211) oriented substrates. No coherent growth was observed for (110) oriented substrates. This is likewise consistent with weak bonding (smaller amplitude W_0 in eq. (20)) on the smoother (110) b.c.c. tungsten surface.

The trends indicated by the above comparisons between measured and calculated values of critical thickness and critical misfit for metal systems is that the agreement is as good as one could expect from the crude models employed. In view of the lack of precision in the data and the approximations involved in the models, the case of metal epitaxial growth is adequately described by equilibrium considerations. The measured and calculated values of equilibrium elastic strain agree less well than critical thickness and misfit comparisons between measured and calculated values [14]. But even there the disagreement is not too severe. Significant discrepancies do occur, however, when one makes similar comparisons to those above but for the case of non-metals [14, 107]. The significant differences between the two cases can be understood in terms of the availability and mobility of MD's. The metal case has relatively higher values of both as compared to the non-metal case.

The growth of GaAs on substrates of GaAs doped with indium showed a measured value of h_c near 0.7 μm [108]. The misfit for this system was measured to be 3.2×10^{-4} which yields a calculated value of h_c from the force balance equation of Matthews [14] of 0.59 μm. The measured value exceeds the calculated value by an acceptable margin in this case. Other experiments that favorably compare with calculated values of h_c may be found in some examples of superlattice growth. In a superlattice of PbTe/PbSnTe where the composition of the PbSnTe corresponds to a misfit of 0.39% a value of h_c near 50 nm was measured. This compares favorably to the value of near 78 nm calculated from an

equilibrium minimum energy balance approach [57]. A superlattice of GaAs and InGaAs for which a misfit of 1.9% exists showed a measured value of h_c as 18 nm [109]. The calculated values for this case are 25 nm from the force balance approach [15] using the Volterra model and 9 nm from the equilibrium approach [57], using a parabolic interaction potential. Other examples of reasonable agreement between measured and calculated h_c values for superlattices could be quoted for non-metals [15, 16]. It is instructive, however, to look at those cases that do not fit the equilibrium calculations.

Silicon epitaxially grown at 570°C on GaP substrates has a misfit of 4.6×10^{-3} at 570°C and of 3.6×10^{-3} at room temperature and a measured value of h_c of 100 nm [110]. The calculated value using the force balance approach is 33 nm [14]. An order of magnitude discrepancy was found for the case of germanium deposited at 350°C on GaAs substrates having a misfit of 0.7% and a measured h_c of 2 μm [107]. The authors showed that smaller values of residual strain and smaller h_c values, close to the calculated values of h_c, were observed when the specimens were annealed at 600°C for 30 min. Similar discrepancies of the magnitude found for Ge on GaAs were observed in magnetic garnet [27, 111]. In this case the nearly dislocation-free substrates and high nucleation barrier to MD generation led to relaxation of the MS through cracking of the epilayer on cleavage planes normal to the interface [27].

Two further examples of non-metal layers deviating substantially from the calculated critical thickness values are provided by InGaAs grown on InP substrates [112] and by SiGe alloys on Si substrates [75]. In the former case a number of thickness–misfit combinations were investigated for the presence or absence of MD's. A map of the data points was constructed from which several deductions can be made. The dependence of h_c on misfit is very steep for both positive and negative misfit values. The data show some scatter, but for a thickness of 4 μm a critical misfit of 1.2×10^{-3} is observed and for a misfit of 2×10^{-3} a critical thickness of less than 1 μm occurs. They showed that cross hatch which loosely correlates with MD's can be imaged several micrometers into the InP substrate showing that MD's penetrate to great depths in the substrate but are not present in the InGaAs deposit. The above values of misfit–thickness pairs are not in agreement with the calculated values on the basis of equilibrium principles, the former being almost an order of magnitude greater than the calculated h_c values. This discrepancy coupled with the very strong dependence of h_c on misfit suggests that the equilibrium model is not applicable to the behavior of this system. A similar discrepancy of an order of magnitude between the observed and calculated h_c values for SiGe alloys on silicon substrates also suggests a model based on other than equilibrium principles.

It is clear from the model of People and Bean [75], and subsequent review by Bean [113] as well as from the success of the incorporation of a thermally activated process in the model [107], that one must include the nucleation and mobility of MD's in an analysis of a misfit accommodation by MD's in non-metal systems. Cabrera was the first to recognize the importance of the barriers to MD acquisition [71]. It is this problem that is receiving renewed attention. The only true nucleation theory approaches that incorporate thermal activation for generat-

ing MD's are those already mentioned [52, 78, 107]. Two of these approaches, however, have not included the effect of the mobility of dislocations once they have been nucleated [52, 78]. A successful model must incorporate the three basic components, nucleation, growth and the ultimate equilibrium state being approached.

