Energetics of misfit- and threading-dislocation arrays in heteroepitaxial films

A. Rockett
The Coordinated Science Laboratory, The Materials Research Laboratory, and the Department of Materials Science and Engineering, University of Illinois, 1101 West Springfield Avenue, Urbana, Illinois 61801

C. J. Kiely
Department of Materials Science and Engineering, University of Liverpool, P.O. Box 147, Liverpool, L69 3BX, United Kingdom
(Received 15 August 1990; revised manuscript received 6 December 1990)

A theory relating the separation of misfit dislocations to lattice mismatch and film thickness in heteroepitaxial thin films is presented. From this, the energy as a function of dislocation spacing is calculated and is shown to include an attractive and repulsive region. The dislocation-formation energy and Peierls barrier to network ordering are shown to be estimable on the basis of measured dispersions in dislocation spacings. The spacing is predicted to be more uniform as the mismatch increases. Thermodynamic functions, such as the compressibility of the dislocation network, can be calculated from the energy–dislocation-spacing relationship. A formula relating the equilibrium dislocation spacing to film thickness, mismatch, and misfit-dislocation character is also derived. Finally, the density of threading dislocations is calculated both at the heterojunction and at the film surface, by assuming a threading-dislocation reaction process. The results are shown to be in good agreement with the experimental data for Si_xGe_{1-x}/Si, InSb/GaAs, and $In_xGa_{1-x}As/GaAs$ structures.

I. INTRODUCTION

One of the principal limitations to progress in the application of semiconductor heterojunctions in state-of-the-art devices is control of the threading-dislocation density in the layers. This is especially important for films grown epitaxially on lattice-mismatched substrates. Improved methods for reducing the density of threading dislocations in the film are waiting for the development of a comprehensive fundamental model for the formation and alignment of such dislocations. Theoretical studies have approached the problem of developing a relationship between the critical thickness at which dislocations form and the bulk lattice mismatch in two ways. The first involves balancing the force required to form an array of noninteracting dislocations against the strain-induced stresses in the film.[1,2] The second method minimizes the total energy which is a sum of the energy of noninteracting dislocations and the lattice strain energy.[3,4] These theories appear to produce conflicting results. However, a recent work by Willis, Jain, and Bullough[5] reconciled these approaches, showing them to produce similar results, while, at the same time, including the possibility of dislocation interactions. It is clear from the fact that ordered dislocation arrays are produced that the interaction energy of dislocations in a misfit array must be significant. This was clearly demonstrated in recent results[6] in which 60°-type misfit dislocations in a Si_xGe_{1-x}/Si heterojunction were pushed apart by new dislocations entering between preexisting ones.

In this paper we present an alternate derivation of the energy minimization formalism which includes the dislocation interaction energies. It results in an explicit relationship giving the equilibrium dislocation separation D as a function of the film thickness h and the bulk lattice mismatch for any character of dislocation. The result is shown to yield a critical thickness which is qualitatively identical to that of Matthews and Blakeslee but is based on the more straightforward energy minimization approach. Thus we confirm the qualitative similarity of the two previously reported approaches. We then proceed to use the formalism developed to calculate the excess energy available for driving dislocation formation or ordering as a function of array spacing. We estimate the allowed variations in dislocation separation in a net, and calculate the rate of introduction of threading dislocations into the film as misfit dislocations are added to remove remnant strain.

II. MODEL

In this section we derive a relationship among total energy, lattice misfit, and misfit-dislocation spacing. The derivation is based on a uniform network of dislocations with an average spacing D in a film of thickness h. Multilayer films are planned to be dealt with in future papers. The spacing is assumed to be the same in the two orthogonal directions of the lattice defined by the misfit-dislocation lines. Local deviations from uniformity are considered in Sec. III B. The dislocations are further assumed to be a single type; for example, all 60° type[7] or all pure edge type. The coordinate system used in the following derivation is defined in Fig. 1.

It is convenient to consider a unit of surface area (in which the energies are calculated) to be a square of side D. Thus the unit surface area always contains a length of

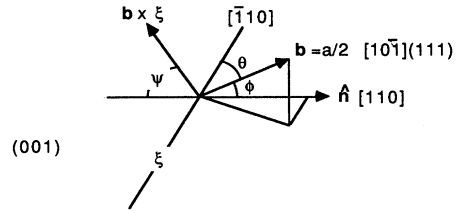

FIG. 1. A schematic diagram giving the definition of the dislocation line direction ξ and Burgers vector **b** with respect to the interfacial plane as well as the angles referenced in the text.

dislocation line $2D$. The choice of boundaries is such that they are high-symmetry lines in the interface and the area represents a unit cell of the surface. Furthermore, the amount of dislocation line need not be calculated in a fixed area, which simplifies development of the subsequent results. The total energy of the system for a square substrate of side L_t is given by the energy per cell E_c multiplied by the number of cells $n=L_t/D$ squared, or

$$E_{\rm tot}=\frac{L_t^2 E_c}{D^2} . \tag{1}$$

The equilibrium dislocation spacing is found by minimizing E_c/D^2 with respect to D. For a two-dimensional planar misfit-dislocation net with a uniform spacing D, E_c is related to the strain energy E_s and the dislocation net energy per unit-cell area by

$$\frac{E_c}{D^2}=\frac{E_s}{D^2}+\left[\frac{2D}{D^2}\right]\frac{E_d}{D}=\frac{E_s}{D^2}+\frac{2E_d}{D^2} , \tag{2}$$

where E_d is the energy of the two dislocations in the area D^2 and includes interactions between dislocations in adjacent unit cells. The energy associated with the intersection of the two orthogonal dislocations in the unit cell is ignored in this treatment. For pure edge-type dislocations this is a reasonable assumption. However, the interaction energy between 60° dislocations with orthogonal line directions may be more important. The interactions among 60°-type dislocations are discussed in Ref. 7.

