Critical Reviews in Solid State and Materials Sciences, 17(6):507–546 (1992)

Misfit Dislocations in Lattice-Mismatched Epitaxial Films

Robert Hull and John C. Bean
AT&T Bell Laboratories, 600 Mountain Ave., Murray Hill, NJ 07974

ABSTRACT: This article reviews current experimental and theoretical knowledge of the relaxation of lattice-mismatch strain via misfit dislocations in heteroepitaxial semiconductor films. The energetics and kinetics of misfit dislocation nucleation, propagation, and interaction processes are described in detail. In addition, there is a brief review of the principal properties of dislocations in bulk semiconductors and an outline of existing models for strained layer stability.

KEY WORDS: misfit dislocations, strained layer, epitaxy, lattice-mismatched.

I. INTRODUCTION

In the last 2 decades, the advent of crystal growth technologies, such as molecular beam epitaxy (MBE), organometallic vapor phase epitaxy (OMVPE), and hybrids such as gas source molecular beam epitaxy (GSMBE), has enabled ultra-thin epitaxial semiconductor layers to be routinely grown with both monolayer precision in thickness and compositional control of the order of one atomic percent. This has opened new regimes in physics and materials science for study. The initial focus was on homoepitaxial films and heteroepitaxial systems that are closely lattice-matched. The two primary examples of the latter class of systems are $Al_xGa_{1-x}As/GaAs$ and $In_{0.53}Ga_{0.47}P/InP$. Such structures have been grown with extremely high quality in terms of background impurity concentrations, and point and extended defect densities. Suitably doped structures in these systems have led to the development of such concepts as modulation doping, the integer and fractional quantum Hall effects, the semiconductor laser, heterojunction bipolar transistors, and many other novel device geometries. Extension of such concepts and the exploration of new physical and device possibilities requires access to a greater range of heteroepitaxial systems. The possibilities for lattice-matched heteroepitaxial systems are relatively limited (see Figure 1) and thus, in the last decade, increasing focus has been placed upon lattice-mismatched combinations of materials. The latter class of systems also offers unique new prospects, such as the use of strain to modify electronic or optical properties,[1–4] and integration of materials with very different optical, electronic, mechanical, or thermal properties (e.g., in the GaAs/Si system).[5] The principal problems associated with lattice-mismatched epitaxy are the formation of misfit-accommodating dislocations and clustered or three-dimensional growth of heteroepitaxial layers. The former problem is absent in lattice-matched epitaxy, and the latter is generally far less prevalent in lattice-matched films. This review is primarily concerned with the limits to which lattice-mismatched films can be grown without the introduction of misfit dislocations, and the properties of misfit dislocations when these limits are exceeded. We concentrate on results from diamond cubic (dc) and zincblende (zb) semiconductors, which constitute the bulk

1040-8436/92/$.50
© 1992 by CRC Press, Inc.

IQE EX1027
U.S. Patent No. 7,250,360

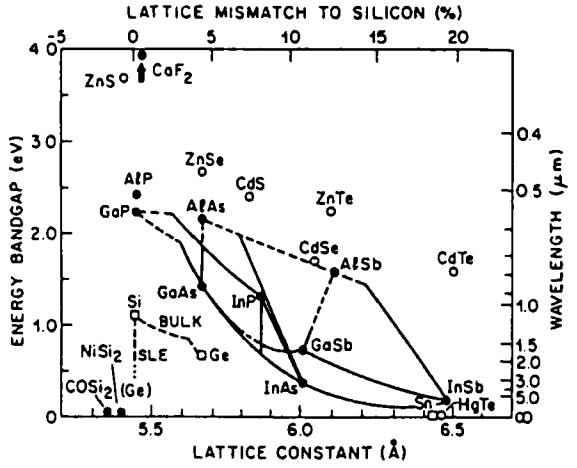


FIGURE 1. Graph of lattice parameters and bandgaps of column IV, II-VI, and III-V semiconductors. Solid and dashed lines show direct and indirect band gaps, respectively, for alloys.

of recent reports in the literature, but also describe salient results from other systems. In addition we concentrate on heterostructures that are essentially planar because relaxation mechanisms in clustered or islanded epitaxy are sufficiently complex to warrant their own review.[6-9]

II. THE CONCEPT OF STRAINED LAYER EPITAXY

Figure 2 illustrates the essential concept of strained layer epitaxy. In Figure 2(a), an epitaxial film is grown upon a substrate with the same lattice structure and lattice parameter. In principle, such films may be grown arbitrarily thick without an energetic requirement for dislocation formation. The principal caveat to this statement arises from the fact that although two materials may have the same lattice parameter at the crystal growth temperature, they will generally have different thermal expansion coefficients, thus producing a finite lattice-mismatch at room temperature. For elastically similar systems, such as $Al_xGa_{1-x}As/GaAs$, this is likely to produce stresses sufficient to generate significant misfit dislocation densities only for layer thicknesses of microns or more. For very dissimilar material combinations, however, thermal expansion mismatches may be far more severe. The $CoSi_2/Si$ system, for example, is lattice-matched at ~1200°C but has a lattice-mismatch of 1.2% ($CoSi_2$ smaller) of the unit cell lattice parameter at room temperature.[10] In Figure 2(b), an epitaxial layer is grown upon an infinitely thick substrate with a different lattice parameter (and possibly different unit cell symmetry). If the layer is thin enough, it is possible to accommodate the difference in lattice parameter by a biaxial stress in the interfacial plane (we assume here and in subsequent discussion a planar layer geometry), causing the lattice parameter of the epilayer to become equal to that of the substrate parallel to the interfacial plane. The epilayer lattice distorts along the growth normal to produce a tetragonal distortion in the same sense as the lattice parameter mismatch, (i.e., for an epilayer with a larger free lattice parameter than the substrate, the in-plane epilayer lattice parameter is reduced and the normal lattice parameter is increased, and vice versa for an epilayer lattice parameter smaller than the substrate).

The magnitude of this tetragonal distortion is given by standard elasticity theory for a biaxial stress:

$$\frac{a_n}{a_p} = 1 + \frac{1+\nu}{1-\nu}\varepsilon \qquad (1)$$

where a_n and a_p are, respectively, the normal and in-plane lattice parameters of the strained epilayer, ν is Poisson's ratio (a compound elastic constant) for the epilayer, and ϵ is the linear strain between epilayer and substrate, defined approximately by:

$$\varepsilon \sim \frac{a_e - a_s}{a_s} \qquad (2)$$

where a_e and a_s are the free in-plane lattice parameters of epilayer and substrate unit cells, respectively.

This geometry stores a high amount of elastic strain energy, because interatomic bond lengths in the epilayer are significantly stretched or compressed with respect to their equilibrium value. In addition, materials with highly directional bonding, such as covalent bonding (e.g., Si and Ge), will have extra energy introduced from bond rotations. At some epilayer thickness, generally

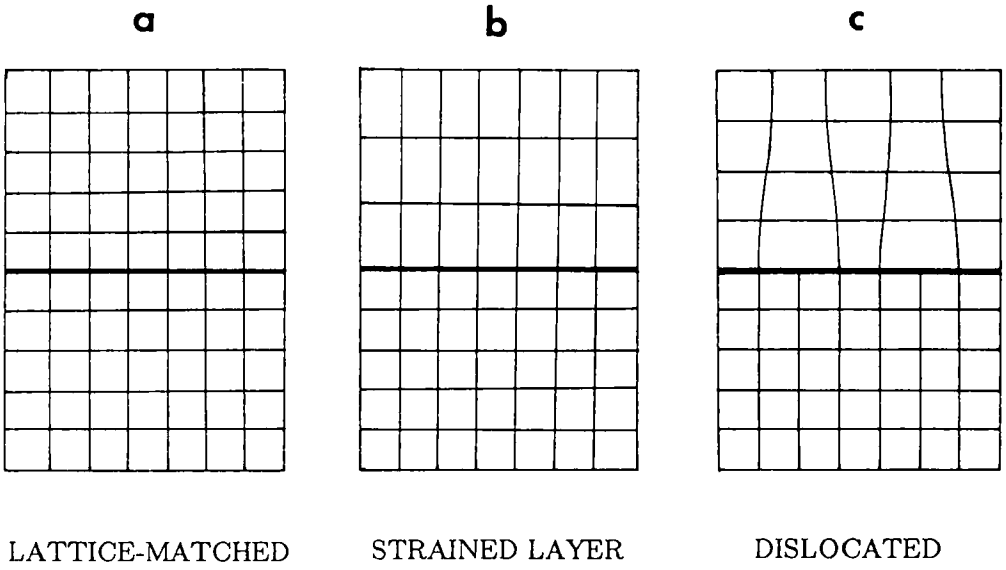


FIGURE 2. Schematic illustration of (a) lattice-matched heteroepitaxy, (b) coherently strained lattice-mismatched heteroepitaxy, and (c) relaxed lattice-mismatched heteroepitaxy.

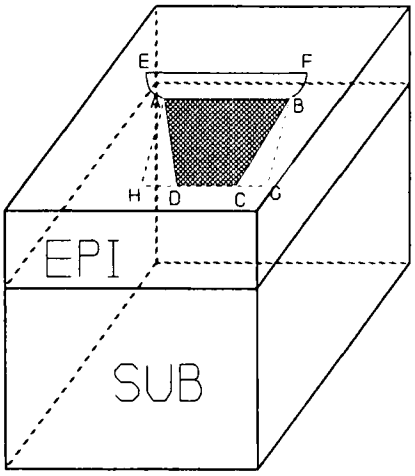
called the *critical thickness*, h_c, it becomes energetically favorable to reduce this elastic strain energy by introduction of misfit dislocations, which, as illustrated schematically in Figure 2(c), allow the epilayer to relax toward its free lattice parameter. The magnitude of this critical thickness essentially depends upon the balance of the strain energy relieved with the self-energy associated with the dislocation. The primary variable in determining the critical thickness is the amount of strain between epilayer and substrate, but many other factors also play a role. These factors, and theoretical predictions of h_c, will be discussed in detail in Section V.

If the structure is to be kept free of misfit-accommodating defects, the epilayer thickness must be less than h_c. This is clearly the ideal configuration for electronic or optical devices, as dislocations have a very deleterious effect on minority-carrier lifetimes and radiative recombination rates. Equilibrium theoretical predictions of the magnitude of h_c in semiconductor heterostructures, however, vary from of the order 10 Å for strains ~4 to 5% (e.g., GaAs/Si, Ge/Si) to hundreds of Å for strains of the order 0.5%. These thicknesses may be too low compared with the compositional abruptness attainable from the crystal growth technique used, or they may be insufficient to allow the doping transitions required for semiconductor devices. If relatively low, but nonzero, densities of misfit defects may be tolerated for a given application, then kinetic barriers to misfit dislocation nucleation and propagation may allow supercritical films to be considered, especially at low strains. These kinetic effects will be considered in later sections.

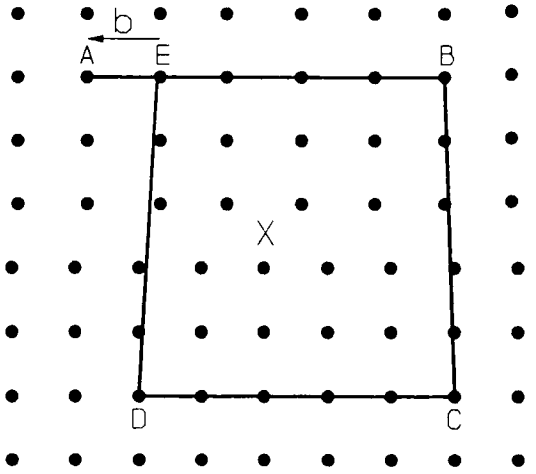
III. REVIEW OF DISLOCATION THEORY AND PROPERTIES

Following is a brief review of the salient properties of dislocations pertinent to understanding their role in relieving lattice mismatch. The geometry specific to semiconductor systems under consideration is highlighted. Also, we briefly summarize existing data and theories regarding kinetic properties of dislocations in semiconductors. For a full and detailed description of these fields, the reader is referred to the volume by Hirth and Lothe.[11]

A *perfect* or *total dislocation* is a line defect bounding a slipped region of crystal (see Figure 3(a)). A circuit drawn around atoms enclosing this line will have a closure failure as illustrated in Figure 3(b). This closure failure is known as the *Burgers vector* of the dislocation. For a perfect or total dislocation, the Burgers vector is a



A



B

FIGURE 3. (a) Illustration of the slipped region in a crystal bounded by a total dislocation. The dislocation half-loop ABCD bounds a slipped region (shaded area) extending from the crystal surface to the substrate-epilayer interface. The region ABGH represents part of the slip plane for the dislocation. The dislocation Burgers vector, which in this figure has a component normal to the surface, produces a step ABFE on the surface. Using the nomenclature developed in Section IV, CD represents the interfacial misfit dislocation, and BC and DA the threading arms. (b) Demonstration of the Burgers vector of a total dislocation. The example shown here is for an edge dislocation, with the line direction $\underline{u}$ running into the page. The dislocation core is at X. A circuit of five atoms square, which would close in perfect material, demonstrates a closure failure when drawn around the dislocation core. This closure failure is the dislocation Burgers vector, $\underline{b}$. The points A and E correspond to S and F, respectively, in the FS/RH notation.

lattice translation vector. Although the line direction of a given dislocation may vary arbitrarily, its Burgers vector is constant, apart from a possible difference in its sign depending upon the convention used. (A widely accepted convention for determining the sign of the Burgers vector is to draw the circuit from start, S, to finish, F, in the direction of a right-handed screw. This is the so-called FS/RH convention.[12] Under this convention, the opposite sides of a dislocation loop have opposite Burgers vector signs.) A total dislocation cannot end within the bulk of a crystal. It must terminate at an interface with noncrystal (generally a free surface, but possibly also an interface with amorphous material), at a node with another defect, or upon itself to form a loop.

The *character* of a dislocation is defined by the relationship between its line direction, $\underline{u}$, and its Burgers vector, $\underline{b}$. If $\underline{b}$ is parallel or antiparallel to $\underline{u}$, the dislocation is said to be of *screw* character. If $\underline{b}$ is perpendicular to $\underline{u}$, the dislocation is said to be of *edge* character. In intermediate configurations, the dislocation is said to be of *mixed* character. As $\underline{u}$ may vary along a dislocation but $\underline{b}$ may not, a nonstraight dislocation will vary in character along its length.

A dislocation has a self-energy arising from the distortions it produces in the surrounding medium. This energy may be divided into two contributions: those arising from inside and outside the dislocation *core*. The distortions of atomic positions inside the core are sufficiently high that linear elasticity theory does not apply; dangling bonds may also exist. The dislocation core energy is not well known, but it depends on the material type (predominantly the nature of the interatomic bonding) and the dislocation Burgers vector and character. The dimensions of the core are also uncertain, but theoretical and experimental estimates of its diameter are of the order of magnitude of the Burgers vector for covalent semiconductors[13] and several Burgers vectors for metals. Outside the core, atomic distortions may be modeled using linear elasticity theory and exact expressions for this energy can be derived. Hirth and Lothe[11] formulate the self-energy per unit length of an infinitely long dislocation parallel to a free surface* a distance, R, away:

$$E_s = \frac{Gb^2(1 - \nu \cos^2 \theta)}{4\pi(1 - \nu)} \ln\left(\frac{\alpha R}{b}\right) \qquad (3)$$

where G is the epilayer shear modulus, Θ is the angle between $\underline{b}$ and $\underline{u}$, and α is a factor intended to account for the dislocation core energy (α is generally estimated to be from 0.5 to 1.0 in metals and from perhaps 1 to 4 in semiconductors, with slightly higher values for edge than screw dislocations). Thus, the dislocation self-energy varies as the square of the magnitude of its Burgers vector. This strongly encourages the Burgers vector of a total dislocation to be the minimum lattice translation vector in a given class of crystal structure. For the diamond cubic (dc) and zincblende (zb) structures (and also for their parent face-centered cubic (fcc) structure), which this review concentrates upon, this minimum vector is $a/2\langle 011 \rangle$. Indeed, this is the total Burgers vector almost always observed in these structures. The value of R in Equation 3 pertinent to calculations of interfacial misfit dislocation energies in uncapped strained epilayers is generally the epitaxial film thickness, h. For interfaces with very high defect densities, a cutoff parameter, R, corresponding to the average distance between defects, if this is less than the distance to the free surface, may be more appropriate.

The self-energy of a dislocation produces a virtual force pulling it toward an *image dislocation* on the opposite side of a free surface. The simplest case is for an infinite straight screw dislocation running parallel to a free surface whose image is a dislocation of opposite Burgers vector *in vacuo* equidistant from the surface. Other configurations have more complex image constructions. Image effects also exist across internal interfaces between materials with different shear moduli. The stress field around a dislocation also produces an interaction force, F_{ij}, between two separate dislocations, and although its magnitude is again configuration dependent, in general the interaction force per unit length is given by:

$$F_{ij} = k \frac{\underline{b}_i \cdot \underline{b}_j}{R'} \qquad (4)$$

* Strictly speaking, this formula is derived for a dislocation within a cylindrical volume of inside radius b/α and outside radius R, but application to a planar free surface a distance R from the dislocation is generally a good approximation.

where R′ is the separation of the two segments and k is the constant of proportionality (equal to $1/2\ \pi$ for the simplest case of parallel screw dislocations). Thus, this force is maximally attractive for antiparallel Burgers vectors, maximally repulsive for parallel Burgers vectors, and zero for orthogonal Burgers vectors.

Motion of dislocations occurs most easily within their *glide* planes, which are the planes containing their line direction and Burgers vectors. For a screw dislocation, $\underline{u}$ is parallel to $\underline{b}$ and, thus, any plane is a potential glide plane. For mixed or edge dislocations, there is only one unique glide plane, whose normal is given by $\underline{b} \cdot \underline{u}$. Glide also occurs most easily on the widest spaced planes in a given system because the Peierls stress[14] for dislocation motion decreases with increasing planar separation. For dc, zb, and fcc crystals this corresponds to the {111} sets of planes, and these are the almost ubiquitously observed glide planes in these structures. Glide occurs by simultaneous reconfiguration of bonds at the dislocation core to effectively move the core one atomic spacing. This process is further aided by nucleation and motion of atomic-scale *kinks*, as will be discussed shortly. No mass transport of point defects is required in the glide process. However, motion out of the glide plane, called *climb*, has to occur by extension or contraction of the half plane terminating at the dislocation core, thereby requiring mass transport of point defects. Such diffusion processes are generally much slower than glide processes at physically reasonable temperatures.