5. Discussion and conclusions

The existence of a natural misfit f between the two crystals of an epicrystal pair AB has important fundamental and technological implications. The misfit can be accommodated in several modes of which misfit strain (MS; $\bar{e}$) and misfit dislocations (MD's; $\bar{f}$) are the most prominent. MD's and MS may coexist ($f \approx \bar{f} + |\bar{e}|$) and knowledge of the criteria for the coherency–incoherency transition (C–IC: $\bar{f} = 0 \rightarrow \bar{f} \neq 0$) is highly desirable. Because of continuity considerations it has been mathematically expedient to study the inverse transition (IC–C; $\bar{f} \neq 0 \rightarrow \bar{f} = 0$). For a given overlayer thickness h the transition occurs at a critical misfit f^c, and conversely, for a given misfit, at a critical thickness h^c. The objectives of this paper were to describe (i) the principles governing these quantities, (ii) the criteria adopted for determining their values and (iii) the models used in the calculations; and also to assess (iv) the successes of (i)–(iii) as judged by their correlation with experiment. Clearly the processes effecting the transitions are given by the free-energy gradients towards the minimum of free energy (A) defining the ultimate equilibrium. Because of its dominance in the free energy (A), A is usually approximated by the energy E; thus effecting significant simplification. There exists a hierarchy of energy barriers impeding progress and the system may thus linger in a metastable configuration for time spans depending on the height U of the barrier and the thermal energy (kT_s) available. An observation may accordingly find the system in a stable or a metastable configuration. The degree of equilibrium affects two entirely different though basic aspects: (i) the growth mode; ML-by-ML ≡ 2D and islands ≡ 3D, and (ii) the status of the C–IC transition.

Because of their simplicity most calculations in the past have been carried out for 2D growth at equilibrium. Materials that grow 3D at equilibrium can be coerced by non-equilibrium conditions into growing layerlike [48, 49]. This provides some credibility for the practice of calculating critical parameters for systems of uniform thickness. The critical misfit and thickness obtained from equilibrium criteria will be denoted by $f_c(h)$ and $h_c(f)$ respectively, or simply by f_c and h_c. Equilibrium theories assume the unimpeded acquisition of MD's and can be analysed by either energy minimization (eqs. (13a) and (13b)) or equivalently by balancing forces (eq. (13c)). The latter criterion usually assumes that crystal dislocations, specifically threading dislocations, are already present. Energy minimization has often in the past been approximated by (eq. (15)) equality of MS energy (at $|\bar{e}| = f$) and MD energy (at $\bar{f} = f$). This assumption may however, overestimate critical parameters by as much as 30%.

The fundamental (thermally activated) processes affecting growth (adatom

diffusion) and C–IC transitions (MD nucleation and motion) are entirely different. The activation energy (U) for the latter are normally vastly in excess of those of the former, but decrease in proportion to the layer thickness as a consequence of the MS induced Peach–Koehler forces. $U(h)$ ultimately vanishes at $h_s(f) > h_c(f)$, whereafter MD's are formed spontaneously. When h approaches h_s from below, U becomes so small that the frequency with which the barrier is surmounted with the aid of thermal energy is large enough for MS relief to become observable. Clearly the rate of equilibration for growth shapes (involving adatom surface migration) and C–IC transitions (involving the introduction of MD's) are not coupled.

Most models employed in the past are of phenomenological nature and have in common that they treat the crystals individually (A–A and B–B interactions) in the harmonic approximation as linear elastic isotropic continua. Mainly two models have been used for the A–B interactions: (i) the Volterra model (VM), exploited by Matthews, that assumes the relative displacement u of atomic planes on either side of the interface to be uniform and equal to $\boldsymbol{b}$ (the Burgers vector) and (ii) the Frenkel–Kontorova model (FKM) that adopts an interaction which varies periodically in u with periodicity b and amplitude W approximately proportional to the AB bond strengths in contrast to the VM where the variation is ignored and $u = b$ throughout the interfacial region between consecutive MD's. Clearly the FKM is more realistic. Whereas the FKM is solvable with the more acceptable sinusoidal variation only for a monolayer (ML) and on a thick substrate and for two "thick" halfcrystals, and for *all* thicknesses only when the less acceptable parabolic representation is used, the VM is applicable to all thicknesses and can handle imperfect MD's, partials and the new surface which is created in the MD acquisition process. The main shortcoming of the VM is its inaccuracy at very small thicknesses: its strain field representation is inadequate and it does not provide for a variable A–B bond strength which is not negligible at small thicknesses.

More recently the phenomenological models have been supplemented by *ab initio* calculations using appropriate potentials. Whereas the continuum models have general qualitative and semiquantitative predictive power the ab initio calculations are limited to the material combinations considered. A prerequisite for *ab initio* calculations is to find accurate interaction potentials. It is foreseen that these calculations which have more quantitative significance will generate new insights particularly for systems with one or more dimensions of small atomic order where the applicability of the continuum models is questionable.

Experimental observations on metals are in reasonable agreement with equilibrium predictions of $f_c(h)$ and $h_c(f)$ and also yield acceptable evidence of the influence of bond strength, specifically the AB bond strength. The situation is somewhat different for strong covalently bonded semiconducting crystals. While some show agreement with equilibrium predictions others yield values of h^c greatly in excess of $h_c(f)$. It is known that the mobility of dislocations is seriously impeded by the relatively high Peierls barriers in some of them, e.g. Si and its alloys. The theoretical supposition that the barrier height may be reduced by MS

induced Peach–Koehler forces to the point where MS relief becomes observable in the presence of the appropriate thermal energy (kT_s, $T_s \equiv$ substrate temperature), seems to be confirmed. The nucleation probability of fresh MD loops is practically zero in an otherwise perfect crystal, and they must come from a "source" of some kind when they occur.

It may be concluded (i) that the phenomenological continuum theories have played a major role in understanding the physical principles underlying the observed critical parameters and will continue to do so, (ii) that *ab initio* calculations will play an increasing role in obtaining more quantitative understanding, and (iii) attempts at solving the dynamics of MS relief, for which there is a great need, will be more frequent.

Acknowledgements

We gratefully acknowledge financial support from the Foundation for Research Development of the South African Council for Scientific and Industrial Research and also from NASA under grant NAG-1-350.

References

To follow . . .