The fractional misfit between the film and substrate lattices, with bulk lattice constants a_f and a_s, respectively, is defined as

$$f_\infty=\frac{2|a_f-a_s|}{a_f+a_s}=\frac{\Delta a}{a} . \tag{3}$$

For the current analysis the misfit is a variable which depends on the spacing D and Burgers vector **b** of misfit dislocations in the interface. The contribution to misfit relief perpendicular to the dislocation lines is proportional to the projected edge component $b\cos\phi$ of **b** in the plane of the interface as viewed along the dislocation line. The fractional misfit relieved is $b\cos\phi/D$ and ϕ is defined as shown in Fig. 1. The total misfit as a function of dislocation spacing is then given by

$$f(D)=f_\infty-\frac{b\cos\phi}{D} . \tag{4}$$

For a fractional misfit $f(D)$, the strain energy per unit area in the film is given by[8]

$$\frac{E_s}{D^2}=2\mu h\,[f(D)]^2\frac{1+\nu}{1-\nu} , \tag{5}$$

where μ is the shear modulus and ν is the Poisson ratio of the film. This formula accounts for the deformation of the film normal to the interface, and assumes that all of the strain energy occurs in the film. A small distortion in the substrate atom positions also occurs which is generally much less than in the film. For roughly equal shear moduli in the film and substrate, the strain energy in the substrate is less than the strain energy in the film, to first order, by the ratio of h/t. Here, t is the substrate thickness. Thus for thin single layer films the assumption of no elastic energy in the substrate is valid.

The energy of a dislocation line per unit length is found by summation of the self-energies and interaction energies of dislocations at a separation of D and with length L. The self-energy per length L for any type of dislocation is[9]

$$E_1=\frac{\mu b^2(1-\nu\cos^2\theta)}{4\pi(1-\nu)}L\ln\left[\frac{\alpha L}{b}\right] . \tag{6}$$

The interaction energy for two parallel dislocations of arbitrary character is

$$E_i=\frac{\mu b^2(1-\nu\cos^2\theta)}{2\pi(1-\nu)}L\ln\left[\frac{D}{L}\right]+L\frac{\mu b^2\sin\theta\cos\Psi}{2\pi(1-\nu)} , \tag{7}$$

where θ and Ψ are defined in Fig. 1. The total dislocation energy is then given by $E_d=2E_1+E_i$. From Eqs. (6) and (7) taking the length of the dislocations to be D,

$$\frac{E_d}{D}=\frac{\mu b^2}{2\pi(1-\nu)}\left[(1-\nu\cos^2\theta)\ln\left[\frac{\alpha r}{b}\right]+\sin\theta\cos\Psi\right] , \tag{8}$$

where r is the effective distance to which the strain field extends and α is a correction term of order unity which accounts for the core energy of the dislocation. For pure edge dislocations, $\Psi=90$. A range of values of α from 1 to 4 have been used previously. We have followed Dixon and Goodhew by selecting $\alpha=3$.[7]

When the dislocations are separated by a large distance $D\gg h$, the strain field is assumed to terminate at the nearest film surface, dislocation interactions are insignificant, and $r=h$. The assumption of a cutoff in strain field at a distance h is reasonable if the shear moduli of the film and substrate are nearly equal. The image force introduced by a dislocation array of opposite type located outside the surface by a distance h nearly eliminates the strain field of the interfacial dislocations far into the substrate. The strain field will include a contribution from the film and a contribution from the substate. Thus the value of μ should be a weighted average of the film and substrate values. For the model presented here the strain field deep in the substrate is assumed to be

eliminated by the image forces at the nearby surface. When the shear moduli of the film and substrate are significantly different the strain field does not cancel and extends to the farthest boundary of the crystal. The calculation to estimate the error is analogous to solving Poisson's equation for a charge at the interface between two materials of different dielectric constant. The error enters only logarithmically and is ignored in this treatment.

When D is less than h the strain field interaction dominates the dislocation energy and the crystal dimension term drops out of the resulting energy expression. Equation (8) remains unchanged with $r=D$. For a dislocation net the interaction energy per unit length is doubled by interaction with dislocations in both adjacent unit cells. However, only half of this energy is distributed to each dislocation in the unit cell with the other half going to the dislocation with which the interaction occurs. Thus the energy per unit length of a dislocation in a regular array of spacing $D \ll h$ is given by

$$\frac{E_d}{D}=\frac{\mu b^2}{2\pi(1-\nu)}\left[(1-\nu\cos^2\theta)\ln\left[\frac{\alpha D}{b}\right]+\sin\theta\cos\Psi\right]. \tag{9}$$

Because there are two orthogonal dislocation nets present in the sample the total amount of dislocation energy is doubled [see Eq. (2)]. Equation (8) and the corresponding expression (9) are in units of energy per length D rather than in energy per unit area D^2. The units are corrected by multiplying both sides of these equations by $2/D$ because a length of dislocation $2D$ is contained in an area D^2. This dislocation line energy term can be combined with Eq. (5) to yield the total strain-related energy of the bilayer,

$$\frac{E_c}{D^2}=2\mu h f(D)^2\frac{1+\nu}{1-\nu}+\frac{\mu b^2}{\pi D(1-\nu)}\left[(1-\nu\cos^2\theta)\ln\left[\frac{\alpha r}{b}\right]+\sin\theta\cos\Psi\right], \tag{10}$$

where $r=h$ for $h \ll D$ and $r=D$ for $D \ll h$. The equilibrium dislocation spacing is found by minimizing this energy per unit area with respect to D using $d(E_c/D^2)/dD=0$. Differentiating Eq. (10),

$$D^2\frac{d(E_c/D^2)}{dD}=4\mu h b\cos\phi\frac{1+\nu}{1-\nu}f(D)-\frac{\mu b^2}{\pi(1-\nu)}\left[(1-\nu\cos^2\theta)\ln\left[\frac{\alpha r}{b}\right]+\sin\theta\cos\Psi\right]+\frac{\mu b^2(1-\nu\cos^2\theta)}{\pi(1-\nu)}\frac{D}{r}\frac{dr}{dD}. \tag{11}$$