As alluded to earlier, the glide process is facilitated by nucleation of atomic-scale kinks along the propagating dislocation line, with lateral motion of the kink arms transverse to the dislocation line, effectively moving the entire dislocation length. In the absence of kinks, dislocations typically lie along well-defined crystallographic directions along minima in the crystal potential (known as *Peierls valleys*). These are the ⟨011⟩ directions for the dc and zb semiconductors considered here. Formation of a kink requires motion of a small length of dislocation line across the potential, or Peierls, barrier between valleys. Subsequent lateral motion of the kink arms also involves motion over a secondary Peierls barrier. This process is illustrated in Figure 4. In metals, the energy required to form and move these kinks is relatively low, typically less than of the order 0.1 eV. This arises from the low Peierls barriers resulting from the relatively weak metallic bonding. This low kink energy in metals means that the dislocations are not strongly constrained to lying within Peierls valleys and, thus, are not very straight. Motion of dislocations in metals is also possible at very low temperatures. There is evidence that dislocations in cubic metals remain mobile due to kink motion down to temperatures approaching absolute zero.[15,16] Dislocation motion and configurations in semiconductors, however, are typically dominated by the Peierls barriers. In materials that are purely covalently bonded, the glide activation energies are particularly high e.g., 1.6 eV in Ge and 2.2 eV in Si for stresses in the tens to hundreds of MPa regime.[17–21] In materials that are only partially covalently bonded, the activation energies are correspondingly lower, e.g., about 1.0 eV in GaAs and InAs.[20,22–24] These glide activation energies arise from the Peierls barriers, which

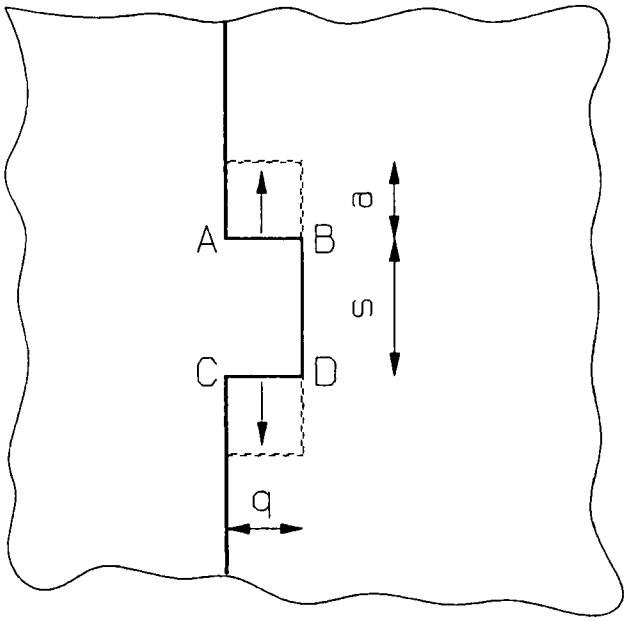


FIGURE 4. Motion of dislocations by kink pairs in a semiconductor crystal. A small length, s, of a straight dislocation line jumps an interatomic distance, q, transverse to the line, to form a kink pair AB and CD. The kinks then run parallel to the dislocation line by successive jumps of interatomic distance, a. The entire dislocation line thus effectively moves a distance q to the right.

have to be overcome in forming and moving kinks. The details of the kink nucleation and propagation process have been derived in some detail by Hirth and Lothe,[11] and corresponding dislocation velocities, v_d, as functions of temperature and applied stress have been calculated. These lead to an expression of the form:

$$v_d = v_0 \sigma_e^m e^{-E_v(\sigma)/kT} \quad (5)$$

where v_0 is a prefactor containing an attempt frequency (commonly taken to be of the order of the Debye frequency), kink jump distances, and an inverse linear dependence upon temperature, σ_e is the effective stress acting on the dislocation (we describe how this is derived for strained epitaxial layers in Sections V and VI) and $E_v(\sigma)$ is the glide activation energy. The prefactor V_0 may also depend upon the propagating dislocation length.[11] The power, m, is predicted to vary from $^7/_8$ to $^5/_4$ in different stress regimes,[25] and is found experimentally to be of the order 1.0 at stresses of the order tens of MPa in very pure Si and Ge.[18] The activation energy $E_v(\sigma)$ also varies in different dislocation length regimes. (This essentially depends upon whether kinks collide with each other before reaching the end of the propagating dislocation segment. It is generally assumed that they do in bulk samples, whereas for very short propagating dislocation lengths, as may occur in thin epilayers, kinks may reach the end of the propagating segment before colliding with each other.) The activation energy also exhibits a stress-dependence at applied stresses that approach a significant fraction of the Peierls stress.

The kink model for dislocation motion is widely accepted, and experimental results are generally described loosely by Equation 5 (in many studies m is often found to be >1.0, perhaps of the order 1.5[17]), but the prefactors derived from measurements on bulk semiconductors typically are 2 to 3 orders of magnitude higher than the Hirth-Lothe theoretical predictions.[17,18] It has been suggested by many authors[18,26–29] that this is due to *obstacles* to dislocation motion, such as point defects, impurities, inhomogeneities in the dislocation core structure, etc. Another significant factor may be the charge state of the dislocation cores: dangling bonds may be present, particularly at the kinks themselves, leading to a charge effect on dislocation motion. This model is supported by observations of a doping dependence of dislocation velocity.[19,30–32]

A final major consideration is that total dislocations may be *dissociated* into *partial* dislocations. A partial dislocation is a dislocation whose Burgers vector is not a lattice vector and is usually smaller than the minimum lattice vector. In fcc, zb, and dc materials, a very common partial dislocation is a stacking fault in the cubic ABC stacking sequence of atoms along ⟨111⟩ directions. These stacking faults have Burgers vectors of a/6⟨112⟩ or a/3⟨111⟩ and are called Shockley and Frank partials, respectively. Only the Shockley partial can glide within {111} planes. Total dislocations may often dissociate because they can lower their energy according to the b^2 criterion. For example, the reaction:

$$a/2[110] = a/6[121] + a/6[2\bar{1}\bar{1}] \quad (6)$$

is energetically favorable. The resulting partial dislocations mutually repel each other and glide apart on the $(\bar{1}1\bar{1})$ plane. This produces a ribbon of stacking fault between them, and at some point the extra cost in energy from the stacking fault balances the interaction energy of the two partials. (Typical dissociation widths in unstressed Si and Ge are of the order of a few nanometers, implying stacking fault energies of the order 50 to 80 $mJ \cdot m^{-2}$.[20,33–35])

The existence of partial dislocations and the possibility of dissociation can dramatically affect the energetics of dislocation motion. For example, the partials may have different core structures and charge states from each other and from the undissociated dislocation. The Peierls barriers for motion of partial dislocations may be very different than for motion of a total dislocation, particularly if kink formation and motion on the two partials is correlated.[36] The existence of partial and dissociated dislocations may also be of significance in dislocation nucleation energetics, as will be discussed later.

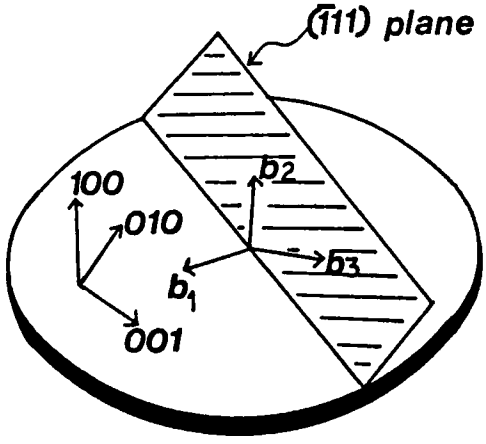
IV. MISFIT DISLOCATION GEOMETRIES AT HETEROEPITAXIAL INTERFACES

In this section we discuss the specific misfit

dislocation geometries pertinent to heteroepitaxial semiconductor interfaces. By far the most common interface orientation is (100), so we initially limit our discussion to this interface.

For the (100) orientation, the four possible {111} glide planes intersect the (100) interface along orthogonal in-plane [011] and [01$\bar{1}$] directions, with a pair of glide planes intersecting along each direction. The orientation of one of these glide planes, and the accompanying possible a/2⟨110⟩ dislocations is shown in Figure 5. The intersections of the glide planes with the interface thus produces a square mesh of interfacial dislocations, as shown experimentally in Figure 6. Only dislocations with Burgers vectors lying within these {111} planes will be able to move by glide. For a given glide plane, three such Burgers vectors exist, e.g., for the ($\bar{1}$11) plane in Figure 5, $\underline{b}$ = a/2[101], a/2[110], or a/2[01$\bar{1}$]. Of these three Burgers vectors, the last is a screw dislocation and, as will be shown in the next section, does not experience any resolved lattice-mismatch stress. The first two are of mixed edge and screw character and are known as *60-degree dislocations*, corresponding to the angle between $\underline{b}$ and $\underline{u}$. Only 50% of the magnitude of their Burgers vectors projects onto the interfacial plane; thus they are only 50% effective at relieving lattice-mismatch. The final possibility to consider is that of edge dislocations, e.g., for the ($\bar{1}$11) plane, $\underline{b}$ = a/2[0$\bar{1}\bar{1}$]. Such dislocations have their Burgers vectors lying within the interfacial plane and are 100% effective at removing lattice-mismatch. The Burgers vectors do not, however, lie within any glide plane, and thus these defects must move by far slower climb processes.

The geometrical requirement that a dislocation must terminate at another dislocation, upon itself, or at a free surface means that something has to happen with the ends of misfit dislocations. They cannot simply terminate within the interface. If the defect density is relatively low and, thus, dislocation interactions are unlikely, the most obvious place for the misfit dislocation to terminate is at the nearest free surface, which will generally be the epilayer surface. This requires *threading* dislocations that traverse the epitaxial layer from interface to surface and, in general, each misfit dislocation will be associated with a threading defect at each end, as illustrated in Figure 3(a). (The necessity for the threading arms is removed if the length of the misfit dislocation grows sufficiently that it can terminate at the wafer edge, or at a node with another defect.) Propagation of misfit dislocations then occurs by lateral propagation of the threading arms. The major stages of the strain relaxation process by misfit dislocations — nucleation, propagation, and interaction — are illustrated schematically in Figure 7.



$\underline{b}_1 = \frac{1}{2}[0\bar{1}\bar{1}]$

$\underline{b}_2 = \frac{1}{2}[110]$

$\underline{b}_3 = \frac{1}{2}[101]$

FIGURE 5. Schematic illustration of the geometry of misfit dislocations at a (100) interface in fcc, dc, or zb crystals. One inclined (1̄11) glide plane and consequent interfacial [01$\bar{1}$] dislocation direction are shown, together with the possible a/2⟨110⟩ Burgers vectors orientations, where $\underline{b}_1$ is of edge (climb) type and $\underline{b}_2$ and $\underline{b}_3$ are of 60° (glide) type.

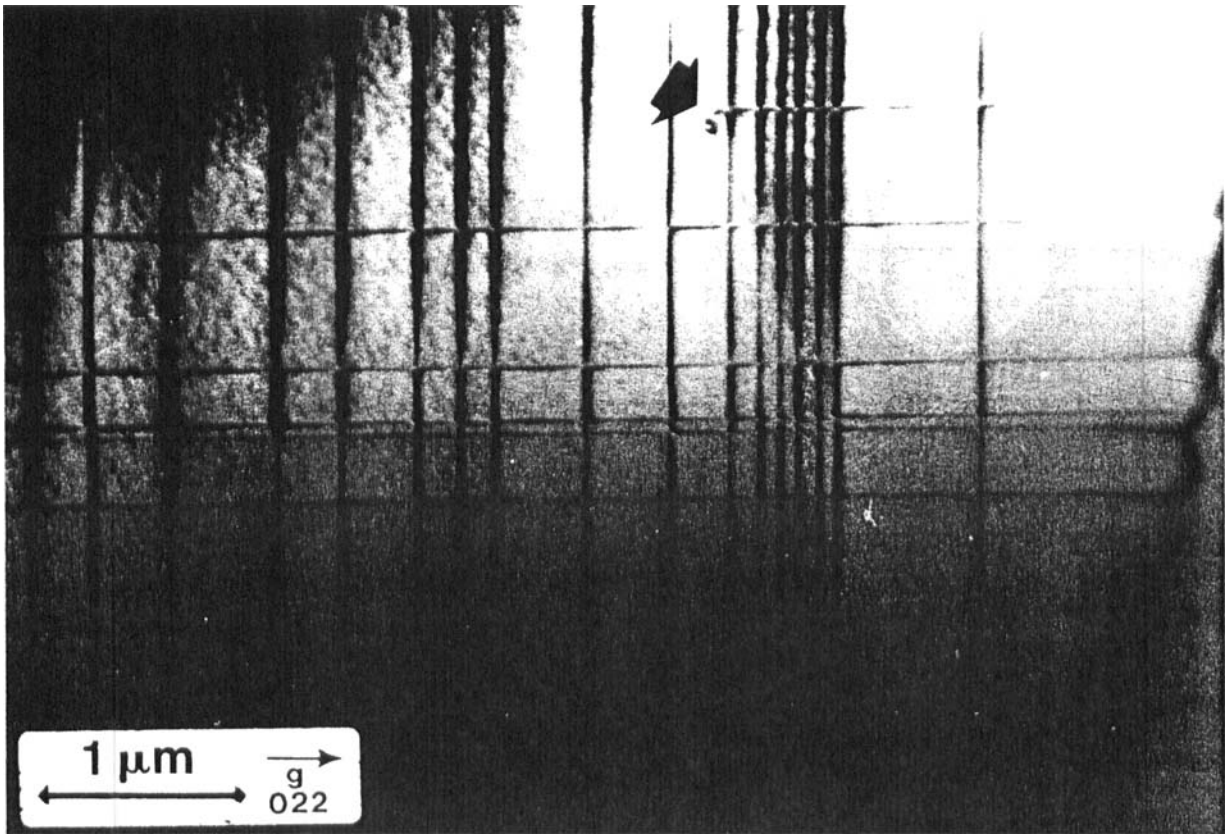


FIGURE 6. Plan view TEM image of the orthogonal misfit dislocation array at a 1500 Å $Ge_{0.15}Si_{0.85}$/Si(100) interface. A threading dislocation end is arrowed.

These threading dislocations are extremely deleterious to practical application of strained layer epitaxy. For many potential device applications, a high interfacial misfit dislocation density is tolerable if the epilayer quality is sufficiently high at some distance from the interface (about a micron or so). Threading dislocations inhibit this possibility. For example, for fully relaxed interfaces in the GaAs/Si system (which has a lattice-mismatch of 4.2%), threading defect densities of the order 10^8 cm^{-2} are typical about 1 μm from the interface. For lower mismatch systems, threading defect densities tend to be lower, perhaps typically 10^6 to 10^7 cm^{-2} at 0.5% mismatch. Techniques for reducing these threading defect densities will be discussed in Section X.

The geometry of misfit dislocations will be different on surfaces other than (100). If the dislocations glide on inclined {111} planes, the interfacial misfit dislocation line directions will be defined by the intersection of these planes with the relevant interface. The expected geometries for (100), (110), and (111) interfaces are illustrated schematically in Figure 8. These geometries have been confirmed by us in the Ge_xSi_{1-x}/Si system, although, as described in the next section, different dislocation structures are observed at (110), (111), and (211) interfaces when compared with (100) interfaces.[37-40]

V. THEORETICAL MODELS FOR THE CRITICAL THICKNESS

There have been many attempts to model the critical thickness for misfit dislocation introduction in strained epitaxial layers. Early models by Frank and Van der Merwe et al.[41,42] attempted to model the critical thickness by minimizing the energy of a misfit dislocation array at the interface. Their approach, though rigorous, is math-

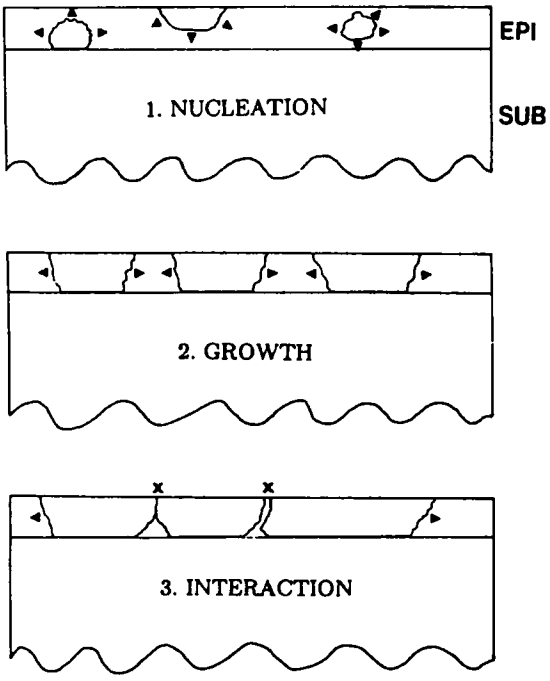


FIGURE 7. Schematic illustration of (1) nucleation, (2) growth, and (3) interaction of misfit dislocations.

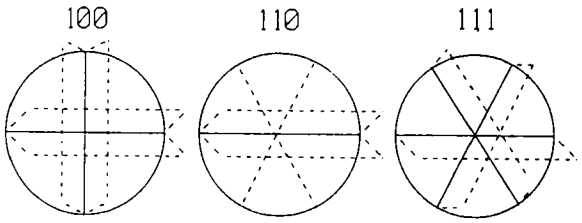


FIGURE 8. Schematic illustrations of the symmetries of interfacial misfit dislocations at (100), (110), and (111) interfaces. Solid straight lines show the interfacial misfit dislocations; dashed lines outline intersecting {111} glide planes.

ematically complex and analytical solutions to the equations constructed exist only in the limits of very thin or very thick films. Their models enjoy considerable success in predicting the critical thickness in bcc metal systems.

The standard approach used to predict the equilibrium critical thickness in strained semiconductor heterostructures is that initially adopted by Matthews and Blakeslee (MB).[43,44] A schematic representation of their model is illustrated in Figure 9. A preexisting dislocation threading through the substrate threads through the epitaxial layer and experiences an applied force, F_a, due to the lattice-mismatch strain. If this force is sufficiently high, the threading dislocation will propagate through the epitaxial layer, generating misfit dislocation length at the substrate/epilayer interface. This generated misfit dislocation length will have a self-energy, as discussed in Section III, that will produce a restoring stress upon the propagating threading arm, F_T. This restoring stress is often loosely referred to as being due to a "line tension". If $F_a > F_T$, then the threading dislocation will propagate and a misfit dislocation is generated. The MB definition of critical thickness is when $F_a = F_T$. Energetically, this is equivalent to saying that misfit dislocations form when their self-energy is less than the elastic strain energy they relax. Note that the MB analysis applies not only to the geometry of dislocations threading up from the substrate, but also to any other geometry with suitable modification of the F_a and F_T terms.

Conceptually, the MB model is now almost universally accepted to be correct in predicting the epilayer thickness at which the first misfit dislocations should appear, and experimentally it has been verified for a wide range of systems. Some confusion and controversy still arises because of differences in the way in which the F_a and F_T terms are modeled, and because of uncertainties in the exact magnitudes of some of the factors incorporated into the expressions for these two forces.

The in-plane stress due to a biaxial strain is given by standard elasticity theory in an elastically isotropic solid to be

$$\sigma_0 = 2G \frac{(1 + \nu)}{(1 - \nu)} \varepsilon \qquad (7)$$

The stress experienced by the epilayer threading dislocation arm, by whose motion the misfit dislocation propagates, is given by the resolution of the stress in Equation 7 onto the dislocation Burgers vector. This resolution is given by the Schmid factor, $S = \cos \lambda \cos \phi$,[45] where, for heteroepitaxial interfaces, λ is the angle between the Burgers vector and the normal in the interface to the dislocation line direction, and ϕ is the angle between the normal to the glide plane and the free surface. For 60-degree glide dislocations at

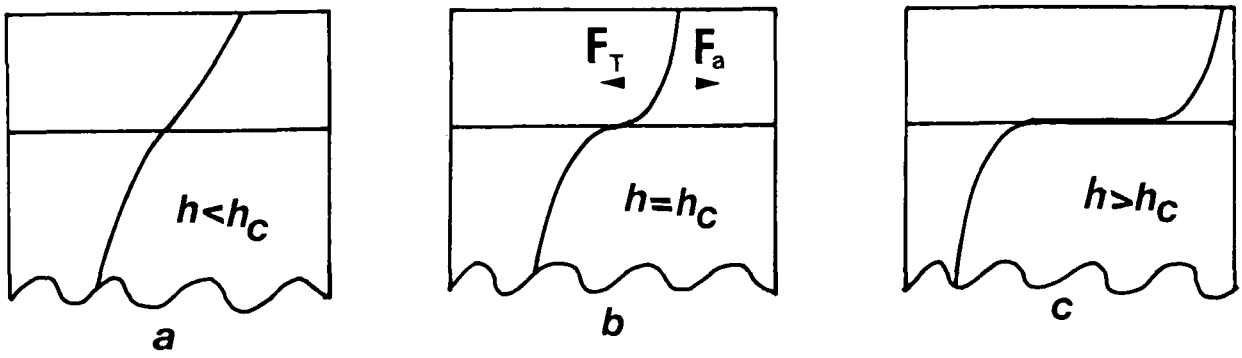


FIGURE 9. Schematic illustration of the Matthews and Blakeslee model of critical thickness.

a (100) interface, $\cos\lambda = 0.5$ and $\cos\phi = \sqrt{2}/\sqrt{3}$. The applied stress, σ_a, experienced by the dislocation is now:

$$\sigma_a = 2G\varepsilon \frac{(1 + \nu)}{(1 - \nu)} \cos\lambda \cos\phi \qquad (8)$$

Note that Equation 8 predicts that dislocations lying within a glide plane perpendicular to the interface (this includes all screw dislocations) will experience no resolved stress due to lattice-mismatch.