Setting this equation to zero and solving for h yields

$$h=\frac{b}{4\pi(1+\nu)\cos\phi}\frac{(1-\nu\cos^2\theta)\ln(\alpha r/b)+\sin\theta\cos\Psi-(D/r)dr/dD}{f(D)}. \tag{12}$$

For $D\to\infty$, $r=h$, and $f(D)=\Delta a/a$, Eq. (12) is qualitatively identical to that of Matthews and Blakeslee[1,2] for the critical thickness at which strain relief occurs but represents an equilibrium strain relief condition. The major differences between Eq. (12) and Matthews and Blakeslee's Eq. (5) in Ref. 1 are a factor of 2 due to the differentiation of $[f(D)]^2$ and the terms involving $\sin\theta\cos\Psi$ and $(D/r)dr/dD$. The former results from dislocation interactions while the latter is due to the effective cutoff of the strain field at the nearest crystal boundary. To this point we have verified the results of Willis, Jain, and Bullough[5] showing that the result of Matthews and Blakeslee[1,2] is qualitatively identical to that obtained by Frank and van der Merwe[3,4] using a total energy minimization and including dislocation interactions.

A more useful result is obtained by solving the energy minimization for $D(h)$. Substituting for $f(D)$ from Eqs. (3) and (4),

$$D(h)=\frac{b\cos\phi}{\Delta a/a-\zeta p/h}, \tag{13}$$

where

$$\zeta=\frac{b}{4\pi(1+\nu)\cos\phi} \tag{14}$$

is a constant and

$$p=(1-\nu\cos^2\theta)\ln\left[\frac{\alpha r}{b}\right]+\sin\theta\cos\Psi-\frac{D}{r}\frac{dr}{dD}. \tag{15}$$

Note that Eq. (13) can be rewritten as

$$D(h)=\frac{D_\infty}{1-h_c/h}, \tag{16}$$

where $h_c=\zeta p/(\Delta a/a)$ is the equilibrium critical thickness for the onset of strain relief [from Eq. (12) with $r=h$ and $f(D)=\Delta a/a$]. For $h\to\infty$ Eq. (16) yields $D_\infty=b\cos\phi/(\Delta a/a)$, as it should. When h is small, $D(h)$ diverges at the critical thickness. Note that p is either $(1-\nu\cos^2\theta)\ln(\alpha h/b)+\sin\theta\cos\Psi$ or $(1-\nu\cos^2\theta)\ln(\alpha D/b)+\sin\theta\cos\Psi-1$ depending on the value of r. When $h \ll D$, $r=h$ and p is increasing with h. A maximum in p occurs in the range where $h\sim D$ above which p decreases to $\ln[\alpha\cos\phi(\Delta a/a)]+\sin\theta\cos\Psi-1$ at large h. In general p will lie in the range of 1–7 with smaller values for higher misfit systems. Variations in p thus represent minor, slowly varying corrections to the dislocation spacing.

The expression for $D(h)$ can be used to calculate the rate at which misfit dislocations are introduced into the

film as a function of h. In a region of wafer surface of dimension L there will be $n=L/D$ dislocations. The rate of addition of dislocations in a length of surface L as the film thickness changes is given by $d(n/L)/dh = dD^{-1}/dh = -D^{-2}\,dD/dh$. Differentiating Eq. (13) with respect to h yields

$$\frac{dD}{dh} = b\zeta\cos\phi\,\frac{(1/h)dp/dh - p/h^2}{(\Delta a/a - \zeta p/h)^2} \quad . \tag{17}$$

For $h \gg h_c = \zeta p/(\Delta a/a)$, $dp/dh = 0$. Substituting from $dD/dh = D^2 d(n/L)/dh$,

$$\frac{d(n/L)}{dh} = \frac{b\zeta p\cos\phi}{D^2(h\Delta a/a - \zeta p)^2} \quad . \tag{18}$$

The second term in the parentheses is small for $h \gg h_c$. Assuming that $D \sim b\cos\phi/(\Delta a/a)$,

$$\frac{d(n/L)}{dh} = \frac{\zeta p}{h^2 b\cos\phi} \tag{19}$$

near the end of the strain relief process. When $h \ll D$, $dp/dh = h^{-1}$, and

$$\frac{d(n/L)}{dh} = \frac{\zeta(p-1)}{b\cos\phi[h - \zeta p/(\Delta a/a)]^2} \quad . \tag{20}$$

It should be noted that these formulas are for an equilibrium dislocation spacing. Variations from equilibrium are considered in the next section.

III. DISCUSSION

The following assumes a heterojunction grown on a (001) surface of a diamond or zinc-blende semiconductor. It does not account for the possibility of misfit dislocations extending out of the plane of the interface. This could be driven, for example, by attraction to the surface due to image forces or by large differences in shear modulus across the interface. The interaction between dislocations will depend on their Burgers vectors and line directions. When the majority of dislocations are of the 60° type, a large number of possible interactions result. Hence the following analysis is valid for locally homogeneous regions of the dislocation net.

A. Nonequilibrium dislocation spacing

As strained layer epitaxy begins with $h < h_c$ there is no physical solution of Eq. (16) and the total energy of the system is increased by adding misfit dislocation at any spacing. Several possibilities exist for misfit relief as the layer growth proceeds beyond $h = h_c$. When the misfit is small and growth is two dimensional, 60°-type dislocations are introduced by extension of existing threading dislocations,[1,6] by operation of a threading-dislocation source, or by expansion of loops by glide on {111} planes between the heterointerface and the surface. When the misfit is large and growth is three dimensional, perfect edge dislocations may be formed at the island perimeters. These glide along the interface to form the misfit network. An intermediate case must occur if the edge dislocation propagation process has only partially relieved the ultimate misfit ($D > D_\infty$) when coalescence occurs. In this case the remainder of the strain relief is carried out by 60°-type dislocations as in the low-strain case. In general, some pure edge-type dislocations and some 60° dislocations which are non-strain-relieving are observed. The model presented in Sec. II is applicable to regions of the interface which are relatively homogeneous in dislocation type. The introduction of, for example, edge-type dislocations into a 60°-type dislocation net will introduce local perturbations in the net.