The applied stress is related to the applied force, F_a, by the standard relation:

$$\frac{F_a}{L} = \sigma_a b \qquad (9)$$

where L is the length of the threading dislocation, equal to $h \cdot \sec\phi$. Note that we have assumed here that the threading dislocation runs along the normal within the glide plane to the interfacial dislocation line direction, i.e., it takes the shortest path length within the glide plane from interface to surface. Any other threading arm orientation introduces another angular factor, as discussed by Chidambarrao et al.[46] Thus:

$$F_a = 2Gbh\varepsilon \frac{(1 + \nu)}{(1 - \nu)} \cos\lambda \qquad (10)$$

The restoring force acting upon the threading dislocation, F_T, may be calculated from the energy required to create a segment of interfacial misfit dislocation of length dx: $dE = E_s dx$, where E_s is given by Equation 3. Using the relation that the total force acting on the threading dislocation is given by dE/dx, we have:

$$F_T = \frac{Gb^2(1 - \nu\cos^2\theta)}{4\pi(1 - \nu)} \ln\left(\frac{\alpha h}{b}\right) \qquad (11)$$

Equating F_a to F_T then yields for the MB definition of critical thickness:

$$h_c = \frac{b(1 - \nu\cos^2\theta)}{8\pi(1 + \nu)\varepsilon\cos\lambda} \ln\left(\frac{\alpha h_c}{b}\right) \qquad (12)$$

Thus, it is seen that application of the MB model to modeling the forces acting on the dislocation and to the critical thickness requires knowledge of the exact dislocation geometry via the terms $\cos\lambda$ and $\cos\theta$, the elastic constants ν and G, and the core-energy parameter, α. The angular factors will be specific to a given Burgers vector direction and interface orientation (although they are fixed for all possible a/2⟨110⟩ glide dislocations at a (100) interface), and thus care must be taken to calculate these accurately. The bulk elastic parameters G and ν are known for elastically isotropic bulk materials, but they may differ in very thin layers or at very high strains (also, elastic anisotropicity may be a significant variable in considering forces on different surfaces). Finally, the core parameter α is known only approximately, particularly in covalent semiconductors. All these factors can contribute to inexactness in applying the MB equation.

There have been many other formulations of the critical thickness in strained layer epi-

taxy (for example, References 47 thru 51). Most of these models represent extensions or refinements of the Frank/Van der Merwe or Matthews/Blakeslee frameworks.

Other effects can influence the critical thickness. As discussed in Section III, perfect dislocations may dissociate into partial pairs separated by ribbons of stacking fault. The modified Burgers vectors and extra stacking fault and integration energies should thus, in principle, be incorporated into the critical thickness calculation. Cammarata and Sieradzki[52] have considered the effects of compressive surface stresses on the critical thickness, and have shown that they can be considerable at large lattice-mismatches. Fox and Jesser[53] have invoked a static Peierls stress as an additional restoring stress upon a dislocation, thereby increasing the critical thickness. (However, it seems to us that the effect of the Peierls barrier enters into dislocation kinetics, as discussed in the next section, rather than for the static case of the equilibrium critical thickness.) Basson and Booyens[54] have considered the effect of substrate relaxation (curvature) upon the critical thickness. People and Bean[55] considered the critical thickness, effectively based upon the local strain energy density necessary for the nucleation of misfit dislocations.

For the common configuration where the strained epilayer is buried beneath a cap of material that has the same lattice parameter as the substrate, the misfit dislocation configuration may involve interfacial segments at either both interfaces or just the bottom interface, as indicated schematically in Figure 10. The conditions for these two configurations and the transition between them has been studied by several groups.[56,57] Essentially, the double interfacial dislocation configuration introduces an extra dislocation self-energy term compared with the single dislocation configuration, but it also ensures that the lattice parameter of the cap is kept at its bulk value. (For the single dislocation configuration, the cap lattice parameter is necessarily strained away from its bulk value.) Thus, the double dislocation configuration is favored if the energy of the dislocation at the top interface is less than the strain energy created in relaxing the cap layer away from its natural lattice parameter, i.e., a kind of reverse critical thickness calculation. This depends primarily upon the thickness of the capping and strained layers. For the common configuration where the capping layer thickness is greater than the strained layer thickness, the double interfacial dislocation configuration is generally preferred. The presence of the two interfacial dislocations now clearly modifies the line tension and critical thickness expressions, Equations 11 and 12. The self-energy per unit interfacial length of the dislocation dipole is given by the separate energies of the top and bottom dislocations minus the interaction energy between them. For a capping layer thickness, h_{cp}, and a strained layer thickness, h_e, this yields:

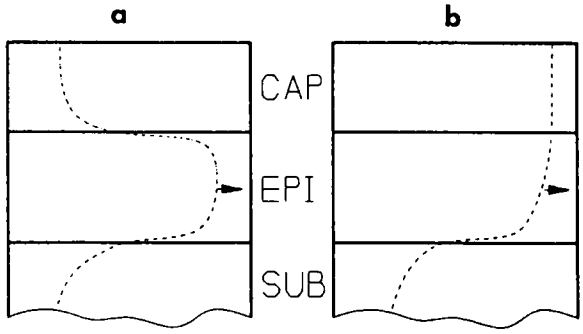


FIGURE 10. Misfit dislocation configurations at buried strained layers. Possible configurations are (a) relaxation of both interfaces, or (b) just the lower interface.

$$E_{pair} = Gb^2 \frac{(1 - \nu \cos^2 \theta)}{4\pi(1 - \nu)} \times \left[\ln\left(\frac{\alpha h_{cp}}{b}\right) + \ln\left(\frac{\alpha(h_{cp} + h_e)}{b}\right) - 2 \ln\left(\frac{\alpha(h_{cp} + h_e)}{b}\right) + 2 \ln\left(\frac{\alpha h_e}{b}\right) \right] \quad (13)$$

The last pair of logarithmic terms in this equation is the interaction energy between the two interfacial dislocations.[11] For the limit $h_{cp} \gg h_e$, the entire logarithmic expression simplifies to $\sim 2 \ln (\alpha h_e/b)$ or simply twice the energy of the single interfacial dislocation at an uncapped epilayer. The line tension term for the buried layer, F_{Tb},

is then a factor of two higher than for the single layer (as given by Equation 11) and the critical thickness for $h_c \gg b$ is also approximately a factor of two higher. In fact, this approximation is reasonably accurate for all $h_{cp} > h_e$ when $h_e \gg b$. This factor of approximately two in critical thickness from capping of strained layers is of great benefit in post-growth processing of practical strained layer devices.

So far, we have limited discussion to the critical thickness on (100) surfaces. Extrapolation of the MB model via Equation 12 to different surfaces requires calculation of the appropriate value of cosλ, and consideration of elastic anistropicity, which is by no means a minor correction for the materials studied here. For example, if we assume elastic isotropicity and consider the (100), (110), and (111) surfaces, cos λ varies as cos $\lambda_{100} = 0.50$, cos $\lambda_{110} = 0.71$, and cos $\lambda_{111} = 0.29$ for a/2⟨110⟩ glide dislocations. This means that for these dislocations, $h_c(111) > h_c(100) > h_c(110)$ for identical epilayers grown on the three different surfaces.[37,39] We have observed, however, that for growth on (110) and (111) surfaces motion of separate a/6⟨112⟩ Shockley dislocations is possible.[37-39] The observed Shockley dislocations are of edge type and glide within the appropriate {111} planes. An illustration of such dislocations is shown at a Ge_xSi_{1-x}/Si(110) interface in Figure 11. In Figure 12, we show calculations of the critical thickness at a Ge_xSi_{1-x}/Si (110) interface for an undissociated a/2⟨110⟩ dislocation, and for the two Shockley partials into which the total dislocation would dissociate. These partials make angles of 30° and 90° with their interfacial line direction. It is seen that the critical thickness is lowest for the 90° partial over most of the alloy composition range. The critical thickness is generally highest for the 30° partial. (Note that in calculations of the critical thickness for the partial dislocations, we have included an extra force arising from the energy of the stacking faults that they crate. We use an average value of 65 mJ · m^{-2} for this stacking fault energy[20]). These calculations would thus predict the observed presence of 90° Shockley partials at the (110) interface (with similar predictions and observations for the (111) interface). Such partials are not, however, observed for compressive strains on the (100) interface because the Thompson tetrahedron construction[58] shows that the passage of such a defect under these conditions would produce a very high-energy stacking fault of the type ABC/CAB, where "/" denotes the stacking fault. Passage of the 30° partial in this instance is necessary to produce the required low-energy ABC/BCA type fault. (The 90° partial will follow the 30° partial as it then removes the stacking fault and produces a total dislocation.) For the (100) surface, the 30° partial again has a higher critical thickness than the a/2⟨110⟩ total, so separate partials are generally not observed. For *tensile* strains on the (100) surface, the situation is reversed and it is the 90° partial that produces the low-energy stacking fault. Thus, in this configuration, isolated partial dislocations are frequently observed.[59-61] For compressive strains on the (110) and (111) surfaces, the Thompson construction again shows that passage of a 90° partial can produce a low-energy ABC/BCA fault and thus, because it has a lower critical thickness than an a/2⟨110⟩ total dislocation, it is observed frequently. This substantially limits technological application of (110) and (111) surfaces, as misfit dislocations thus generally imply stacking faults throughout the epitaxial layer.

VI. KINETICALLY DEPENDENT CRITICAL THICKNESSES

In Figure 13, we show calculated values of the MB equilibrium critical thickness and experimentally determined critical thickness for (a) the Ge_xSi_{1-x}/Si and (b) the $In_xGa_{1-x}As$/GaAs (100) interfaces. In Figure 13(a) we calculate the MB critical thickness using the following values: $\nu = 0.28$; cos $\lambda = {}^1/_2$; cos $\theta = {}^1/_2$; $b = 3.9$ Å; $\alpha = 4$. Also shown are experimental measurements of critical thickness by Bean et al.[62] for an MBE growth temperature of 550°C, and Kasper et al.[63] for an MBE growth temperature of 750°C. In Figure 13(b), we reproduce from Fritz et al.[64] calculations and measurements from superlattices in the $In_xGa_{1-x}As$/GaAs(100) system. (Note that slightly different materials parameters were used by Fritz et al.[64] from those that we used in calculating the MB critical thickness in Figure 13(a).)

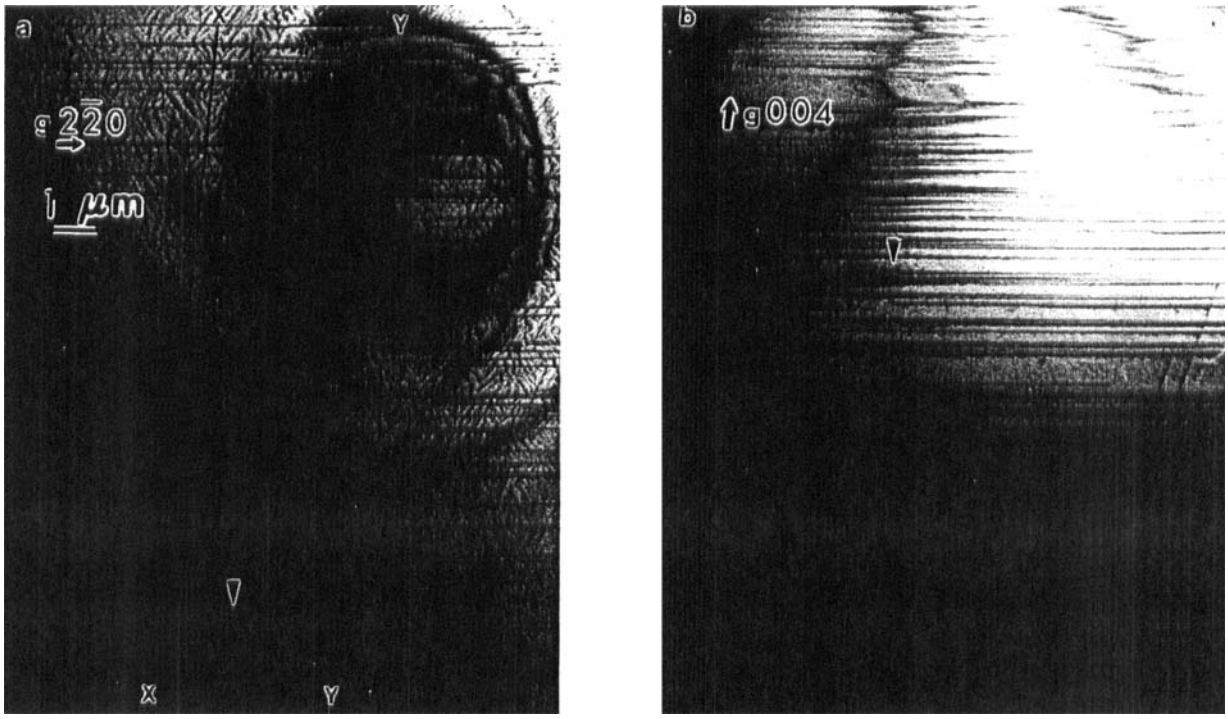


FIGURE 11. Bright field plan view TEM images of a 250 Å $Ge_{0.33}Si_{0.67}$/Si(110) structure with (a) $\underline{g}$ = [220] and (b) $\underline{g}$ = [004] in the (110) pole. The "missing" dislocations in (a) with $\underline{g}$ parallel to $\underline{u}$ correspond to edge Shockley partials. An arrow is shown as a locating mark for the same area in the two images.

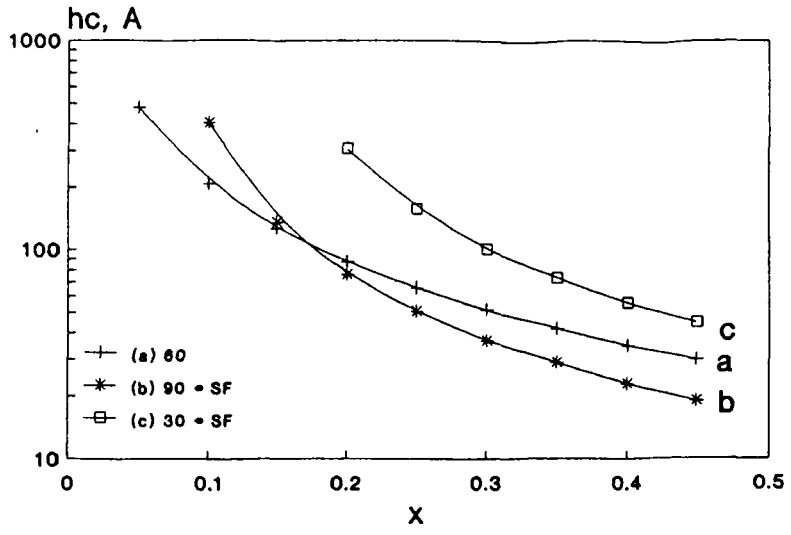


FIGURE 12. Equilibrium critical thickness for misfit dislocations at the Ge_xSi_{1-x}/Si(110) interface, as calculated from the Matthews-Blakeslee model. (a) Undissociated 60° a/2⟨110⟩, (b) 90-degee a/6⟨112⟩, and (c) 30-degree / 6⟨112⟩. The a/6⟨112⟩ Shockley partial calculations include an extra term arising from a stacking fault energy of 65 mJ · m^{-2}.

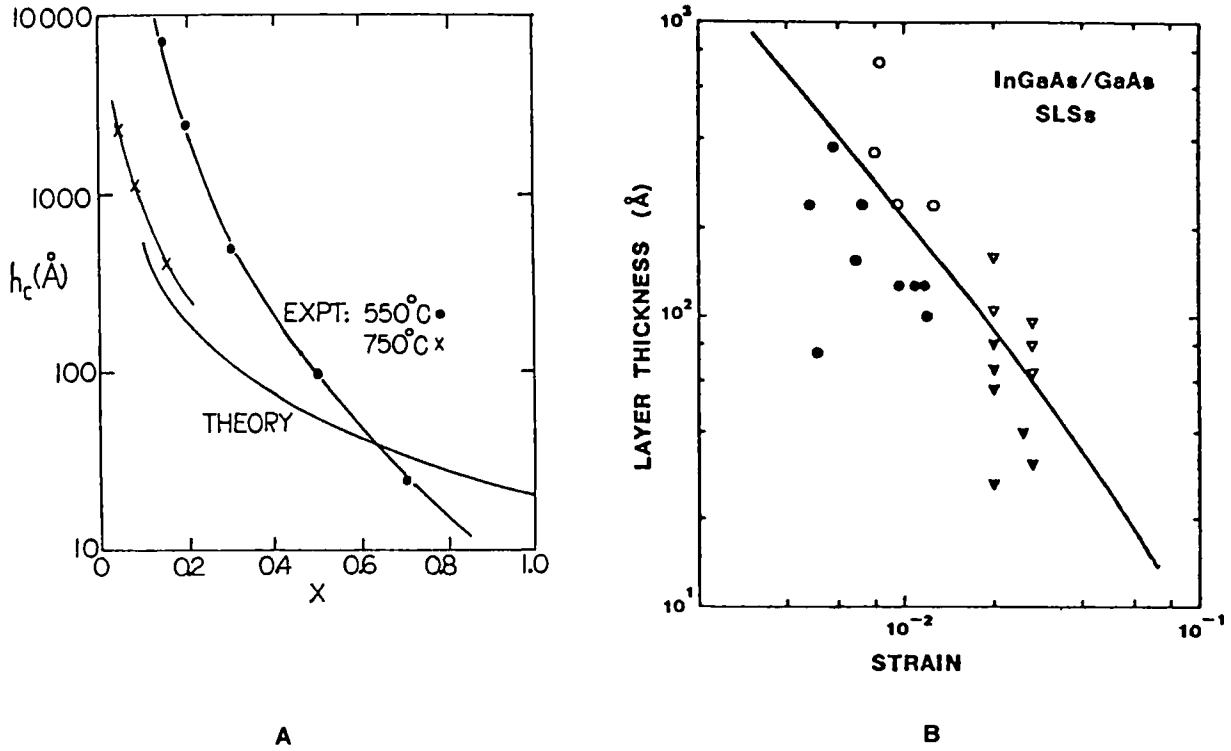


FIGURE 13. Plots of the experimentally observed critical thickness in (a) the Ge_xSi_{1-x}/Si(100)[62,63] and (b) the $In_xGa_{1-x}As$/GaAs(100) systems,[64] together with predictions of the Matthews-Blakeslee theory. In (b) the closed points indicate undislocated structures, whereas the open points indicate dislocated structures (triangular and circular points denote structures grown in two different chambers[64]). The solid line in (b) corresponds to calculations of the MB theory.[64]

The following trends are observed in these and other similar sets of data:[65-67]

1. The experimentally determined critical thickness is often greater than the MB equilibrium prediction.
2. The discrepancy between experiment and theory increases with decreasing growth temperature.
3. The discrepancy between experiment and theory is generally greater for covalently bonded heterostructures (e.g., Ge_xSi_{1-x}/Si) than mixed ionic/covalent bonding (e.g., $In_xGa_{1-x}As$/GaAs).

The ability to grow greater thicknesses of strained epilayer than predicted by the MB theory before misfit dislocations are observed has led to the concept of *metastable strained layer growth*. This observed metastable regime is not, however, inconsistent with the MB model. The apparent differences arise from the following sources:

1. There are errors and uncertainties in applying the MB model.
2. There is the sensitivity of the characterization technique used for observing misfit dislocations and/or strain relaxation. As first pointed out by Fritz,[68] the observed critical thickness will crucially depend on the minimum detectable strain relaxation or misfit dislocation density for a given technique. Techniques such as photoluminescent imaging[69] or electron beam induced current,[70] that are able to detect arbitrarily low defect densities, should yield critical thicknesses closer to agreement with the MB model than less sensitive techniques, such as X-Ray Diffraction (XRD), Rutherford Backscattering (RBS), and transmission electron microscopy (TEM), which will only measure the "critical thickness" when a certain threshold defect density or strain relaxation has been exceeded.
3. As the equilibrium critical thickness is ex-

ceeded, the total length of misfit dislocation and therefore the magnitude of strain relaxation will increase with epilayer thickness. The rate of dislocation length generation or strain relaxation will depend upon misfit dislocation nucleation and propagation rates, which in turn will depend upon growth temperature, growth rate, and the stress/strain levels in the film. Thus, the strain relaxation rate will be greater at higher temperatures and strain. The necessary growth or post-growth anneal "time at temperature" (after exceeding the equilibrium critical thickness) for a detectable defect density to be attained will then decrease with increasing temperature and strain. Therefore, the apparent discrepancy between equilibrium and measured critical thickness will also decrease with increasing temperature and strain. Also, interatomic bonds are typically significantly weaker in compound semiconductors than in covalent elemental semiconductors, thereby favoring faster strain relaxation rates in the former materials, and thus an apparent closer agreement to the MB theory. These trends are evident in Figure 13.

Note that even in the Ge_xSi_{1-x}/Si system, where the relaxation kinetics are particularly sluggish, measurements of critical thicknesses for structures that have been either grown or post-growth annealed at relatively high temperature (~900°C) are in reasonable agreement with predictions of the MB theory.[71,72] This is because the very high temperature in these experiments dramatically accelerates the relaxation process, even for moderate excess stresses.

The most widely used model for the kinetics of strain relaxation in semiconductor heterostructures is that of Dodson and Tsao (DT).[73] (An earlier model was proposed by Matthews et al.[74] but does not provide nearly as good a fit to existing data as the DT model.) DT considered the rates of both dislocation nucleation and propagation, and combined expressions for these rates to predict an overall strain relaxation rate as a function of time and temperature. The dislocation propagation velocity was assumed to follow an expression analogous to Equation 5 for bulk semiconductors:

$$v = v_0 \sigma_{ex} e^{-E_v/kT} \tag{14}$$

Bulk values for E_v were initially assumed, although Dodson[75] later considered the possibility that E_v might be stress dependent as the applied stress in the material approached the Peierls stress. The parameter v_0 was obtained essentially by fitting to experimental data. The quantity σ_{ex} is the "effective" or "excess" stress driving defect motion. It is obtained by converting the difference in Matthews-Blakeslee forces, $F_a - F_T$ Equations 10 and 11, to equivalent stresses using Equation 9. This yields for an uncapped epilayer:

$$\begin{aligned}\sigma_{ex} &= \sigma_a - \sigma_T \\ &= \frac{G.\cos\phi}{(1-\nu)}\left[2\varepsilon(1+\nu)\cos\lambda - \frac{b(1-\nu\cos^2\theta)}{4\pi h}\ln\left(\frac{\alpha h}{b}\right)\right]\end{aligned} \tag{15}$$

For a capped epilayer, an expression for σ_T derived from the expression for F_T in Equation 13 should be used. The concept of this excess stress as the driving stress for dislocation motion is of considerable significance for dislocation dynamics in lattice-mismatched films. It is consistent with the Matthews-Blakeslee definition of critical thickness, h_c, in that below h_c the excess stress is less than zero, and no extra dislocation length will be generated (indeed, any existing interfacial dislocations would be expected to shrink).

Dislocation nucleation in the Dodson-Tsao framework is proposed to arise from dislocation multiplication/interaction mechanisms, with the initial defects arising from a background "source" density preexisting in the structure. (This source density can arise from threading dislocations present in the substrate, particulates, incomplete substrate cleaning, and other growth "accidents".) Possible multiplication mechanisms will be discussed in the next section, but we note here that many groups, including ours, have found little or no evidence of multiplicative mechanisms in strained layer epitaxy, particularly in the Ge_xSi_{1-x}/Si system.

The DT model thus predicts the following stages in strain relaxation during growth of increasing thickness of a lattice-mismatched epitaxial layer:

(i). No strain relaxation until the critical thickness is reached.

(ii). After the critical thickness has just been passed, strain relaxation is initially very sluggish. This "incubation period" arises because the excess stress at this stage is relatively low and dislocation multiplication is retarded by the low initial source density.

(iii). As the epilayer thickness increases, the excess stress gets progressively larger. This increases dislocation propagation velocities and thus strain relaxation rates. Also, more and more defects are generated by multiplication as the defect population increases, effectively producing an approximately exponential increase in the number of dislocations. Hence, the strain relaxation rate dramatically increases.

(iv). Eventually, a large fraction of the original lattice mismatch strain is relaxed. The excess stress decreases (because ϵ in Equation 15 has decreased) and, because dislocations are now moving more slowly, they interact less frequently and thus dislocation multiplication slows. The structure now asymptotically approaches its equilibrium strain state, ϵ_0. Note that ϵ_0 is not equal to zero, as may be confirmed by setting the excess stress in Equation 15 to zero. For a given lattice mismatch in a given system, the magnitude of ϵ_0 asymptotically approaches zero as the layer thickness approaches infinity. The different stages of this relaxation process are indicated schematically in Figure 14.

In the original Dodson Tsao formulation, kinetic parameters, such as the prefactor and activation energy for dislocation motion, and the dislocation multiplication rate are determined either from measurements in bulk semiconductors or from fitting to existing experimental data of strained

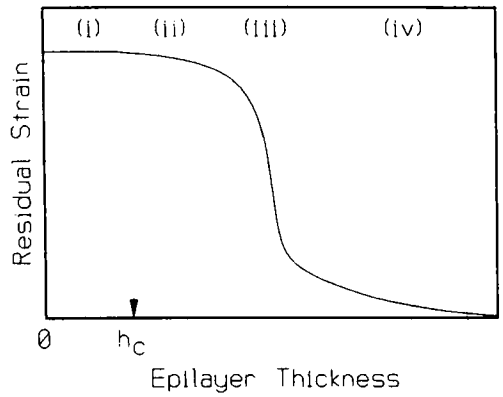


FIGURE 14. Schematic illustration of the different stages of strain relaxation in the Dodson-Tsao model.

layer relaxation. Within this framework, DT were able to demonstrate excellent agreement between their model and existing experimental data in the Ge_xSi_{1-x}/Si system.[73] Two subsequent attempts[67,76–80] to model strain relaxation in the Ge_xSi_{1-x}/Si system have incorporated direct measurements of the necessary kinetic parameters. Hull et al.[76–80] have measured misfit dislocation nucleation, propagation, and interaction rates directly via *in situ* TEM strain relaxation measurements (see Sections VII and VIII). In their model, the total length of misfit dislocation generated at a given time is calculated and then converted to a macroscopic strain relaxation via a simple geometric transformation. L(t), the misfit dislocation length generated at time t, is given by the time integral of the product of the number of growing dislocations, $N(t)_g$, and their velocity, v(t):

$$L(t) = \int N(t)_g v(t)\, dt \qquad (16)$$

For relaxation during crystal growth, the integral is evaluated between the time at which the MB critical thickness is attained and the end of the growth cycle. For post-growth annealing cycles, the integral is evaluated between the start and end of the anneal. The dislocation velocity is constructed via the Dodson-Tsao relation: Equation 14 with measured prefactors and activation energies and a calculated excess stress. The number of growing dislocations is determined by the expression:

$$N(t)_g = N_0 + tN_t - N(t)_i \qquad (17)$$

Here N_0 is a fixed population of preexisting dislocation sources in the structure, N_t is the rate of thermal activation of dislocations (note that the factor t multiplying N_t was inadvertently omitted from Equation 7 in Hull et al.,[78] but was included in all calculations, it is present in the corresponding published Equation 4 in Hull and Bean[80]) and $N(t)_i$ is the loss from the growing defect population due to pinning during dislocation interactions (see Section IX). Of these parameters, N_0 and N_t are measured experimentally and $N(t)_i$ is calculated from the calculable rate of dislocation intersections and an effective "pinning thickness" for the epitaxial layer (see Section IX).[78,81] The strain state of the system was calculated by iterative solution of Equation 16 at time intervals of 1 s. This model is capable of simultaneously calculating both densities and average lengths of the misfit dislocation population (hence, also residual strains and threading dislocation densities), and thus, given the correct experimental input, should be effective at estimating the material quality for device applications. Comparison of the model with experimental data for stages 2, 3, and 4 above are shown in Figure 15.

The above model attempts to calculate the strain state of the system throughout the relaxation process. This is particularly complex in the later stages of relaxation when misfit dislocations start to interact. A later model by Houghton[67,82] models the strain state of the Ge_xSi_{1-x}/Si system during the earlier stages of relaxation, when dislocation interactions are not significant. As pointed out by Houghton, this low-density regime is the regime of interest for practical application of strained layer electronic devices. In Houghton's model, the total dislocation length is given by Equation 16, but the number of growing dislocations is effectively taken to be only the expression tN_t in Equation 17 because dislocation interactions — (the $N(t)_i$ term) — are not) significant in this regime and the N_0 term was effectively taken to be the "source" of the tN_t term via the relation:

$$tN_t = tN_0(\sigma_{ex})^{2.5}e^{-E_n/kT} \qquad (18)$$

Note that in the above equation, we have modified notations in Houghton[67,82] to be consistent with those in this review. The quantity E_n is an activation energy for dislocation nucleation. The parameters in this equation are measured by dislocation counting by optical microscopy of structures that have been post-growth annealed and subjected to a defect revealing etch.[67,82] The quantity σ_{ex} is the Dodson-Tsao excess stress, with the power of 2.5 implying a strong stress

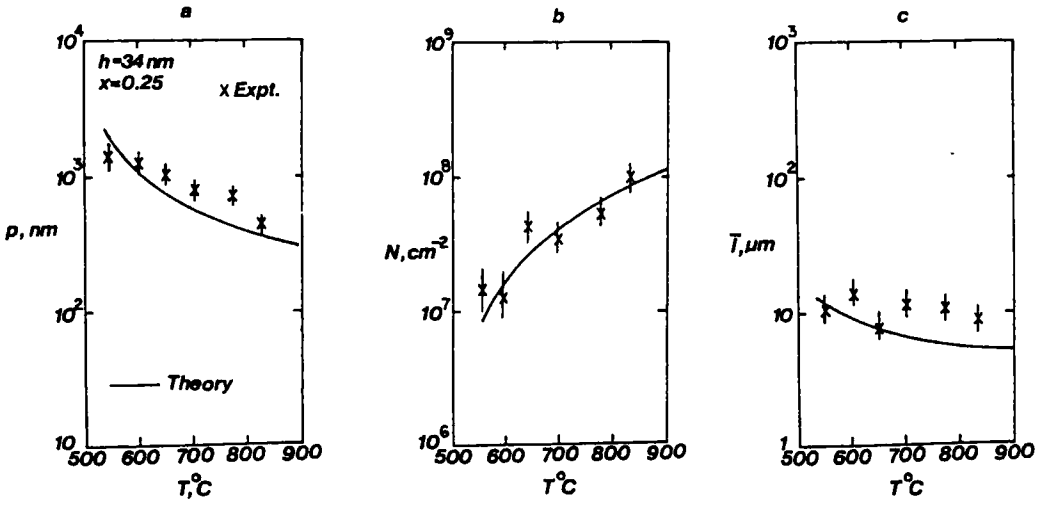


FIGURE 15. Comparison of the predictive strain relaxation model of Hull et al.[78] with experimental data for a 350 Å $Ge_{0.25}Si_{0.75}$ layer annealed for ~4 min at successively higher temperatures for (a) average distance between misfit dislocations, p, (b) areal density of threading dislocations, N, and (c) average misfit dislocation length, l.

dependence of the dislocation nucleation rate. Note that the initial source density, N_0, was itself found to vary quite significantly (by up to two orders of magnitude), generally increasing with the Ge concentration x in Ge_xSi_{1-x}.

The expression used for dislocation velocity in the Houghton model, deduced from experimental measurement of the maximum dislocation lengths following defined annealing cycles, was

$$v = v_0(\sigma_{ex})^2 e^{-E_v/kT} \quad (19)$$

This again implies a stronger stress dependence of the velocity than assumed by the Dodson-Tsao Equation 14. Houghton's model thus predicts that the strain relaxation rate in the noninteracting regime varies as the 4.5 power of excess stress. The relaxation is predicted to be quadratic with time, with an activation energy of ~4.6 eV.[67,82] Houghton was able to demonstrate good agreement between experiment and theory for the initial stages of strain relaxation.

In contrast, the DT model[73] predicts a relaxation rate linear with time, varying as the square of the excess stress and with an overall activation energy ~2.0 eV. The model of Hull et al.[78] predicts a quadratic dependence upon time, a linear dependence upon excess stress in the prefactor, a stress dependence in the activation energy, and a total activation energy varying from 1.4 to 2.5 eV. The apparent quantitative discrepancies between these three models may arise partially from:

1. Differences in the interacting vs. noninteracting regimes.[79] Based upon the formulations, the model of Hull et al. should be better suited to the former, and that of Houghton to the latter.
2. Different assumptions about nucleation mechanisms.
3. Whether stress dependences should exist in the prefactor or within the activation energy. (As discussed in Hull et al.,[79] expansion of a stress-dependent activation energy over a limited temperature range can be approximated by a preexponential power law of the stress. Differences in the stress formulation for dislocation propagation velocities in Hull et al.[79] and Houghton[67] may be reconciled in this fashion.)

Nevertheless, all three models have qualitatively similar features: they all build upon the MB and DT frameworks and upon the central concept of excess stress. In addition, all models appear to converge reasonably well in regimes where they are trying to model the same data (e.g., the measured critical thickness at 550°C in the Ge_xSi_{1-x}/Si(100) system).

A recent, fascinating result suggests strong departures in the $In_xGa_{1-x}As$/GaAs system from the assumptions and models described above (which appear to work well in the Ge_xSi_{1-x}/Si system). Whaley and Cohen[83] have shown, via *in situ* RHEED monitoring of the surface lattice parameter of $In_xGa_{1-x}As$/GaAs, that strain relaxation during and after growth proceeds at very different rates. An example of their results is shown in Figure 16, where it is observed that the relaxation of the epilayer strain as a function of epilayer thickness proceeds at a much slower rate (in fact, it is virtually halted) when growth is stopped but the structure is held at the growth temperature in the MBE chamber. In the framework of the Dodson-Tsao, Hull-Bean, and Houghton models discussed above, annealing (inside or outside of the chamber) of a metastably strained structure should continue to promote strain relaxation. (Note that *ex situ* annealing experiments of metastable structures in the $In_xGa_{1-x}As$/GaAs system *have* shown strain relaxation at temperatures equal to the structure growth temperature.[84] Thus, the Whaley-Cohen results appear to suggest either a difference in UHV vs. non-UHV anneals, a dependence upon the impinging molecular beams, or a dependence upon surface stoichiometry or reconstructions. These different potential mechanisms have not been resolved, but the Whaley-Cohen experiments suggest hitherto unrevealed facets to the strain relaxation process.

VII. DISLOCATION NUCLEATION

In the next few sections we consider separately what is known about the mechanisms by which misfit dislocations nucleate, propagate, and interact.

The precise mechanisms of nucleation of misfit dislocations remain elusive and contro-

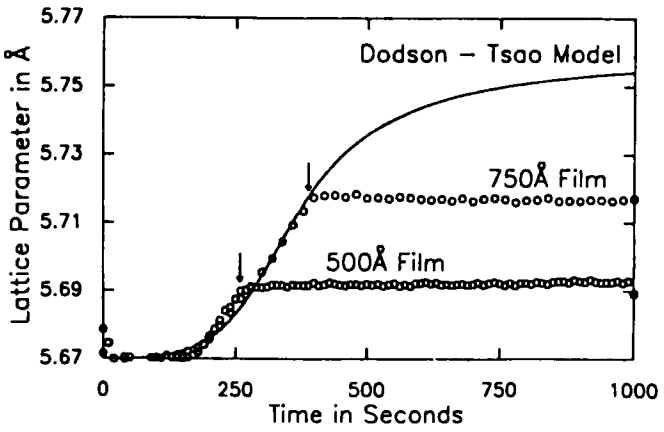


FIGURE 16. Relaxation of the surface lattice parameter of $In_{0.26}Ga_{0.74}As/GaAs(100)$ heterostructures at 450°C. The $In_{0.26}Ga_{0.74}As$ layers are grown at 2 Å · s^{-1} from t = 0 until the times indicated by arrows, at which points to As fluxes are switched off. (From Whaley, G. J. and Cohen, P. I., *Appl. Phys. Lett.*, 57, 144, 1990. With permission.)

versial. The generic candidates for misfit dislocation sources are

1. Multiplication mechanisms arising from dislocation pinning and/or interaction processes.
2. Heterogeneous nucleation at specific local stress concentrations due, for example, to growth artifacts or preexisting substrate defects.
3. Homogeneous or spontaneous nucleation of dislocation loops or half-loops.

A simple geometrical calculation shows that high densities of dislocation nucleation events are required. Consider a 10-cm-diameter (100) substrate upon which an epilayer is grown with a lattice mismatch of 2% with respect to the substrate. For a sufficiently thick epilayer grown at a sufficiently high temperature, the residual strain will be almost zero. If the interfacial misfit dislocations are of the 60° a/2⟨110⟩ type with a Burgers vector of 4 Å, complete strain relaxation requires an interfacial misfit dislocation spacing of ~100 Å. The total length of the required orthogonal interfacial misfit dislocation grid will be of the order 10^6 m, and even if each dislocation makes a chord from edge to edge of the wafer, ~10^7 separate dislocations will be required, or of the order 10^5 per square centimeter of substrate area. This should be regarded as an absolute minimum required source density, and consideration of kinetic effects implies substantially higher required densities. For example, as discussed in the next section, typical misfit dislocation velocities in Si-based heterostructures at strains ~1% and excess stresses in the range 500 MPa to 1 GPa are of the order 1 μm · s^{-1} at 550°C. At a typical molecular beam epitaxy (MBE) growth rate of 1 Å · s^{-1}, it takes 1000 s to grow a 1000 Å epilayer, by which time a strain of 0.01 is typically >10% relaxed. A 10% relaxation of the original strain requires a total interfacial dislocation length of ~10^5 m. The maximum length of individual dislocations, based upon their propagation velocity, is 1 mm (the average will be significantly less than this). Thus, a minimum number of 10^8 or density of ~10^6 cm^{-2} dislocations are required. These numbers again represent a lower limit to the required source density. This is because not all dislocations will nucleate at the start of the relaxation process, propagation velocities will be substantially lower than the number quoted above until the critical thickness is significantly exceeded, and misfit dislocation interactions (see Section IX) will slow propagation.

The original formulations of Matthews and Blakeslee[43,44] often approximated the required density of defect sources via existing threading dislocations in the substrate. This may have been a reasonable assumption for the GaAs substrates of the 1970s that they were considering, but contemporary substrates have defect densities in the ranges ~10 to 10^2 cm^{-2} for Si and ~10^3 to 10^4 cm^{-2} for GaAs and InP. Clearly the substrate itself cannot provide sufficient densities of defect sources.

Other defect sources may be provided by artifacts, accidents, or errors in growth, such as incomplete substrate cleaning, particulates, source "spitting", and contaminants. Each of these problems can certainly provide sufficient source densities if the growth quality is poor enough, but each should be controllable to $<10^3$ cm^{-2}. Of course, each source may emit more than one misfit dislocation, but for finite source dimensions (which for isotropic source dimensions must be less than of the order of the epilayer thickness) generation of more than a certain number of misfit dislocations will be energetically unfavorable, due to local overrelaxation of the strain.

The above growth "errors" constitute the most likely candidates for heterogeneous sources, but other more subtle heterogeneous sources have been observed or postulated. Higgs et al.[85] have provided convincing evidence recently of the role of trace impurities of Cu (at about the 10^{14} cm^{-3} level) in surface half-loop nucleation in strained epilayers. Eaglesham et al.[86] have reported heterogeneous distributions of 1/6⟨114⟩ diamond-shaped stacking faults, with dimensions typically on the order of 1000 Å, which act as sources for classic a/2⟨110⟩ glide dislocations in Ge_xSi_{1-x}/Si structures. No other reports have been made of such "diamond defects" in strained layer epitaxy, however, and it seems possible that the observed defects are due to trace metallic impurities. (For example, similar defects have been observed resulting from precipitation of iron in bulk Si.[87]) Hull and Bean[88] have calculated that the alloy nature of strained epilayers, e.g., in the Ge_xSi_{1-x}/Si and $In_xGa_{1-x}As$/GaAs systems, may play a central role in dislocation nucleation because local (random or nonrandom) fluctuations in the alloy composition may produce localized stress concentrations.

Houghton[67,82] has reported the most complete direct experimental measurements to date of dislocation nucleation densities and sources in the Ge_xSi_{1-x}/Si system, via dislocation counting as a function of time and temperature following post-growth annealing of strained heterostructures. Only the early stages of strain relaxation were studied, where less than about 0.1% of the initial lattice mismatch strain was relaxed. It was reported that the number of observed dislocations followed the trend:

$$N = BN_0\left(\frac{\sigma_{ex}}{G}\right)^{2.5} e^{-E_n/kT} \qquad (20)$$

In this equation, B is a constant of the order $10^{18}s^{-1}$. The activation energy for dislocation nucleation, E_n, is of the order 2.5 eV for Ge_xSi_{1-x}, with $0.0 < x < 0.3$. N_0 is the density of preexisting heterogeneous nucleation sites, reported to be in the range 10^3 to 10^5 cm^{-2} and generally increasing for higher x. This equation implies that, at least in the early stages of relaxation, *all* misfit dislocations are nucleated from heterogeneous sources. As indicated above, this appears to be become geometrically impossible for later stages of strain relaxation.