We examine first the case where growth is two dimensional and strain relief is almost entirely by 60°-type dislocations. This is the case for the $In_xGa_{1-x}As/GaAs$ and Ge_xSi_{1-x}/Si heterostructures. When $h > h_c$, thermodynamics favors the introduction of dislocations to relieve strain but a kinetic barrier for formation and/or propagation of the dislocations remains.[10,11] When h rises sufficiently above h_c, preexisting threading dislocations can be displaced to produce misfit-dislocation segments.[1] Because the dislocation segments are present in the film, only the energy to extend the dislocation through the film, no formation energy, is required. The energy for such dislocation extension can be estimated from Eq. (10) using the initial thickness of films at which threading-dislocation extension is observed.

Experimental values for dislocations spacing as a function of misfit are available for the $In_xGa_{1-x}As/GaAs$ system for $x = 0.10$, 0.15, 0.20, and 0.25.[7] We have estimated the critical thicknesses as a function of misfit based on these data. The procedure was as follows. The best fit was obtained to the data using Eq. (16) with D_∞ calculated based on bulk lattice parameters. The value of h_c was allowed to vary as a fitting parameter. The calculated curve was found to lie below the observed dislocation spacings. This is due to the material being in a metastable excess strain state,[10] $D_{x\to\infty} > b\cos\phi/(\Delta a/a)$. The fit value of h_c was compared with values calculated from Eq. (12). Critical thicknesses, fitted to experimental values of $D(h)$ for two orthogonal directions in the interface, and calculated values are presented in Table I. The calculated values of h_c are in good agreement with the fitted values for most of the experimental data points especially in the case of the [110] direction. These critical thicknesses correspond to strain energies of 0.65 and 0.82 eV/atom assuming an atomic density of $\sim 3\times 10^{14}$ atoms cm^{-2} and give a measure of the excess energy required to drive threading-dislocation extension.

Dislocation generation from operation of a point source produces misfit dislocations which must force themselves between those generated in previous cycles of the source. An example of the spreading of preexisting misfit dislocations during the introduction of a new dislocation is shown in Fig. 2. As for the threading-dislocation extension case discussed above, dislocations forcing their way between existing segments require a net excess strain energy to drive their formation. However, for the dislocation to squeeze into the network this driving force must also be sufficient to overcome forces resisting spreading of the network. This is shown schematically in Fig. 3. The preexisting dislocations move apart until the available excess strain energy plus the energy gained by reduction of repulsive interactions is less than

TABLE I. Estimated values for critical thickness for the $In_xGa_{1-x}As/GaAs$ system based on experimental data from Ref. 7.

In fraction x	Misfit $\Delta a/a$	Experimental estimates $h_c([110])$ (nm)	Experimental estimates $h_c([1\bar{1}0])$ (nm)	Calculated from Eq. (12) h_c (nm)
0.10	0.007	45	50	36.5
0.15	0.010	35	40	22.3
0.20	0.014	14	19	15.7
0.25	0.017	11	15	11.7

the Peierls barrier for motion of the outermost dislocation. To force a dislocation into a preexisting net, the strain energy must be sufficient to drive dislocation line creation [from Eq. (9)] *and* to create it in a network of spacing equal to that at which spreading stops. This means that the film must be thicker to introduce dislocations into a network by forcing them between preexisting segments than would be needed for creating isolated segments. Clearly, this can only happen when dislocation nucleation is slow. When the network spreads as a dislocation is introduced we refer to it as being in compression, since the expansion stops before it would if the Peierls barrier was zero. When the network could add dislocations without spreading the preexisting net but is limited by the dislocation nucleation rate, we refer to the network as being in tension.

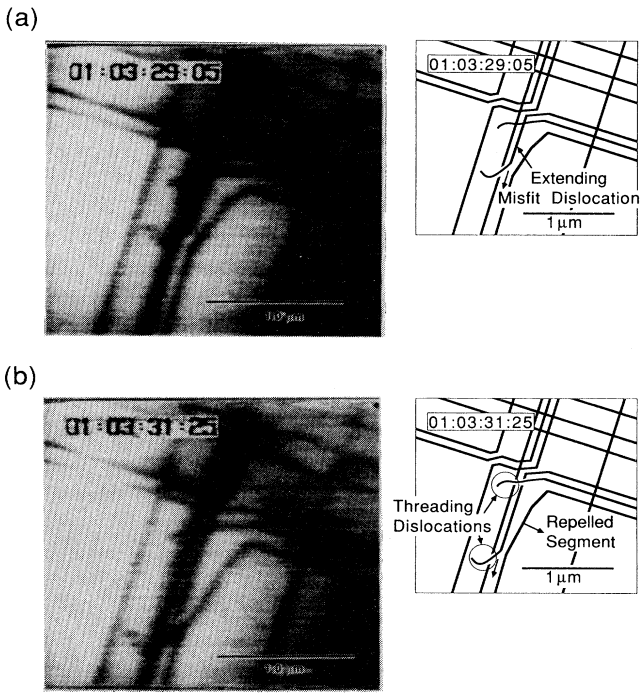

FIG. 2. Two transmission electron micrographs obtained by digitization of video images from Ref. 6. The micrographs show a propagating misfit-threading-dislocation unit at a strained $Ge_{0.2}Si_{0.8}/Si$ heterojunction forcing itself between two preexisting segments of misfit dislocation. The dislocation line segments in the micrographs are shown schematically in the insets. In (a), a misfit dislocation generated by motion of a threading dislocation has interacted with two orthogonal dislocations in passing and has begun to repel a parallel segment. (b), taken 1.2 sec later, shows the parallel segment continuing to move away from the new misfit dislocation due to the strain field interactions. The scale marks are accurate to approximately ±10%.