The only other published report of which we are aware that attempts to directly measure dislocation nucleation activation energies is by our group[77] via *in situ* TEM heating experiments, also in the Ge_xSi_{1-x}/Si system but at a later stage in the relaxation process. Activation energies of the order 0.3 eV were reported for a 350 Å $Ge_{0.25}Si_{0.75}$/Si(100) structure. The prefactor of 2.2×10^6 $cm^{-2}s^{-1}$ resulted in overall nucleation rates of the same order of magnitude as those reported by Houghton at comparable excess stress at a temperature of 650°C (e.g., for a structure of 1000 Å and $x = 0.23$, Houghton reports a nucleation rate ~3×10^4 $cm^{-2}s^{-1}$, whereas for a structure of 350 Å and $x = 0.25$, we reported a nucleation rate ~5×10^4 $cm^{-2}s^{-1}$). In other strained layer structures in the Ge_xSi_{1-x}/Si system, with $0.15 < x < 0.30$, we have measured activation energies in the range 0.3 to 1.1 eV. The activation energies we have measured are thus significantly lower than those reported by Houghton. These differences may be due to dif-

ferent activation mechanisms in the later (which we study) and earlier (which Houghton studies) stages of relaxation, or to possible surface damage/scratching during TEM sample preparation (although we are careful in trying to avoid this).

Preliminary measurements of misfit dislocation nucleation densities have also been reported using laser reflectance measurements of wafer curvature.[89-91] The obvious generic candidate for a source mechanism producing high densities of misfit dislocations is a multiplication mechanism, by analogy to stress experiments in bulk semiconductors.[17] The idea of a regenerative dislocation source goes as far back as the Frank-Read source.[92] Probably the first application of this concept to semiconductor strained layer epitaxy was reported by Hagen and Strunk for growth of Ge on GaAs.[93] Their proposed mechanism is shown in Figure 17. In Figure 17(a), dislocations of equal Burgers vectors, moving on different inclined slip planes along orthogonal interfacial ⟨011⟩ directions at a (100) interface, intersect and cross. At the intersection, the very high energy configuration at the initially sharp right-angle intersection[11] is lowered by formation of two interfacial segments with localized climb, producing rounding near the original intersection, Figure 17(b), thereby reducing the very high radius of curvature and lowering the dislocation configurational energy. Indeed, this geometry is observed almost ubiquitously at intersections of orthogonal dislocations with identical Burgers vectors. Local stresses near the dislocation intersection now modify the local energy balance for the dislocation and one of the rounded segments of dislocation moves toward the surface, Figure 17(c). When the tip intersects the surface, Figure 17(d), a new dislocation segment is formed that can act as a new misfit dislocation. The rounded dislocation segments at the interfacial intersections may continually move to the surface, thereby producing a regenerative dislocation source.

The evidence for this mechanism in the original Hagen-Strunk paper was compelling. Several authors have since invoked this mechanism as a dislocation source in strained layer epitaxy.[94-96] However, the experimental evidence for the Hagen-Strunk source is generally based upon dislocation configurations, as shown schematically in Figure 17(e), where dislocation segments align along ⟨011⟩ directions across the original intersection event. The true "footprint" of the Hagen-Strunk source is as shown in Figure 17(f), where alignment across the intersection does not occur. The configuration of Figure 17(e)

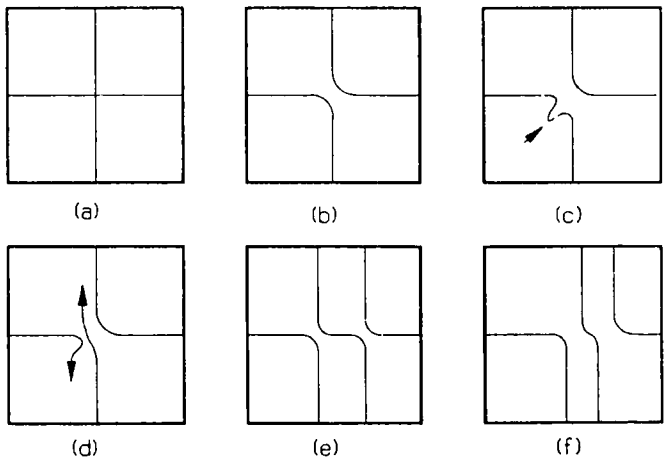


FIGURE 17. Schematic illustration of the Hagen-Strunk mechanism of dislocation multiplication, parts a–d. Part e shows the expected configuration of intersecting dislocations of equal Burgers vector without the Hagen-Strunk mechanism operating, whereas f shows the expected configuration after operation of the Hagen-Strunk mechanism.

is simply that of Figure 17(b), where many orthogonal dislocation intersection events have occurred (see also Dixon and Goodhew[65]). Note that it has been theoretically and experimentally demonstrated by Lefebvre et al.[97] that the Hagen-Strunk source *does not* operate at an order of magnitude-higher stress (than the original Hagen-Strunk experiments) in the $In_xGa_{1-x}As/GaAs$ system (x = 0.08 to 0.18) and, indeed, dislocation tips in stages Figure 17(c) and (d) actually move away from, rather than toward, the surface.

Other proposed dislocation multiplication sources in strained layer epitaxy have been reported. LeGoues et al.[98] reported the operation of a Frank-Read source in graded Ge_xSi_{1-x}/Si structures. A Frank-Read mechanism has also been reported by Lefebvre et al.[99] in $In_xGa_{1-x}As/GaAs$ that operates following interaction of orthogonal 60° dislocations and cross-slip. Washburn and Kvam[100] have proposed a similar mechanism based upon cross-slip and pinning.

The remaining major generic candidate for misfit dislocation sources is homogeneous nucleation of dislocation loops or half-loops. This situation has been modeled by Hirth and Lothe,[11] Matthews et al.,[101] and subsequently by many authors.[86,88,102-104] The general concept behind the energetics of this source is that a growing loop of appropriate Burgers vector relaxes strain energy within the epilayer, E_{strain}, but balancing this is the self-energy of the dislocation loop itself, E_{loop} (which is closely analogous to the Matthews-Blakeslee criterion for growth of interfacial dislocations, as discussed in Section V). Other energy terms that need to be considered are the energies of steps created or removed in the nucleation process, E_{step} (a 60° glide dislocation at a (100) surface, for example, has a Burgers vector component normal to the surface and, thus, must always be associated with step creation/removal), and the energy of any stacking fault created, E_{sf}, either as a result of dislocation dissociation or as a result of separate nucleation of partial dislocations. These different energy contributions lead to a total energy of the form:

$$E_{total} = E_{loop} - E_{strain} \pm E_{step} \pm E_{sf} \quad (21)$$

Note that E_{sf} may either increase or decrease the total energy: a stacking fault clearly increases the system energy, but a partial dislocation could nucleate along the path of an existing partial dislocation. This could remove the preexisting stacking fault, as the combination of the Burgers vectors of the two partials can produce a total dislocation. Such a process has indeed been observed in strained metallic systems by Cherns and Stowell.[105]

In general, the total system energy will pass through a maximum value, δE, at a critical loop radius, R_c. The quantity δE may be regarded as an activation barrier for loop nucleation. The exact magnitude of δE as a function of strain depends upon the elastic parameters used (and hence the system being modeled), the exact form of the expression used for E_{loop} (one important consideration here is the contact angle between the dislocation loop and the free surface[103,104]), the Burgers vector of the dislocation, the magnitude of the dislocation core energy assumed, and whether or not E_{step} and E_{sf} are considered. Variations between these different terms cause a range of predicted activation barriers at a given strain in calculations by different authors,[60,86,88,101-104] but most calculations agree that activation barriers are very low (<1 eV) at strains >4 to 5%, approach a physically attainable limit (say, 5 to 10 eV) at typical growth temperatures for strains of the order 1 to 3%, and are unphysically high ($\gtrsim$100 eV) for operation of this source at strains much below 1%.

A common expression used for the loop self-energy is that due to Bacon and Crocker[106] for a complete loop of radius R:

$$E_{loop} = \frac{Gb^2R}{2(1-\nu)} \times \left[\frac{b_z^2}{b^2} + \left(1 - \frac{\nu}{2}\right)\left(1 - \frac{b_z^2}{b^2}\right)\right] \times \left[\ln\left(\frac{2\pi R\alpha}{b} - 1.758\right)\right] \quad (22)$$

Here, b_z is the component of the dislocation Burgers vector normal to the loop. The original Bacon and Crocker expression[106] contained an extra term due to surface tractions; the magnitude of this extra expression is relatively low and is generally omitted by most authors.

The strain energy relaxed by the loop is given by:

$$E_{strain} = \frac{2\pi R^2 G(1 + \nu)\varepsilon}{(1 - \nu)} \times (b_g \cos\phi \cos\lambda + b_e \cos^2\phi) \quad (23)$$

Here b_g and b_e are the glide and climb components, respectively, of the dislocation Burgers vector.

Considering these two energy terms only, the critical radius is found, setting $\delta E_{total}/\delta R = 0$, to be

$$R_c = \frac{b^2}{8\pi(1 + \nu)\varepsilon} \times \frac{\left[\frac{b_e^2}{b^2} + \left(1 - \frac{\nu}{2}\right)\left(1 - \frac{b_e^2}{b^2}\right)\right]\left[\ln\left(\frac{2\pi R\alpha}{b} - 1.758\right)\right]}{(b_g \cos\phi \cos\lambda + b_e \cos^2\phi)} \quad (24)$$

For simplicity, and to shorten equations, we will now assume a specific dislocation loop configuration, i.e., 60° a/2⟨110⟩ dislocations moving on {111} glide planes with a (100) interface. This yields: $b_z = b_e = 0$, $b_g = b$, $\cos\lambda = 1/2$, $\cos\phi = (2/3)^{1/2}$. Equation 24 now simplifies to:

$$R_c = \frac{b}{8\pi(1 + \nu)\varepsilon} \times \frac{\left(1 - \frac{\nu}{2}\right)\ln\left(\frac{2\pi R\alpha}{b} - 1.758\right)}{6^{-1/2}} \quad (25)$$

Substituting this value for R_c back into the total energy expression yields:

$$\delta E = \frac{Gb^3\sqrt{6}}{32\pi(1 + \nu)(1 - \nu)\varepsilon} \left[\left(1 - \frac{\nu}{2}\right)\ln\left(\frac{2\pi R\alpha}{b} - 1.758\right)\right]^2 \quad (26)$$

Thus, the activation barrier is proportional to b,[3] and will therefore be significantly reduced for lower dislocation Burgers vector magnitude.[60,105] It will also strongly depend upon the core energy parameter, which appears as $(\ln \alpha)^2$. It is inversely proportional to the strain. The corresponding factors for a dislocation half-loop, relative to a full loop, are E_{strain}, halved; E_{loop}, approximately halved (this is a questionable assumption because of the complexity of the general dislocation image construction, but is generally used); E_{total}, approximately halved; R_c, unchanged; and δE, halved. Note that the expressions for critical radii for half and full loops correspond closely to the critical thickness and half the critical thickness, respectively, for capped and uncapped layers in the MB formalism. As mentioned above, the critical radius is very closely analogous to critical thickness, because it denotes that loop dimension at which the system energy decreases with increasing loop size. The small quantitative differences between R_c and h_c for, say, an uncapped epilayer arise because the dislocation configuration is somewhat different, thus modifying the expression for dislocation self-energy and because, generally, R_c is measured in an inclined {111} plane, whereas h_c is measured along the surface normal.

In Figure 18 we show the variation of E_{loop}, E_{strain}, E_{total}, and E_{step} for a dislocation half-loop nucleating at the surface of a $Ge_{0.4}Si_{0.6}$/Si(100) structure. The dislocations are assumed to be of 60° a/2⟨110⟩ type, removing surface steps as they nucleate. (E_{step} is calculated as $E_{step} = 2\,Rb\gamma_s \sin\beta$, where β is the angle between the dislocation Burgers vector and the free surface, and γ_s is the epilayer surface energy, which following Matthews et al.[101] we approximate by $\gamma_s \sim Gb/8$.) The dislocation core energy parameter, α, is taken to be one. Note that the results of Equations 24 through 26 are extremely sensitive to α and, as α is not well known, this leads to considerable uncertainties in predicting δE. We have suggested[88] that the alloy nature of Ge_xSi_{1-x}, for example, can significantly modify (lower) α, and in the same reference we have argued that either random or nonrandom clusters of Ge-rich regions (relative to the Ge_xSi_{1-x} matrix) can locally, substantially enhance ε within the dislocation nucleation volume, thereby lowering δE.

Thus, it appears that homogeneous nucleation can be very effective in producing the required densities of dislocation sources at sufficiently high strains, say of the order 2% or greater. At strains less than about 1%, homogeneous nucleation would be expected to be insignificant. In this regime, the nucleation energetics must be enhanced by preexisting local stress concentra-

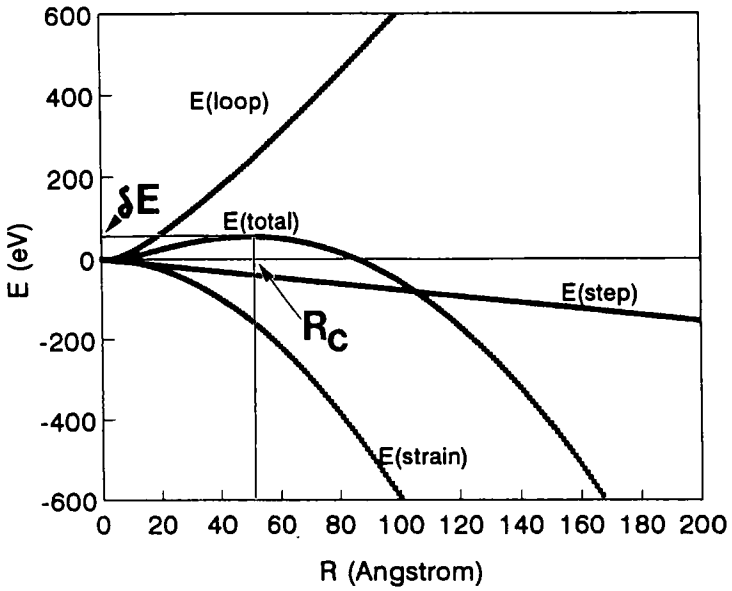


FIGURE 18. Variation of E_{loop}, E_{total}, E_{strain}, and E_{step} for a 60° glide dislocation half-loop annihilating a step at the surface of a $Ge_{0.4}Si_{0.6}$ grown on a S:(100) substrate.

tions, i.e., heterogeneous sources such as particulates, macroscopic surface/interface steps, interfacial oxide, etc. Such sources will only be present in relatively low densities, explaining, at least in part, the substantial discrepancies often observed between measured critical thicknesses and the MB model at low strains (e.g., see Figure 13). Dislocation multiplication mechanisms remain an attractive alternative, and as described above have been proposed by many authors. However, a consensus is still lacking on whether or not multiplication is a ubiquitous or even a significant mechanism.

VIII. DISLOCATION PROPAGATION

We now turn to the question of dislocation propagation kinetics. From Section VI it is apparent that the dislocation velocity is a central parameter in determining overall strain relaxation rates. In the past few years there have been several direct measurements of dislocation velocities in strained layer heterostructures, primarily in the Ge_xSi_{1-x}/Si system, which are complementary to existing measurements in bulk semiconductors.

For the Ge_xSi_{1-x}/Si system, we are aware of four sets of measurements of misfit dislocation velocity: by our group,[76-79] Tuppen and Gibbings,[107] Houghton,[67,82] and Nix et al.[108,109] The first three sets of measurements are primarily on MBE-grown material and the last set primarily on material grown by limited reaction processing (LRP),[110] a variant of chemical vapor deposition (CVD). In the work of Houghton, and Tuppen and Gibbings, dislocation velocities are generally inferred from the maximum lengths of misfit dislocations, as determined by chemical etching and optical microscopy, following post-growth annealing cycles. In the work of Hull and Bean, and Nix et al. misfit dislocation velocities are determined via direct observations during *in situ* annealing inside the TEM.

In Figure 19, we plot measured dislocation velocities by all groups at 550°C (interpolated, where necessary, from linear fits to velocities measured at other temperatures) as a function of excess stress. It is apparent that absolute velocities are in reasonable agreement between the different groups and techniques. (Note, however, that some earlier reports[111-113] included a few dislocation velocity measurements that were found to be inconsistent with later measurements, arising in 111, 112 from uncertainties in layer composition calibrations as detailed in Reference 79. Only measurements in accurately calibrated sam-

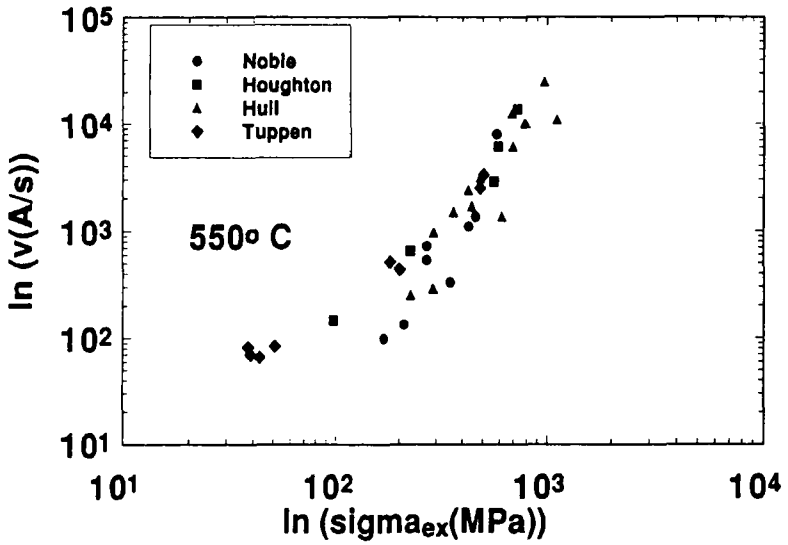


FIGURE 19. Measured misfit dislocation velocities as a function of excess stress in the Ge_xSi_{1-x}/Si(100) system from several different groups and measurement techniques. (From data given in references in the text.[67,76-79,82,107-109])

ples are included in Figure 19.) Initial measurements of velocities in LRP-grown material[108] were also found to be somewhat lower than velocities shown in Figure 19 due to very high oxygen concentrations: $\sim 10^{20}$ cm^{-3} in the original LRP material.[114] The measurements from LRP material in Figure 19 are from material with oxygen concentrations comparable to those typical in MBE material, $\sim 10^{18}$ cm^{-3}.

Despite reasonable quantitative agreement in absolute dislocation velocities, different groups have interpreted their data somewhat differently. Tuppen and Gibbings[107] were able to fit their data by an expression identical to the Dodson-Tsao expression:

$$v = v_0 \sigma_{ex} e^{-E_v/kT} \qquad (27)$$

Here, v_0 was taken to be a similar prefactor to that determined from a survey of experiments on bulk Si and bulk Ge:[17] 2.81×10^{-3} m^2 kg^{-1} s. E_v was found by fitting to the experimental data to vary with Ge concentration in Ge_xSi_{1-x} as $2.156 - 0.7x$ eV, which is in reasonable agreement with bulk activation energies. Tuppen and Gibbings also found a dependence of dislocation velocity upon layer thickness for layers less than a few thousand Angstroms thick. Indeed, this is predicted by the Hirth-Lothe diffusive double kink theory as discussed in Section III, where for dislocation lengths with $L \ll X$ (the average distance between kinks), the dislocation velocity is predicted to be proportional to L. An increase in activation energy is also expected compared to the bulk regime (where $L \gg X$); this was not observed by Tuppen and Gibbings, however.