The energy of a uniform network as a function of D may be calculated from Eq. (10). Representative curves of $E_c(D)/D^2$ are plotted in Fig. 4. The plots clearly resemble the interaction between particles in a lattice. The attractive portion of the curves results from strain relief while the repulsive portion combines the strain field repulsion of the dislocations with strain over-relief. As the mismatch between the materials increases [Fig. 4(a)] the energy minimum becomes sharper and the minimum energy becomes greater. A similar behavior is found as the thickness of the film increases, as shown in Fig. 4(b). When $D=D_\infty$ and $h>D_\infty$, $f(D)=0$ and E_c does not depend on h. Hence all $E_c(D,h)$ curves pass through a common point at $E_c(D_\infty, h>D_\infty)$. Note that for sufficiently thin films there is little difference in network energy as the network expands (there is little strain energy to relieve) and a wide range of dislocation spacings may exist with little penalty in total system energy.

If the energy barrier to network ordering is constant, the dispersion in dislocation spacings will be smaller in

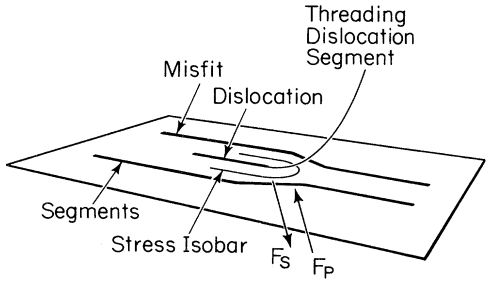

FIG. 3. A schematic diagram showing the forces controlling expansion of a dislocation network as a new dislocation is introduced from a remote point source between preexisting misfit dislocations of the same type. The strain field results in a force F_s repelling surrounding dislocations while the Peierls energy barrier results in a force F_p which opposes dislocation motion. Expansion of the network continues until $F_s=F_p$.

more strongly mismatched systems or in thicker films. However, as the number of dislocations which must be displaced to produce ordering increases the energy barrier is likely to increase. Although Figs. 4(a) and 4(b) are calculated for an infinite, well-ordered dislocation net the primary energy terms will be relatively local. Hence the illustrated behavior is probably correct to order of magnitude even for very small groups of dislocations. The energy barrier to dislocation generation and ordering must be less than the maximum value, from Fig. 4, corresponding to the largest or smallest observed spacings.

The explicit form of $E(D)$ in Eq. (10) makes it possible to calculate thermodynamic properties of the network. As an example we calculate the network compressibility. Defining the pressures as $-dE/dV$, where $V=D^2$ is the network area, the compressibility $\kappa=[-V(d^2E/dV^2)]^{-1}$ may be calculated by differentiating Eq. (10) twice with respect to D^2. This yields

$$\kappa=\frac{D}{b\cos\phi\left\{B+C\left[\ln\left[\frac{\alpha r}{b}\right]+\left[\frac{D}{r}\frac{dr}{dD}\right]+\left[\frac{D}{r}\frac{dr}{dD}\right]^2\right]\right\}}, \tag{21}$$

where

$$B=\mu h\left[\frac{1+\nu}{1-\nu}\right]\left[\frac{\Delta a}{a}\right] \tag{22}$$

and

$$C=\frac{\mu b(1-\nu\cos^2\theta)}{4\pi(1-\nu)\cos\phi} \tag{23}$$

are constants. The compressibility is positive and decreasing with decreasing D at small spacings. This provides a quantitative prediction of the amount by which the network is more resistant to a given amount of compression than to the same amount of extension. If the entropy of the dislocation network is estimated, free energies can be determined and other thermodynamic quantities such as the equilibrium fluctuations in net spacing can be determined.

In heavily mismatched materials more efficient strain relief is possible as pure edge-type dislocations form along the perimeter of three-dimensional islands (pure edge-type dislocations relieve twice as much strain as 60°-type dislocations). Strain relief generally occurs for very thin layers prior to island coalescence. The introduction of dislocations from the film edge results in a network in tension. The dislocations are only added when an excess strain energy sufficient to drive their formation at the film edge is available. This energy is expected to be very low since the dislocation can be formed during the addition of adatoms to the island.[12,13] Thus the dislocation network in these materials can be expected to have nearly perfect spacing at island coalescence [the $E(D)$ curve exhibits a sharp minimum] and to be in relatively uniform tension.

The formation of a pure edge-type dislocation network has been observed for the InSb/GaAs system for which the bulk lattice mismatch is 14.6%.[12] A network within precoalescence islands of 30 nm height was observed to be well ordered and at nearly the spacing required for complete strain relief, as shown in Fig. 5. Furthermore, the dislocation spacing is nearly perfectly uniform to the edge of the islands even though the height of the island is changing rapidly there. The final dislocation spacing for this mismatch at large thickness is only 6.85**b** where **b** is the Burgers vector. Since the actual spacing must be an integral value, relatively large local fractional fluctuations in D will be built into the network in order to achieve nonintegral values of average D. This raises the energy of the real network (analogous to zero-point vibrations in a quantum oscillator). However, the majority of the fluc-

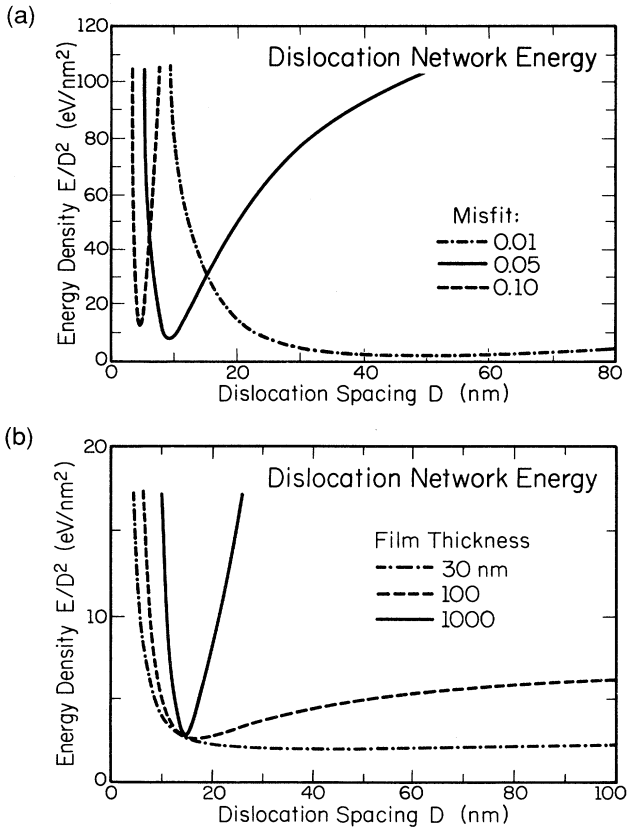

FIG. 4. A plot of the calculated system energy $E(D)$ per unit D^2 area as a function of dislocation spacing D. The energy is normalized by the shear modulus of the film and the Burgers vector of the dislocation. (a) shows the effect of misfit on the $E(D)$ curve. (b) shows the effect of film thickness on $E(D)$.

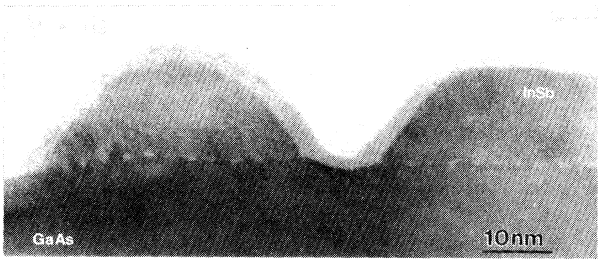

FIG. 5. Cross-sectional transmission electron micrograph showing a lattice image of a 13-nm-high InSb island on a GaAs surface. The zone axis of the image is (110) and fringes in the image are (111) planes. The bright points along the heterojunction indicate the presence of misfit dislocations. The dislocations are pure edge type with an average spacing of one every seven (111) planes.

tuations in D will occur due to misalignment of the networks in adjacent islands upon coalescence. The effects of island coalescence on the network are planned to be considered in a separate publication.

To complete this section, we consider the factors determining the kinetic barrier to misfit relief. In all cases, dislocation motion can be thermally activated. Thus growth of mismatched systems at elevated temperatures increases the rate of strain relief and decreases the remnant strain. A detailed theory relating this temperature dependence to remnant strain has been developed by Dodson and Tsao and may be found in Refs. 10 and 11. When strain relief occurs by loop nucleation, the presence of stress concentrators such as steps and other defects reduces the energy barrier substantially. Thus growth of smoother surfaces and interfaces on relatively well oriented substrates may increase the thickness at which strain relief would occur by loop expansion.

B. Density of misfit dislocations with film thickness

Following the initial strain relief achieved through sources requiring low energies to operate, the relief process is thought to proceed through the formation of dislocation loops. These typically include a segment of misfit dislocation, a surface trace, and two threading segments if the loop does not reach the sample perimeter. The misfit segments must fit into the existing network and are likely to encounter some obstacle to extension before reaching the edge of the crystal. Hence the number of threading dislocations introduced by addition of a segment of misfit dislocation is roughly 2. Assuming this to be the case, the density of threading dislocations in the film can be estimated as a function of thickness from Eq. (19).

The rate of addition of misfit dislocations per unit length (orthogonal to the added dislocation lines) at large thickness is inversely proportional to h^2. The rate of addition of threading dislocations per "unit length" L is twice the rate of addition of misfit dislocations or

$$\frac{d(N_+/L)}{dh}=\frac{2\zeta p}{h^2 b\cos\phi}\ . \tag{24}$$

L represents the average separation of the threading dislocations when the loop ceases to expand and is assumed to be much larger than D and much smaller than the wafer dimension. Clearly, the fewer pinning points in the crystal available to halt loop expansion the fewer threading dislocations will be introduced. The density of threading dislocations in a fixed unit of area will scale roughly as the square of the pinning site density halting loop expansion. It is assumed in the following treatment that the pinning site density is constant and loops expand to a fixed size independent of threading-dislocation density. An alternate assumption would be that the loop dimension scales with threading-dislocation density, which may be more accurate at high dislocation densities near the interface.

The rate of addition of threading dislocations decreases as film thickness increases. However, in the absence of a mechanism for eliminating threading dislocations the total population of threading dislocations in the film would always increase. From transmission electron microscopy evidence it is clear that the density of these dislocations decreases rapidly with distance from the heterointerface.[14] For this to occur it is necessary that threading dislocations react to be eliminated. This is equivalent, for threading dislocations propagating to the surface, to requiring that dislocation trace steps on the surface be eliminated by reaction. A dislocation loop originating at the heterojunction but ceasing to expand prior to reaching the surface would lead to threading dislocations appearing to terminate in the film. This could happen either spontaneously or due to a reaction with another loop. The order of the process of terminating threading-dislocation propagation is given by the linear density of dislocations involved in the termination. If the order of the process for terminating the propagation of threading dislocations is n, the rate of elimination of threading dislocations per unit length $d(N_-/L)/dh$ will be approximately

$$\frac{d(N_-/L)}{dh}=R\left[\frac{N_t}{L}\right]^n, \tag{25}$$

where N_t is the total number of threading dislocations available to interact and R is the probability per unit film thickness for this elimination. The net rate of change of threading-dislocation density at the film surface is thus $d(N_t/L)dh=d(N_+/L)/dh-d(N_-/L)/dh$, or combining Eqs. (24) and (25), for $h\gg h_c$

$$\frac{d(N_t/L)}{dh}=\frac{2\zeta p}{h^2 b\cos\phi}-R\left[\frac{N_t}{L}\right]^n. \tag{26}$$