In our work[79] we have modeled our *in situ* TEM data using the Hirth-Lothe model, but have included a correction due to Seeger and Schiller,[115] which predicts that the glide activation energy will be stress dependent because of the work done by the kink pair during its nucleation against the applied stress field. The magnitude of this effect is relatively insignificant for typical applied stresses (of the order tens to one hundred MPa) used in bulk experiments, but is very significant in the higher stresses of the present experiments, reducing the kink nucleation energy by of the order 0.5 eV at stresses in the range 0.5 GPa. This correction accounts for the apparent anomaly in the Tuppen-Gibbings results where the activation energy was not observed to change from the $L \gg X$ to the $L \ll X$ regimes. We have also postulated that in thin uncapped epilayers, kinks may nucleate singly at the free surface, rather than as pairs along the dislocation line, producing a substantially lower kink nucleation energy and

glide activation energy. This implies a mechanism for increased dislocation propagation velocities and hence strain relaxation rates in uncapped vs. capped epilayers. Indeed, we have measured significantly higher velocities (with enhancements significantly greater than the factor ~2x predicted by the difference in excess stress) in equivalent uncapped vs. capped structures. This is consistent with other reports of increased stability in capped layers.[116,117] Houghton,[67,82] however, did not observe a significant difference between velocities (normalized to the same excess stress) in capped vs. uncapped structures. Such differences, however, might be explained partially by differences in velocity formulations, particularly the squared dependence of velocity upon excess stress in the prefactor of Houghton's formulation. He interpreted his velocity data using the equation:

$$v = v_0\left(\frac{\sigma_{ex}}{G}\right)^2 e^{-E_v/kT} \qquad (28)$$

Here, E_v was taken to be constant over the Ge composition range $0.0 < x < 0.25$ at 2.25 eV, and v_0 was measured to be of the order 10^{13} cm $\cdot s^{-1}$. The substantially greater power of excess stress, $m = 2.0$, in the present equation (as opposed to the linear dependence assumed in the Dodson-Tsao equation and observed in the work of other groups) might be linked to the assumption of a composition-independent activation energy. With reference to Figure 19, the linear regression fit of ln v vs. ln σ to all data does have a gradient of the order 2. However, higher excess stresses generally mean higher Ge concentrations, which other groups and the interpolation of bulk measurements have found to be consistent with lower activation energies. If the data in Figure 19 is normalized to the same activation energy, say the activation energy of Si, and assuming a linear dependence of activation energy upon x and a difference of 0.7 eV between Si and Ge, a gradient of the order 1.3 for the ln v vs. ln σ regression fit is obtained. This is much closer to the predicted Dodson-Tsao linear dependence. Significant differences in the value of M will affect comparisons between velocities in capped and uncapped structures.

It should be stressed that the data discussed previously in this section are essentially for misfit dislocations that are *noninteracting*, that is the dislocation velocities are measured during the early stages of the strain relaxation process where the defect density is relatively low and dislocations typically do not interact. Dislocation velocities that are measured in the later stages of strain relaxation, where defect densities are higher, are characteristic of an *interacting* regime where the defect motion is substantially influenced by the presence of other defects. This may be due to the interaction between the stress fields of adjacent or intersecting defects (see Section IX), or to glide retarding elements (known as *jogs*[11]), which can be produced on the propagating dislocation line as it intersects orthogonal defects. This causes the misfit dislocation velocity, normalized to the excess stress, to become progressively lower (at a fixed temperature) in the later stages of strain relaxation. As greater strain relaxation is, in general, concomitant with higher annealing temperatures, this effectively reduces dislocation velocities at higher temperatures in the interacting regime relative to a noninteracting regime. This produces lower apparent energies as discussed by Hull et al.[79] and illustrated in Figure 20.

There has been far less work on the motion of misfit dislocations in strained layer compound semiconductor heterostructures. Experiments on glide activation energies in bulk compound semiconductors (see George and Rabier[20] for a review) reveal significantly lower activation energies than in Ge and Si. This corresponds to the weaker interatomic bonds and hence lower Peierls barriers in the mixed covalently and ionically bonded compound semiconductors, as opposed to the pure covalent bonding of Ge and Si. For example, the glide activation energy is of the order 1 eV in GaAs and InAs.[22-24] We have measured misfit dislocation velocities in the $In_xGa_{1-x}As/GaAs$ system[84] using our *in situ* TEM technique and typical data, which reveal an activation energy ~1eV and substantially higher velocities normalized to excess stress than in Ge_xSi_{1-x}/Si structures, are shown in Figure 21. Note that modification of the dislocation velocity by the electron beam (which we have been able to rule out at typical beam intensities of ~1 mA $\cdot cm^{-2}$ used in these experiments in the Ge_xSi_{1-x}/Si system[79]) may be significant in the $In_xGa_{1-x}As/GaAs$ system. However, we have observed that overall strain relaxation rates are very similar

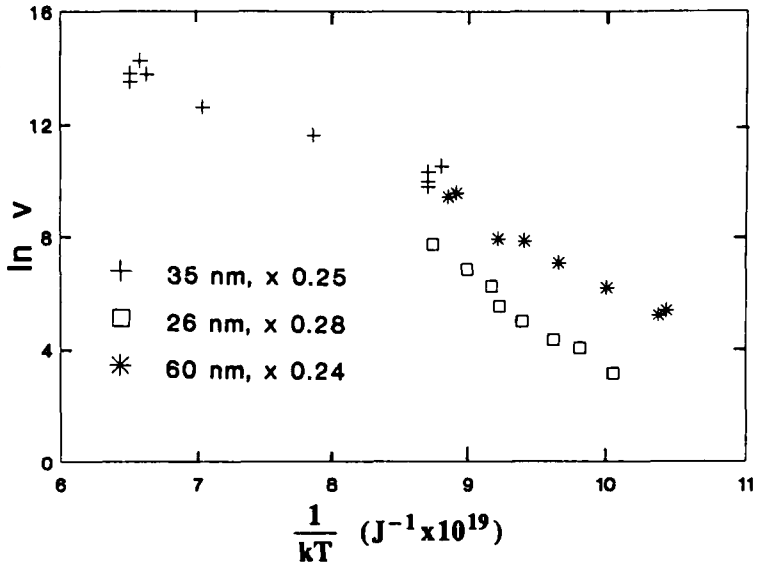


FIGURE 20. Graph of measured dislocation velocities vs. inverse energy for similar uncapped Ge_xSi_{1-x}/Si(100) structures, illustrating how dislocation interactions lower dislocation velocities and apparent activation energies (relative to a noninteracting regime) at higher temperatures.

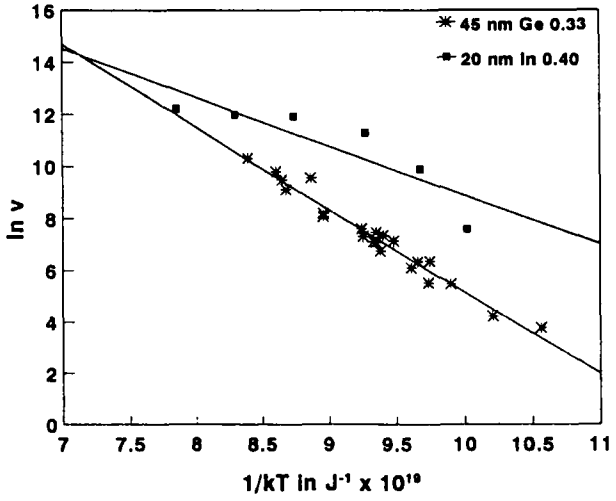


FIGURE 21. Measured dislocation velocities for capped 50 nm $Ge_{0.3}Si_{0.7}$/Si(100) and 20 nm $In_{0.4}Ga_{0.6}As$/GaAs(100) from *in situ* TEM measurements.

with and without electron beam irradiation[84] and thus believe that the beam is *probably* not a dominating effect in the $In_xGa_{1-x}As$/GaAs relaxation experiments.

IX. MISFIT DISLOCATION INTERACTIONS

Interactions between misfit dislocations are generally the dominant mechanisms in the later stages of strain relaxation. As described in Section III, the strain field around dislocations produces a force between two dislocation segments that is inversely proportional to the segment separation and proportional to the dot product of the Burgers vectors of the two dislocations. Hence, this force is maximally repulsive for parallel

Burgers vectors, maximally attractive for antiparallel Burgers vectors, and zero for orthogonal Burgers vectors.

To illustrate the effect of misfit dislocation interactions, consider the case of a propagating dislocation about to intersect a preexisting orthogonal interfacial dislocation, as illustrated in Figure 22. The MB force (discussed in Section V) driving misfit dislocation motion is $F_a - F_T$. For parallel Burgers vectors of the two dislocations there will be a repulsive interdislocation force that will act against the MB force prior to dislocation intersection. If this repulsive force is sufficiently high, it may be larger in magnitude than the MB force, thereby making the net driving force zero and pinning the dislocation. For this to occur, the interdislocation force must be greater than the MB force along the entire threading arm, and this is clearly more likely to occur (because of the 1/r dependence of the interdislocation force) in thinner rather than thicker layers. Thus, as discussed in Hull and Bean,[118] dislocation pinning events are more likely in more highly strained semiconductor combinations, because the critical thickness is lower and dislocation motion occurs in thinner layers. This is one reason why, for a given amount of relaxation, high strain systems have higher threading dislocation densities than low strain systems,[118,119] as more pinning events have occurred in the former. Simple modeling of the interdislocation force suggests that it is effective in pinning parallel Burgers vectors dislocations at epilayer thicknesses up to of the order twice the critical thickness.[112] Note that the analysis is equivalent for antiparallel dislocations, as in this case the interdislocation force acting against the MB force occurs when the propagating dislocation has crossed the orthogonal dislocation and is attracted back toward it. In a particularly complete paper, Freund has analyzed the interdislocation force for all a/2⟨110⟩ glide Burgers vectors configurations at a (100) interface.[81]

Dodson has also modeled dislocation interactions in more complex compound semiconductor structures and compared it with experimental data.[120] The importance of dislocation interactions is also being suggested by laser reflectance measurements of wafer curvature during thermal relaxation of Ge_xSi_{1-x}/Si(100) structures.[91] In these experiments, it is shown that the final strain state attained (i.e., after all measurable relaxation has finished) is strongly dependent upon anneal temperature. Finally, as noted

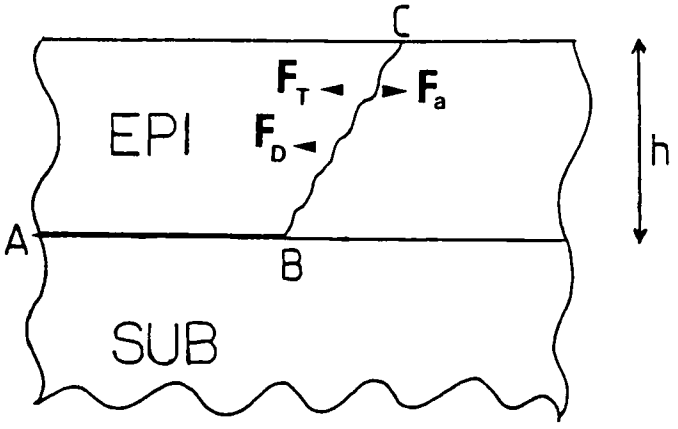


FIGURE 22. Schematic illustration of the forces acting on a propagating threading dislocation (AB) when it encounters a preexisting orthogonal misfit dislocation (D) of parallel Burgers vector. F_a and F_T are the Matthews-Blakeslee lattice-mismatch and line tension forces, respectively. F_D is the horizontal component of the inter dislocation force between D and BC.

in the previous two sections, misfit dislocation interactions can play a central role in dislocation nucleation and propagation processes.

X. DISLOCATION REDUCTION TECHNIQUES, AND LIMITATIONS OF STRAINED LAYER EPITAXY IN DEVICE APPLICATIONS

In this section, we describe and critically review reported techniques for reducing misfit and threading dislocation densities. We also briefly highlight dislocation-induced limitations of strained layer epitaxy in practical electronic device applications.

Semiconductor substrate wafers can now be grown with remarkable structural perfection, e.g., less than 10 dislocations per square centimeter in Si, and of the order 1000 per square centimeter in GaAs. (In-doped GaAs wafers can approach defect densities comparable to Si wafers.) Homoepitaxial growth techniques can, in principle, replicate this substrate defect density. If strained layer epitaxy is to compete with existing technologies, either it must offer unique and overwhelming advantages of homoepitaxy or it must be of sufficient structural perfection to be compatible with, and improve, existing device technologies. The magnitude of defect densities these requirements translate into is somewhat subjective (and controversial), but maximum defect densities of the order 10^5 cm^{-2} in majority carrier devices, 10^3 cm^{-2} for discrete minority carrier devices, and 10 cm^{-2} for Si integrated circuit technology are typically quoted permissible limits.

The best way to avoid misfit and threading dislocations in strained layer epitaxy is to remain below the critical thickness. This, however, places extremely severe limitations on layer thicknesses for most material combinations, and such layers may be too thin compared with growth-induced compositional grading at the interface or too thin to allow sufficient doping transitions between different layers. However, some practical niches for such "mechanically stable" layers have been exploited, such as the $Al_xGa_{1-x}As/GaAs$ system, where the difference in lattice parameters across the interface is extremely small, or heterojunction bipolar transistors (HBTS) in the Ge_xSi_{1-x}/Si system,[121-126] where a fortuitous combination of strain-induced bandgap lowering and a high valence band offset[3,4] have combined to allow usable band gap variation at relatively low x and thus a relatively high critical thickness. An additional restriction to the critical thickness calculated from the lattice-mismatch at room temperature is the question of different thermal expansion coefficients for the two materials: lattice match at room temperature may translate into a significant mismatch at the growth temperature. In terms of modifying the critical thickness of semiconductor systems, however, this is likely to be a second-order effect except at very high growth temperatures or for materials with very different forms of interatomic bonding.

Tolerably low defect levels in strained layer heterostructures may also, in principle, be achieved by "metastable growth", that is, growth beyond the MB equilibrium critical thickness, but at sufficiently low temperatures, strains, and excess stresses that the misfit defect density is still low enough for practical device fabrication. As discussed in Sections VII and VIII, however, post-growth thermal processing during device fabrication will cause further nucleation and propagation of dislocations. To prevent further significant degradation, thermal time-temperature cycles during processing would thus have to remain less than or comparable to the original growth cycle (particularly when one considers that this post-growth annealing effectively occurs at the maximum excess stress at the final growth thickness). This may be an impractical restraint. Other device fabrication processes may be incompatible with metastable strained layers. For example, in Hull et al.[127] and Figure 23 we show how implantation, and subsequent thermal activation, into a strained $Si/Ge_xSi_{1-x}/Si(100)$ heterostructure causes dramatically enhanced strain relaxation rates compared with unimplanted structures. This is believed to be due to condensation of point defects during postimplantation annealing. By analogy to bulk semiconductors, these point defects condense to form small dislocation loops. In bulk semiconductors, these loops shrink eventually and annihilate upon further annealing, but in metastable strained layers they act as ideal misfit dislocation sources.

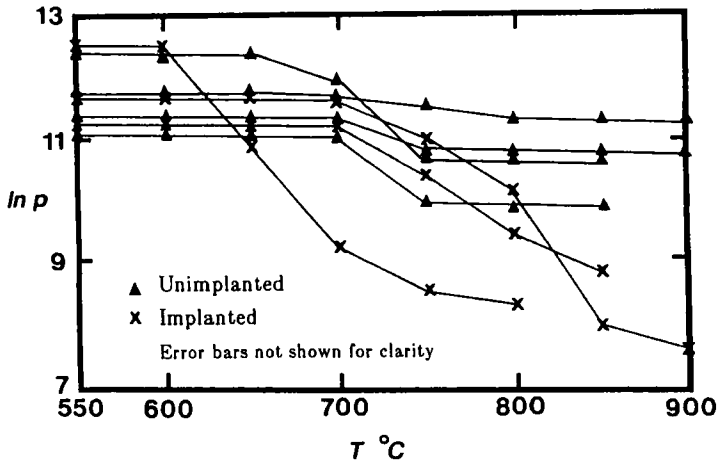


FIGURE 23. Relaxation of a 1000 Å $Ge_{0.2}Si_{0.8}$ layer buried beneath a 3000 Å Si cap. The quantity, p, is the average measured distance between misfit dislocations in Angstroms, and T is the *in situ* annealing temperature. Data points are given for successive anneals of 3 min at each temperature. Triangles correspond to four separate experiments on the unimplanted structure; crosses correspond to three separate experiments on a 1×10^{15} cm $^{-2}$, 100 kV B, implanted structure.

Another crucial concept in strained layer epitaxy is that of the sacrificial *buffer layer* (see Kasper and Schaffler[128] for a review). The basic premise of the buffer layer is that interfacial misfit dislocations need not adversely affect devices formed in the epilayer far from the interface. In this case, the density of threading dislocations will be the important metric for structural quality. A variety of techniques have thus been developed for *filtering* threading dislocations, i.e., to reduce their density to acceptable levels. The dual challenges here are (1) completely relax the strain in the epitaxial layer (so that dislocations do not form and grow during device processing or even device operation), and (2) make the residual threading defect densities as low as possible.

The ideal relaxed epilayer configuration would be if each misfit dislocation segment was sufficiently long to terminate at the edge of the wafer, producing two orthogonal sets of parallel, equally spaced chords across the interfacial ⟨011⟩ directions of a (100) wafer. As discussed in the previous two sections, finite dislocation propagation rates and dislocation interactions act to prevent this. Dislocation propagation can be enhanced either by growing at higher temperatures or by post-growth annealing (the latter may be preferred due to likely morphology problems in strained layer growth at higher temperatures[6,129]). Lateral motion to the edges of a 10-cm wafer, however, is still extremely unlikely due to dislocation interactions. (Recall, also, that in the last stages of strain relaxation the MB force is very low, as the residual strain is also very low. This makes propagation velocities progressively lower and dislocation pinning events progressively more likely.)

Dislocation interactions, however, may have one substantial beneficial property: dislocation annihilation. As illustrated in Figure 24, dislocations of opposite Burgers vectors can attract each other and annihilate, transforming two dislocation loops into one. This process can be extremely effective at high threading dislocation densities, but unfortunately as the defect population decreases, so does the probability of further dislocation interactions and further defect reduction. The annihilation process thus becomes virtually ineffective at lower defect densities. Nevertheless, in films with high-defect densities, both annealing[130] and growing to greater thicknesses is effective in reducing dislocation dens-

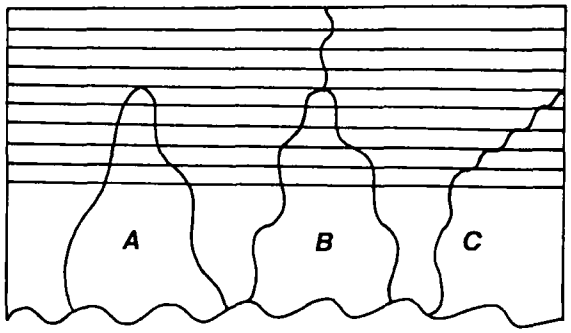


FIGURE 24. Schematic illustrations of possible mechanisms for dislocation threading arm interaction/annihilation events in a strained layer superlattice.

ities. (As will become increasingly clearer, higher temperature and greater epilayer thickness is beneficial in almost all the threading defect reduction schemes discussed in this section.)