Near the heterojunction little interaction of threading-dislocation segments has occurred and few are eliminated. There can be no elimination of threading dislocations very near the interface as this would also necessarily eliminate the strain relieving dislocation. Thus the relative recombination probability R decreases to zero near the interface. Equation (26) with $R=0$ is exactly soluble and yields

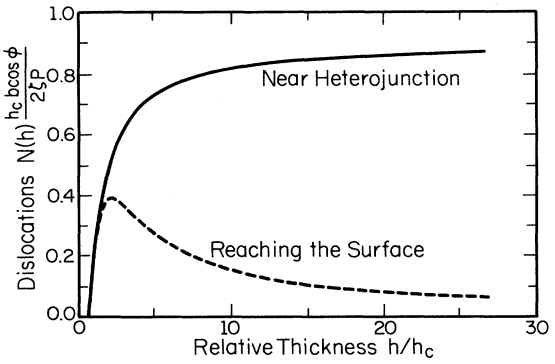

FIG. 6. Threading-dislocation densities at the heterojunction where no dislocation reaction has occurred and at the film surface calculated by solution of Eq. (26) assuming a second-order dislocation recombination process, $n=2$.

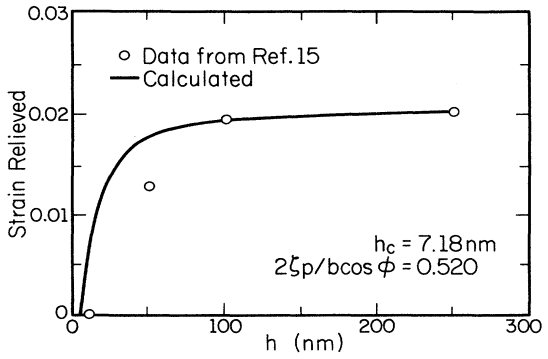

FIG. 7. Shows the calculated strain relieved (which is proportional in this model to the threading-dislocation density at the heterojunction) and the experimentally observed values from Ref. 15 for a $Si_{0.5}Ge_{0.5}$/Si(001) structure.

$$\frac{N(h)}{L}=\frac{2\zeta p(1-h_c/h)}{h_c b\cos\phi} \tag{27}$$

at the interface. This threading-dislocation density always increases as the film grows.

When $h \gg h_c$ the rate of addition of dislocations becomes nearly zero. In this limit, for finite R, Eq. (26) reduces to $d(N_t/L)/dh=-R(N_t/L)^n$, yielding

$$\frac{N_t}{L}=[R(h-h_c)(n-1)]^{1/(1-n)} \tag{28}$$

for $n>1$, at the film surface. When $n=1$,

$$\frac{N(h)}{L}=e^{-Rh}. \tag{29}$$

The threading-dislocation density at the film surface initially rises as dislocations are formed but then decreases as the rate of threading-dislocation removal becomes greater than the rate of dislocation introduction. The dislocation distributions at the surface and near the heterojunction are plotted in Fig. 6 assuming $n=2$. The distribution between these two limits will fall between the plotted curves and will decrease according to Eq. (28) at large h. The progression of curves from the surface limit to the heterojunction limit ($R=0$) depends on the way in which R changes with distance from the interface and may be affected by changes in L. There are undoubtedly a large number of mechanisms by which threading dislocations can interact and cease to propagate as the film grows. The exponent of N_t in Eq. (25) depends on the order of the interaction while R is an empirical constant which accounts for all of the reactive mechanisms. Thus both R and the exponent N_t will be difficult to calculate *a priori* without a detailed understanding of the mechanisms by which threading dislocations are eliminated. However, by fitting experimental data it should be possible to determine both R and the exponent n. With information concerning R and n, a dominant mechanism of threading-dislocation formation and removal could be deduced.

Two variables of strain relief are easily observed in mismatched systems, the total strain relieved (by measuring the film lattice parameter in the interface plane), and the etch-pit density at the surface giving the number of threading dislocations reaching the surface. The latter is probably composed of a set of dislocations propagating from the substrate of essentially fixed quantity plus a number of threading dislocations introduced during strain relief. Hence it should be possible to estimate $N(h)$ from the etch-pit density after subtracting a constant number of intrinsic defects propagating out of the substrate. The number of threading dislocations at the interface will be proportional to the misfit relieved and increases according to Eq. (27) at large film thicknesses.

Based on the misfit relieved $N(h)/L$ at the interface can be estimated. To provide the proper asymptotic behavior, the value of $2\zeta p/(h_c b\cos\phi)$ in Eq. (27) must equal $1/D_\infty$, which is calculable for a given system. The remaining parameter R is determined by fitting the solution of Eq. (26) to $N(h)$ at the surface estimated from etch-pit densities. We have carried out half of this process for $Ge_{0.5}Si_{0.5}$ for which values of fractional misfit relieved are available.[15] The result is shown in Fig. 7. The fit is excellent at large values of h but deviates at lower values for which Eq.(26) is not valid, L is probably not constant, and D is far from D_∞. To determine the recombination mechanism for threading dislocations it will be necessary to have measurements of strain relieved and etch-pit density in a single set of films.

IV. CONCLUSIONS

The energy minimization formalism originally proposed by van der Merwe for calculating strain relief in lattice-mismatched heteroepitaxial systems has been extended to include dislocation interactions. From this the energy as a function of dislocation spacing has been calculated and shown to include an attractive and repulsive region. This permits estimations of the dislocation formation energy and Peierls barrier to network ordering from measured dispersions in dislocation spacings. It

further explains why the dislocation network is better ordered in highly mismatched systems than in weakly mismatched heterostructures. It has been shown that thermodynamic functions such as the compressibility of the dislocation network can be calculated from the energy–dislocation-spacing relationship. A formula relating the equilibrium dislocation spacing to film thickness, mismatch, and misfit-dislocation character is also calculated. Finally, the density of threading dislocations is calculated both at the heterojunction and at the film surface assuming a threading-dislocation reaction process. The interface threading-dislocation behavior is shown to be in good agreement with the fractional strain relieved in Si_xGe_{1-x}/Si structures.

ACKNOWLEDGMENTS

The authors wish to thank P. J. Goodhew for many useful discussions and advice and R. Hull for his comments and experimental data. The authors gratefully acknowledge the support of the Department of Energy under Contract No. DEAC-76-ER01198.

[1]J. W. Matthews and A. E. Blakeslee, J. Cryst. Growth **27**, 118 (1975).
[2]J. W. Matthews, J. Vac. Sci. Technol. **12**, 126 (1975).
[3]F. C. Frank and J. van der Merwe, Proc. R. Soc. London, Ser. A **198**, 261 (1949).
[4]F. C. Frank and J. van der Merwe, Proc. R. Soc. London, Ser. A **198**, 2205 (1949).
[5]J. R. Willis, Suresh C. Jain, and R. Bullough, Proc. Mat. Res. Soc. (to be published).
[6]R. Hull and J. C. Bean (unpublished).
[7]R. H. Dixon and P. J. Goodhew, J. Appl. Phys. **68**, 3168 (1990).
[8]*Epitaxial Growth Part B,* edited by J. W. Matthews (Academic, New York, 1975).
[9]Equations (6)–(9) are developed from more general forms given in John Price Hirth and Jens Lothe, *Theory of Dislocations,* 2nd ed. (Wiley, New York, 1982).
[10]Brian W. Dodson and Jeffrey Y. Tsao, Appl. Phys. Lett. **51**, 1325 (1987).
[11]Brian W. Dodson and Jeffrey Y. Tsao, Appl. Phys. Lett. **53**, 2498 (1988), and references cited therein.
[12]C. J. Kiely, J. -I. Chyi, A. Rockett, and H. Morkoç, Philos. Mag. A **60**, 321 (1989).
[13]C. J. Kiely, A. Rockett, J. -I. Chyi, and H. Morkoç, in *Atomic Scale Structure of Interfaces,* edited by R. D. Brigans, R. M. Feenstra, and J. M. Gibson, MRS Symposia Proceedings No. 159 (Materials Research Society, Pittsburgh, in press).
[14]See, for example, Chin-An Chang, C. M. Serrano, L. L. Chang, and L. Esaki, J. Vac. Sci. Technol. **17**, 603 (1980); or G. R. Booker, J. M. Titchmarsh, J. Fletcher, D. B. Darby, M. Hockly, and M. Al-Jassim, J. Cryst. Growth **45**, 407 (1987).
[15]J. C. Bean, L. C. Feldman, A. T. Firoy, S. Nakahura, and I. K. Robinson, J. Vac. Sci. Technol. A **2**, 436 (1984).

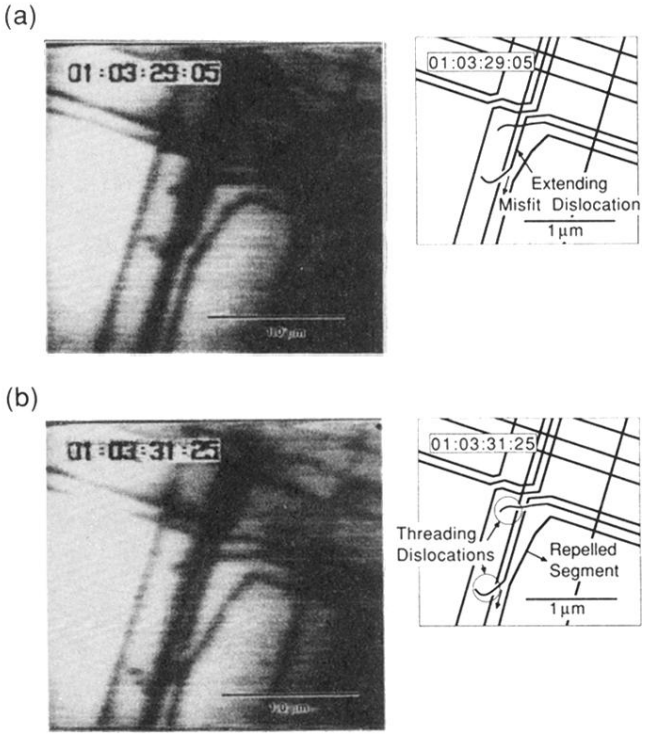

FIG. 2. Two transmission electron micrographs obtained by digitization of video images from Ref. 6. The micrographs show a propagating misfit-threading-dislocation unit at a strained $Ge_{0.2}Si_{0.8}/Si$ heterojunction forcing itself between two preexisting segments of misfit dislocation. The dislocation line segments in the micrographs are shown schematically in the insets. In (a), a misfit dislocation generated by motion of a threading dislocation has interacted with two orthogonal dislocations in passing and has begun to repel a parallel segment. (b), taken 1.2 sec later, shows the parallel segment continuing to move away from the new misfit dislocation due to the strain field interactions. The scale marks are accurate to approximately $\pm 10\%$.

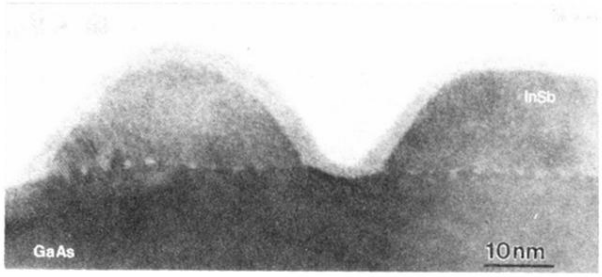

FIG. 5. Cross-sectional transmission electron micrograph showing a lattice image of a 13-nm-high InSb island on a GaAs surface. The zone axis of the image is (110) and fringes in the image are (111) planes. The bright points along the heterojunction indicate the presence of misfit dislocations. The dislocations are pure edge type with an average spacing of one every seven (111) planes.