In all the previous examples, the sacrificial buffer layer is essentially the near-interface region (i.e., near the original substrate or homoepitaxial buffer layer) of the epilayer. Typical defect densities in abrupt single interface systems are of the order 10^7 to 10^9 cm^{-2} after growth of 1 μm for strains >1%.[128] Lower strain systems may allow perhaps one order of magnitude reduction in this density because of reduced dislocation interactions.[118,119] An extreme example of the buffer layer scheme is in the potentially technologically important GaAs/Si system. For this materials combination, growth of GaAs at its usual homoepitaxy temperature (~600°C) on the Si surface causes extreme clustering and rough surface morphology persisting to layer thicknesses of several microns. This is typically solved by growth of an intermediate lower temperature layer of GaAs at ~400°C and with a thickness ~1000 Å onto the Si surface. The GaAs nucleation density is progressively higher at lower temperatures[131,132] and this allows a continuous epilayer to form at much lower thicknesses. The structural quality of the low-temperature GaAs is, however, very poor. Subsequent GaAs growth is thus performed at the homoepitaxial temperature and, since it is deposited onto a continuous GaAs prelayer, its morphology is good and the higher temperature growth helps anneal out defects in the prelayer. Defect densities as low as 10^7 cm^{-2}, with good surface morphology, after growth of several microns of GaAs have been reported by many groups using this method.[5]

Note that differential thermal expansion coefficients are again significant in the buffer layer geometry, as they may cause substantial defect generation in structures that are fully relaxed at the growth temperature. In particular, this thermal stress may be of different sign to the lattice mismatch stress. Thus, a structure that is in compressive strain during growth, such as Ge_xSi_{1-x}/Si may experience a net tensile strain during cool-down if the lattice mismatch is sufficiently relaxed at the growth temperature. This can cause extended, even infinite, splitting of the partial dislocations, which constitute a dissociated total dislocation, as discussed in Section III, leaving a 90-degree a/6⟨112⟩ partial at a (100) interface. This produces a stacking fault extending throughout the epitaxial layer, as has been observed commonly in GaAs/Si.[5]

A promising technique for minimizing the nonbeneficial effect of threading/misfit dislocation interactions is continuous grading of strain. This is clearly most easily effected in a strained alloy system, such as Ge_xSi_{1-x}/Si or $In_xGa_{1-x}As$/GaAs. Recent reports of compositional grading in the Ge_xSi_{1-x}/Si system[133,134] have demonstrated defect densities as low as 10^6 cm^{-2} in layers 5 to 10 μm thick at total lattice mismatches as high as 2%. This is claimed to be an order of magnitude below that for nongraded layers of the same thickness. The main potential benefit of grading is the ability to separate dislocations vertically (i.e., along the growth direction) through the structure as they adjust to the continuously varying strain field. This gives an extra degree of freedom for dislocations to propagate past each other and thereby minimize pinning events. The vertical distribution of dislocations can also vary during specimen cool-down following growth, minimizing the effects of differential thermal expansion coefficients.

The probability of defect interactions is substantially enhanced by providing definite directions for threading dislocation motion. This is the basis of *strained layer superlattice filtering*, as pioneered by Matthews and Blakeslee.[44] The technique consists of growth of a stack of strained layers on top of the relaxed epilayer. The thickness and strain of each individual layer within

the stack is insufficient to allow nucleation of additional dislocation but is sufficient to deflect threading dislocations into being misfit dislocations at the interfaces. This translates into a criterion that the MB critical thickness be exceeded, but not by too much. For a (100) interface, the interfacial misfit dislocations must run along the interfacial ⟨011⟩ directions. This increases the probability of their meeting, interacting, and annihilating compared with a less correlated motion of threading dislocations within their slip planes in the absence of the superlattice. Many groups have claimed successful application of this technique,[44,135-137] typically reducing threading dislocation densities from $\sim 10^8$ to 10^9 cm^{-2} to of the order 10^6 to 10^7 cm^{-2}. However, analysis of dislocation interaction probabilities in this geometry[138] still suggests that defect densities much below 10^6 cm^{-2} are unlikely to be achieved.

To date, the most promising approach to reducing threading dislocation densities appears to be *patterned or mesa epitaxy*. In this geometry, strained layers are only grown over a finite delimited growth area, typically with dimensions in the range 10 to 100 μm. In principle, the number of heterogeneous dislocation nucleation sites within this reduced growth area may be very low, and this will even reduce interfacial dislocation densities (although this situation will cease when the strain becomes sufficiently high to homogeneously nucleate dislocation loops, particularly at the mesa edges). However, the most important advantage to this technique is in reduction of threading dislocation densities, because the dislocation now has only to propagate a far more limited distance to reach the mesa edge than it would have to reach the wafer edge. The potential of this technique has been illustrated elegantly in the $In_xGa_{1-x}As/GaAs$ system by Fitzgerald et al.[102,139] and Figure 25. Other groups have introduced variants of this technique.[140-143]

The primary disadvantages of the mesa technique are its limited growth area available for device processing and its nonplanar geometry, which may be incompatible with some device-processing steps. We have attempted to address these shortcomings by devising a new substrate patterning technique that allows finite dislocation propagation lengths, connectivity across an entire wafer, and planar geometries.[144] The basic concept is illustrated in Figure 26. A two-dimensional array of oxide dots of height, H, and interfacial dimension, d, is lithographically defined on a Si(100) substrate. The dots are offset from each other with respect to the interfacial ⟨011⟩ directions by an amount, s, where $s<d$. The distance between dots, measured along an interfacial ⟨011⟩ direction, is L. Thus, if a lattice-mismatched epitaxial layer of thickness h, such that $h_c<h<H$, is grown upon the patterned substrate, any dislocation traveling along an ⟨011⟩ direction must intercept a dot within a path length, Λ, given by

$$\Lambda = \frac{L^2}{s} \tag{29}$$

We have defined a series of patterns with a range of d, s, and L upon a single 4″ mask, with a typical set of dimensions being: d = 2 μm, L = 10 μm, and s = 1 μm, giving Λ = 100 μm. We believe that dislocation propagation lengths of this magnitude are easily achievable for growth of moderately mismatched systems (such that dislocation interactions are minimized) at sufficiently high temperatures, or with sufficient post-growth annealing. Preliminary results using this technique[144] show that we have been able to obtain defect densities of the order 10^5 cm^{-2} for Ge_xSi_{1-x} layers grown by MBE with x = 0.15 to 0.20 and thicknesses 2000 to 8000 Å on Si(100) surfaces. The growth temperatures used varied from 550 to 700°C. Growth of similar structures on unpatterned wafers typically yields dislocation densities in the range 10^6 to 10^7 cm^{-2}. Post-growth annealing of these structures to temperatures as high as 900°C causes the structures to almost completely relax, resulting in dramatic enhancements in the interfacial misfit dislocation density, but the threading dislocation density still typically remains below 10^6 cm^{-2}. In Hull et al.,[144] we also describe how we are able to produce planar structures via selective growth using LRP, and how lateral overgrowth over the oxide pillars can produce a continuous single crystal surface.

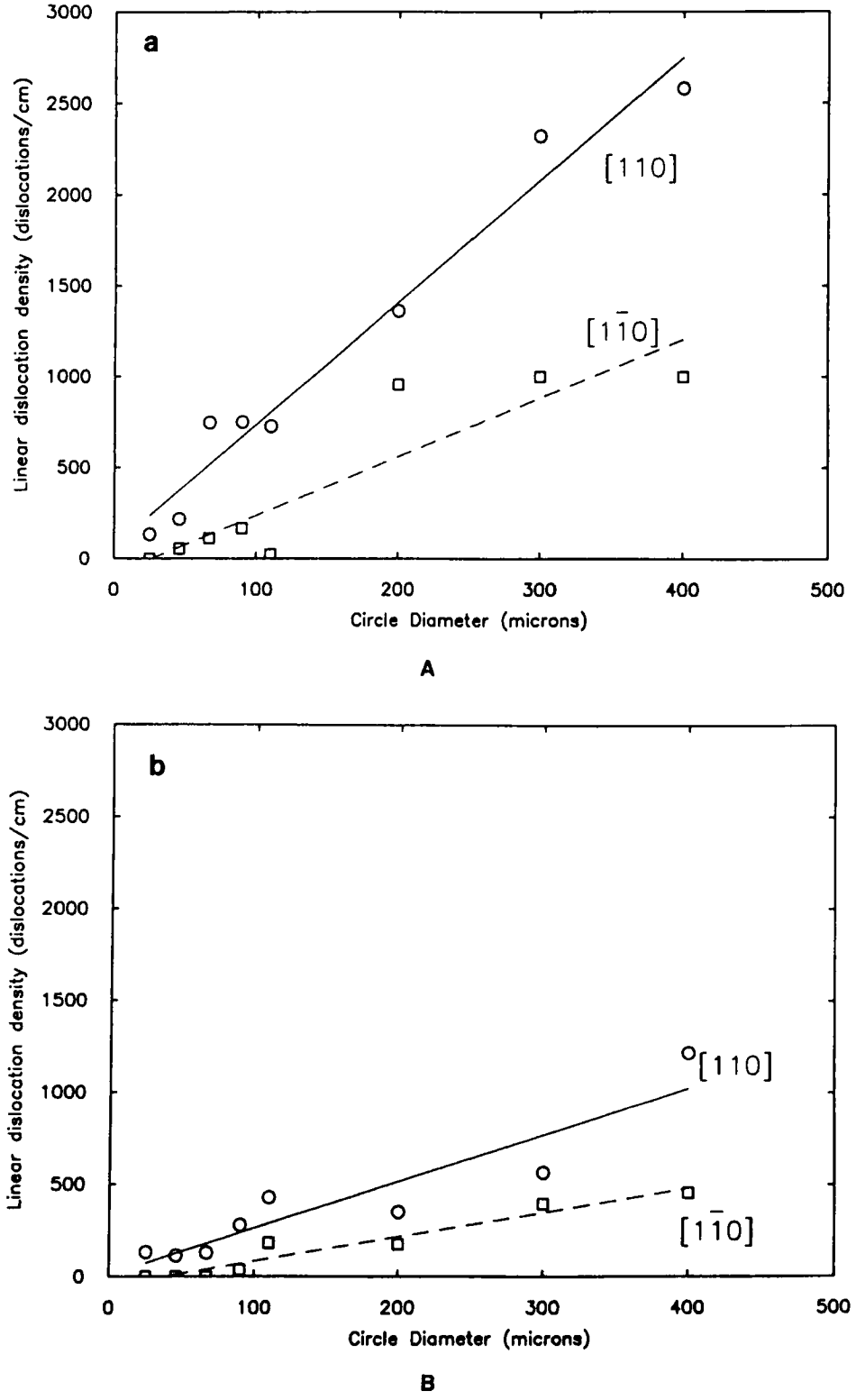


FIGURE 25. Linear interface dislocation density vs. circular mesa diameter for samples with 3500 Å of $In_xGa_{1-x}As$ (x = 0.05) grown onto GaAs substrates with (a) 1.5×10^5 cm^{-2} and (b) 10^4 cm^{-2} preexisting dislocations in the substrate. (From Fitzgerald et al., *J. Appl. Phys.*, 65, 2688, 1989. With permission.)

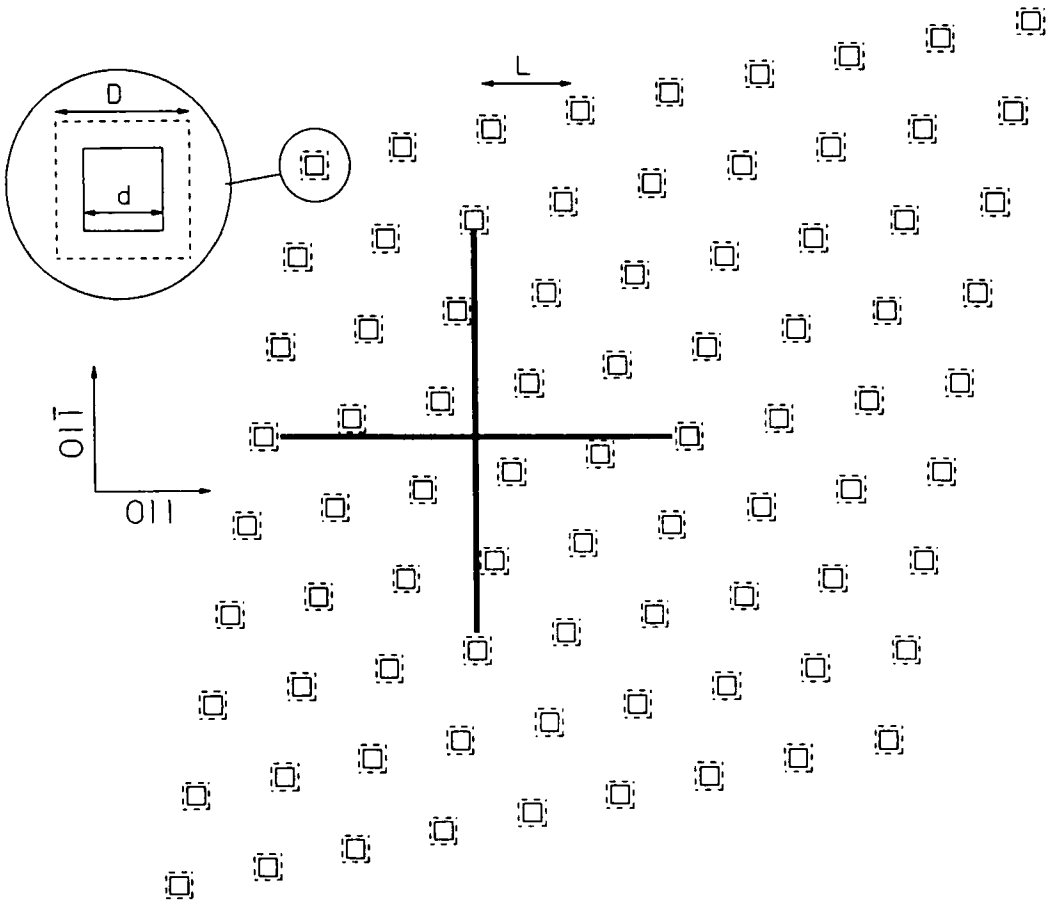


FIGURE 26. Schematic illustration of the oxide pillar patterning arrangement used to reduce threading defect densities on a (100) surface.[144] The dashed squares illustrate the intended size of oxide pillars, whereas the smaller solid squares illustrate their actual size after etching. (The pillar shape actually also becomes approximately circular after etching.) The cross of wide solid lines is intended to schematically represent how dislocations terminate on the pillars. The inset in the top left-hand corner shows an enlargement of one pillar.

XI. SUMMARY

We have attempted to review the central issues in the energetics and kinetics of misfit dislocations in planar lattice-mismatched semiconductor epitaxy. The enduring criterion for the stability of such structures against misfit dislocation formation remains the Matthews-Blakeslee "Mechanical Equilibrium" model.[43,44] This model analyzes the forces acting on a dislocation threading through the epitaxial layer and predicts the critical thickness, h_c, at which it first becomes energetically favorable for the threading dislocation to extend along the lattice-mismatched interface to form a misfit dislocation. The Matthews-Blakeslee model is generally recognized to be conceptually correct and quantitatively accurate to the limit in which certain parameters in the appropriate force equations are known. As argued by Fritz,[133] however, measurement of the true Matthews-Blakeslee critical thickness requires an experimental technique that is able to detect very low misfit dislocation densities or small amounts of strain relaxation. Less-sensitive techniques will generally result in measurement of higher values of critical thickness because of kinetic limitations in producing a sufficiently high misfit dislocation density that is detectable experimentally. In the limit of techniques that are able to detect very low dislocation densities, or

in structures that have undergone prolonged annealing, investigations of h_c have generally validated the Matthews-Blakeslee model.

The kinetics of strain relaxation are generally based upon the Dodson-Tsao model,[73] which introduces the concept of excess stress driving dislocation motion. The major kinetic misfit dislocation processes involved in strain relaxation are nucleation, propagation, and interaction.

The nucleation process is probably the least understood. The major problem is to invoke a source that is energetically plausible and present in sufficient densities. The generic candidates for misfit dislocation sources are homogeneous, or spontaneous nucleation, heterogeneous nucleation (i.e., nucleation associated with a specific type of site or defect in the structure), and multiplication. The latter mechanism was assumed by Dodson and Tsao[73] by analogy to stress experiments in bulk semiconductors.[17] Hagen and Strunk[93] proposed a detailed mechanism for dislocation multiplication in strained layer semiconductors, but this does not appear to be a general mechanism. Other authors, for example, LeGoues et al.[98] and Lefebvre et al.,[99] have reported dislocation multiplication mechanisms in the Ge_xSi_{1-x}/Si and $In_xGa_{1-x}As/GaAs$ systems, respectively, but most groups, including ours, have failed to observe reproducible multiplication processes. Therefore, the generality of dislocation multiplication in strained layer relief must be questioned. Several authors have calculated the energetics of homogeneous nucleation,[86,88,101-104] concluding that strains $\gtrsim 2\%$ are necessary for this process. In the absence of multiplication, nucleation at strains substantially below 2% is likely to be limited by the paucity of heterogeneous sources in high-quality crystal growth. This leads to the concept of nucleation-limited relaxation, which is evident, for example, in the low-temperature Ge_xSi_{1-x}/Si growth in Figure 13.

Several groups have measured misfit dislocation propagation velocities in the Ge_xSi_{1-x}/Si system,[67,79,82,107,108] and their results are in general agreement with each other and with expressions used to describe dislocation motion in bulk Si and Ge. However, some modifications of bulk theory due to the very high stresses in these epitaxial layers[78] and finite dislocation lengths[79,107] have been suggested. Glide activation energies are generally observed to be of the order 2.0 eV, and absolute dislocation velocities are of the order 1 $\mu m \cdot s^{-1}$ at 550°C and excess stresses ~0.5 GPa. Preliminary measurements of velocities in the $In_xGa_{1-x}As/GaAs$ system have also been made.[84]

Dislocation interactions are also of crucial importance in strain relaxation. They are implicated in dislocation nucleation (via multiplication mechanisms), in reducing propagation velocities,[79] and in pinning mechanisms.[81,118] Dislocation interactions are also likely to provide the fundamental limit to which threading dislocations can be eliminated from epitaxial layers, with obvious implications for device applications.

Several techniques for limiting defect (particularly threading dislocation) densities in strained layer epitaxial semiconductor systems have been proposed. The obvious approach is to keep layer thicknesses below the critical thickness, although this will generally result in layers too thin for practical device applications. Metastable layers, grown beyond the Matthews-Blakeslee critical thickness but with growth terminated before substantial dislocation densities have had time to form, in principle, may also be of practical use. They are likely to be very unstable to subsequent device processing, such as contact implantation, however.[127] Strained layer superlattice filtering[44] has been shown to be very effective at reducing threading dislocation densities to moderate levels, but also becomes progressively ineffective as the defect density increases.[138] Compositional grading of strained layers also shows promise in reducing threading dislocation densities,[133] primarily by reducing dislocation interaction and pinning processes. Finally, mesa growth[102] and substrate patterning[144] offer perhaps the greatest promise of reduced threading defect densities due to limited propagation lengths required before termination at a surface or amorphous feature.

In our opinion, the primary remaining challenges in understanding strain relaxation via misfit dislocations are

1. The accurate modeling or measurement of dislocation core energies and Peierls barriers to defect motion.

2. The detailed understanding of dislocation nucleation mechanisms.
3. A successful many-body approach to modeling dislocation interactions.

The Whaley-Cohen results[83] regarding anomalous stability in the $In_xGa_{1-x}As/GaAs$ system must also be understood. The primary challenges to be overcome for successful exploitation of strained layers for general device applications are

1. A scheme for greatly reducing threading dislocation densities, consistent with surface planarity and reasonable epilayer thicknesses.
2. A scheme for controlling thermal mismatch stresses.
3. A structural design consistent with subsequent device processing steps.
4. A detailed understanding of the electrical properties of misfit dislocations and the effect of device operation on structural stability.

ACKNOWLEDGMENTS

We would like to acknowledge many colleagues who have made substantial theoretical and experimental contributions to our own work and the ideas expressed in this review. They are Don Bahnck, Phil Cohen, Brian Dodson, Janet Bonar, David Eaglesham, Gene Fitzgerald, Ben Freund, Marcia Grabow, Derek Houghton, Yong-Fen Hsieh, Eric Kvam, Ron Leibenguth, Dennis Maher, David Noble, Larry Peticolas, Pirouz Pirouz, Frances Ross, Jeff Tsao, and Chris Tuppen.

REFERENCES

1. **Osbourn, G.,** *J. Appl. Phys.*, 53, 1586, 1982.
2. **Osbourn, G.,** *Phys. Rev.*, B27, 5126, 1983.
3. **People, R.,** *Phys. Rev.*, B32, 1405, 1985.
4. **Lang, D. V., People, R., Bean, J. C., and Sergent, A. M.,** *Appl. Phys. Lett.*, 47, 1333, 1985.
5. See papers in, Volumes on Heteroepitaxy on Si, *Mater. Res. Soc. Proc.*, Vols. 67, 91, 116.
6. **Grabow, M. and Gilmer, G.,** *Mater. Res. Soc. Proc.*, 94, 13, 1987.
7. **Eaglesham, D. J. and Cerullo, M.,** *Phys. Rev. Lett.*, 64, 1943, 1990.
8. **Bruinsma, R. and Zangwill, A.,** *Europhys. Lett.*, 4, 729, 1987.
9. **Synder, C. W., Orr, B. G., Kessler, D., and Sander, L. M.,** *Phys. Rev. Lett.*, 66, 3032, 1991.
10. **Ditchek, B. M.,** *J. Cryst. Growth*, 69, 207, 1984.
11. **Hirth, J. P. and Lothe, J.,** *Theory of Dislocations*, McGraw-Hill, New York, 1968.
12. **Bilby, B. A., Bullough, R., and Smith, E.,** *Proc. Soc.*, A 231, 263, 1955.
13. **Nandedkhar, A. S. and Narayan, J.,** *Phil. Mag.*, A61, 873, 1990.
14. **Peierls, R. E.,** *Proc. Phys. Soc.*, 52, 23, 1940.
15. **Alers, G. A. and Thompson, D. O.,** *J. Appl. Phys.*, 32, 283, 1961.
16. **Lothe, J.,** *J. Appl. Phys.*, 33, 2116, 1962.
17. **Alexander, H. and Haasen, P.,** *Solid State Phys.*, Vol. 22, 1968.
18. **Imai, M. and Sumino, K.,** *Philos. Mag.*, A47, 599, 1983.
19. **Patel, J. R. and Chaudhuri, A. R.,** *Phys. Rev.*, 143, 601, 1966.
20. **George, A. and Rabier, J.,** *Rev. Phys. Appl.*, 22, 1941, 1987.
21. **Schaumberg, H.,** *Philos. Mag.*, 25, 1429, 1972.
22. **Osvenski, V. B. and Kholodnyi, L. P.,** *Sov. Phys. Solid State*, 14, 2822, 1973.
23. **Erofeeva, S. A. and Osipyan, Y. A.,** *Sov. Phys. Solid State*, 15, 538, 1973.
24. **Choi, S. K., Mihara, H., and Ninomiya, M.,** *Jpn. J. Appl. Phys.*, 16, 73, 1977; 17, 32, 1978.
25. **Bouchaud, J. P. and Georges, A.,** *Comments Solid State Phys.*, 1991, in press.
26. **Hirsch, P. B., Ourmazd, A., and Pirouz, P.,** *Inst. Phys. Conf. Ser.*, ser. 60, p. 29, 1981.
27. **Maeda, K. and Yamashita, Y.,** *Inst. Phys. Conf. Ser.*, ser. 104, p. 269, 1989.
28. **Nititenho, V. I., Farber, B. Y., and Iunin, Y. L.,** *Sov. Phys. JETP*, 66, 738, 1988.
29. **Moller, H.-J.,** *Acta Metall.*, 26, 963, 1978.
30. **Hirsch, P. B.,** *J. Phys. Paris Colloq.*, 44, C3-149, 1981.
31. **Hirsch, P. B.,** *Mater. Sci. Res.*, 18, 1, 1985.
32. **Gibbings, C. J., Tuppen, C. G., and Higgs, V.,** *Mater. Res. Soc. Proc.*, 220, 205, 1991.
33. **Gomez, A., Cockayne, D. J. H., Hirsch, P. B., and Vitek, V.,** *Philos. Mag.*, 31, 105, 1975.
34. **Cockayne, D. J. H. and Hons, A.,** *J. Phys. Paris Colloq.*, 40, C6, 1979.
35. **Bourret, A. and Desseaux, J.,** *Philos. Mag.*, A39, 405, 1979.
36. **Heggie, M. and Jones, R.,** *Inst. Phys. Conf. Ser.*, ser. 87, 367, 1987.
37. **Hull, R., Bean, J. C., Bonar, J. M., and Peticolas, L. J.,** Proc. Int. Conf. Microscopy of Semiconducting Materials, Oxford, England, March 1991, Institute of Physics, Bristol, England, 1991, p. 497.
38. **Hull, R., Bean, J. C., Peticolas, L. J., and Bahnck, D.,** *Appl. Phys. Lett.*, 9, 964, 1991.

39. **Hull, R., Bean, J. C., Peticolas, L. J., Xie, Y. H., and Hsieh, Y. F.,** *Mater. Res. Soc. Proc.*, 220, 153, 1991.
40. **Mitchell, T. E. and Unal, O.,** *J. Electron. Mater.*, 20, 723, 1991.
41. **Frank, F. C. and van der Merwe, J. H.,** *Proc. R. Soc.*, A198, 205, 1949; A198, 216, 1949; A200, 125, 1949.
42. **Van der Merwe, J. H. and Ball, C. A. B.,** in *Epitaxial Growth*, Part b, Matthews, J. W., Ed., Academic Press, New York, 1975, 493–528.
43. **Matthews, J. W. and Blakeslee, A. E.,** *J. Cryst. Growth*, 27, 118, 1974; 32, 265, 1976.
44. **Matthews, J. W.,** *J. Vac. Sci. Technol.*, 12, 126, 1975.
45. **Schmid, E.,** *Z. Elektrochem.*, 37, 447, 1931.
46. **Chidambarrao, D., Srinivasan, G. R., Cunningham, B., and Murthy, C. S.,** *Appl. Phys. Lett.*, 57, 1001, 1990.
47. **Ball, C. A. B.,** *Phys. Status Solidi*, 42, 357, 1990.
48. **Jesser, W. A. and Kuhlmann-Wilsdorf, D.,** *Phys. Status Solidi*, 19, 95, 1967.
49. **Markov, I. and Milchev, A.,** *Surf. Sci.*, 36, 519, 1984.
50. **Ball, C. A. and Laird, C.,** *Thin Solid Films*, 41, 9, 1977.
51. **Willis, J. R., Jain, S. C., and Bullough, R.,** *Philos. Mag.*, A62, 115, 1990; A64, 629, 1991.
52. **Cammarata, R. C. and Sieradzki, K.,** *Appl. Phys. Lett.*, 55, 1197, 1989.
53. **Fox, B. A. and Jesser, W. A.,** *J. Appl. Phys.*, 68, 2801, 1990.
54. **Basson, J. H. and Booyens, H.,** *Phys. Status Solidi*, 64, 777, 1981.
55. **People, R. and Bean, J. C.,** *Appl. Phys. Lett.*, 47, 322, 1985.
56. **Tsao, J. Y. and Dodson, B. W.,** *Appl. Phys. Lett.*, 53, 848, 1988.
57. **Twigg, M. E.,** *J. Appl. Phys.*, 68, 5109, 1990.
58. **Thompson, N.,** *Proc. Phys. Soc.*, 66B, 481, 1953.
59. **Maree, P. M. J., Barbour, J. C., van der Veen, J. F., Kavanagh, K. L., Bulle-Lieuwma, C. W. T., and Viegers, M. P. A.,** *J. Appl. Phys.*, 62, 4413, 1987.
60. **Wegscheider, W., Eberl, K., Menczigar, U., and Abstreiter, G.,** *Appl. Phys. Lett.*, 57, 875, 1990.
61. **Hwang, D. M., Bhatt, R., Schwarz, S. A., and Chen, C. Y.,** *Phys. Rev. Lett.*, 66, 739, 1991.
62. **Bean, J. C., Feldman, L. C., Fiory, A. T., Nakahara, S., and Robinson, I. K.,** *J. Vac. Sci. Technol.*, A2, 436, 1984.
63. **Kasper, E., Herzog, H.-J., and Kibbel, H.,** *Appl. Phys.*, 8, 199, 1975.
64. **Fritz, I. J., Picreaux, S. T., Dawson, L. R., Drummond, T. J., Laidig, W. D., and Anderson, N. G.,** *Appl. Phys. Lett.*, 46, 967, 1985.
65. **Dixon, R. H. and Goodhew, P. J.,** *J. Appl. Phys.*, 68, 3163, 1990.
66. **Grundmann, M., Dienert, U., Bimberg, D., Fischer-Colbrie, A., and Miller, J. N.,** *Appl. Phys. Lett.*, 55, 1765, 1989.
67. **Houghton, D. C.,** *J. Appl. Phys.*, 70, 2136, 1991.
68. **Fritz, I. J.,** *Appl. Phys. Lett.*, 51, 1080, 1987.
69. **Gourley, P. L., Fritz, I. J., and Dawson, L. R.,** *Appl. Phys. Lett.*, 52, 377, 1988.
70. **Kohama, Y., Fukuda, Y., and Seki, M.,** *Appl. Phys. Lett.*, 52, 380, 1988.
71. **Houghton, D. C., Gibbings, C. J., Tuppen, C. G., Lyons, M. H., and Halliwell, M. A. G.,** *Appl. Phys. Lett.*, 56, 460, 1990.
72. **Green, M. L., Weir, B. E., Brasen, D., Hsieh, Y. F., Higashi, G., Feygenson, A., Feldman, L. C., and Headrick, R. L.,** *J. Appl. Phys.*, 69, 745, 1991.
73. **Dodson, B. W. and Tsao, J. Y.,** *Appl. Phys. Lett.*, 51, 1325, 1987.
74. **Matthews, J. W., Mader, S., and Light, T. B.,** *J. Appl. Phys.*, 41, 3800, 1970.
75. **Dodson, B. W.,** *Phys. Rev.*, B38, 12383, 1988.
76. **Hull, R., Bean, J. C., Werder, D. J. and Leibenguth, R. E.,** *Appl. Phys. Lett.*, 52, 1605, 1988.
77. **Hull, R., Bean, J. C., Werder, D. J., and Leibenguth, R. E.,** *Phys. Rev.*, B40, 1681, 1989.
78. **Hull, R., Bean, J. C., and Buescher, C.,** *J. Appl. Phys.*, 66, 5837, 1989.
79. **Hull, R., Bean, J. C., Bahnck, D., Peticolas, L. J., Short, K. T., and Unterwald, F. C.,** *J. Appl. Phys.*, 70, 2052, 1991.
80. **Hull, R. and Bean, J. C.,** *Mater. Res. Soc. Proc.*, 148, 309, 1989.
81. **Freund, L. B.,** *J. Appl. Phys.*, 68, 2073, 1990.
82. **Houghton, D. C.,** *Appl. Phys. Lett.*, 57, 1434/2124, 1990.
83. **Whaley, G. J. and Cohen, P. I.,** *Appl. Phys. Lett.*, 57, 144, 1990.
84. **Bonar, J. M., Hull, R., Malik, R. J., Ryan, R. W., and Walker, J. F.,** *Mater. Res. Soc. Proc.*, 160, 117, 1990.
85. **Higgs, V., Kightley, P., Augustus, P. D., and Goodhew, P. J.,** *Appl. Phys. Lett.*, 57, 829, 1991.
86. **Eaglesham, D. J., Kvam, E. P., Maher, D. M., Humphreys, C. J., and Bean, J. C.,** *Philos. Mag.*, A59, 1059, 1989.
87. **DeCoteau, M. D., Wilshaw, P. R., and Falster, R.,** *Solid State Phen.*, 19–20, 27, 1991.
88. **Hull, R. and Bean, J. C.,** *J. Vac. Sci. Tech.*, A7, 2580, 1989.
89. **Gillard, V. T., Nobel, D. B., and Nix, W. D.,** *Mater. Res. Soc. Proc.*, Vol. 239, 395, 1992.
90. **Volkert, C. A., Hull, R., Fitzgerald, E. A., and Xie, Y. H.,** paper presented at the 1991 Fall Materials Research Society Meeting, Symp. D.
91. **Volkert, C. A., Fitzgerald, E. A., Hull, R., Xie, Y. H., and Mii, Y. J.,** *J. Electron. Mater.*, 20, 833, 1991.

92. **Frank, F. C. and Read, W. T.,** in *Symposium on Plastic Deformation of Crystalline Solids*, Carnegie Institute of Technology, Pittsburgh, 1950, 44.
93. **Hagen, W. and Strunk, H.,** *Appl. Phys.*, 17, 85, 1978.
94. **Rajan, K. and Denhoff, M.,** *J. Appl. Phys.*, 62, 1710, 1987.
95. **Kvam, E. P., Eaglesham, D. J., Maher, D. M., Humphreys, C. J., Bean, J. C., et al.,** *Mater. Res. Soc. Proc.*, 104, 623, 1988.
96. **Chang, K. H., Berger, P. R., Gibala, R., Bhattacharya, P. K., Singh, J., et al.,** in *Dislocations and Interfaces in Semiconductors*, Proc. of TMS/AIME Symp. TMS, Warrendale, PA, 1988, 157.
97. **Lefebvre, A., Herbeaux, C., and Di Persio, J.,** *Philos. Mag.*, A63, 471, 1991.
98. **LeGoues, F. K., Myerson, B. S., and Morar, J. F.,** *Phys. Rev. Lett.*, 66, 2903, 1991.
99. **Lefebvre, A., Herbeaux, C., Boillet, C., and Di Persio, J.,** *Philos. Mag. Letter.*, 63, 23, 1991.
100. **Washburn, J., and Kvam, E. P.,** *Appl. Phys. Lett.*, 57, 1637, 1991.
101. **Matthews, J. W., Blakeslee, A. E., and Mader, S.,** *Thin Solid Films*, 33, 253, 1976.
102. **Fitzgerald, E. A., Watson, G. P., Proano, R. E., Ast, D. G., Kirchner, P. D., Pettit, G. D., and Woodall, J. M.,** *J. Appl. Phys.*, 65, 2688, 1989.
103. **Kamat, S. V. and Hirth, J. P.,** *J. Appl. Phys.*, 67, 6844, 1990.
104. **Dregia, S. A. and Hirth, J. P.,** *J. Appl. Phys.*, 69, 2169, 1991.
105. **Cherns, D. and Stowell, M. J.,** *Thin Solid Films*, 29, 107, 1975; 29, 127, 1975; 37, 249, 1976.
106. **Bacon, D. J. and Crocker, A. G.,** *Philos. Mag.*, 12, 195, 1965.
107. **Tuppen, C. G. and Gibbings, C. J.,** *J. Appl. Phys.*, 68, 1526, 1990.
108. **Nix, W. A., Noble, D. B., and Turlo, J. F.,** *Mater. Res. Soc. Proc.*, 188, 315, 1990.
109. **Noble, D.,** unpublished results.
110. **Gronet, C. M., King, C. A., Opyd, W., Gibbons, J. F., Wilson, S. D., and Hull, R.,** *J. Appl. Phys.*, 61, 2047, 1987.
111. **Hull, R. and Bean, J. C.,** *Appl. Phys. Lett.*, 54, 925, 1989.
112. **Hull, R., Bean, J. C., Eaglesham, D. J., Bonar, J. M., and Buescher, C.,** *Thin Solid Films*, 183, 117, 1989.
113. **Houghton, D. C., Timbrell, P. Y., and Baribeau, J. M.,** *Mater. Res. Soc. Proc.*, 160, 77, 1990.
114. **Hull, R., Bean, J. C., Noble, D., Hoyt, J. L., and Gibbons, J. F.,** *Appl. Phys. Lett.*, 59, 1585, 1991.
115. **Seeger, A. and Schiller, P.,** *Acta Metall.*, 10, 348, 1962.
116. **Noble, D. B., Hoyt, J. L., Gibbons, J. F., Scott, M. P., Laderman, S. S., Rosner, S. J., and Kamins, T. I.,** *Appl. Phys. Lett.*, 55, 1978, 1989.
117. **Scott, M. P., Laderman, S. S., Kamins, T. I., Rosner, S. J., Nauka, K., Noble, D. B., Hoyt, J. L., King, C. A., Gronet, C. M., and Gibbons, J. F.,** *Mater. Res. Soc. Proc.*, 130, 179, 1989.
118. **Hull, R. and Bean, J. C.,** *Appl. Phys. Lett.*, 54, 925, 1989.
119. **Kvam, E. P.,** *Philos. Mag. Lett.*, 62, 167, 1990.
120. **Dodson, B. W.,** *Appl. Phys. Lett.*, 53, 37, 1988.
121. **Tatsumi, T., Hiriyama, H., and Aizaki, N.,** *Appl. Phys. Lett.*, 52, 895, 1988.
122. **Temkin, H., Bean, J. C., Antreasyan, A., and Leibenguth, R.,** *Appl. Phys. Lett.*, 52, 1089, 1988.
123. **Patton, G. L., Iyer, S. S., Delage, S. L., Tiwari, S., and Stork, J. M. C.,** *IEEE Electron. Dev. Lett.*, 9, 165, 1988.
124. **King, C. A., Hoyt, J. L., Gronet, C. M., Gibbons, J. F., Scott, M. P., Rosner, S. J., Reid, G., Laderman, S., Nauka, K., and Kamins, T. I.,** IEEE 46th Annual Device Research Conf., p. VB-7, 1988.
125. **Xu, D.-X., Shen, G.-D., Willander, M., Ni, W.-X., and Hansson, G. V.,** *Appl. Phys. Lett.*, 52, 2239, 1988.
126. **Patton, G. L., Comfort, J. H., Meyerson, B. S., Crabbe, E. F., Scilla, G. J., DeFresart, E., Stork, J. M. C., Sun, J. Y. C., Harame, D. L., and Burghartz, J. M.,** *IEEE Electron. Dev. Letter.*, 11, 171, 1990.
127. **Hull, R., Bean, J. C., Bonar, J. M., Higashi, G. S., Short, K. T., Temkin, H., and White, A. E.,** *Appl. Phys. Lett.*, 56, 2445, 1990.
128. **Kasper, E. and Schaffler, F.,** Strained Layer Superlattices: Materials Science and Technology, in *Semiconductors and Semimetals*, Vol. 33, Pearsall, T. P., Ed., Academic Press, San Diego, CA, 1991, 240–244.
129. **Bean, J. C., Sheng, T. T., Feldman, L. C., Fiory, A. T., and Lynch, R. T.,** *Appl. Phys. Lett.*, 44, 102, 1984.
130. **Chand, N., People, R., Baiocchi, F. A., Wecht, K. W., and Cho, A. Y.,** *Appl. Phys. Lett.*, 49, 81, 1986.
131. **Biegelsen, D. K., Ponce, F. A., Smith, A. J., and Tramontana, J. C.,** *J. Appl. Phys.*, 61, 1856, 1987.
132. **Biegelsen, D. K., Ponce, F. A., Krusor, B. S., Tramontana, J. C., Yingling, R. D., Bringans, R. D., and Fenner, D. B.,** in *Mater. Res. Soc. Proc.*, 116, 33, 1988.
133. **Fitzgerald, E. A., Xie, Y. H., Green, M. L., Brasen, D., Kortan, A. R., Michel, J., Mie, Y. J., and Weir, B. E.,** *Appl. Phys. Lett.*, 59, 811, 1991.
134. **Tuppen, C. G., Gibbings, C. J., and Hockly, M.,** *Mater. Res. Soc. Proc.*, 220, 205, 1991.
135. **Lilliental-Weber, Z., Weber, E. R., Washburn, J., Liu, T. Y., and Kroemer, H.,** in *Mater. Res. Soc. Proc.*, 91, 91, 1987.
136. **Dupuis, R. D., Bean, J. C., Brown, J. M., Macrander, A. T., Miller, R. C., and Hopkins, L. C.,** *J. Electron. Mater.*, 16, 69, 1986.

137. **Olsen, G. H., Abrahams, M. S., Buiocchi, G. J., and Zamerowski, T. J.,** *J. Appl. Phys.*, 46, 1643, 1975.

138. **Hull, R., Bean, J. C., Leibenguth, R. E., and Werder, D. J.,** *J. Appl. Phys.*, 65, 4723, 1989.

139. **Fitzgerald, E. A., Kirchner, P. D., Proano, R., Petit, G. D., Woodall, J. M., and Ast, D. G.,** *Appl. Phys. Lett.*, 52, 1496, 1988.

140. **Guha, S., Madhukar, A., Kaviani, K., and Kapre, R.,** *J. Vac. Sci. Technol.*, B8, 149, 1990.

141. **Guha, S., Madhukar, A., and Chen, L.,** *Appl. Phys. Lett.*, 56, 2304, 1990.

142. **Matyi, R. J., Shichijo, H., and Tsai, H. L.,** *J. Vac. Sci. Technol.*, B6, 699, 1988.

143. **Lee, H. P., Huang, Y.-H., Liu, X., Lin, H., Smith, J. S., Weber, E. R., Yu, P., Wang, S., and Lilliental-Weber, Z.,** *Mater. Res. Soc. Proc.*, 116, 219, 1988.

144. **Hull, R., Bean, J. C., Peticolas, L. C., Higashi, G. S., Green, M. L., Bahnck, D., and Brasen, D.,** *Appl. Phys. Lett.*, 60, 1468, 1992.